

Neutrino oscillations as a probe of macrorealism

Kathrine Mørch Groth^{✉,*}, Johann Ioannou-Nikolaides[✉], D. Jason Koskinen[✉], and Markus Ahlers[✉]
Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, 2100 Copenhagen, Denmark



(Received 12 April 2025; accepted 12 November 2025; published 25 November 2025)

The correlations between successive measurements of a quantum system can violate a family of Leggett-Garg Inequalities (LGIs) that are analogous to the violation of Bell's inequalities of measurements performed on spatially separated quantum systems. These LGIs follow from a *macrorealistic* point of view, imposing that a classical system is at all times in a definite state and that a measurement can, at least in principle, leave this state undisturbed. Violations of LGIs can be probed by neutrino flavor oscillations if the correlators of consecutive flavor measurements are approximately stationary. We discuss here several improvements of the methodology used in previous analyses based on accelerator and reactor neutrino data. We argue that the strong claims of LGI violations made in previous studies are based on an unsuitable modeling of macrorealistic systems in statistical hypothesis tests. We illustrate our improved methodology via the example of the MINOS muon-neutrino survival data, where we find revised statistical evidence for violations of LGIs at the $(2-3)\sigma$ level, depending on macrorealistic background models.

DOI: [10.1103/y2v2-ttj9](https://doi.org/10.1103/y2v2-ttj9)

I. INTRODUCTION

The statistical properties of measurements performed on a quantum system defy our probabilistic intuition from a macroscopic, i.e., classical, point of view. One well-known example is the violation of Bell's inequalities in entangled quantum systems [1], which challenges the probabilistic interpretation of quantum mechanics as a local theory of hidden variables. While Bell's inequalities relate simultaneous measurements in spatially separated systems, *Leggett & Garg* [2] derived analogous inequalities that relate successive measurements performed on a local system. These Leggett-Garg inequalities (LGIs) follow from three *macrorealistic* postulates that: (i) a system is at any time in a definite state of its observables (*macroscopic realism per se*), that (ii) a measurement on this system can, in principle, be performed without disturbance (*noninvasive measurability*), and that (iii) the properties of the system are exclusively determined by its initial condition (*induction*) [3]. These three postulates appear to be reasonable from a classical perspective and they define our notation of the term *macrorealism* in the broader sense; for a review see Ref. [4].

It has been argued that the violation of LGIs—and therefore macrorealism in the broader sense—can be

observed in neutrino flavor oscillations [5,6]. This first appears counterintuitive since neutrinos are only weakly interacting and therefore a successive measurement of a neutrino system is experimentally unfeasible. However, if the correlation of neutrino flavors measured at different times is stationary and therefore only depends on the time difference, a sequence of measurements can be represented by neutrino flavor survival and transition probabilities that are inferred from a large ensemble of equally prepared neutrinos, e.g., a beam of muon neutrinos from pion decays. In this way, previous work has studied violations of LGIs in accelerator neutrino beams from MINOS [7] and NO ν A [8] as well as reactor neutrinos from Daya Bay [9], RENO [8], and KamLAND [10]; see also Refs. [11–13].

The standard LGIs used in previous analyses of neutrino data are based on Leggett-Garg *strings* (called K_n in the following; see Ref. [4]) which can be constructed from a combination of n neutrino flavor measurements at oscillation phases (ϕ) related to the neutrino baseline (L) and energy (E) as $\phi \propto L/E$ with the requirement that the sum over the first $n - 1$ phases matches the n th phase within some tolerance. A test statistic for the departure from macrorealism can then be defined as the fraction of violations of LGIs inferred from all of these matching phase combinations. We discuss in this paper a generalization of this method, that allows us to probe LGIs between arbitrary sequences of data with matching phase sums and that always yields the strongest violation of LGIs for these matching sequences.

A statistical hypothesis test of LGIs needs to be able to quantify in what way the observed neutrino flavor correlations are incompatible with the predictions from

*Contact author: kathrine.groth@nbi.ku.dk

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

macrorealism. Previous work has approached this question by introducing a *classical* LG string (K_n^C) and testing its violation of LGIs on resampled oscillation data [7,9], rather than simulating pseudodata following a macrorealistic background model; other studies have neglected to study the behavior of background models [8,10]. We argue here that this approach does not yield reliable results and the significance of the violation of LGIs reported in previous studies should be scrutinized. In particular, the *classical* LG string K_n^C when applied to oscillation data will—by construction—only violate LGIs if the flavor correlations become unphysical, related to uncertainties of experimental data. In this paper, we provide an alternative statistical method that offers a more robust estimate of the significance.

The paper is organized as follows: in Sec. II we first review the general form of LGIs based on LG strings and their violation in quantum-mechanical systems. We then introduce an improved method for identifying LGIs for stationary correlators in neutrino flavor oscillation analyses in Sec. III. In Sec. IV we discuss our revised method for testing macrorealism in neutrino oscillations based on pseudodata, that is robust against unphysical flavor correlators, $|\mathcal{C}| > 1$, unlike previously proposed methods. We then apply our method to MINOS/MINOS+ data in Sec. V in order to demonstrate the differences and improvements compared to previous studies. We discuss our results in Sec. VI before concluding in Sec. VII. Throughout the paper we are working in natural units where $c = \hbar = 1$ and boldface quantities indicate vectors.

II. LEGGETT-GARG INEQUALITIES

In the following, the quantity $Q(t)$ represents a *dichotomic* property of a system evolving in time t with values normalized to ± 1 . For instance, we can assign $Q(t) = 1$ for the case that a neutrino is in the flavor state ν_μ and $Q(t) = -1$ for all other flavor states, i.e., ν_e or ν_τ . We will use the abbreviation $Q_i = Q(t_i)$ for different observation times t_i in the following. The expected correlation of measurements at different times is expressed by the correlators $\mathcal{C}_{ij} \equiv \langle Q_i Q_j \rangle$ which are bounded as $|\mathcal{C}_{ij}| \leq 1$ with $\mathcal{C}_{ij} = \pm 1$ representing full correlation and anticorrelation, respectively. The theoretical prediction of the expectation value $\langle \cdot \rangle$ is fundamentally different in quantum mechanics (QM) and macrorealism (MR).

Leggett-Garg inequalities express relations between the correlators $\mathcal{C}_{ij}$ under the assumption of MR [2]. From a macrorealistic point of view, the expectation values $\langle \cdot \rangle$ are understood as ensemble averages following from underlying joint probability densities $\rho(Q_i, Q_j)$ of measuring the outcomes Q_i and Q_j at times t_i and t_j , respectively [2]. More explicitly, the macrorealistic correlator is expected to take the form

$$\mathcal{C}_{ij} \equiv \sum_{Q_i, Q_j} Q_i Q_j \rho(Q_i, Q_j) \quad (\text{in MR}). \quad (1)$$

Note that MR requires that the correlator in Eq. (1) is symmetric under the exchange of measurement times, $\mathcal{C}_{ij} = \mathcal{C}_{ji}$. One can now define n -measurement Leggett-Garg strings K_n with $n \geq 3$ as [4]

$$K_n \equiv \sum_{i=1}^{n-1} \mathcal{C}_{i(i+1)} - \mathcal{C}_{1n}. \quad (2)$$

Using the macrorealistic correlator in Eq. (1) it can be shown that these LG strings obey the LG inequalities $K_n \leq n - 2$ (see Appendix A).

