Statistical significance of the sterile-neutrino hypothesis in the context of reactor and gallium data

Jeffrey M. Berryman,[a,b,c] **Pilar Coloma,**[d] **Patrick Huber,**[e] **Thomas Schwetz**[f] **and Albert Zhou**[f]

[a]*Department of Physics, University of California,*
366 Physics North MC 7300, Berkeley, CA, U.S.A.

[b]*Department of Physics and Astronomy, University of Kentucky,*
505 Rose St., Lexington, KY, U.S.A.

[c]*Institute for Nuclear Theory, University of Washington,*
3910 15th Ave. NE, Seattle, WA, U.S.A.

[d]*Instituto de Física Teórica UAM-CSIC,*
Calle Nicolás Cabrera 13-15, Madrid 28049, Spain

[e]*Center for Neutrino Physics, Physics Department, Virginia Tech,*
850 W Campus Dr, Blacksburg, Virginia, U.S.A.

[f]*Institut für Astroteilchen Physik, Karlsruher Institut für Technologie (KIT),*
Hermann-von-Helmholtz-Platz 1, Eggenstein-Leopoldshafen 76344, Germany

E-mail: jeffberryman@berkeley.edu, pilar.coloma@ift.csic.es,
pahuber@vt.edu, schwetz@kit.edu, albert.zhou@kit.edu

Abstract: We evaluate the statistical significance of the 3+1 sterile-neutrino hypothesis using ν_e and $\bar{\nu}_e$ disappearance data from reactor, solar and gallium radioactive source experiments. Concerning the latter, we investigate the implications of the recent BEST results. For reactor data we focus on relative measurements independent of flux predictions. For the problem at hand, the usual χ^2-approximation to hypothesis testing based on Wilks' theorem has been shown in the literature to be inaccurate. We therefore present results based on Monte Carlo simulations, and find that this typically reduces the significance by roughly 1σ with respect to the naïve expectation. We find no significant indication in favor of sterile-neutrino oscillations from reactor data. On the other hand, gallium data (dominated by the BEST result) show more than 5σ of evidence supporting the sterile-neutrino hypothesis, favoring oscillation parameters in agreement with constraints from reactor data. This explanation is, however, in significant tension ($\sim 3\sigma$) with solar neutrino experiments. In order to assess the robustness of the signal for gallium experiments we present a discussion of the impact of cross-section uncertainties on the results.

Contents

1 Introduction

There is a number of long-standing experimental results in neutrino physics that defy Standard Model (SM) expectations, and at least individually find an explanation by new oscillations with a mass-squared splitting of around $1\,\text{eV}^2$; for reviews see e.g., refs. [1, 2]. This scale is much larger than the mass-squared splittings associated with the three neutrinos of the SM, and hence points to the existence of a forth neutrino, which must be a singlet under the SM gauge group (i.e., sterile), in order to agree with the invisible decay width of the Z-boson [3]. A sterile neutrino would open a fermionic portal to a dark sector and, as such, is well motivated.

In 2011, a new piece of evidence appeared in the form of the so-called *reactor antineutrino anomaly* [4]: new evaluations of the prediction of the antineutrino flux from a reactor [5, 6] resulted in a 6% increase of the expected neutrino signal. This implied

that past experiments (which were in agreement with previous calculations) were in fact observing a deficit of the same magnitude. In parallel, renewed analyses of radiochemical source-calibration measurements done for the solar-neutrino experiments SAGE [7, 8] and GALLEX [9, 10] also indicated a deficit compatible with the result from reactors: the so-called *gallium anomaly* [11–13]. Thus, for the first time, there was considerable evidence of anomalous ν_e and $\bar{\nu}_e$ disappearance.

Because of this, the past decade has seen a plethora of experimental efforts (both with reactor neutrinos and radioactive sources), as well as theoretical efforts in nuclear physics, to improve source and cross-section predictions. On the experimental side, a number of reactor experiments at short distances (6–25 m) from the reactor core have been conducted: DANSS [14], NEOS [15], PROSPECT [16], STEREO [17] and Neutrino-4 [18]. These experiments derive their results for sterile-neutrino oscillations by comparing the measured neutrino spectra at different distances from the reactor, and thus their results do not rely on reactor-flux modeling. While some of these reported evidence favoring the sterile-neutrino hypothesis, others did not, and thus the question of the statistical compatibility of the different data sets and their interpretation in the form of a global fit is relevant [19, 20]. All of these experiments look for an oscillatory distortion in the ratio of measured spectra, and there are a number of subtleties that arise in their statistical interpretation [21–24], which render the usual approach based on Wilks' theorem [25] unreliable.

In parallel, in order to test the gallium anomaly, the BEST collaboration recently presented a new result of a two-zone gallium radiochemical measurement employing a chromium-51 source [26]. Although they did not observe a different rate in the two volumes, their results confirmed the earlier deficit with much smaller error bars. This new result brings the gallium anomaly to a new level, making it highly significant and raising the question of its compatibility with other data [27, 28].

Since 2011, significant effort has gone into improving our understanding of reactor antineutrino fluxes, both from a theoretical perspective and from many measurements of beta-feeding functions of individual isotopes. These efforts have led to a much better understanding of the underlying physics, see e.g., ref. [29], but have not resolved the reactor antineutrino anomaly. Global analyses of past reactor neutrino data [30–32], precise neutrino-spectrum measurements [33] and fuel-evolution measurements [34] have been indicating for a while that there might be a problem with the prediction for uranium-235. The recent absolute flux determination from a reactor with only uranium-235 fission by the STEREO experiment [35] confirms this. The basis for the Huber-Mueller reactor fluxes are integral beta-spectrum measurements conducted in the 1980s at the Institute Laue-Langevin (ILL) [36–38]. Recently, the ratio of uranium-235 to plutonium-239 integral beta spectra was measured [39], indicating the possibility that the normalization of the original ILL data for uranium-235 was too high. Taking this new information into account, a global flux analysis [28] finds that there is no evidence for the reactor antineutrino anomaly in total rates. All data sets and methods of prediction agree in total rate if one replaces the normalization of uranium-235 with the new result of ref. [39]. Therefore, any support for the sterile-neutrino hypothesis from reactor neutrino data now would have to come from the observation of an actual oscillation by DANSS, NEOS, PROSPECT, STEREO or

Neutrino-4. Note that the spectrum discrepancy between flux models and reactor neutrino measurements [40–42] around 5 MeV of neutrino energy, colloquially known as the 5 MeV bump, has no bearing on the question of a sterile neutrino for these experiments, since in taking the ratio of spectra measured at different distances, the bump effectively cancels.

Finally, a number of electron-neutrino disappearance results from accelerator experiments may also be interpreted in the context of a sterile neutrino. Specifically, both KARMEN and LSND [43–45] have provided bounds from a charged-current reaction on ^{12}C. A measurement of the T2K experiment reported in ref. [46], its main purpose being a determination of the charged-current electron-neutrino cross section and characterization of the ν_e beam component, also leads to a mild indication of disappearance. A somewhat similar result appeared recently based on MicroBooNE data [47], see also ref. [48]. All of these results are not very statistically significant, as of yet, and do not strongly favor oscillation over no-oscillation nor vice versa; we therefore do not include them in our fits. Similar remarks apply to the tritium beta-decay experiment KATRIN [49], whose exclusion depends strongly on whether the light neutrino masses can be neglected with respect to the mass of the much heavier fourth neutrino [49, 50].

Let us also remark that additional eV-scale neutrinos face severe constraints from cosmology, both in regard to their early-time contributions as relativistic degrees of freedom and their late-time contributions to the energy density as hot dark matter, suppressing the formation of small-scale structure, see e.g., ref. [51] for a recent global analysis and refs. [1, 2] for reviews and further references. However, any cosmological bound relies on the assumption of thermalization of the sterile neutrino in the early universe and may thus be evaded in an alternative cosmology. Our goal here is to present results that are robust with respect to cosmological assumptions.

In this paper, we present a global fit of up-to-date results from the DANSS, NEOS, PROSPECT, STEREO and Neutrino-4 reactor experiments, the past gallium results from SAGE and GALLEX, as well as the recent BEST result and solar-neutrino results in the framework of one additional sterile neutrino (the so-called 3+1-oscillation hypothesis). The results presented are independent of reactor-flux predictions and are thus complementary to the results of ref. [28]. Since Wilks' theorem can not be used reliably here [21–24], we perform a full Feldman-Cousins analysis [53].

The analysis results are presented in the main text; individual experiments, as well as details of the Feldman-Cousins analysis, are described in the appendices. Section 2 contains a description of the reactor data, as well as the results of their combination. In section 3, we discuss the gallium experiments and the dependence of their results with the charged-current cross section for neutrino capture. In section 4, we evaluate the statistical compatibility of the various data sets, paying particular attention to the compatibility between solar and gallium data, and present our global-fit results. In section 5, we present a summary of our findings and draw our conclusions.

2 Reactor experiments

2.1 Description of the data used

In the following, we provide a brief summary of the reactor experiments whose data we have used; detailed descriptions of our analyses — including the precise forms of the χ^2 functions and our techniques for generating pseudo data — are presented in appendix A.

DANSS. At DANSS, the detector is located on a movable platform under an industrial $3.1\,\mathrm{GW_{th}}$ reactor [14]. It measures the neutrino spectrum at three baselines: bottom, middle and top ($I \equiv B, M, T$, respectively). The DANSS collaboration presents their data with the following ratios

$$R_1^i = \frac{n_B^i}{n_T^i} \qquad\qquad R_2^i = \frac{n_M^i}{\sqrt{n_B^i n_T^i}}, \qquad\qquad (2.1)$$

where i indicates the energy bin and $n_I^i = N_I^i/\Delta t_I$. Here, N_I^i stands for the event rates the energy bin i and Δt_I is the corresponding exposure time period for each detector location I. A total of 36 energy bins are used in our analysis, spanning energies between 1.5–6 MeV. Exposure times and the background-subtracted rates for the top baseline were extracted from ref. [14, p. 12]. Doing this, we get a total event rate for the top baseline of $\sum_i n_T^i = 4132.61$ events per day, for energies between 1.5–6 MeV (in prompt energy). Details on our χ^2 implementation can be found in appendix A.1.

NEOS. The NEOS detector consists of approximately 1000 L of homogeneous 0.5% Gd-loaded liquid scintillator. It is located at about $23.7\,\mathrm{m}$ from a commercial reactor core with a thermal power of $2.8\,\mathrm{GW}$ in the Hanbit nuclear-power complex in Yeonggwang, South Korea. The detector operated for 46 days with the reactor off and 180 days with the reactor on. NEOS measured 1976.7 ± 3.4 inverse-beta-decay candidates per day (with prompt energies between 1 and $10\,\mathrm{MeV}$) during the reactor-on period, as opposed to 85.1 ± 1.4 events in the reactor-off period. For our analysis, we take the data from ref. [15, figure 3a]: the black points are assumed to be the background-subtracted counts and are normalized to the quoted total signal rate, $1976.7 - 85.1 = 1891.6$ events/day.

The NEOS collaboration does a shape-only analysis using the neutrino spectrum from ref. [54] plus corrections due to different fission fractions (see appendix A.2 for details); therefore, we leave the normalization completely free in our fit. We build a binned χ^2, with 61 energy bins[1] between 1–10 MeV.

PROSPECT. The PROSPECT experiment uses a large, highly segmented detector. Segments are grouped into 10 different baselines (see figure 39 of ref. [16]). Data are then reported as 10 energy spectra — one for each cluster of segments — with 16 energy bins. Our analysis benefits from the thorough data release accompanying ref. [16], including the background-subtracted spectra for each baseline bin and the covariance matrix describing them, which includes both statistical and systematic uncertainties. The χ^2 is calculated following ref. [16, eq. (11)] (see also appendix A.3).

---

[1]Note that the last bin is *not* included in the total rate stated in the paper, but *is* used in the analysis.

STEREO. The STEREO detector contains six cells, each with a different effective baseline. The experiment was divided up into two phases: for each cell, phase-I data are presented in ten bins from 1.625 to 6.625 MeV of prompt energy, while phase-II data are presented in eleven bins from 1.625 to 7.125 MeV (in both cases the bin sizes are set to 0.5 MeV). The data have been published alongside ref. [17]; however, the raw event rates have been normalized to the prediction under the no-oscillation hypothesis (after minimization over systematic uncertainties). Therefore, we also rescaled our predictions in order to compare to their data, using as the nuisance parameters (for the prediction under the no-oscillation hypothesis) the values from figure 31 of ref. [17]. The STEREO analysis contains a total of 40 pulls: we are able to minimize 14 of these analytically, while the remaining 26 are minimized numerically. Further details on our implementation of STEREO can be found in appendix A.4.

Neutrino-4. Our analysis of Neutrino-4 [18] is based on the previous analysis performed in ref. [24]. Neutrino-4 also uses a segmented detector. The data are binned using nine 0.5 MeV energy bins from 2.3 to 6.8 MeV, and twenty-four 0.235 m baseline bins from 6.25 to 11.89 m. The resulting 216 bins in L/E are then clustered into consecutive groups of 8 bins, leading to a total of 27 data points. The analysis is performed using the ratios

$$R_i = \frac{d_i}{\frac{1}{N}\sum_j d_j},$$ (2.2)

where d_i is the number of events in the ith bin. In our analysis, we use the first 19 bins for the combined phase-1 and phase-2 data sets (the blue points in figure 47 of the v2 arXiv preprint of ref. [18]). Note that although the event counts d_i are uncorrelated, the ratios R_i are correlated through the denominator in eq. (2.2). However, we have estimated these correlations and find that they do not change our results, so they will be neglected here. Further details on our χ^2 implementation and analysis can be found in appendix A.5 as well as in ref. [24].

2.2 Results of the reactor analysis

For all the reactor data listed above, we can reproduce the results of the corresponding collaborations with good accuracy when adopting the same assumptions and statistical method. This is shown in table 1, which compares our best-fit points to the ones from the respective collaborations, showing reasonable agreement in all cases. The χ^2 values of the global minimum ($\chi^2_{\min}$) are also shown: as can be seen, our results match very well those obtained by the different collaborations in all cases. Notice the slightly different number of dof between our analysis and the collaboration's for the NEOS and PROSPECT analyses, which is due to a different number of nuisance parameters used in the fit.

In the right-most part of the table, we evaluate the data's preference for the presence of sterile-neutrino oscillations by considering the test statistic[2]

$$\Delta\chi^2_{3\nu} = \chi^2_{3\nu} - \chi^2_{\min},$$ (2.3)

---

[2]This test statistic has been denoted by T in ref. [24].

	$\chi^2_{\min}/\text{dof}$	$\Delta m^2_{\min}$	$\sin^2 2\theta_{\min}$	$\Delta\chi^2_{3\nu}$	p_0	$\#\sigma$	$\#\sigma^{(W)}$
DANSS		$1.3\ \text{eV}^2$	0.014	3.2			1.3
Our Fit	77.6/70	$1.32\ \text{eV}^2$	0.011	3.6	43.8%	0.8	1.4
NEOS	57.5/59	$1.73\ \text{eV}^2$	0.05	6.5	22%	1.2	2.1
Our Fit	59/58	$2.95\ \text{eV}^2$	0.16	7.4	19.4%	1.3	2.2
PROSPECT	119.3/142	$1.78\ \text{eV}^2$	0.11	4.0	57%	0.6	1.5
Our Fit	118.1/158	$1.75\ \text{eV}^2$	0.11	4.3	63.3%	0.5	1.6
STEREO	128.4/112	$8.95\ \text{eV}^2$	0.63	9.0	9%	1.7	2.5
Our Fit	128.6/112	$8.72\ \text{eV}^2$	0.59	7.8	15.8%	1.4	2.3
Nu-4	14.7/17	$7.26\ \text{eV}^2$	0.38	15.3			3.5
Our Fit	16.1/17	$7.31\ \text{eV}^2$	0.38	12.6	1.5%	2.4	3.1
REACTORS	428/421	$8.86\,\text{eV}^2$	0.26	7.3	27.4%	1.1	2.2
W/ Solar	432/425	$1.30\,\text{eV}^2$	0.014	6.6	17.8%	1.3	2.1
W/ Gallium[a]	433/427	$8.86\,\text{eV}^2$	0.32	38.8	$(0.14 \to 1.4)\times 10^{-7}$	$5.7 \to 5.3$	5.9

[a]The reactor-with-gallium p_0 values and significances are extrapolations and not MC calculations. The lower p-value assumes a maximum-Gauss distribution with $n = 60$ [24], the upper p-value corresponds to a linear extrapolation of the tail of the distribution; see figure 6.

