Statistics of baryon correlation functions in lattice QCD

Michael L. Wagman,[1,2] and Martin J. Savage[1]

(NPLQCD Collaboration)

[1]*Institute for Nuclear Theory, Box 351550, Seattle, Washington 98195-1550, USA*
[2]*Department of Physics, University of Washington, Box 351560, Seattle, Washington 98195, USA*

A systematic analysis of the structure of single-baryon correlation functions calculated with lattice QCD is performed, with a particular focus on characterizing the structure of the noise associated with quantum fluctuations. The signal-to-noise problem in these correlation functions is shown, as long suspected, to result from a sign problem. The log-magnitude and complex phase are found to be approximately described by normal and wrapped normal distributions respectively. Properties of circular statistics are used to understand the emergence of a large time noise region where standard energy measurements are unreliable. Power-law tails in the distribution of baryon correlation functions, associated with stable distributions and "Lévy flights," are found to play a central role in their time evolution. A new method of analyzing correlation functions is considered for which the signal-to-noise ratio of energy measurements is constant, rather than exponentially degrading, with increasing source-sink separation time. This new method includes an additional systematic uncertainty that can be removed by performing an extrapolation, and the signal-to-noise problem reemerges in the statistics of this extrapolation. It is demonstrated that this new method allows accurate results for the nucleon mass to be extracted from the large-time noise region inaccessible to standard methods. The observations presented here are expected to apply to quantum Monte Carlo calculations more generally. Similar methods to those introduced here may lead to practical improvements in analysis of noisier systems.

I. INTRODUCTION

Modern nuclear physics research relies upon large-scale high-performance computing (HPC) to predict the properties of a diverse array of many-body systems, ranging from the properties of hadrons computed from the dynamics of quarks and gluons to the form of gravitational waves emitted from inspiraling binary neutron star systems. In many cases, the entangled quantum nature of these systems and the nonlinear dynamics that define them preclude analytic calculation of their properties. In these cases, precise numerical evaluations of high-dimensional integrations that systematically approach the quantum path integral are required. Typically, it is average quantities that are determined by Monte Carlo (MC) path integral evaluations. These average values are to be used subsequently in direct comparison with experiment, as input to analytic frameworks with outputs that can then be compared with experiment, or as predictions for critical components of systems that are inaccessible to experiment such as the equation of state of dense matter in explosive astrophysical environments. Enormous amounts of HPC resources are used in such MC calculations to determine average values of quantities and their uncertainties. The central limit theorem and, in particular, the $1/\sqrt{N}$ scaling anticipated for the uncertainties associated with average values are used to make estimates of projected resource requirements. When a system has a "sign problem," for which the average value of a quantity of interest results from cancellations of (relatively) large contributions, such as those found when averaging the phase $e^{i\theta}$ over θ, the HPC resources required for accurate numerical estimates of the average(s) are prohibitively large. This is the case for numerical evaluations of the path integrals describing strongly interacting systems with even a modest nonzero net baryon number.

While the quantum fluctuations (noise) of many-body systems contain a wealth of information beyond average values, only a relatively small amount of attention has been paid to refining calculations based upon the structure of the noise. This statement, of course, does not do justice to the fact that all observables (S-matrix elements) in quantum field theory calculations can be determined from vacuum expectation values of products of quantum fields. However, in numerical calculations, it is generally the case that noise is treated as a nuisance, something to reduce as much as needed, as opposed to a feature that may reveal aspects of systems that are obscured through distribution among many expectation values. In the area of lattice quantum chromodynamics (LQCD), which is the numerical technique used to evaluate the quantum path integral associated with quantum chromodynamics (QCD) that defines the dynamics of quarks and gluons, limited progress has been made

toward understanding the structure of the noise in correlation functions and the physics that it contains.

Strongly interacting quantum systems can be described through path integral representations of correlation functions. In principle, MC evaluation of lattice regularized path integrals can solve QCD as well as many strongly interacting atomic and condensed matter theories. In practice, conceptual obstacles remain and large nuclei and nuclear matter are presently inaccessible to LQCD. In the grand canonical formulation, LQCD calculations with nonzero chemical potential face a sign problem where MC sampling weights are complex and cannot be interpreted as probabilities. In the canonical formulation, calculations with nonzero baryon number face a signal-to-noise (StN) problem where statistical uncertainties in MC results grow exponentially at large times. Like the sign problem, the StN problem arises when the sign of a correlation function can fluctuate, at which point cancellations allow for a mean correlation function of much smaller magnitude than a typical MC contribution.

The nucleon provides a relatively simple and well-studied example of a complex correlation function with a StN problem. The zero-momentum Euclidean nucleon correlation function $G(t)$ is guaranteed to be real by existence of a Hermitian, bounded transfer matrix and the spectral representation

$$G(t) = \langle C_i(t) \rangle = \sum_{\mathbf{x}} \langle N(\mathbf{x}, t) \bar{N}(0) \rangle$$

$$= \sum_{n=0}^{\infty} \tilde{Z}_n Z_n^\dagger e^{-E_n t} \sim e^{-M_N t}, \qquad (1)$$

where C_i denotes an individual nucleon correlation function calculated from quark propagators in the presence of the ith member of a statistical ensemble U_i of $i = 1, \ldots, N$ gauge field configurations; $\langle \cdot \rangle$ denotes an average over gauge field ensembles in $\langle C_i(t) \rangle$ and an average over quark and gluon fields in the middle term; $\bar{N}$ and N are nucleon creation and annihilation interpolating operators; $\tilde{Z}_n^\dagger$ and Z_n represent the overlap of these interpolating operators onto the nth QCD eigenstates with quantum numbers of the nucleon; E_n is the energy of the corresponding eigenstate; t is Euclidean time; M_N is the nucleon mass; and $\sim$ denotes proportionality in the limit $t \to \infty$. A phase convention for creation and annihilation operators is assumed so that $C_i(0)$ is real for all correlation functions in a statistical ensemble. At small times $C_i(t)$ is approximately real, but at large times it must be treated as a complex quantity. The equilibrium probability distribution for $C_i(t)$ can be formally defined as

$$\mathcal{P}(C_i(t)) = Z^{-1} \int \mathcal{D}U e^{-S(U)} \delta(C(U;t) - C_i(t))$$

$$\text{with} \quad Z = \int \mathcal{D}U e^{-S(U)}, \qquad (2)$$

where U is a gauge field, $C(U;t)$ is the nucleon correlation function in the presence of a background gauge field U, $\mathcal{D}U$ is the Haar measure for the gauge group, and $S(U)$ is the gauge action arising after all dynamical matter fields have been integrated out. For convenient comparison with LQCD results, a lattice regulator with a lattice spacing equal to unity will be assumed throughout. Unless specified, results will not depend on details of the ultraviolet regularization of $\mathcal{P}(C_i(t))$.

MC integration of the path integral representation of a partition function, as performed in LQCD calculations, provides a statistical ensemble of background quantum fields. Calculation of $C_i(t)$ in an ensemble of QCD-vacuum-distributed gauge fields U_i provides a statistical ensemble of correlation functions distributed according to $\mathcal{P}(C_i(t))$. Understanding the statistical properties of this ensemble is essential for efficient MC calculations, and significant progress has been achieved in this direction since the early days of lattice field theory. Following Parisi [1], Lepage [2] argued that $C_i(t)$ has a StN problem where the noise, or square root of the variance of $C_i(t)$, becomes exponentially larger than the signal, or average of $C_i(t)$, at large times. It is helpful to review the pertinent details of Parisi-Lepage scaling of the StN ratio.

Higher moments of $C_i(t)$ are themselves field theory correlation functions with well-defined spectral representations.[1] Their large-time behavior is a single decaying exponential whose scale is set by the lowest energy state with appropriate quantum numbers. Assuming that matter fields have been integrated out exactly rather than stochastically, $C_i(t)^\dagger C_i(t)$ will contain three valence quarks and three valence antiquarks whose net quark numbers are separately conserved. This does not imply that $|C_i(t)|^2$ will only contain nucleon-antinucleon states, as nothing prevents these distinct valence quarks and antiquarks from forming lower energy configurations such as three pions. Quadratic moments of the correlation function, therefore, have the asymptotic behavior

$$\langle C_i(t)^2 \rangle \sim e^{-2M_N t}, \qquad \langle |C_i(t)|^2 \rangle \sim e^{-3m_\pi t}. \qquad (3)$$

At large times, the nucleon StN ratio is determined by the slowest-decaying moments at large times, taking the form

$$\frac{\langle C_i(t) \rangle}{\sqrt{\langle |C_i(t)|^2 \rangle}} \sim e^{-(M_N - \frac{3}{2} m_\pi) t}, \qquad (4)$$

---

[1]The nth moment $\langle C_i(t)^n \rangle$ represents the n-nucleon nuclear correlation function in the absence of Pauli exchange between quarks in different nucleons. This is formally a correlation function in a partially quenched theory with nN_f valence quarks and N_f sea quarks. In general, such a theory is guaranteed to have a bounded, but not necessarily Hermitian, transfer matrix [3].

and is therefore exponentially small.[2] The quantitative behavior of the variance of baryon correlation function in LQCD calculations was investigated in high-statistics studies by the NPLQCD Collaboration [9–11] and more recently by Detmold and Endres [12,13], and was found to be roughly consistent with Parisi-Lepage scaling. One of us [14] extended Parisi-Lepage scaling to higher moments of $C_i(t)$ and showed that all odd moments of $C_i(t)$ are exponentially suppressed compared to even moments at large times; see Refs. [11,15] for further discussion. Nucleon correlation function distributions are increasingly broad and symmetric with exponentially small StN ratios at large times, as seen, for example, in histograms of the real parts of LQCD correlation functions in Ref. [11].

Beyond moments, the general form of correlation function distributions has also been investigated. Endres *et al.* [16] found that correlation functions in the nonrelativistic quantum field theory describing unitary fermions possess approximately log-normal distributions. They presented general arguments, that are discussed below, suggesting that this behavior might be a generic feature of quantum field theories. Knowledge of the approximate form of the correlation function distribution was exploited to construct an improved estimator, the cumulant expansion, that was productively applied to subsequent studies of unitary fermions [17–20]. Correlation function distributions have been studied analytically in the Nambu–Jona-Lasinio model [15,21], where it was found that real correlation functions were approximately log-normal but complex correlation functions in a physically equivalent formulation of the theory were broad and symmetric at large times with qualitative similarities to the QCD nucleon distribution. DeGrand [22] observed that meson, baryon, and gauge-field correlation functions in $SU(N_c)$ gauge theories with several choices of N_c are also approximately log-normal at small times where imaginary parts of correlation functions can be neglected. Log-normal distributions have also been observed and applications of the cumulant expansion have been explored in theories relevant to condensed matter systems [23,24]. These various observations provide strong empirical evidence that the distributions of real correlation functions in generic quantum field theories are approximately log-normal.

A generalization of the log-normal distribution for complex random variables that approximately describes the QCD nucleon correlation function at large times is presented in this work. To study the logarithm of a complex correlation function, it is useful to introduce the magnitude-phase decomposition

$$C_i(t) = |C_i(t)|e^{i\theta_i(t)} = e^{R_i(t)+i\theta_i(t)}. \tag{5}$$

At small times where the imaginary part of C_i is negligible, previous observations of log-normal correlation functions [22] demonstrate that R_i is approximately normally distributed. It is shown below that R_i is approximately normal at all times, and that θ_i is approximately normal at small times. Statistical analysis of θ_i is complicated by the fact that it is defined modulo 2π. In particular, the sample mean of a phase defined on $-\pi < \theta_i \leq \pi$ does not necessarily provide a faithful description of the intuitive average phase (consider a symmetric distribution peaked around $\pm\pi$ with a sample mean close to zero). Suitable statistical tools for analyzing θ_i are found in the theory of circular statistics and it will be seen below that θ_i is described by an approximately wrapped normal distribution.[3] This work is based on a high-statistics analysis of 500,000 nucleon correlation functions generated on a single ensemble of gauge-field configurations by the NPLQCD Collaboration [28] with LQCD. This ensemble has a pion mass of $m_\pi \sim 450$ MeV, physical strange quark mass, lattice spacing ~ 0.12 fm, and spacetime volume $32^3 \times 96$. The Lüscher-Weisz gauge action [29] and $N_f = 2 + 1$ clover-improved Wilson quark actions [30] were used to generate these ensembles, details of which can be found in Ref. [28]. Exploratory data analysis of this high-statistics ensemble plays a central role below.

Section II discusses standard statistical analysis methods in LQCD that introduce concepts used below. In Sec. III, the magnitude-phase decomposition of the nucleon correlation function and connections to the StN problem are discussed. Section III A describes the distributions of the log-magnitude and its time derivative in more detail, while Sec. III B describes the distribution of the complex phase and its time derivative and explains how their features lead to systematic bias in standard estimators during a large-time region that is dominated by noise. Section IV draws on these observations to propose an estimator for the nucleon mass in which accurate results can be extracted from the large-time noise region with a precision that is constant in source-sink separation time t but exponentially degrading in an independent time parameter Δt. Section V conjectures about applications to the spectra of generic complex correlation functions and concludes.

II. RELEVANT ASPECTS OF STANDARD ANALYSIS METHODS OF CORRELATION FUNCTIONS

Typically, in calculations of meson and baryon masses and their interactions, correlation functions are generated from combinations of quark- and gluon-level sources and sinks with the appropriate hadron-level quantum numbers.

---

[2]QCD inequalities [4–8] have been used to prove that $M_N \geq \frac{3}{2} m_\pi$. Assuming that interaction energy shifts in the three-pion states contributing to the variance correlation function are negligible, the nucleon StN ratio is therefore exponentially small for all quark masses.

[3]See Refs. [25–27] for textbook introductions to circular statistics.

Linear combinations of these correlation functions are formed, either using variational-method type techniques [31], the matrix-Prony technique [9], or other less automated methods, in order to optimize overlap onto the lowest lying states in the spectrum and establish extended plateaus in relevant effective mass plots (EMPs). In the limit of an infinite number of independent measurements, the expectation value of the correlation function is a real number at all times, and the imaginary part can be discarded as it is known to average to zero. The large-time behavior of such correlation functions becomes a single exponential (for an infinite time direction) with an argument determined by the ground-state energy associated with the particular quantum numbers, or more generally the energy of the lowest-lying state with non-negligible overlap.

The structure of the source and sink play a crucial role in determining the utility of sets of correlation functions. For many observables of interest, it is desirable to optimize the overlap onto the ground state of the system, and to minimize the overlap onto the correlation function dictating the variance of the ground state. In the case of the single nucleon, the sources and sinks, $\mathcal{O}$, are tuned in an effort to have maximal overlap onto the ground-state nucleon, while minimizing overlap of $\mathcal{O}\mathcal{O}^\dagger$ onto the three-pion ground state [12]. NPLQCD uses momentum projected hadronic blocks [32] generated from quark propagators originating from localized smeared sources to suppress the overlap into the three-pion ground state by a factor of $1/\sqrt{V}$, where V is the lattice volume, e.g. Ref. [9]. For such constructions, the variance of the average scales as $\sim e^{-3m_\pi t}/(VN)$ at large times, where N is the number of statistically independent correlation functions, while the nucleon correlation function scales as $\sim e^{-M_N t}$. For this setup, the StN ratio scales as $\sim \sqrt{VN}e^{-(M_N-3m_\pi/2)t}$, from which it is clear that exponentially large numbers of correlation functions or volumes are required to overcome the StN problem at large times. The situation is quite different at small and intermediate times in which the variance correlation function is dominated, not by the three-pion ground state, but by the "connected" nucleon-antinucleon excited state, which provides a variance contribution that scales as $\sim e^{-2M_N t}/N$.

This time interval where the nucleon correlation function is in its ground state and the variance correlation function is in a nucleon-antinucleon excited state has been called the "golden window" [9] (GW). The variance in the GW is generated, in part, by the distribution of overlaps of the source and sink onto the ground state, that differs at each lattice site due to variations in the gluon fields. In the work of NPLQCD, correlation functions arising from Gaussian-smeared quark-propagator sources and pointlike or Gaussian-smeared sinks have been used to form single-baryon hadronic blocks. Linear combinations of these blocks are combined with coefficients (determined using the matrix-Prony technique of Ref. [9] or simply by minimizing the χ^2/dof in fitting a constant to an extended

plateau region) that extend the single-baryon plateau region to earlier times, eliminating the contribution from the first excited state of the baryon and providing access to smaller time slices of the correlation functions where StN degradation is less severe. High-statistics analyses of these optimized correlation functions have shown that GW results are exponentially more precise and have a StN ratio that degrades exponentially more slowly than larger time results [9,33,34] (for a review, see Ref. [10]). In particular StN growth in the GW has been shown to be consistent with an energy scale close to zero, as is expected from a variance correlation function dominated by baryon, as opposed to meson, states. Despite the ongoing successes of GW analyses of few-baryon correlation functions, the GW shrinks with increasing baryon number [9,33,34] and calculations of larger nuclei may require different analysis strategies suitable for a correlation function without a GW.

EMPs, such as that associated with the Ξ-baryon shown in Fig. 1, are formed from ratios of correlation functions, which become constant when only a single exponential is contributing to the correlation function,

$$M(t) = \frac{1}{t_J}\ln\left[\frac{\langle C_i(t)\rangle}{\langle C_i(t+t_J)\rangle}\right] \to E_0, \qquad (6)$$

where E_0 is the ground-state energy in the channel with appropriate quantum numbers. The average over gauge field configurations is typically over correlation functions derived from multiple source points on multiple gauge-field configurations. This is well-known technology and is a "workhorse" in the analysis of LQCD calculations. Typically, t_J corresponds to one temporal lattice spacing, and the jackknife and bootstrap resampling techniques are used to generate covariance matrices in the plateau interval used to extract the ground-state energy from a correlated χ^2-minimization [10,11,35].[4] The energy can be extracted from an exponential fit to the correlation function or by a direct fit to the effective mass itself. Because correlation functions generated from the same, and nearby, gauge-field configuration are correlated, typically they are blocked to form one average correlation function per configuration, and blocked further over multiple configurations, to create a smaller ensemble containing (approximately) statistically independent samplings of the correlation function.

It is known that baryon correlation functions contain strong correlations over $\sim m_\pi^{-1}$ time scales, and that these correlations are sensitive to the presence of outliers. Figure 2 shows the distribution of the real part of small-time nucleon correlation functions, which resembles a heavy-tailed log-normal distribution [22]. Log-normal

---

[4]For pedagogical introductions to LQCD uncertainty quantification with resampling methods, see Refs. [11,35–37].
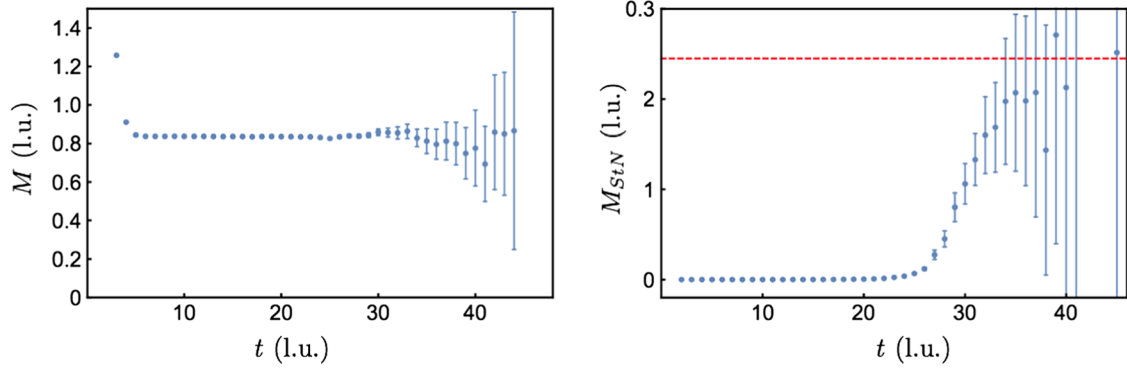
FIG. 1.   The EMP associated with the Ξ-baryon correlation function with $t_J = 2$ (left panel) and the energy scale associated with the standard deviation of the ground-state energy (right panel). This correlation function is a tuned linear combination of those resulting from localized smeared and point sinks and from a localized smeared source at a pion mass of $m_\pi \sim 450$ MeV calculated from 96 sources per configuration on 3538 statistically independent isotropic clover gauge-field configurations [28]. They have been blocked together to form 100 independent samplings of the combined correlation function. These blocked correlation functions have been bootstrap resampled, and averages across these bootstrap ensembles have been used to determine the mean results and standard error intervals displayed. Bootstrap averages and covariance matrices for this and other figures in this work are provided as supplemental material [38] and can be used to reproduce the least-squares fits used to determine results quoted in the main text. The red dashed line in the right panel corresponds to the lowest energy contributing to the StN ratio that is expected to dominate at large times.
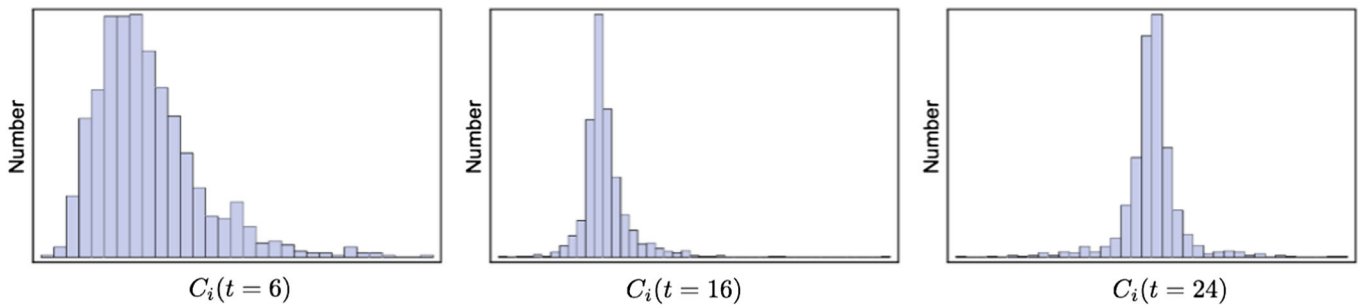

FIG. 2.   The distribution of the real part of 10^3 nucleon correlation functions at time slices $t = 6$ (left panel), $t = 16$ (middle panel) and $t = 24$ (right panel).

distributions are associated with a larger number of "outliers" that arise when sampling a Gaussian distribution, and the sample mean of these small-time correlation function will be strongly affected by the presence of these outliers. The distribution of baryon correlation functions at very large source-sink separations is also heavy tailed; Kaplan has analyzed the real parts of NPLQCD baryon correlation functions and found that they resemble a stable distribution [39]. Cancellations between positive and negative outliers occur in determinations of the sample mean of this large-time distribution, leading to different statistical issues that are explored in detail in Sec. III.