Conversely to a classical system, in a quantum system the correlation of two consecutive measurements can be expressed as

$$\mathcal{C}_{ij} \equiv \frac{1}{2} \langle \{\hat{Q}_i, \hat{Q}_j\} \rangle \quad (\text{in QM}), \quad (3)$$

where $\hat{Q}_i$ and $\hat{Q}_j$ are understood as Heisenberg operators and the expectation value $\langle \cdot \rangle$ is now in terms of a trace over basis states weighted by the density matrix (see Appendix B). Again, the correlator in Eq. (3) is symmetric, $\mathcal{C}_{ij} = \mathcal{C}_{ji}$.

For instance, a stationary quantum system that maximally oscillates between the eigenvalues $Q = \pm 1$ has correlators $\mathcal{C}_{ij} = \cos(\phi_i - \phi_j)$ where the oscillation phase ϕ depends on the oscillation frequency and time. If we consider a fixed phase step between consecutive measurements, $\phi_{i+1} - \phi_i = \Delta\phi$, the LG strings in Eq. (2) become $K_n = (n-1) \cos \Delta\phi - \cos((n-1)\Delta\phi)$. These can be shown to become maximal at $\Delta\phi = \pi/n$ with $K_n(\pi/n) = n \cos(\pi/n)$ and violate the LGIs for any $n \geq 3$; see Appendix B.

The LG string in Eq. (2) is not the only (and not necessarily the optimal) quantity for a test of MR between n measurements. First, the measurement times t_i that appear in the LG string do not need to be time-ordered (or even different), as was assumed in previous analyses of the violation of macroscopic realism in neutrino oscillations [7–9]. And second, the expression in Eq. (2) can be generalized by arbitrarily redefining the observables as $Q_i \rightarrow -Q_i$, which introduces a sign flip in the corresponding correlators. As we show in Appendix A, this leads to a generalization of the LG string that can be written in the form

$$K_n(\sigma) \equiv \sum_{i=1}^n \sigma_i \mathcal{C}_{i(i+1)}, \quad (4)$$

where we define $t_{n+1} \equiv t_1$ for later convenience and introduce the n sign assignments $\sigma_i = \pm 1$, which can take any combination subject to the condition

$$\prod_{i=1}^n \sigma_i = -1. \quad (5)$$

This results in 2^{n-1} independent LG strings (4), all subject to the inequality $K_n(\sigma) \leq n - 2$.

The original LG string in (2) corresponds to the assignment $\sigma_i = 1$ for all $i < n$ and $\sigma_n = -1$. For even n we can also invert any sign vector, $\sigma \rightarrow -\sigma$, leading to an overall sign flip $K_n \rightarrow -K_n$ of the LG string and the two-sided inequality $-n + 2 \leq K_n \leq n - 2$. For odd n we can only give the lower bound $-n \leq K_n$, which follows from $|K_n| \leq n$.

The freedom of choosing σ in Eq. (4) allows us to define the *optimal* LG string for a test of MR as the maximum:

$$K_n^{\max} \equiv \max_{\sigma} K_n(\sigma). \quad (6)$$

From Eq. (4) it is clear that the sign assignment $\sigma_i = \text{sign} \mathcal{C}_{i(i+1)}$ maximizes $K_n(\sigma)$ as long as this combination obeys the condition in Eq. (5). If this is not the case, we can find the maximum by alternating the sign for the smallest member of the set $\{|\mathcal{C}_{i(i+1)}|\}$.

As an illustration, Fig. 1 shows the example of a LG string with $n = 4$ and its violation of the LGI in maximally oscillating two-level quantum systems. We consider four measurements Q_i that are distributed with relative phase

steps ϕ_a, ϕ_b , and ϕ_c with $\phi_a + \phi_b + \phi_c = 3\pi/2$. The three different columns show different sign assignments and different time orders of measurements. The left column assumes that the measurements appear time-ordered in the LG string as indicated in the graph. The ternary diagram shows the value of the conventional LG string K_4 in Eq. (2); none of the phase configurations violate the LGI $K_4 \leq 2$ (nor the lower bound $-2 \leq K_4$).

In the center column of Fig. 1 we use the same time-ordered measurements, but now the optimized LG string of Eq. (6). With the appropriate choice of σ , we now find regions where the LGI is violated (red-shaded regions). Finally, the last column of Fig. 1 shows an example where the measurements Q_i are not time-ordered, but follow the sequence shown in the top graph. In the previous two cases, the relative phase differences $\Delta\phi_i \equiv \phi_{i+1} - \phi_i$ were positive for $i \leq 4$ and negative for $i = 4$ (“1 + 3”). Now, we have two positive ($i = 1$ and 2) and two negative ($i = 3$ and 4) phase differences between observations (“2 + 2”). Note that the corresponding LG strings of Eq. (6) now finds violations of LGIs that were not visible under the previous time order (left and middle columns). We will discuss in the next section that these different sequences, 1 + 3 and 2 + 2, correspond to different matching conditions for the phase differences in stationary correlators in order to construct LG strings for neutrino oscillations.

