Future searches for light sterile neutrinos at nuclear reactors

Jeffrey M. Berryman[1,2,*] Luis A. Delgadillo,[3,†] and Patrick Huber[3,‡]

[1]*Department of Physics and Astronomy, University of Kentucky, Lexington, Kentucky 40506, USA*
[2]*Department of Physics, University of California, Berkeley, California 94720, USA*
[3]*Center for Neutrino Physics, Department of Physics, Virginia Tech, Blacksburg, Virginia 24061, USA*

We study the optimization of a green-field, two-baseline reactor experiment with respect to the sensitivity for electron antineutrino disappearance in search of a light sterile neutrino. We consider both commercial and research reactors and identify as key factors the distance of closest approach and detector energy resolution. We find that a total of 5 tons of detectors deployed at a commercial reactor with a closest approach of 25 m can probe the mixing angle $\sin^2 2\theta$ down to $\sim 5 \times 10^{-3}$ around $\Delta m^2 \sim 1$ eV^2. The same detector mass deployed at a research reactor can be sensitive up to $\Delta m^2 \sim 20$–30 eV^2 assuming a closest approach of 3 m and excellent energy resolution, such as that projected for the Taishan Antineutrino Observatory. We also find that lithium doping of the reactor could be effective in increasing the sensitivity for higher Δm^2 values.

I. INTRODUCTION

Neutrino physics as an experimental science started with the discovery of Cowan and Reines using neutrinos from a reactor in 1956 [1]. Since then, reactor measurements have been a mainstay in the quest to understand neutrino properties. KamLAND [2], Double Chooz [3], Daya Bay [4], and RENO [5] have played central roles in establishing the oscillation of the three active neutrino flavors. In the near future, the JUNO experiment will provide the best measurement of the so-called solar parameters θ_{12} and Δm^2_{21} as well as determine the mass ordering [6]. In the run-up to the 2011/2012 measurement of θ_{13}, the question of how to predict the reactor antineutrino spectrum received renewed attention [7,8] and the surprising result was an upward correction to the predicted inverse beta decay (IBD) event rate by approximately 6%. This, in turn, gave rise to the so-called reactor antineutrino anomaly (RAA) [9], which would be naturally explained if the $\bar{\nu}_e$ mixed with an additional species of neutrino with a mixing angle $\sin^2 2\theta \simeq 0.1$ and $\Delta m^2 \geq 1$ eV^2. For a review of the status of the field at that time, see Ref. [10]. Enormous experimental progress has been made since then and we refer the

reader to the global fitting literature for details, see, e.g., Ref. [11–17]; for a more experiment-centered review, see Ref. [18].

As far as reactor neutrino experiments are concerned, modern experiments (i.e., those conducted after 2011) [19–23] all rely on a comparison of measured event rate spectra at different baselines and are thus *independent* of flux predictions. Nonetheless, reactor flux predictions have been subject to intense study [24–27]; see also Refs. [11,12,14,15,17,28–32]. Moreover, the rate anomaly has not, as yet, been resolved, though it may be less significant than originally suggested.[1] The spectral anomaly known as the 5- MeV bump [22,23,34–36] is entirely unresolved; solutions grounded in nuclear physics have been proposed (see, e.g., Refs. [27,37–40]) and it is likely not due to new physics [41]. The experiments we will discuss here will contribute data on neutrino fluxes, but this is not a focus of our work. Instead, we ask what the best possible reactor neutrino experiment would look like to either find a light sterile neutrino or to exclude a sizable mixing. We also focus exclusively on electron (anti) neutrinos and their potential mixing with a sterile neutrino. This choice is motivated, to some degree, by the fact that all data on $\nu_e \to \nu_e$ and $\bar{\nu}_e \to \bar{\nu}_e$ at this point are mutually consistent. The same cannot be said of the $\nu_\mu \to \nu_e$ and $\bar{\nu}_\mu \to \bar{\nu}_e$ data sets; additionally, the global disappearance

*jeffberryman@berkeley.edu
†luisd@vt.edu
‡pahuber@vt.edu

[1]Recent work [33], if confirmed, would provide a simple and consistent solution to the RAA by shifting the ratio of ^{239}Pu to ^{235}U in the integrated beta spectrum. In this case, beta and neutrino data would be in good agreement with both summation and conversion calculations of the reactor flux.

data are known to be inconsistent with the global appearance data when interpreted in the context of a truly sterile fourth neutrino [13,18,42,43]. On top of all of this, an eV-scale sterile neutrino of the sort indicated by terrestrial oscillation anomalies is severely constrained by cosmological observations [44–47]. Our hope is that a careful study of neutrino disappearance at reactors will facilitate future analyses of the global neutrino data set.

The reason nuclear reactors are the electron neutrino[2] source of choice is threefold:

(1) They are free, in the sense that they are constructed for purposes other than neutrino physics.

(2) They are very bright, with experimentally accessible fluxes of up to 10^{13} cm^{-2} s^{-1}.

(3) They produce antineutrinos which can be cleanly detected via IBD, which both provides a flavor tag and has a large (relative to other weak-interaction processes) cross section of approximately 6×10^{-43} cm^2 per fission.

Other sources of electron neutrinos which have been considered for sterile neutrino searches include radioactive sources [48–55],[3] beta-beams [59,60], kaon beams [61], beam-driven beta-decay sources [62] and stored muon beams [63]. None of these other electron neutrino sources shares all of the advantages of nuclear reactors—in particular they all would be quite costly and require detectors much larger than what will be considered here. On the other hand, some of these approaches like beam-driven beta-decay sources and stored muon beams would yield far better sensitivities than possible at a reactor. But the question remains: what sensitivities are a *possible* with a reactor source? This is the subject of the present study.

II. GREEN-FIELD STUDIES

For the neutrino energies available at nuclear reactors and for the baselines that we will consider (less than 100 m), oscillations among the three neutrinos of the Standard Model (SM) are negligible. Therefore, if a sterile neutrino[4] exists and participates in oscillations, then any

oscillations observed over these distances would be attributable to the new state. We may describe these oscillations in the two-flavor limit, writing the $\bar{\nu}_e$ survival probability as

$$P_{\bar{e}\bar{e}} = 1 - \sin^2 2\theta \sin^2\left(\frac{\Delta m^2 L}{4E_\nu}\right). \qquad (2.1)$$

Here, L is the baseline over which the neutrinos propagate and E_ν is their energy. The parameter Δm^2 is the difference in the squares of the masses of two neutrino mass eigenstates; $\sin^2 2\theta$ is the mixing angle describing the amplitude of the oscillations. Probing sterile neutrinos amounts to determining $\sin^2 2\theta$ and Δm^2.

The overarching purpose of this study is to determine the optimal configuration(s) for a two-baseline reactor neutrino experiment with realistic operational assumptions. We categorize these by the class of reactor at which they are conducted—at either a *commercial* or a *research* reactor facility. The salient difference between these is that the former is much larger ($R \sim 2$ m) than the latter ($R \sim 20$ cm); commercial cores are also more powerful and produce a larger flux of antineutrinos. However, the layout of the facilities at which these are housed prevents an experiment from operating within a certain distance of the core, typically of the order 10–25 m. Research cores, though less powerful, are less constrained in this regard: one can operate an experiment as close as 3–6 m from the core. The increase in flux from the shorter allowed baseline partially offsets the lower absolute rate of antineutrino production. We will consider each of these scenarios in turn.

We note similar previous works presented in Refs. [64,65]. Our work differs from these in two primary ways. Firstly, our focus is to systematically study the optimal physical configuration of the two detectors, depending on the type of reactor core. In contrast, these references are oriented more towards understanding the effects of, e.g., backgrounds and systematic uncertainties in dictating sensitivities. Secondly, our work benefits from nearly a decade of additional experimental effort. In particular, these references largely predate measurements of spectral ratios [19–23,66–68], which favor decidedly different regions of the sterile neutrino parameter space than the traditional RAA. Our work serves as a complement to these early efforts.

A. Methodology

We consider pseudoexperiments in which the spectrum of reactor antineutrinos is observed at two baselines. We treat the detector as pointlike, but the reactor core has finite physical extent; we treat the latter as a perfect sphere. Because of this, we cannot use the oscillation probability in Eq. (2.1) as is—we must flux average $P_{\bar{e}\bar{e}}$ over the production region. The effect of this is to average out high-frequency oscillations: if the oscillation wavelength is comparable to or smaller than the size of the production

---

[2]This paper deals only with electron flavor neutrinos and/or antineutrinos and we rely on context to disambiguate these two cases.

[3]Since this work first appeared, the BEST Collaboration has reported an ∼20% deficit in the antineutrino rate from a chromium source relative to prediction [56]. These results can be interpreted [57] as an ∼5σ indication of the existence of a sterile neutrino. It would be premature to declare that a sterile neutrino has been discovered; however, given the tension between this result and solar neutrino data [58]. This underscores the importance of fully addressing the anomalies in the reactor sector in order to form a coherent picture of $\nu_e/\bar{\nu}_e$ disappearance.

[4]We will use the term "sterile" neutrino to refer to a generic fourth neutrino that participates in oscillations but not in weak interactions. This state may have interactions beyond those of the SM, but if these do not affect oscillations, then there is no need to distinguish it from a truly sterile neutrino.

region, then the experiment will have muted sensitivity to the associated value of Δm^2. Details on this procedure and the resulting average survival probability are given in the Appendix.

The analysis window is taken to be $E_{\text{prompt}} \subset [2.0, 8.0]$ MeV; we generally assume a bin size of 250 keV, but occasionally consider finer spacing. We assume a constant, nonzero background in each energy bin, ignoring any systematic uncertainty on this background and assuming that the near and far baselines experience the same background rate per unit exposure.[5] We consider two energy response models as limiting cases of the detector resolution. The first is the response model of the PROSPECT experiment, which the collaboration has provided in the Supplemental Material to Ref. [69]. The response may be roughly described as a Gaussian with width $\sim 7\%/\sqrt{E/[\text{MeV}]}$, though it includes a non-Gaussian tail down to low energies. On the other end of the spectrum is the upcoming Taishan Antineutrino Observatory (TAO) experiment [70], for which the resolution will be of the order $\sim 1\%/\sqrt{E[\text{MeV}]}$. We assume that the TAO detector response is perfectly Gaussian.