To analyze temporal correlations in baryon correlation functions in more detail, results for inverse covariance matrices generated through bootstrap resampling of the Ξ baryon effective mass are shown in Fig. 3. The size of off-diagonal elements in the inverse covariance matrix directly sets the size of contributions to the least-squares fit result from temporal correlations in the effective mass, and so it

is appropriate to use their magnitude to describe the strength of temporal correlations. The inverse covariance matrix is seen to possess large off-diagonal elements associated with small time separations that appear to decrease exponentially with increasing time separation at a rate somewhat faster than m_π^{-1}. Mild variation in the inverse covariance matrix is seen when t_J is varied. Taking $t_J \gg m_\pi^{-1}$ decreases correlations between $C_i(t)$ and $C_i(t - t_J)$ and is expected to reduce off-diagonal correlations between $M(t)$ and $M(t')$, but this effect is not clearly visible in the inverse covariance matrix on the logarithmic scale shown in Fig. 3.

The role of outliers in temporal correlations on time scales $\lesssim m_\pi^{-1}$ is highlighted in Fig. 4, where inverse covariance matrices determined with the Hodges-Lehmann estimator are shown. The utility of robust estimators, such as the median and the Hodges-Lehmann estimator, with reduced sensitivity to outliers, has been explored in Ref. [11]. When the median and average of a function
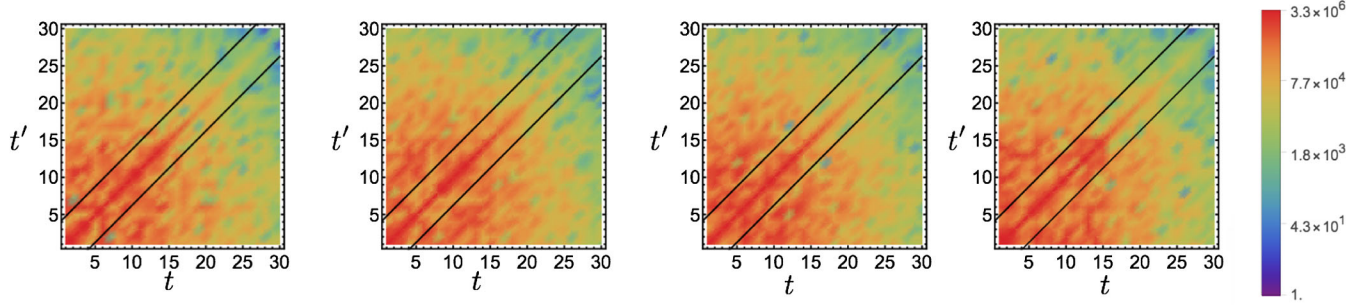
FIG. 3.   The logarithm of the inverse of the Ξ-baryon effective mass covariance matrix for $t_J = 1, 2, 3, 16$ determined using bootstrap resampling of the sample mean. Lines with $|t - t'| = m_\pi^{-1}$ are shown to demonstrate expected hadronic correlation lengths. The correlation function is the same as in Fig. 1. The normalization of the color scale is identical for all t_J.
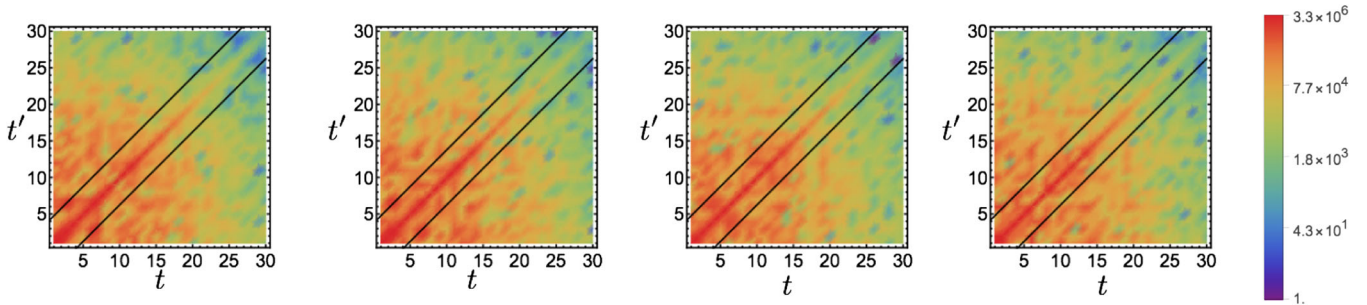

FIG. 4. Ξ-baryon effective mass inverse covariance matrices with the same range of t_J, color scale normalization, and lines at $|t - t'| = m_\pi^{-1}$ as Fig. 3. In contrast to Fig. 3, the covariance matrices are determined using bootstrap resampling of the Hodges-Lehman estimator.

are known to coincide, there are advantages to using the median or Hodges-Lehmann estimator to determine the average of a distribution. The associated uncertainty can be estimated with the "median absolute deviation" (MAD) and be related to the standard deviation with a well-known scaling factor. Off-diagonal elements in the inverse covariance matrix associated with time scales $\lesssim m_\pi^{-1}$ are visibly smaller on a logarithmic scale when the covariance matrix is determined with the Hodges-Lehmann estimator instead of the sample mean. This decrease in small-time correlations when a robust estimator is employed strongly suggests that short-time correlations on scales $\lesssim m_\pi^{-1}$ are associated with outliers.

III. A MAGNITUDE-PHASE DECOMPOSITION

In terms of the log-magnitude and phase, the mean nucleon correlation function is

$$\langle C_i(t) \rangle = \int \mathcal{D}C_i \mathcal{P}(C_i(t)) e^{R_i(t) + i\theta_i(t)}. \qquad (7)$$

In principle, $e^{R_i(t)}$ could be included in the MC probability distribution used for importance sampling. With this approach, $R_i(t)$ would contribute as an additional term in a new effective action. The presence of nonzero $\theta_i(t)$ demonstrates that this effective action would have an imaginary part. The resulting weight therefore could not be interpreted as a probability and importance sampling could not proceed; importance sampling of $C_i(t)$ faces a sign problem. In either the canonical or grand canonical approach, one-baryon correlation functions are described by complex correlation functions that cannot be directly importance sampled without a sign problem, but it is formally permissible to importance sample according to the vacuum probability distribution, calculate the phase resulting from the imaginary effective action on each background field configuration produced in this way, and average the results on an ensemble of background fields. This approach, known as reweighting, has a long history in grand canonical ensemble calculations but has been generically unsuccessful because statistical averaging is impeded by large fluctuations in the complex phase that grow exponentially with increasing spacetime volume [40–42]. Canonical ensemble nucleon calculations averaging $C_i(t)$ over background fields importance sampled with respect to the vacuum probability distribution are in effect solving the sign problem associated with nonzero $\theta_i(t)$ by reweighting. As emphasized by Ref. [15], similar chiral physics is responsible for the exponentially hard StN problem appearing in canonical calculations and exponentially large fluctuations of the complex phase in grand canonical calculations.

Reweighting a pure phase causing a sign problem generically produces a StN problem in theories with a mass gap. Suppose $\langle e^{i\theta_i(t)}\rangle \sim e^{-M_\theta t}$ for some $M_\theta \neq 0$. Then because $|e^{i\theta_i(t)}|^2 = 1$ by construction, $\theta_i(t)$ has the StN ratio

$$\frac{\langle e^{i\theta_i(t)}\rangle}{\sqrt{\langle |e^{i\theta_i(t)}|^2\rangle}} = \langle e^{i\theta_i(t)}\rangle \sim e^{-M_\theta t}, \qquad (8)$$

which is necessarily exponentially small at large times. Nonzero M_θ guarantees that statistical sampling of $e^{i\theta_i(t)}$ has a StN problem. Strictly, this argument applies to a pure phase but not to a generic complex observable such as $C_i(t)$ which might receive zero effective mass contribution from $\theta_i(t)$ and could have important correlations between $R_i(t)$ and $\theta_i(t)$. MC LQCD studies are needed to understand whether the pure phase StN problem of Eq. (8) captures some or all of the nucleon StN problem of Eq. (4).

To determine the large-time behavior of correlation functions, it is useful to consider the effective-mass estimator commonly used in LQCD spectroscopy, a special case of Eq. (6),

$$M(t) = \ln\left[\frac{\langle C_i(t)\rangle}{\langle C_i(t+1)\rangle}\right]. \qquad (9)$$

As $t \to \infty$, the average correlation function can be described by a single exponential whose decay rate is set by the ground-state energy, and therefore $M(t) \to M_N$. The uncertainties associated with $M(t)$ can be estimated by resampling methods such as bootstrap. The variance of $M(t)$ is generically smaller than that of $\ln\langle C_i(t)\rangle$ due to cancellations arising from correlations between $\ln[\langle C_i(t)\rangle]$ and $\ln[\langle C_i(t+1)\rangle]$ across bootstrap ensembles. Assuming that these correlations do not affect the asymptotic scaling of the variance of $M(t)$, propagation of uncertainties for

bootstrap estimates of the variance of $\ln[\langle C_i(t)\rangle]$ shows that the variance of $M(t)$ scales as

$$\mathrm{Var}(M(t)) \sim \frac{\mathrm{Var}(C_i(t))}{\langle C_i(t)\rangle^2} \sim e^{2(M_N - \frac{3}{2}m_\pi)t}. \qquad (10)$$

An analogous effective-mass estimator for the large-time exponential decay of the magnitude is

$$M_R(t) = \ln\left[\frac{\langle e^{R_i(t)}\rangle}{\langle e^{R_i(t+1)}\rangle}\right], \qquad (11)$$

and an effective-mass estimator for the phase is

$$M_\theta(t) = \ln\left[\frac{\langle e^{i\theta_i(t)}\rangle}{\langle e^{i\theta_i(t+1)}\rangle}\right] = \ln\left[\frac{\langle \cos(\theta_i(t))\rangle}{\langle \cos(\theta_i(t+1))\rangle}\right], \qquad (12)$$

where the reality of the average correlation function has been used.

Figure 5 shows EMPs for $M(t)$, $M_R(t)$, and $M_\theta(t)$ calculated from the LQCD ensemble described previously. The mass of the nucleon, determined from a constant fit in the shaded plateau region $15 \leq t \leq 25$ indicated in Fig. 5, is found to be $M_N = 0.7253(11)(22)$, in agreement with the mass obtained from the golden window in previous studies [28] of $M_N = 0.72546(47)(31)$. $M_R(t)$ and $M_\theta(t)$ do not visually plateau until much larger times. For the magnitude, a constant fit in the shaded region $30 \leq t \leq 40$ gives an effective mass $M_R(t) \to M_R = 0.4085(2)(13)$ which is close to the value $\frac{3}{2}m_\pi = 0.39911(35)(14)$ [28] indicated by the red line. For the phase, a constant fit to the shaded region $25 \leq t \leq 29$ gives an effective mass $M_\theta(t) \to M_\theta = 0.296(20)(12)$, which is consistent with the value $M_N - \frac{3}{2}m_\pi = 0.32636(58)(34)$ [28] indicated by the red line. It is unlikely that the phase has reached its asymptotic value by this time, but a signal cannot be established at larger times. For $t \geq 30$, large fluctuations lead to complex
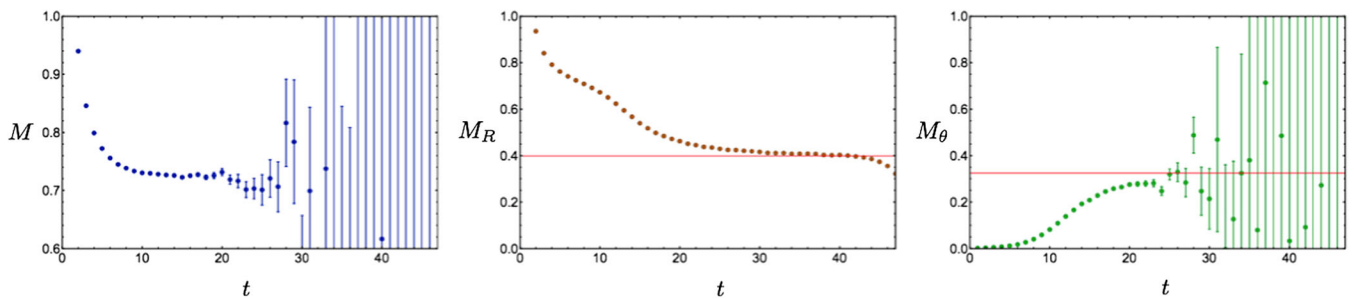

FIG. 5. The left panel shows the nucleon effective mass $M(t)$ as a function of Euclidean time in lattice units. The middle and right panels show the effective masses $M_R(t)$ and $M_\theta(t)$ of the magnitude and phase respectively. The asymptotic values of $M_R(t)$ and $M_\theta(t)$ are close to $\frac{3}{2}m_\pi$ and $M_N - \frac{3}{2}m_\pi$ respectively, whose values are indicated for comparison with horizontal red lines. The uncertainties are calculated using bootstrap methods. Past $t \gtrsim 30$ the imaginary parts of $\langle C_i(t)\rangle$ and $\langle \cos\theta_i(t)\rangle$ are not negligible compared to the real part. Here and below we display the real part of the complex log in Eqs. (9)–(12); taking the real part of the average correlation functions before taking the log or some other prescription would modify the results after $t \gtrsim 30$ in the left and right panels. All definitions are equivalent in the infinite statistics limit where $\langle C_i(t)\rangle$ is real.
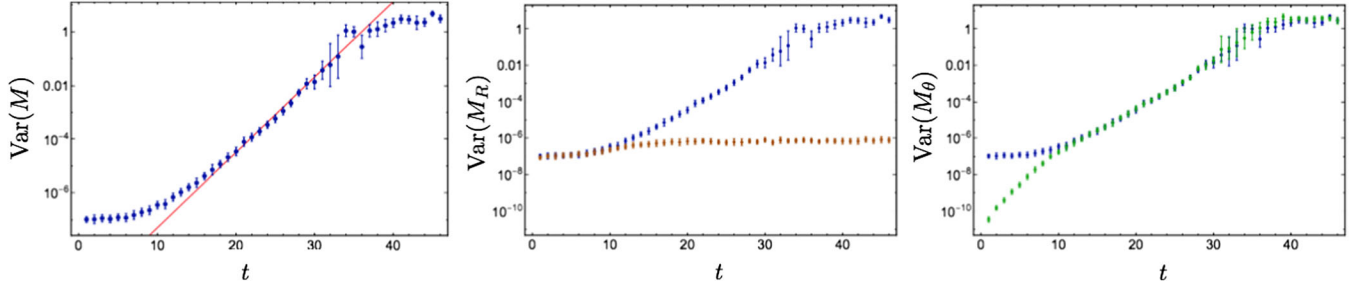
FIG. 6.   Variances of the effective mass estimates shown in Fig. 5. The blue points common to all panels show the variance of $M(t)$. The red line in the left plot shows a fit to $e^{2(M_N-\frac{3}{2}m_\pi)t}$ variance growth, where the normalization has been fixed to reproduce the observed variance at $t = 22$. The orange points in the middle panel show the variance associated with $M_R(t)$. The green points in the right panel show the variance associated with $M_\theta(t)$.

effective mass estimates for $M(t)$ and $M_\theta(t)$ and unreliable estimates and uncertainties. $M_R(t) + M_\theta(t)$ agrees with $M(t)$ up to $\lesssim 5\%$ corrections at all times, demonstrating that the magnitude and cosine of the complex phase are approximately uncorrelated at the few percent level. This suggests the asymptotic scaling of the nucleon correlation function can be approximately decomposed as

$$\langle C_i(t)\rangle \approx \langle e^{R_i(t)}\rangle \langle e^{i\theta_i(t)}\rangle \sim \left(e^{-\frac{3}{2}m_\pi t}\right)\left(e^{-(M_N-\frac{3}{2}m_\pi)t}\right). \quad (13)$$

For small times $t \lesssim 10$, the means and variances of $M(t)$ and $M_R(t)$ agree up to a small contribution from $M_\theta(t)$. This indicates that the real part of the correlation function is nearly equal to its magnitude at small times. At intermediate times $10 \lesssim t \lesssim 25$, the contribution of $M_\theta(t)$ grows relative to $M_R(t)$, and for $t \gtrsim 15$ the variance of the full effective mass is nearly saturated by the variance of $M_\theta(t)$, as shown in Fig. 6. At intermediate times a linear fit normalized to $\text{Var}(M(t = 22))$ with slope $e^{2(M_N-\frac{3}{2}m_\pi)t}$ provides an excellent fit to bootstrap estimates of $\text{Var}(M(t))$, in agreement with the scaling of Eq. (10). $\text{Var}(M_\theta(t))$ is indistinguishable from $\text{Var}(M(t))$ in this region, and $m_\theta(t)$ has an identical StN problem.

$\text{Var}(M_R(t))$ has much more mild time variation, and $M_R(t)$ can be reliably estimated at all times with no significant StN problem. At intermediate times, the presence of nonzero $\theta_i(t)$ signaling a sign problem in importance sampling of $C_i(t)$ appears responsible for the entire nucleon StN problem.

$M(t)$ approaches its asymptotic value much sooner than $M_R(t)$ or $M_\theta(t)$. This indicates that the overlap of $\bar{N}(0)N(0)$ onto the three-pion ground state in the variance correlation function is greatly suppressed compared to the overlap of $\bar{N}(0)$ onto the one-nucleon signal ground state. Optimization of the interpolating operators for high signal overlap contributes to this. Another contribution arises from momentum projection, which suppresses the variance overlap factor by $\sim 1/(m_\pi^3 V)$ [33]. A large hierarchy between the signal and noise overlap factors provides a GW visible at intermediate times $10 \lesssim t \lesssim 25$. In the GW, $M(t)$ approaches its asymptotic value but $\text{Var}(M(t))$ begins to grow exponentially and $M_\theta(t)$ is suppressed compared to $M_R(t)$. Reliable extractions of $M(t)$ are possible in the GW.

The effects of blocking, that is averaging subsets of correlation functions and analyzing the distribution of the averages, are shown in Figs. 7 and 8. $M_\theta(t)$ is suppressed compared to $M_R(t)$ for larger times in the
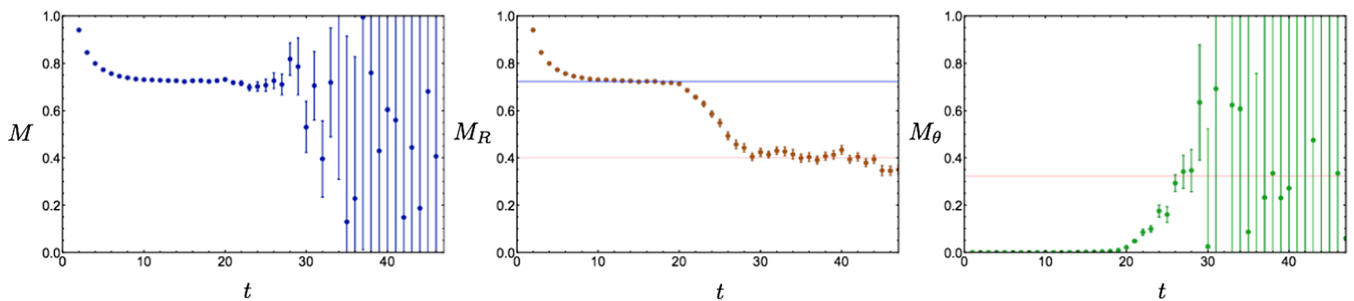

FIG. 7.   EMPs from an ensemble of 500 blocked correlation functions, each of which is equal to the sample mean of 1000 nucleon correlation functions. The left panel shows the effective mass $M(t)$ of the blocked correlation functions. The middle panel shows the magnitude contribution $m_R(t)$ and, for reference, a red line at $\frac{3}{2}m_\pi$ and a blue line at M_N are shown. The right panel shows the phase mass $m_\theta(t)$ of the blocked correlation functions along with a red line at $M_N - \frac{3}{2}m_\pi$.
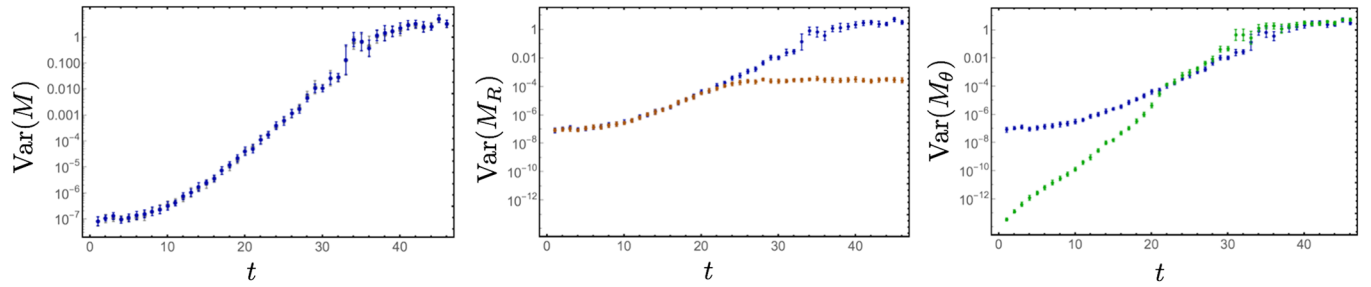
FIG. 8.   Bootstrap estimates of the variance of the effective mass using blocked correlation functions. The left panel shows the variance of $M(t)$ for blocked data in blue and the almost indistinguishable variance of $M(t)$ for unblocked data in gray. The middle panel shows the variance of blocked estimates of $m_R(t)$ in orange and the right panel shows the variance of blocked estimates of $m_\theta(t)$ in green.

blocked ensemble, and the log-magnitude saturates the average and variance of $M(t)$ through intermediate times $t \lesssim 25$. Blocking does not actually reduce the total variance of $M(t)$. Variance in $M(t)$ is merely shifted from the phase to the log-magnitude at intermediate times. This is reasonable, since the imaginary part of $C_i(t)$ vanishes on average and so blocked correlation functions will have smaller imaginary parts. Still, blocking does not affect $\langle C(t) \rangle$ and only affects bootstrap estimates of $\mathrm{Var}(M(t))$ at the level of correlations between correlation functions in the ensemble. Blocking also does not delay the onset of a large-time noise region $t \gtrsim 35$ where $M(t)$ and $m_\theta(t)$ cannot be reliably estimated.

Eventually the scaling of $\mathrm{Var}(M(t))$ begins to deviate from Eq. (10), and in the noise region $t \gtrsim 35$ the observed variance remains approximately constant (up to large fluctuations). This is inconsistent with Parisi-Lepage scaling. While the onset of the noise region is close to the midpoint of the time direction $t = 48$, a qualitatively similar onset occurs at earlier times in smaller statistical ensembles. Standard statistical estimators therefore do not reproduce the scaling required by basic principles of quantum field theory in the noise region. This suggests

systematic deficiencies leading to unreliable results for standard statistical estimation of correlation functions in the noise region. The emergence of a noise region where standard statistical tools are unreliable can be understood in terms of the circular statistics describing $\theta(t)$ and is explained in Sec. III B. A more straightforward analysis of the distribution of $R_i(t)$ is first presented below.