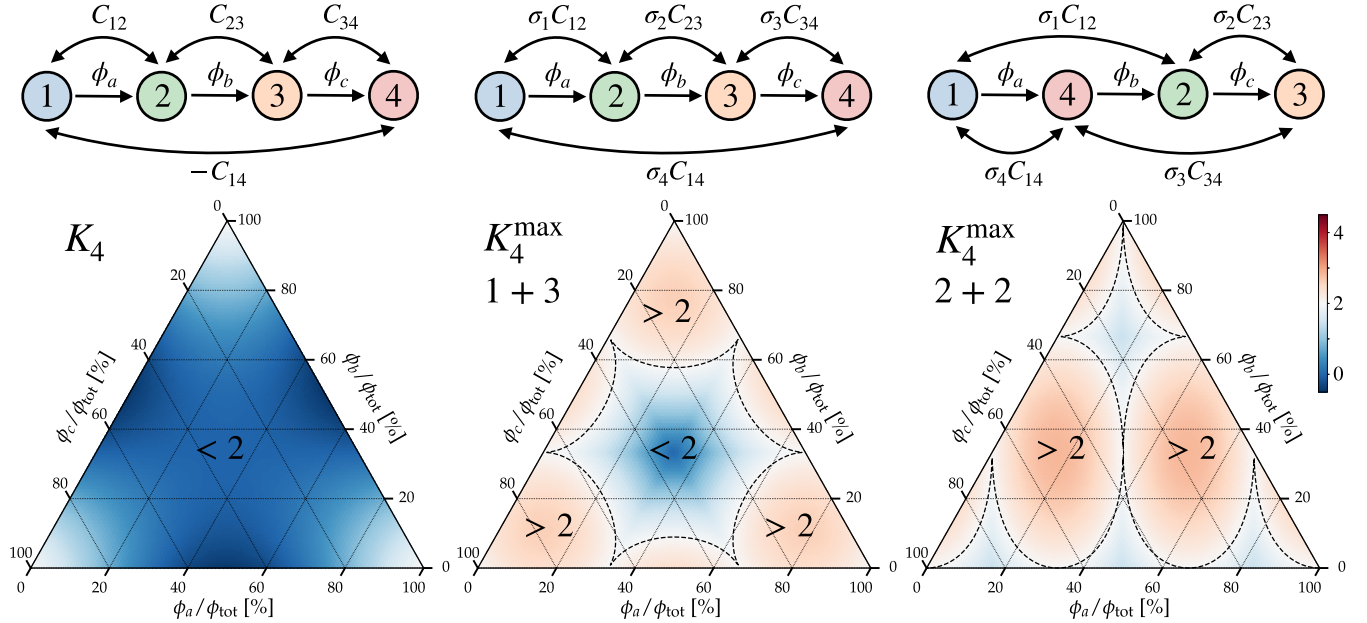


FIG. 1. Illustration of LG strings and their violation of LGIs in QM. We consider $n = 4$ consecutive measurements (see top panels) with phase steps ϕ_a, ϕ_b , and ϕ_c and assume that the correlators follow $\mathcal{C}_{ij} = \cos(\phi_i - \phi_j)$ for maximal two-level oscillations in QM. The total phase $\phi_{\text{tot}} = \phi_a + \phi_b + \phi_c$ is fixed to $\phi_{\text{tot}} = 3\pi/2$, comparable to the phase range covered by the MINOS/MINOS+ data. The ternary plots indicate the relative combinations of ϕ_a, ϕ_b , and ϕ_c where we expect that the corresponding LG strings violate (obey) the LGIs indicated by the red-shaded (blue-shaded) regions. The left panel shows K_4 as in Eq. (2) which does not violate the LGI $K_4 \leq 2$ (nor $-2 \leq K_4$) for this example. The center (right) panel shows $K_4^{\max}$ in Eq. (6) for two different orders of measurements (“1 + 3” or “2 + 2”) as indicated in the top panel.

III. STATIONARY CORRELATORS

For the test of LGIs in neutrino oscillations, we consider situations where the flavor evolution can be approximated as that of a two-level system with an effective Hamiltonian that is independent of time. This is appropriate for neutrino oscillations in vacuum or in uniform matter as long as the observation times are much shorter than the oscillation period induced by the solar mass splitting. Under these conditions, the correlators take on the form

$$C_{ij} = p_+(t_i)[2P_{++}(t_j - t_i) - 1] + p_-(t_i)[2P_{--}(t_j - t_i) - 1], \quad (7)$$

where $p_{\pm}(t)$ is the probability of observing the system in the state $Q(t) = \pm 1$ and $P_{\pm\pm}(\tau)$ is the conditional (“survival”) probability that the state is observed in the same state after a time difference τ . If the evolution is invariant under time reversal, $P_{+-} = P_{-+}$, the survival probabilities become identical, $P_{++} = P_{--}$, and the correlators are independent of the system’s initial condition with the simple form

$$C_{ij} = C(t_j - t_i) = 2P_{++}(t_j - t_i) - 1. \quad (8)$$

We see that, under above conditions, the correlators C_{ij} become stationary, i.e., only depend on the time between observations, and can be simply related to the stationary survival probability.

Consider now oscillation data obtained from an ensemble of neutrinos that are initialized in the (pure) state $Q = +1$ and observed at times $\tau_i > 0$. Assuming stationarity as in Eq. (8), the LG strings defined in Eq. (4) can be expressed in terms of n time differences¹ $\tau_i \equiv |t_{i+1} - t_i|$ and $C_{i(i+1)} = C(\tau_i)$. We can now consider a sequence $\mathbf{s} = (s_1, \dots, s_n)$ representing the indices of n time differences τ_{s_i} (including repeated entries) that can be written as a union of two subsequences, $\mathbf{s} = \mathbf{s}_a \cup \mathbf{s}_b$, with length n_a and n_b such that

$$\left| \sum_{i=1}^{n_a} \tau_{s_{a,i}} - \sum_{i=1}^{n_b} \tau_{s_{b,i}} \right| \leq \frac{\varepsilon}{2} \sum_{i=1}^n \tau_{s_i}, \quad (9)$$

where $\varepsilon \ll 1$ is the relative tolerance of timescale; to be discussed later. In the limit $\varepsilon \rightarrow 0$ the sequence $\mathbf{s}$ allows us to define a LG string

$$K(\mathbf{s}, \sigma) \equiv \sum_{i=1}^n \sigma_i C(\tau_{s_i}), \quad (10)$$

which is subject to the inequality $K(\mathbf{s}, \sigma) \leq n - 2$. For instance, Fig. 1 shows two examples of sequences

¹Since in both QM and MR we have $C_{ij} = C_{ji}$ we can focus on the absolute scale of time differences.

for $n = 4$ with $n_a = 1$ and $n_b = 3$ (“1 + 3”; middle column) and $n_a = 2$ and $n_b = 2$ (“2 + 2”; right column), where the time differences τ_i are equivalent to phase differences $\Delta\phi_i = \omega\tau_i$ in maximal two-level oscillations with some angular frequency ω .

For the test of MR in neutrino oscillation, we can now determine the fraction of all sequences $\mathbf{s}$ that give at least one violation of LGIs. This is equivalent to testing the violation of the LGI for the maximum $K(\mathbf{s}, \sigma)$ for a given sequence $\mathbf{s}$

$$K_{\max}(\mathbf{s}) \equiv \max_{\sigma} K(\mathbf{s}, \sigma), \quad (11)$$

analogous to the construction in Eq. (6). Note that the number of sequences satisfying the time-matching condition in Eq. (9) increases drastically with n . To avoid duplicates and trivial extensions of shorter sequences, e.g., by simply adding the same time step to both subsequences, we also demand that: (a) $\mathbf{s}$ is unique up to permutations, (b) $\mathbf{s}_a$ and $\mathbf{s}_b$ are disjoint ($\mathbf{s}_a \cap \mathbf{s}_b = \emptyset$), and (c) $\mathbf{s}_a$ and $\mathbf{s}_b$ are not equivalent to unions $\mathbf{s}'_a \cup \mathbf{s}''_a$ and $\mathbf{s}'_b \cup \mathbf{s}''_b$, respectively, corresponding to shorter sequences $\mathbf{s}'$ and $\mathbf{s}''$ passing the time-matching condition.

IV. MACROREALISTIC BACKGROUND

A challenging task for a statistical test of violation of LGIs is the definition of a “realistic” background hypothesis. Previous analyses of neutrino oscillation data approached this challenge by replacing the LG string in Eq. (2) by a “classical” LG string defined as [7,9]

$$K_n^C \equiv \sum_{i=1}^{n-1} C_{i(i+1)} - \prod_{i=1}^{n-1} C_{i(i+1)}. \quad (12)$$

This expression is motivated by the assumption that a “classical” system behaves Markovian, which implies that the classical neutrino transition/survival probability $P_{Q_i Q_j}^C$ obeys the Chapman-Kolmogorov equation [14]

$$P_{Q_1 Q_3}^C(\tau_1 + \tau_2) = \sum_{Q_2} P_{Q_1 Q_2}^C(\tau_1) P_{Q_2 Q_3}^C(\tau_2). \quad (13)$$

Assuming a two-level system obeying time-reversal symmetry, this implies that $C(\tau_1 + \tau_2) = C(\tau_1)C(\tau_2)$ and, more generally, $C(\sum_i \tau_i) = \prod_i C(\tau_i)$.

