PHYSICAL REVIEW D **96,** 073001 (2017)

Global Bayesian analysis of neutrino mass data

Allen Caldwell,[*] Alexander Merle,[†] Oliver Schulz,[‡] and Maximilian Totzauer[§]

Max-Planck-Institut für Physik (Werner-Heisenberg-Institut), Föhringer Ring 6, 80805 München, Germany

(Received 9 May 2017; published 2 October 2017)

We perform a global Bayesian analysis of currently available neutrino data, putting data from oscillation experiments, neutrinoless double beta decay ($0\nu\beta\beta$), and precision cosmology on an equal footing. We evaluate the discovery potential of future $0\nu\beta\beta$ experiments and the Bayes factor of the two possible neutrino mass ordering schemes for different prior choices. We show that the indication for normal ordering is still very mild and does not strongly depend on realistic prior assumptions or different combinations of cosmological data sets. We find a wide range for $0\nu\beta\beta$ discovery potential, depending on the absolute neutrino mass scale, mass ordering and achievable background level.

DOI: 10.1103/PhysRevD.96.073001

I. INTRODUCTION

Neutrino physics is one of the most attractive fields to look for new physics, and many parameters in the neutrino sector are not yet determined. A series of oscillation experiments has established the fact that at least two neutrino masses are distinct from zero, but their smallness cannot be accommodated within the Standard Model of Particle Physics. Furthermore, quantities like the Dirac phase δ_{CP} describing the difference between matter and antimatter, the absolute neutrino mass scale, or the mass ordering (i.e., which neutrino is the lightest) are currently only poorly restricted [1]. The even more fundamental question of whether neutrinos are of Majorana or of Dirac nature (i.e., whether they are identical to their antiparticles, or not) is also still unanswered. A global analysis, combining all relevant experimental results and using the current data as efficiently as possible, is the most suitable way to address these open issues. It additionally informs upcoming experimental choices.

The first fully comprehensive Bayesian analysis of this kind is presented in this paper. We use information from oscillation experiments, precision cosmology, and neutrinoless double beta decay ($0\nu\beta\beta$). While previous works focus mostly on one of these aspects, we adopt a fully global view. Our analysis is based on the minimal framework of three light Majorana neutrinos, the most predictive setting for neutrino physics. We use global oscillation data from the `nu-fit` Collaboration [1], cosmological data from the Planck Legacy Archive (PLA),[1] and data from the $0\nu\beta\beta$-experiments KamLAND-Zen [2], EXO-200 [3], and GERDA [4]. We do not include single β decay results as they currently do not provide additional constraints.

We extract the implications for future $0\nu\beta\beta$-experiments, and also address whether current data may already exhibit a tendency towards the normal ordering scheme—which is currently under intense debate [5–8].

Bayes' Theorem provides the logical path from the probability of data under different hypotheses H_i, $P(D|H_i)$, to the probability of a hypothesis being correct given the data. It requires the explicit definition of prior probabilities. Any other approach is either incoherent (for a discussion of p-values as evidence, see e.g. [9]), requires predefined error rates (Neyman frequentist approach [10]), or contains implicit prior choices—all of which are to be disfavored. If the outcome depends strongly on the choice of prior, this is simply an indication that the data is not powerful enough to draw a conclusion. In this context, it is important to find the implications from a range of reasonable prior choices.

For our analysis we assign equal prior probabilities to the two mass orderings. For the absolute neutrino mass scale, the choice of the prior on the mass, m_{lightest}, of the lightest neutrino is critical. At present, only an upper limit is available. We consider two choices: A prior flat in m_{lightest} (this will tend to favor larger values and reflects the sizes of the measured solar and atmospheric mass square differences $\Delta m^2_{\odot} \equiv m_2^2 - m_1^2$ and $\Delta m^2_A \equiv |m_3^2 - m_1^2|$) and a scale-free prior (flat in the logarithm of m_{lightest}, which tends to favor small values and reflects the fact that the absolute neutrino mass scale is still unknown). We then evaluate the consequences of both prior choices on the posterior probabilities.

Technical details regarding priors and likelihoods are provided in the Appendix.

II. PARAMETER SPACE, LIKELIHOODS, PRIORS

A. Parameter space

Our analysis depends on eight parameters, aggregated into the vector $\boldsymbol{\theta}$,

[*]caldwell@mpp.mpg.de
[†]amerle@mpp.mpg.de
[‡]oschulz@mpp.mpg.de
[§]totzauer@mpp.mpg.de

[1]Based on observations by Planck (http://www.esa.int/Planck).

$$\boldsymbol{\theta} = (m_{\text{lightest}}, \Delta m^2_{\odot}, \Delta m^2_A, s^2_{12}, s^2_{13}, \alpha_1, \alpha_2, \mathcal{G}). \quad (1)$$

Here, m_{lightest} is the smallest neutrino mass eigenvalue, while s^2_{ij} denotes the sine squared of the mixing angle θ_{ij}. The Majorana phases α_1 and α_2 take trivial values if neutrinos are of Dirac nature. The nuclear matrix elements (NMEs) required to calculate rates of $0\nu\beta\beta$ for different considered isotopes are condensed into $\mathcal{G}$. Note that $\boldsymbol{\theta}$ does *not* contain the Dirac phase δ_{CP} or s^2_{23}, as they only affect neutrino oscillations, but neither $0\nu\beta\beta$ nor cosmology. Furthermore we allow for the choice of either normal (NO: $m_1 < m_2 < m_3$) or inverted (IO: $m_3 < m_1 < m_2$) ordering.[2]

B. Likelihood

The global likelihood function $\mathcal{L}_{\text{glob}}$ is a function that depends nontrivially on all eight parameters. Since every suite of experiment constitutes an independent data set, we can factorize the likelihood,

$$\mathcal{L}_{\text{glob}} = \mathcal{L}_{\text{cosmo}} \times \mathcal{L}_{0\nu\beta\beta} \times \mathcal{L}_{\text{osc}}. \quad (2)$$

The cosmological term $\mathcal{L}_{\text{cosmo}}$ only depends on the sum of neutrino masses,

$$\Sigma \equiv m_1 + m_2 + m_3, \quad (3)$$

and hence on m_{lightest}, $\Delta m^2_{\odot}$, and Δm^2_A. We reconstructed the likelihood from the Markov chains that are publicly available on the PLA. As those chains were sampled on a prior flat in Σ, one can construct a function of Σ that is directly proportional to $\mathcal{L}_{\text{cosmo}}$ from these chains.[3] To be agnostic about the choice of data sets used in our analysis, we chose two models: both contain the *TT* correlation of the cosmic microwave background (CMB), a joint likelihood on *TT*, *EE*, *BB*, and *TE* correlations[4] for low multipoles ($2 \leq l \leq 29$), as well as weak gravitational lensing data. We refer to this combination of data as *conservative model*. Augmenting the data by data from baryonic accoustic oscillations (BAO) defines a *restrictive model*, which puts tighter constraints onto Σ.[5]

The term $\mathcal{L}_{0\nu\beta\beta}$ is factorized into three terms for GERDA, KamLAND-Zen, and EXO, respectively. Each term is well parametrized by the form

$$\mathcal{L}_{0\nu\beta\beta,\text{i}} = \exp\left[-(a_{0,i} + a_{1,i}/T_{1/2} + a_{2,i}/T^2_{1/2})\right], \quad (4)$$

where $T_{1/2}$ is the neutrinoless double beta decay half-life in units of 10^{25} a. The parameters entering this formula have either been determined directly from published data or confirmed by the experimental collaborations.