A. The magnitude

Histograms of the nucleon log-magnitude are shown in Fig. 9. Particularly at large times, the distribution of $R_i(t)$ is approximately described by a normal distribution. Fits to a normal distribution are qualitatively good but not exact, and deviations between normal distribution fits and $R_i(t)$ results are visible in Fig. 9. Cumulants of $R_i(t)$ can be used to quantify these deviations, which can be recursively calculated from its moments by

$$\kappa_n(R_i(t)) = \langle R_i(t)^n \rangle - \sum_{m=1}^{n-1} \binom{n-1}{m-1} \kappa_m(R_i(t)) \langle R_i(t)^{n-m} \rangle.$$
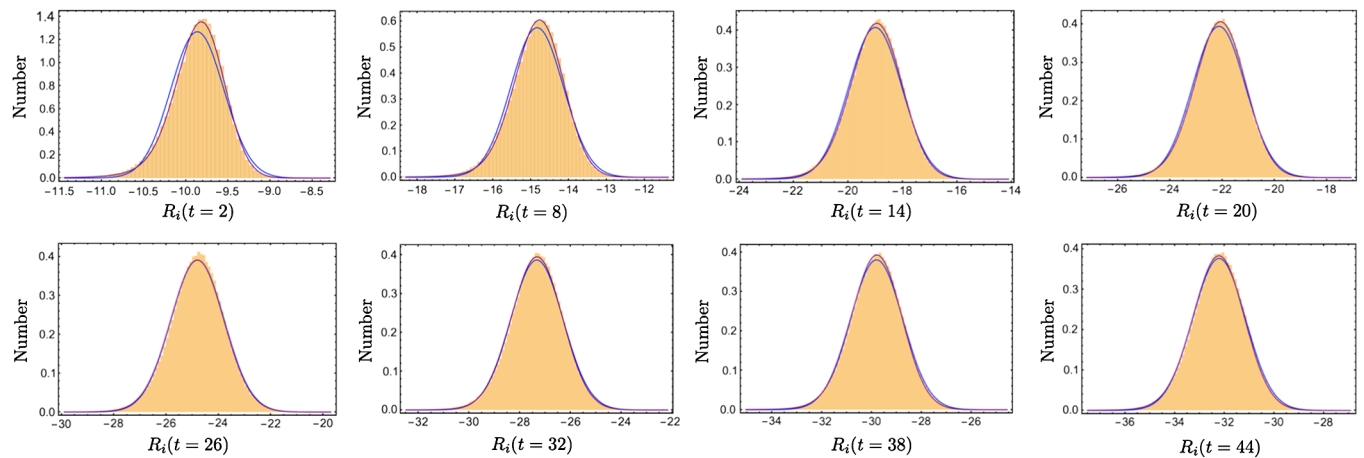
$$(14)$$

FIG. 9.   Normalized histograms of $R_i(t)$ derived from the LQCD results. The blue curves correspond to best-fit normal distributions determined from the sample mean and variance, while the purple curves correspond to maximum likelihood fits to generic stable distributions. See the main text for more details.
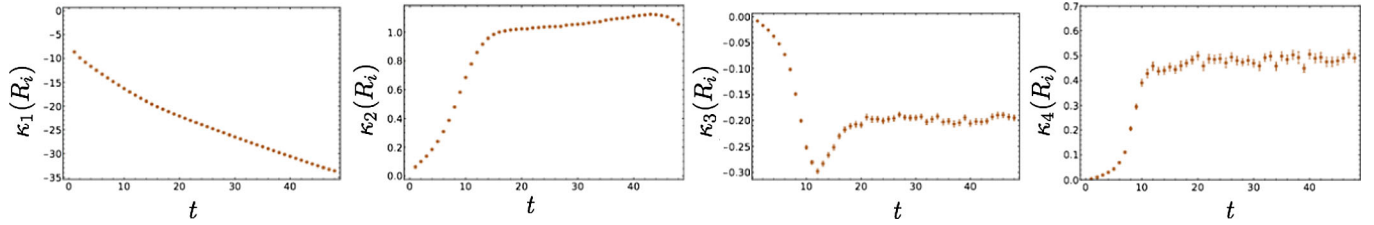
FIG. 10.   The first four cumulants of $R(t)$ as functions of t. Cumulants are calculated from sample moments using Eq. (14) and the associated uncertainties are estimated by bootstrap methods. From left to right, the panels show the cumulants $\kappa_1(R(t))$ (mean), $\kappa_2(R(t))$ (variance), $\kappa_3(R(t))$ (characterizing skewness) and κ_4 (characterizing kurtosis).

The first four cumulants of a probability distribution characterize its mean, variance, skewness, and kurtosis respectively. If $|C_i(t)|$ were exactly log-normal, the first and second cumulants of $R_i(t)$, its mean and variance, would fully describe the distribution. Third and higher cumulants of $R_i(t)$ would all vanish for exactly log-normal $|C_i(t)|$. Figure 10 shows the first four cumulants of $R_i(t)$. Estimates of higher cumulants of $R_i(t)$ become successively noisier.

The cumulant expansion of Ref. [16] relates the effective mass of a correlation function to the cumulants of the logarithm of the correlation function. The derivation of Ref. [16] is directly applicable to $M_R(t)$. The characteristic function $\Phi_{R(t)}(k)$, defined as the Fourier transform of the probability distribution function of $R_i(t)$, can be described by a Taylor series for $\ln[\Phi_{R(t)}(k)]$ whose coefficients are precisely the cumulants of $R_i(t)$,

$$\Phi_{R(t)}(k) = \langle e^{ikR_i(t)} \rangle = \exp\left[\sum_{n=1}^{\infty} \frac{(ik)^n}{n!} \kappa_n(R_i(t)) \right]. \quad (15)$$

The average magnitude of $C_i(t)$ is given in terms of this characteristic function by

$$\langle e^{R_i(t)} \rangle = \Phi_{R(t)}(-i) = \exp\left[\sum_{n=1}^{\infty} \frac{\kappa_n(R_i(t))}{n!} \right]. \quad (16)$$

This allows application of the cumulant expansion in Ref. [16] to the effective mass in Eq. (11) to give

$$M_R(t) = \sum_{n=1}^{\infty} \frac{1}{n!} [\kappa_n(R_i(t)) - \kappa_n(R_i(t+1))]. \quad (17)$$

Since $\kappa_n(R_i(t))$ with $n > 2$ vanishes for normally distributed $R_i(t)$, the cumulant expansion provides a rapidly convergent series for correlation functions that are close to, but not exactly, log-normally distributed. Note that the right-hand side of Eq. (17) is simply a discrete approximation suitable for a lattice regularized theory of the time derivative of the cumulants.

Results for the effective mass contributions of the first few terms in the cumulant expansion of Eq. (17) are shown in Fig. 11. The contribution $\kappa_1(R_i(t)) - \kappa_1(R_i(t+1))$, representing the time derivative of the mean, provides an excellent approximation to $M_R(t)$ after small times. $(\kappa_2(R_i(t)) - \kappa_2(R_i(t+1)))/2$ provides a very small negative contribution to $M_R(t)$, and contributions from $\kappa_3(R_i(t))$ and $\kappa_4(R_i(t))$ are statistically consistent with zero. As $M_R(t)$ approaches its asymptotic value, the log-magnitude distribution can be described to high accuracy by a nearly normal distribution with very slowly increasing variance and small, approximately constant $\kappa_{3,4}$. The slow increase of the variance of $R_i(t)$ is consistent with observations above that $|C_i(t)|$ has no severe StN problem. It is also consistent with expectations that $|C_i(t)|^2$ describes a (partially quenched) three-pion correlation function with a very mild StN problem, with a scale set by the attractive isoscalar pion interaction energy.

As Eq. (17) relates $M_R(t)$ to time derivatives of moments of $R_i(t)$, it is interesting to consider the distribution
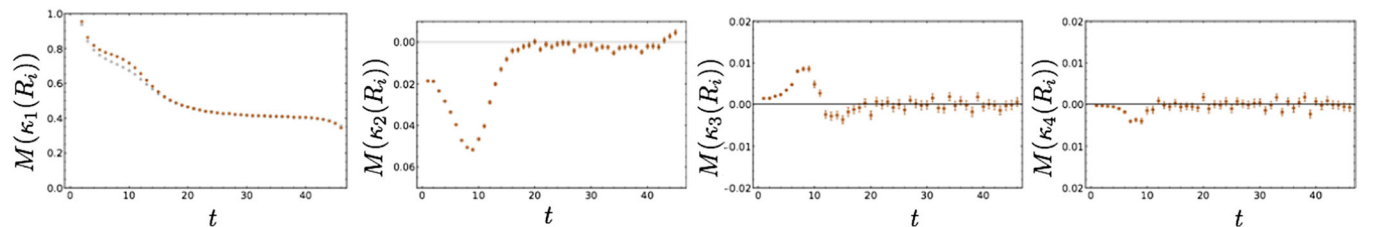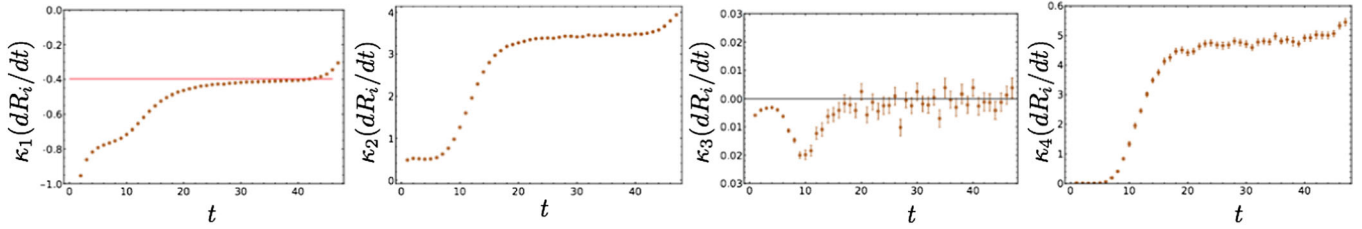

FIG. 11.   Contributions to $M_R(t)$ from the first four terms in the cumulant expansion of Ref. [16] given in Eq. (17). In the leftmost panel, the gray points correspond to the unapproximated estimate for $M_R(t)$ (that are also shown in Fig. 5), while the orange points show the contribution from the mean $\kappa_1(R(t))$. The other panels show the contributions to Eq. (17) associated with the higher cumulants $\kappa_2(R_i(t))$, $\kappa_3(R(t))$, and $\kappa_4(R(t))$, respectively.

FIG. 12.   The first four cumulants of $\frac{dR_i}{dt}$, determined analogously to the cumulants in Fig. 10.

of the time derivative $\frac{dR_i}{dt}$. Defining generic finite differences,

$$\Delta R_i(t, \Delta t) = R_i(t) - R_i(t - \Delta t), \qquad (18)$$

the time derivative of lattice regularized results can be defined as the finite difference,

$$\frac{dR_i}{dt} = \Delta R_i(t, 1). \qquad (19)$$

If $R_i(t)$ and $R_i(t-1)$ were statistically independent, it would be straightforward to extract the time derivatives of the moments of $R_i(t)$ from the moments of $\frac{dR_i}{dt}$. The presence of correlations in time, arising from nontrivial QCD dynamics, obstructs a naive extraction of $M_R(t)$ from moments of $\frac{dR_i}{dt}$. For instance, without knowledge of $\langle R_i(t)R_i(t-1)\rangle$ it is impossible to extract the time derivative of the variance of $R_i(t)$ from the variance of $\frac{dR_i}{dt}$. While the time derivative of the mean of $R_i(t)$ is simply the mean of $\frac{dR_i}{dt}$, time derivatives of the higher cumulants of $R_i(t)$ cannot be extracted from the cumulants of $\frac{dR_i}{dt}$ without knowledge of dynamical correlations.

The cumulants of $\frac{dR_i}{dt}$ are shown in Fig. 12. As expected, the mean of $\frac{dR_i}{dt}$ approaches $\frac{3}{2}m_\pi$ at large times. The variance of $\frac{dR_i}{dt}$ is tending to a plateau which is approximately one third of the variance of $R_i(t)$. This implies there are correlations between $R_i(t)$ and $R_i(t-1)$ that are on the same order of the individual variances of $R_i(t)$ and $R_i(t-1)$. This is not surprising, given that the QCD correlation length is larger than the lattice spacing. No statistically significant κ_3 is seen for $\frac{dR_i}{dt}$ at large times, but a statistically significant positive κ_4 is found. Normal distribution fits to $\frac{dR_i}{dt}$ are found to be poor, as shown in Fig. 13, as they underestimate both the peak probability and the probability of finding "outliers" in the tails of the distribution. Interestingly, Fig. 12, and histograms of $\frac{dR_i}{dt}$ shown in Fig. 13, suggest that the distribution of $\frac{dR_i}{dt}$ becomes approximately time independent at large times.

Stable distributions are found to provide a much better description of $\frac{dR_i}{dt}$, and are consistent with the heuristic arguments for log-normal correlation functions given in Ref. [16]. Generic correlation functions can be viewed as products of creation and annihilation operators with many transfer matrix factors describing Euclidean time evolution. It is difficult to understand the distribution of products of transfer matrices in quantum field theories, but following
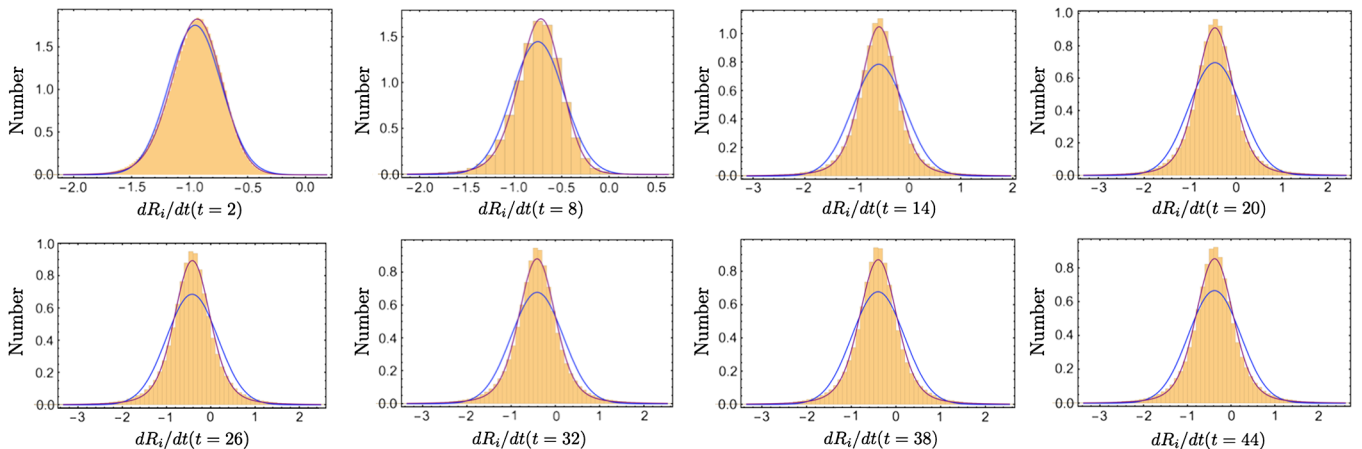

FIG. 13.   Histograms of $\frac{dR}{dt}$, defined as the finite difference $\Delta R(t, 1)$ given in Eq. (18). The blue curves in each panel correspond to the best-fit normal distribution, while the purple curves correspond to the best-fit stable distribution.
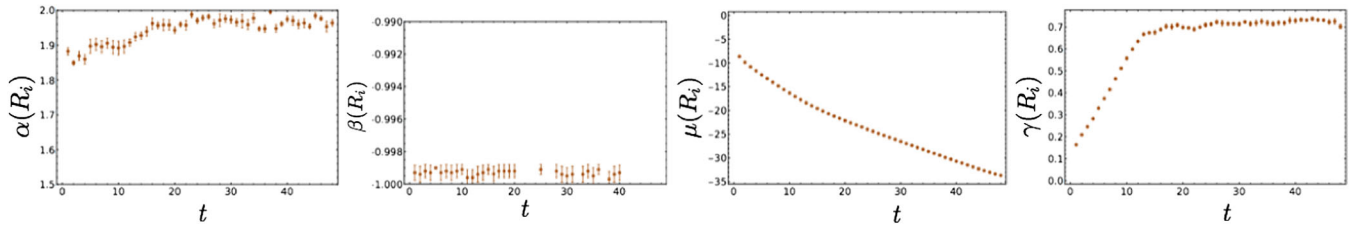
FIG. 14.   Maximum likelihood estimates for stable distribution fits of $R_i(t)$ in terms of the parameters of Eqs. (20)–(21). $\alpha = 2$ corresponds to a normal distribution. The associated uncertainties are estimated by bootstrap methods. Changes in β do not affect the likelihood when $\alpha = 2$, and reliable estimates of $\beta(R_i(t))$ are not obtained at all times.

Ref. [16] insight can be gained by considering products of random positive numbers. As a further simplification, one can consider a product of independent, identically distributed positive numbers, each schematically representing a product of many transfer matrices describing time evolution over a period much larger than all temporal correlation lengths. Application of the central limit theorem to the logarithm of a product of many independent, identically distributed random numbers shows that the logarithm of the product tends to become normally distributed as the number of factors becomes large. The central limit theorem in particular assumes that the random variables in question originate from distributions that have a finite variance. A generalized central limit theorem proves that sums of heavy-tailed random variables tend to become distributed according to stable distributions (that include the normal distribution as a special case), suggesting that stable distributions arise naturally in the logs of products of random variables.

Stable distributions are named as such because their shape is stable under averaging of independent copies of a random variable. Formally, stable distributions form a manifold of fixed points in a Wilsonian space of probability distributions where averaging independent random variables from the distribution plays the role of renormalization group evolution. A parameter α, called the index of stability, dictates the shape of a stable distribution and remains fixed under averaging transformations. All probability distributions with finite variance evolve under averaging towards the normal distribution, a special case of the stable distribution with $\alpha = 2$. Heavy-tailed distributions with ill-defined variance evolve towards generic stable distributions with $0 < \alpha \leq 2$. In particular, stable distributions with $\alpha < 2$ have power-law tails; for a stable random variable X the tails decay as $X^{-(\alpha+1)}$. The heavy-tailed Cauchy, Levy, and Holtsmark distributions are special cases of stable distributions with $\alpha = 1, 1/2$, and $3/2$, respectively, that arise in physical applications.[5]

Stable distributions for a real random variable X are defined via Fourier transform,

$$\mathcal{P}_S(X; \alpha, \beta, \mu, \gamma) = \int \frac{dk}{2\pi} e^{-ikX} \Phi_X(k; \alpha, \beta, \mu, \gamma), \quad (20)$$

of their characteristic functions

$$\Phi_X(k; \alpha, \beta, \mu, \gamma) = \exp\left(i\mu k - |\gamma k|^\alpha \left[1 - i\beta \frac{k}{|k|} \tan(\pi\alpha/2) \right] \right), \quad (21)$$

where $0 < \alpha \leq 2$ is the index of stability, $-1 \leq \beta \leq 1$ determines the skewness of the distribution, μ is the location of peak probability, and γ sets the width. For $\alpha = 1$, the above parametrization does not hold and $\tan(\pi\alpha/2)$ should be replaced by $-\frac{2}{\pi}\ln|k|$. For $\alpha > 1$ the mean is μ, and for $\alpha \leq 1$ the mean is ill defined. For $\alpha = 2$ the variance is $\sigma^2 = \gamma^2/2$ and Eq. (21) implies the distribution is independent of β, while for $\alpha < 2$ the variance is ill defined.

The distributions of $R_i(t)$ obtained from the LQCD calculations can be fit to stable distributions through maximum likelihood estimation of the stable parameters α, β, μ, and γ, obtaining the results that are shown in Fig. 14. Estimates of $\alpha(R_i)$ are consistent with 2, corresponding to a normal distribution. This is not surprising, because higher moments of $|C_i(t)|$ would be ill defined and diverge in the infinite statistics limit if $R_i(t)$ were literally described by a heavy-tailed distribution. $\beta(R_i)$ is strictly ill defined when $\alpha(R_i) = 2$, but results consistent with $\beta(R_i) = -1$ indicate negative skewness in agreement with observations above. Estimates of $\mu(R_i)$ and $\gamma(R_i)$ are consistent with the cumulant results above if a normal distribution $[\alpha(R_i) = 2]$ is assumed. Fits of $R(t)$ to generic stable distributions are shown in Fig. 9, and are roughly consistent with fits to a normal distribution, though some skewness is captured by the stable fits.

Stable distribution fits to $\frac{dR_i}{dt}$ indicate statistically significant deviations from a normal distribution ($\alpha = 2$), as seen in Fig. 15. The large-time distribution of $\frac{dR_i}{dt}$ appears time independent, and fitting $\alpha(\frac{dR_i}{dt})$ in the large-time plateau region gives an estimate of the large-time index

---

[5]Further details can be found in textbooks and reviews on stable distributions and their applications in physics. See, for instance, Refs. [43–47] and references within.
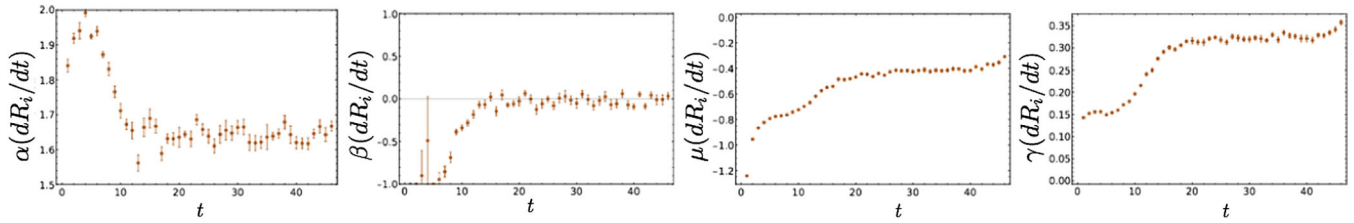
FIG. 15. Maximum likelihood estimates for stable distribution fits of $\frac{dR_i}{dt}$ similar to Fig. 14. The associated uncertainties are estimated by bootstrap methods.

of stability. Recalling that $\frac{dR_i}{dt}$ describes a finite difference over a physical time interval of one lattice spacing, we have the estimated index of stability

$$\alpha(\Delta R(t \to \infty, \Delta t \sim 0.12 \text{ fm})) \to 1.639(4)(1). \quad (22)$$

Maximum likelihood estimates for $\mu(\frac{dR_i}{dt})$ are consistent with the sample mean, and $\beta(\frac{dR_i}{dt})$ is consistent with zero in agreement with observations of vanishing skewness. Therefore, the distribution of $\frac{dR_i}{dt}$ is symmetric, as observed in Fig. 13, with power-law tails scaling as $\sim(\Delta R_i)^{-2.65}$ over this time interval of $\Delta t \sim 0.12$ fm.