For a given experimental configuration, specified by the near and far baselines (L_{near} and L_{far}, respectively), we form the following χ^2 function:

$$\chi^2 = \sum_i^{N_{\text{bins}}} \left[\left(\frac{M_i^N - \phi_i P_i^N(\sin^2 2\theta, \Delta m^2, \{\eta_j\})}{\sigma_i^N} \right)^2 \right. \\ \left. + \left(\frac{M_i^F - \phi_i P_i^F(\sin^2 2\theta, \Delta m^2, \{\eta_j\})}{\sigma_i^F} \right)^2 \right] \\ + \sum_j^{N_{\text{sys}}} \left(\frac{\eta_j}{\sigma_j} \right)^2. \qquad (2.2)$$

The components of this expression are as follows:

(i) The M_i^N and M_i^F represent the numbers of pseudo-data IBD counts in energy bin i in the near and far detectors, respectively. These pseudodata are generated without oscillations.

(ii) The P_i^N and P_i^F are the predicted numbers of events in energy bin i for oscillation parameters $\{\sin^2 2\theta, \Delta m^2\}$ and nuisance parameters $\{\eta_j\}$, which we discuss more below.

(iii) The σ_i^N and σ_i^F are the statistical uncertainties associated with the predictions. We account for the nonzero background in forming σ_i^N and σ_i^F; we write

$$\sigma_i^{N,F} = \sqrt{P_i^{N,F} + \frac{1+r}{r} B^{N,F}}, \qquad (2.3)$$

where B^N (B^F) is the number of background events per energy bin in the near (far) detector. The factor of r in Eq. (2.3) is the ratio of reactor-off to reactor-on times; this arises from background subtraction. The expected raw number of events in each bin during the reactor-on period is $P_i^{N,F} + B^{N,F}$; during the reactor-off period, this is $rB^{N,F}$. Reweighting the (Poissonian) uncertainty on the latter by r^{-1} and adding this in quadrature with the uncertainty on the former leads to Eq. (2.3). We have simplified our analysis by taking $r = 1$, i.e., that the reactor-on and reactor-off times are equal.

(iv) The ϕ_i are nuisance parameters for the flux in each energy bin. Our analysis is *flux-free* in that we do not introduce a prior on these nuisance parameters; while we use the Huber-Mueller (HM) flux model [7,8] to generate our pseudodata, our analysis is not sensitive to this choice. These unconstrained nuisance parameters reduce the number of degrees of freedom in the fit, leading to lower values of the χ^2 and, ultimately, less statistical significance. However, sensitivities derived in this fashion are more robust than those that rely on a given flux model. We may analytically minimize the χ^2 with respect to these nuisance parameters to obtain

$$\phi_i = \frac{\frac{M_i^N P_i^N}{(\sigma_i^N)^2} + \frac{M_i^F P_i^F}{(\sigma_i^F)^2}}{\left(\frac{P_i^N}{\sigma_i^N} \right)^2 + \left(\frac{P_i^F}{\sigma_i^F} \right)^2}. \qquad (2.4)$$

(v) Lastly, the $\{\eta_j\}$ are the nuisance parameters describing our systematic uncertainties, which are assumed to be consistent with the systematics described by the STEREO Collaboration [71]:

(1) A 1.2% uncorrelated normalization uncertainty.
(2) A 1.0% uncorrelated energy scale uncertainty.
(3) A 0.3% correlated energy scale uncertainty.

We expect this to characterize the capabilities of next-generation reactor experiments. We ignore the possibility of (uncorrelated) systematic uncertainties in the measured spectral shape in our initial studies; we discuss this further in Sec. II E.

We only consider simple $\Delta \chi^2$ statistics here. In recent years, it has been repeatedly emphasized that searches for sterile neutrino oscillations do not satisfy Wilks' theorem: simple $\Delta \chi^2$ statistics do not provide the proper coverage in the $\sin^2 2\theta$–Δm^2 parameter space, so converting $\Delta \chi^2$ values to confidence levels using Gaussian statistics is formally incorrect (see, e.g., Refs. [72,73]). Consequently, we will avoid rigorously assigning confidence levels. That said, statistical methods that aim to address these deficiencies (e.g., the CL_s method [74] and Feldman-Cousins

---

[5]In truth, the accidental background rate ought to be higher closer to the reactor core. We are ignoring this for simplicity, but note that this would require a more careful study under real-world conditions. Given the commensurate increase in signal rate, however, we do not expect this to dramatically change our findings.

method [75]) are computationally intensive; performing these at scale is beyond the scope of this work. Therefore, we report our sensitivities as contour(s) along which $\Delta\chi^2 = 11.83$—for two degrees of freedom, this would correspond to 3σ if Wilks' theorem applied. In most cases, the true significance will be less than this.

B. Research reactor optimization

We begin with research reactors. Because of their smaller size and because they are not attached to the heat-extraction apparatus of a commercial core, one could operate an experiment with much shorter baselines at a research core. For instance, the MINER experiment [76] can operate within 13 m of a low-power (1 MW$_{\text{th}}$) reactor. However, it is more typical of the current generation of short-baseline experiments to operate in the range 6–10 m. We consider baselines as close as 3 m and as far as 50 m from a core whose radius is taken to be 20 cm and whose power is set to 100 MW$_{\text{th}}$. The total exposure is fixed at 1 yr, which we assume corresponds to 90 d of effective reactor-on time, accounting for a finite detector efficiency ($\sim$50%). Moreover, because research reactors use high-enriched uranium fuel, the flux of antineutrinos is entirely attributable to ^{235}U fission. We calculate the total number of events in each detector to be

$$N_{\text{total}}^{\text{research}} \approx 1.5 \times 10^6$$
$$\times \left(\frac{10 \text{ m}}{L}\right)^2 \left(\frac{m_{\text{det}}T\epsilon}{5 \text{ ton yr}}\right)\left(\frac{P}{100 \text{ MW}_{\text{th}}}\right), \quad (2.5)$$

where m_{det} is the detector mass, T the total exposure time and ϵ an efficiency factor, which accounts for detector efficiency and reactor up time.

The downside of operating so close to the core is backgrounds: on one hand, operating close to the core subjects, the detector to reactor-correlated neutron and gamma backgrounds. On the other, the detector is necessarily close to the surface subjecting the detector to cosmogenic backgrounds. It is possible to mitigate either source of background, but this presents a fundamental challenge to the operation of such an experiment. We will assume that a background rate of 250 events per ton-day of reactor-on time spread uniformly over the analysis region at both detector sites. This is broadly consistent with the background rate achieved at PROSPECT [69].

In Figs. 1(a) and 1(b), we show representative sensitivities to oscillations with $\Delta m^2 = 5$ eV^2 for a PROSPECT-like and TAO-like detector response, respectively. The color conveys the sensitivity ($\Delta\chi^2 = 11.83$) to $\sin^2 2\theta$ for each pair of near and far baselines; white space indicates baseline pairs for which there is no sensitivity. The central conclusion is that getting as close as possible to the reactor core is the key factor in improving the sensitivity. Indeed, the sensitivity is maximized when the spatial separation of the two detectors corresponds to half the oscillation wavelength L_{osc} at $E_\nu \approx 4$ MeV, where the event rate is maximized:

$$L_{\text{osc}} \equiv \frac{4E_\nu}{\pi\Delta m^2} \approx (2 \text{ m}) \times \left(\frac{E_\nu}{4 \text{ MeV}}\right)\left(\frac{5 \text{ eV}^2}{\Delta m^2}\right). \quad (2.6)$$
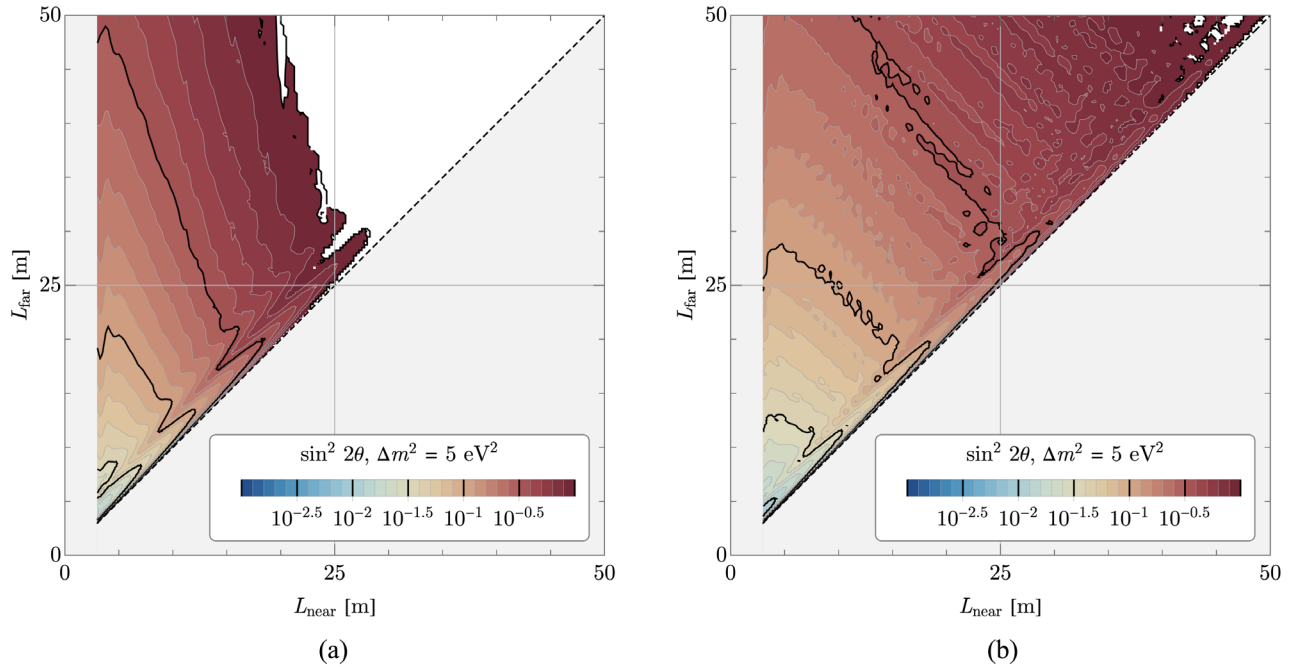

FIG. 1.   The sensitivity ($\Delta\chi^2 = 11.83$) of a hypothetical two-baseline research reactor experiment to oscillations with $\Delta m^2 = 5$ eV^2 as a function of its near and far baselines, L_{near} and L_{far}. Panel (a) shows results for PROSPECT-like response, while panel (b) shows results for TAO-like response.