The expression from Eq. (12) obeys the LGI $K_n^C \leq n - 2$ for arbitrary flavor correlations with a maximum of K_n^C appearing at $C_{i(i+1)} = 1$. Therefore, regardless of the structure and precision of the neutrino oscillation data, the quantity from Eq. (12) will *never* violate the LGI and does not allow to quantify the level of chance violations in background data. The background distributions of LGI violations found in previous studies [7,9] depend on the relatively rare cases where the neutrino oscillation data

assumes *unphysical* values with $P > 1$ or $P < 0$ related to measurement uncertainties. This is the only case where the classical LG string in Eq. (12) can violate LGI. This demonstrates that an estimate of the statistical significance of LGI violations based on the classical LG string is not a robust method.

In this analysis we develop a more realistic approach. In general, we will assume that a macrorealistic system follows a classical Markovian process. The classical correlator of an (effective) two-level systems follows then the model $\mathcal{C}(\tau) = e^{-\Gamma\tau}$ where $\Gamma > 0$ is the decorrelation rate; see also Ref. [13]. We will consider two background models assuming: (a) a fully correlated classical system with $\mathcal{C}(\tau) = 1$ assuming $\Gamma = 0$, and (b) a decorrelating system where the rate Γ is fit to the experimental data. For each background model, denoted $\mathcal{H}_0^a$ and $\mathcal{H}_0^b$ in the following, we generate pseudodata with the same variance $\Delta\mathcal{C}(\tau_i)$ as observed in the experimental data. In contrast to previous analyses, we will then determine the chance violations of LGIs based on statistical fluctuations of the macrorealistic background model.

V. PROBING LGIS IN NEUTRINO DATA

In the following, we illustrate our revised test for violations of LGIs in neutrino oscillations using the muon neutrino survival probabilities reported by the MINOS experiment and its upgrade MINOS+ [15]. The oscillation data was inferred from muon neutrinos with energies up to 120 GeV produced by the NuMI beamline at Fermilab and observed as charged current events by near and far detectors over a baseline of $L \simeq 735$ km. The MINOS/MINOS+ data in the form $P_{\text{obs}} \pm \Delta P_{\text{obs}}$ correspond to the ratio of observed ν_μ events in the far detector compared to those expected from observations in the near detector and is provided in terms of 39 bins of reconstructed muon neutrino energy E [15].

The observation of highly relativistic neutrinos at distance $L \simeq ct$ corresponds to neutrinos that have evolved over the proper time $t/\gamma \propto L/E$ since production, assuming that muon neutrinos have an (effective) mass $m \ll E$ and Lorentz factor $\gamma = E/m$. Hence, the role of time in neutrino data is played by the ratio L/E and we will in general refer to this as the neutrino oscillation *phase*. Figure 2 shows the MINOS/MINOS+ data in terms of the flavor correlations from Eq. (8) and L/E bins. The solid green line shows the best-fit neutrino oscillation model using best-fit oscillations parameters assuming normal ordering [16]. The baseline L of the far detector is well below the oscillation length induced by the solar mass splitting, $\lambda_{\text{sol}} \simeq 3.3 \times 10^4 (E/\text{GeV})$ km and we can hence treat the oscillation as an effective two-level system driven by the oscillation length of the atmospheric mass splitting, $\lambda_{\text{atm}} \simeq 990 (E/\text{GeV})$ km [16].

Assuming stationarity, we first determine the number of phase combinations ($\tau_i \propto L/E_i$) of the MINOS/MINOS+ data that allow us to construct LG strings. For comparison

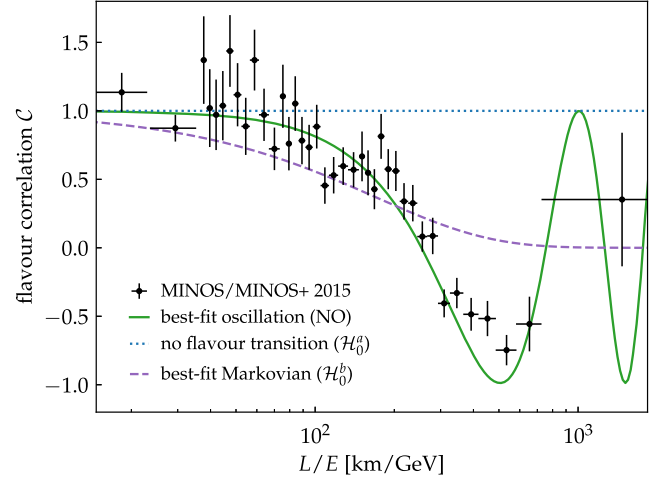


FIG. 2. Flavor correlations observed in MINOS/MINOS+ data [15] (black data) in comparison to the prediction of the best-fit oscillation models assuming normal ordering [16] (in green) and two classical models: a constant model (in dotted blue) and the best-fit Markovian model (in dashed purple).

with the earlier study [7], we choose our nominal phase tolerance in Eq. (9) to be $\varepsilon = 0.5\%$. Table I shows the number of unique phase combinations following the prescription outlined in Sec. III. We show the results for up to $n = 5$ combinations of phases in terms of their split into subsequences with length n_a and n_b (“ $n_a + n_b$ ”). Note that the conventional LG strings in Eq. (2) used in previous studies are based on the phase combinations with $n_a = 1$ and $n_b = n - 1$. However, as we pointed out in Sec. II, it is also possible to construct additional phase combinations in the case $n > 3$ from the matching condition in Eq. (9).

It is apparent in Fig. 2 that some of the correlations inferred from MINOS/MINOS+ data (black data) extend into the unphysical range $C > 1$, related to reconstructed survival probabilities with $P_{\text{obs}} > 1$. Note that the muon neutrino survival data reported by MINOS/MINOS+ correspond to ratios of observed muon neutrino events to those expected from background expectation assuming no flavor transitions. This quantity is subject to statistical

TABLE I. The total number of unique phase combinations satisfying Eq. (9) based on the MINOS/MINOS+ data. The combinations are shown for three values of the tolerance parameter ε and increasing number of phases n , split into different combinations of subsequences s_a and s_b with $n = n_a + n_b$. Previous analyses of MINOS/MINOS+ data [7,8] are based on 1 + 2 (in K_3) and 1 + 3 (in K_4).

	$n = 3$		$n = 4$		$n = 5$	
ε	1 + 2	1 + 3	2 + 2	1 + 4	2 + 3	
5%	731	6,902	9,479	45,990	218,542	
0.5%	92	699	964	4,646	24,017	
0.05%	24	76	81	495	2,350	

fluctuations, reconstruction uncertainties, background subtraction, systematic uncertainties, etc. and can lead to best-fit values of $P_{\text{obs}} > 1$ or even $P_{\text{obs}} < 0$. But the LGIs rely on the condition that the survival probabilities are bound to the physical region $0 \leq P \leq 1$ (see Appendix A). Therefore, in order to test for the violation of LGI it is necessary to *account for unphysical correlations* related to measurement uncertainties.