The value of $\alpha(\frac{dR_i}{dt})$ depends on the physical time separation used in the finite difference definition Eq. (18), and stable distribution fits can be performed for generic finite differences $\Delta R_i(t, \Delta t)$. For all Δt, the distribution of ΔR_i becomes time independent at large times. Histograms of the large-time distributions ΔR for $\Delta t = 4, 8$ are shown in Fig. 16, and the best-fit large-time values for $\alpha(\Delta R_i)$ and $\gamma(\Delta R_i)$ are shown in Fig. 17. Since QCD has a finite correlation length, $\Delta R_i(t, \Delta t)$ can be described as the difference of approximately normally distributed variables at large Δt. In the large Δt limit, ΔR_i is therefore necessarily almost normally distributed, and correspondingly, $\alpha(\Delta R_i)$, shown in Fig. 17, increases with Δt and begins to approach the normal distribution

value $\alpha(\Delta R_i) \to 2$ for large Δt. A large Δt plateau in $\alpha(\Delta R_i)$ is observed that demonstrates small but statistically significant departures from $\alpha(\Delta R_i) < 2$. This deviation is consistent with the appearance of small but statistically significant measures of non-Gaussianity in $R_i(t)$ seen in Fig. 10. Heavy-tailed distributions are found to be needed only to describe the distribution of ΔR_i when Δt is small enough such that $R_i(t)$ and $R_i(t - \Delta t)$ are physically correlated. In some sense, the deviations from normally distributed differences, i.e. $\alpha(\Delta R_i) < 2$, are a measure of the strength of dynamical QCD correlations on the scale Δt.

The heavy-tailed distributions of ΔR_i for dynamically correlated time separations correspond to time evolution $\frac{dR_i}{dt}$ that is quite different from that of diffusive Brownian motion describing the quantum mechanical motion of free point particles. Rather than Brownian motion, heavy-tailed jumps in $R_i(t)$ correspond to a superdiffusive random walk or Lévy flight. Power-law, rather than exponentially suppressed, large jumps give Lévy flights a qualitatively different character than diffusive random walks, including fractal self-similarity, as can be seen in Fig. 18. The dynamical features of QCD that give rise to superdiffusive time evolution are presently unknown; however, we conjecture that instantons play a role. Instantons are associated with large, localized fluctuations in gauge fields, and we expect that instantons may also be responsible for infrequent, large
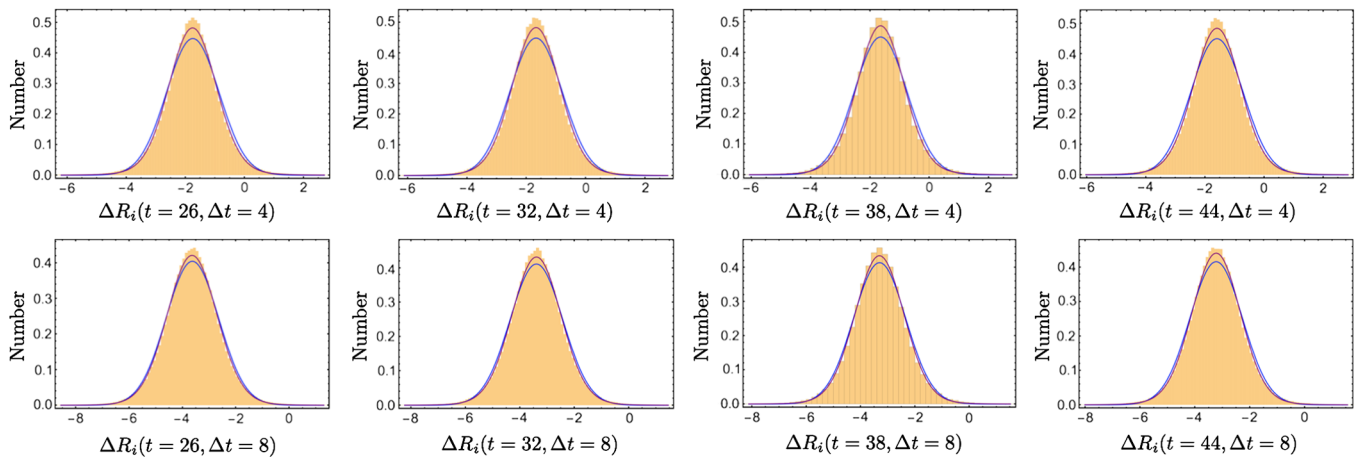

FIG. 16. Histograms of $\Delta R_i(t, \Delta t)$ for selected large-time values of t. The top row shows results for $\Delta t = 4$, the bottom row shows results for $\Delta t = 8$, and Fig. 13 shows the results for $\Delta t = 1$. The blue curves represent fits to a normal distribution, while the purple curves represent fits to a stable distribution.
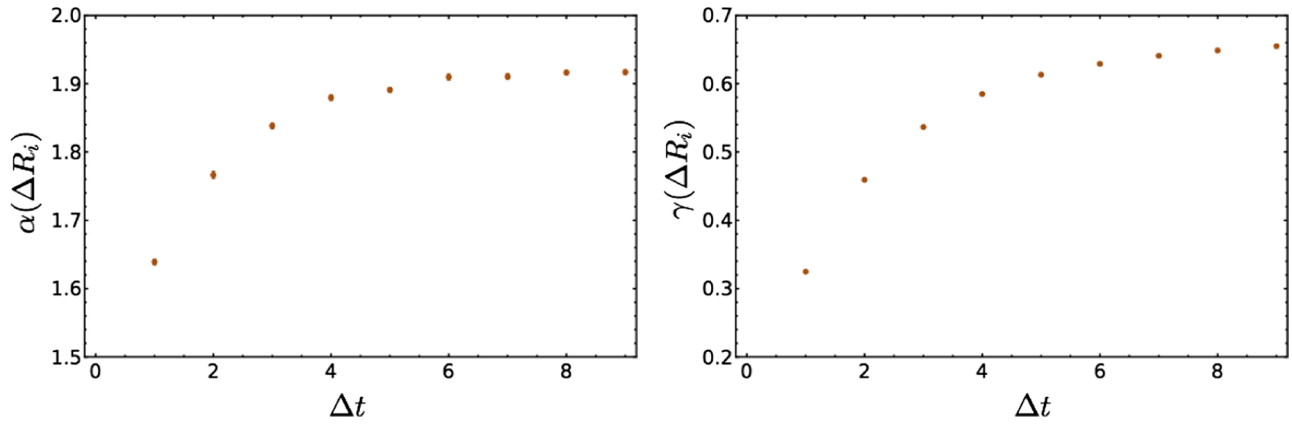
FIG. 17.   Maximum likelihood estimates for the index of stability $\alpha(\Delta R_i(t, \Delta t))$ and width $\gamma(\Delta R_i(t, \Delta t))$, in the large-time plateau region as a function of Δt. Associated uncertainties are estimated with bootstrap methods.
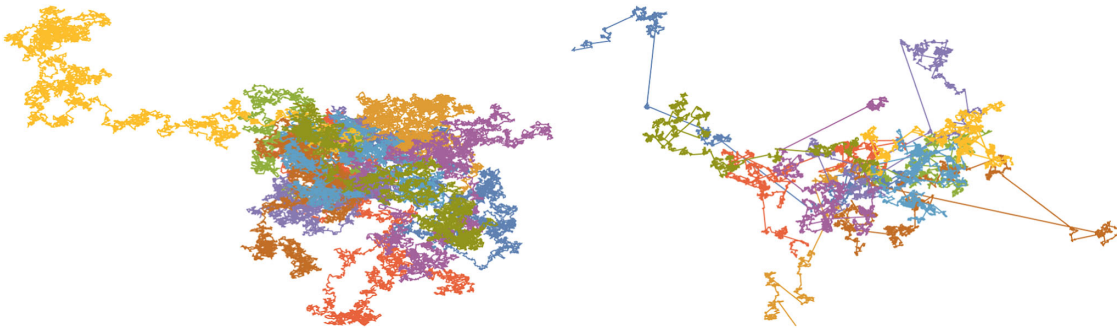

FIG. 18.   The two-dimensional motion of test particles with their random motion taken from symmetric stable distributions. At each time step, the angle of the outgoing velocity is chosen randomly with respect to the incident velocity while the magnitude of the velocity is chosen from a symmetric stable distribution with $\alpha = 2$ corresponding to Brownian motion (left panel), and $\alpha = 1.5$ corresponding to a Holtsmark distribution (right panel). In the right panel, the large separations between clusters achieved during one time interval correspond to Lévy flights.

fluctuations in hadronic correlation functions generating the tails of the dR_i/dt distribution. It would be interesting to understand if $\alpha(\frac{dR_i}{dt})$ can be simply related to observable properties of the nucleon. It is also not possible to say from this single study whether $\alpha(\frac{dR_i}{dt})$ has a well-defined continuum limit for infinitesimal Δt. Further LQCD studies are required to investigate the continuum limit of $\alpha(\frac{dR_i}{dt})$. Lattice field theory studies of other systems and calculations of $\alpha(\frac{dR_i}{dt})$ in perturbation theory, effective field theory, and models of QCD could provide important insights into the dynamical origin of superdiffusive time evolution.[6]

One feature of LQCD $\frac{dR_i}{dt}$ results is not well described by a stable distribution. The variance of heavy-tailed distributions is ill defined, and were $\frac{dR_i}{dt}$ truly described by a heavy-

tailed distribution then the variance and higher cumulants of $\frac{dR_i}{dt}$ would increase without bound as the size of the statistical ensemble is increased. This behavior is not observed. While the distribution of $\frac{dR_i}{dt}$ is well described by a stable distribution near its peak, the extreme tails of the distribution of $\frac{dR_i}{dt}$ decay sufficiently quickly that the variance and higher cumulants of $\frac{dR}{dt}$ shown in Fig. 12 give statistically consistent results as the statistical ensemble size is varied. This suggests that $\frac{dR_i}{dt}$ is better described by a truncated stable distribution, a popular model for, for example, financial markets exhibiting high volatility but with a natural cutoff on trading prices, in which some form of sharp cutoff is added to the tails of a stable distribution [46]. Note that the tails of the $\frac{dR_i}{dt}$ distribution describe extremely rapid changes in the correlation function and are sensitive to ultraviolet properties of the theory. One possibility is that $\frac{dR_i}{dt}$ describes a stable distribution in the presence of a (perhaps smooth) cutoff arising from ultraviolet regulator effects that damps the stable distribution's power-law decay at very large $\frac{dR_i}{dt}$. Further studies at different lattice spacings will be needed

---

[6]For example, an analysis of pion correlation functions from the same ensemble of gauge-field configurations shows that R_i and $\frac{dR_i}{dt}$ are both approximately normally distributed, with $\alpha = 1.96(1)$ and $\alpha = 1.97(1)$, respectively. We conclude that the pion shows only small deviations from free particle Brownian motion.

to understand the form of the truncation and whether the truncation scale is indeed set by the lattice scale. It is also possible that there is a strong interaction length scale providing a modification to the distribution at large $\frac{dR_i}{dt}$, and it is further possible that stable distributions only provide an approximate description at all $\frac{dR_i}{dt}$. For now we simply observe that a truncated stable distribution with an unspecified high-scale modification provides a good empirical description of $\frac{dR_i}{dt}$.

Before turning to the complex phase of $C_i(t)$, we summarize the main findings about the log-magnitude:

(i) The log-magnitude of the nucleon correlation function in LQCD is approximately normally distributed with small but statistically significant negative skewness and positive kurtosis.

(ii) The magnitude effective mass $M_R(t)$ approaches $\frac{3}{2}m_\pi$ at large times, consistent with expectations from Parisi-Lepage scaling for the nucleon variance $|C_i(t)|^2 \sim e^{-3m_\pi t}$. The plateau of $M(t)$ marks the start of the golden window where excited-state systematics are negligible and statistical uncertainties are increasing slowly. The plateau of $M_R(t)$, which begins at much larger t than the plateau of $M(t)$, roughly coincides with the plateau of $M_\theta(t)$ to $M_N - \frac{3}{2}m_\pi$ and occurs after variance growth of $M(t)$ reaches the Parisi-Lepage expectation $e^{2(M_N-\frac{3}{2}m_\pi)t}$. Soon after, a noise region begins where the variance of $M(t)$ stops increasing and the effective mass cannot be reliably estimated.

(iii) The log-magnitude does not have a severe StN problem, and $M_R(t)$ can be measured accurately across all 48 time steps of the present LQCD calculations. The variance of the log-magnitude distribution only increases by a few percent in 20 time steps after visibly plateauing.

(iv) The cumulant expansion describes $M_R(t)$ as a sum of the time derivatives of the cumulants of the log of the correlation function. At large times, the time derivative of the mean of $R_i(t)$ is constant and approximately equal to $M_R(t)$. Contributions to $M_R(t)$ from the variance and higher cumulants of $R_i(t)$ are barely resolved in the sample of 500,000 correlation functions.

(v) Finite differences in $R_i(t)$, $\Delta R_i(t, \Delta t)$, are described by time-independent distributions at large times. For large Δt compared to the QCD correlation length, ΔR describes a difference of approximately independent normal random variables and is therefore approximately normally distributed. For small Δt, ΔR_i describes a difference of dynamically correlated variables. The mean of $\frac{dR_i}{dt}$ is equal to the time derivative of the mean of $R_i(t)$ and therefore provides a good approximation to $M_R(t)$. The time derivatives of higher cumulants of $R_i(t)$ cannot be readily extracted from cumulants of $\frac{dR_i}{dt}$ without knowledge of dynamical correlations.

(vi) At large times, $\frac{dR_i}{dt}$ is well described by a symmetric, heavy-tailed, truncated stable distribution. The presence of heavy tails in $\frac{dR_i}{dt}$ indicates that $R_i(t)$ is not described by free particle Brownian motion but rather by a superdiffusive Lévy flight. Deviations of the index of stability of $\frac{dR_i}{dt}$ from a normal distribution quantify the amount of dynamical correlations present in the nucleon system, the physics of which is not yet understood. Further studies are required to determine the continuum limit value of the index of stability associated with $\frac{dR_i}{dt}$ and the dynamical origin and generality of superdiffusive Lévy flights in quantum field theory correlation functions.

B. The phase

The reality of average correlation functions requires that the distribution of $\theta_i(t)$ be symmetric under $\theta_i(t) \to -\theta_i(t)$. Cumulants of $\theta_i(t)$ calculated from sample moments in analogy to Eq. (14) are shown in Fig. 19. The mean and κ_3 are noisy but statistically consistent with zero as expected. The variance and κ_4 are small at small times since every sample of $\theta_i(t)$ is defined to vanish at $t = 0$, and grow linearly at intermediate times $10 < t < 20$ around the golden window. After $t = 20$, this linear growth slows and they become constant at large times, and are consistent with results from a uniform distribution. Histograms of $\theta_i(t)$ shown in Fig. 20 qualitatively suggest that $\theta_i(t)$ is described by a narrow, approximately normal distribution at
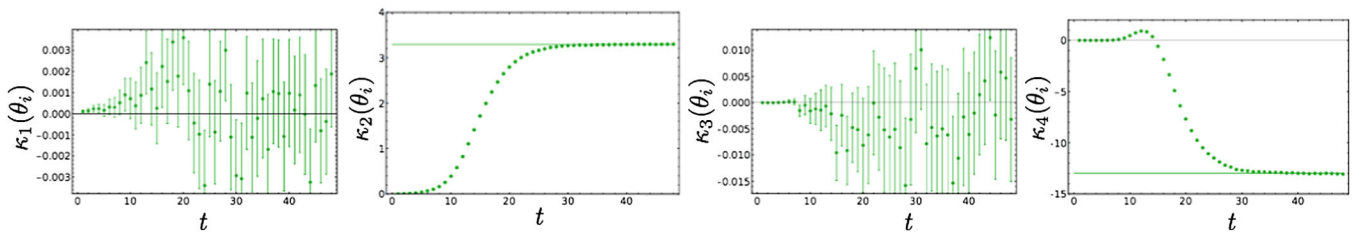

FIG. 19.   The first four cumulants of $\theta_i(t)$. In these fits, no special care is given to the fact that $\theta_i(t)$ is a phase defined on $-\pi < \theta_i(t) \leq \pi$ and standard sample moments are used to determine these cumulants in analogy to Eq. (14). Uniform distribution results of $\frac{\pi^2}{3}$ variance and $-\frac{2\pi^4}{15}$ fourth cumulant are shown as green lines for reference.
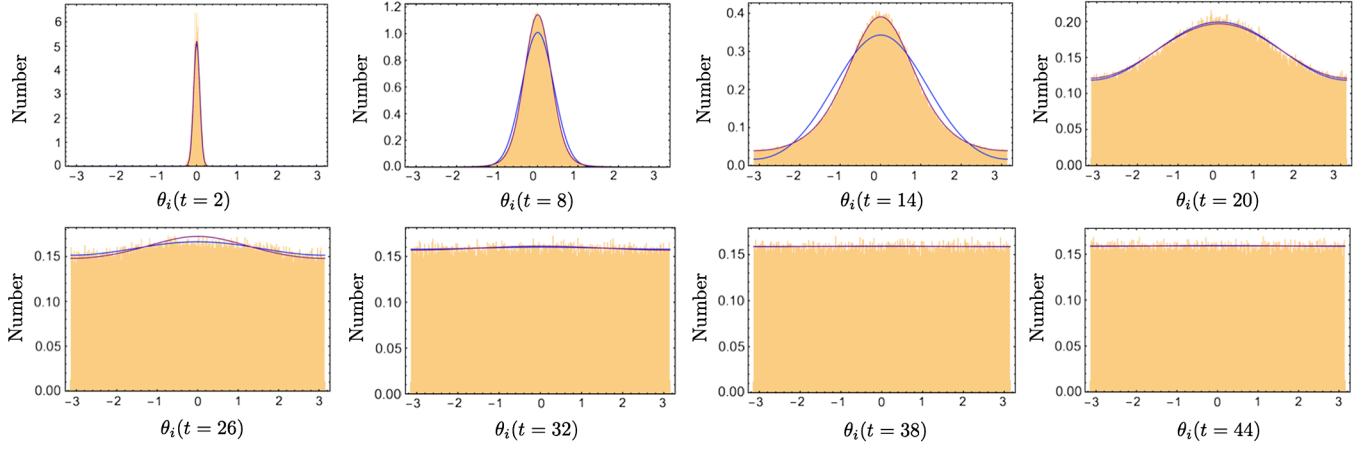
FIG. 20. Histograms of $\theta_i(t)$ with fits to wrapped normal distributions using Eq. (27) shown in blue and fits to wrapped stable distributions using maximum likelihood estimation of the parameters of Eq. (37) shown in purple. See the main text for details.

small times and an increasingly broad, approximately uniform distribution at large times. $\theta_i(t)$ is only defined modulo 2π and can be described as a circular variable defined on the interval $-\pi < \theta_i \leq \pi$. The distribution of $\theta_i(t)$ can therefore be described with angular histograms, as shown in Fig. 21. Again, $\theta_i(t)$ resembles a uniform circular random variable at large times.

A cumulant expansion can be readily constructed for $M_\theta(t)$. The mean phase is given in terms of the characteristic function and cumulants of $\theta_i(t)$ by

$$\langle e^{i\theta_i(t)} \rangle = \Phi_{\theta(t)}(1) = \exp\left[\sum_{n=0}^{\infty} \frac{i^n}{n!} \kappa_n(\theta_i(t)) \right], \quad (23)$$

and the appropriate cumulant expansion for $M_\theta(t)$ is therefore, using Eq. (12),

$$M_\theta(t) = \sum_{n=0}^{\infty} \frac{i^n}{n!} [\kappa_n(\theta_i(t)) - \kappa_n(\theta_i(t+1))]. \quad (24)$$

Factors of i^n dictate that a linearly increasing variance of $\theta_i(t)$ makes a positive contribution to $M_\theta(t)$, in contradistinction to the slight negative contribution to $M_R(t)$ made by linearly increasing variance of $R_i(t)$. Since the mean of $\theta_i(t)$ necessarily vanishes, the variance of $\theta_i(t)$ makes the dominant contribution to Eq. (24) for approximately normally distributed $\theta_i(t)$. For this contribution to be positive, the variance of $\theta_i(t)$ must increase, indicating
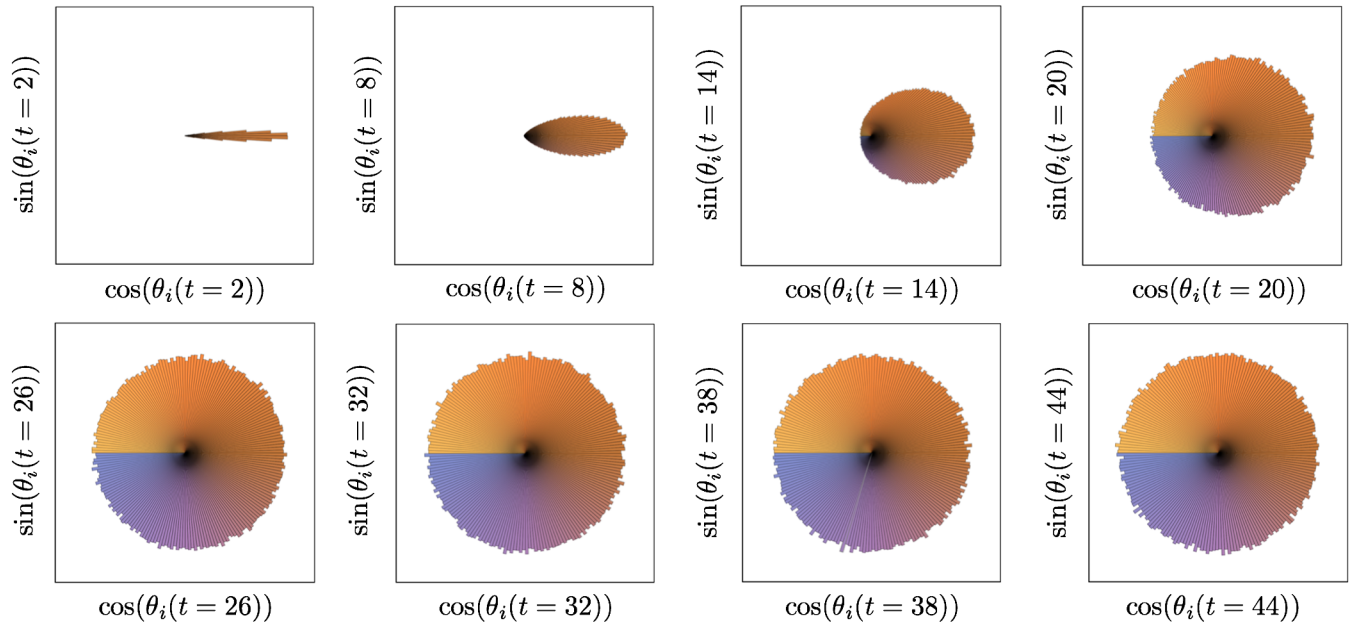

FIG. 21. Angular histograms of $\theta_i(t)$. The unit circle is split into a uniform sequence of bins, and the number of $\theta_i(t)$ samples falling in each bin sets the radial length of a bar at that angle. Colors ranging from orange to blue also denote angle, and are included to indicate the $\theta_i = \pi$ location of the branch cut in $\theta_i = \arg C_i(t)$.
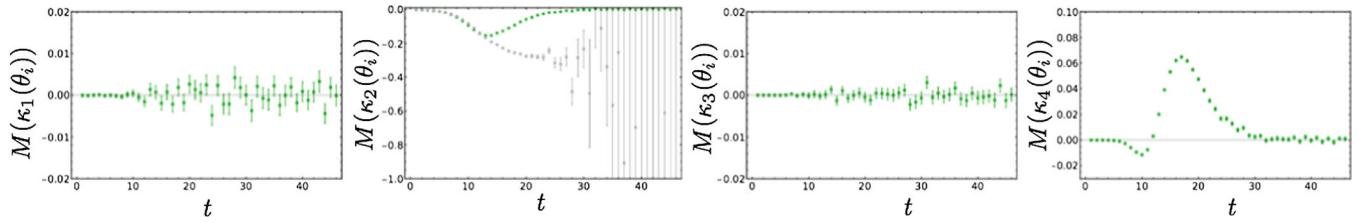
FIG. 22.   Contributions from the first four terms in the cumulant expansion of Eq. (17). The variance, shown second from left, is expected to provide the dominant contribution if a truncation of Eq. (17) is reliable. Standard estimates of $M_\theta(t)$ from Eq. (12) are shown as the gray points, alongside the cumulant contribution (green points) in the second from left panel. Other panels only show cumulant contributions (green points).

that $\theta_i(t)$ has a StN problem. For the case of approximately normally distributed $\theta_i(t)$, nonzero M_θ requires a StN problem for the phase.