Table 1. Comparison of our fit results with those from the individual collaborations (where available) for DANSS [14, p. 17], NEOS [15], PROSPECT [16, section VIII.C], STEREO [17, table 5] and Neutrino-4 [18, section 21 of v2 of arXiv preprint]. The first three columns provide the χ^2 value at the global minimum and the corresponding oscillation parameters. The remaining columns show the $\Delta\chi^2$ of the null hypothesis (no new oscillations), the p-value of the null hypothesis calculated from MC simulations (p_0), and its equivalent number of Gaussian standard deviations, $\#\sigma$. The last column shows the corresponding number of standard deviations when Wilks' theorem is assumed, i.e., assuming $\Delta\chi^2_{3\nu}$ is distributed as $\chi^2(2\,\text{dof})$. We also show our combined results for all reactors, as well as reactors with solar, and reactors with gallium data. The gallium analysis uses the Kostensalo et al. cross section [55].

where $\chi^2_{3\nu}$ is the χ^2 value for the null hypothesis, $\sin^2 2\theta = 0$ (θ being the angle that parametrizes the mixing between the sterile neutrino and the light states), and $\chi^2_{\min}$ is the χ^2 value at the best-fit point. We use this test statistic to evaluate the p-value of the null hypothesis in the following way. For a given experiment, the distribution that $\Delta\chi^2_{3\nu}$ follows can be obtained from Monte Carlo (MC) simulations: a set of pseudo experiments is randomly generated according to the expected statistical fluctuations of event rates in each bin (see appendix B.2 for details). The obtained distribution of the test statistic can then be compared to the experimental value in order to obtain the corresponding p-value. The p-values we obtain, as well as the corresponding number of Gaussian standard deviations, are given in table 1. Again, the comparison with the collaborations' corresponding results shows good agreement.
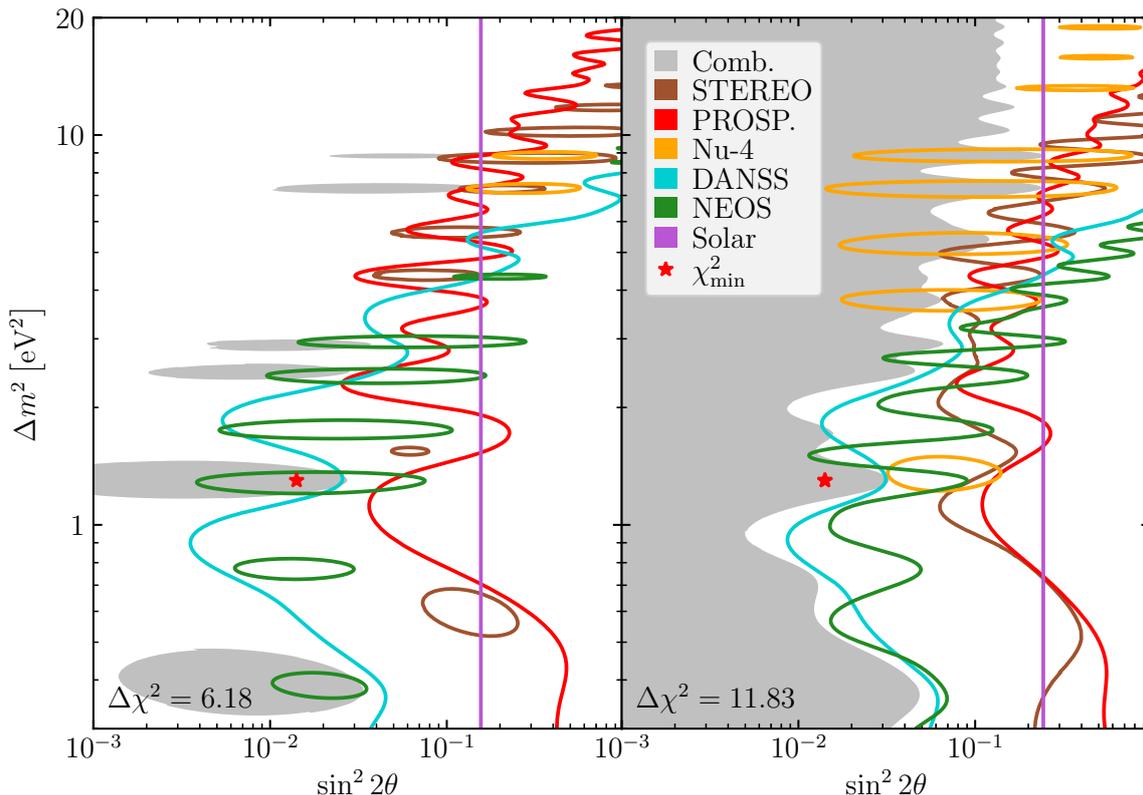
Figure 1. Isocontours of $\Delta\chi^2 = 6.18$ (left) and 11.83 (right) for reactor experiments and the solar constraint, corresponding to 2, 3σ levels under Wilks' theorem, respectively. The grey regions correspond to the combined analysis and the red star is the best-fit point of all data combined.

If Wilks' theorem held, then $\Delta\chi^2_{3\nu}$ would follow a χ^2 distribution with 2 dof (corresponding to the two parameters, $\sin^2 2\theta$ and Δm^2, which we minimize over). In the last column of table 1, we show the number of standard deviations corresponding to that assumption, $\#\sigma^{(W)}$. Comparing with the previous column, we find that Wilks' theorem consistently overestimates the significance by about one standard deviation, confirming previous studies [21–24]. Therefore, in order to assess the significance for the presence of sterile neutrinos from these data, it is essential to study the distribution of the test statistic numerically.

Next, let us discuss the differences in the allowed confidence regions obtained in the $\sin^2 2\theta$–Δm^2 plane. Figure 1 shows the contours for constant $\Delta\chi^2 = 6.18$ and 11.83 for the various reactor experiments individually (as well as their combination) together with the constraint from solar neutrinos (see below), where

$$\Delta\chi^2(\sin^2 2\theta, \Delta m^2) = \chi^2(\sin^2 2\theta, \Delta m^2) - \chi^2_{\min}. \qquad (2.4)$$

If Wilks' theorem were valid, then these contours would correspond to 95.45% and 99.73% confidence regions, respectively. However, since Wilks' theorem cannot be applied here, the confidence regions should be determined from simulation, according to the Feldman-Cousins (FC) prescription [53] as follows. For a given pair of assumed true values of $\sin^2 2\theta$

and Δm^2, MC simulations of statistical fluctuations in the data can be used to determine the expected distribution of $\Delta \chi^2$ as defined in eq. (2.4) and compared to the value measured experimentally, in order to determine the confidence level (CL) at which a given point of parameter space can be rejected. The full allowed region in the $\sin^2 2\theta$–Δm^2 plane is then obtained by repeating this procedure for all points in parameter space. Confidence regions obtained in this way have the correct coverage by construction.

The colored bands in figure 2 show such FC confidence regions, corresponding to 1, 2, and 3σ significance, for the individual reactor experiments (as well as for their combination, in the last panel), where the width of the bands indicates the 99% spread due to the finite size of the MC sample (see appendix B.4 for details). Comparing these bands to the lines of the same color, which correspond to $\Delta \chi^2$ contours assuming that Wilks' theorem applies, we find that in most cases the true CL is reduced approximately by one Gaussian standard deviation, similar to the null-hypothesis case. Also, we do not find closed confidence regions at 99.73% CL for any of the reactor experiments, in agreement with the p-values of the null hypothesis reported in table 1.

Although none of the individual experiments shows a clear signal at high significance, several of them favor the sterile-neutrino-oscillation hypothesis at a mild confidence level, $\lesssim 2\sigma$. Therefore, the question arises, whether these indications add up and lead to a significant result in a combined analysis. We saw already in figure 1 that the allowed regions for the different experiments do not show a consistent overlapping pattern and therefore an enhancement of the various hints is not expected from their combination. This is confirmed and quantified by the results of our global reactor MC analysis, reported in table 1 as well as in the bottom-right panel of figure 2: the combined reactor fit gives $\Delta \chi^2_{3\nu} = 7.3$ with a p-value corresponding to 1.1σ. Hence, we find that reactor data are statistically compatible with no sterile-neutrino oscillations. Again, we note that assuming Wilks' theorem we would still have a $\approx 2\sigma$ hint for sterile neutrinos, whereas the MC analysis reduces this to $\approx 1\sigma$.

This effect is further illustrated in figure 3, where we show the probability distribution function (PDF) and the survival function[3] (SF) of the $\Delta \chi^2_{3\nu}$ distribution for the combination of all reactor experiments, obtained numerically by generating 10^5 MC data sets. We see from the plot that the distribution deviates considerably from the χ^2 distribution with 2 dof. On the other hand, in ref. [24] it has been shown that for this type of oscillation search, under certain idealizations, the SF for $\Delta \chi^2_{3\nu}$ follows the so-called maximum-Gauss distribution, which is defined as the maximum value of n independent standard-normal variables. Therefore, in figure 3 we also compare the numerical SF with this distribution for $n = 80$ and 120 (here n plays the role of an effective number of degrees of freedom, and has been chosen in order to match the shape of the SF). Note that small deviations from the maximum-Gauss distribution occur since some of the assumptions under which it has been derived are not fulfilled exactly; for instance, due to non-trivial systematic uncertainties, the data points in our fit are not all statistically independent.

---

[3]The survival function is defined as $\mathrm{SF} = 1 - \mathrm{CDF}$, where CDF denotes the cumulative distribution function.
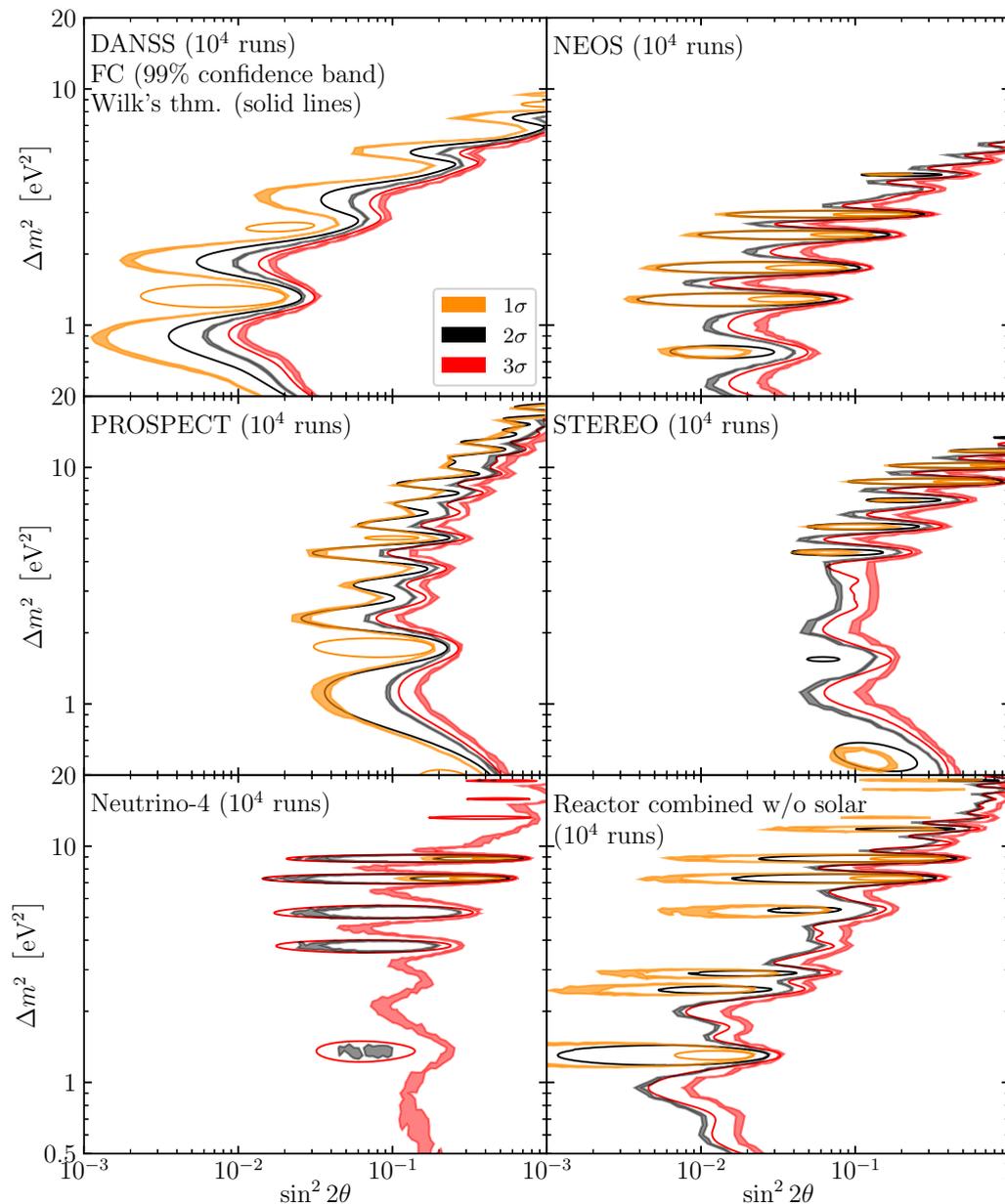
Figure 2. Confidence regions at 68.27%, 95.45%, 99.73% CL (1, 2, 3σ, respectively) for the individual reactor experiments and all reactors combined (bottom right). The bands correspond to the Feldman-Cousins analysis, where their width indicates the 99% confidence spread of the limit due to the finite size of our MC sample (see appendix B.4). Thin curves are obtained under the assumption of Wilks' theorem (assuming that $\Delta\chi^2$ follows a χ^2 distribution with 2 dof).

3 Gallium radioactive-source experiments

The gallium solar-neutrino detectors GALLEX and SAGE have been used to measure the neutrino rate from radioactive ^{51}Cr and ^{37}Ar sources [7–10], leading to results consistently lower than expected. This so-called gallium anomaly could potentially find its explanation
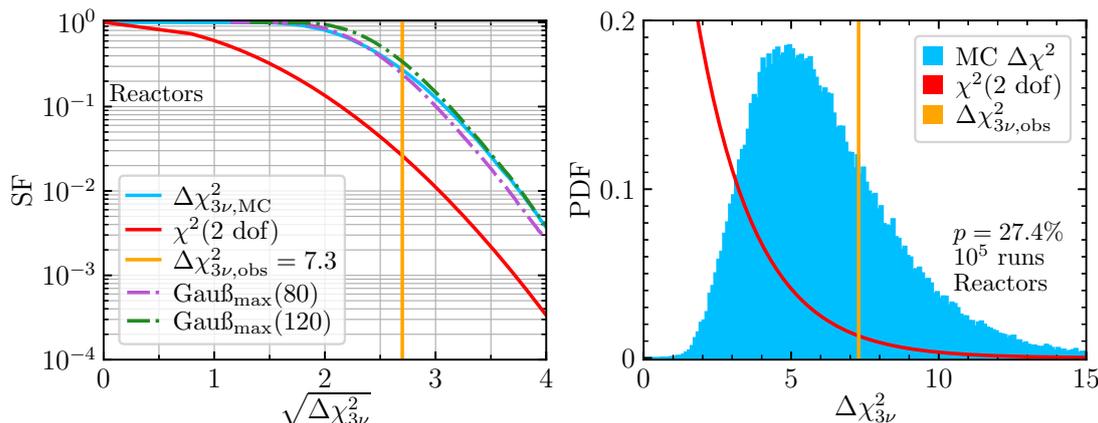
Figure 3. Survival function (SF, left) and probability distribution function (PDF, right) of $\Delta\chi^2$ under the null hypothesis for the combination of all reactor experiments for 10^5 runs (blue). We also plot the χ^2(2 dof) SF and PDF for comparison (red), as well as the maximum-Gauss distribution for $n = 80, 120$, see ref. [24].

GALLEX [10]	GALLEX [10]	SAGE [7]	SAGE [8]	BEST [26] (inner)	BEST [26] (outer)
0.953 ± 0.11	0.812 ± 0.10	0.95 ± 0.12	0.791 ± 0.084	0.791 ± 0.044	0.766 ± 0.045

Table 2. The ratio of observed to expected event rates for the gallium experiments. The quoted 1σ errors include statistical and uncorrelated experimental errors. The SAGE measurement [8] in the fourth column uses an ^{37}Ar source; all other measurements are from a ^{51}Cr source. The ratios are based on the Bahcall cross sections [57].

by sterile-neutrino oscillations [12, 56]. Recently, the BEST collaboration has performed a dedicated source experiment using a ^{51}Cr source embedded in a two-volume gallium detector, confirming the previous hints at high significance [26]; see also ref. [27].

We summarize the gallium data in table 2, quoting the ratios of observed to expected event rates. The errors in the table include only statistical and uncorrelated experimental errors. The errors for the two BEST data points are obtained by adding the statistical error given in table 1 of ref. [26] and the uncorrelated experimental systematic error of 2.6% in quadrature. We observe that the reported ratios deviate from unity by about 20%, confirmed by BEST at high significance.