To quantify the level of LGI violations in the MINOS/MINOS+ data we now proceed as follows. We determine the fraction of LGI violations of the data for the five different phase combinations of LG strings listed in Table I. We then create pseudosamples of data by sampling the

muon neutrino survival probabilities from model predictions treating the (symmetric) measurement uncertainties ΔP_{obs} of the MINOS/MINOS+ as the standard deviation of normal distributions. The $\phi \propto L/E$ bins of the pseudo-samples are identical to those of the MINOS/MINOS+ data and we average the model predictions over the bin width. As a first check, we can test if the data is consistent with the expectation from best-fit neutrino oscillation models assuming normal order [16], predicting a flavor correlation that is shown as the green line in Fig. 2.

The top left plot in Fig. 3 shows the fraction of LGI violations observed in the MINOS/MINOS+ data (solid lines with $\star$ symbol) compared to the normalized

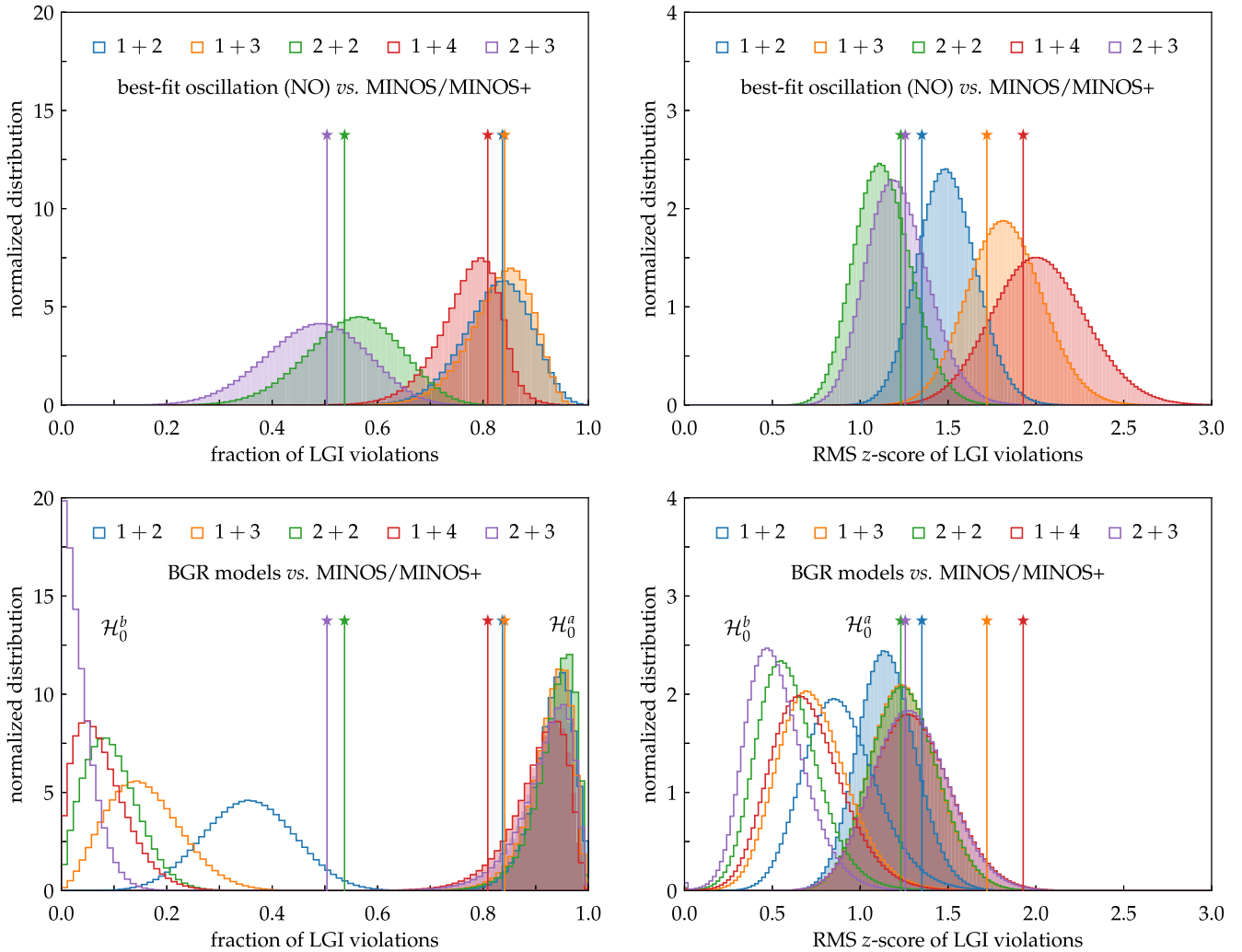


FIG. 3. Normalized distributions of the fraction of LGI violations (left panels) and corresponding RMS z -scores (right panels) based on 10^7 pseudosamples. Results are shown separately for LG strings in Eq. (11) for five different combinations of n_a and n_b with $n_a + n_b = 3, 4$ and 5 listed in Table I. The results from MINOS/MINOS+ data are shown as vertical lines with $\star$ symbols. The top panels show the distributions expected from neutrino oscillations assuming normal ordering (NO) [16]. The bottom panels show those expected for the background (BGR) models $\mathcal{H}_0^a$ with $\mathcal{C}(\tau) = 1$ (filled histograms) and $\mathcal{H}_0^b$ with $\mathcal{C}(\tau) = e^{-\Gamma\tau}$ (open histograms). The BGR distributions of RMS z -scores allow us to estimate the chance probabilities (p -values) of LGI violations under the BGR hypotheses at the level observed in MINOS/MINOS+ data. Treating the five p -values as independent trials, we estimate the post-trial significance at the level of 2.1σ (3.7σ) for the BGR model $\mathcal{H}_0^a$ ($\mathcal{H}_0^b$).

distribution of those from 10^7 pseudosamples following the prediction of neutrino oscillation (filled histograms). The average fraction of LGI violations appears to be relatively high, reaching about 80% on average for the case $n = 3$ (blue histogram for $1 + 2$), consistent with Ref. [7]. The observed fraction of LGI violations lies well within the distributions of those from pseudosamples in all case. However, to understand the significance of this result, we need to test the level of LGI violations that would be observed in macrorealistic data. This can be done by creating distributions from background expectations following our macrorealistic models, $\mathcal{H}_0^a$ and $\mathcal{H}_0^b$, introduced in the previous section and shown in terms of the corresponding flavor correlation in Fig. 2 (dotted and dashed lines, respectively).