Contributions to Eq. (24) from the first four cumulants of $\theta_i(t)$ are shown in Fig. 22. Contributions from odd cumulants are consistent with zero, as expected by $\theta_i(t) \rightarrow -\theta_i(t)$ symmetry. The variance provides the dominant contribution to $M_\theta(t)$ at small and intermediate times, and is indistinguishable from the total $M_\theta(t)$ calculated using the standard effective mass estimator for $t \lesssim 15$. Towards the end of the golden window $15 \lesssim t \lesssim 25$, the variance contribution to the effective mass begins to decrease. At very large times $t \gtrsim 30$ contributions to $M_\theta(t)$ from the variance are consistent with zero. The fourth cumulant makes smaller but statistically significant contributions to $M_\theta(t)$ at intermediate times. Contributions from the fourth cumulant also decrease and are consistent with zero at large times. The vanishing of these contributions results from the distribution becoming uniform at large times, and time independent as a consequence. These observations signal a breakdown in the cumulant expansion at large times $t \gtrsim 25$ where contributions from the variance do not approximate standard estimates of $M_\theta(t)$. Notably, the breakdown of the cumulant expansion at $t \gtrsim 25$ coincides with plateaus to uniform distribution cumulants in Fig. 19 and with the onset of the noise region discussed in Sec. III.

Observations of these unexpected behaviors of $\theta_i(t)$ in the noise region hint at more fundamental issues with the statistical description of $\theta_i(t)$ used above. A sufficiently localized probability distribution of a circular random variable peaked far from the boundaries of $-\pi < \theta_i(t) \leq \pi$ can be reliably approximated as a standard probability distribution of a linear random variable defined on the real line. For broad distributions of a circular variable, the effects of a finite domain with periodic boundary conditions cannot be ignored. While circular random variables are not commonly encountered in quantum field theory, they arise in many scientific contexts, most notably in astronomy, biology, geography, geology, meteorology and oceanography. Familiarity with circular statistics is not assumed here, and a few basic results relevant for

understanding the statistical properties of $\theta_i(t)$ will be reviewed without proof. Further details can be found in Refs. [25–27] and references therein.

A generic circular random variable θ_i can be described by two linear random variables $\cos(\theta_i)$ and $\sin(\theta_i)$ with support on the line interval $[-1, 1]$ where periodic boundary conditions are not imposed. It is the periodic identification of $\theta_i = \pm\pi$ that makes sample moments poor estimators of the distribution of θ_i and, in particular, allows the sample mean of a distribution symmetrically peaked about $\theta_i = \pm\pi$ to be opposite the actual location of peak probability. Parameter estimation for circular distributions can be straightforwardly performed using trigonometric moments of $\cos(\theta_i)$ and $\sin(\theta_i)$. For an ensemble of N random angles θ_i, the first trigonometric moments are defined by the sample averages,

$$\bar{\mathcal{C}} = \frac{1}{N} \sum_i \cos(\theta_i), \qquad \bar{\mathcal{S}} = \frac{1}{N} \sum_i \sin(\theta_i). \quad (25)$$

Higher trigonometric moments can be defined analogously but will not be needed here. The average angle can be defined in terms of the mean two-dimensional vector $(\bar{\mathcal{C}}, \bar{\mathcal{S}})$ as

$$\bar{\theta} = \arg\left(\bar{\mathcal{C}} + i\bar{\mathcal{S}}\right). \quad (26)$$

A standard measure of a circular distribution's width is given in terms of trigonometric moments as

$$\bar{\rho}^2 = \bar{\mathcal{C}}^2 + \bar{\mathcal{S}}^2 \quad (27)$$

where $\bar{\rho}$ should be viewed as a measure of the concentration of a circular distribution. Smaller $\bar{\rho}$ corresponds to a broader, more uniform distribution, while larger $\bar{\rho}$ corresponds to a more localized distribution.

One way of defining statistical distributions of circular random variables is by "wrapping" distributions for linear random variables around the unit circle. The probability of a circular random variable equaling some value in $-\pi < \theta \leq \pi$ is equal to the sum of the probabilities of the linear random variable equaling any value that is

equivalent to θ modulo 2π. Applying this prescription to a normally distributed linear random variable gives the wrapped normal distribution

$$
\begin{aligned}
\mathcal{P}_{WN}(\theta_i; \mu, \sigma) &= \frac{1}{\sqrt{2\pi}\sigma} \sum_{k=-\infty}^{\infty} \exp\left[-\frac{(\theta_i - \mu + 2\pi k)^2}{2\sigma^2}\right] \\
&= \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{in(\theta_i - \mu) - \sigma^2 n^2/2},
\end{aligned}
\tag{28}
$$

where the second form follows from the Poisson summation formula. Wrapped distributions share the same characteristic functions as their unwrapped counterparts, and the second expression above can be derived as a discrete Fourier transform of a normal characteristic function. The second sum above can also be compactly represented in terms of elliptic-ϑ functions. For $\sigma^2 \lesssim 1$ the wrapped normal distribution qualitatively resembles a normal distribution, but for $\sigma^2 \gtrsim 1$ the effects of wrapping obscure the localized peak. As $\sigma^2 \to \infty$, the wrapped normal distribution becomes a uniform distribution on $(-\pi, \pi]$. Arbitrary trigonometric moments and therefore the characteristic function of the wrapped normal distribution are given by

$$
\langle e^{in\theta_i} \rangle_{WN} = e^{in\mu - n^2\sigma^2/2}.
\tag{29}
$$

Parameter estimation in fitting a wrapped normal distribution to LQCD results for $\theta_i(t)$ can be readily performed by relating $\bar{\theta}$ and $\bar{\rho}$ above to these trigonometric moments as

$$
\mu = \bar{\theta} \quad \text{and} \quad e^{-\sigma^2} = \bar{\rho}^2.
\tag{30}
$$

Note that Eq. (30) holds only in the limit of infinite statistics. Estimates for the average of a wrapped normal distribution are consistent with zero at all times, as expected. Wrapped normal probability distribution functions with $\sigma^2(\theta_i(t))$ determined from Eq. (30) are shown with the histograms of Fig. 20 and provide a good fit to the data at all times.

The appearance of a uniform distribution at large times is consistent with the heuristic argument that the logarithm of a correlation function should be described by a stable distribution. The uniform distribution is a stable distribution for circular random variables, and in fact is the only stable circular distribution [27]. The distribution describing a sum of many linear random variables broadens as the number of summands is increased, and the same is true of circular random variables. A theorem of Poincaré proves that as the width of any circular distribution is increased without bound, the distribution will approach a uniform distribution. One therefore expects that the sum of many well-localized circular random variables might initially tend towards a narrow wrapped normal distribution while boundary effects are negligible. Eventually as more terms are added to the sum this wrapped normal distribution will broaden and approach a uniform distribution. This intuitive picture appears consistent with the time evolution of $\theta_i(t)$ shown in Figs. 20 and 21.

The wrapped normal variance estimates for $\theta_i(t)$ that are shown in Fig. 23 require further discussion. At intermediate times, the wrapped normal variance calculated from Eq. (30) rises linearly with a slope consistent with $M_N - \frac{3}{2}m_\pi$. This is not surprising because assuming an exactly wrapped normal $\theta_i(t)$, $M_\theta(t)$ becomes

$$
\begin{aligned}
M_\theta^{WN}(t) &= \ln\left[\frac{\langle e^{i\theta_i(t)} \rangle_{WN}}{\langle e^{i\theta_i(t+1)} \rangle_{WN}}\right] \\
&= -\frac{1}{2}[\sigma^2(\theta_i(t)) - \sigma^2(\theta_i(t+1))].
\end{aligned}
\tag{31}
$$

Equation (31) resembles the first nonzero term in the cumulant expansion given in Eq. (24) adapted for circular random variables. Results for $M_\theta^{WN}(t)$ are also shown in Fig. 23, where it is seen that $M_\theta^{WN}(t)$ is indistinguishable from $M_\theta(t)$ at small and intermediate times. In the noise region, both $M_\theta^{WN}(t)$ and standard estimates for $M_\theta(t)$ are consistent with zero. $M_\theta(t)$ has smaller variance than $M_\theta(t)$ in the noise region, but this large-time noise is the only
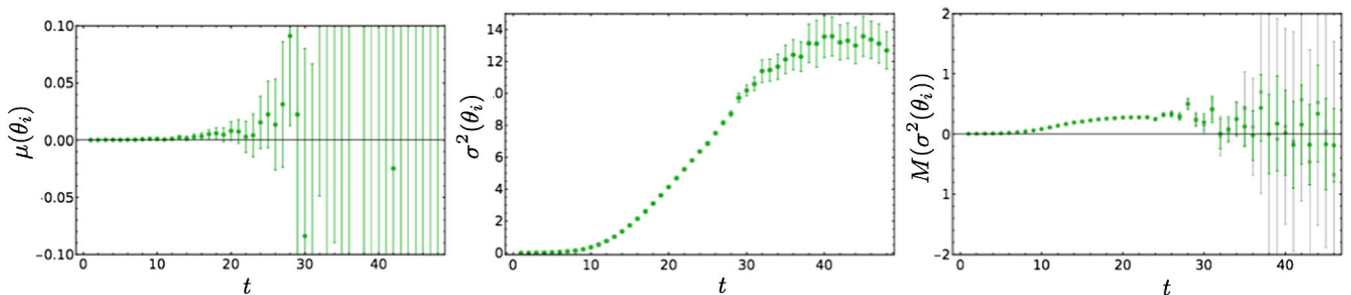

FIG. 23. The left panel shows estimates of the wrapped normal mean $\mu(\theta_i(t))$ calculated from Eq. (30) as a function of time. The center panel shows analagous estimates of the wrapped normal variance, $\sigma^2(\theta_i(t))$. The right panel shows the wrapped normal effective mass, $M_\theta^{WN}(t)$, defined in Eq. (31) (green points) along with the standard complex phase effective mass $M_\theta(t)$ defined in Eq. (24) (gray points).

visible signal of deviation between the two. This is not surprising, because $M_\theta^{WN}(t)$ is actually identical to $M_\theta(t)$ when $\bar{\mathcal{S}}(\theta(t)) = 0$. Since $\bar{\mathcal{S}}(\theta_i(t))$ vanishes in the infinite statistics limit by $\theta_i(t) \to -\theta_i(t)$ symmetry, $M_\theta^{WN}(t)$ must agree with $M_\theta(t)$ up to statistical noise. At large times $t \gtrsim 30$, the wrapped normal variance shown in Fig. 23 becomes roughly constant up to sizable fluctuations. The region where $\sigma^2(\theta_i(t))$ stops increasing coincides with the noise region previously identified.

The time at which the noise region begins depends on the size of the statistical ensemble N. Figure 24 shows estimates of $\sigma^2(\theta_i(t))$ from Eq. (30) for statistical ensemble sizes $N = 50, 5,000, 500,000$ varying across four orders of magnitude. The time of the onset of the noise region varies logarithmically as $t \sim 20, 27, 35$. The constant noise region value of $\sigma^2(\theta_i(t))$ is also seen to vary logarithmically with N. Equality of $M_\theta^{WN}(t)$ and $M_\theta(t)$ up to statistical fluctuations shows that $M_\theta(t)$ must be consistent with zero in the noise region. Since corrections to $M(t) \approx M_\theta(t) + M_R(t)$ from magnitude-phase correlations appear small at all times, it is reasonable to conclude that standard estimators for the nucleon effective mass are systematically biased in the noise region and that exponentially large increases in statistics are required to delay the onset of the noise region.

Besides these empirical observations, the inevitable existence and exponential cost of delaying the noise region can be understood from general arguments of circular statistics. The expected value of the sample concentration $\bar{\rho}^2$ can be calculated by applying Eq. (29) to an ensemble of

independent wrapped normal random variables θ_i in Eq. (27). The result shows that $\bar{\rho}^2$ is a biased estimate of $e^{-\sigma^2}$, and that the appropriate unbiased estimator is [25,27]

$$e^{-\sigma^2} = \frac{N}{N-1}\left(\bar{\rho}^2 - \frac{1}{N}\right). \tag{32}$$

For $\bar{\rho}^2 < 1/N$, Eq. (32) would lead to an imaginary estimate for σ^2 and therefore no reliable unbiased estimate can be extracted. A similar calculation shows that the expected variance of $\bar{\rho}^2$ is

$$\mathrm{Var}(\bar{\rho}^2) = \frac{N-1}{N^3}(1 - e^{-\sigma^2})^2[(1 - e^{-\sigma^2})^2 + 2Ne^{-\sigma^2}]. \tag{33}$$

In the limit of an infinitely broad distribution, all circular distributions tend towards uniform and the variance of $\bar{\rho}^2$ is set by the $\sigma^2 \to \infty$ limit of Eq. (33) regardless of the form of the true underlying distribution. When analyzing any very broad circular distribution, measurements of $\bar{\rho}^2$ will therefore include fluctuations on the order of $1/N$. For $e^{-\sigma^2} < 1/N$, the expected error from finite sample size effects in statistical inference based on $\bar{\rho}^2$ is therefore larger than the signal to be measured. In this regime $\bar{\rho}^2$ has both systematic bias and expected statistical errors that are larger than the value $e^{-\sigma^2}$ that $\bar{\rho}^2$ is supposed to estimate. $\bar{\rho}^2$ cannot provide accurate estimates of $e^{-\sigma^2}$ in this regime.

Inability to perform statistical inference in the regime $e^{-\sigma^2} < 1/N$ matters for the nucleon correlation function because $e^{-\sigma^2(\theta_i(t))} = \bar{\rho}^2(\theta_i(t)) = \langle\cos(\theta_i(t))\rangle^2$ and therefore $e^{-\sigma^2(\theta_i(t))}$ decreases exponentially with time. At large times there will necessarily be a noise region where $e^{-\sigma^2(\theta_i(t))} < 1/N$ is reached and $\bar{\rho}^2(\theta_i(t))$ is not a reliable estimator. Keeping $e^{-\sigma^2(\theta_i(t))}$ larger than the bias and expected fluctuations of $\bar{\rho}^2(\theta_i(t))$ requires

$$N > e^{\sigma^2(\theta_i(t))} \sim e^{2(M_N - \frac{3}{2}m_\pi)t}. \tag{34}$$

Equation (34) demonstrates that exponential increases in statistics are required to delay the time where statistical uncertainties and systematic bias dominate physical results estimated from $\bar{\rho}^2(\theta_i(t))$. Formally, the noise region can be defined as the region where Eq. (34) is violated. Lines at $\sigma^2(\theta_i(t)) = \ln N$ are shown on Fig. 24 for the ensembles with $N = 50, 5,000, 500,000$ shown. By this definition, the noise region formally begins once $\sigma^2(\theta_i(t))$ (extrapolated from reliable estimates in the golden window) crosses above the appropriate line. Excellent agreement can be seen between this definition and the above empirical characterizations of the noise region based on constant $\sigma^2(\theta_i(t))$ and unreliable effective mass estimates with constant errors.

Breakdown of statistical inference for sufficiently broad distributions is a general feature of circular distributions.
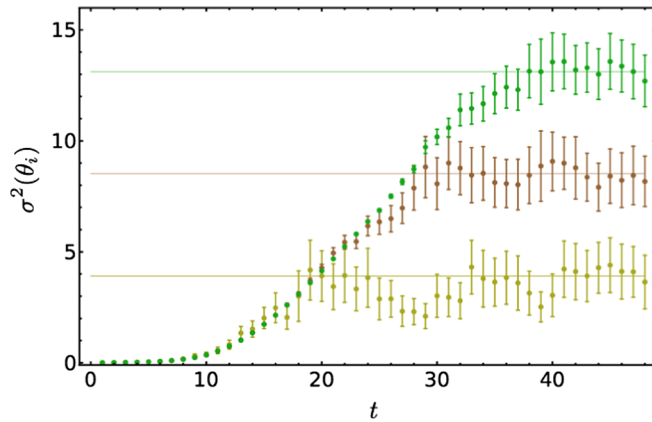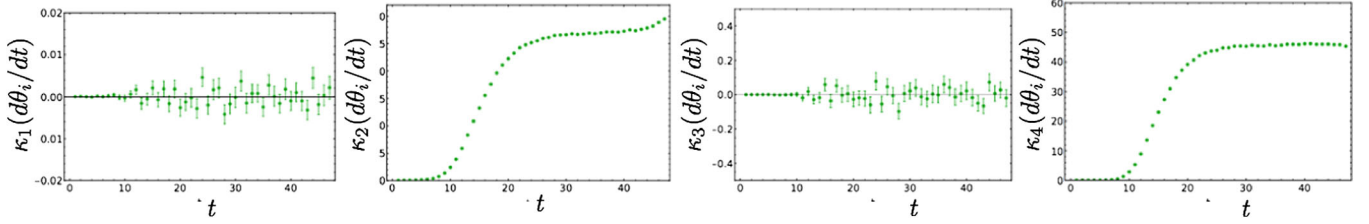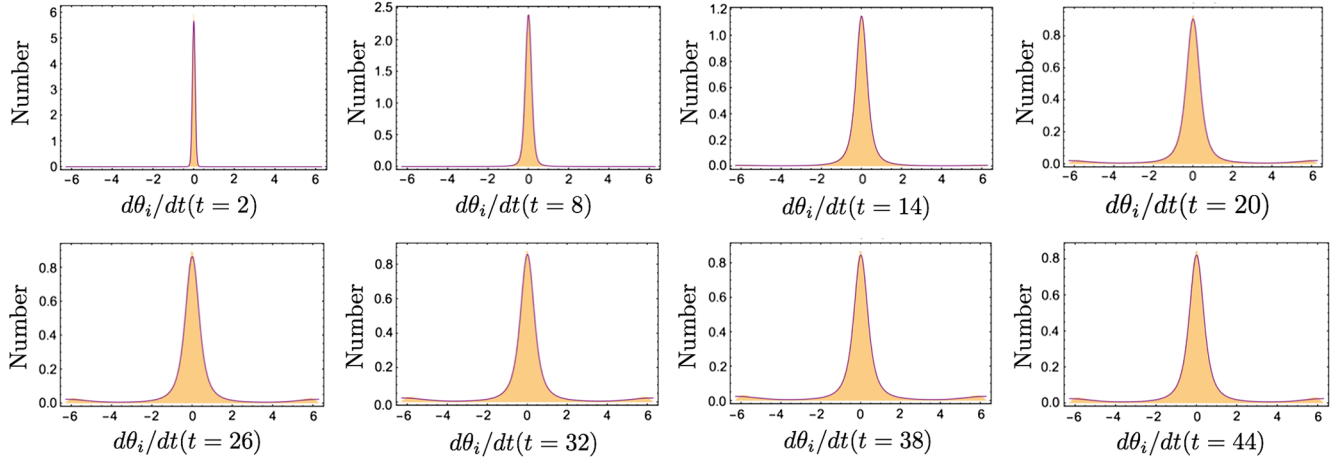

FIG. 24. Wrapped normal variance of the phase $\sigma^2(\theta_i(t))$ for statistical ensembles of various sizes. Results for an ensemble of $N = 50$ nucleon correlation functions are shown in yellow, $N = 5,000$ in brown, and $N = 500,000$ in green. Lines of each color are also shown at $\sigma^2(\theta_i(t)) = \ln(N)$. Above the relevant line, Eq. (34) is violated for each ensemble and measurements of $\sigma^2(\theta_i(t))$ are expected to be roughly equal to $\ln(N)$ instead of the underlying physical value of $\sigma^2(\theta_i(t))$. Estimates of $\sigma^2(\theta_i(t))$ reaching these lines mark the beginning of the noise region defined by violations of Eq. (34) for each ensemble.

FIG. 25. The first four cumulants of $\frac{d\theta_i}{dt}$.

FIG. 26. Histograms of $\frac{d\theta_i}{dt}$ with fits to a wrapped stable mixture distribution shown as the purple curves. See the main text for details.

Fisher notes that circular distributions are distinct from more familiar linear distributions in that "formal statistical analysis cannot proceed" for sufficiently broad distributions [25]. The arguments above do not rely on the particular form of the wrapped normal model assumed for $\theta_i(t)$, and the basic cause for the onset of the noise region for broad $\theta_i(t)$ is that $\bar{\rho}^2$ has an uncertainty of order $1/N$ for any broad circular distribution that begins approaching a uniform distribution.[7] Analogs of Eq. (34)

can be expected to apply to statistical estimation of the mean of any complex correlation function. As long as the asymptotic value of M_θ is known, Eq. (34) and analogs for other complex correlation functions can be used to estimate the required statistical ensemble size necessary to reliably estimate the mean correlation function up to a desired time t.