An important ingredient in the interpretation of these results is the ν_e-capture cross section on gallium. The ratios in table 2 are based on the traditional Bahcall 1997 cross sections [57], but for alternative cross sections they have to be rescaled accordingly. Here we have considered two recent re-evaluations of the cross sections from Kostensalo et al. [55] and Semenov [58], summarized in table 3. Note that in table 3 we give the best estimates for the total cross section for the ν_e-induced ^{71}Ga $\rightarrow^{71}$Ge transition, which receives contributions from the ground-state-to-ground-state (g.s.) transition as well as from transitions into excited states of ^{71}Ge. Two remarks should be made at this point. First, the matrix element of the g.s. Transition can be directly inferred from the electron-capture lifetime

Reference	$\sigma(\mathrm{Cr})$	$\sigma_{\mathrm{g.s.}}(\mathrm{Cr})$	$\sigma(\mathrm{Ar})$	$\sigma_{\mathrm{g.s.}}(\mathrm{Ar})$
Bahcall [57]	58.1 ± 2.1	55.2	70.0 ± 4.9	66.2
Kostensalo et al. [55]	56.7 ± 1.0	55.3 ± 0.7	68.0 ± 1.2	66.2 ± 0.9
Semenov [58]	59.38 ± 1.16	55.39 ± 0.19	71.69 ± 1.47	66.25 ± 0.23

Table 3. Cross section for ν_e detection on gallium, weighted by the energy lines for neutrinos produced from a Cr and an Ar source, for our three representative cross-section models. Units are $10^{-46}\,\mathrm{cm}^2$. We give the best estimate for the total cross sections as well as the cross section corresponding to the ground state (g.s.) transition only. For the Bahcall numbers we quote the (larger) upper error (see text); Bahcall does not provide an error estimate on $\sigma_{\mathrm{g.s.}}$.

of $^{71}\mathrm{Ge}$ and the corresponding phase-space factor. Second, the excited-state contributions add incoherently and thus can only increase the cross section, that is, the g.s. transition provides a *lower bound* on the total cross section. Finally, note that the relatively large error in the g.s.-only cross section of Kostensalo et al. arises because it is based on their shell-model calculation and not on the observed electron-capture parameters. Therefore, the choice of this calculation as default will produce the most conservative result for the g.s.-only analysis. At the same time, this choice is not fully consistent, since the measured ground state strength is used to tune g_A for the shell model calculation of Kostensalo et al.. Semenov's ground state cross section and uncertainty would be a more consistent choice, but would lead to a much larger significance for the gallium anomaly (see below).

In order to analyze gallium data, we define the following χ^2 function:

$$\chi^2_{\mathrm{Gallium}} = \min_{\xi}\left[\sum_{i=1}^{6}\frac{(R_i - (1+\xi)P_i)^2}{\sigma_i^2} + \frac{\xi^2}{\sigma_\xi^2}\right] . \tag{3.1}$$

Here, R_i and σ_i are the observed ratios and their uncertainties from table 2, rescaled according to the adopted cross section, and P_i are the oscillation probabilities, averaged over the detector volumes and weighted by the neutrino-energy lines from the corresponding source. The pull parameter ξ describes the correlated uncertainty on the cross section.[4] In order to set the uncertainty σ_ξ we proceed as follows: for a given cross-section model, we first minimize with respect to ξ, adopting the total uncertainty given in table 3; we then check whether the resulting cross section at the pull minimum is smaller than the ground-state contribution; if this is the case, then we switch to the smaller uncertainty of $\sigma_{\mathrm{g.s.}}$, making the pull stiffer. In this way, we take into account the asymmetries of cross-section uncertainties. An analogous procedure is adopted when generating random values for the pull parameter for the MC studies.

In table 4, we report the results of our analysis of gallium data under the sterile-neutrino-oscillation hypothesis adopting different assumptions about the detection cross section. We give the value of $\Delta\chi^2_{3\nu}$ as defined in eq. (2.3) and the corresponding significance

---

[4]We assume the uncertainties of $\sigma(\mathrm{Cr})$ and $\sigma(\mathrm{Ar})$ to be fully correlated. The combined fit is dominated by BEST, and the impact of the SAGE argon datum is small; therefore this assumption has little impact on the result.

	GALLEX & SAGE		BEST		All gallium combined			
Cross section	$\Delta\chi^2_{3\nu}$	$\#\sigma^{(W)}$	$\Delta\chi^2_{3\nu}$	$\#\sigma^{(W)}$	$\sin^2 2\theta_{\min}$	$\Delta m^2_{\min}$	$\Delta\chi^2_{3\nu}$	$\#\sigma^{(W)}$
Bahcall [57]	3.7	1.4	31.3	5.2	0.35	$1.3\,\mathrm{eV}^2$	31.7	5.3
Kostensalo [55]	4.9	1.7	31.5	5.2	0.32	$1.3\,\mathrm{eV}^2$	32.9	5.4
Semenov [58]	9.4	2.6	42.4	6.2	0.39	$1.3\,\mathrm{eV}^2$	44.7	6.4
Ground state	3.4	1.3	29.7	5.1	0.29	$1.3\,\mathrm{eV}^2$	31.5	5.3

Table 4. Significance of the gallium radioactive-source measurements for the previous GALLEX and SAGE results combined, the recent BEST result, and the combination of all gallium data using different evaluations of the detection cross section. For each data set, we give the $\Delta\chi^2$ of the null hypothesis and the corresponding significance in terms of Gaussian standard deviations, which we evaluate using Wilks' theorem, i.e., assuming that the $\Delta\chi^2$ is distributed as $\chi^2(2\,\mathrm{dof})$. For the combined data, we also give the location of the oscillation parameters at the global minimum. For the ground-state-only analysis (last row) we assume the Semenov g.s. value and uncertainty (cf. table 3).

in units of Gaussian standard deviations. While previous results from GALLEX and SAGE show only a weak hint for sterile-neutrino oscillations, the significance of the BEST result is $> 5\sigma$, independent of the assumption on the cross section. Note that even the conservative ground-state-only analysis leads to more than 5σ significance. Let us stress that in order to calculate the significances in table 4, we assume that $\Delta\chi^2_{3\nu}$ is distributed as $\chi^2(2\,\mathrm{dof})$. For the combined analysis and the Kostensalo et al. cross section we have tested this assumption with a high-statistics MC study (10^9 pseudo experiments). We find that the $\Delta\chi^2_{3\nu}$ distribution is more similar to $\chi^2(1\,\mathrm{dof})$. This reduction of the dof is related to the presence of the physical boundary $\sin^2 2\theta \geq 0$, see e.g., ref. [59] for a discussion of this effect. The null hypothesis, when evaluated by Monte Carlo, yields

$$p_0 = 2.7 \times 10^{-8}\,(5.6\sigma)\,, \qquad \text{(gallium data, Kostensalo et al. cross section)} \qquad (3.2)$$

which is slightly higher than the corresponding value for a χ^2 with 2 dof (5.4σ, see table 4) and more similar to the value for 1 dof, which is 5.7σ.

In the following, we will adopt the Kostensalo et al. [55] cross section as our default analysis for gallium data. In figure 4 (left panel), we show the allowed regions in the sterile-neutrino parameter space for gallium data. An explanation of these measurements in terms of sterile-neutrino oscillations requires rather large mixing angles, $0.21 \leq \sin^2 2\theta \leq 0.47$ at 2σ. We observe that in this case the FC confidence regions are in excellent agreement with $\Delta\chi^2$ contours under Wilks' theorem assuming 2 dof. Many of the arguments leading to deviations from Wilks' theorem in the reactor case [24] do not apply here. In particular, the effect is basically a constant suppression of the rate, without oscillatory behaviour. Since the allowed region appears far away from the physical boundary (which dominates the distribution under the null-hypothesis, see above), we recover here the full freedom corresponding to the 2 parameters $\sin^2 2\theta$ and Δm^2.
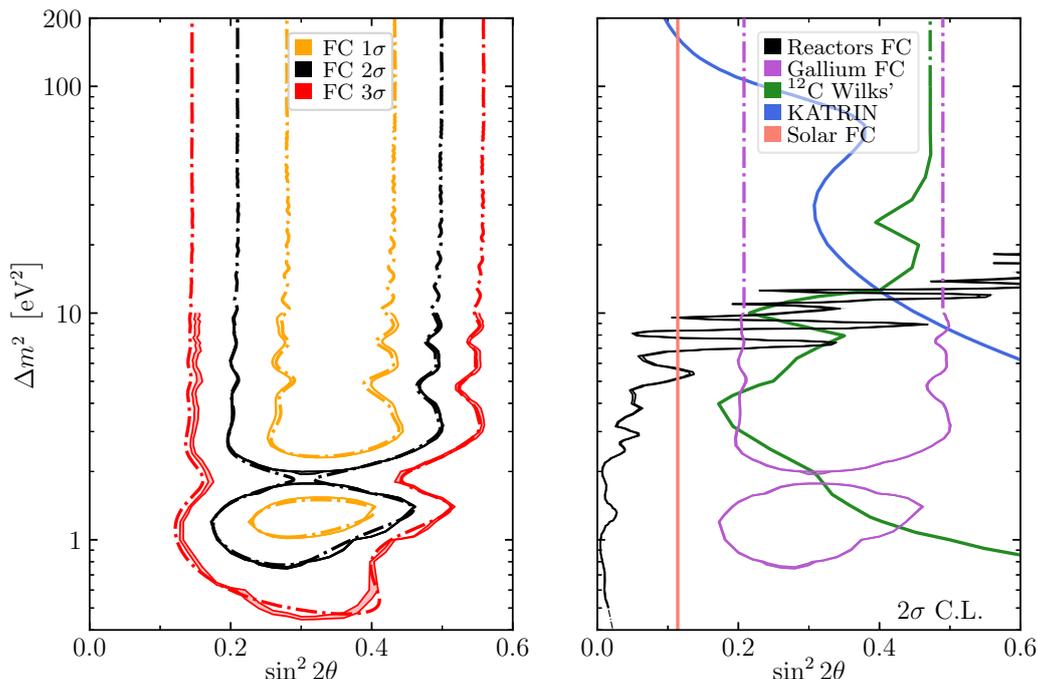
Figure 4. Left: confidence regions at 68.27%, 95.45%, 99.73% CL (1, 2, 3σ, respectively) for the Feldman-Cousins analysis (bands), where the bands indicate the 99% confidence spread of the contour for 10^5 runs due to the finite size of our MC sample (see appendix B.4). Dash-dotted curves are obtained under the assumption of Wilks' theorem (2 dof). Right: FC confidence regions for gallium, reactor, and solar data at 2σ. We superimpose the 95% exclusion limit from the KATRIN collaboration [49] (this curve fixes the lightest neutrino mass to zero) and the Wilks' 95% exclusion limit (2 dof) from ν_e–^{12}C scattering from the LSND and KARMEN experiments [43, 44], taken from ref. [56]. Dash-dotted lines are extrapolations assuming constant sensitivity in the mixing.

In the right panel of figure 4 we compare the region preferred by gallium with constraints from reactor data, solar neutrinos, the KATRIN experiment, and data on ν_e–^{12}C scattering from the LSND and KARMEN experiments. In the next section, we are going to investigate in detail the compatibility of gallium, reactor and solar neutrino data. The KATRIN constraint shown in figure 4 is taken from [49], where this curve has been derived by fixing the lightest neutrino mass to zero. This constraint cuts off the allowed parameter space at large values of $\Delta m^2 \gtrsim 100\,\mathrm{ev}^2$, but becomes quickly weaker for smaller mass-squared differences. We note that the KATRIN limit would be further relaxed if the assumption of neglecting the light neutrino masses is dropped [49, 50]. The constraint from ν_e scattering on ^{12}C from LSND [44] and KARMEN [43] is taken from the re-analysis performed in ref. [56]. This limit is in tension (at $\sim 2\sigma$ under Wilks' theorem) with the gallium region for $\Delta m^2 \lesssim 10\,\mathrm{eV}^2$. However, note that the results from LSND/KARMEN and KATRIN shown here were obtained assuming Wilks' theorem is applicable, which may lead to an overestimation of the exclusion regions, similar to the data sets studied here. Thus, due to the limited statistical power of the KATRIN and LSND/KARMEN constraints in the region of interest, we focus below on gallium, reactor, and solar data.

– 13 –

4 Global fit results and consistency tests

In this section, we combine the reactor and gallium data discussed in sections 2 and 3, respectively, and study their consistency. In addition, we take into account information from solar neutrinos, which provides an important constraint on large mixing angles. The solar-neutrino analysis adopted here is based on the simplified χ^2 construction from ref. [60], which offers an efficient way to include solar neutrino data in a MC study. We refer the interested reader to that reference for details; a brief summary is provided in appendix A.6.

Table 1 shows the value of the χ^2 per dof for the combination of reactor data, and the combination of reactor + solar, and reactor + gallium. The mild differences between the values of the $\chi^2_{\min}$ in the last three rows of the table indicate that the usual goodness-of-fit test would not provide any insights concerning the mutual consistency of the different data sets, given the very large number of dof at hand. Therefore, in order to quantify the consistency between different data sets, we use the parameter goodness-of-fit (PG) test [61, 62], based on the test statistic

$$\chi^2_{\mathrm{PG}} = \chi^2_{\mathrm{min,comb}} - \sum_k \chi^2_{\mathrm{min},k} \,, \tag{4.1}$$

where the index k labels the data sets to be tested for consistency. The first term on the right-hand side corresponds to the global χ^2 minimum, whereas the second term is the sum of the χ^2 minima of each data set individually. Under Wilks' theorem, χ^2_{PG} will follow a χ^2 distribution with $\sum_k P_k - P$ dof, where P_k is the number of parameters, on which the data set k depends, and P is the total number of parameters of the model [61, 62]. For our cases of interest, we have $P_{\mathrm{reactor}} = P_{\mathrm{gallium}} = P = 2$, corresponding to $\sin^2 2\theta$ and Δm^2, while $P_{\mathrm{solar}} = 1$ — in the limit relevant for us here, solar data are independent of Δm^2 and depend only on $\sin^2 2\theta$.

Let us first investigate the consistency of the solar and gallium data sets. Figure 5 shows the $\Delta \chi^2$ profiles for gallium data under various assumptions of the detection cross section, compared to the one from solar data. The dash-dotted curves correspond to $\Delta \chi^2_{\mathrm{gallium}} + \Delta \chi^2_{\mathrm{solar}}$; the value of χ^2_{PG} for the consistency of solar and gallium data is given by the minimum of the dashed curves. We see that the Bahcall, Kostensalo and g.s. cross sections all give very similar results for χ^2_{PG}. Even under the ground-state-only assumption, we would obtain $\chi^2_{\mathrm{PG}} = 11$, corresponding to 3.3σ tension. For the Semenov cross section, we find $\chi^2_{\mathrm{PG}} = 18$ and the tension is at the 4.2σ level. Also, note that here we use the GS98 solar model [63], which leads to conservative limits. If the AGSS09 model [63] were used instead, then the solar constraint would get significantly stronger [60], making the tension with gallium even more severe.

Next, let us study the consistency when reactor data are also considered. The results of various consistency checks are reported in table 5, where, for concreteness, we have adopted the Kostensalo et al. cross section for the gallium experiments. The last two columns show our results for the PG test, where the distribution of χ^2_{PG} is calculated using MC simulations. In doing so, the pseudo-data fluctuations are generated by using the prediction for the corresponding best-fit parameters as the pseudo-data mean. We
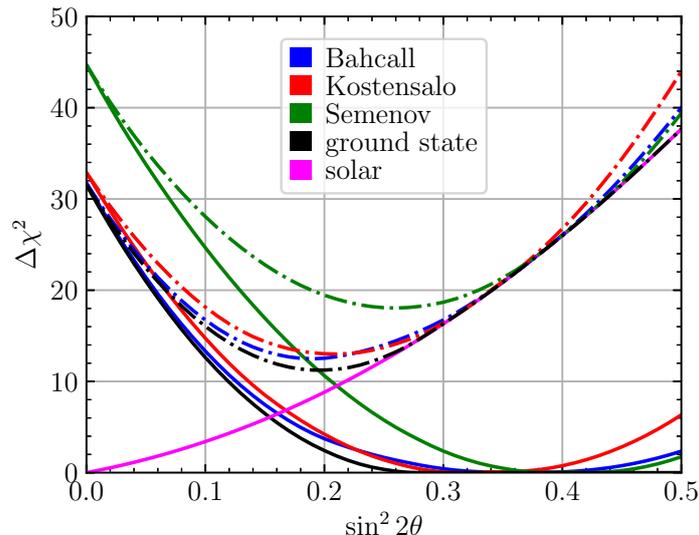
Figure 5. Solid: $\Delta\chi^2$ profiles (marginalizing Δm^2) for combined gallium data for four different assumptions on the detection cross section, see table 3. The magenta curve corresponds to solar neutrino data, and the dashed curves show the sum $\Delta\chi^2_{\text{gallium}} + \Delta\chi^2_{\text{solar}}$.