Note that even in the case of a detector with TAO-like energy resolution, the sensitivity barely reaches $\sin^2 2\theta \sim 10^{-2}$ for the closest conceivable baseline of 3 m. Constructing an experiment this close to a core will be challenging; a less difficult case is a near baseline of ~ 6 m. In this case, the optimal sensitivity worsens slightly to $\sin^2 2\theta \sim 3 \times 10^{-2}$.

C. Commercial reactor optimization

Commercial reactors can be several orders of magnitude more powerful than research reactors, meaning that one does not need to operate an experiment at as short a baseline to have an appreciable event rate. This is fortunate, since one *cannot* conduct an experiment at such close baselines, given the complicated layout of commercial reactor plants. We take the radius of the core to be 2.0 m, the core power to be 4.5 GW_{th} and the exposure to be 1 yr, which we assume corresponds to 365 d of reactor-on time. Moreover, the fuel at commercial reactors is typically low-enriched uranium, meaning the antineutrino flux is a nontrivial combination of the fluxes from the four main fissile isotopes (^{235}U, ^{238}U, ^{239}Pu, ^{241}Pu). We take the following values for the effective isotopic fission fractions:

$$\{^{235}U, ^{238}U, ^{239}Pu, ^{241}Pu\} = \{0.56, 0.07, 0.31, 0.06\}.$$

We calculate the total number of events in each detector to be

$$N_{total}^{commercial} \approx 1.9 \times 10^7$$
$$\times \left(\frac{20 \text{ m}}{L}\right)^2 \left(\frac{m_{det} T \epsilon}{5 \text{ ton yr}}\right) \left(\frac{P}{4.5 \text{ GW}_{th}}\right). \quad (2.7)$$

The closest baseline that any experiment has attained at a commercial reactor is ~ 10 m at DANSS experiment [77], the result of the specific construction of the Kalinin Nuclear Power Plant. A more typical scenario is akin to the NEOS experiment [19], which sits ~ 24 m from a commercial core in a tendon gallery at the Hanbit Nuclear Power Plant. We consider baselines between 10–100 m in our simulations. The impact of this lack of closeness on sterile neutrino searches is twofold. Firstly, because shorter baselines allow for searches for larger oscillation frequencies, the sensitivity of commercial reactor experiments to larger values of Δm^2 will be muted. Also, the larger core size also causes high-frequency oscillations to average out. Secondly, because the experiment is further from the core, the backgrounds are less severe: reactor-correlated neutron and gamma backgrounds are substantially attenuated and higher overburdens can be achieved to reduce cosmogenic backgrounds. We assume a background rate of 90 events per ton-day of reactor-on time, roughly

corresponding to the background rate measured at NEOS [19].[6]

In Figs. 2(a) and 2(b), we show sensitivities to oscillations with $\Delta m^2 = 1$ eV^2. The color scale is the same as in Figs. 1(a) and 1(b). The sensitivity is improved for commercial reactors over research reactors, owing to the much larger event rate: for a given common baseline, the factors of ~ 45 in assumed power and ~ 4 in effective run time lead to a factor ~ 100–200 in signal event rate at a commercial core relative to a research core. As before, the key factor in determining the sensitivity is allowing for as short a baseline as achievable, subject to the constraints of the facility. Moreover, we again observe that the separation of the detectors should be roughly half the oscillation wavelength in Eq. (2.6). The best sensitivity shown for the TAO-like response corresponds to $\sin^2 2\theta \sim 3 \times 10^{-3}$, but even restricting the baselines to be no less than 25 m or using the PROSPECT response model (or both) yields a maximum sensitivity better than $\sin^2 2\theta \sim 10^{-2}$ for this value of Δm^2.

D. Comparisons

We aggregate our results for both the commercial and research reactor cases in Fig. 3. The figure shows the baseline-optimized sensitivity ($\Delta\chi^2 = 11.83$) to $\sin^2 2\theta$ particular to a given Δm^2. We note that there is no achievable configuration that maximizes the sensitivity at every such value. The purple curves correspond to commercial reactors, whereas green curves are for research reactors. The lighter shade of either color exploits the full range of baselines, i.e., that a detector can be placed as close as possible to a given type of core (10 m for commercial, 3 m for research). The darker shade, in contrast, truncates the nearest baseline at what we believe is a more realistic distance (25 m for commercial, 6 m for research). The shading indicates the effect of the detector resolution, ranging between the PROSPECT (lower) and TAO (upper) responses.

For a fixed exposure, the sensitivity of a commercial reactor experiment exceeds that of a research reactor experiment by some $\mathcal{O}(1$–10$)$ factor, a consequence of the increased statistics at commercial experiments. However these experiments are most sensitive to oscillations with different values of Δm^2, a consequence of the different core sizes and permissible baselines. If one were optimizing a sterile neutrino search for $\Delta m^2 \lesssim 2$–3 eV^2, then one would ultimately elect to operate the experiment at

---

[6]NEOS was able to achieve this background rate with an overburden of ~ 20 m.w.e. Generically, a detector at such short distances would require a decent amount of shielding in order to keep background rates low, but given that this has been achieved at NEOS, we assume it will be possible in future experiments, too. Given the large signal rates, we do not expect that our results would be appreciably different if the background rate were modestly larger than this.
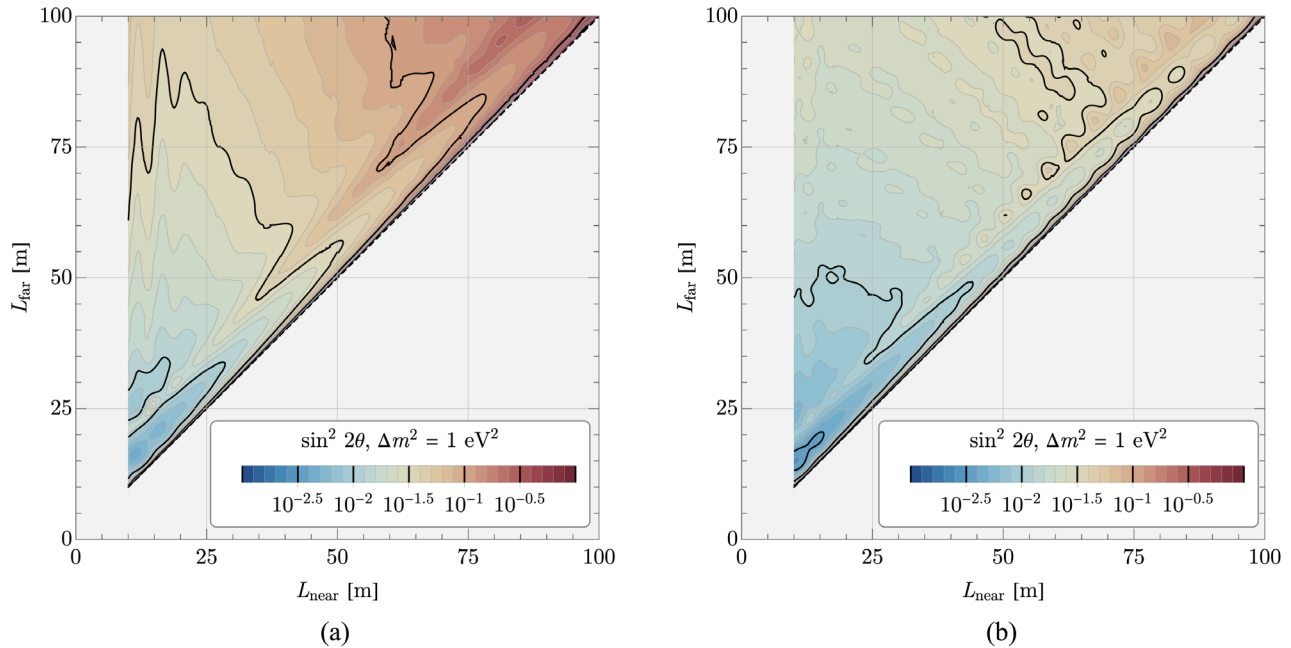
FIG. 2. The sensitivity ($\Delta\chi^2 = 11.83$) of a hypothetical two-baseline commercial reactor experiment to oscillations with $\Delta m^2 = 1$ eV^2 as a function of its near and far baselines, L_{near} and L_{far}. Panel (a) shows results for PROSPECT-like response, while panel (b) shows results for TAO-like response.

a commercial reactor; conversely, for larger values of Δm^2, one would choose a research reactor. This figure concretely demonstrates the benefit of putting the detector closer to the core for probing high-frequency oscillations—a transition occurs around $\Delta m^2 \sim 0.5$ eV^2 (~ 2 eV^2) for commercial (research) reactors where the advantage of higher statistics and shorter baselines dissipates.

To contextualize these sensitivities, we also show the projected final 95% C.L. sensitivity of the ongoing KATRIN tritium-decay experiment [79] in dashed gray. While a research core would be the preferred method up to $\Delta m^2 \sim 30$ eV^2 if one could realize ultrashort baselines, KATRIN is ostensibly the preferred technique for searches above $\Delta m^2 \sim 15$ eV^2 for more realistic configurations. However, there is an important caveat in interpreting the KATRIN sensitivity: this depends on the prior on the effective neutrino mass squared, m_ν^2, that one extracts from the experiment. If one insists that $m_\nu^2 \geq 0$—a sensible physical criterion—then one derives the sensitivity shown in the figure. However, one may reasonably insist on a different prior on this quantity, given that historical tritium-decay experiments have often found negative best-fit values of this parameter; see Ref. [80] and references therein. If one were to allow m_ν^2 to take any value—positive or negative—to avoid a biased measurement, then the resulting sensitivity changes; this is depicted in Fig. 3 of Ref. [79] for the exclusion based on current data. We are unaware of any study of this effect on the ultimate sensitivity of KATRIN and do not attempt one here, but this suggests that the constraint may not be robust below

$\Delta m^2 \sim 30$ eV^2. If this is the case, then research reactor experiments may be the best path forward to exploring this portion of the parameter space.