The pathological features of the large-time distribution of θ_i are not shared by $\frac{d\theta_i}{dt}$. As with the log-magnitude, it is useful to define general finite differences,

$$\Delta\theta_i(t, \Delta t) = \theta_i(t) - \theta_i(t - \Delta t), \quad (35)$$

and a discrete (lattice) time derivative,

$$\frac{d\theta_i}{dt} = \Delta\theta_i(t, 1). \quad (36)$$

The sample cumulants of $\frac{d\theta_i}{dt}$ are shown in Fig. 25, histograms of $\frac{d\theta_i}{dt}$ are shown in Fig. 26, and angular histograms are shown in Fig. 27. Much like $\frac{dR_i}{dt}$, $\frac{d\theta_i}{dt}$ appears to have a time-independent distribution at large times. While $\frac{d\theta_i}{dt}$ is a circular random variable, its distribution is still well localized at large times and can be clearly visually distinguished from a uniform distribution. This suggests that statistical inference of $\frac{d\theta_i}{dt}$ should be reliable in the noise region.
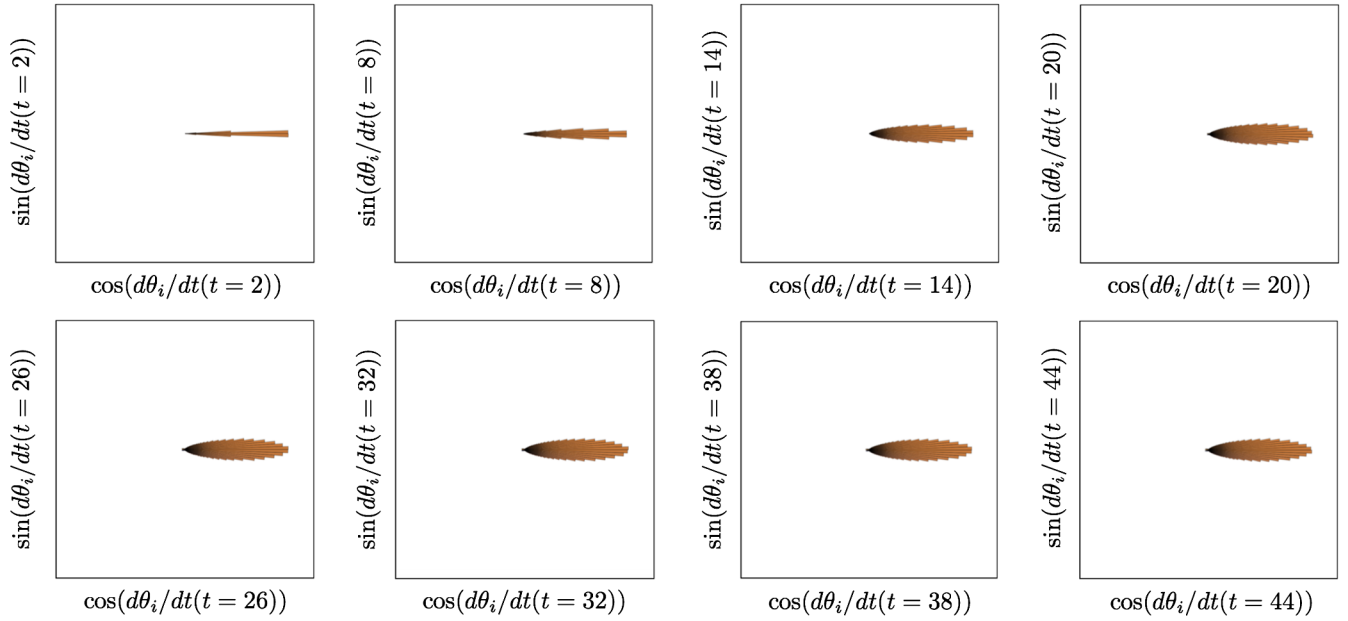
---

[7]One may wonder whether there is a more optimal estimator than $\bar{\rho}^2$ that could reliably calculate the width of broad circular distributions with smaller variance. While this possibility cannot be discarded in general, it is interesting to note that it can be in one model. The most studied distribution in one-dimensional circular statistics is the von Mises distribution, which has a simpler analytic form than the wrapped normal distribution. The von Mises distribution is also normally distributed in the limit of a narrow distribution, uniform in the limit of a broad distribution, and in general a close approximation but not identical to the wrapped normal distribution. Von Mises distributions provide fits of comparable qualitative quality to $\theta_i(t)$ as wrapped normal distributions. For the von Mises distribution, $\frac{N}{N-1}(\bar{\rho}^2 - \frac{1}{N})$ is an unbiased maximum likelihood estimator related to the width. By the Cramér-Rao inequality, a lower mean-squared error cannot be achieved if $\theta_i(t)$ is von Mises. Particularly in the limit of a broad distribution where all circular distributions tend towards uniform, it would be very surprising if an estimator could be found that satisfied this bound for the von Mises case but could reliably estimate the width of $\theta_i(t)$ in the noise region if a different underlying distribution is assumed.

FIG. 27.   Angular histograms of $\frac{d\theta_i}{dt}$. Since $\frac{d\theta_i}{dt}$ is defined on $-2\pi < \frac{d\theta_i}{dt} \leq 2\pi$, normalizations are such that $\frac{1}{2}\frac{d\theta_i}{dt}$ is mapped to the unit circle in analogy to Fig. 21.
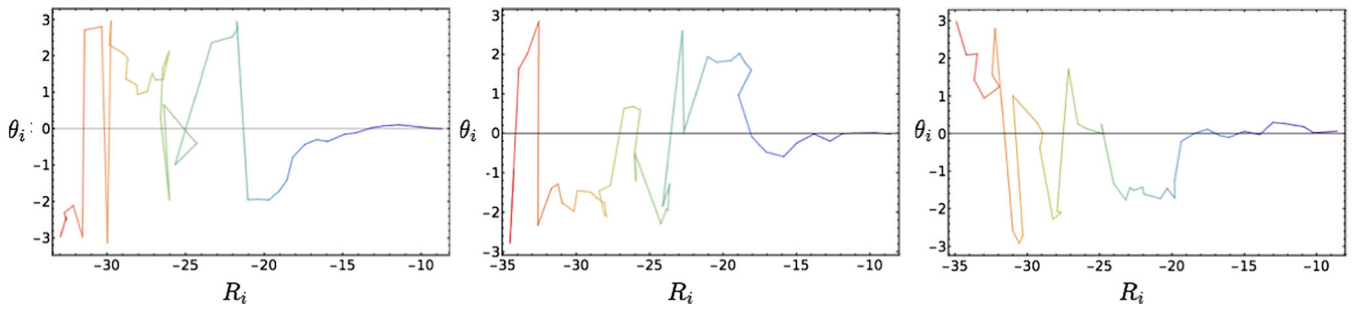

FIG. 28.   Time series showing $R_i(t)$ on the horizontal axis and $\theta_i(t)$ on the vertical axis for three individual nucleon correlation functions, where the color of the line shows the time evolution from violet at $t = 0$ to red at $t = 48$. The evolution of $R_i(t)$ shows a clear drift towards increasingly negative $R_i(t)$. Some large jumps where $\theta_i(t)$ changes by nearly $\pm 2\pi$ correspond to crossing the branch cut in $\theta_i(t)$. There are also sizable jumps where $\theta_i(t)$ changes by nearly $\pm\pi$ which likely do not correspond to crossing a branch cut.

Like $\frac{dR_i}{dt}$, $\frac{d\theta_i}{dt}$ shows evidence of heavy tails. The time evolution of $R_i(t)$ and $\theta_i(t)$ for three (randomly selected) correlation functions are shown in Fig. 28, and exhibit large jumps in both $R_i(t)$ and $\theta_i(t)$ more characteristic of Lévy flights than Brownian motion, leading us to consider stable distributions once again. Wrapped stable distributions can be constructed analogously to wrapped normal distributions as

$$\mathcal{P}_{WS}(\theta_i; \alpha, \beta, \mu, \gamma)$$

$$= \sum_{k=-\infty}^{\infty} \mathcal{P}_S(\theta_i + 2\pi k; \alpha, \beta, \mu, \gamma)$$

$$= \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \exp\left(i\mu n - |\gamma n|^\alpha \left[1 - i\beta \frac{n}{|n|}\tan(\pi\alpha/2)\right]\right),$$

$$(37)$$

where, as in Eq. (21), $\tan(\pi\alpha/2)$ should be replaced by $-\frac{2}{\pi}\ln|n|$ for $\alpha = 1$. This wrapped stable distribution is still not appropriate to describe $\frac{d\theta_i}{dt}$ for two reasons. First, $\frac{d\theta_i}{dt}$ describes a difference of angles and so is defined on a periodic domain $-2\pi < \frac{d\theta_i}{dt} \leq 2\pi$. This is trivially accounted for by replacing 2π by 4π in Eq. (37). Second, $\theta_i(t)$ is determined from a complex logarithm of $C_i(t)$ with a branch cut placed at $\pm\pi$. Whenever $\theta_i(t)$ makes a small jump across this branch cut, $\frac{d\theta_i}{dt}$ will be measured to be around 2π even though the distance traveled by $\theta_i(t)$ along its full Riemann surface is much smaller. This behavior results in the small secondary peaks near $\frac{d\theta_i}{dt} = \pm 2\pi$ visible in Fig. 26. This can be accommodated by fitting $\frac{d\theta_i}{dt}$ to a mixture of wrapped stable distributions peaked at zero and 2π. Since $\theta_i(t) \to -\theta_i(t)$ symmetry demands that both of these distributions are symmetric, a probability
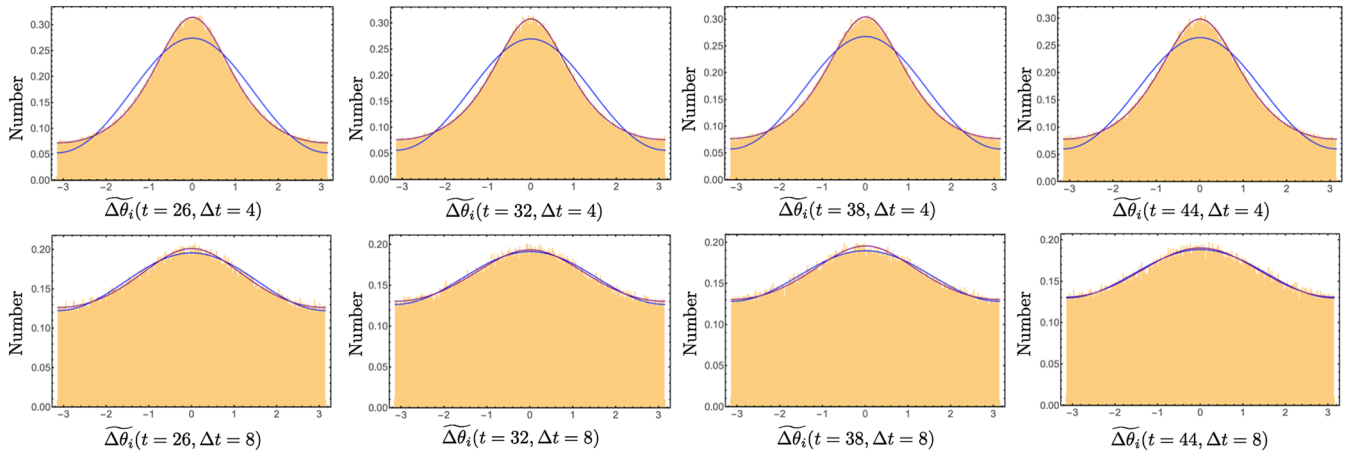
FIG. 29.   Histograms of $\widetilde{\Delta\theta_i}$ along with fits to wrapped normal distributions in blue and wrapped stable distributions in purple.

distribution able to accommodate all observed features of $\frac{d\theta_i}{dt}$ is given by the wrapped stable mixture distribution

$$
\tilde{\mathcal{P}}_{WS}(\theta_i; \alpha_1, \alpha_2, \gamma_1, \gamma_2, f)
$$

$$
= \frac{1}{4\pi} \left[1 + 2 \sum_{n=1}^{\infty} (1-f) e^{-|\gamma_1 n|^{\alpha_1}} \cos(n\theta_i) \right.
$$

$$
\left. + f e^{-|\gamma_2 n|^{\alpha_2}} \cos(n(\theta_i - 2\pi)) \right], \tag{38}
$$

where f represents the fraction of $\frac{d\theta_i}{dt}$ data in the secondary peaks at $\frac{d\theta_i}{dt} = \pm 2\pi$ representing branch cut crossings. Fits of $\frac{d\theta_i}{dt}$ to this wrapped stable mixture model performed with maximum likelihood estimation are shown in Fig. 26 and are in good qualitative agreement with the LQCD results.

If the widths of the main and secondary peaks in $\frac{d\theta_i}{dt}$ were sufficiently narrow, it would be possible to unambiguously associate each $\frac{d\theta_i}{dt}$ measurement with one peak or the other and "unwrap" the trajectory of $\theta_i(t)$ across its full Riemann surface by adding $\pm 2\pi$ to measured values of $\frac{d\theta_i}{dt}$ whenever the branch cut in $\theta_i(t)$ is crossed. This should become

increasingly feasible as the continuum limit is approached. However, the presence of heavy tails in the $\frac{d\theta_i}{dt}$ primary peak prevents unambiguous identification of branch cut crossings in the LQCD correlation functions considered here. Due to the power-law decay of the primary peak, there is no clear separation visible between the main and secondary peaks, and in particular, points near $\frac{d\theta_i}{dt} = \pm\pi$ cannot be unambiguously identified with one peak or another.

For descriptive analysis of $\frac{d\theta_i}{dt}$, it is useful to shift the secondary peak to the origin by defining

$$
\widetilde{\Delta\theta_i} = \text{Mod}(\Delta\theta_i + \pi, 2\pi) - \pi. \tag{39}
$$

$\widetilde{\Delta\theta_i}$ is well described by the wrapped stable distribution of Eq. (37). Histograms of the large-time behavior of $\widetilde{\Delta\theta_i}$ are shown in Fig. 29 for $\Delta t = 4$, 8 and fits of the index of stability of $\widetilde{\Delta\theta_i}$ are shown in Fig. 30. The large-time distribution of $\widetilde{\Delta\theta_i}(t, \Delta t)$ appears time independent for all Δt. Heavy tails are visible at all times, even as Δt becomes large. The large Δt behavior visible here is consistent with a wrapped Cauchy distribution. The estimated index of
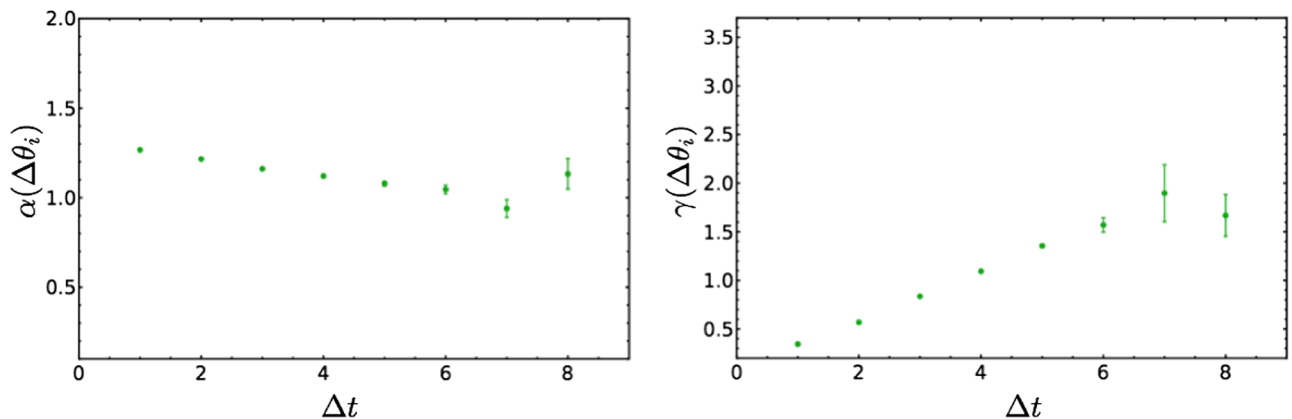

FIG. 30.   Maximum likelihood estimates for the wrapped stable index of stability $\alpha(\widetilde{\Delta\theta_i})$ (left) and width $\gamma(\widetilde{\Delta\theta_i})$ (right) extracted from the large-time plateau region as functions of Δt.

stability of $\widetilde{\Delta\theta_i}$ differs significantly from that of ΔR, and for $\Delta t = 1$, the large-time behavior is found to have

$$\alpha(\widetilde{\Delta\theta_i}(t \to \infty, \Delta t \sim 0.12 \text{ fm})) \to 1.267(4)(1). \quad (40)$$

This result is consistent with maximum likelihood estimates of $\alpha_1(\frac{d\theta_i}{dt})$ in the wrapped stable mixture model of Eq. (38). $\alpha_2(\frac{d\theta_i}{dt})$, associated with the peak shifted from $\theta_i = \pm\pi$ in the wrapped stable mixture model, cannot be reliably estimated from the available LQCD correlation functions. The continuum limit index of stability of $\frac{d\theta_i}{dt}$ cannot be determined without additional LQCD studies at finer lattice spacings.

As seen in Fig. 30, the large-time width of $\widetilde{\Delta\theta_i}(t, \Delta t)$ increases with increasing Δt. This behavior is shared by $\Delta\theta_i(t, \Delta t)$. In accordance with the observations above that the wrapped normal variance of $\theta_i(t)$ increases linearly with t, the constant large-time wrapped normal variance of $\Delta\theta_i(t, \Delta t)$ increases linearly with Δt. This is consistent with a picture of $\Delta\theta_i(t, \Delta t)$ as the sum of Δt single time step differences, $\frac{d\theta_i}{dt}$, that make roughly equal contributions to $\Delta\theta_i(t, \Delta t)$. In accordance with the scaling $\sigma^2(\theta_i(t)) \sim (M_N - \frac{3}{2} m_\pi)t$ discussed previously, this linear scaling gives $\sigma^2(\Delta\theta_i(t, \Delta t)) \sim 2(M_N - \frac{3}{2} m_\pi)\Delta t$.

We summarize our observations on the phase of $C(t)$:

(i) The phase of the nucleon correlation function is described by an approximately wrapped normal distribution whose width increases with time. At small times the distribution is narrow and resembles a normal distribution. At large times the distribution becomes broad compared to the 2π range of definition of $\theta_i(t)$ and resembles a uniform distribution.

(ii) The phase effective mass $M_\theta(t)$ appears to plateau to a value close to $M_N - 3/2m_\pi$. Since $|e^{i\theta_i(t)}|^2 = 1$ is time independent by construction, this nonzero asymptotic value of M_θ implies that $\theta_i(t)$ has a severe StN problem.

(iii) $M_\theta(t)$ can be determined from the time derivative of the wrapped normal variance of $\theta_i(t)$ in analogy to the cumulant expansion. The effective mass extracted from growth of the wrapped normal variance is identical to $M_\theta(t)$ up to statistical fluctuations. This leads to scaling of the wrapped normal variance of $\theta_i(t)$ consistent with $\sigma^2(\theta_i(t)) \sim 2(M_N - \frac{3}{2} m_\pi)t$.

(iv) Standard estimators for the wrapped normal variance have a systematic bias and for a sufficiently broad distribution the minimum expected statistical uncertainty is set by finite sample size $1/N$ effects. Once the wrapped normal variance becomes larger than $\ln N$, finite sample size fluctuations become larger than the signal required to extract $M_\theta(t)$. Since the width of $\theta_i(t)$ increases with time, a region where finite sample size errors prevent reliable extractions of $M_\theta(t)$ will inevitably occur at sufficiently large times. This is the noise region

empirically identified above. Standard effective mass estimates are systematically biased in the noise region. Exponentially large increases in statistics are necessary to delay the onset of the noise region.

(v) Finite differences, $\Delta\theta_i(t, \Delta t)$, are described by time-independent distributions at large times. $\Delta\theta_i$ is heavy tailed for all Δt considered here, and $\frac{d\theta_i}{dt}$ is well described by a wrapped stable mixture distribution. Further studies will be needed to understand the continuum limit of the index of stability of $\frac{d\theta_i}{dt}$.

IV. AN IMPROVED ESTIMATOR

The preceding observations suggest that difficulties in statistical analysis of nucleon correlation functions arise from difficulties in statistical inference of $\theta_i(t)$. The same exponentially hard StN and noise region problems obstruct large-time estimation of the wrapped normal variance of $\theta_i(t)$ and of $M(t)$. Conversely, the width of $\Delta\theta_i(t, \Delta t)$ distributions does not increase with time, and there is no StN problem impeding statistical inference of $\Delta\theta_i(t, \Delta t)$. This suggests that it would be preferable to construct an effective mass estimator relying on statistical inference of $\Delta\theta_i(t, \Delta t)$.

First consider the magnitude for simplicity. The mean correlation function magnitude can be expressed in terms of ΔR_i as

$$\langle e^{R_i(t)} \rangle = \left\langle \exp\left(R_i(0) + \sum_{t'=1}^{t} \frac{dR_i}{dt}\bigg|_{t'} \right) \right\rangle$$

$$= \left\langle \exp\left(R_i(0) + \sum_{t'=1}^{t-\Delta t} \frac{dR_i}{dt}\bigg|_{t'} \right) \exp\left(\sum_{t'=t-\Delta t+1}^{t} \frac{dR_i}{dt}\bigg|_{t'} \right) \right\rangle$$

$$= \langle e^{R_i(0)+\Delta R_i(t-\Delta t, t-\Delta t)} e^{\Delta R_i(t, \Delta t)} \rangle. \quad (41)$$

The last expression above shows that $e^{R_i(t)}$ can be expressed as a product of two factors involving the evolution of $R_i(t)$ in the regions $[0, t - \Delta t]$ and $[t - \Delta t, t]$ respectively. Because QCD has a finite correlation length, these two factors should be approximately decorrelated. Correlations should only arise from contributions involving points near the boundary at $t - \Delta t$. At large times, t can be assumed to be much larger than Δt and than any QCD correlation length, so boundary effects can be assumed to be negligible for the first region. Boundary effects cannot be neglected for the smaller region of length Δt. Treating these boundary effects as a systematic uncertainty allows the correlation function to be factorized between the regions $[0, t - \Delta t]$ and $[t - \Delta t, t]$ as

$$\langle e^{R_i(t)} \rangle = \langle e^{R_i(0)+\Delta R_i(t-\Delta t, t-\Delta t)} \rangle \langle e^{\Delta R_i(t, \Delta t)} \rangle [1 + O(e^{-\delta E_R \Delta t})]. \quad (42)$$

where δE_R is the smallest energy scale responsible for nontrivial correlations between the factors on the rhs

associated with $[0, t - \Delta t]$ and $[t - \Delta t, t]$, and terms suppressed by $e^{-\delta E_R(t-\Delta t)}$ are neglected.[8] If both factors on the rhs of Eq. (42) only receive contributions from the ground state and have single-exponential time evolution, then the product of the independently averaged factors on the rhs has the same single-exponential behavior as the lhs. If excited states make appreciable contributions to either factor on the rhs, then the product of sums of exponentials representing multistate evolution over $[0, t - \Delta t]$ and $[t - \Delta t, t]$ respectively will not exactly equal the sum of exponentials representing multistate evolution over $[0, t]$.