The neutrino oscillation data is adopted from the global oscillation fit provided by `nu-fit.org` [1]. This collaboration does not provide a public likelihood function depending on all oscillation parameters. We thus approximate the likelihood using a spline-fit of the one-dimensional projections of their $\Delta\chi^2$ fit onto the relevant parameters, shifting the distributions so they have a minimum at $\Delta\chi^2 = 0$, and taking the likelihood to be a product of the resulting four factors. Each such factor is given by

$$\exp\left(-\frac{\Delta\chi^2(\theta_i)}{2}\right), \quad (5)$$

where $\chi^2(\theta_i)$ is the one dimensional projection onto the parameter θ_i. To account for the slight tendency of oscillation data to favor NO, we then added *one* offset $\Delta\chi^2 = 0.83$ [1] to the product value for IO after recalibrating the projections. This procedure avoids counting the slight inclination towards NO each time it gets projected onto one parameter. The quoted value of $\Delta\chi^2$ corresponds to a Bayes factor of 1.5 in favor of NO.[6]

C. Priors

Our choice of priors is intended to be informative whenever possible, and otherwise agnostic for parameters with little or no information available. For mass square differences and mixing angles, we choose flat priors extending well beyond the best-fit values for both orderings, since the likelihood functions severely constrain their values in any case. The choice of range does not influence our results. For the Majorana phases α_1 and α_2, we choose priors flat on the entire range $[0, 2\pi]$, as we do not have any prior knowledge or theoretical motivation for an informative distribution. The prior on the NME factor $\mathcal{G}$ of a certain isotope is discrete, equally weighing the different theoretical calculations, which deviate from one another by factors of up to 3. This procedure will be further explained and justified in the Appendix (cf. Fig. 4 and subsequent discussion).

[2]Such an "external" parameter is called *hyperparameter*.

[3]Note that the choice of prior used by the Planck Collaboration is, in fact, unphysical, since Σ has a minimal value $\Sigma_{\text{min}} \neq 0$. As shown in [5], using a more realistic prior, the resulting constraints become stronger, but they constrain $(\Sigma - \Sigma_{\text{min}})$, making the corresponding limit on Σ appear looser.

[4]Here, *T* refers to the temperature of the CMB, while *E* and *B* denote the basic patterns of its polarization.

[5]We use the PLA data sets base_mnu_plikHM_TT_lowTEB_lensing (conservative) and base_mnu_plikHM_TT_lowTEB_lensing_BAO (restrictive).

[6]In future analyses, this procedure could be refined by using the full multiparameter likelihoods or by directly fitting the oscillation data.

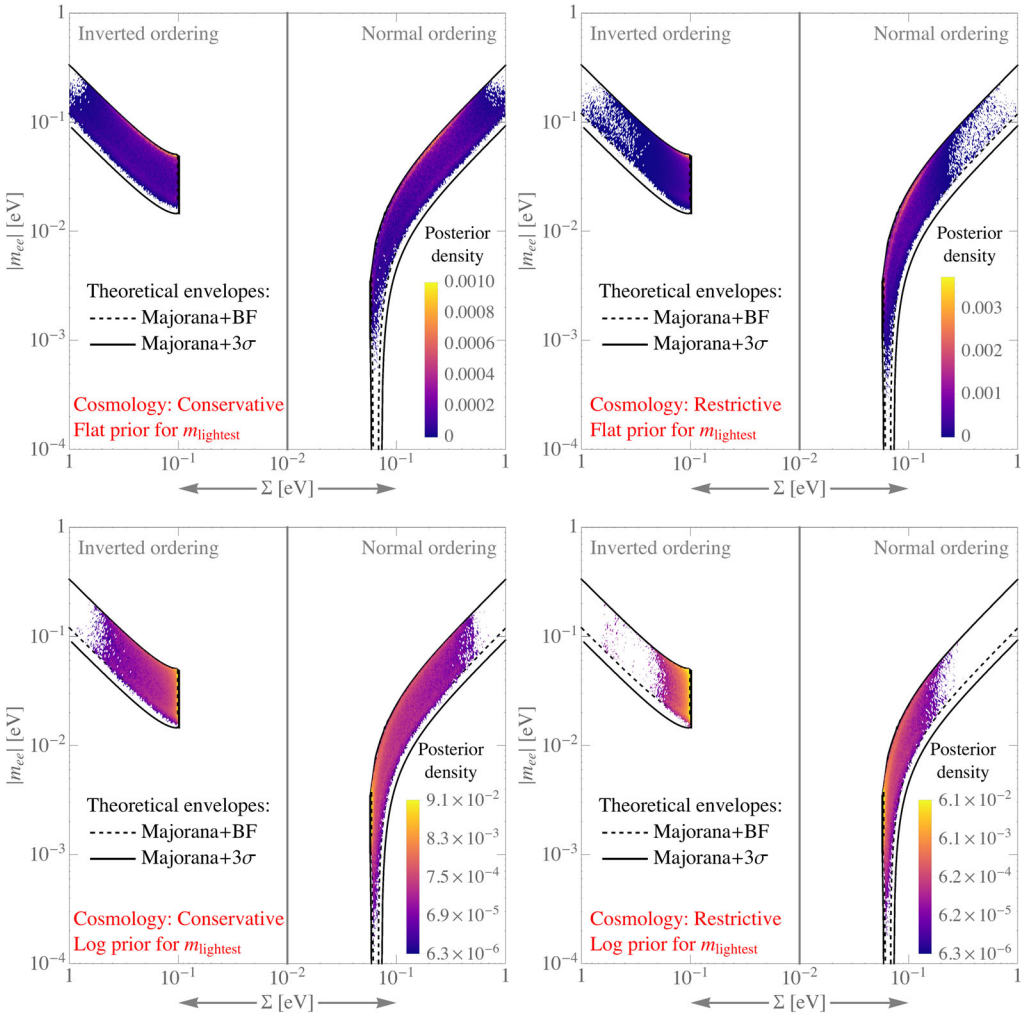

FIG. 1. Heat map of posterior probability density for both combinations of cosmological data sets and choices of the prior on m_{lightest}. The upper panels depict the flat prior while the lower panels show the log prior. Note that the color coding also changes from linear to logarithmic. See text for definitions of the different quantities.

We choose two very different priors for m_{lightest}, to explicitly show how results depend on the respective *assumption*. Both choices span the range $[10^{-7}\text{eV}, 0.6\text{ eV}]$.[7] The flat prior allocates 90% of the probability mass at values greater than 60 meV, while the scale-free log prior allocates about 85% in a region where $m_{\text{lightest}} < 60$ meV.

[7]The upper end of our range is motivated by the restrictive power of the cosmological data sets: Extending it to larger values does not alter the results significantly. The lower end of the range can be argued to be as low as 10^{-13} eV [11], using a general theoretical argument from perturbation theory. As we are interested in implications for $0\nu\beta\beta$, there will however be a point where lowering the minimal value does not anymore alter any implications. In order to back up this statement numerically, we also define a second logarithmic prior on m_{lightest}, spanning the range $[10^{-13}\text{eV}, 0.6\text{ eV}]$ and denoted by $\log^*$. We will compare the numerical results of this prior choice to the somewhat restricted log prior where appropriate.

III. RESULTS OF OUR ANALYSIS

For all combinations of neutrino mass ordering, cosmological data set, and prior for m_{lightest}, we used the BAT Markov chain Monte Carlo code in version `v1.0.0-RC1` [12] to sample from the posterior probability density. For each scenario, we ran four Markov chains with a length of 4×10^6 samples each. The results derived from the posterior samples are presented in Figs. 1, 2, and 3.

A. Posterior probabilities

For fixed mass ordering, cosmology, and prior on m_{lightest}, we show the *posterior probability density* in the plane of the effective mass $|m_{ee}|$ controlling the $0\nu\beta\beta$ rate and the sum Σ of neutrino masses in Fig. 1. We combine the NO and the IO cases into subplots in a "back-to-back" manner. We also show theoretical envelopes bounding the regions in the Σ–$|m_{ee}|$ parameter space if

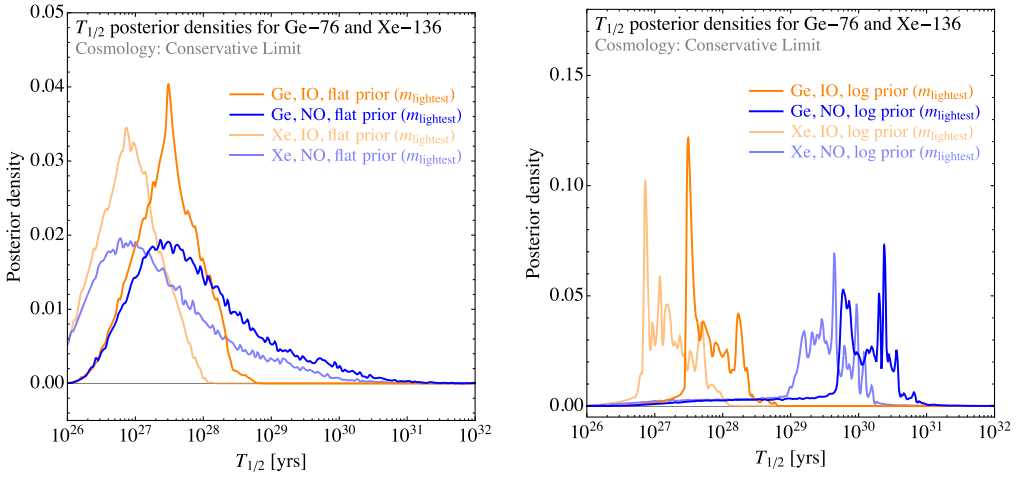

FIG. 2. Probability distribution for $T_{1/2}$ for both ^{76}Ge and ^{136}Xe (conservative cosmology). The distributions on the left feature a prior flat in $m_{\rm lightest}$, while the ones on the right feature a prior flat in $\log(m_{\rm lightest})$.