The last two components of Fig. 3 are the regions of parameter space preferred by analyses of separate existing datasets. The first is the 3σ-preferred region from an analysis of the Neutrino-4 experiment [78], shown in red shading. The second is the 3σ-preferred[7] region from a recent global analysis of Bugey-3, DANSS, Daya Bay, Double Chooz, NEOS and RENO from Ref. [15], shown in blue shading. These are the regions that next-generation reactor experiments will target—indeed, the presence of closed 3σ contours in existing analyses is what inspires the next generation of experiments in the first place.[8] Our results suggest that the optimal strategy to probe these regions would be to build a commercial reactor experiment to probe below $\Delta m^2 \sim 2$ eV^2 and a research reactor experiment to probe above this, and that they should both be as close as can be achieved to their respective cores.

We have considered the effect of bin size on Fig. 3. Specifically, we have looked at bins of width 100 keV for both a PROSPECT-like and TAO-like response model at both a commercial and research reactor facility. On balance,

---

[7]What this reference calls "3σ" should more appropriately be called the "$\Delta\chi^2 = 11.83$" exclusion.

[8]There are, of course, practical applications of antineutrino physics that will require research and development; see, e.g., Ref. [81] and references therein. However, these technologies depend on a precise understanding of isotopic antineutrino fluxes and are thus inextricably coupled to the oscillation anomalies.
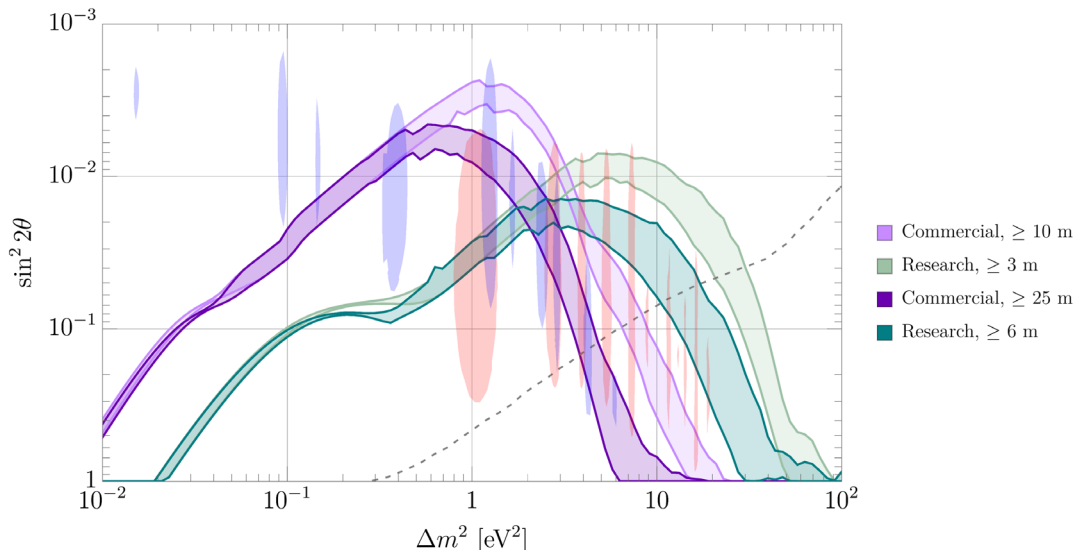
FIG. 3. The baseline-optimized sensitivity ($\Delta\chi^2 = 11.83$) to $\sin^2 2\theta$ as a function of Δm^2. The purple (green) curves are for a commercial (research) reactor. The lighter band of either color represents the sensitivity with optimistic assumptions about the possible closest baseline; the darker bands represent more realistic assumptions. The weakest (strongest) sensitivity of each color corresponds to a PROSPECT-like (TAO-like) energy resolution. Also shown is the 95% C.L. sensitivity from KATRIN (dashed gray), as well as the 3σ-preferred regions from Neutrino-4 [78] (red) and from the global analysis of Ref. [15] (blue).

the changes induced are not markedly different from Fig. 3. We have also investigated the effect of 30-keV bins on analyses with our TAO-like response and again find no appreciable shift in sensitivity.

E. Exposure and systematics limitation

One can ask how the sensitivities shown in Fig. 3 can be improved by either operating the experiment for a longer period of time or by constructing a more massive detector. To study this, we must include an effect that we have ignored to this point. While our flux-free analysis does not introduce any prior uncertainties on the flux, one should account for systematic uncertainties on the measured spectral shape, in addition to the systematics that we have already included. We assume that this systematic is uncorrelated between energy bins and that next-generation experiments can achieve shape uncertainties as low as 0.5%, consistent with the value claimed by TAO [70]. This is introduced to our χ^2 in Eq. (2.2) via the replacement

$$(\sigma_i^{N,F})^2 \to (\sigma_i^{N,F})^2 + (P_i^{N,F}\sigma_{sh})^2, \qquad (2.8)$$

where σ_{sh} represents the shape systematic.

To demonstrate the interplay between the exposure and a finite shape systematic, we consider a pseudoexperiment at a research reactor with a near baseline of 6 m and a far baseline of 9 m, roughly corresponding to the near and far baselines at PROSPECT, that employs the TAO response model. In Fig. 4, we show the sensitivity to $\sin^2 2\theta$ for fixed values of $\Delta m^2 - 1$ eV^2 in cyan and 5 eV^2 in magenta—as a function of exposure. The vertical, gray line corresponds to the

5-ton-years exposure used in our previous analyses. The dashed curves correspond to vanishing shape uncertainty. For these, as the exposure is increased, the sensitivity improves without bound: the increase in statistics leads to an increasingly precise determination of the event rate at each position and in each energy bin, resulting in sensitivity that scales as $\sim(\text{exposure})^{-1/2}$. In contrast, the solid curves show the sensitivity including a 0.5% shape uncertainty. As the number of raw counts is increased, the fractional statistical uncertainty eventually becomes eclipsed by the
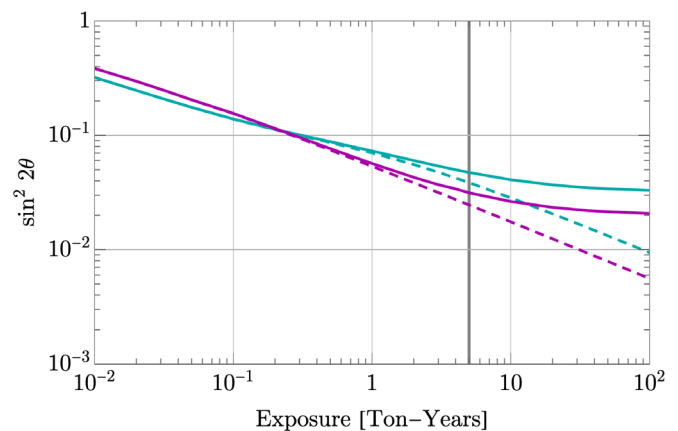

FIG. 4. The sensitivity ($\Delta\chi^2 = 11.83$) to $\sin^2 2\theta$ as a function of exposure for a research reactor experiment with $L_{near} = 6$ m and $L_{far} = 9$ m. Cyan (magenta) curves correspond to $\Delta m^2 = 1$ eV^2 ($\Delta m^2 = 5$ eV^2). Dashed curves correspond to vanishing shape systematic uncertainty; solid curves take this to be 0.5%. The vertical, gray line indicates our benchmark 5 ton-years exposure, at which value each detector observes 3.2×10^5 events.

shape uncertainty—the experiment becomes systematics limited. Past this point, the sensitivity saturates at some finite value. Inspection of Fig. 4 reveals that for our nominal 5-ton-year exposure, the experiment is already transitioning from being statistically limited to being systematically limited. Consequently, increasing the exposure beyond this value results in a negligible increase in sensitivity to $\sin^2 2\theta$.

III. STERILE NEUTRINO SENSITIVITY AT TAO

The TAO is a future experiment whose main purpose is to produce a model-independent reference spectrum for the physics program at JUNO [6]. It is planned to start taking data in 2022 and is expected to achieve an unprecedented experimental resolution of the order $\sim 1.0\%/\sqrt{E[\text{MeV}]}$. Though it is a *single-volume* detector, its fiducial volume may be *virtually* segmented, allowing for a multiple-baseline measurement of the sort outlined in the previous section. The segmentation is possible due to improvements in the coverage and efficiency of the photo sensors, as well as the precise timing resolution; see, e.g., Sec. 8.5 of Ref. [70]. This is a novel capability that has not been employed by existing searches for sterile neutrino. We have previously considered two detectors located at arbitrary baselines; here, on the other hand, we consider a fixed (core-to-detector) baseline position of 30 m and compare the spectra measured in each of its virtual segments. We will virtually segment the detector into either two or four equal-volume segments; the displacements between the barycenters and the center of the detector are $(\pm 3/8\ R)$ and $(\pm 0.170R, \pm 0.580R)$, respectively, $R = 65$ cm being the radius of the fiducial volume at TAO. While virtual segmentation is an impressive capability of this detector, the fixed geometry is a nontrivial restriction on a multiple-baseline measurement.

We assume a total of 5-ton-year exposure at a commercial core with a 2.0-m radius and a power of 4.5 GW$_{\text{th}}$, in broad agreement with the TAO conceptual design report (CDR) [70]. The fuel fractions of the core are assumed to be the same as our green-field studies (Sec. II). Moreover, a total of 60 bins between 2.0 to 8.0 MeV in prompt energy are considered. Background events are spread uniformly over the energy bins, assuming a total of 90 events per ton-day, consistent with those at NEOS [19]; we do not consider shape uncertainties on the background. Systematic uncertainties in this experiment are assumed to be consistent with those at STEREO [71]. Furthermore, a 0.5% bin-to-bin uncorrelated shape uncertainty is taken into account.