$e^{R_i(t+1)}$ can similarly be split into an approximately decorrelated product. Performing this split with regions $[0, t - \Delta t]$ and $[t - \Delta t, t + 1]$ gives

$$\langle e^{R_i(t+1)} \rangle = \langle e^{R_i(0)+\Delta R_i(t-\Delta t, t-\Delta t)} \rangle \langle e^{\Delta R_i(t+1, \Delta t+1)} \rangle$$
$$\times [1 + O(e^{-\delta E_R \Delta t})]. \tag{43}$$

The common term in both expressions cancels when constructing the magnitude effective mass, allowing us to define

$$\tilde{M}_R(t, \Delta t) = \ln \left[\frac{\langle e^{\Delta R_i(t, \Delta t)} \rangle}{\langle e^{\Delta R_i(t+1, \Delta t+1)} \rangle} \right]$$
$$= M_R(t) + O(e^{-\delta E_R \Delta t}). \tag{44}$$

Identical steps can be applied to the phase, leading to

$$\tilde{M}_\theta(t, \Delta t) = \ln \left[\frac{\langle e^{i\Delta\theta_i(t, \Delta t)} \rangle}{\langle e^{i\Delta\theta_i(t+1, \Delta t+1)} \rangle} \right]$$
$$= M_\theta(t) + O(e^{-\delta E_\theta \Delta t}). \tag{45}$$

The same steps can also be applied to the full correlation function $C_i(t) = e^{R_i(t)+i\theta_i(t)}$. Noting that

$$e^{\Delta R_i(t, \Delta t)+i\Delta\theta_i(t, \Delta t)} = \frac{C_i(t)}{C_i(t - \Delta t)}, \tag{46}$$

the analogous relation for the full effective mass takes the simple form

$$\tilde{M}(t, \Delta t) = \ln \left[\frac{\langle C_i(t)/C_i(t - \Delta t) \rangle}{\langle C_i(t+1)/C_i(t - \Delta t) \rangle} \right]$$
$$= M(t) + O(e^{-\delta E \Delta t}). \tag{47}$$

The correlation function ratio effective mass estimator $\tilde{M}(t, \Delta t)$ has different statistical properties than the traditional effective mass $M(t)$ when Δt is treated as an independent t. Note that although Δt appears in the numerator and denominator of correlation function ratios superficially similarly to t_J in Eq. (6), these two parameters induce quite different statistical behavior. $C_i(t + 1)$ in Eq. (47) could be replaced by $C_i(t + t_J)$ (with an appropriate $1/t_J$ overall normalization added). Taking $t_J > 1$ increases the time separation between $C_i(t - \Delta t)$ and $C_i(t + t_J)$ in the correlator ratio in the denominator of Eq. (47), resulting in larger statistical uncertainties in effective mass results, and will not be pursued further here.

The approximate factorization leading to Eq. (47) can be understood from a quantum field theory viewpoint without reference to the magnitude and phase individually. Inserting a complete set of states in a correlation function at $t - \Delta t$ allows the correlation function to be expressed as a sum of exponentials $e^{-E_n \Delta t}$ times prefactors representing the amplitude for the system being in the nth state at time $t - \Delta t$. These prefactors for each $e^{-E_n \Delta t}$ term are proportional to $e^{-E_n(t-\Delta t)}$, enhancing the amplitude for finding the system in its ground state at large $t - \Delta t$. In this way, the contribution to the correlation function from the region $[0, t - \Delta t]$ can be thought of as an effective source for the correlation function in the region $[t - \Delta t, t]$ whose ground-state overlap is dynamically improved compared to the overlap of the original source at time zero. The prefactors for each $e^{-E_n \Delta t}$ will depend on the structure of this effective source, but the exponents are fixed by the QCD spectrum. The factor of $C_i(t - \Delta t)^{-1}$ in Eq. (47) can be considered to be a modification of the effective source in the region $[0, t - \Delta t]$. The presence of $C_i(t - \Delta t)^{-1}$ will modify the prefactor of each $e^{-E_n \Delta t}$ term, but it should not affect time evolution of the system in the region $[t - \Delta t, t]$. This suggests that an effective mass designed to extract the ground-state energy from the sum of $e^{-E_n \Delta t}$ terms, as in Eq. (47), should provide the exact ground-state mass at large Δt up to corrections arising from excited-state contributions to the $e^{-E_n \Delta t}$ sum. These corrections should decrease exponentially with increasing Δt at a rate set by the energy gap between the ground and first excited state in the system of interest. The size of this energy gap will be set by the lowest-lying excitation consistent with the quantum numbers of the system, a derivatively coupled pion for the case of the nucleon,[9] leading to the expectation $\tilde{M}(t, \Delta t) = M(t) + O(e^{-m_\pi \Delta t})$.

It is not straightforward to construct a representation of $C_i(t - \Delta t)^{-1}$ in terms of local quark and gluon operators that

---

[8]It is not proven that the magnitude of a correlation function can be expressed as a sum of exponentials; however, the square of the magnitude contributes to the variance correlation function and must have a spectral representation as a sum of exponentials. Results of Sec. III demonstrate numerically that the magnitude decays exponentially at large times with a ground-state energy equal to half the ground-state energy of the variance correlation function. Equation (44), which further supposes exponential magnitude excited-state contamination, is investigated numerically below; see Fig. 31.

[9]Multihadron correlation functions contain additional low-lying excitations that may introduce larger correlation lengths than m_π^{-1} associated for instance with near-threshold bound states. Such multihadron systems are outside the scope of this work.
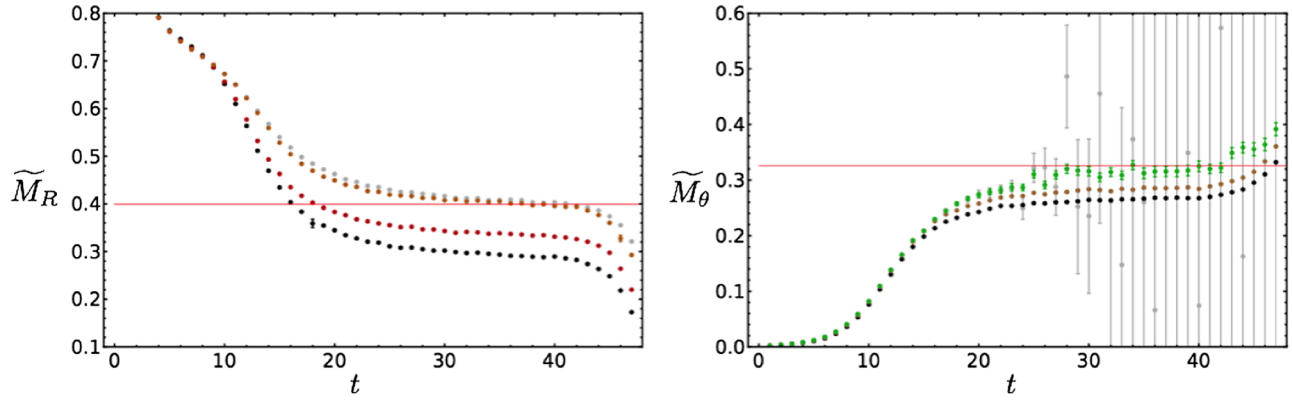
FIG. 31.   Results for the correlation-function-ratio-based estimators $\tilde{M}_R(t, \Delta t)$ and $\tilde{M}_\theta(t, \Delta t)$ with $\Delta t = 1, 2, 8$. The left panel shows results for $m_R(t, \Delta t)$ with $\Delta t = 1$ in black, $\Delta t = 2$ in red, and $\Delta t = 8$ in orange. The standard estimator $m_R(t)$ is shown in gray, and a red line is shown for reference at $\frac{3}{2} m_\pi$. The right panel shows results for $m_\theta(t, \Delta t)$ with $\Delta t = 1$ in black, $\Delta t = 2$ in brown, and $\Delta t = 8$ in green. The standard estimator $m_\theta(t)$ is shown in gray and a red line is shown for reference at $M_N - \frac{3}{2} m_\pi$.

would allow a rigorous proof of these statements, and so numerical LQCD calculations are used to investigate the validity of Eq. (47). Exponential reduction of systematic error is numerically demonstrated, but at a faster rate than m_π^{-1}. This suggests that the structure of the effective source plays an important role in determining which $e^{-E_n \Delta t}$ terms are appreciable at the large but finite Δt accessible to LQCD calculations in the same way that the structure of the source at time zero determines which excited states make appreciable contributions to the standard effective mass at small t.

The LQCD results for $\tilde{M}_R(t, \Delta t)$ and $\tilde{M}_\theta(t, \Delta t)$ with $\Delta t = 1, 2, 8$ are shown in Fig. 31, and results for $\tilde{M}(t, \Delta t)$ are shown in Fig. 32. The statistical uncertainties associated with $\tilde{M}(t, \Delta t)$ are the same as those of $M(t)$ within the golden window, but at large times they become constant in time rather than exponentially increasing. This is in accord with our observations about the form of the statistical

distributions associated with $\Delta R_i(t, \Delta t)$ and $\Delta \theta_i(t, \Delta t)$, which, up to small magnitude-phase correlations, indicate that

$$\mathrm{Var}(\tilde{M}(t, \Delta t)) \sim \frac{\mathrm{Var}(e^{R_i(t, \Delta t) + i\theta_i(t, \Delta t)})}{\langle e^{R_i(t, \Delta t) + i\theta_i(t, \Delta t)} \rangle^2} \sim e^{2(M_N - \frac{3}{2} m_\pi) \Delta t}. \quad (48)$$

The statistical uncertainties associated with $\tilde{M}(t, \Delta t)$ are constant in t, although they do increase exponentially with increases in Δt. Since $\Delta \theta_i(t, \Delta t)$ has constant width at large times, the inevitable onset of the noise region where statistical inference fails for $\theta_i(t)$ can be avoided. The constraint required for reliable statistical inference of $\tilde{M}(t, \Delta t)$ at large times is that the wrapped normal variance of $\Delta \theta_i(t, \Delta t)$ can be extracted without large finite sample size errors. This constraint can be expressed as a bound on the statistical sample size required for a particular choice of Δt,

$$N > e^{\sigma^2(\Delta \theta_i(t, \Delta t))} \sim e^{2(M_N - \frac{3}{2} m_\pi) \Delta t}. \quad (49)$$

The statistical uncertainties of $\tilde{M}(t, \Delta t)$ determined from the LQCD correlation functions are shown in Fig. 33, from which it can be seen that they become constant at large times for all fixed Δt. For small and moderately large values of $\Delta t = 1, 7, 15$, the expected exponential increase in large-time statistical uncertainties is observed, consistent with Eq. (49). Once Eq. (49) is violated, exponential scaling of statistical uncertainties with Δt ceases. For $\Delta t \lesssim \frac{\ln(N)}{2(M_N - \frac{3}{2} m_\pi)}$, the relative statistical uncertainty in $\tilde{M}(t, \Delta t)$ compared to $\tilde{M}(t, \Delta t = 1)$ is approximately equal to N rather than $e^{2(M_N - \frac{3}{2} m_\pi)(\Delta t - 1)}$.[10] This is seen in Fig. 33 in the large-time behavior of the standard effective mass.
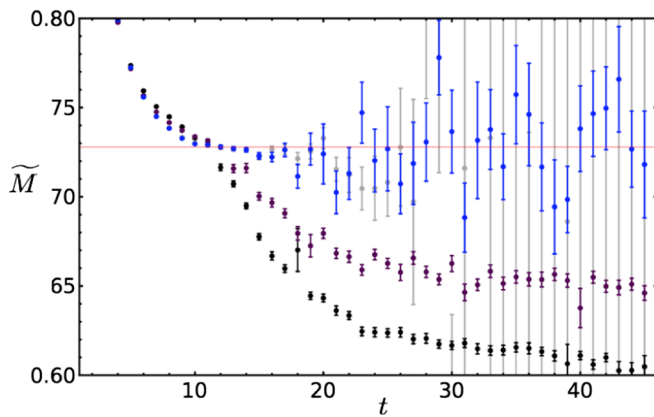

FIG. 32.   Results for the correlation-function-ratio-based estimator $\tilde{M}(t, \Delta t)$. The left panel shows results with $\Delta t = 1$ in black, $\Delta t = 2$ in purple, and $\Delta t = 8$ in blue, along with the traditional effective mass estimator $M(t)$ shown in gray and a red line at M_N shown for reference.

---

[10]These bounds only indicate scaling with N. To be made more precise, proportionality constants can be computed using the scaling indicated in Eq. (49).
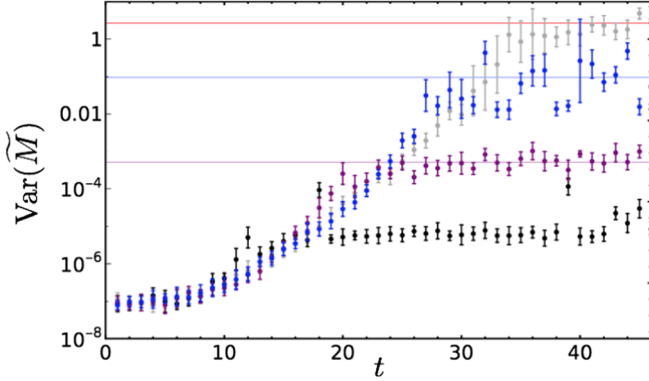
FIG. 33. Variance in the estimates of $\tilde{M}(t, \Delta t)$ as a function of time t for various choices of Δt. The black points show $\Delta t = 1$, the purple show $\Delta t = 7$, and the blue show $\Delta t = 15$. The gray points show uncertainties in the standard effective mass estimator equivalent to $\Delta t = t$. The purple and blue lines show the expected large-time variance of $\tilde{M}(t, \Delta t)$ with $\Delta t = 7, 15$ predicted by Eq. (48) with the overall normalization fixed by the $\Delta t = 1$ case. The red line shows the bound of Eq. 33 with overall normalization again fixed by the $\Delta t = 1$ case. Breakdown of statistical inference of broad circular distributions predicts that the large-time variance of $\tilde{M}(t, \Delta t)$ will not systematically rise above the red line for any Δt.

When Eq. (49) is violated, $\Delta\theta_i(t, \Delta t)$ cannot be reliably estimated at large times and increasing Δt does not improve the accuracy of $\tilde{M}(t, \Delta t)$. The standard effective mass estimator can be thought of as evolving with $t \sim \Delta t$, and will become unreliable because of finite sample size effects
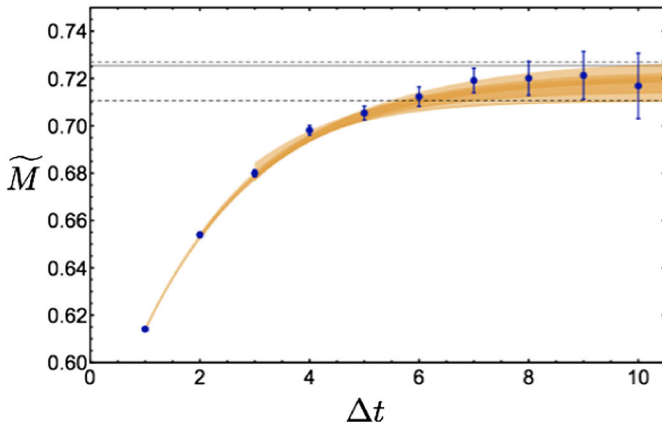
at large times scaling as $t \gtrsim \ln N / (2(M_N - \frac{3}{2}m_\pi))$. Similarly, our improved effective mass becomes unreliable for $\Delta t \gtrsim \ln N / (2(M_N - \frac{3}{2}m_\pi))$. In this extreme case, the bias associated with neglected correlations in $\tilde{M}(t, \Delta t)$ becomes less important than the bias associated with statistical inference of overly broad circular random variables. Exponential growth of statistical uncertainties with Δt suggests that smaller choices of Δt where Eq. (49) holds likely lead to smaller overall statistical plus systematic uncertainties.

The systematic bias of $\tilde{M}(t, \Delta t)$ can be explored through calculations at various Δt. Figure 34 shows results for $\Delta t = 1, ..., 9$. For $\Delta t \gtrsim 7$, results for the $\tilde{M}(t, \Delta t)$ fit during the large-time noise region $25 \leq t \leq 40$ are statistically consistent with fits extracted from the golden window $15 \leq t \leq 25$. Late-time fits with $\tilde{M}(t, \Delta t)$ have larger statistical uncertainties than golden window fits. More precise fits than either could be made by including both the golden window and the noise region in fits of $\tilde{M}(t, \Delta t)$. There is only a minor advantage in including the noisier large-time points in fits that include a precise golden window, and this exploratory work does not aim for a more precise extraction of the nucleon mass. Practical advantages of large-time fits of $\tilde{M}(t, \Delta t)$ compared to golden window fits of $\tilde{M}(t)$ are more likely to be found in systems where a reliable golden window cannot be unambiguously identified. Large-time fits of $\tilde{M}(t, \Delta t)$ would also be more advantageous for lattices with larger time directions.
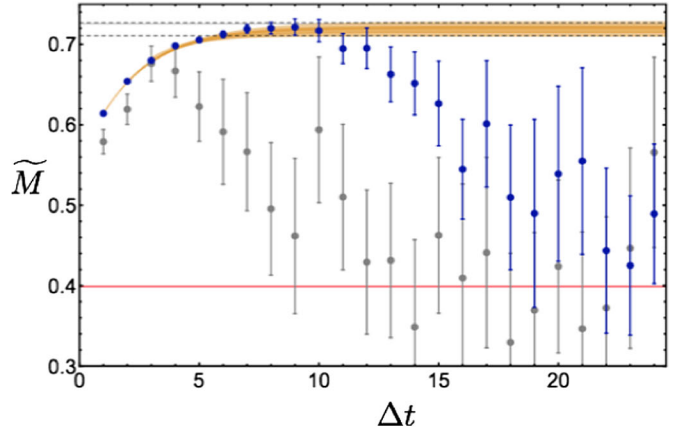

FIG. 34. In both the left and right panels, results for $\tilde{M}(\Delta t)$ taken from correlated χ^2-minimization fits of $\tilde{M}(t, \Delta t)$ to a constant in the region $25 \leq t \leq 40$ with fixed Δt are shown as blue points. The tan bands show the results of correlated χ^2-minimization fits of $\tilde{M}(t, \Delta t)$ in various rectangles of t and Δt to the three-parameter (constant plus exponential) form shown in Eq. (50). The three light-brown bands all use data from $25 \leq t \leq 40$ and then $1 \leq \Delta t \leq 10$, $2 \leq \Delta t \leq 10$, and $3 \leq \Delta t \leq 10$. The black dashed lines show the extrapolated prediction for the nucleon mass including statistical errors from the $2 \leq \Delta t \leq 10$ fit added in quadrature with a systematic error calculated as half the maximum difference in central values given by the three fits shown. The horizontal gray bands show $M_N \pm \delta M_N$ from the precision NPLQCD calculation of Ref. [28], which used a high-statistics ensemble of correlation functions with optimized sources generated on the same gauge configurations used here. The right panel shows a much larger range of Δt and also includes results calculated with a smaller ensemble of $N = 5,000$ correlation functions as gray points. Deviations from the asymptotic prediction due to finite statistics are clearly visible and lead to incorrect results at much earlier Δt in the smaller ensemble.

Results for a range of Δt shown in Fig. 34 can also be used to fit the systematic bias in $\tilde{M}(t, \Delta t)$ and formally extrapolate to the unbiased $\Delta t \to t \to \infty$ result. During the development of a refined version of this improved estimator [48], it was realized that the parametric form of the bias can be deduced by considering a decomposition of $[0, t]$ into an extended "source region" $[0, t - \Delta t]$ involving $C_i(t)$ and $C_i^{-1}(t - \Delta t)$ and an "evolution region" $[t - \Delta t, t]$ only involving $C_i(t)$. Standard QCD time evolution should apply after the boundary of the source region at $t - \Delta t$, and so at large Δt correlation function ratios should scale with $\sim e^{-M_N \Delta t}$ relative to their $t - \Delta t$ boundary values. Corrections to this ground-state scaling will arise from excited states, which will make contributions to $\langle C_i(t) C_i^{-1}(t - \Delta t) \rangle$ scaling as $\sim e^{-(M_N + \delta E)\Delta t}$, where δE is the gap between the nucleon ground and first excited-state energies. This allows the dominant contribution to the bias in $\tilde{M}(t, \Delta t)$ to be parametrized as

$$\tilde{M}(t \to \infty, \Delta t) = \ln\left[\frac{e^{-M_N \Delta t}(1 + ce^{-\delta E \Delta t} + \cdots)}{e^{-M_N(\Delta t + 1)}(1 + ce^{-\delta E(\Delta t + 1)} + \cdots)}\right]$$
$$= M_N + c\delta E e^{-\delta E \Delta t} + \cdots, \qquad (50)$$

where c is the ratio of excited- to ground-state overlaps produced by the effective boundary at $t - \Delta t$. At sufficiently light quark masses and large Δt, this excited-state gap will be set by m_π. However, it is noteworthy that Eq. (47) involves products of momentum-projected unaveraged correlation functions. It is familiar from studies of two-baryon correlation functions formed from products of momentum-projected one-baryon blocks that summing over all points in the spatial volume separately for each factor in a product leads to a suppression by $O(m_\pi^{-3} V^{-1})$ in the fraction of points in the product where the nucleons are within one-pion Compton wavelength of one another. It is expected that correlations between $\tilde{C}_i(t)$ and $C_i^{-1}(t - \Delta t)$ described by one-pion excitations will be similarly volume suppressed. The dominant excited-state bias is then expected to arise from excitations that could be produced throughout the lattice volume at the boundary of the source region. Such excitations are generically far from the nucleon and any other sources of conserved charge, so they should have quantum numbers of the vacuum. The dominant excited-state bias contributing to Eq. (50) is therefore expected to be $e^{-M_\sigma \Delta t}$, where M_σ is the mass of the σ-meson, the lightest excited state with quantum numbers of the vacuum. Performing a correlated χ^2-minimization three-parameter fit of $\tilde{M}(t, \Delta t)$ to the constant plus exponential form shown in Eq. (50) for noise region data $25 \leq t \leq 40$ gives

$$M_N = 0.7192(49)(42), \qquad c = -0.358(26)(17),$$
$$\delta E = 0.512(65)(73), \qquad\qquad\qquad (51)$$

where the first uncertainty is the statistical uncertainty and the second uncertainty is a measure of systematic uncertainty taken from the variation in the central value of the fit as the fitting range in Δt varies. The extrapolated result in Eq. 34 agrees within uncertainties with the intermediate-time plateau result $M_N = 0.7253(11)(22)$ and with the high-precision GW result $M_N = 0.72546(47)(31)$ of Ref. [28]. For the extrapolated large-time result, the total statistical and systematic uncertainty in quadrature is $\delta M_N = 0.0064$, which is larger than the total uncertainty of the plateau region determination $\delta M_N = 0.0025$. The large-time plateau considered effectively comprises a two-dimensional region $1 \leq \Delta t \leq 10$ and $25 \leq t \leq 40$ with 150 points. The value of the χ^2-minimization fit to this two-dimensional region is most sensitive to points with smaller Δt and therefore exponentially smaller uncertainties but is equally sensitive to points with all t that are expected to be approximately decorrelated over intervals $t \gtrsim m_\pi^{-1}$. The intermediate-time plateau region $10 \leq t \leq 25$ includes 15 points that are expected to be approximately decorrelated over intervals $t \gtrsim m_\pi^{-1}$. The value of the standard effective mass fit is most sensitive to points with smaller t and therefore exponentially smaller uncertainties, though the variance correlation function is not dominated by the three-pion ground state until $t \gtrsim 20$. This indicates that results from the intermediate-time plateau have smaller point-by-point uncertainties than points from the large-time noise region. The total uncertainty of the noise region result could be reduced by increasing the length of the lattice time direction, while the length of the smaller-time plateau available to standard estimators is restricted by the StN problem. The proof-of-principle calculation presented here demonstrates that accurate results can be extracted from the noise region. It remains to be seen in future calculations of single- and multibaryon systems optimized for large-time analysis whether the methods introduced in this work can be used to achieve significantly higher precision with the same resource budget as calculations optimized for smaller-time analysis.