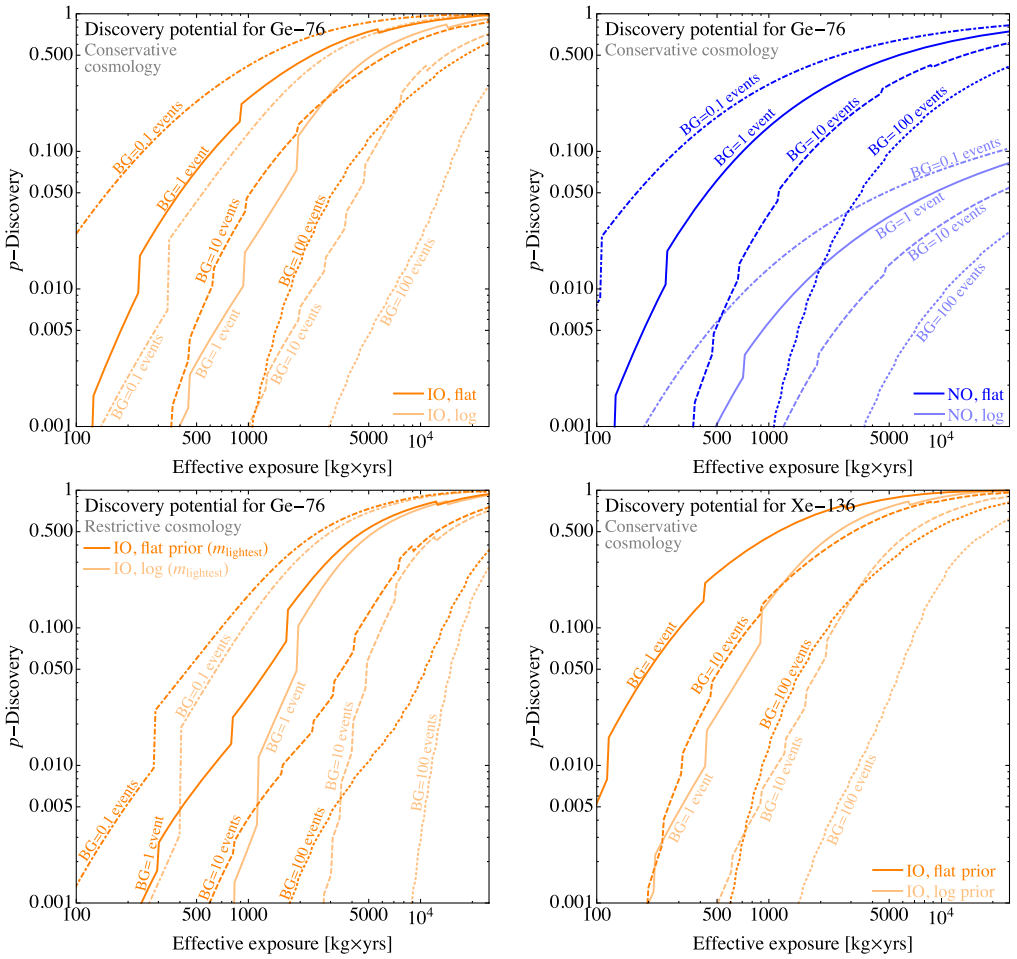

FIG. 3. Discovery potential as a function of effective exposure $E\epsilon$ for different background levels. The upper left panel defines a *benchmark case* (germanium, IO, conservative cosmology), while for the other panels one of these parameters is changed: upper right—NO; lower left—restrictive cosmology; lower right—^{136}Xe. The kinks in the curves are a consequence of the integer nature of Poissonian statistics.

the Majorana phases are varied within their full ranges while the relevant oscillations parameters are kept at their best fit values (dashed lines) or varied within their 3σ ranges (solid lines).

Choosing the restrictive cosmological data sets forces Σ (and by means of oscillation data also m_{lightest}) to be smaller, resulting in more probability mass shifted towards smaller values of $|m_{ee}|$. Another clear feature is the fact that a scale-free prior on m_{lightest} leaves almost no probability mass in regions where $\Sigma \gtrsim 0.1$ eV. Finally, we note that the highest probability density tends to be at the *upper* ends of the allowed $|m_{ee}|$ ranges for a given Σ. This reflects the flat prior assumption on the Majorana phases, which makes it highly unlikely to get a cancellation of terms leading to vanishing $|m_{ee}|$ [13–15].

B. Double beta decay

The posterior samples from our global fits can be used to infer the discovery potential for $0\nu\beta\beta$ experiments. The relevant quantity is the half-life probability distribution, examples of which are shown in Fig. 2.[8] As expected, the probability mass at smaller half-life is always bigger in the case of IO. However, for the log prior (right panel), favoring small m_{lightest}, the effect is much more pronounced than in the case of the flat prior (left panel). Additionally, the log prior leads to a spikier structure of the probability distribution, caused by the discrete set of values for $\mathcal{G}$.

The relationship between the signal expectation and the half-life is

$$\nu = \frac{N_A \ln 2}{m_{\mathrm{enr}}} \frac{E\epsilon}{T_{1/2}}, \tag{6}$$

where N_A is Avogadro's number and m_{enr} is the molar mass of the relevant enriched isotope. E is the exposure and ϵ is the efficiency to find a signal in the region of interest. The latter includes active/fiducial mass considerations as well as signal reconstruction efficiencies.

Selected results for the discovery potential for $0\nu\beta\beta$, assuming light Majorana neutrinos as the source of a signal, are presented as a function of effective exposure $E\epsilon$ for four different background levels $(\mathrm{BG} = 0.1, 1, 10, 100$ events) for ^{76}Ge and three for ^{136}Xe $(\mathrm{BG} = 1, 10, 100)$[9] in Fig. 3. The discovery potential can be used to judge the merits of individual experiments. In general, we see that a prior flat in m_{lightest} gives a much larger probability for discovery than a prior flat in $\log(m_{\mathrm{lightest}})$. The importance of keeping the background small is obvious from these plots.

[8]The full set of plots is available in the Appendix.

[9]Discovery is defined as a Bayes factor of at least 100 in favor of the presence of $0\nu\beta\beta$. The details of the statistical analysis are available in the Appendix.

TABLE I. Bayes factors of posterior odds for NO vs IO, for all combinations of cosmologies and priors. For completeness, we show the negligible changes when allowing the log prior to extend down to 10^{-13} eV. The slight change for the conservative cosmology appears larger than it actually when rounding to two significant digits.

Cosmology	Prior on m_{lightest}	Bayes factor NO/IO
restrictive	flat	2.5
conservative	flat	1.7
restrictive	log	2.0
conservative	log	1.6
restrictive	log*	2.0
conservative	log*	1.7

C. Mass ordering

As noted earlier, the oscillation data give a Bayes factor of 1.5 in favor of the normal mass ordering. We have evaluated the Bayes factor that results from including the data of cosmology and $0\nu\beta\beta$ into the analysis for the different cosmological data sets used as well as for our two choices of priors on m_{lightest}. All evidence integrals were computed using the `Cuhre` quadrature algorithm provided by the `CUBA` library [16]. The results are shown in Table I. As can be seen, the additional data and our choices of priors do not change the conclusions significantly. Also the choice of the cut-off scale at 10^{-7} eV is perfectly reasonable, as lowering it does not change the numerical results significantly anymore.