The left bottom plot of Fig. 3 shows the fraction of LGI violations of the background model $\mathcal{H}_0^a$ with no transitions between neutrino flavors with $\mathcal{C}(\tau) = 1$ (filled histograms) and the alternative background model $\mathcal{H}_0^b$ with $\mathcal{C}(\tau) = e^{-\Gamma\tau}$. One can observe that the expected LGI violations in the background models disagree with those of the oscillation model and, moreover, can drastically exceed those observed in the MINOS/MINOS+ data in the case of the model $\mathcal{H}_0^a$. The relatively large number of LGI violations for $\mathcal{H}_0^a$ can be understood from the definition of the LG string in Eq. (10): The LGI is saturated for $\mathcal{C} = 1$ and pseudodata accounting for the MINOS/MINOS+ measurement uncertainties can frequently combine to lift the LG string above the boundary. In comparison, the chance violation of LGIs in the alternative background model $\mathcal{H}_0^b$ with $\mathcal{C}(\tau) = e^{-\Gamma\tau}$ are suppressed compared to those in the model $\mathcal{H}_0^a$ since the correlators are $|\mathcal{C}| \ll 1$ for most of the MINOS/MINOS+ data range.

The numerous LGI violations in the background model $\mathcal{H}_0^a$ indicate that the simple count of LGI violation is—in general—not a good statistical measure for tests of MR: small fluctuations in the pseudodata derived for macrorealistic background models can frequently combine to violate LGIs. This motivates us to consider the (one-sided) z -score of a LGI violation for the phase combination $\mathbf{s}$ of length n as

$$z(\mathbf{s}) = \max \left(0, \frac{K_{\max}(\mathbf{s}) - n + 2}{\Delta K_{\max}(\mathbf{s})} \right), \quad (14)$$

where $K_{\max}(\mathbf{s})$ is the maximum LG string in Eq. (11) and $\Delta K_{\max}(\mathbf{s}, \boldsymbol{\sigma})$ its uncertainty following from the uncertainty of the MINOS/MINOS+ data via error propagation. The root-mean-square (RMS) of the z -scores for a given pseudosample with $N_{\text{LGV}} > 0$ defines our test statistic (TS)

$$z_{\text{RMS}} \equiv \sqrt{\frac{1}{N_{\text{LGV}}} \sum_{\mathbf{s}} z^2(\mathbf{s})}, \quad (15)$$

where N_{LGV} are the number of sequences that *violate* the LGI, i.e., sequences with $z(\mathbf{s}) > 0$. In the case $N_{\text{LGV}} = 0$ we instead use $z_{\text{RMS}} = 0$.

The resulting distributions of RMS z -scores of LGI violations are shown in the right column of Fig. 3. The values derived from the MINOS/MINOS+ data (vertical lines with $\star$ symbols) show an excess of $z_{\text{RMS}} > 1$ for all five combinations of LG sequences, most noticeably for the cases $1 + 3$ and $1 + 4$ and consistent with the expected TS distribution from the best-fit oscillation model (top right panel). The bottom right panel shows the corresponding TS distribution of the two background models. Despite the numerous violations of LGIs in the background case $\mathcal{H}_0^a$ (filled histograms in bottom left panel), the corresponding RMS z -scores only marginally exceed unity (filled histograms in bottom right panel). Note that the TS distributions of this time-independent background model ($\mathcal{C} = 1$) depend on the total number of phases, $n = n_a + n_b$, rather than the individual split in n_a and n_b . This is noticeable as the near-identical distributions for $1 + 3$ and $2 + 2$ as well as $1 + 4$ and $2 + 3$. On the other hand, the TS distributions of the background case $\mathcal{H}_0^b$ (open histograms) are centered around much lower z_{RMS} values and those of the MINOS/MINOS+ data appear in their high-TS tails.

As we discuss in Appendix C, our TS in Eq. (15) is related to a maximum log-likelihood ratio of the set of LG strings $K_{\max}(\mathbf{s})$, assuming that their true value is limited by the boundary $n - 2$ under the background hypothesis. The RMS z -scores are then equivalent to the reduced χ^2 where the degree of freedom (dof) corresponds to the number of LGI violations N_{LGV} . For LGI violations driven by quantum mechanics (and not by measurement uncertainties), we expect that the reduced χ^2 exceeds 1. This is clearly noticeable in the TS distributions of the oscillation model and the MINOS/MINOS+ data. However, also the TS distributions for the background model $\mathcal{H}_0^a$ have a median that marginally exceed 1, related to the optimization of the sign vector $\boldsymbol{\sigma}$ in Eq. (11) introducing a bias toward larger z -scores above the boundary $n - 2$.

VI. DISCUSSION

The RMS z -score from Eq. (15) allows us to estimate the significance of LGI violations in neutrino oscillation data. For the TS distribution of individual families of LG sequences “ $n_a + n_b$ ”, the p -value of the violation of LGI can be determined by the fraction of background pseudosamples with a TS larger than the TS observed for the actual data, as shown in the bottom right panel of Fig. 3. The five p -values that we determine this way for each background model can—in general—not be considered as independent trials, since they are based on the same pseudosamples, each consisting of only 39 data points. Conservatively, we estimate the post-trial p -value from the minimum p -value $p_{\min}$ as $p_{\text{post}} = 1 - (1 - p_{\min})^5$.

This yields a post-trial significance of LGI violations of 2.1σ (3.7σ) in comparison to the background model $\mathcal{H}_0^a$ ($\mathcal{H}_0^b$) with $p_{\min} \simeq 2.0 \times 10^{-2}$ (8.9×10^{-5}).

We argue that the background case $\mathcal{H}_0^a$ corresponds to the *most robust estimate* for the significance of LGI violations in neutrino oscillations, reflected by the smaller post-trial significance. A background expectation value $C = 1$ guarantees that the maximal LG string $K_{\max}$ falls closed to the boundary $n - 2$ of LGI for any sequence $\mathbf{s}$. Consequently, the pseudosamples derived from this background model allow us to directly estimate the impact of measurement uncertainties on the chance violation of LGIs. On the other hand, the background case $\mathcal{H}_0^b$ with exponentially decaying correlations has a lower chance for LGI violations, as visible in the bottom panels of Fig. 3. In particular, for larger n the chance of violation of LGIs in the background cases reduces significantly.

It is important to emphasize that a test of LGI violation in neutrino flavor evolution is not equivalent to a test of neutrino oscillations *per se*. A constant flavor correlation $C = 1$ is not consistent with the MINOS/MINOS+ data in Fig. 2. Describing the goodness of fit of the three models shown in Fig. 2 to the original MINOS/MINOS+ data by χ^2 values, we see that the best-fit Markovian model with $\chi^2/\text{dof} \simeq 5.8$ is a significantly better description of the MINOS/MINOS+ data than the constant model, which has $\chi^2/\text{dof} \simeq 31$. The best-fit oscillation model assuming normal ordering has $\chi^2/\text{dof} \simeq 1.8$. These different goodness-of-fit values are also reflected by the different number N_{LGV} of LGI violations of these models (left column of Fig. 3). The normalization of the TS of Eq. (15) via N_{LGV} allows us to focus on the *quality* of the LGI violation, not their *quantity*.