Data set	$\chi^2_{\text{PG}}/\text{dof}$	$p^{(W)}$	$\#\sigma^{(W)}$	$p_{\text{b.f.}}$	$\#\sigma_{\text{b.f.}}$
Reactor vs. Solar	0.65/1	0.42	0.8	0.39	0.9
Reactor vs. Gallium	1.4/2	0.50	0.67	0.62	0.5
Solar vs. Gallium	13.0/1	3.1×10^{-4}	3.6	1.6×10^{-3}	3.2
Reactor vs. Solar vs. Gallium	15.6/3	1.4×10^{-3}	3.2	5.1×10^{-3}	2.8

Table 5. Consistency test of the various data sets based on the parameter goodness-of-fit [61, 62]. In the middle columns, the p-values and number of Gaussian standard deviations are evaluated under the assumption of Wilks' theorem. The final two columns show results derived from MC simulations, generating pseudo-data fluctuations around the best-fit (b.f.) prediction.

find that both the reactor and solar data, as well as reactor and gallium data, are fully compatible. For comparison, the middle columns in table 5 show the expected results obtained under the assumption of Wilks' theorem. Although the differences in p-value with respect to the MC result are small in the case of consistent data sets, we find a larger discrepancy in the case of inconsistent data sets. This is shown in the last two rows of the table, where our MC simulations show a reduction of $\sim 0.5\sigma$ with respect to the expected result under the assumption that Wilks' theorem holds.

Given the strong tension between solar and gallium data, we will not combine these two data sets in the following but consider only the combinations reactor+solar and reactor+gallium separately. The best-fit points and p-values for the null hypothesis of these combinations can be found in the last rows of table 1. While the combination of solar and reactor data shows no significant hint for sterile-neutrino oscillations, the reactor+gallium combination prefers oscillations over the null hypothesis with $\Delta\chi^2_{3\nu} = 38.8$. Assuming
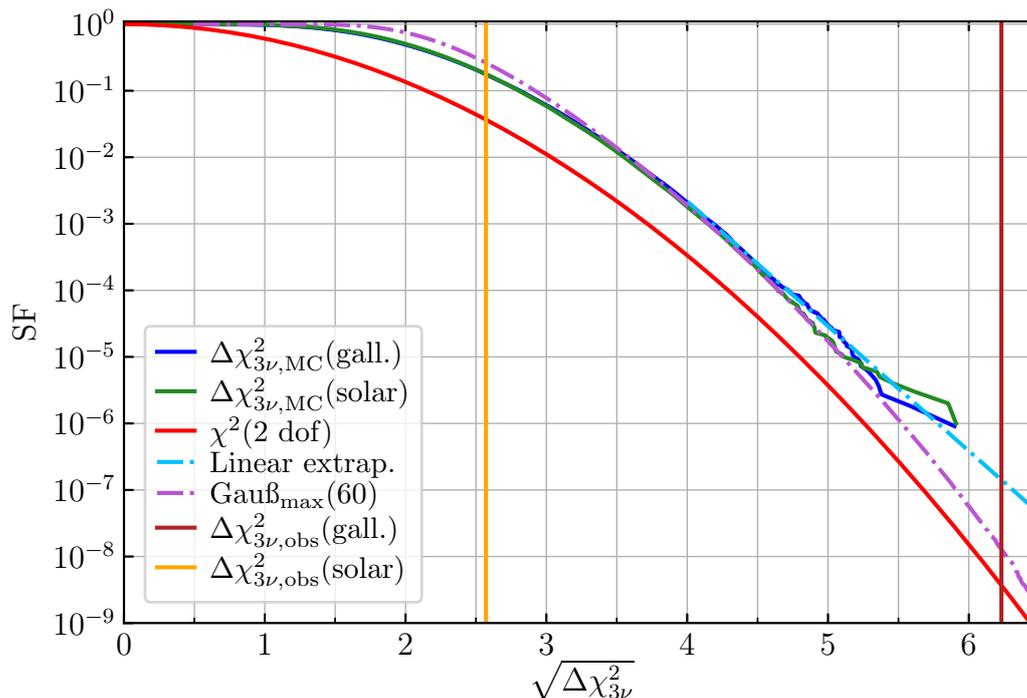
Figure 6. Survival function of $\sqrt{\Delta\chi^2_{3\nu}}$ under the null hypothesis for reactor data combined with solar (green) or with gallium (blue), for 10^6 runs. For comparison we also show the $\chi^2(2 \text{ dof})$ SF (solid red line), the maximum-Gauss distribution for $n = 60$ (dash-dotted purple), see [24], and a linear extrapolation of the tail of the gallium SF (dash-dotted light blue).

Wilks' theorem (2 dof), this would correspond to $p_0^{(W)} = 3.8 \times 10^{-9}$ (5.9σ). Such a high significance makes it impossible to calculate a p-value by MC.

In figure 6, we show the MC distribution of the test statistic for the null hypothesis, $\Delta\chi^2_{3\nu}$, defined in eq. (2.3). Both the reactor+solar and reactor+gallium datasets have an almost identical distribution which differs significantly from a χ^2 distribution, but has a similar shape as the maximum-Gauss distribution discussed in [24]. This suggests that in both cases the statistical properties are dominated by the reactor data set. We can use this observation to extrapolate the SF of $\Delta\chi^2_{3\nu}$ to the observed value for reactor+gallium. Assuming a maximum-Gauss distribution with $n = 60$, which gives a good fit to the MC distribution in the region where it can be simulated (cf., purple dash-dotted curve in figure 6), we find

$$p_0 = 1.4 \times 10^{-8} \, (5.7\sigma) \,, \qquad \text{(reactor+gallium, max.-Gauss extrapolation)} \,. \qquad (4.2)$$

If we instead use a linear extrapolation of the tail in the log-plot, (light-blue dash-dotted line in figure 6) we would obtain

$$p_0 = 1.4 \times 10^{-7} \, (5.3\sigma) \,, \qquad \text{(reactor+gallium, linear extrapolation)} \,. \qquad (4.3)$$

In any case we confirm the $\gtrsim 5\sigma$ indication in favor of sterile-neutrino oscillations of combined reactor+gallium data, which is of course driven by the BEST result. Under the adopted assumptions for the BEST analysis, this is a statistically robust result.
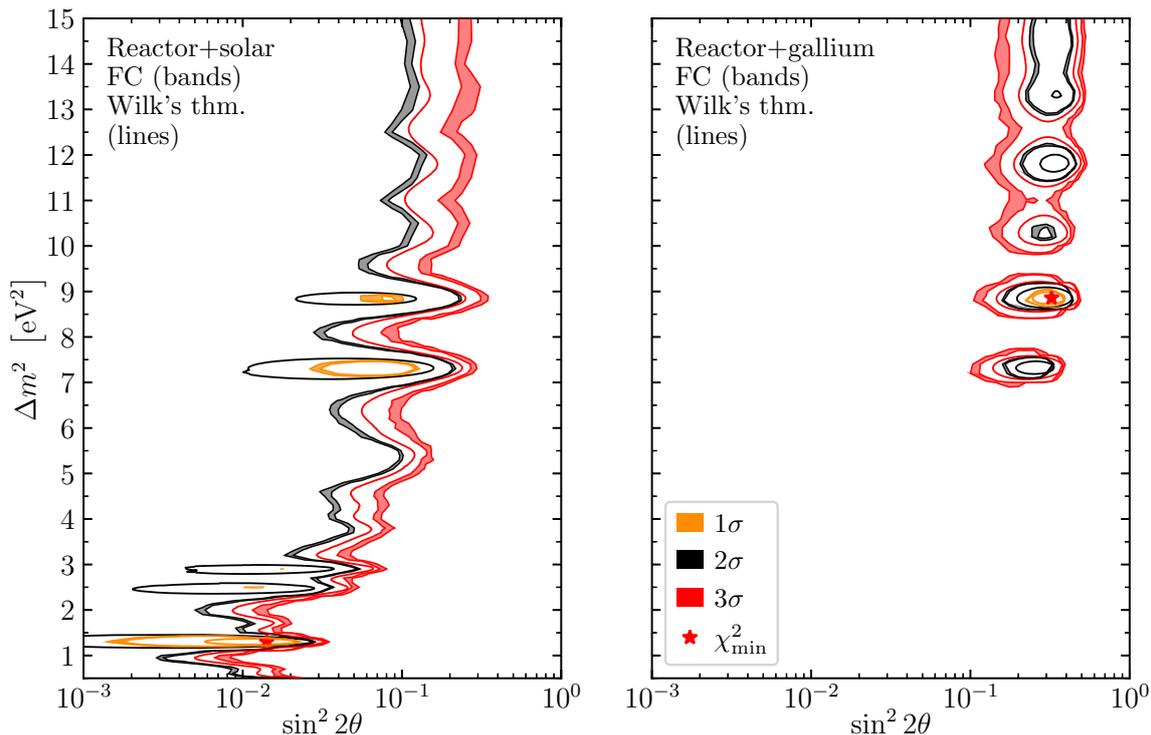
Figure 7. Confidence regions at 68.27%, 95.45%, 99.73% CL (1, 2, 3σ, respectively) for reactor+solar data (left), and reactor+gallium data (right). The bands correspond to the Feldman-Cousins analysis, where the width of the bands indicates the 99% confidence spread of the contour for 10^4 runs due to the finite size of our MC sample (see appendix B.4). Thin curves are obtained under the assumption of Wilks' theorem.

In figure 7 we show the allowed regions for reactor+solar data (left) and reactor + gallium data (right). For the reactor+solar data we find no closed contours at 2σ for the FC analysis, consistent with the null-hypothesis being allowed at 1.3σ. The main effect of solar neutrino data is to exclude the large mixing angles in the region $\Delta m^2 \gtrsim 10$. For the reactor+gallium analysis the determination of the mixing angle $0.11 < \sin^2 2\theta < 0.53$ (3σ) is dominated by gallium data, whereas Δm^2 is determined by Neutrino-4 reactor data. The best-fit point is located at

$$\sin^2 2\theta = 0.32 \,, \quad \Delta m^2 = 8.86 \quad \text{(best fit, reactor+gallium data)}\,, \tag{4.4}$$

however, several local minima are present at 2σ. In both panels we note again that the FC confidence regions are larger by roughly 1σ compared to contours based on Wilks' theorem.

5 Summary and conclusions

We have performed a global Feldman-Cousins analysis of the reactor short-baseline experiments DANSS, NEOS, PROSPECT, STEREO and Neutrino-4, past gallium data from SAGE and GALLEX and the recent result of the BEST experiment, as well as solar neutrino data. The analysis is performed in a 3+1-neutrino-oscillation framework. In general, our results show that a full Feldman-Cousins treatment reduces all *p*-values by about 1σ

as compared to a naïve application of Wilks' theorem, as shown in table 1. Our main findings are that no reactor experiment individually favors oscillation at more than $2.4\,\sigma$, with the strongest hint coming from Neutrino-4. Combined reactor data are consistent with no sterile-neutrino oscillations at $1.1\,\sigma$, and solar neutrino data are fully compatible with this result. On the other hand, gallium data favor a rate deficit at more than $5\,\sigma$, even under conservative choices for the cross-section uncertainty. This result is compatible with reactor data, and would then point to values of $\Delta m^2 \simeq 7$–$12\,\mathrm{eV}^2$, driven by the Neutrino-4 result. At the same time, the gallium and solar results are in tension, according to a parameter goodness-of-fit test [62] by more than $3\,\sigma$. One important caveat to our findings is that we take each collaboration's statements about systematic uncertainties at face value.

In comparison to earlier global fits of reactor and gallium data, for instance refs. [19, 20, 32], we find that the newest data from DANSS, NEOS, PROSPECT and STEREO are consistent with the no-oscillation hypothesis; the evidence in favor of oscillation stems entirely from the latest BEST gallium data, consistent with hints at much weaker significance from previous gallium measurements and the Neutrino-4 reactor data, which all point to larger mixing angles and mass-squared splittings. These data sets are in mutual agreement, with a parameter goodness-of-fit p-value of 0.5. The reactor experiments not seeing any evidence for oscillation are simply insensitive to the region preferred by Neutrino-4 and the gallium data. Although the Neutrino-4 results have been challenged on experimental grounds [23, 64–66], we take the results published by the collaboration as the basis for our analysis. Neutrino-4 alone prefers oscillation at $2.4\,\sigma$ (with a p-value of 1.5%), whereas the gallium data exclude the no-oscillation hypothesis at more than $5\,\sigma$. For gallium data, the main systematic is the cross section for the reaction $^{71}\mathrm{Ga} + \nu_e \to \,^{71}\mathrm{Ge} + e^-$; we have performed a literature survey of several cross-section calculations. The difficulty arises from excited-state contributions, since the ground-state-to-ground-state transition matrix element can be directly determined from the observation of the electron capture of $^{71}\mathrm{Ge}$. However, the excited-state contributions add incoherently and thus can only increase the cross section, making the effect larger. Even considering only the ground-state contribution to the cross section, we find the result is $> 5\,\sigma$ significant, see table 4. If this result is due to some systematic effect, then it seems unlikely that it is related to the cross section. Note that the results from the two detector volumes in BEST are practically identical and thus this result (as was the case for past gallium results) is an absolute rate measurement, leaving room for some common, as-yet-unidentified, systematic affecting the overall normalization of the event rates.

On the other hand, solar data strongly reject sterile-neutrino oscillations with large mixing angles for the whole range of mass-squared splittings relevant here. As a consequence, solar data are in significant tension with the gallium data (with a parameter goodness-of-fit of at least $3\,\sigma$, see table 5 and figure 5). The evidence in favor of oscillation from reactors is much weaker and hence there is no tension in a global fit between solar and reactor data. The solar bound presumably could be evaded by constructing a model in which sterile neutrinos experience some anomalous matter effect in the Sun, but this goes beyond the 3+1 model considered here.

This leaves the question of how to test our findings. In the context of 3+1 oscillations, note that the oscillation length of a 4 MeV neutrino with $\Delta m^2 = 10\,\mathrm{eV}^2$ is only about $0.5\,\mathrm{m}$, making the direct observation of oscillations in a reactor experiment very challenging. The KATRIN experiment is, in principle, sensitive to sterile neutrinos down to $\Delta m^2 \sim 10\,\mathrm{eV}^2$, but a strong exclusion limit can be only obtained if a prior is put on the effective parameter $m^2_{\nu_e,\mathrm{eff}} > 0$. This prior is well motivated by physics, but systematic effects in beta-decay endpoint experiments have consistently led to measured values of $m^2_{\nu_e,\mathrm{eff}}$ being negative (for a recent review see e.g., ref. [67]). Given the strong tension with solar data and the fact that large values of $\Delta m^2 \sim 10\,\mathrm{eV}^2$ are also in tension with standard cosmology, the notion that we are dealing with new physics different from a vanilla sterile neutrino is worthwhile to investigate, as is the search for as-yet-unidentified experimental systematics.

Acknowledgments

A Details on the experiment simulations and data analyses

In this appendix, we describe our analyses of the reactor experiments, as well as our treatment of solar neutrino data. While our analyses of reactor experiments are insensitive to the modeling of the (differential) flux $\mathrm{d}\Phi/\mathrm{d}E_\nu$ and inverse-beta-decay cross section σ_{IBD}, we must make some choices for these in order to perform our simulations. When necessary, we simulate event rates using the Huber-Mueller spectra [5, 6] and the Beacom-Vogel cross

---

[5] https://n3as.berkeley.edu.

section [68], unless noted otherwise. We reiterate that our results are insensitive to these choices.

A.1 DANSS

The flux of neutrinos is determined by using the fuel fission fractions provided by DANSS [14, p. 5]. We assume the reconstructed energy E_{rec} is distributed as a Gaussian with mean E_ν and width

$$\frac{\sigma(E_\nu)}{E_\nu} = 0.352 + 0.024 \cdot (E_\nu/\mathrm{MeV}) - 0.148\sqrt{E_\nu/\mathrm{MeV}} + \frac{0.040}{\sqrt{E_\nu/\mathrm{MeV}}}, \qquad \text{(A.1)}$$

encoded by the reconstruction matrix $R(E_{\mathrm{rec}}, E_\nu)$. The width is taken from ref. [69, figure 5] but increased by 10% to better match the results of the collaboration.