IV. CONCLUSIONS

In this work, we have conducted a global Bayesian analysis of neutrino mass parameters within the minimal framework of three light Majorana neutrinos, combining data from oscillation experiments, $0\nu\beta\beta$ decay, and precision cosmology. Working with one prior flat in m_{lightest} and another one flat in $\log(m_{\mathrm{lightest}})$, we have investigated both extreme cases of uninformative priors for the critical parameter m_{lightest}. Combining these priors with a more conservative cosmological data set on one hand and a more restrictive one on the other hand, we conclude that the posterior probability for NO is still very mild, even in the extreme case of a restrictive cosmology and a flat prior on m_{lightest}. In the other cases, the slight inclination towards NO is almost entirely driven by neutrino oscillation experiments.

Furthermore we have evaluated the posterior distributions of $T_{1/2}$ in the different combinations of priors and cosmological data sets as well as for different isotopes. This allowed us to infer the discovery potential of different experimental approaches for $0\nu\beta\beta$. Depending on the neutrino mass ordering, the achievable effective exposure, and the background level, the discovery potential spans a wide range. Assuming a flat prior for m_{lightest} is in all cases

favorable for $0\nu\beta\beta$ searches as is the inverted mass ordering.

Our approach of putting all relevant experimental insights on a consistent and equal footing helps to pave the way for future comprehensive analyses of neutrino data. Additional data can be easily accommodated and probabilities updated in this scheme.

ACKNOWLEDGMENTS

We would like to thank Matteo Costanzi, Matteo Viel, and in particular, Massimiliano Lattanzi for valuable discussions concerning cosmology. Furthermore, we are grateful to Itaru Shimizu and Luciano Pandola for checking our likelihood parametrizations for KamLAND-Zen and GERDA, respectively. We thank Frederik Beaujean for advice regarding the implementation of our model in BAT. AM acknowledges partial support by the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie Grant Agreements No. 690575 (InvisiblesPlus RISE) and No. 674896 (Elusives International Training Network), as well as by the Micron Technology Foundation, Inc. MT acknowledges support by the IMPRS-EPP and by the Studienstiftung des deutschen Volkes.

Note added.—A few days after this manuscript appeared on the arXiv, a similar Bayesian analysis by Agostini *et al.* [17] also became available on the arXiv. While their work is qualitatively similar to ours in some respect, it also differs in a few key aspects: their analysis does not take cosmological data into account and assumes a flat prior on the sum of neutrino masses. This is qualitatively similar to our case of a flat prior for m_{lightest}. For these reasons, their discovery potential is in general somewhat more optimistic than ours. Furthermore, their treatment of uncertainties in the NMEs is qualitatively different from ours using a discrete prior on the different theoretical predictions.

APPENDIX TECHNICAL DETAILS

This supplemental material describes how we obtained the likelihood functions, it illustrates the procedure we used to determine discovery potentials, and it also presents our main result (namely the posterior probability densities) in an alternative coordinate system, which only uses observables accessible in laboratory experiments.

1. Statistical formulation

The analysis was performed using Bayes' Theorem,

$$P(H|D,I) = \frac{P(D|H,I)P_0(H|I)}{P(D|I)}, \tag{A1}$$

where H is the hypothesis to which we are assigning a probability, D represents the data being used to update the probability, and I is the additional information and constraints that have been used to define the probabilities. $P_0(H|I)$ is the prior probability assignment to H. The denominator can be rewritten using the law of total probability as

$$P(D|I) = \sum_i \int d\boldsymbol{\theta} P(D|M_i,\boldsymbol{\theta},I)P(\boldsymbol{\theta}|M_i,I)P(M_i|I), \tag{A2}$$

where M_i represent different models allowed under I subject to $\sum_i P(M_i|I) = 1$ and $\boldsymbol{\theta}$ are the (possibly different) parameters present in the model. $P(\boldsymbol{\theta}|M_i,I)$ is also normalized.

In our analysis, we assume that there are three light Majorana neutrinos and have two models representing the two allowed mass ordering schemes: normal (NO: $m_1 < m_2 < m_3$) or inverted (IO: $m_3 < m_1 < m_2$). We assign equal prior probabilities to the two mass orderings: $P(M_1 = \text{NO}|I) = P(M_2 = \text{IO}|I) = 1/2$. The parameters used are

$$\boldsymbol{\theta} = (m_{\text{lightest}}, \Delta m^2_{\odot}, \Delta m^2_A, s^2_{12}, s^2_{13}, \alpha_1, \alpha_2, \mathcal{G}), \tag{A3}$$

where m_{lightest} is the smallest neutrino mass eigenvalue and s^2_{ij} denotes the sine squared of the mixing angle θ_{ij}. The Majorana phases α_1 and α_2 take trivial values if neutrinos are of Dirac nature. The nuclear matrix elements (NMEs) required to calculate rates for $0\nu\beta\beta$ are collectively denoted by $\mathcal{G}$. Note that $\boldsymbol{\theta}$ does *not* contain the Dirac phase δ_{CP} or s^2_{23}, as they only affect neutrinos oscillations but not $0\nu\beta\beta$ or cosmology. We assign flat probabilities to all parameters except for m_{lightest} and $\mathcal{G}$. The latter are discussed in more detail below. For the parameters with flat priors, the ranges were defined either to be the maximal range (for the phases) or to cover a sufficiently wide range so as not to affect the results. For example, in the case of the mixing angles and squared mass differences, the parameters are so strongly constrained by the oscillation data that values even remotely close to the edges are already strongly disfavored.

We consider two prior probability assignments for m_{lightest}: A prior flat in m_{lightest} (this will tend to favor larger values, and it reflects the sizes of the measured mass square differences $\Delta m^2_{\odot} \equiv m_2^2 - m_1^2$ and $\Delta m^2_A \equiv |m_3^2 - m_1^2|$) and a scale-free prior (flat in the logarithm of m_{lightest}, which tends to favor small values and reflects the fact that the absolute neutrino mass scale is still unknown). We note that

$$P(\log(m_{\text{lightest}})) = \text{const} \Leftrightarrow P(m_{\text{lightest}}) \propto 1/m_{\text{lightest}}.$$

The prior probability for $\mathcal{G}$ was defined by assigning equal probabilities of $1/9$ to the different matrix element computations, the results of which are collected in Table II,

TABLE II. NMEs as extracted from the references given. See text for details.

Method [reference]	Value for ^{76}Ge	Value for ^{136}Xe
Energy density functional method [25]	4.60	4.20
Interacting shell model [21]	2.30	1.77
Microscopic interacting boson model [26]	5.47	3.36
Proton-neutron QRPA [27]	5.52	3.35
Self-consistent renormalized QRPA (Bonn potential) [23]	5.82	3.36
Self-consistent renormalized QRPA (Argonne V18 potential) [23]	5.44	2.75
QRPA (Bonn potential) [22]	5.40	2.38
QRPA (Argonne V18 potential) [22]	5.00	2.11
Deformed self-consistent Skyrme QRPA [28]	5.53	1.68

using the normalization corresponding to the following half-life formula [18]:

$$\frac{1}{T^{0\nu}_{1/2}} = G_{0\nu}|\mathcal{M}_{0\nu}|^2\left(\frac{|m_{ee}|}{m_e}\right)^2, \tag{A4}$$

where $G_{0\nu}$ is a phase space factor which can be easily computed for any isotope under consideration (we make use of the values from Ref. [19], which were slightly updated by Ref. [20]). $\mathcal{M}_{0\nu}$ is the NME, which encodes all nuclear physics that goes into the process. Some remarks