For our experimental configuration, we form the following χ^2 function:

$$\chi^2 = \sum_i^{N_{\text{bin}}} \sum_j^{N_{\text{seg}}} \left(\frac{M_{i,j} - \phi_i P_{i,j}(\sin^2 2\theta, \Delta m^2, \{\eta_\ell\})}{\sigma_{i,j}} \right)^2$$

$$+ \sum_\ell^{N_{\text{sys}}} \left(\frac{\eta_\ell}{\sigma_\ell} \right)^2. \tag{3.1}$$

This expression closely resembles Eq. (2.2), the salient difference being that we now index the N_{seg} (=2, 4) segments with j. We account for the nonzero background and shape uncertainty in a similar fashion as Eq. (2.3) by defining $\sigma_{i,j}$ as

$$\sigma_{i,j}^2 = P_{i,j}(1 + \sigma_{\text{sh}}^2 \times P_{i,j}) + 2B_{i,j}, \tag{3.2}$$

were σ_{sh} is the bin-to-bin shape uncertainty and $B_{i,j}$ is the (uniform) background in each energy bin i of segment j.[9] The ϕ_i are flux nuisance parameters in each energy bin. As before, we analytically minimize the χ^2 with respect to these nuisance parameters to obtain

$$\phi_i = \frac{\sum_j \frac{M_{i,j} P_{i,j}}{(\sigma_{i,j})^2}}{\sum_j \left(\frac{P_{i,j}}{\sigma_{i,j}} \right)^2}, \tag{3.3}$$

where j runs over the detector segments. Lastly, the $\{\eta_\ell\}$ are the nuisance parameters describing our systematic uncertainties.

In Ref. [70], the HM flux model [7,8] is used to obtain the antineutrino energy spectrum with an introduction of a 5% bin-to-bin uncorrelated shape uncertainty on the flux. While we, too, use the HM fluxes, our analysis is not sensitive to this choice; by using the ϕ_i, our analysis employs a data-to-data spectral comparison, as compared to a data-to-model one. This removes any model dependence on our analysis, but reduces the number of degrees of freedom in the fit, leading to lower values of the χ^2 and, ultimately, less inferred statistical significance. Therefore, we expect the sensitivities obtained in this study to be less aggressive but more robust. Another important difference between our analysis and that of Ref. [70] is the assessment of the backgrounds: the collaboration publishes a combined rate of accidental, fast neutron and ^{9}Li/^{8}He decay backgrounds of ~ 450 events per ton-day, a more conservative figure than that assumed in our study. We have consciously elected to be more optimistic, given that the 90 events per ton-day we consider has been achieved at NEOS [19]. We also remind the reader that we only communicate sensitivities as contours of constant $\Delta\chi^2$, whereas Ref. [70] provides the results of Gaussian CL$_s$ analysis. This is not, strictly speaking, an apples-to-apples comparison of sensitivities; however, the gross features of the analyses are distinct enough that one can still draw important qualitative conclusions.

We show our results in Fig. 5. The dashed and solid cyan curves show the sensitivity ($\Delta\chi^2 = 11.83$) to sterile neutrinos for our analyses with two and four segments, respectively. The dotted-dashed, magenta curve corresponds to the 99.7% C.L.$_s$ sensitivity from the TAO

---

[9] We are again assuming that the reactor-off and reactor-on times are equal, leading to the factor of 2 in Eq. (3.2).
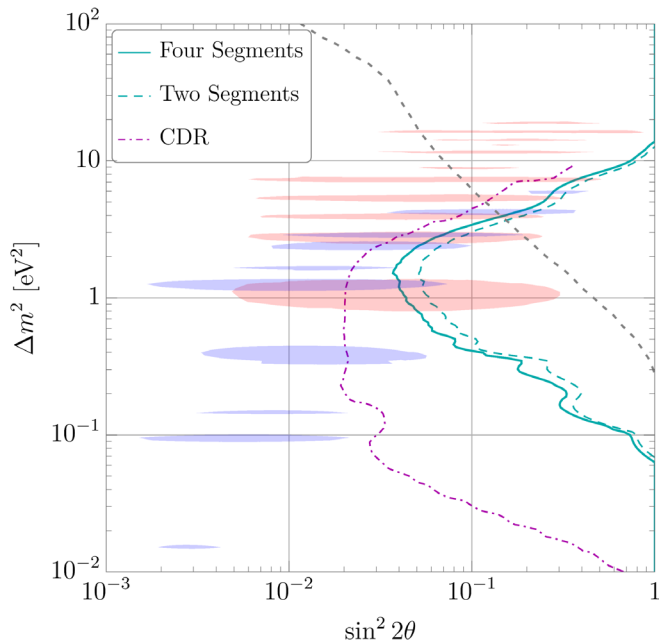
FIG. 5. The sensitivity ($\Delta\chi^2 = 11.83$) to a sterile neutrino at TAO. Cyan curves correspond to our analysis with four (solid) or two (dashed) virtual segments. The dotted-dashed, magenta curve corresponds to the 99.7% C.L.$_s$ sensitivity presented in Ref. [70]. Also shown are the 95% C.L. sensitivity from KATRIN (dashed gray), as well as the 3σ-preferred regions from Neutrino-4 [78] (red shading) and from Ref. [15] (blue shading).

CDR [70]. We also show the 95% C.L. sensitivity from KATRIN in dashed gray, as well as the 3σ-preferred regions from Neutrino-4 [78] and from Ref. [15], in red and blue shading contours, respectively.

In the region $\Delta m^2 \leq 1$ eV^2, our analysis is substantially less sensitive than that of the CDR. This is a consequence of the (in)dependence on the flux model in these analyses. By comparing directly to the HM flux model, the CDR analysis is sensitive to changes in the antineutrino spectrum induced over the $\sim$30-m baseline between the core and the detector, affording sensitivity to lower values of Δm^2. In contrast, by comparing data to data, our analysis is sensitive only to changes induced over the $\sim$1-m length scale spanned by the detector. Therefore, sensitivity to longer-wavelength oscillations is inevitably muted, and the barycenter of the sensitivity curve shifts to higher values of Δm^2. Furthermore, the addition of more segments improves the overall sensitivity; the gains, however, are modest. On one hand, increasing the number of segments increases the number of degrees of freedom in the fit. On the other, this also increases the (statistical) uncertainty of the contents of any given energy bin. The former evidently outweighs the latter, though only marginally. Lastly, we note that the results of this analysis are largely independent of the binning, despite the exceptional resolution available. The reason for this is clear: there exists a trade-off between increasing the number of bins and the statistical power of a given bin. For the nominal operation parameters we have employed, decreasing the bin size beyond 100 keV does not lead to a meaningful gain in sensitivity. Of course, increasing the exposure beyond the nominal 5 ton-years would lead to an improvement, but the gains in $\sin^2 2\theta$ are expected to be $\sim\mathcal{O}(1)$ before systematics become dominant.

On balance, the sensitivity to sterile neutrinos presented in this study is less optimistic than the TAO CDR, but this is a consequence of having fundamentally different search strategies. Flux model independence is central to our analysis, as it was in the previous section. We believe this is conceptually more robust, though we have made some assumptions that are more optimistic than those presented in the CDR.

IV. LITHIUM-8 DOPING OF RESEARCH REACTORS

As we have seen in the previous sections, the sensitivity to oscillations with $\Delta m^2 \gtrsim 2$ eV^2 is significantly improved if the near baseline is minimized as much as possible, due to (1) the increased event rate, and (2) high-frequency oscillations not averaging out. However, the physical layout of the facility severely constrains the possible location of a detector. Notice that because oscillations depend on the quantity $\Delta m^2 L/E$, the effect of reducing the baseline can be emulated by increasing the energies of the interacting antineutrinos. However, the energy spectrum of the antineutrinos emitted in fission is, of course, immutable.[10] In order to reach higher energies, then, one would need to consider a separate source of antineutrinos.

We consider precisely such a source, in the form of antineutrinos produced via the decays of ^{8}Li, produced by neutron capture on ^{7}Li. The main advantage is that the endpoint energy of these decays is $\sim$13 MeV; this provides the high energies desired to extend the sensitivity of a sterile neutrino search to higher Δm^2. Using lithium at a reactor to probe neutrino oscillations was previously suggested in Ref. [82]; we aim to robustly study the sensitivity to a sterile neutrino within this scheme. A similar proposal exists in the form of IsoDAR [62,83], but implementing a ^{8}Li source at a nuclear reactor has the advantage that no new facilities are required to produce the neutrons that ultimately drive the source. If one can load the reactor core with enough ^{7}Li, then the excess neutrons produced in the core can, in principle, be used to generate these β decays.

In Fig. 6, we show the flux-weighted IBD cross section per ^{8}Li decay against the same per ^{235}U fission. Because the IBD cross section grows roughly quadratically in energy,

---

[10]This is not strictly speaking true: the spectrum changes slightly as the composition of the reactor's fuel evolves. Here, we are referring to the impossibility of manipulating a given β decay into yielding more energetic neutrinos.
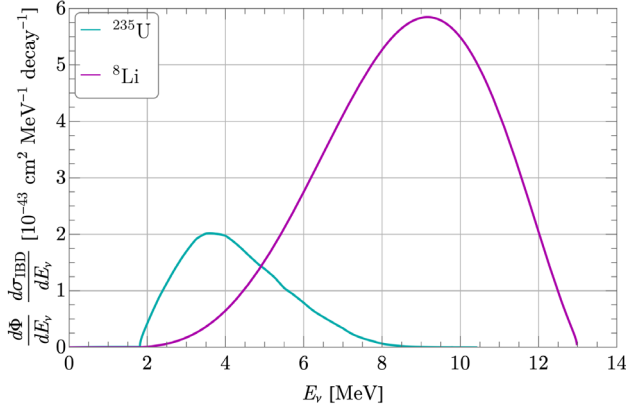
FIG. 6. The flux-weighted cross section per decay for IBD detection of antineutrinos from ^{235}U fission (cyan) and from ^{8}Li β decay (magenta).

the events associated with ^{8}Li β decay are disproportionately skewed to higher energies than those from fission. This results in a larger integrated flux-weighted cross section *per decay*; fewer antineutrinos are emitted per decay, but they are much more inclined to interact with the detector. However, while the per-decay rate of ^{8}Li events exceeds that of ^{235}U, it is unlikely that the absolute rate of ^{8}Li decay in an experiment could be arranged to be as large as the rate of ^{235}U fission. More broadly, there are a number of challenges in implementing ^{8}Li decays at nuclear reactors:

  (i) Inserting ^{8}Li into a commercial core is essentially out of the question. This would require a significant interruption of operation of a core used to generate electricity and, consequently, income for the operator.