Because the baseline is so short, the finite size of the reactor core and detector needs to be taken into account. The detector is assumed to be a cube of $1\,\mathrm{m}^3$ volume [70] and the reactor size is taken to be $4.5\,\mathrm{m}$. This number is set to be slightly higher than the value quoted in ref. [14] in order to obtain a better match to the results of the collaboration, and we assume that this is due to diagonally travelling neutrinos having a longer path length. Thus, for each baseline $I \equiv B, M, T$ (bottom, middle, top) the predicted number of events is computed as

$$N_I^i(\Delta m^2) = N_{0,I}^i - \sin^2 2\theta \int_{L_I-0.5}^{L_I+0.5} \mathrm{d}x \int_{x-2.25}^{x+2.25} \mathrm{d}L \int_{E_i}^{E_{i+1}} \mathrm{d}E_{\mathrm{rec}} \int_0^\infty \mathrm{d}E_\nu$$
$$\left[\sigma_{\mathrm{IBD}}(E_\nu) \frac{d\Phi}{dE_\nu}(E_\nu) R(E_{\mathrm{rec}}, E_\nu) \frac{\sin^2 \frac{\Delta m^2 L}{4E_\nu}}{L^2} \right]. \qquad \text{(A.2)}$$

The DANSS analysis is performed in terms of the ratios between the number of events at the different locations as outlined in section 2.1, see eq. (2.1). The predicted values for the ratios can be written in terms of the event rates in eq. (A.2) as

$$P_1^i = R_1^{3\nu} \frac{N_B^i}{N_T^i}, \quad P_2^i = R_2^{3\nu} \frac{N_M^i}{\sqrt{N_B^i N_T^i}}, \qquad \text{(A.3)}$$

where the no-oscillation predicted ratios $R_1^{3\nu}, R_2^{3\nu}$ can be found in ref. [14, p. 17].

A 2×2 covariance matrix (with indices $a, b \in [1, 2]$) is defined for each energy bin i:

$$V_{ab}^i = \sum_{I=B,M,T} \frac{\partial R_a^i}{\partial N_I^i} \frac{\partial R_b^i}{\partial N_I^i} \sigma_{N_I^i}^2 = \sum_{I=B,M,T} \frac{\partial R_a^i}{\partial n_I^i} \frac{\partial R_b^i}{\partial n_I^i} \frac{n_I^i}{\Delta t_I} \qquad \text{(A.4)}$$

where we have assumed that the uncertainty of the event rates in each bin is statistics dominated ($\sigma_{N_I^i} = \sqrt{N_I^i}$). This describes correlations due to the fact that $P_{1,2}^i$ contain common factors of $N_{B,T}^i$.

The final ingredient in the computation of the χ^2 function for DANSS is the effect of systematic uncertainties. We introduce two nuisance parameters $\kappa_{1,2}$ for each of the ratios

measured, associated with the relative efficiency, and add a pull term for each of them. The χ^2 function for DANSS is then defined as

$$\chi^2_{\text{DANSS}} = \min_{\kappa_1, \kappa_2} \left\{ \sum_{i=1}^{36} \sum_{a,b=1}^{2} [D_a^i - (1+\kappa_a)P_a^i](V^i)_{ab}^{-1}[D_b^i - (1+\kappa_b)P_b^i] + \frac{\kappa_1^2 + \kappa_2^2}{\sigma_{\text{sys}}^2} \right\}, \quad \text{(A.5)}$$

where $\sigma_{\text{sys}} = 0.2\%$, and D_a^i corresponds to the ratios measured by the experiment. The minimization is performed analytically, noting that the problem is to minimize a degree-two multinomial in the nuisance parameters.

By following this procedure, we find good agreement between our χ^2 map and the result reported in ref. [14, p. 16] by the collaboration. A comparison of the best-fit points is given in table 1.

To generate the pseudo data, we first compute the mean value for each bin as

$$\left\langle n_T^i \right\rangle = n_T^{i,\text{obs}}, \quad \left\langle n_M^i \right\rangle = n_T^{i,\text{obs}} P_2^i \sqrt{P_1^i}, \quad \left\langle n_B^i \right\rangle = n_T^{i,\text{obs}} P_1^i. \quad \text{(A.6)}$$

Statistical fluctuations are then injected by randomly sampling a Gaussian distribution for each bin, centered around its mean value and with standard deviation $\sigma_{I,i} = \sqrt{\langle n_I^i \rangle / \Delta t_I}$.

A.2 NEOS

The NEOS analysis uses the neutrino spectrum reported from the Daya Bay collaboration, obtained from ref. [54]. Due to this complication, and their unknown error response, we obtain our prediction by applying a ratio to the collaboration's no-oscillation prediction.

Let us first denote the collaboration's prediction in the absence of oscillations by $S_{i,0}$. We obtain $S_{i,0}$ by dividing the ratios from ref. [15, figure 3c] through the extracted signal points (corresponding to the observed data minus the expected background) from ref. [15, figure 3a]. Doing this normalizes the sum $\sum_i S_{i,0} = 1891.6$.

To obtain the prediction S_i for the sterile-neutrino hypothesis, we then rescale[6] the no-oscillation prediction by the ratio

$$S_i = S_{i,0} \frac{\text{DB}_i(\sin^2 2\theta, \Delta m^2)}{\text{DB}_i(\sin^2 2\theta = 0)}, \quad \text{(A.7)}$$

where DB_i is a function defined for each bin i in reconstructed energy as

$$\text{DB}_i(\sin^2 2\theta, \Delta m^2) = \left(1 - \frac{\sin^2 2\theta}{2}\right)^{-1} \times$$

$$\int dE_\nu \mathcal{R}_i(E_\nu) \times \left[\sigma_{\text{IBD}} \frac{d\Phi}{dE_\nu}\right](E_\nu) \times \left\langle \frac{P_{ee}}{L^2} \right\rangle (\sin^2 2\theta, \Delta m^2, E_\nu), \quad \text{(A.8)}$$

and

$$\mathcal{R}_i(E_\nu) = \int_{E_i}^{E_{i+1}} R(E_{\text{rec}}, E_\nu) dE_{\text{rec}}. \quad \text{(A.9)}$$

---

[6]This rescaling is only approximate, due to the incomplete information on the detector response. However it should be accurate as long as the effects due to our mismodeling of the detector response are comparable for the numerator and the denominator.

Here, $R(E_{\rm rec}, E_\nu)$ stands for the reconstruction matrix (see below for an exact definition), while $[\sigma_{\rm IBD} {\rm d}\Phi/{\rm d}E_\nu]\,(E_\nu)$ is the flux-weighted cross section in terms of the true neutrino energy, taken from ref. [54]. In eq. (A.8), we have also defined a weighted oscillation probability with the inverse baseline squared as

$$\left\langle \frac{P_{ee}}{L^2} \right\rangle (\sin^2 2\theta, \Delta m^2, E_\nu) = \int_{L_{\min}}^{L_{\max}} P_{ee}\left(\sin^2 2\theta, \frac{\Delta m^2 L}{E_\nu}\right) \frac{\rho(L)}{L^2} {\rm d}L \,, \qquad (A.10)$$

where P_{ee} is the 3+1-oscillation probability at NEOS; $L_{\min, \max}$ are the minimum, maximum baselines respectively; and $\rho(L)$ is the baseline distribution (extracted from ref. [71, p. 8]).

Note that the Daya Bay spectrum from ref. [54] unfolds the three-neutrino oscillations, and, since Daya Bay would be sensitive to both 3ν and 4ν oscillations, we need to remove distortions that would be introduced under a different hypothesis. For the large values of Δm^2 considered in this work, we assume that the oscillations due to the new frequency, introduced by the sterile neutrino, average out at Daya Bay, which leads to the factor $[1 - \sin^2 2\theta/2]^{-1}$ in eq. (A.8). Corrections from different isotope fractions are applied as per ref. [15].

The reconstruction matrix $R(E_{\rm rec}, E_\nu)$ has two components. For the first, we take the approximation from ref. [72] for positron energy loss,

$$R_{e^+}(E_{\rm rec}) = \begin{cases} Z + \left(\frac{1}{\sqrt{2\pi}\sigma(E_{\rm avg})} - Z\right) \exp\left\{-\frac{(E_{\rm rec} - E_{\rm avg})^2}{2\sigma^2(E_{\rm avg})}\right\} & {\rm for}\ \ E_{\rm rec} < E_{\rm avg} \,, \\ \frac{1}{\sqrt{2\pi}\sigma(E_{\rm avg})} \exp\left\{-\frac{(E_{\rm rec} - E_{\rm avg})^2}{2\sigma^2(E_{\rm avg})}\right\} & {\rm for}\ \ E_{\rm rec} > E_{\rm avg} \,, \end{cases} \qquad (A.11)$$

where $Z = 0.01$. Here, the energy resolution is parametrized as [73, figure 3]

$$\sigma(E) = \left[2 \cdot 10^{-2} + 1.2 \cdot 10^{-3} \frac{E}{\rm MeV} + 0.2 \cdot 10^{-3} \left(\frac{E}{\rm MeV}\right)^2\right]^{1/2} \,, \qquad (A.12)$$

and the average energy $E_{\rm avg}$ includes the non-linear detector response,

$$E_{\rm avg} = 1.01 \cdot E_p f(0.41 \cdot E_p - 0.05), \qquad (A.13)$$

where $f(\,\cdot\,)$ is taken from the supplementary material of ref. [74] and $E_p = E_\nu - \Delta_{np} + m_e$, with m_e being the electron mass and Δ_{np} the neutron-proton mass difference. The choice of parametrization for the function f was simply made by tuning $AE_p f(aE_p + b)$ to the results of the collaboration, as no detector response was provided. Besides eq. (A.11) we also consider a second contribution from escaping $511\,{\rm keV}$ annihilation photons, with a relative height 0.5 as per ref. [71, p. 15]:

$$R_\gamma(E_{\rm rec}) = 0.5 \times \frac{1}{\sqrt{2\pi}\sigma(E_{\rm avg})} \exp\left\{-\frac{(E_{\rm rec} + m_e - E_{\rm avg})^2}{2\sigma^2(E_{\rm avg})}\right\} \,. \qquad (A.14)$$

The final reconstruction matrix $R(E_{\rm rec}, E_\nu)$ is computed by adding the contributions from eqs. (A.11) and (A.14). However, for the last bin we find that this approach does not provide good agreement with ref. [15]. Thus, the entry in this last bin is estimated as

a piecewise constant (see the inset of [15, figure 3a]) so that the data/Daya Bay ratio is correctly reproduced. Finally, once the reconstruction matrix has been computed, the integral over reconstructed energy in eq. (A.9) is done analytically.

Following ref. [15], we define the NEOS covariance matrix as $V_{ij} = V_{ij}^{\text{syst}} + V_{ij}^{\text{stat}}$. The matrix V^{stat} accounts for the statistical errors and is defined as

$$V_{ij}^{\text{stat}} = \delta_{ij} \left[\frac{S_i^{\text{obs}} + B_i}{\Delta t_{\text{on}}} + \frac{B_i}{\Delta t_{\text{off}}} \right] , \tag{A.15}$$

where B_i corresponds to the background events and S_i^{obs} corresponds to the observed background-subtracted events per bin i (that is, the black points in [15, figure 3a]), while Δt_{on} and Δt_{off} correspond to the data-taking periods in days, with the reactor on and off, respectively. In eq. (A.15) the first term is the Poisson error, while the second one stems from the fact that the background rate was measured during the (relatively short) reactor-off time period and therefore is not known precisely. On the other hand, the contribution V^{sys} is defined as

$$V_{ij}^{\text{syst}} = \mathcal{N}^2 \iint \mathrm{d}E_\nu \mathrm{d}E_\nu' \, V^{\text{DB}}(E_\nu, E_\nu') \mathcal{R}_i(E_\nu) \mathcal{R}_j(E_\nu') \tag{A.16}$$

where V^{DB} is constructed through a bicubic interpolation of the covariance matrix in ref. [54, table 13], and the normalization constant $\mathcal{N}$ is determined as

$$\mathcal{N} = \frac{\sum_i S_{i,0}}{\sum_i \mathrm{DB}_i(0)} \int_{L_{\min}}^{L_{\max}} \frac{\rho(L)}{L^2} \mathrm{d}L \, . \tag{A.17}$$

As for systematic uncertainties, we include a nuisance parameter for the overall flux normalization, ξ, but we add no penalty term to the χ^2: it is left free in the fit. Thus, the NEOS χ^2 function reads:

$$\chi_{\text{NEOS}}^2 = \min_\xi \sum_{i,j} (S_i^{\text{obs}} - \xi S_i) V_{ij}^{-1} (S_j^{\text{obs}} - \xi S_j), \tag{A.18}$$

where the minimization over ξ is done analytically.

We compared our Bayesian exclusion curve (computed as in ref. [15, eq. (4)]) to the result of the collaboration; in this method, the unnormalized probability distribution for a fixed Δm^2 is $\exp(-\Delta\chi^2/2)$. Our curve has extremely good agreement in the range 1–2 eV^2.

Due to our incomplete knowledge of the detector response, our fit does not perfectly reproduce the results from the collaboration: in particular, we get more local minima in our fit, and the location of our global minimum differs from the one obtained by the collaboration in table 1. We assume this to be due to the unknown detector response, combined with the impact of systematic uncertainties (e.g., energy scale) which are technically difficult to implement outside the collaboration. However, we have checked that the differences in χ^2 are relatively small. For example, the collaboration obtains two degenerate minima at $(\Delta m^2, \sin^2 2\theta) = (0.05, 1.73 \, \text{eV}^2)$ and $(0.04, 1.30 \, \text{eV}^2)$. For these two points of parameter space, our fit yields a higher χ^2 by approximately two units, see table 6.

NEOS	Our Fit χ^2	Collab. χ^2	Δm^2	$\sin^2 2\theta$
Minimum 1	59.73	~ 57.5	1.30 eV^2	0.04
Minimum 2	59.60	57.5	1.73 eV^2	0.05

Table 6. Comparison of our χ^2 to the values obtained by the collaboration at their quasi-degenerate global minima.

A.3 PROSPECT

We approximate the detector response from the extensive public data release accompanying ref. [16], which contains efficiencies (ϵ_s) and response matrices (R_{ij}^s) for each segment s of the detector. For our analysis, we group the segments into baseline bins, which we index with l, according to the segment map given by the collaboration. The fit is performed using the GLoBES [75, 76] libraries. For each bin l, the detector mass is set to the sum of the relative efficiencies of the segments contained in that bin. Moreover, its effective baseline is defined by adding the contribution of each segment, weighted by its relative efficiency $w_s = \frac{\epsilon_s}{\sum_s \epsilon_s}$ as

$$\left\langle \frac{1}{L^2} \right\rangle_l = \sum_s w_s \left\langle \frac{1}{L^2} \right\rangle_s , \qquad \text{with} \qquad \left\langle \frac{1}{L^2} \right\rangle_s = \int dL \, \frac{\rho_s(L)}{L^2} . \qquad (A.19)$$

Here, the baseline distribution $\rho_s(L)$ for each segment s is constructed by randomly selecting points in the reactor core and detector segment to form a density histogram (the segment map provided by the collaboration determines the detector geometry, while the reactor core is modeled as a cylinder of height 51 cm and radius 22 cm).

The four-neutrino oscillation probability for a given bin l and neutrino energy E_ν is computed as

$$P_{ee}^l \left(\sin^2 \theta, \frac{\Delta m^2}{4E_\nu} \right) = 1 - \sin^2 2\theta \cdot \Pi_l \left(\frac{\Delta m^2}{4E_\nu} \right) , \qquad (A.20)$$

where for each baseline bin l we define

$$\Pi_l(q) = \left[\sum_s w_s \left\langle \frac{\sin^2(qL)}{L^2} \right\rangle_s \right] \Big/ \left\langle \frac{1}{L^2} \right\rangle_l \qquad (A.21)$$

with

$$\left\langle \frac{\sin^2(qL)}{L^2} \right\rangle_s = \int dL \, \frac{\rho_s(L) \sin^2(qL)}{L^2} . \qquad (A.22)$$

At a given baseline bin l and for an energy bin i, the predicted event rates for a set of oscillation parameters are proportional to

$$N_{(il)}(\sin^2 2\theta, \Delta m^2) = \sum_j R_{ij}^l \int_{E_j}^{E_{j+1}} dE_\nu \left[\sigma_{\text{IBD}} \frac{d\Phi}{dE_\nu} \right] (E_\nu) \times P_{ee}^l \left(\sin^2 2\theta, \frac{\Delta m^2}{4E_\nu} \right) \quad (A.23)$$

where $\mathrm{d}\Phi/\mathrm{d}E_\nu$ is the predicted neutrino flux and σ is the cross section, while R^l_{ij} stands for the effective energy response matrix for bin l, given by

$$R^l_{ij} = \sum_s W_s R^s_{ij}, \qquad \text{where} \qquad W_s = \frac{w_s \langle L^{-2} \rangle_s}{\langle L^{-2} \rangle_l}. \tag{A.24}$$

We have grouped the energy index i and the baseline index l into a combined index (il) as a bookkeeping device. The weights W_s account for the relative expected event rates in the segments. In principle, these vary as a function of the sterile-neutrino parameters, but we assume that these differences are negligible.