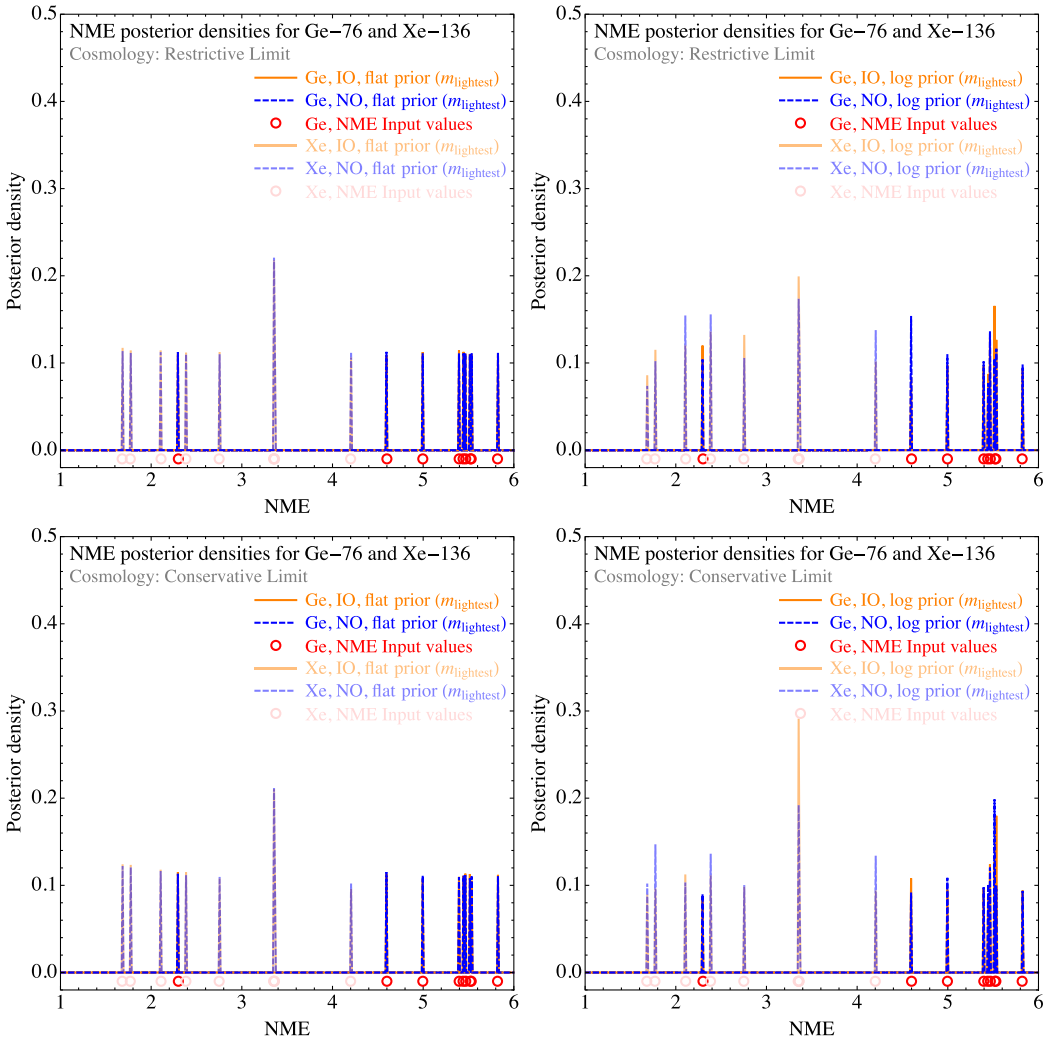

FIG. 4. Posterior probability distributions of the NMEs, sorted into 500 bins, for both ^{76}Ge and ^{136}Xe, along with the input values. As can be seen, the posterior probabilities basically reproduce the input, i.e., the current constraints are not sufficiently strong to impact NME computations.

are at order concerning the values shown in Table II. For definiteness, we have adopted the standard value $g_A = 1.25$ for the axial vector coupling, which means that we had to rescale some of the NMEs [21,22]. Also the phase space factors used [19,20] correspond to $g_A = 1.25$. In cases where different versions of a computations are available, we have for definiteness always chosen the most optimistic result. For example, Ref. [23] reports the results for both, an intermediate size model space and a large size single particle space, the latter of which tends to yield larger values of the NME. Hence, we have decided to use the large size results. This differs from the choices e.g. made in Ref. [24], where in some cases the smaller and in others the larger value has been chosen. Nevertheless, none of the treatments is wrong in the sense that at the moment the values still suffer from nuclear physics uncertainties, and there exists no way to decide which value is closer to reality.

The NMEs are shown together with their posterior probabilities in Fig. 4. While we only show results for ^{76}Ge and ^{136}Xe, it is straightforward to extend the analysis to any isotope for which NMEs have been calculated. Note that the posterior distribution for the NMEs tracks the prior distribution quite closely. This is to be expected, as the data from the $0\nu\beta\beta$ experiments does not yet hint towards a signal but only gives lower bounds on the lifetimes of the isotopes. Thus, no insights about the NMEs mediating the decays can be generated, and our prior knowledge does not get updated in a relevant way. Due to this argument, we can also justify taking a prior distribution sharply centered around discrete values. Even if each calculation were equipped with an uncertainty, which could be accounted for by modeling the prior by a sum of narrow Gaussians, this would not change the results significantly. As the spread between the different calculations is presumably much larger than this uncertainty, the overall picture would stay unaltered. The choice of assigning a equal normalization of N_{calc}^{-1} to each calculation just formalizes our current state of ignorance concerning the best method—a state that can only be changed by observation of $0\nu\beta\beta$ decay.

The probability of the data, $P(D|M_i, \boldsymbol{\theta}, I)$, viewed as function of the parameters with the data fixed, is known as the likelihood $\mathcal{L}(\boldsymbol{\theta})$. In the following, we describe which prescriptions were used to define these likelihoods.

a. Neutrino oscillation data

The neutrino oscillation likelihood is taken from `nu-fit.org` [1], of which we have used `v3.0` of their neutrino oscillation global fit results. As described in the main text, we have converted the $\Delta\chi^2$ projections from `nu-fit.org` into likelihood functions, $\mathcal{L}(\theta_i) \propto \exp[-\Delta\chi^2(\theta_i)/2]$, where $\Delta\chi^2(\theta_i)$ is the $\Delta\chi^2$ projection for the parameter θ_i. The likelihoods are shown in Fig. 5. Note that the fit by `nu-fit.org` very slightly disfavors IO, by an offset $\Delta\chi^2 = 0.83$. However, while this value appears in each of the projections shown by that collaboration, we have to be careful not to penalize IO more than once in our analysis, because the value of 0.83 also appears only once in the full χ^2-function, before projecting out individual parameters. We have accounted for this by adding the offset only *once*, to avoid an artificial worsening of IO which would even increase with the number of parameters used. Note that this results in a global factor in front of the likelihood function for IO, which drops out when computing posterior probabilities. However, once we compute a Bayes factor, both likelihoods need to share a common normalization to make sense of the values obtained, and in that case this constant prefactor does play a role, and it corresponds to the factor of 1.5 in favor of NO, as mentioned in the main text.

b. Double beta decay data

The most strongly constraining experiments to date are the GERDA, KamLAND-Zen, and EXO experiments, and

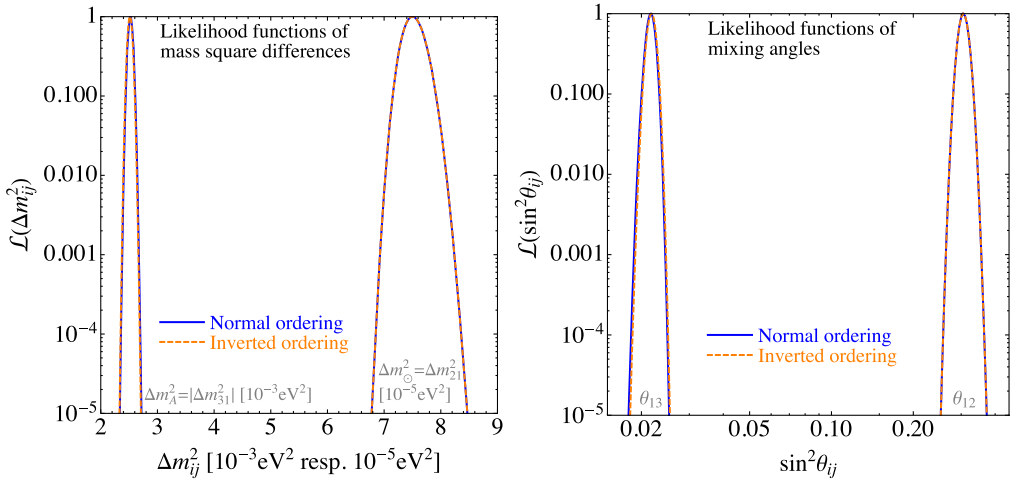

FIG. 5. Likelihood functions for the squared mass differences (left) and for the relevant mixing angles (right).