 (ii) Research reactor cores, in contrast, are constructed such that one may insert a sample of some material into the core to absorb some excess neutrons without disrupting the core. However, if one wanted to achieve lithium decay rates equal to the fission rate, then one would essentially need to absorb *every* excess neutron over the operation of the experiment. Apart from reactor physics constraints, this would monopolize the facility, that is, no neutrons would be available for any other user.

(iii) Lastly, while natural lithium is dominantly (~92.4%) ^{7}Li, the thermal neutron capture cross section for ^{6}Li is roughly three orders of magnitude larger than that of ^{7}Li. Therefore, one would need high-purity ^{7}Li—around 99.9%—to ensure that the majority of neutrons will ultimately produce ^{8}Li; to ensure that 90% of neutrons capture on ^{7}Li, this increases to nearly 99.99% purity. Scaling from previous results on Li loading of HFIR [84] a 10% fraction of neutrinos may be possible in this case. Note that high-purity ^{7}Li is available in large

quantities and considered as coolant in molten salt reactors.

In light of these restrictions, we expect that it may be possible to achieve a rate of one ^{8}Li decay for every five to ten fission events at a research core.

We employ a pseudoexperiment with $L_{\rm near} = 6$ m and $L_{\rm far} = 9$ m. The total exposure of the experiment is fixed to be 5 ton-years, apportioned between the two detectors according to their inverse-squared baseline, and we continue to assume that one year at a research reactor corresponds to three months of reactor-on time with an average power of 100 MW$_{\rm th}$. The rate of background events is still assumed to be 250 events per ton-day *in the region 2.0–8.0 MeV prompt energy*, i.e., 41.7 counts per ton-day per MeV; we assume that this rate extends out to 13 MeV. We further assume a TAO-like energy resolution and employ 100-keV bins between 2.0 and 12.0 MeV in prompt energy to take full advantage of this resolution. The systematics budget is the same as in previous studies and we ignore any bin-uncorrelated shape uncertainties.

The results are shown in Fig. 7. The cyan curve shows the sensitivity ($\Delta\chi^2 = 11.83$) in the absence of any added lithium. We note that this curve is slightly more sensitive to oscillations with $\Delta m^2 \gtrsim 20$ eV^2 than the configuration-optimized sensitivity in Fig. 3 suggests, owing to the smaller bin size used here. Aside from the region around $\Delta m^2 \sim 3$–4 eV^2, this sensitivity is not too far from the
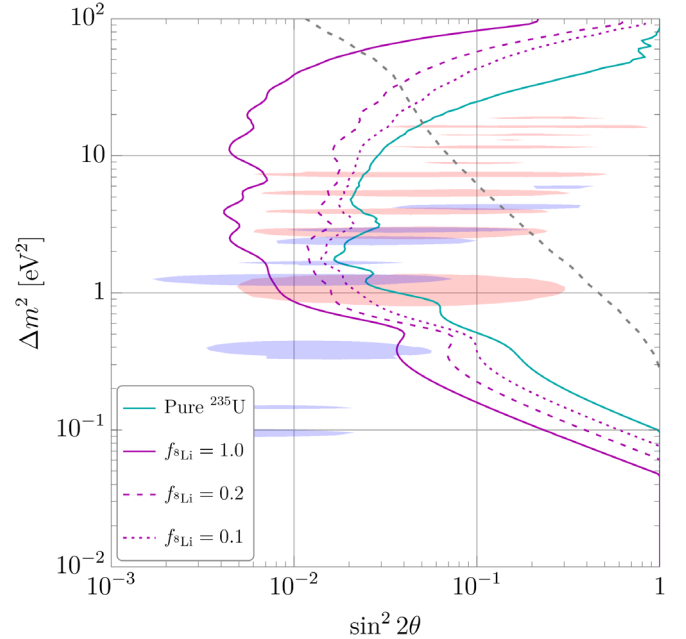

FIG. 7. The sensitivity ($\Delta\chi^2 = 11.83$) of a lithium-loaded research reactor to a sterile neutrino. The sensitivity without ^{8}Li is shown in cyan while the magenta curves assume varying numbers of lithium decays per fission, $f_{^8{\rm Li}}$: 1.0 (solid), 0.2 (dashed) and 0.1 (dotted). We also show the 95% C.L. sensitivity from KATRIN (dashed gray), as well as the 3σ-preferred regions from Neutrino-4 [78] (red shading) and Ref. [15] (blue shading).

optimal sensitivity for a 5-ton-year exposure; we expect this to be representative of the capabilities of a multiple-baseline experiment with baselines of this order. The magenta curves introduce varying numbers of ^{8}Li decays per ^{235}U fission, which we denote $f_{^8\text{Li}}$: solid corresponds to $f_{^8\text{Li}} = 1$, dashed to $f_{^8\text{Li}} = 0.2$ and dotted to $f_{^8\text{Li}} = 0.1$. We also show the projected sensitivity from KATRIN (dashed gray) and the 3σ-preferred regions from Neutrino-4 (red shading) and from Ref. [15] (blue shading). The benefit of including lithium decays is clear: for $f_{^8\text{Li}} = 1$, the sensitivity improves by as much as an order of magnitude, particularly above $\Delta m^2 \gtrsim 5$ eV^2. In light of the aforementioned issues in the interpretation of the KATRIN constraint in the region $\Delta m^2 \lesssim 30$ eV^2, reactor experiments may be the only means by which to robustly probe a sterile neutrino in this region of parameter space—and a highly lithium-loaded reactor may even exceed the sensitivity of KATRIN above this value.

However, for reasons laid out above, there are a number of impediments to achieving $f_{^8\text{Li}} = 1$. It would be more realistic to expect $f_{^8\text{Li}}$ in the range 0.1–0.2, in which case the dashed and dotted magenta lines should guide our interpretation of Fig. 7. The gains in sensitivity are more modest in this case. To compensate for this, one could envision increasing the exposure of the experiment. In Fig. 8, we show how the sensitivity evolves with the exposure. Cyan and orange curves correspond to the sensitivities for normal reactor conditions (i.e., pure ^{235}U) for $\Delta m^2 = 1$ eV^2 and 30 eV^2, respectively, whereas the magenta and red curves correspond to lithium loading with $f_{^8\text{Li}} = 0.1$. Dashed curves assume negligible shape systematics, whereas the solid curves introduce a 0.5% shape uncertainty. Our calculations with negligible shape uncertainty for a 5-ton-year exposure are in rough agreement with the sensitivities in the systematics-dominated limit. Moreover, the uranium-only and lithium-loaded curves asymptote to their ultimate sensitivities in lockstep. From Fig. 8, it is clear that increasing the exposure for pure uranium does not recover the sensitivity to higher Δm^2 obtained from adding even a small amount of lithium. Simply put, the increase in statistics between 2.0 and 8.0 MeV prompt energy is not enough to compensate for the total absence of events between 8.0 and 12.0 MeV. Given the choice between increasing the exposure of the experiment and increasing the lithium loading of the reactor, the latter is more effective in increasing the ultimate sensitivity.

In summary, our analysis suggests that loading a research core with highly enriched lithium is, from a physics perspective, an attractive possibility for enhancing the sensitivity of a reactor antineutrino experiment to a sterile neutrino, especially for moderately large values of Δm^2. That said, there are several practical hurdles that make any realization of such a concept technically demanding.
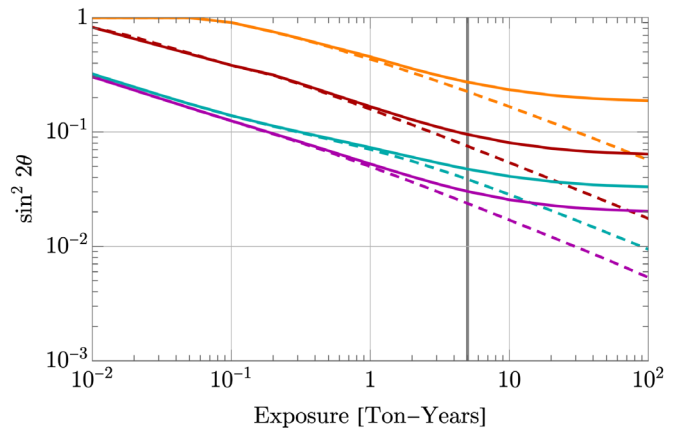

FIG. 8. The evolution of the sensitivity ($\Delta\chi^2 = 11.83$) as a function of exposure for our lithium-enhanced research reactor pseudoexperiment. Cyan and orange curves assume only ^{235}U is present for $\Delta m^2 = 1$ eV^2 and 30 eV^2, respectively; magenta and red curves introduce lithium with $f_{^8\text{Li}} = 0.1$. Dashed curves ignore shape systematics and solid curves assume a 0.5% shape uncertainty. The vertical gray line indicates a 5-ton-year exposure.