The result from eq. (A.25) is then used to rescale the predicted signal event rates by the collaboration in the absence of oscillations ($S^0_{(il)}$, provided in the data release), as

$$S_{(il)} = S^0_{(il)} \frac{N_{(il)}(\sin^2 2\theta, \Delta m^2)}{N_{(il)}(\sin^2 2\theta = 0)}. \tag{A.25}$$

Once the event rates have been computed, the χ^2 is calculated as in ref. [16, eq. (11)]:

$$\chi^2_{\text{PROSPECT}} = \sum_{l,m} \sum_{i,j} \left(O_{(il)} - S_{(il)} \frac{O_i}{S_i} \right) V^{-1}_{(il)(jm)} \left(O_{(jm)} - S_{(jm)} \frac{O_j}{S_j} \right), \tag{A.26}$$

where V is the covariance matrix provided by the collaboration,[7] $O_{(il)}$ is the observed number of events in energy bin i for baseline bin l, and we have defined

$$O_i = \sum_l O_{(il)} \qquad \text{and} \qquad S_i = \sum_l S_{(il)}. \tag{A.27}$$

In figure 8 (left), we show our Feldman-Cousins allowed regions with statistical fluctuations generated as explained in appendix B; we find very good agreement with the results reported by the collaboration.

A.4 STEREO

For each of the six cells in the STEREO detector, we calculate the effective baseline and oscillation probability in a similar fashion to PROSPECT (described in appendix A.3), according to the geometry described in ref. [17]. We index these cells with $l = 1, \ldots, 6$. In principle, light produced in a given cell can leak into its neighbors, potentially converting a single event of a given energy into multiple events with lesser energy in multiple segments. However, we assume that this cross-talk is negligible in our analysis.

The STEREO data set is divided into two phases; we index these phases with $\lambda = \text{I, II}$. Here we assume that all cells have the same energy resolution and that the cell masses are identical for both phases. The energy-resolution function can be read from ref. [77, figure 2.19] for phase I, and from ref. [17, figure 11] for phase II. It is parametrized as

$$\frac{\sigma_\lambda(E)}{E} = a_\lambda + \frac{b_\lambda}{E/\text{MeV}}, \quad \text{where} \quad \begin{aligned} a_I &= 0.031, & b_I &= 0.059; \\ a_{II} &= 0.043, & b_{II} &= 0.050. \end{aligned} \tag{A.28}$$

---

[7]Our notation reflects the intuition that V is naturally expressed a 160×160 matrix, rather than as a $10 \times 16 \times 10 \times 16$ tensor, by flattening the rates into a 160-element vector. In particular, the former lends itself to inversion more readily than the latter.
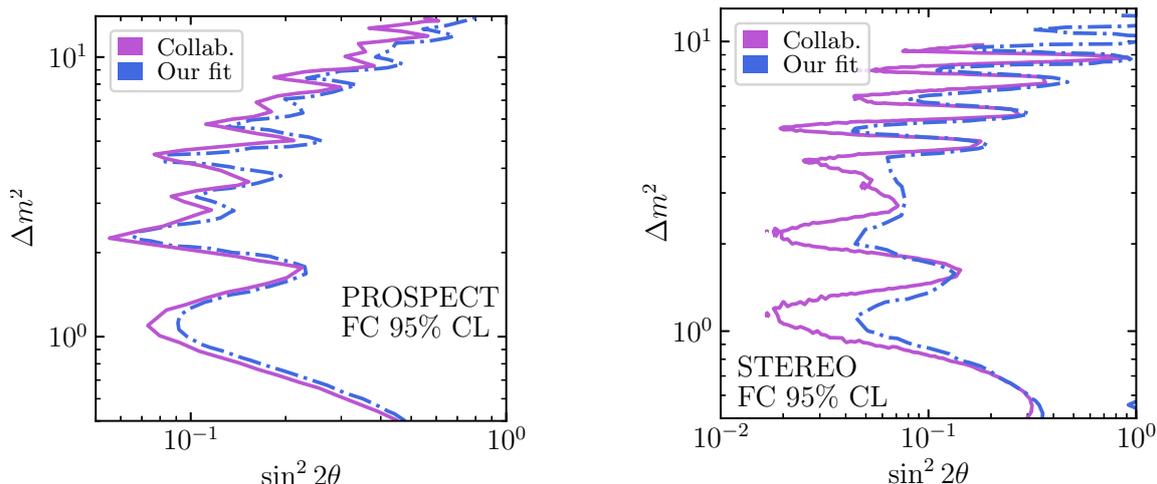
Figure 8. Comparison of our 95% FC curve for PROSPECT (left) and STEREO (right), both evaluated with 10^4 pseudo experiments, to the result of the collaboration [16, 80].

l	Phase I			Phase II		
	c_l^I [%]	d_l^I [%]	e_l^I [%]	c_l^{II} [%]	d_l^{II} [%]	e_l^{II} [%]
1	2.2	-0.97	0.092	-0.23	0.25	-0.029
2	1.7	-0.98	0.10	-0.054	0.043	-0.010
3	1.1	-0.69	0.074	-0.11	0.22	-0.027
4	1.65	-0.96	0.01	0.15	-0.076	0.0026
5	2.0	-0.99	0.10	-0.10	0.095	-0.020
6	1.2	-0.70	0.070	0.35	-0.38	0.022

Table 7. STEREO non-linearity coefficients (eq. (A.29)) for each cell l, as obtained from a quadratic fit to the points shown in ref. [77, figure 2.19] for phase I, and from ref. [17, figure 11] for phase II.

The detector non-linearities can also be found in the same references. The extracted data for the non-linear detector responses are fit to quadratic functions,

$$f_l^\lambda(E) = c_l^\lambda + d_l^\lambda \frac{E}{\text{MeV}} + e_l^\lambda \left(\frac{E}{\text{MeV}}\right)^2 , \tag{A.29}$$

and are applied to the reconstructed energies as $E_{\text{rec}} \mapsto E_{\text{rec}}[1 + f_l^\lambda(E_{\text{rec}})]$. The values obtained for the coefficients for each cell are provided in table 7 for convenience.

The energy resolution and non-linear response functions are then combined to build a Gaussian response matrix $R_l^\lambda(E_{\text{rec}}^i, E_\nu^j)$, used to map events in *true* antineutrino energy E_ν (distributed across 70 bins between 1.85 and 7.95 MeV) onto the *reconstructed* energy E_{rec} used in the analysis, listed in section 2.1. The corresponding event rates are

$$N_{l,i}^\lambda(\sin^2 2\theta, \Delta m^2) = \sum_j R_l^\lambda(E_{\text{rec}}^i, E_\nu^j) \times$$

$$\int_{E_\nu^j}^{E_\nu^{j+1}} dE_\nu \left[\sigma_{\text{IBD}} \frac{d\Phi}{dE_\nu}\right](E_\nu) \times P_{ee}^l\left(E_\nu; \sin^2 2\theta, \Delta m^2\right). \tag{A.30}$$

As mentioned in the main text, the STEREO collaboration has provided their data normalized with respect to their best-fit no-oscillation hypothesis for each phase. As such, in order to compare our predictions with their data, we must normalize our predictions in the same way. To wit, for each phase λ, the prediction we employ for energy bin i and cell l can be written as [17]:

$$S_{l,i}^{\lambda} = \frac{\left(1 + \zeta_l^{\lambda}\right) \times N_{l,i}^{\lambda} \left(\eta_l^{\lambda} + \xi^{\lambda}\right)}{\left(1 + \bar{\zeta}_l^{\lambda}\right) \times N_{l,i}^{0,\lambda}(\bar{\eta}_l^{\lambda} + \bar{\xi}_l^{\lambda})}, \tag{A.31}$$

where η_l^{λ} is a pull introduced for the cell-uncorrelated energy-scale uncertainty, ξ^{λ} is its cell-correlated counterpart, and ζ_l^{λ} is a pull for the signal-normalization uncertainty. Here, $N^{0,\lambda}$ is the prediction for the no-oscillation hypothesis for the fixed (optimized) pulls $\bar{\eta}_l^{\lambda}, \bar{\xi}_l^{\lambda}, \bar{\zeta}_l^{\lambda}$ provided in ref. [17, figure 31].

The χ^2 function for STEREO is computed by adding the contributions of data from the two phases. Following ref. [17, eqs. (15, 19)], we include a set of free flux normalizations for each bin i (ϕ_i, common to the two phases)[8] as well as a relative normalization uncertainty between the two phases (Φ^{I}, common to all bins). The final χ^2 function is obtained after minimization over all nuisance parameters:

$$\chi_{\mathrm{STEREO}}^2 = \min_{\substack{\eta_l^{I,II};\, \xi^{I,II} \\ \zeta_l^{I,II};\, \phi_i;\, \Phi}} \left\{ \sum_{l=1}^{N_{\mathrm{cells}}} \sum_{i=1}^{N_{\mathrm{Ebins}}^I} \left(\frac{O_{l,i}^I - \Phi^I \phi_i S_{l,i}^I (\eta_l^I + \xi^I, \zeta_l^I)}{\sigma_{l,i}^I} \right)^2 \right.$$

$$+ \sum_{l=1}^{N_{\mathrm{cells}}} \sum_{i=1}^{N_{\mathrm{Ebins}}^{II}} \left(\frac{O_{l,i}^{II} - \phi_i S_{l,i}^{II}(\eta_l^{II} + \xi^{II}, \zeta_l^{II})}{\sigma_{l,i}^{II}} \right)^2 \tag{A.32}$$

$$\left. + \sum_{\substack{l=1 \\ \lambda=I,II}}^{N_{\mathrm{cells}}} \left(\frac{\eta_l^{\lambda}}{\sigma_{\eta}^{\lambda}} \right)^2 + \left(\frac{\zeta_l^{\lambda}}{\sigma_{\zeta}^{\lambda}} \right)^2 + \sum_{\lambda=I,II} \left(\frac{\xi^{\lambda}}{\sigma_{\xi}^{\lambda}} \right)^2 \right\}.$$

Here, $O_{l,i}^{\lambda}$ are the observed data points in baseline bin l and energy bin i [17, 78] while $\sigma_{l,i}^{\lambda}$ are the statistical errors on the data. The uncertainties of the nuisance parameters are:

$$\begin{aligned} \sigma_{\eta}^I = 1.06\%, \quad & \sigma_{\xi}^I = 0.35\%, \quad & \sigma_{\zeta}^I = 1.18\%, \\ \sigma_{\eta}^{II} = 1.02\%, \quad & \sigma_{\xi}^{II} = 0.30\%, \quad & \sigma_{\zeta}^{II} = 1.18\%, \end{aligned} \tag{A.33}$$

while both ϕ_i and Φ^I are left completely *unconstrained* in the fit.

Finally, we make the transformations

$$\alpha_l^{\lambda} = \eta_l^{\lambda} + \xi^{\lambda}, \qquad\qquad \Phi^I(1 + \zeta_l^I) = 1 + \beta_l^I, \tag{A.34}$$

and for consistency $\beta_l^{II} = \zeta_l^{II}$. This isolates the dependence of ξ^{λ}, Φ^I in the quadratic pull terms, which we can analytically minimize. We numerically minimize the remaining pulls with a non-linear conjugate-gradient algorithm [79].

---

[8]Note that while the ten lowest-energy bins are identical for the two phases, there is an additional high-energy bin for the phase-II data, (i.e., $N_{\mathrm{Ebins}}^{II} = N_{\mathrm{Ebins}}^{I} + 1$). As such, the phase II analysis depends on an additional flux nuisance parameter ϕ to which phase I is insensitive.

Using our χ^2 implementation, we are able to reproduce the results of the collaboration to a reasonable degree of accuracy, both for the $\Delta\chi^2$ as well as for the 95% CL$_s$,[9] curves published in their data release [80]. A comparison of our FC 95% CL curve is shown in figure 8 (right). While the shape of our exclusion curve is similar to the result of the collaboration, we do find a general reduction in our calculated exclusion. We think, however, that the level of agreement achieved is reasonable, given that:

1. the χ^2 is quite flat in the low-Δm^2 region. Consequently, small numerical differences — in either the low-Δm^2 region or around the global minimum — can lead to sizeable shifts in the locations of contours of constant $\Delta\chi^2$.

2. Our minimization procedure has finite resolution. If the global minimum is missed during the minimization procedure, then this would manifest as an apparent reduction in sensitivity over the entire parameter space. Increasing the resolution beyond the current level would make our simulation computationally too expensive.

3. Our detector response is only approximate. We use cell- and phase-specific descriptions of the non-linearity, but assume a common resolution for all cells for a given phase. This may affect the high-Δm^2 region in particular.

Finally note that, when generating pseudo data for STEREO, for some of the bins in the high-energy tail, we sometimes obtain negative fluctuations. Whenever that happens, we simply throw these away and redraw fluctuations for all bins.

A.5 Neutrino-4

Our analysis of Neutrino-4 is based on ref. [24]. As outlined in section 2.1, the event rates are binned in L/E and grouped into 27 bins. In our analysis, we use the first 19 bins of such a data set, shown as blue triangles in figure 47 in v2 of the preprint of ref. [18].

The collaboration performed their analysis in terms of the ratio of events in each bin to the total rates in all bins, as in eq. (2.2). The predicted event rates in each bin (in the absence of statistical fluctuations) can be computed as[10]

$$R_i^{\mathrm{pred}} = \frac{1 - \sin^2 2\theta \left\langle \sin^2 \frac{\Delta m^2 L}{4E} \right\rangle_i}{1 - \sin^2 2\theta \frac{1}{N} \sum_{i=1}^{N} \left\langle \sin^2 \frac{\Delta m^2 L}{4E} \right\rangle_i} \,, \tag{A.35}$$

where the average over each bin i is indicated by $\langle \cdot \rangle_i$ and takes into account both energy and baseline uncertainties.

The fit to Neutrino-4 data is performed with a simple Gaussian χ^2 definition:

$$\chi^2_{\mathrm{Neutrino\text{-}4}} = \sum_{i=1}^{19} \frac{(R_i^{\mathrm{obs}} - R_i^{\mathrm{pred}})^2}{\sigma_i^2} \,, \tag{A.36}$$

---

[9]See [81] for an explanation of the CL$_s$ procedure.

[10]Note the slightly different definition in the denominator with respect to ref. [24]. While the expression used in ref. [24] is theoretically correct, in practice the average over different bins is not perfect and can lead to small differences in the region of small Δm^2. Thus, here we adopt the exact expression instead.

where σ_i corresponds to the statistical uncertainty read off figure 47 of ref. [18], while R_i^{obs} and R_i^{pred} refer to the observed and predicted ratios in each bin.

Once the theoretical prediction has been computed for each bin, we generate statistical fluctuations in each bin according to a Gaussian centered at the theoretical prediction, $\mathcal{N}(R_i^{\text{pred}}, \sigma_i)$, where σ_i is the uncertainty in each bin extracted from the same figure[11] in ref. [18].

Let us mention that Neutrino-4's treatment of the detector energy resolution has been criticized in refs. [64–66]. In particular, a proper treatment of the energy resolution may reduce the reported signal significance by approximately 0.5σ [66], and require even larger values for the mixing angle. Additionally, the authors of ref. [23] have claimed that unaccounted correlations between energy and efficiency may induce an oscillation-like signature at Neutrino-4. In our analysis we do take into account finite energy (as well as spatial) resolution, see eq. (A.35). In choosing the value for the resolution we follow our general strategy, to reproduce as close as possible the official oscillation results published by the collaboration. Specifically, the average in each bin is performed using an effective energy uncertainty $\Delta E_{\text{eff}} = E\sqrt{(\Delta L/L)^2 + (\Delta E/E)^2}$, where $\Delta E = 500\,\text{keV}$ and $\Delta L = 0.48\,\text{m}$.

A.6 Solar neutrinos

The solar-neutrino analysis performed here is based on the simplified χ^2 construction from ref. [60], which in turn is based on the global fit from ref. [82]. While we refer the interested reader to the aforementioned references for details, here we provide a brief explanation of the solar analysis for completeness.

Our fit considers four data points as in ref. [60], corresponding to the four oscillation probabilities:

$$r = (P_{ee}^{\text{LE}}, P_{ee}^{\text{HE}}, P_{ex}^{\text{LE}}, P_{ex}^{\text{HE}}). \tag{A.37}$$

Here, P_{ee} is the electron-neutrino survival probability and $P_{ex} = P_{e\mu} + P_{e\tau}$ is the transition probability of electron neutrinos to the other active neutrino flavours. The indices LE (low energy) and HE (high energy) refer to the energy regions below and above the MSW resonance, respectively.