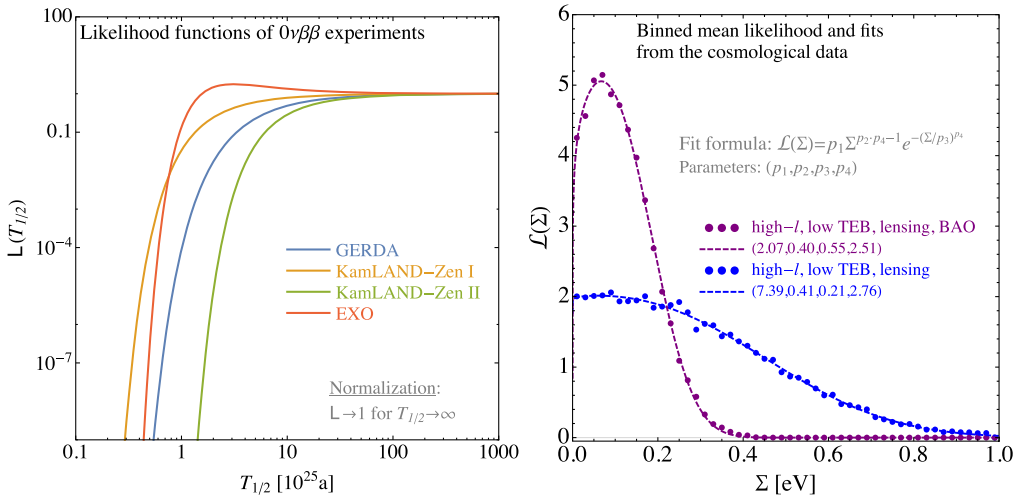

FIG. 6. Likelihood functions for the different $0\nu\beta\beta$ experiments (left) and cosmological data sets (right).

we have included likelihood terms for these experiments in our analysis. The data were analyzed and presented in different ways by the three experiments. We discuss these in turn to explain how we included these data into our Bayesian analysis. Note that $T_{1/2}$ values are always given in units of 10^{25} years. The functions we will present in the following are plotted in the left panel of Fig. 6 for illustration.

Gerda

The GERDA Collaboration [29] performed both a Bayesian and a frequentist limit setting analysis on their data. We use the posterior probability density for the inverse of the half-life of ^{76}Ge as determined by the GERDA Collaboration, which is well parametrized by the form[10]

$$\mathcal{P}(1/T^{\mathrm{Ge}}_{1/2}) \propto \exp\left[-\frac{1}{2}\frac{(1/T^{\mathrm{Ge}}_{1/2}+1.48)^2}{0.461^2}\right]. \quad \text{(A5)}$$

Since GERDA used a flat prior for $1/T^{\mathrm{Ge}}_{1/2}$, we use

$$\mathcal{L}_{\mathrm{GERDA}}(T^{\mathrm{Ge}}_{1/2}) \propto \exp\left[-\frac{1}{2}\frac{(1/T^{\mathrm{Ge}}_{1/2}+1.48)^2}{0.461^2}\right]. \quad \text{(A6)}$$

KamLAND-Zen

The KamLAND-Zen Collaboration [30] performed frequentist analyses based on a profile likelihood test statistic. They presented their results in the form of $\Delta\chi^2$ as a function of $T^{\mathrm{Xe}}_{1/2}$ (Wilk's Theorem was assumed to apply). We have interpreted these $\Delta\chi^2$ as likelihoods for the half-life of ^{136}Xe by setting

$$\mathcal{L}_{\text{KamLAND-Zen}}(T^{\mathrm{Xe}}_{1/2}) \propto \exp\left[-\frac{1}{2}\Delta\chi^2(T^{\mathrm{Xe}}_{1/2})\right]. \quad \text{(A7)}$$

The phase-I and phase-II results were used separately, and the following parametrizations were found to represent the KamLAND-Zen results well[11]:

$$\Delta\chi^2 = \begin{cases} 2[2.3/T^{\mathrm{Xe}}_{1/2} + 1.09/(T^{\mathrm{Xe}}_{1/2})^2] & \text{phase-I,} \\ 2[9.71/T^{\mathrm{Xe}}_{1/2} + 28.1/(T^{\mathrm{Xe}}_{1/2})^2] & \text{phase-II.} \end{cases} \quad \text{(A8)}$$

EXO

The EXO Collaboration also performed a frequentist analysis and presented confidence level limits on $T^{\mathrm{Xe}}_{1/2}$. However, the background expectation, background uncertainty, and observed numbers of events were presented in the paper [31], allowing for the construction of the likelihood as

$$\mathcal{L}_{\mathrm{EXO}}(T^{\mathrm{Xe}}_{1/2}) \propto \int e^{-(\nu+\lambda)}\frac{(\nu+\lambda)^n}{n!} \times \mathcal{G}(\lambda;\lambda_0=31.1,\sigma_\lambda=3.3)d\lambda, \quad \text{(A9)}$$

where the signal expectation ν depends on $T^{\mathrm{Xe}}_{1/2}$ as well as on the exposure and the efficiency, which are both given in Ref. [31]. The background level is given by λ and is centered on λ_0. The uncertainty is introduced by smearing the expectation with a Gaussian of width σ_λ. The observed number of events is $n = 39$. The resulting likelihood is well parametrized by the form

[10]The parametrizations have been cross-checked by the collaboration [L. Pandola (private communication)].

[11]The parametrizations have been cross-checked by the collaboration [I. Shimizu (private communication)].

TABLE III. Fit parameters for the cosmological likelihood factor parametrized by Eq. (A11) for both data sets used in our analysis.

Level	Data set	p_1	p_2	p_3	p_4
Conservative	base_mnu_plikHM_TT_lowTEB_lensing	2.073	0.401	0.552	2.514
Restrictive	base_mnu_plikHM_TT_lowTEB_lensing_BAO	7.385	0.407	0.207	2.765

$$\mathcal{L}_{\mathrm{EXO}}(T^{\mathrm{Xe}}_{1/2}) \propto \exp\left[-\frac{1}{2}\frac{(1/T^{\mathrm{Xe}}_{1/2} - 0.32)^2}{0.30^2}\right]. \quad \text{(A10)}$$

We note that the peak is at finite $T^{\mathrm{Xe}}_{1/2}$ since more events were observed than expected from background processes.

c. Cosmological data sets

The cosmological factor $\mathcal{L}_{\mathrm{cosmo}}$ contains extensive information about cosmic structure formation and cannot be given in a simple closed form. Instead, we have used Markov chains sampling the posterior for different combinations of data sets that are publicly available on the Planck Legacy Archive (PLA).[12] These posterior samples were obtained from a uniform prior in Σ, such that the posterior is identical to the likelihood except for some unknown normalization that is irrelevant in our implementation. The total number of samples for each model is about 50×10^3. The samples were binned into 50 bins to obtain a likelihood estimator, which was then fit to the following four-parameter template:

$$\mathcal{L}_{\mathrm{cosmo}}(\Sigma) = p_1 \Sigma^{p_2 p_3 - 1} \exp\left[-\left(\frac{\Sigma}{p_3}\right)^{p_4}\right]. \quad \text{(A11)}$$

This functional form is related to the generalized Gamma distribution, and it is a good empirical fit to likelihood, as the right panel of Fig. 6 shows. Table III lists the fit coefficients we obtained for the two cosmological data sets used in our analysis.

2. Double beta decay: Discovery probability analysis

We define two hypotheses:

H_0 There are only background processes producing the data;

H_1 There is, additionally to the background processes, also a signal from neutrinoless double beta decay.

The full information is contained in the posterior probability for the hypothesis given the data, $P(H_1|D)$. If this probability is large enough, then a "discovery" can be claimed. We will use the related concept of posterior odds

$$\mathcal{O}_1 = \frac{P(H_1|D)}{P(H_0|D)}$$

and place requirements on this quantity to claim a discovery. Assuming the two hypotheses are exhaustive[13] we have

$$P(H_0|D) = \frac{P(D|H_0)P_0(H_0)}{P(D|H_1)P_0(H_1) + P(D|H_0)P_0(H_0)},$$
$$P(H_1|D) = \frac{P(D|H_1)P_0(H_1)}{P(D|H_1)P_0(H_1) + P(D|H_0)P_0(H_0)}.$$

The posterior odds are then given by

$$\mathcal{O}_1 = \frac{P(D|H_1)}{P(D|H_0)}\mathcal{O}_0,$$

where

$$\mathcal{O}_0 = \frac{P_0(H_1)}{P_0(H_0)}$$

are the prior odds. The factor $\frac{P(D|H_1)}{P(D|H_0)}$ is known as the "Bayes factor".