The best-fit excitation scale $\delta E = 866(110)(124)$ MeV in Eq. (34) can be compared with the σ-meson mass extracted from mesonic sector calculations to test the heuristic arguments above that lighter excitations will make volume-suppressed contributions. Calculations of the σ-meson face a severe StN problem, particularly at light quark masses where the σ-meson describes a broad $\pi\pi$ isoscalar resonance rather than a compact QCD bound state, but a recent calculation by the Hadron Spectrum Collaboration has precisely determined $M_\sigma = 758(4)$ MeV at $m_\pi \sim 391$ MeV where the σ-meson is weakly bound [49]. Similarly precise results at slightly higher quark masses are not available for interpolation to $m_\pi \sim 450$ MeV, but a crude extrapolation can be made using the Hadron Spectrum result and the (real part of the) physical position of the σ-meson pole obtained from dispersive analysis of
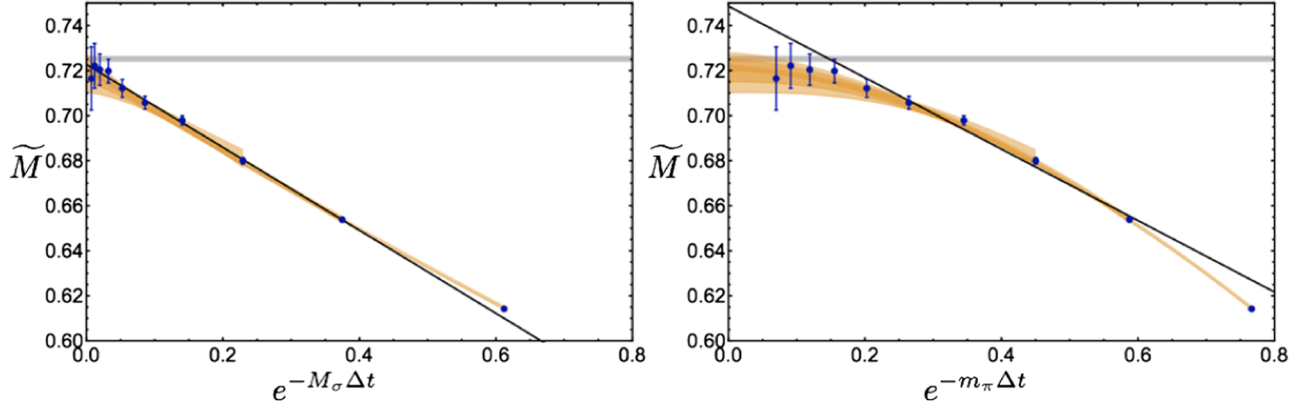
FIG. 35. The blue points and light-brown bands show the same χ^2-minimization fit results to large-time $\tilde{M}(t, \Delta t)$ plateaus as Fig. 34. The horizontal axis has been rescaled to coordinates that would show a linear bias for excited-state contributions from σ-mesons (left) and pions (right). Black lines show the central values of χ^2-minimization fits to constrained versions of Eq. (50) where δE is fixed to be M_σ (left) or m_π (right). The horizontal gray bands correspond to $M_N \pm \delta M_N$ from the high-precision NPLQCD calculation of Ref. [28].

experimental data: $M_\sigma = 457(14)$ MeV [50,51]. An extrapolation linear in the pion mass gives $M_\sigma \sim 830$ MeV at $m_\pi \sim 450$ MeV, in rough agreement with the best-fit excitation scale determined above. This agreement is insensitive to the form of the extrapolation used, as the Hadron Spectrum σ-meson mass result at $m_\pi \sim 391$ MeV is itself less than one standard deviation smaller than the best-fit nucleon excitation scale. Fits where $\delta E = M_\sigma$ is explicitly assumed can be performed more precisely and lead to consistent results with smaller uncertainties for the nucleon mass $M_N = 0.7226(18)$, as shown in Fig. 35. These fits provide another consistency check on δE but do not appropriately capture the systematic uncertainties of explicit assumptions about the excited-state spectrum.

The improved estimator proposed here exploits physical locality and finite correlation lengths to extract the effective mass from the evolution of $C_i(t)$ between times $t - \Delta t$ and t rather than the full evolution between source time $t = 0$ and sink time t. The correlation function at time $t - \Delta t$ is effectively treated as a new source so that the effective source/sink separation is fixed to be a constant length Δt rather than an increasing separation t. The effective source at $t - \Delta t$ still incorporates the dynamical evolution of the system between time 0 and $t - \Delta t$, and in particular has exponentially reduced excited-state contamination in magnitude compared to the original source. In principle t can be taken arbitrarily large with Δt fixed in order to extract a plateau in $\tilde{M}(t, \Delta t)$ with arbitrarily small excited-state contamination in the magnitude and constant statistical uncertainties across the plateau. The length of the lattice time direction becomes the only factor limiting the length of the plateau in this case.

Similar physical ideas underlie the hierarchical integration approach of Ref. [52]. In that approach, locality is exploited to decompose correlation functions into products of factors that can be computed on subsets of a lattice volume with exponentially reduced StN problems.

Hierarchical integration has been successfully implemented in studies of gluonic observables [53–57] and recently explored for baryon correlation functions in the quenched approximation [58] and beyond [59]. For baryon correlators, the method of Ref. [58] implements approximate factorization with systematically reducible uncertainties, as in the method proposed here. The benefits of the two methods are distinct. Hierarchical integration also employs standard statistical estimators for observables defined on subvolumes to determine correlation functions at large t with exponentially slower StN degradation. The new estimators introduced here allow data to be extracted from large-t correlation functions with constant StN, but removing all systematic uncertainties requires an extrapolation to large Δt with exponential StN degradation of the same severity as the original correlation function. Investigations of the compatibility of and relations between these methods are left to future work. In addition, this method also has similarities to the generalized pencil-of-functions method introduced to LQCD in Ref. [60], where correlation functions involving shifted source and sink times are combined in a variational basis. In the generalized pencil-of-functions approach, shifted source and sink times have primarily been investigated to reduce excited-state contamination rather than StN improvement.

In some sense, Δt can be considered a "factorization" scale in the time direction. The LQCD calculations are valid for all energy scales below that defined by the inverse lattice spacing, π/a. While well defined, the MC sampling of the path integral and analysis of baryon correlation functions fails to converge in the noise region because of the quantum fluctuations encountered along the paths from the source to large times, which include many incoherent hadronic volumes. The new estimator provides exponentially improved signal extraction at large times through limiting the number of contributing hadronic volumes to those within Δt, but does not provide a complete

description of the IR behavior of QCD, introducing a bias in the extracted mass of the nucleon. An extrapolation in Δt, using a form motivated by low-energy pion physics, is used to remove this bias. While different, this reminds one of matching LQCD calculations to the p-regime of chiral perturbation theory to remove finite-spatial-volume effects. The idea of performing an extrapolation to overcome a sign problem is not new. It was introduced 30 years ago to deal with the sign problem in MC calculations of modest size nuclei [61], and recently used in lattice effective field theory calculations to continuously evolve between the eigenvalues of nuclear many-body systems described by a Hamiltonian without a sign problem to one that does have a sign problem [62].

V. SUMMARY AND CONCLUSIONS

This work presents observations about the nucleon correlation function in LQCD that highlight the role of the complex phase in the signal-to-noise problem. The magnitude is found to have no StN problem and has the large-time scaling $\langle |C_i(t)| \rangle \sim e^{-\frac{3}{2}m_\pi t}$. The nucleon log-magnitude, $R_i(t)$, is approximately described by a normal distribution with linearly increasing mean and almost constant variance. The complex phase, which gives the direct importance sampling of $C_i(t)$ a sign problem, has the large-time scaling of approximately $\langle e^{i\theta_i(t)} \rangle \sim e^{-(M_N - \frac{3}{2}m_\pi)t}$. The StN problem arising from reweighting the complex phase of the nucleon correlation function matches the nucleon StN problem.

We present evidence that nucleon correlation functions are statistically described by a nearly decorrelated product of an approximately log-normal magnitude and wrapped normal phase. Complex correlation functions with log-normal magnitudes and wrapped normal phase factors are consistent with the arguments of Endres *et al.* [16], who suggested stable distributed correlation function logarithms may be a generic feature of quantum field theory and pursued a systematic statistical analysis of unitary fermion correlation functions that provides inspiration for this work. The wrapped normal phase distribution broadens with time, and at large times cannot be reliably distinguished from a uniform distribution. A noise region begins at this point where the sample mean phase becomes biased and systematically deviates from the true mean phase. In contrast, and importantly, $\frac{dR_i}{dt}$ and $\frac{d\theta_i}{dt}$ are described by approximately stable and wrapped stable distributions respectively that become constant at large times and can be estimated in the noise region with no StN problem.

It is remarkable that the Euclidean-time derivative of the logarithm of the correlation function is described by a heavy-tailed distribution while the logarithm itself is nearly normally distributed at all times. Further studies will be needed to understand the dynamical origin, continuum limit behavior, and universality of heavy-tailed

Euclidean-time evolution of correlation functions in quantum field theory. LQCD calculations at finer lattice spacings are needed to explore the continuum limit of the index of stability describing time evolution of the nucleon correlation function. Perturbative QCD and model calculations will provide useful insights into the dynamical origin of the heavy-tailed time evolution of the nucleon correlation function. Lattice and continuum studies of other quantum field theories are required to understand the universality of the heavy-tailed Euclidean-time evolution of correlation functions. Implications for real-time evolution are also left for future investigations.

Building on the observation that $\frac{d\theta_i}{dt}$ has constant width at large times, we have proposed a new estimator in Eq. (47) for the effective mass of the nucleon correlation function that relies on statistically sampling ratios of correlation functions at different times. This estimator has a StN ratio that is constant in t, the source-sink separation time, and the StN problem instead leads to an exponentially degrading StN ratio in Δt, the difference between the numerator and denominator sink times. The independence of t and Δt in this estimator allows similarly precise results to be extracted from all sufficiently large t rather than from a window of intermediate t with standard estimators. The new estimator effectively includes Δt time steps of time evolution following $t - \Delta t$ time steps of dynamical source improvement and it includes a systematic uncertainty that must be eliminated by extrapolating to the limit $\Delta t \to t \to \infty$. The systematic uncertainty of the new estimator is expected to decrease as $e^{-\delta E \Delta t}$ for large Δt, where δE is the energy gap between the ground state and the first excited state with appropriate quantum numbers and appreciable overlap with the effective source at $t - \Delta t$. Statistical uncertainties increase with increasing Δt as $\sim e^{2(M_N - \frac{3}{2}m_\pi)\Delta t}$. For $\Delta t \gtrsim \frac{\ln(N)}{2(M_N - \frac{3}{2}m_\pi)}$ additional systematic uncertainties associated with finite-sample-size effects in the statistical inference of circular random variables leads to unreliable results in the same way that $t \gtrsim \frac{\ln(N)}{2(M_N - \frac{3}{2}m_\pi)}$ leads to unreliable results in the noise region of standard estimators.

The properties of the new estimator that we have introduced may prove advantageous in the analysis of LQCD calculations of nuclei. Such systems are plagued by a reduced golden window compared to the single nucleon, presently limiting the length of plateaus from which to extract energy eigenvalues. A reanalysis of existing nuclear correlation functions generated by the NPLQCD Collaboration [28,63,64] is planned in order to determine the utility of this work for such systems. Binding momenta and other scales appearing in multibody hadronic systems may affect the form of the extrapolation used to remove the bias of the new estimator. It will be important to verify and further understand the scaling of the new estimator with pion mass and lattice spacing, as well as to investigate dependences on smearing scales and other

scales appearing in LQCD calculations. Studies of the vacuum channel including glueballs and scalar mesons and analyses of disconnected diagrams provide additional directions for further studies. Forming ratios of position space, rather than momentum space, correlation functions may be advantageous in future studies. Other types of LQCD calculations may also benefit from the new estimator, for instance in the isoscalar meson sector and those at nonzero baryon chemical potential.

It is not expected that the statistical properties of $\theta_i(t)$ discussed here and, in particular, the constant large-time width of $\frac{d\theta_i}{dt}$, are unique to single-nucleon correlation functions. If analogous statistical properties apply to generic complex correlation functions in quantum field theory, then estimators analogous to Eq. (47) can be constructed to extract the spectra of complex correlation functions and reweighted complex actions without StN problems. It remains to be seen if the approaches developed in this work can be fruitfully applied to other systems in particle, nuclear, and condensed matter physics that encounter sign and StN problems.

ACKNOWLEDGMENTS

[1] G. Parisi, Phys. Rep. **103**, 203 (1984).
[2] G. P. Lepage, in *From actions to answers. Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics, Boulder, USA, 1989*, edited by T. A. DeGrand and D. Toussaint (World Scientific, Singapore, 1990), pp. 97–120.
[3] C. Bernard and M. Golterman, Phys. Rev. D **88**, 014004 (2013).
[4] D. Weingarten, Phys. Rev. Lett. **51**, 1830 (1983).
[5] E. Witten, Phys. Rev. Lett. **51**, 2351 (1983).
[6] S. Nussinov, Phys. Rev. Lett. **51**, 2081 (1983).
[7] T. D. Cohen, Phys. Rev. Lett. **91**, 032002 (2003).
[8] W. Detmold, Phys. Rev. Lett. **114**, 222001 (2015).
[9] S. R. Beane, W. Detmold, T. C. Luu, K. Orginos, A. Parreno, M. J. Savage, A. Torok, and A. Walker-Loud, Phys. Rev. D **79**, 114502 (2009).
[10] S. R. Beane, W. Detmold, K. Orginos, and M. J. Savage, Prog. Part. Nucl. Phys. **66**, 1 (2011).
[11] S. R. Beane, W. Detmold, K. Orginos, and M. J. Savage, J. Phys. G **42**, 034022 (2015).
[12] W. Detmold and M. G. Endres, *Proc. Sci.*, LATTICE2014 (2015) 170 [arXiv:1409.5667].
[13] W. Detmold and M. G. Endres, Phys. Rev. D **90**, 034503 (2014).
[14] M. J. Savage (private communication).
[15] D. Grabowska, D. B. Kaplan, and A. N. Nicholson, Phys. Rev. D **87**, 014504 (2013).
[16] M. G. Endres, D. B. Kaplan, J.-W. Lee, and A. N. Nicholson, Phys. Rev. Lett. **107**, 201601 (2011).
[17] M. G. Endres, D. B. Kaplan, J.-W. Lee, and A. N. Nicholson, Phys. Rev. A **84**, 043644 (2011).
[18] M. G. Endres, D. B. Kaplan, J.-W. Lee, and A. N. Nicholson, *Proc. Sci.*, LATTICE2011 (2011) 017 [arXiv:1112.4023].
[19] J.-W. Lee, M. G. Endres, D. B. Kaplan, and A. N. Nicholson, *Proc. Sci.*, LATTICE2011 (2011) 203 [arXiv:1111.3793].
[20] M. G. Endres, D. B. Kaplan, J.-W. Lee, and A. N. Nicholson, Phys. Rev. A **87**, 023615 (2013).
[21] A. N. Nicholson, D. Grabowska, and D. B. Kaplan, J. Phys. Conf. Ser. **432**, 012032 (2013).

[22] T. DeGrand, Phys. Rev. D **86**, 014512 (2012).

[23] J. E. Drut and W. J. Porter, Phys. Rev. E **93**, 043301 (2016).

[24] W. J. Porter and J. E. Drut, Phys. Rev. A **95**, 053619 (2017).

[25] N. Fisher, *Statistical Analysis of Circular Data* (Cambridge University Press, Cambridge, England, 1995), ISBN 9780521568906.

[26] G. Borradaile, *Statistics of Earth Science Data: Their Distribution in Time, Space and Orientation* (Springer, Berlin Heidelberg, 2003), ISBN 9783540436034.

[27] K. Mardia and P. Jupp, *Directional Statistics*, Wiley Series in Probability and Statistics (Wiley, New York, 2009), ISBN 9780470317815.

[28] K. Orginos, A. Parreno, M. J. Savage, S. R. Beane, E. Chang, and W. Detmold, Phys. Rev. D **92**, 114512 (2015).

[29] M. Lüscher and P. Weisz, Commun. Math. Phys. **97**, 59 (1985); **98**, 433(E) (1985).

[30] B. Sheikholeslami and R. Wohlert, Nucl. Phys. **B259**, 572 (1985).

[31] M. Lüscher and U. Wolff, Nucl. Phys. **B339**, 222 (1990).

[32] S. R. Beane, P. F. Bedaque, K. Orginos, and M. J. Savage, Phys. Rev. Lett. **97**, 012001 (2006).

[33] S. R. Beane, W. Detmold, T. C. Luu, K. Orginos, A. Parreno, M. J. Savage, A. Torok, and A. Walker-Loud, Phys. Rev. D **80**, 074501 (2009).

[34] S. R. Beane, W. Detmold, H.-W. Lin, T. C. Luu, K. Orginos, M. J. Savage, A. Torok, and A. Walker-Loud (NPLQCD Collaboration), Phys. Rev. D **81**, 054505 (2010).

[35] T. A. DeGrand and D. Toussaint, *From actions to answers. Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics, Boulder, USA, 1989* (World Scientific, Singapore, 1990), p. 750.

[36] P. Young, Jackknife, and Bootstrap Resampling Methods in Statistical Analysis to Correct for Bias, http://young.physics.ucsc.edu/jackboot.pdf.

[37] M. Lüscher, in *Modern perspectives in lattice QCD: Quantum field theory and high performance computing. Proceedings, International School, 93rd Session, Les Houches, France, August 3-28, 2009*, edited by L. Lellouch, R. Sommer, B. Svetitsky, A. Vladikas, and L. F. Cugliandolo (2010), pp. 331–399, https://inspirehep.net/record/846344/files/arXiv:1002.4232.pdf.

[38] See supplemental material at http://link.aps.org/supplemental/10.1103/PhysRevD.96.114508 which includes raw histogram data, resampled averages, and covariance matrices for all results presented in figures throughout the article. This data allows readers to verify claims in the article based on numerical results and to test alternative statistical hypotheses for describing numerical results.

[39] D. Kaplan, Noise, Statistics and Sign Problems at Conceptual Advances in Lattice Gauge Theory (LGT14), 2014, https://indico.cern.ch/event/284433/contributions/1634883/.

[40] P. E. Gibbs, Phys. Lett. B **182**, 369 (1986).

[41] K. Splittorff and J. J. M. Verbaarschot, Phys. Rev. Lett. **98**, 031601 (2007).

[42] K. Splittorff and J. J. M. Verbaarschot, Phys. Rev. D **75**, 116003 (2007).

[43] S. Chandrasekhar, Rev. Mod. Phys. **15**, 1 (1943).

[44] J.-P. Bouchaud and A. Georges, Phys. Rep. **195**, 127 (1990).

[45] F. Bardou, Nature (London) **408**, 661 (2000).

[46] J. Voit, *The Statistical Mechanics of Financial Markets*, Texts and Monographs in Physics (Springer, New York, 2005), ISBN 9783540262855.

[47] J. P. Nolan, *Stable Distributions: Models for Heavy Tailed Data* (Birkhauser, Boston, 2015), in progress, Chapter 1 online at http://fs2.american.edu/jpnolan/www/stable/stable.html.

[48] M. L. Wagman and M. J. Savage, arXiv:1704.07356.

[49] R. A. Briceno, J. J. Dudek, R. G. Edwards, and D. J. Wilson, Phys. Rev. Lett. **118**, 022002 (2017).

[50] I. Caprini, G. Colangelo, and H. Leutwyler, Phys. Rev. Lett. **96**, 132001 (2006).

[51] R. Garcia-Martin, R. Kaminski, J. R. Pelaez, and J. Ruiz de Elvira, Phys. Rev. Lett. **107**, 072001 (2011).

[52] M. Lüscher and P. Weisz, J. High Energy Phys. 09 (2001) 010.

[53] H. B. Meyer, J. High Energy Phys. 01 (2003) 048.

[54] M. Della Morte and L. Giusti, Comput. Phys. Commun. **180**, 813 (2009).

[55] M. Della Morte and L. Giusti, Comput. Phys. Commun. **180**, 819 (2009).

[56] M. Della Morte and L. Giusti, J. High Energy Phys. 05 (2011) 056.

[57] M. García Vera and S. Schaefer, Phys. Rev. D **93**, 074502 (2016).

[58] M. Cè, L. Giusti, and S. Schaefer, Phys. Rev. D **93**, 094507 (2016).

[59] M. Cè, L. Giusti, and S. Schaefer, Phys. Rev. D **95**, 034503 (2017).

[60] C. Aubin and K. Orginos, AIP Conf. Proc. **1374**, 621 (2011).

[61] Y. Alhassid, D. J. Dean, S. E. Koonin, G. Lang, and W. E. Ormand, Phys. Rev. Lett. **72**, 613 (1994).

[62] T. A. Lähde, T. Luu, D. Lee, U.-G. Meißner, E. Epelbaum, H. Krebs, and G. Rupak, Eur. Phys. J. A **51**, 92 (2015).

[63] S. R. Beane, E. Chang, S. D. Cohen, W. Detmold, H. W. Lin, T. C. Luu, K. Orginos, A. Parreno, M. J. Savage, and A. Walker-Loud (NPLQCD Collaboration), Phys. Rev. D **87**, 034506 (2013).

[64] S. R. Beane *et al.* (NPLQCD Collaboration), Phys. Rev. C **88**, 024003 (2013).

[65] R. G. Edwards and B. Joo (SciDAC Collaboration, LHPC Collaboration, UKQCD Collaboration), Nucl. Phys. B, Proc. Suppl. **140**, 832 (2005).