Under the same set of assumptions adopted in [60], the probabilities in eq. (A.37) only depend on the angles $\theta_{12}, \theta_{13}, \theta_{14}$. Since it has been shown in ref. [56] that the determination of θ_{13} is unaffected by the presence of a sterile neutrino, throughout this work we will fix the value of θ_{13} to the three-neutrino best-fit point from ref. [82], $\sin^2\theta_{13} = 0.0223$. Moreover, in ref. [60] it was also shown that the most conservative bound for θ_{14} is achieved for $\sin^2\theta_{12} = 0.3125$. Thus, this will also be fixed throughout, for simplicity (we have explicitly checked that allowing θ_{12} to vary has a negligible impact on our results).

The χ^2 function for solar data is computed as:

$$\chi^2_{\text{Solar}} = \sum_{r,s}(O_r - P_r)V_{rs}^{-1}(O_s - P_s). \tag{A.38}$$

where O_r and P_r stand for the observed and predicted values for the probabilities in eq. (A.37), respectively, while V_{rs}^{-1} denotes the inverse covariance matrix, see eq. (A.40)

---

[11]In order to reproduce the results in ref. [18] better, we scale down their uncertainties by 3%.

below. The observations O_r, as well as the corresponding uncertainties σ_r and correlation matrix ρ, are taken from ref. [60, table 1]. In this work we use the values for the GS98 solar model [63], corresponding to the upper half of the table. Let us define the relative covariance matrix for the observations as

$$S_{rs} = \rho_{rs}\frac{\sigma_r}{O_r}\frac{\sigma_s}{O_s} \tag{A.39}$$

(no sum over repeated indices). Following ref. [60], a good approximation to the full solar-neutrino fit is obtained by splitting the inverse covariance matrix V_{rs}^{-1} into a theoretical and an experimental part, as

$$V_{rs}^{-1} = \frac{S_{rs}^{-1}}{2}\left[\frac{1}{O_r}\frac{1}{O_s} + \frac{1}{P_r}\frac{1}{P_s}\right] = \frac{1}{2}\left[X_{rs}^{-1} + S_{rs}^{-1}\frac{1}{P_r}\frac{1}{P_s}\right], \quad \text{where} \quad X_{rs} = \rho_{rs}\sigma_r\sigma_s \tag{A.40}$$

(no sum over repeated indices), and S^{-1} is the inverse of S. Using the inverse covariance matrix defined in eq. (A.40) the χ^2 for solar data can be computed as in eq. (A.38).

Regarding generation of pseudo data, note when $P_r = O_r$, we have $V_{rs} = X_{rs}$ and the dependence on P_r drops out (see [60, eq. (2.12)]). Therefore, we use the Cholesky decomposition of X in order to generate the statistical fluctuations for the pseudo data, following the procedure outlined in appendix B.2.

B   Details of the Feldman-Cousins analysis

B.1   Treatment of nuisance parameters

When generating pseudo data, in addition to statistical fluctuations on the event rates, we need to include fluctuations on the nuisance parameters as well. This means that the pseudo data should be generated using a set of "true" nuisance parameters, ξ_{true}, as the mean for ξ. The question is what value should be taken for ξ_{true}. Within a frequentist approach, *a priori* one should scan all possible values of ξ and study the distribution of the test statistic accordingly (that is, treat the nuisance parameters as if they were additional parameters of the model). However, in practice this is often not possible and a choice has to be made regarding the assumed values for ξ_{true}.

For all reactor experiments considered in this work, we will leave the overall normalization of the event rates completely free; as such, we will *not* consider fluctuations of these free pulls, fixing them to the values that provide the best fit to the observed data. We do so, because a free normalization implies that no prior information on that parameter is being used. Therefore, since the only data to consider are those measured by the experiment, it seems plausible to use them in order to determine the flux normalization. In other words: we allow each experiment to "choose" their preferred value of the overall normalization for the event rates, given that we are only interested in the observation of an oscillatory signal in the spectrum. Note that this is done independently for each set of assumed "true" oscillation parameters, $\sin^2 2\theta, \Delta m^2$.

For the rest of the nuisance parameters considered here, prior measurements are typically available and can be used to determine the relevant range of values to consider when

generating the fluctuations. Thus, for each nuisance parameter we will generate Gaussian fluctuations around the best-fit values from previous measurements (i.e., the mean value as reported by the collaboration), taking the reported uncertainties as the width of the Gaussian.

B.2 Generation of pseudo data

We use the following procedure to generate pseudo data in our analysis:

(i) choose the parameters Δm^2, $\sin^2 2\theta$ for the hypothesis of new oscillations;

(ii) calculate the normalization-related nuisance parameters $\{\eta^{\mathrm{min}}\}$ that provide the best fit to the observed data for the assumed hypothesis;

(iii) generate fluctuations for the nuisance parameters subject to prior constraints, taking as central values their reported best-fit values $\overline{\xi}_k$, i.e.,

$$\xi_k^{\mathrm{pseudo}} = \overline{\xi}_k + \sigma_{\xi_k}\delta_k \,, \tag{B.1}$$

where δ_k are standard-normal fluctuations and $\overline{\xi}_k$ are the values to which the pulls are constrained by the penalty terms;

(iv) compute the prediction for the assumed hypothesis, using the fluctuated nuisance parameters and calculated flux normalizations, and set this as the mean for the generation of pseudo data, i.e., for each bin i set

$$\overline{D}_i^{\mathrm{pseudo}} = P_i \left(\Delta m^2, \sin^2 2\theta; \{\xi^{\mathrm{pseudo}}\}, \{\eta^{\mathrm{min}}\} \right); \tag{B.2}$$

(v) if the χ^2 function contains a covariance matrix V, first recalculate (if necessary) by replacing all data with $\overline{D}_i^{\mathrm{pseudo}}$ in V;

(vi) calculate the Cholesky decomposition L of V^{pseudo} (note if χ^2 does not contain a covariance matrix, then L is diagonal with the standard deviations of each bin as the diagonal entries);

(vii) inject fluctuations around the pseudo-data mean with L, i.e.,

$$D_i^{\mathrm{pseudo}} = \overline{D}_i^{\mathrm{pseudo}} + \sum_j L_{ij}\delta_j, \tag{B.3}$$

where δ_j are standard-normal fluctuations.

B.3 Comment on the χ^2 distribution related to randomized nuisance parameters

Here we provide some justification for the way we generate pseudo data in the presence of constrained nuisance parameters. For this aim we consider the simplified case of a single

pull ξ describing a normalization uncertainty. Without loss of generality we set $\overline{\xi} = 0$, such that eqs. (B.1) and (B.2) become:

$$\xi^{\text{pseudo}} = \sigma_\xi \, \delta \,, \tag{B.4}$$

$$\overline{D}_i^{\text{pseudo}} = (1 + \xi^{\text{pseudo}}) P_i(\theta) \,, \tag{B.5}$$

where θ denotes generic model parameters. For simplicity, let us also assume that the data is a priori uncorrelated with variances σ_i^2 and eq. (B.3) becomes

$$D_i^{\text{pseudo}} = \overline{D}_i^{\text{pseudo}} + \sigma_i \delta_i \,. \tag{B.6}$$

Then we obtain for our test statistic for the pseudo data:

$$\chi^2_{\text{pseudo}} = \min_\xi \left[\sum_{i=1}^N \frac{[D_i^{\text{pseudo}} - (1 + \xi) P_i(\theta)]^2}{\sigma_i^2} + \frac{\xi^2}{\sigma_\xi^2} \right] \,. \tag{B.7}$$

Let us now prove that for fixed parameters θ, the test statistic χ^2_{pseudo} is distributed as a χ^2 with N dof. First, it follows from eqs. (B.4) to (B.6) that D_i^{pseudo} are multivariant Gaussian variables with mean and covariances of

$$\langle D_i^{\text{pseudo}} \rangle = P_i(\theta) \,, \quad \text{Var}(D_i^{\text{pseudo}}, D_j^{\text{pseudo}}) \equiv S_{ij} = \delta_{ij} \sigma_i^2 + P_i(\theta) P_j(\theta) \sigma_\xi^2 \,, \tag{B.8}$$

with $\text{Var}(X_i, X_j) = \langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle$ and δ_{ij} is the Kronecker delta. On the other hand, eq. (B.7) has the standard form of a pull-χ^2, which is known[12] to be equivalent to $\chi^2_{\text{pseudo}} = \sum_{ij} [D_i^{\text{pseudo}} - P_i(\theta)] S_{ij}^{-1} [D_j^{\text{pseudo}} - P_j(\theta)]$ with S_{ij} given in eq. (B.8). Hence, by a suitable variable transformation (involving the diagonalization of S) one can write $\chi^2_{\text{pseudo}} = \sum_{i=1}^N \zeta_i^2$, with ζ_i being N independent standard-normal random variables, and therefore χ^2_{pseudo} follows a χ^2 distribution with N dof.

Remark 1. This proof generalizes in a straight-forward way if there is a nontrivial correlation matrix between the data points, and to the case of several pulls, as long as they enter the prediction linearly.

Remark 2. In this proof it has been essential to use the reported best-fit value $\overline{\xi}$ as central value to fluctuate pull, as in eq. (B.1). (Remember that we have set $\overline{\xi} = 0$ in this section, without loss of generality.) If for generating the random pulls a mean value different from $\overline{\xi}$ was used, the pseudo-data would be biased and χ^2_{pseudo} would follow a non-central χ^2 distribution.

B.4   MC uncertainty of FC confidence intervals

We describe how we estimate the uncertainty on our FC confidence regions due to the finite size of the MC sample. Consider a point in the parameter space, which has a true p-value p. Then for a total of N MC trials, the number n of pseudo-data sets with $\Delta\chi^2_{\text{pseudo}} > \Delta\chi^2_{\text{obs}}$ is distributed as a binomial distribution, with parameters N and p. Hence, we can define a
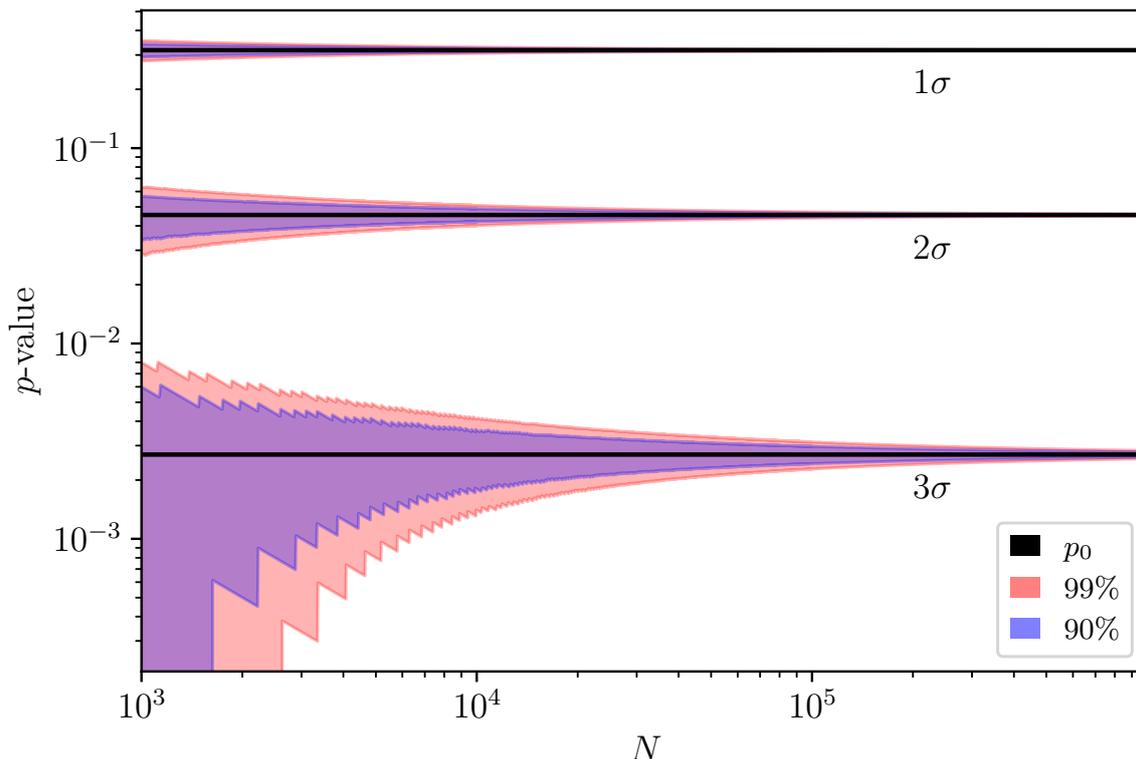
---

[12]See ref. [83] for an explicit proof.

Figure 9. 90% and 99% uncertainty on the p-value for $p = 0.32$ (1σ), 0.045 (2σ), and 0.0027 (3σ) as a function of the MC sample size N.

p-value distribution via $f(q) = N B\left(n; N, p\right)$, where $q = n/N$ and $B\left(\cdot\,;\,\cdot\,,\,\cdot\right)$ is the binomial PDF. An $\alpha = 99\%$ confidence interval $[q_0, q_1]$ for this p-value can be defined as

$$F(q_0) = \frac{1-\alpha}{2}, \qquad F(q_1) = \frac{1+\alpha}{2}, \tag{B.9}$$

where $F(\cdot)$ is the CDF of $f(\cdot)$. For small p-values, the distribution is skewed. Note that since the distribution f is discrete, the relations in eq. (B.9) are only approximate; they become exact in the large-N limit.

The MC approximation for the p-value is n/N. We take this as the p parameter for the binomial distribution (i.e., we assume the MC approximation is exact). To obtain the 99%-confidence uncertainty bands for a confidence region, we create a p-value map in the $\sin^2 2\theta$–Δm^2 plane. Then we interpret the area between the contours at q_0 and q_1 as the 99% confidence uncertainty on the β CL contour with $\beta = 1 - p$.

Note that this prescription, and in particular eq. (B.9), is independent of the specific experiment. It follows from the properties of the binomial distribution and depends only on the CL $\beta = 1 - p$, the MC sample size N, and the uncertainty α. We plot some numerical values in figure 9. The discontinuities visible in the bands for 3σ CL are due to the discrete nature of the binomial distribution. Note that for a MC sample size of 10^3 it is not possible to calculate a meaningful 3σ confidence region at the accuracy of $\alpha = 0.99$ or 0.9, as follows from $q_0 = 0$ for these parameters.

References

[1] S. Böser et al., *Status of Light Sterile Neutrino Searches*, *Prog. Part. Nucl. Phys.* **111** (2020) 103736 [`arXiv:1906.01739`] [InSPIRE].

[2] B. Dasgupta and J. Kopp, *Sterile Neutrinos*, *Phys. Rept.* **928** (2021) 63 [`arXiv:2106.05913`] [InSPIRE].

[3] ALEPH, DELPHI, L3, OPAL, SLD collaborations, LEP Electroweak Working Group, SLD Electroweak Group and SLD Heavy Flavour Group, *Precision electroweak measurements on the Z resonance*, *Phys. Rept.* **427** (2006) 257 [`hep-ex/0509008`] [InSPIRE].

[4] G. Mention et al., *The Reactor Antineutrino Anomaly*, *Phys. Rev. D* **83** (2011) 073006 [`arXiv:1101.2755`] [InSPIRE].

[5] T.A. Mueller et al., *Improved Predictions of Reactor Antineutrino Spectra*, *Phys. Rev. C* **83** (2011) 054615 [`arXiv:1101.2663`] [InSPIRE].

[6] P. Huber, *On the determination of anti-neutrino spectra from nuclear reactors*, *Phys. Rev. C* **84** (2011) 024617 [*Erratum ibid.* **85** (2012) 029901] [`arXiv:1106.0687`] [InSPIRE].

[7] SAGE collaboration, *Measurement of the response of the Russian-American gallium experiment to neutrinos from a ^{51}Cr source*, *Phys. Rev. C* **59** (1999) 2246 [`hep-ph/9803418`] [InSPIRE].

[8] J.N. Abdurashitov et al., *Measurement of the response of a Ga solar neutrino experiment to neutrinos from an ^{37}Ar source*, *Phys. Rev. C* **73** (2006) 045805 [`nucl-ex/0512041`] [InSPIRE].

[9] GALLEX collaboration, *Final results of the ^{51}Cr neutrino source experiments in GALLEX*, *Phys. Lett. B* **420** (1998) 114 [InSPIRE].

[10] F. Kaether, W. Hampel, G. Heusser, J. Kiko and T. Kirsten, *Reanalysis of the GALLEX solar neutrino flux and source experiments*, *Phys. Lett. B* **685** (2010) 47 [`arXiv:1001.2731`] [InSPIRE].

[11] M. Laveder, *Unbound neutrino roadmaps*, *Nucl. Phys. B Proc. Suppl.* **168** (2007) 344 [InSPIRE].