We consider the case of a single experiment measuring n events in a given region of interest defined by the Q value of the decay, the background level, and the energy resolution. We use simple Poisson statistics to evaluate the discovery potential. Small improvements are possible by making use of the background and signal spectral shapes, but the additional sensitivity gained by doing this is not significant for our discussion. We will also ignore uncertainties in the background and assume it is well-known. The uncertainty in the background typically has little effect on the discovery calculation if the background is small, but it can become important if this is not the case.

We use the symbol λ to represent the background expectation in the region of interest, and ν is the signal expectation given a half-life for the decay, the exposure, and the detection efficiency.

Given these definitions, we have

$$P(D|H_1) = P(n|H_1) = \int_0^\infty P(n|\lambda + \nu')P(\nu')d\nu', \quad \text{(A12)}$$

[12]Based on observations by Planck (http://www.esa.int/Planck).

[13]We exclude here the possibility that there are background sources not included in the background model expectations. We also assume that the signal would come from the exchange of a light Majorana neutrino.

where $P(\nu')$ is the probability to have the signal expectation ν' given the experimental conditions and prior knowledge on $T_{1/2}$. The probability of the data for H_0 is simply

$$P(D|H_0) = P(n|H_0) = e^{-\lambda}\frac{\lambda^n}{n!}. \tag{A13}$$

The relationship between the signal expectation and the half-life is

$$\nu = \frac{N_A \ln 2}{m_{\rm enr}}\frac{E\epsilon}{T_{1/2}}, \tag{A14}$$

where N_A is Avogadro's number and $m_{\rm enr}$ is the molar mass of the relevant enriched isotope. E is the exposure and ϵ is the efficiency to find a signal in the region of interest. The latter includes active/fiducial mass considerations as well as signal reconstruction efficiencies.

We evaluate the probability of a claim of discovery assuming three light Majorana neutrinos by

$$P_{\rm discovery} = \sum_{n=0}^{\infty} P(n|H_1) I(\mathcal{O}_1),$$

$$P_{\rm discovery} = \sum_{n=0}^{\infty} \left[\int_0^{\infty} P(n|\nu'+\lambda)P(\nu')d\nu'\right] I(\mathcal{O}_1),$$

where I is the indicator function,

$$I = 1 \quad \text{if } \mathcal{O}_1 \geq 100,$$
$$I = 0 \quad \text{if } \mathcal{O}_1 < 100.$$

In addition to the exposure and the signal efficiency, the discovery potential depends on the background level and on the cut value for the posterior odds. We take the prior odds to be 1. The probability distribution for ν is given by the results of our neutrino parameter analysis, and depends on the mass orderings, on the probability distribution for the effective neutrino mass in these orderings, and on the matrix element values. We give the probability distributions

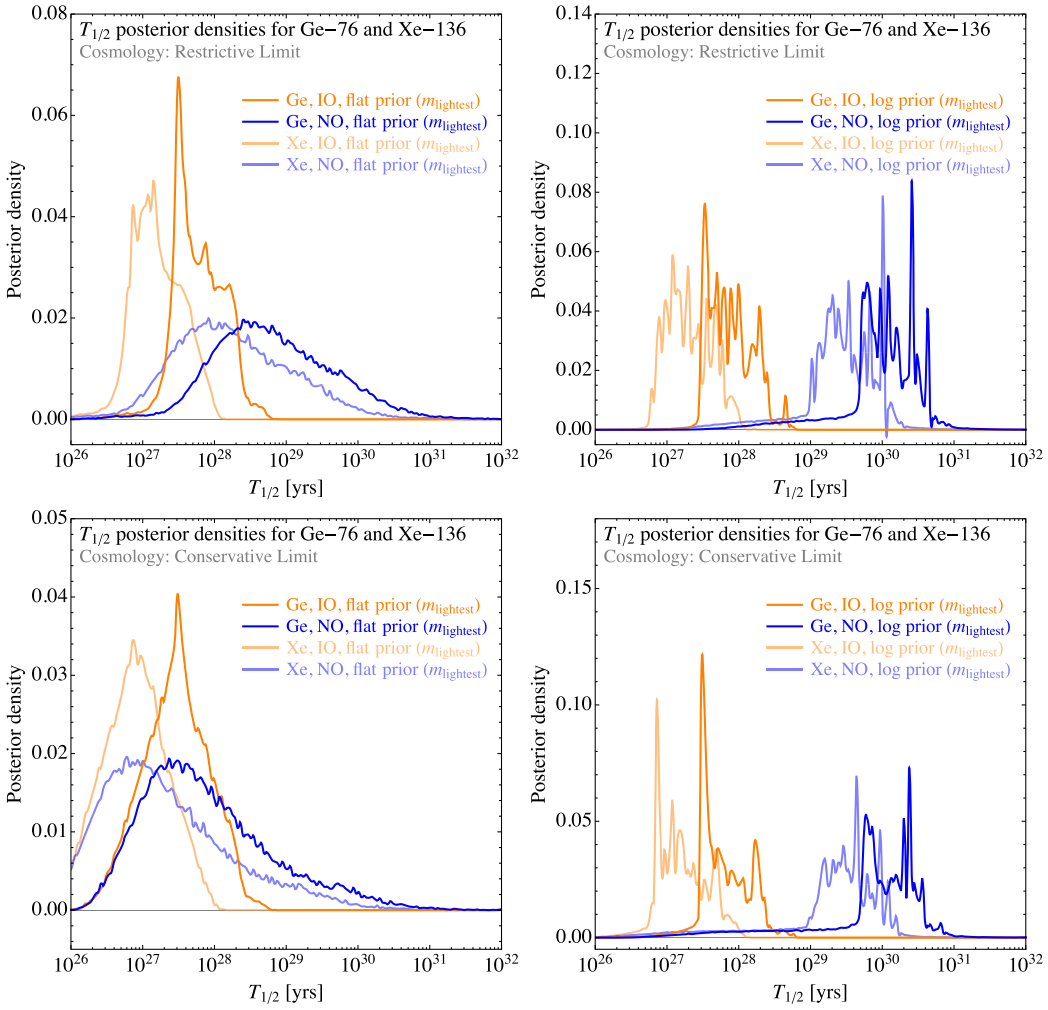

FIG. 7. Posterior probability distributions for $T_{1/2}$ (300 bins) for both ^{76}Ge and ^{136}Xe, featuring both mass orderings and both priors on the lightest neutrino mass. As to be expected, the logarithmic prior disfavors NO more strongly than the flat prior.