V. CONCLUSIONS

In this paper, we have studied the green-field site optimization of short-baseline reactor experiments for their sensitivity to neutrino oscillations, i.e., to $\bar{\nu}_e$ disappearance. The key to a sensitive experiment is the comparison of neutrino spectra measured at different baselines, which renders the results independent from reactor neutrino flux predictions. We perform the optimization by placing detectors at two baselines simultaneously and by finding the combination of baselines which provides the best overall sensitivity. We also find that the optimized difference of baselines is $\mathcal{O}(1)$ m and thus can, in principle, be accommodated within a single detector with position resolution, which we specifically illustrate with TAO. We look at two types of facilities: commercial reactors, with large, high-power cores and a closest approach of 10-25 m; and research reactors with compact, low-power cores and a closest approach of 3–6 m. Experiments at commercial reactors clearly outperform research reactor for $\Delta m^2 < 1$–2 eV^2 under a variety of scenarios, whereas for larger $\Delta m^2 > 2$ eV^2 experiments at research reactors do better. For all cases, getting as close as possible with the best possible energy resolution is valuable. These results are summarized in Fig. 3.

We have also investigated how the sensitivity scales with exposure and find that is saturates at around 10 ton-years even for systematics as small as 0.5%. We further extend the analysis by looking at ^{7}Li doping of a reactor to create higher-energy neutrinos. This results in enhanced sensitivity to large values of Δm^2 ($\sim$30 eV^2) even for a modest, and potentially realistic, lithium fraction, of the order 10%.

Reactor neutrino experiments can offer a very good sensitivity to oscillations from $\Delta m^2 = 0.1$–30 eV^2 with ton-scale experiments, where two or more baselines may be realized in the same detector. Sensitivities below $\sin^2 2\theta < 10^{-2}$ are possible even with reasonable systematics assumptions; pushing towards $\sin^2 2\theta < 10^{-3}$ however seems challenging. TAO has very good sensitivity but does not compare to a purpose-built experiment because of its relatively large distance to the reactor. Therefore, there is a strong case for a dedicated, short-baseline, commercial-reactor experiment with good energy resolution which currently does not exist. To cover the full range of mass-squared splittings, it is essential to use both research and commercial reactors.

APPENDIX: AVERAGING OSCILLATIONS OVER PRODUCTION REGION

If the geometry of an experiment is such that the extent of the core or detector (or both) is not small compared to the distance between them, or if the wavelength associated with a particular oscillation is smaller than either of these, then the expression for the oscillation probability in Eq. (2.1) must be flux averaged over the production and detection regions. In this case, the proper oscillation probability is formally given by

$$\langle P_{\bar{e}\bar{e}} \rangle = \frac{\int d^3\vec{x} d^3\vec{y} \frac{P_{\bar{e}\bar{e}}}{|\vec{y}-\vec{x}|^2}}{\int d^3\vec{x} d^3\vec{y} \frac{1}{|\vec{y}-\vec{x}|^2}}, \tag{A1}$$

where $\vec{x}$ and $\vec{y}$ are the integration variables over the volumes of the core and detector, respectively, and the factor of $\frac{1}{|\vec{y}-\vec{x}|^2}$ in either integral accounts for the inverse-square dependence of the flux. For the pseudoexperiments we consider, the detector is assumed pointlike; the integral over $d^3\vec{y}$ results in a trivial replacement $\vec{y} \to \vec{L}$, the displacement between the center of the core and the detector. Moreover, we approximate the core to be a sphere of radius R. Therefore, we may rewrite Eq. (A1) as

$$\langle P_{\bar{e}\bar{e}} \rangle = 1 - \sin^2 2\theta \times \frac{\int \frac{dr d\cos\phi \, r^2 \sin^2(q\sqrt{L^2+r^2+2rL\cos\phi})}{L^2+r^2+2rL\cos\phi}}{\int \frac{dr d\cos\phi \, r^2}{L^2+r^2+2rL\cos\phi}}$$
$$\equiv 1 - \sin^2 2\theta \times F(q), \tag{A2}$$

where we have employed Eq. (2.1) and abbreviated $q \equiv \frac{\Delta m^2}{4E_\nu}$. The function $F(q)$ may be evaluated analytically:

$$F(q) = \frac{1}{2} + \{4qR\sin(2qL)\cos(2qR) - 4qL\sin(2qR)\cos(2qL) - 2\sin(2qL)\sin(2qR)$$
$$+ 4q^2(L^2-R^2)[\text{Ci}(2q(L+R)) - \text{Ci}(2q(L-R))]\} \times \left\{16q^2RL - 8q^2(L^2-R^2)\ln\left(\frac{L+R}{L-R}\right)\right\}^{-1}, \tag{A3}$$

where $\text{Ci}(x)$ is the cosine integral function, defined as

$$\text{Ci}(x) = -\int_x^\infty dt \frac{\cos(t)}{t}. \tag{A4}$$

One can verify from Eq. (A3) that $F(q) \to \sin^2(qL)$ as $R \to 0$.

[1] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire, Detection of the free neutrino: A confirmation, Science **124**, 103 (1956).

[2] K. Eguchi *et al.* (KamLAND Collaboration), First Results from KamLAND: Evidence for Reactor Anti-Neutrino Disappearance, Phys. Rev. Lett. **90**, 021802 (2003).

[3] Y. Abe *et al.* (Double Chooz Collaboration), Indication of Reactor $\bar{\nu}_e$ Disappearance in the Double Chooz Experiment, Phys. Rev. Lett. **108**, 131801 (2012).

[4] F. P. An *et al.* (Daya Bay Collaboration), Observation of Electron-Antineutrino Disappearance at Daya Bay, Phys. Rev. Lett. **108**, 171803 (2012).

[5] J. K. Ahn *et al.* (RENO Collaboration), Observation of Reactor Electron Antineutrino Disappearance in the RENO Experiment, Phys. Rev. Lett. **108,** 191802 (2012).

[6] F. An *et al.* (JUNO Collaboration), Neutrino Physics with JUNO, J. Phys. G **43,** 030401 (2016).

[7] Th. A. Mueller *et al.*, Improved Predictions of Reactor Antineutrino Spectra, Phys. Rev. C **83,** 054615 (2011).

[8] P. Huber, On the determination of anti-neutrino spectra from nuclear reactors, Phys. Rev. C **84,** 024617 (2011); Erratum, Phys. Rev. C **85,** 029901 (2012).

[9] G. Mention, M. Fechner, Th. Lasserre, Th. A. Mueller, D. Lhuillier, M. Cribier, and A. Letourneau, The reactor antineutrino anomaly, Phys. Rev. D **83,** 073006 (2011).

[10] K. N. Abazajian *et al.*, Light sterile neutrinos: A white paper, arXiv:1204.5379.

[11] C. Giunti, X. P. Ji, M. Laveder, Y. F. Li, and B. R. Littlejohn, Reactor fuel fraction information on the antineutrino anomaly, J. High Energy Phys. 10 (2017) 143.

[12] M. Dentler, A. Hernández-Cabezudo, J. Kopp, M. Maltoni, and T. Schwetz, Sterile neutrinos or flux uncertainties?— Status of the reactor anti-neutrino anomaly, J. High Energy Phys. 11 (2017) 099.

[13] M. Dentler, A. Hernández-Cabezudo, J. Kopp, P. A. N. Machado, M. Maltoni, I. Martinez-Soler, and T. Schwetz, Updated global analysis of neutrino oscillations in the presence of eV-scale sterile neutrinos, J. High Energy Phys. 08 (2018) 010.

[14] 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,** 073005 (2019).

[15] J. M. Berryman and P. Huber, Sterile neutrinos and the global reactor antineutrino dataset, J. High Energy Phys. 01 (2021) 167.

[16] C. Giunti, Statistical significance of reactor antineutrino active-sterile oscillations, Phys. Rev. D **101,** 095025 (2020).

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

[18] S. Böser, C. Buck, C. Giunti, J. Lesgourgues, L. Ludhova, S. Mertens, A. Schukraft, and M. Wurm, Status of light sterile neutrino searches, Prog. Part. Nucl. Phys. **111,** 103736 (2020).

[19] Y.-J. Ko *et al.* (NEOS Collaboration), Sterile Neutrino Search at the NEOS Experiment, Phys. Rev. Lett. **118,** 121802 (2017).

[20] I. Alekseev *et al.* (DANSS Collaboration), Search for sterile neutrinos at the DANSS experiment, Phys. Lett. B **787,** 56 (2018).

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

[22] M. Andriamirado *et al.* (PROSPECT Collaboration), Improved short-baseline neutrino oscillation search and energy spectrum measurement with the PROSPECT experiment at HFIR, Phys. Rev. D **103,** 032001 (2021).

[23] H. Almazán *et al.* (STEREO Collaboration), First antineutrino energy spectrum from ²³⁵U fissions with the STEREO detector at ILL, J. Phys. G **48,** 075107 (2021).

[24] A. C. Hayes, J. L. Friar, G. T. Garvey, G. Jungman, and G. Jonkmans, Systematic Uncertainties in the Analysis of the Reactor Neutrino Anomaly, Phys. Rev. Lett. **112,** 202501 (2014).

[25] L. Hayen, J. Kostensalo, N. Severijns, and J. Suhonen, First forbidden transitions in the reactor anomaly, Phys. Rev. C **100,** 054323 (2019).

[26] M. Estienne *et al.*, Updated Summation Model: An Improved Agreement with the Daya Bay Antineutrino Fluxes, Phys. Rev. Lett. **123,** 022502 (2019).

[27] L. Hayen, J. Kostensalo, N. Severijns, and J. Suhonen, First-forbidden transitions in the reactor anomaly, Phys. Rev. C **100,** 054323 (2019).

[28] F. P. An *et al.* (Daya Bay Collaboration), Evolution of the Reactor Antineutrino Flux and Spectrum at Daya Bay, Phys. Rev. Lett. **118,** 251801 (2017).

[29] C. Giunti, Improved determination of the ²³⁵U and ²³⁹Pu reactor antineutrino cross sections per fission, Phys. Rev. D **96,** 033005 (2017).

[30] Y. Gebre, B. Littlejohn, and P. Surukuchi, Prospects for improved understanding of isotopic reactor antineutrino fluxes, Phys. Rev. D **97,** 013003 (2018).

[31] G. Bak *et al.* (RENO Collaboration), Fuel-Composition Dependent Reactor Antineutrino Yield at RENO, Phys. Rev. Lett. **122,** 232501 (2019).