[12] M.A. Acero, C. Giunti and M. Laveder, *Limits on ν_e and $\bar{\nu}_e$ disappearance from Gallium and reactor experiments*, *Phys. Rev. D* **78** (2008) 073009 [`arXiv:0711.4222`] [InSPIRE].

[13] C. Giunti and M. Laveder, *Statistical Significance of the Gallium Anomaly*, *Phys. Rev. C* **83** (2011) 065504 [`arXiv:1006.3244`] [InSPIRE].

[14] M. Danilov, *New results from the DANSS experiment*, *PoS* **ICHEP2020** (2021) 121 [`arXiv:2012.10255`] [InSPIRE] and online at https://indico.desy.de/event/28202/contributions/105957/.

[15] NEOS collaboration, *Sterile Neutrino Search at the NEOS Experiment*, *Phys. Rev. Lett.* **118** (2017) 121802 [`arXiv:1610.05134`] [InSPIRE].

[16] PROSPECT collaboration, *Improved short-baseline neutrino oscillation search and energy spectrum measurement with the PROSPECT experiment at HFIR*, *Phys. Rev. D* **103** (2021) 032001 [`arXiv:2006.11210`] [InSPIRE].

[17] STEREO collaboration, *Improved sterile neutrino constraints from the STEREO experiment with 179 days of reactor-on data*, *Phys. Rev. D* **102** (2020) 052002 [arXiv:1912.06582] [INSPIRE].

[18] A.P. Serebrov et al., *Search for sterile neutrinos with the Neutrino-4 experiment and measurement results*, *Phys. Rev. D* **104** (2021) 032003 [arXiv:2005.05301] [INSPIRE].

[19] M. Dentler, A. Hernández-Cabezudo, J. Kopp, M. Maltoni and T. Schwetz, *Sterile neutrinos or flux uncertainties? — Status of the reactor anti-neutrino anomaly*, *JHEP* **11** (2017) 099 [arXiv:1709.04294] [INSPIRE].

[20] S. Gariazzo, C. Giunti, M. Laveder and Y.F. Li, *Model-independent $\bar{\nu}_e$ short-baseline oscillations from reactor spectral ratios*, *Phys. Lett. B* **782** (2018) 13 [arXiv:1801.06467] [INSPIRE].

[21] M. Agostini and B. Neumair, *Statistical Methods Applied to the Search of Sterile Neutrinos*, *Eur. Phys. J. C* **80** (2020) 750 [arXiv:1906.11854] [INSPIRE].

[22] C. Giunti, *Statistical Significance of Reactor Antineutrino Active-Sterile Oscillations*, *Phys. Rev. D* **101** (2020) 095025 [arXiv:2004.07577] [INSPIRE].

[23] PROSPECT and STEREO collaborations, *Note on arXiv:2005.05301, 'Preparation of the Neutrino-4 experiment on search for sterile neutrino and the obtained results of measurements'*, arXiv:2006.13147 [INSPIRE].

[24] P. Coloma, P. Huber and T. Schwetz, *Statistical interpretation of sterile neutrino oscillation searches at reactors*, *Eur. Phys. J. C* **81** (2021) 2 [arXiv:2008.06083] [INSPIRE].

[25] S.S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, *Annals Math. Statist.* **9** (1938) 60 [INSPIRE].

[26] V.V. Barinov et al., *Results from the Baksan Experiment on Sterile Transitions (BEST)*, arXiv:2109.11482 [INSPIRE].

[27] V. Barinov and D. Gorbunov, *BEST Impact on Sterile Neutrino Hypothesis*, arXiv:2109.14654 [INSPIRE].

[28] C. Giunti, Y.F. Li, C.A. Ternes and Z. Xin, *Reactor antineutrino anomaly in light of recent flux model refinements*, arXiv:2110.06820 [INSPIRE].

[29] A.C. Hayes and P. Vogel, *Reactor Neutrino Spectra*, *Ann. Rev. Nucl. Part. Sci.* **66** (2016) 219 [arXiv:1605.02047] [INSPIRE].

[30] C. Giunti, X.P. Ji, M. Laveder, Y.F. Li and B.R. Littlejohn, *Reactor Fuel Fraction Information on the Antineutrino Anomaly*, *JHEP* **10** (2017) 143 [arXiv:1708.01133] [INSPIRE].

[31] C. Giunti, Y.F. Li, B.R. Littlejohn and P.T. Surukuchi, *Diagnosing the Reactor Antineutrino Anomaly with Global Antineutrino Flux Data*, *Phys. Rev. D* **99** (2019) 073005 [arXiv:1901.01807] [INSPIRE].

[32] J.M. Berryman and P. Huber, *Sterile Neutrinos and the Global Reactor Antineutrino Dataset*, *JHEP* **01** (2021) 167 [arXiv:2005.01756] [INSPIRE].

[33] DAYA BAY collaboration, *Antineutrino energy spectrum unfolding based on the Daya Bay measurement and its applications*, *Chin. Phys. C* **45** (2021) 073001 [arXiv:2102.04614] [INSPIRE].

[34] Daya Bay collaboration, *Evolution of the Reactor Antineutrino Flux and Spectrum at Daya Bay*, *Phys. Rev. Lett.* **118** (2017) 251801 [arXiv:1704.01082] [InSPIRE].

[35] STEREO collaboration, *Accurate Measurement of the Electron Antineutrino Yield of ^{235}U Fissions from the STEREO Experiment with 119 Days of Reactor-On Data*, *Phys. Rev. Lett.* **125** (2020) 201801 [arXiv:2004.04075] [InSPIRE].

[36] F. Von Feilitzsch, A.A. Hahn and K. Schreckenbach, *Experimental beta spectra from ^{239}Pu and ^{235}U thermal neutron fission products and their correlated anti-neutrinos spectra*, *Phys. Lett. B* **118** (1982) 162 [InSPIRE].

[37] K. Schreckenbach, G. Colvin, W. Gelletly and F. Von Feilitzsch, *Determination of the anti-neutrino spectrum from ^{235}U thermal neutron fission products up to 9.5 MeV*, *Phys. Lett. B* **160** (1985) 325 [InSPIRE].

[38] A.A. Hahn, K. Schreckenbach, G. Colvin, B. Krusche, W. Gelletly and F. Von Feilitzsch, *Anti-neutrino Spectra From ^{241}Pu and ^{239}Pu Thermal Neutron Fission Products*, *Phys. Lett. B* **218** (1989) 365 [InSPIRE].

[39] V. Kopeikin, M. Skorokhvatov and O. Titov, *Reevaluating reactor antineutrino spectra with new measurements of the ratio between ^{235}U and ^{239}Pu β spectra*, *Phys. Rev. D* **104** (2021) L071301 [arXiv:2103.01684] [InSPIRE].

[40] Daya Bay collaboration, *Measurement of the Reactor Antineutrino Flux and Spectrum at Daya Bay*, *Phys. Rev. Lett.* **116** (2016) 061801 [*Erratum ibid.* **118** (2017) 099902] [arXiv:1508.04233] [InSPIRE].

[41] Double Chooz collaboration, *Measurement of θ_{13} in Double CHOOZ using neutron captures on hydrogen with novel background rejection techniques*, *JHEP* **01** (2016) 163 [arXiv:1510.08937] [InSPIRE].

[42] RENO collaboration, *Observation of Energy and Baseline Dependent Reactor Antineutrino Disappearance in the RENO Experiment*, *Phys. Rev. Lett.* **116** (2016) 211801 [arXiv:1511.05849] [InSPIRE].

[43] B. Armbruster et al., *KARMEN limits on $\nu_e \to \nu_\tau$ oscillations in 2-ν and 3-ν mixing schemes*, *Phys. Rev. C* **57** (1998) 3414 [hep-ex/9801007] [InSPIRE].

[44] LSND collaboration, *Measurements of charged current reactions of ν_e on ^{12}C*, *Phys. Rev. C* **64** (2001) 065501 [hep-ex/0105068] [InSPIRE].

[45] J.M. Conrad and M.H. Shaevitz, *Limits on Electron Neutrino Disappearance from the KARMEN and LSND ν_e — Carbon Cross Section Data*, *Phys. Rev. D* **85** (2012) 013017 [arXiv:1106.5552] [InSPIRE].

[46] T2K collaboration, *Search for short baseline ν_e disappearance with the T2K near detector*, *Phys. Rev. D* **91** (2015) 051102 [arXiv:1410.8811] [InSPIRE].

[47] P.B. Denton, *Sterile Neutrino Searches with MicroBooNE: Electron Neutrino Disappearance*, arXiv:2111.05793 [InSPIRE].

[48] C.A. Argüelles et al., *MicroBooNE and the ν_e Interpretation of the MiniBooNE Low-Energy Excess*, arXiv:2111.10359 [InSPIRE].

[49] KATRIN collaboration, *Bound on $3 + 1$ Active-Sterile Neutrino Mixing from the First Four-Week Science Run of KATRIN*, *Phys. Rev. Lett.* **126** (2021) 091803 [arXiv:2011.05087] [InSPIRE].

[50] C. Giunti, Y.F. Li and Y.Y. Zhang, *KATRIN bound on* $3 + 1$ *active-sterile neutrino mixing and the reactor antineutrino anomaly*, *JHEP* **05** (2020) 061 [arXiv:1912.12956] [INSPIRE].

[51] S. Hagstotz et al., *Bounds on light sterile neutrino mass and mixing from cosmology and laboratory searches*, *Phys. Rev. D* **104** (2021) 123524 [arXiv:2003.02289] [INSPIRE].

[52] PLANCK collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [*Erratum ibid.* **652** (2021) C4] [arXiv:1807.06209] [INSPIRE].

[53] G.J. Feldman and R.D. Cousins, *A Unified approach to the classical statistical analysis of small signals*, *Phys. Rev. D* **57** (1998) 3873 [physics/9711021] [INSPIRE].

[54] DAYA BAY collaboration, *Improved Measurement of the Reactor Antineutrino Flux and Spectrum at Daya Bay*, *Chin. Phys. C* **41** (2017) 013002 [arXiv:1607.05378] [INSPIRE].

[55] J. Kostensalo, J. Suhonen, C. Giunti and P.C. Srivastava, *The gallium anomaly revisited*, *Phys. Lett. B* **795** (2019) 542 [arXiv:1906.10980] [INSPIRE].

[56] J. Kopp, P.A.N. Machado, M. Maltoni and T. Schwetz, *Sterile Neutrino Oscillations: The Global Picture*, *JHEP* **05** (2013) 050 [arXiv:1303.3011] [INSPIRE].

[57] J.N. Bahcall, *Gallium solar neutrino experiments: Absorption cross-sections, neutrino spectra, and predicted event rates*, *Phys. Rev. C* **56** (1997) 3391 [hep-ph/9710491] [INSPIRE].

[58] S.V. Semenov, *Cross Section of Neutrino Absorption by the Gallium-71 Nucleus*, *Phys. Atom. Nucl.* **83** (2020) 1549 [INSPIRE].

[59] J. Elevant and T. Schwetz, *On the determination of the leptonic CP phase*, *JHEP* **09** (2015) 016 [arXiv:1506.07685] [INSPIRE].

[60] K. Goldhagen, M. Maltoni, S. Reichard and T. Schwetz, *Testing sterile neutrino mixing with present and future solar neutrino data*, arXiv:2109.14898 [INSPIRE].

[61] M. Maltoni, T. Schwetz, M.A. Tortola and J.W.F. Valle, *Ruling out four neutrino oscillation interpretations of the LSND anomaly?*, *Nucl. Phys. B* **643** (2002) 321 [hep-ph/0207157] [INSPIRE].

[62] M. Maltoni and T. Schwetz, *Testing the statistical compatibility of independent data sets*, *Phys. Rev. D* **68** (2003) 033020 [hep-ph/0304176] [INSPIRE].

[63] N. Vinyoles et al., *A new Generation of Standard Solar Models*, *Astrophys. J.* **835** (2017) 202 [arXiv:1611.09867] [INSPIRE].

[64] M. Danilov, *Searches for sterile neutrinos at very short baseline reactor experiments*, *J. Phys. Conf. Ser.* **1390** (2019) 012049 [arXiv:1812.04085] [INSPIRE].

[65] M.V. Danilov and N.A. Skrobova, *Comment on "Analysis of the Results of the Neutrino-4 Experiment on the Search for the Sterile Neutrino and Comparison with Results of Other Experiments" (JETP Letters 112, 199 (2020))*, *JETP Lett.* **112** (2020) 452 [*Pisma Zh. Eksp. Teor. Fiz.* **112** (2020) 484] [INSPIRE].

[66] C. Giunti, Y.F. Li, C.A. Ternes and Y.Y. Zhang, *Neutrino-4 anomaly: oscillations or fluctuations?*, *Phys. Lett. B* **816** (2021) 136214 [arXiv:2101.06785] [INSPIRE].

[67] J.A. Formaggio, A.L.C. de Gouvêa and R.G.H. Robertson, *Direct Measurements of Neutrino Mass*, *Phys. Rept.* **914** (2021) 1 [arXiv:2102.00594] [INSPIRE].

[68] P. Vogel and J.F. Beacom, *Angular distribution of neutron inverse beta decay,* $\bar{\nu}_e + p \to e^+ + n$, *Phys. Rev. D* **60** (1999) 053003 [hep-ph/9903554] [INSPIRE].

[69] DANSS collaboration, *Sensitivity of DANSS detector to short range neutrino oscillations*, arXiv:1412.0817 [ɪɴSPIRE].

[70] I. Alekseev et al., *DANSS: Detector of the reactor AntiNeutrino based on Solid Scintillator*, 2016 *JINST* **11** P11011 [arXiv:1606.02896] [ɪɴSPIRE].

[71] Y.N. Oh, *NEOS result and reactor SBL-ν project*, talk given at the *26th International Workshop on Weak Interactions and Neutrinos (WIN2017)*, University of California, Irvine, California, U.S.A., 19–24 June 2017.

[72] P. Huber, *NEOS Data and the Origin of the* 5 *MeV Bump in the Reactor Antineutrino Spectrum*, *Phys. Rev. Lett.* **118** (2017) 042502 [arXiv:1609.03910] [ɪɴSPIRE].

[73] Y. Ko et al., *Sterile neutrino search at the NEOS experiment*, *PoS* **NuFact2017** (2017) 042 [ɪɴSPIRE].

[74] Daya Bay collaboration, *Measurement of electron antineutrino oscillation based on 1230 days of operation of the Daya Bay experiment*, *Phys. Rev. D* **95** (2017) 072006 [arXiv:1610.04802] [ɪɴSPIRE].

[75] P. Huber, M. Lindner and W. Winter, *Simulation of long-baseline neutrino oscillation experiments with GLoBES (General Long Baseline Experiment Simulator)*, *Comput. Phys. Commun.* **167** (2005) 195 [hep-ph/0407333] [ɪɴSPIRE].

[76] P. Huber, J. Kopp, M. Lindner, M. Rolinec and W. Winter, *New features in the simulation of neutrino oscillation experiments with GLoBES 3.0: General Long Baseline Experiment Simulator*, *Comput. Phys. Commun.* **177** (2007) 432 [hep-ph/0701187] [ɪɴSPIRE].

[77] L. Bernard, *Recherche d'un neutrino stérile avec l'expérience STEREO: détermination des spectres neutrinos et caractérisation du bruit de fond*, Ph.D. Thesis, LPSC, Grenoble France (2019).

[78] STEREO collaboration, *Sterile Neutrino Constraints from the STEREO Experiment with 66 Days of Reactor-On Data*, *Phys. Rev. Lett.* **121** (2018) 161801 [arXiv:1806.02096] [ɪɴSPIRE].

[79] R. Fletcher, *Practical Methods of Optimization*, in *Practical Methods of Optimization*, second edition, John Wiley & Sons, Ltd (2000), chapter 4, pp. 80–94.

[80] STEREO collaboration, *Improved sterile neutrino constraints from the STEREO experiment with 179 days of reactor-on data*, (2020) [https://doi.org/10.17182/hepdata.92323.v3].

[81] X. Qian, A. Tan, J.J. Ling, Y. Nakajima and C. Zhang, *The Gaussian CL_s method for searches of new physics*, *Nucl. Instrum. Meth. A* **827** (2016) 63 [arXiv:1407.5052] [ɪɴSPIRE].

[82] I. Esteban, M.C. Gonzalez-Garcia, M. Maltoni, T. Schwetz and A. Zhou, *The fate of hints: updated global analysis of three-flavor neutrino oscillations*, *JHEP* **09** (2020) 178 [arXiv:2007.14792] [ɪɴSPIRE].

[83] G.L. Fogli, E. Lisi, A. Marrone, D. Montanino and A. Palazzo, *Getting the most from the statistical analysis of solar neutrino oscillations*, *Phys. Rev. D* **66** (2002) 053010 [hep-ph/0206162] [ɪɴSPIRE].