Before concluding, we would also like to comment on the phase tolerance ε that determines the number of sequences $\mathbf{s}$ of LG strings. For comparison to earlier studies we choose $\varepsilon = 0.5\%$ as our benchmark value. The flavor correlation data is effectively averaged over a timescale corresponding to the size of the energy bins and this sets an upper limit for the required resolution of the phase matching. The relative energy uncertainty $\Delta E/E$ of the MINOS/MINOS+ data reaches 3–6% in the 2–20 eV energy region, so the choice of $\varepsilon = 0.5\%$ is conservative. Switching to $\varepsilon = 5\%$ with a larger number of phase combinations (see Table I) leaves our results practically unchanged.

VII. CONCLUSIONS

We discussed a test of Leggett-Garg inequalities (LGIs) in neutrino flavor oscillations. These inequalities relate the correlation of consecutive measurements in an evolving system and are analogous to Bell's inequalities of measurements performed on spatially separated systems. The violation of LGIs in neutrino oscillations serves as a test of the assumption of *macrorealism* in the broader sense and

constitutes a fundamental test of the quantum nature of neutrinos.

Our work improves the methodology of previous analyses of LGIs in several ways. We derived a generalized family of Leggett-Garg strings that lead to an optimal test for the violation of LGIs for a given set of neutrino flavor measurements. We provided an improved method to simulate pseudodata following a statistical background hypothesis of macrorealistic systems, which is not biased toward unphysical flavor correlations and their corresponding spurious violations of LGIs. We introduced a test statistic that allows us to give a robust estimate of the significance of LGI violations in neutrino data.

We illustrated our revised methodology based on muon neutrino survival data by the MINOS detector. Using our method we find that MINOS/MINOS+ data [15] violates LGI with a significance at the level of $(2\text{--}3)\sigma$, depending on background model. The main reason for the reduced significance compared to the earlier claim of 6.2σ of Ref. [7] comes from our frequentist approach to quantify the level of LGI violation on data following a background hypothesis, rather than introducing a classical LG string as a revised test statistic. Other work based on accelerator and reactor neutrino data has either used the same background method [9] or neglected a background model altogether [8,10]. Our findings indicate that also these results should be critically reexamined with respect to the statistical level of LGI violations.

ACKNOWLEDGMENTS

M. A., K. M. G., and J. I.-N. acknowledge support by Villum Fonden (No. 18994). D. J. K. acknowledges support from the Carlsberg Foundation (No. 117238). J. I.-N. performed part of this work as a visiting researcher supported by the Erasmus+ program of the European Union.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available. The data are available from the authors upon reasonable request.

APPENDIX A: SEQUENTIAL LEGGETT-GARG INEQUALITIES

By definition, MR assumes that the n noninvasive measurements Q_i follow a probability distribution $\rho(Q_1, \dots, Q_n)$. Since $Q_i = \pm 1$, we can label each of the 2^n possible outcomes by an integer:

$$\ell = \sum_{i=1}^n 2^{i-1} \frac{1 + \sigma_i Q_i}{2}, \quad (\text{A1})$$

where $\sigma_i = \pm 1$ is an arbitrary sign convention for *each* measurement Q_i . The terms $(1 + \sigma_i Q_i)/2 \in \{0, 1\}$ are the

bits ℓ_{i-1} in the binary representation of ℓ . The macro-realistic correlator $\mathcal{C}_{ij}$ from Eq. (1), of the measurements Q_i and Q_j is then related to the 2^n probabilities P_ℓ of observing the ℓ th configuration as:

$$\sigma_i \sigma_j \mathcal{C}_{ij} = \sum_{\ell=0}^{2^n-1} (-1)^{\ell_{i-1} + \ell_{j-1}} P_\ell, \quad (\text{A2})$$

with $\ell_i = \lfloor \ell/2^i \rfloor \bmod 2$.

We first start with the usual sign convention $\sigma_i = \pm 1$. For three observables Q_1 , Q_2 , and Q_3 the explicit expressions of the three different correlators are

$$\mathcal{C}_{12} = P_0 - P_1 - P_2 + P_3 + P_4 - P_5 - P_6 + P_7, \quad (\text{A3})$$

$$\mathcal{C}_{23} = P_0 + P_1 - P_2 - P_3 - P_4 - P_5 + P_6 + P_7, \quad (\text{A4})$$

$$\mathcal{C}_{13} = P_0 - P_1 + P_2 - P_3 - P_4 + P_5 - P_6 + P_7. \quad (\text{A5})$$

The LG string K_3 defined in Eq. (2) is now

$$K_3 = \mathcal{C}_{12} + \mathcal{C}_{23} - \mathcal{C}_{13} = 1 - 4(P_2 + P_5) \leq 1, \quad (\text{A6})$$

where we have used $\sum_\ell P_\ell = 1$ in the last step.

Consider now the n th LG string for $n > 3$. We can relate K_n to the previous string K_{n-1} via

$$K_n = K_{n-1} + \mathcal{C}_{1n-1} + \mathcal{C}_{n-1n} - \mathcal{C}_{1n}. \quad (\text{A7})$$

Using the LGI for K_3 we know that

$$\mathcal{C}_{1n-1} + \mathcal{C}_{n-1n} - \mathcal{C}_{1n} \leq 1. \quad (\text{A8})$$

We therefore arrive at the LGI for the n th LG string

$$K_n \leq K_{n-1} + 1 \leq n - 2, \quad (\text{A9})$$

where we have used the LGI for K_{n-1} in the last step.

Finally, instead of the sign convention $\sigma_i = 1$ that we used in the preceding derivation, we would arrive at the same result by replacing the classical correlators in the LG string by $\sigma_i \sigma_j \mathcal{C}_{ij}$ which appears on the left-hand side of Eq. (A2). Since the sign vector σ yields the same expression as the reverse $-\sigma$ we can focus on the number of sign combinations with $\sigma_1 = 1$. If we now define $\sigma'_i \equiv \sigma_i \sigma_{i+1}$ we arrive at the generalized LG strings of Eq. (4) where the signs are arbitrary but need to obey the relation in Eq. (5).

APPENDIX B: VIOLATION OF LEGGETT-GARG INEQUALITIES

Consider the (pure) neutrino state $|\psi\rangle$ that is initially ($t = 0$) in an arbitrary superposition of flavor states. A first measurement at time t_1 can yield a neutrino of flavor state ν_μ ($|+\rangle$) or either ν_e or ν_τ ($|-\rangle$). A second measurement at

$t_2 > t_1$ yields one of the two flavor groups ($\pm'$). According to the Born rule, the probabilities of measuring the four flavor group combinations at times t_1 and t_2 are:

$$P(\pm, \pm') = \langle \psi | \hat{\pi}_\pm(t_1) \hat{\pi}_{\pm'}(t_2) \hat{\pi}_\pm(t_1) | \psi \rangle, \quad (\text{B1})$$

where we define the Heisenberg operator $\hat{\pi}_\pm(t) = U(t)^\dagger \hat{\pi}_\pm U(t)$ with time evolution operator $U(t)$ and projection operator $\hat{\pi}_\pm \equiv (1 \pm \hat{Q})/2$. The quantum-mechanical correlator at times t_1 and t_2 can then be written as

$$\mathcal{C}_{12} = P(+, +) + P(-, -) - P(+, -) - P(-, +), \quad (\text{B2})$$

which is equivalent to Eq. (3) with $\hat{Q}_i \equiv U^\dagger(t_i) \hat{Q} U(t_i)$. We can generalize this to a mixed ensemble of states as $\mathcal{C}_{12} = \text{Tr}(\{\hat{Q}_1, \hat{Q}_2\} \rho)/2$ with density matrix ρ .