[32] J. M. Berryman and P. Huber, Reevaluating reactor antineutrino anomalies with updated flux predictions, Phys. Rev. D **101,** 015008 (2020).

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

[34] Y. Abe *et al.* (Double Chooz Collaboration), Improved measurements of the neutrino mixing angle θ_{13} with the Double Chooz detector, J. High Energy Phys. 10 (2014) 086; 02 (2015) 074.

[35] F. P. An *et al.* (Daya Bay Collaboration), Measurement of the Reactor Antineutrino Flux and Spectrum at Daya Bay, Phys. Rev. Lett. **116,** 061801 (2016); Erratum, Phys. Rev. Lett. **118,** 099902 (2017).

[36] J. H. Choi *et al.* (RENO Collaboration), Observation of Energy and Baseline Dependent Reactor Antineutrino Disappearance in the RENO Experiment, Phys. Rev. Lett. **116,** 211801 (2016).

[37] A. C. Hayes, J. L. Friar, G. T. Garvey, D. Ibeling, G. Jungman, T. Kawano, and R. W. Mills, Possible origins and implications of the shoulder in reactor neutrino spectra, Phys. Rev. D **92,** 033015 (2015).

[38] A. A. Sonzogni, E. A. McCutchan, and A. C. Hayes, Dissecting Reactor Antineutrino Flux Calculations, Phys. Rev. Lett. **119,** 112501 (2017).

[39] G. Mention, M. Vivier, J. Gaffiot, T. Lasserre, A. Letourneau, and T. Materna, Reactor antineutrino shoulder explained by energy scale nonlinearities?, Phys. Lett. B **773,** 307 (2017).

[40] B. R. Littlejohn, A. Conant, D. A. Dwyer, A. Erickson, I. Gustafson, and K. Hermanek, Impact of fission neutron energies on reactor antineutrino spectra, Phys. Rev. D **97,** 073007 (2018).

[41] J. M. Berryman, V. Brdar, and P. Huber, Particle physics origin of the 5 MeV bump in the reactor antineutrino spectrum?, Phys. Rev. D **99,** 055045 (2019).

[42] M. H. Moulai, C. A. Argüelles, G. H. Collin, J. M. Conrad, A. Diaz, and M. H. Shaevitz, Combining sterile neutrino fits to short baseline data with iceCube data, Phys. Rev. D **101,** 055020 (2020).

[43] P. Adamson *et al.* (MINOS+, Daya Bay Collabortions), Improved Constraints on Sterile Neutrino Mixing from Disappearance Searches in the MINOS, MINOS+, Daya Bay, and Bugey-3 Experiments, Phys. Rev. Lett. **125,** 071801 (2020).

[44] J. M. Berryman, Constraining sterile neutrino cosmology with terrestrial oscillation experiments, Phys. Rev. D **100,** 023540 (2019).

[45] S. Gariazzo, P. F. de Salas, and S. Pastor, Thermalisation of sterile neutrinos in the early Universe in the $3 + 1$ scheme with full mixing matrix, J. Cosmol. Astropart. Phys. 07 (2019) 014.

[46] S. Hagstotz, P. F. de Salas, S. Gariazzo, M. Gerbino, M. Lattanzi, S. Vagnozzi, K. Freese, and S. Pastor, Bounds on light sterile neutrino mass and mixing from cosmology and laboratory searches, Phys. Rev. D **104,** 123524 (2021).

[47] M. Adams, F. Bezrukov, J. Elvin-Poole, J. J. Evans, P. Guzowski, B. O. Fearraigh, and S. Söldner-Rembold, Direct comparison of sterile neutrino constraints from cosmological data, ν_e disappearance data and $\nu_\mu \to \nu_e$ appearance data in a $3 + 1$ model, Eur. Phys. J. C **80,** 758 (2020).

[48] P. Anselmann *et al.* (GALLEX Collaboration), First results from the Cr-51 neutrino source experiment with the GALLEX detector, Phys. Lett. B **342,** 440 (1995).

[49] D. N. Abdurashitov *et al.*, The Russian-American Gallium Experiment (SAGE) Cr Neutrino Source Measurement, Phys. Rev. Lett. **77,** 4708 (1996).

[50] W. Hampel *et al.* (GALLEX Collaboration), Final results of the Cr-51 neutrino source experiments in GALLEX, Phys. Lett. B **420,** 114 (1998).

[51] J. N. Abdurashitov *et al.* (SAGE Collaboration), Measurement of the response of the Russian-American gallium experiment to neutrinos from a Cr-51 source, Phys. Rev. C **59,** 2246 (1999).

[52] J. N. Abdurashitov *et al.*, Measurement of the response of a Ga solar neutrino experiment to neutrinos from an Ar-37 source, Phys. Rev. C **73,** 045805 (2006).

[53] C. Grieb, J. Link, and R. S. Raghavan, Probing active to sterile neutrino oscillations in the LENS detector, Phys. Rev. D **75,** 093006 (2007).

[54] 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,** 47 (2010).

[55] G. Bellini *et al.* (Borexino Collaboration), SOX: Short distance neutrino Oscillations with BoreXino, J. High Energy Phys. 08 (2013) 038.

[56] V. V. Barinov *et al.*, Results from the Baksan experiment on sterile transitions (BEST), arXiv:2109.11482.

[57] V. Barinov and D. Gorbunov, BEST impact on sterile neutrino hypothesis, arXiv:2109.14654.

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

[59] C. Albright *et al.* (Neutrino Factory, Muon Collider Collaborations), The neutrino factory and beta beam experiments and development, arXiv:physics/0411123.

[60] S. K. Agarwalla, P. Huber, and J. M. Link, Constraining sterile neutrinos with a low energy beta-beam, J. High Energy Phys. 01 (2010) 071.

[61] L. Delgadillo and P. Huber, Sterile neutrino searches at tagged kaon beams, Phys. Rev. D **103,** 035018 (2021).

[62] J. M. Conrad, M. H. Shaevitz, I. Shimizu, J. Spitz, M. Toups, and L. Winslow, Precision $\bar{\nu}_e$-electron scattering measurements with IsoDAR to search for new physics, Phys. Rev. D **89,** 072010 (2014).

[63] D. Adey *et al.* (nuSTORM Collaboration), Light sterile neutrino sensitivity at the nuSTORM facility, Phys. Rev. D **89,** 071301 (2014).

[64] K. M. Heeger, B. R. Littlejohn, H. P. Mumm, and M. N. Tobin, Experimental parameters for a reactor antineutrino experiment at very short baselines, Phys. Rev. D **87,** 073008 (2013).

[65] K. M. Heeger, B. R. Littlejohn, and H. P. Mumm, Multiple detectors for a short-baseline neutrino oscillation search near reactors, arXiv:1307.2859.

[66] G. Bak *et al.* (RENO Collaboration), Measurement of Reactor Antineutrino Oscillation Amplitude and Frequency at RENO, Phys. Rev. Lett. **121,** 201801 (2018).

[67] D. Adey *et al.* (Daya Bay Collaboration), Measurement of the Electron Antineutrino Oscillation with 1958 Days of Operation at Daya Bay, Phys. Rev. Lett. **121,** 241805 (2018).

[68] H. De Kerret *et al.* (Double Chooz Collaboration), First Double Chooz θ_{13} measurement via total neutron capture detection, Nat. Phys. **16,** 558 (2020).

[69] M. Andriamirado *et al.* (PROSPECT Collaboration), Improved short-baseline neutrino oscillation search and energy spectrum measurement with the PROSPECT experiment at HFIR, Phys. Rev. D **103,** 032001 (2021).

[70] A. Abusleme *et al.* (JUNO Collaboration), TAO conceptual design report: A precision measurement of the reactor antineutrino spectrum with sub-percent energy resolution, arXiv:2005.08745.

[71] H. Almazán Molina *et al.* (STEREO Collaboration), Improved sterile neutrino constraints from the STEREO experiment with 179 days of reactor-on data, Phys. Rev. D **102,** 052002 (2020).

[72] M. Agostini and B. Neumair, Statistical methods for the search of sterile neutrinos, Eur. Phys. J. C **80,** 750 (2020).

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

[74] X. Qian, A. Tan, J. J. Ling, Y. Nakajima, and C. Zhang, The Gaussian CL$_s$ method for searches of new physics, Nucl. Instrum. Methods Phys. Res., Sect. A **827,** 63 (2016).

[75] G. J. Feldman and R. D. Cousins, A unified approach to the classical statistical analysis of small signals, Phys. Rev. D **57,** 3873 (1998).

[76] G. Agnolet *et al.* (MINER Collaboration), Background studies for the MINER coherent neutrino scattering reactor experiment, Nucl. Instrum. Methods Phys. Res., Sect. A **853,** 53 (2017).

[77] M. Danilov (DANSS Collaboration), Recent results of the DANSS experiment, Proc. Sci., EPS-HEP2019 (**2020**) 401 [arXiv:1911.10140].

[78] A. P. Serebrov, Present status of Neutrino-4 experiment search for sterile neutrino, https://indico.cern.ch/event/833568/contributions/3655173/attachments/1957823/3252790/2-12_China_Serebrov_Neutrino-4.pdf.

[79] M. Aker *et al.* (KATRIN Collaboration), Bound on $3 + 1$ Active-Sterile Neutrino Mixing from the First Four-Week Science Run of KATRIN, Phys. Rev. Lett. **126**, 091803 (2021).

[80] J. A. Formaggio, A. L. C. de Gouvêa, and R. G. H. Robertson, Direct measurements of neutrino mass, Phys. Rep. **914**, 1 (2021).

[81] M. Bergevin *et al.*, *Applied Antineutrino Physics 2018 Proceedings* (2019).

[82] V. I. Lyashuk, High flux lithium antineutrino source with variable hard spectrum. How to decrease the errors of the total spectrum?, arXiv:1612.08096.

[83] J. R. Alonso and K. Nakamura (IsoDAR Collaboration), IsoDAR@KamLAND: A conceptual design report for the conventional facilities, arXiv:1710.09325.

[84] A. Conant, Ph.D. thesis, Georgia Tech, 2018.