The time-evolution of $\hat{Q}$ is equivalent to a rotation of a Pauli spinor. A general time-independent two-flavor Hamiltonian, e.g., neglecting spatial variations of the matter potential during neutrino propagation, takes the form $\hat{H} = (a + \mathbf{b} \cdot \boldsymbol{\sigma})/2$ where $\boldsymbol{\sigma}$ represents a vector of Pauli matrices and a ($\mathbf{b}$) a real scalar (vector) with dimensions of energy. From this, we obtain

$$\hat{Q}(t) = (\mathbf{n} \cdot \boldsymbol{\sigma}) n_z (1 - \cos \phi) + \cos \phi \sigma_z + \sin \phi (\mathbf{n} \times \boldsymbol{\sigma})_z, \quad (\text{B3})$$

with $\mathbf{n} = \mathbf{b}/|\mathbf{b}|$ and $\phi = t|\mathbf{b}|$. The correlator in Eq. (B2) is then

$$\mathcal{C}_{12} = n_z^2 + (1 - n_z^2) \cos(\phi_1 - \phi_2). \quad (\text{B4})$$

For vacuum oscillations with mixing $|\nu_1\rangle = \cos \theta |+\rangle + \sin \theta |-\rangle$ and $|\nu_2\rangle = \cos \theta |-\rangle - \sin \theta |+\rangle$ between two effective mass states $|\nu_i\rangle$ with mass splitting Δm^2 and energy E_ν we have $\mathbf{b} = \omega(\sin 2\theta, 0, \cos 2\theta)$ with oscillation frequency $\omega = |\mathbf{b}| = \Delta m^2/2E_\nu$ and:

$$\mathcal{C}_{12} = \cos^2 2\theta + \sin^2 2\theta \cos(\omega(t_2 - t_1)). \quad (\text{B5})$$

The maximal violation of LGIs can be observed under maximal mixing, i.e., mixing angle $\theta = \pi/4$.

APPENDIX C: BOUNDARY TEST STATISTIC

We motivate our test statistic in Eq. (15) as follows. Assuming that n data x_i are normal distributed with standard deviation Δx_i and expectation values μ_i , we define the likelihood as the product of normal distributions

$$\mathcal{L}(\boldsymbol{\mu} | \mathbf{x}) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi} \Delta x_i} \exp\left(-\frac{(x_i - \mu_i)^2}{2(\Delta x_i)^2}\right). \quad (\text{C1})$$

In the following we are agnostic about the exact model predictions μ_i . However, under our background hypothesis $\mathcal{H}_0$ we assume that the background model is bounded to

values $\mu_i \leq B$. Treating the μ_i as independent parameters, we can maximize the likelihood for the case of background ($\mathcal{H}_0$) and signal ($\mathcal{H}$) hypotheses as $\hat{\mu}_i = x_i$ and $\hat{\mu}_{0,i} = \min(B, x_i)$, respectively. The maximum log-likelihood-ratio is then

$$t \equiv -2 \ln \frac{\mathcal{L}(\hat{\mu}_0|\mathbf{x})}{\mathcal{L}(\hat{\mu}|\mathbf{x})} = \sum_{i=1}^n z_i^2, \quad (\text{C2})$$

with one-sided z -scores defined as

$$z_i \equiv \max\left(0, \frac{x_i - B}{\Delta x_i}\right), \quad (\text{C3})$$

which reproduces Eq. (14) after the replacements $B \rightarrow n - 2$ and $x_i \rightarrow K_{\max}(\mathbf{s})$.

For background data that follow normal distributions with mean $\mu_i = B$ we expect that Eq. (C2) follows a χ^2 -distribution with degree of freedom corresponding to the number $N_V \leq n$ of boundary violations, $x_i > B$. The number of violations N_V depends not only on the signal and background hypotheses, but also on the variance $(\Delta x_i)^2$ related to measurement uncertainties. This motivates us to define a test statistic as

$$t_V \equiv \begin{cases} t/N_V & N_V > 0, \\ 0 & N_V = 0, \end{cases} \quad (\text{C4})$$

equivalent to a reduced χ^2 and identical to the square of the RMS z -score in Eq. (15). For general background data sampled from expectation values $\mu_i \leq B$ the median t_V value will lie below 1. On the other hand, for data sampled from signal models where μ_i can (but not must) exceed the boundary, the median t_V value can significantly exceed 1.

The background model $\mathcal{H}_0^a$ introduced in Sec. IV with $C = 1$ produces LG strings that are expected to scatter around the boundary $B = n - 2$. We therefore consider this model as the most conservative estimate for the TS distribution of data following macrorealistic background hypotheses. Note that we determine the significance of LGI violations from the TS distributions of pseudosamples following the background and signal hypotheses. This construction does not rely on detailed statistical properties of Eq. (C4), but rather on its tendency to increase beyond 1 in the presence of signal regardless of the overall number of violations.

-
- [1] J. S. Bell, *Phys. Phys. Fiz.* **1**, 195 (1964).
 - [2] A. J. Leggett and A. Garg, *Phys. Rev. Lett.* **54**, 857 (1985).
 - [3] A. J. Leggett, *J. Phys. Condens. Matter* **14**, R415 (2002).
 - [4] C. Emary, N. Lambert, and F. Nori, *Rep. Prog. Phys.* **77**, 016001 (2013).
 - [5] D. Gangopadhyay, D. Home, and A. S. Roy, *Phys. Rev. A* **88**, 022115 (2013).
 - [6] D. Gangopadhyay and A. S. Roy, *Eur. Phys. J. C* **77**, 260 (2017).
 - [7] J. A. Formaggio, D. I. Kaiser, M. M. Murskyj, and T. E. Weiss, *Phys. Rev. Lett.* **117**, 050402 (2016).
 - [8] R. Z. Barrios and M. A. Acero, *arXiv:2401.00240*.
 - [9] Q. Fu and X. Chen, *Eur. Phys. J. C* **77**, 775 (2017).
 - [10] X.-Z. Wang and B.-Q. Ma, *Eur. Phys. J. C* **82**, 133 (2022).
 - [11] Xue K. Song, Y. Huang, J. Ling, and Man H. Yung, *Phys. Rev. A* **98**, 050302 (2018)(R).
 - [12] J. Naikoo, A. K. Alok, S. Banerjee, and S. U. Sankar, *Phys. Rev. D* **99**, 095001 (2019).
 - [13] S. Shafaq, T. Kushwaha, and P. Mehta, *arXiv:2112.12726*.
 - [14] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, New York, 2007).
 - [15] A. B. Sousa (MINOS, MINOS+ Collaborations), *AIP Conf. Proc.* **1666**, 110004 (2015).
 - [16] I. Esteban, M. C. Gonzalez-Garcia, M. Maltoni, T. Schwetz, and A. Zhou, *J. High Energy Phys.* **09** (2020) 178.