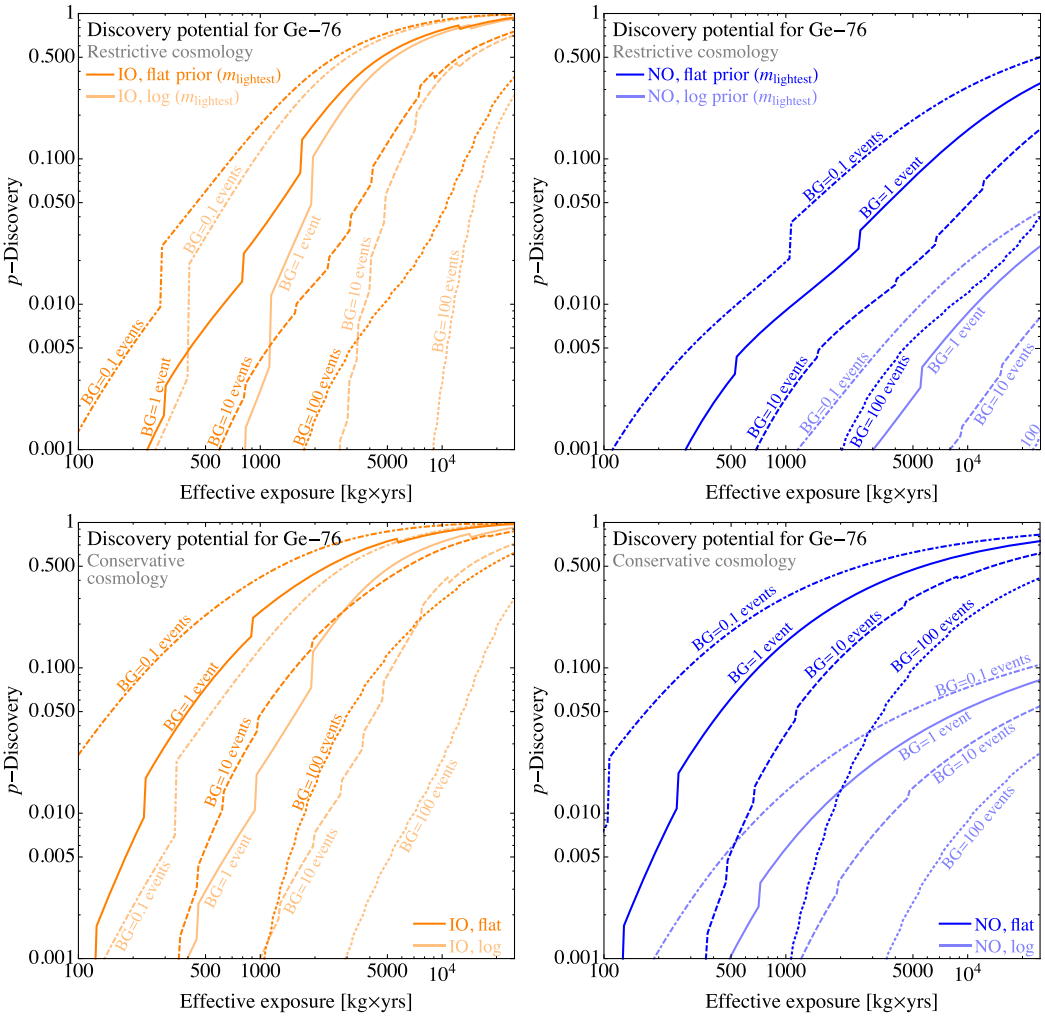

FIG. 8. Discovery potential for ^{76}Ge: IO/NO (left/right) and restrictive/conservative cosmology (top/bottom).

for $T_{1/2}$ for ^{76}Ge and ^{136}Xe for the different orderings, cosmological data sets, and mass scale priors in Fig. 7.

We use the output of BAT to carry out the integrals,

$$P(n|H_1) = \int_0^\infty P(n|\lambda+\nu')P(\nu')d\nu' \tag{A15}$$

$$= E[P(n|\lambda+\nu)]_{P(\nu)} \tag{A16}$$

$$\approx \frac{1}{N}\sum_{i=1}^{N} P(n|\lambda+\nu)|_{P(\nu)}, \tag{A17}$$

with N being the number of samples of $T_{1/2}$ taken from the Markov chain. Recall that the background level λ is fixed.

The discovery potential is presented as a function of efficiency reduced exposure, $E\epsilon$, for four different background levels ($\lambda = 0.1, 1, 10, 100$) for Ge and three for Xe (1,10,100). The results are shown in Figs. 8 and 9.[14]

[14]Results for other isotopes can be requested from the authors.

The cut value for the Bayes factor was taken to be 100—i.e., we would claim a discovery if H_1 was 100 times more probable than H_0. This is an arbitrary choice, and varying it by a factor of 10 is certainly possible, but it does not have strong influence on the results of this analysis. The jumps in the curve are due to the Poisson nature of the statistical fluctuations.

3. Posterior probability for lab-based experiments

In the main text, we have shown the heat map of posterior probability density in the Σ–$|m_{ee}|$ plane, i.e., in terms of the observables that are probed by cosmology and by $0\nu\beta\beta$. This choice is inspired by these two observables being the most promising both to identify the neutrino mass ordering and to get information on the absolute neutrino mass scale. In the theoretical literature, one often finds plots of m_{lightest} vs $|m_{ee}|$, see e.g. the literature on how to probe certain neutrino mass models by $0\nu\beta\beta$ [32–35]. However, this choice is problematic in our case because not only is m_{lightest} only a theoretical

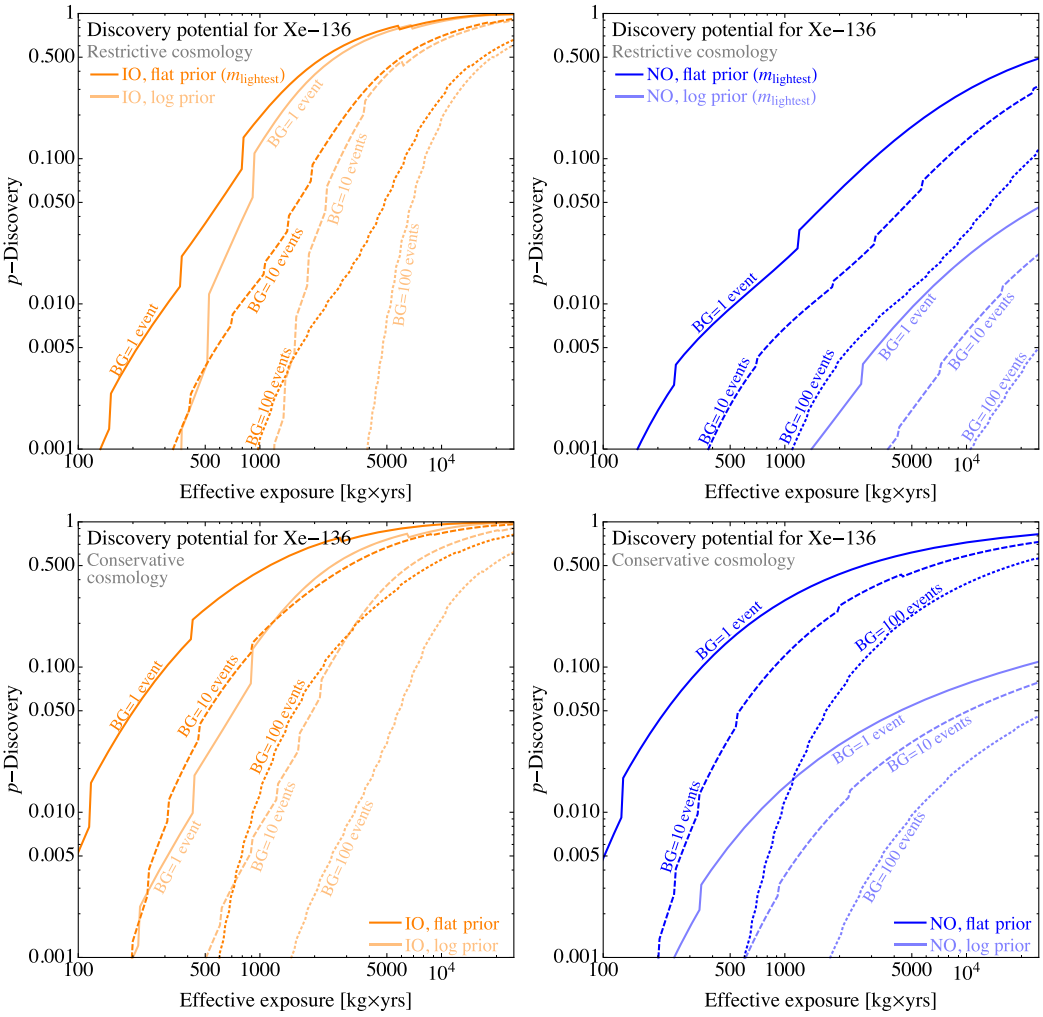

FIG. 9. Discovery potential for ^{136}Xe: IO/NO (left/right) and restrictive/conservative cosmology (top/bottom).

parameter rather than an actual observable, but it also exhibits a very strong dependence on the prior, which is poorly constrained.

When considering laboratory-based experiments, there is a long history of attempts to measure the neutrino mass in single beta decay experiments. These types of experiments are looking for deviations in the kinematic endpoint of the resulting electron spectrum, and what they constrain (at least with realistic resolution) is the so-called *effective electron neutrino mass* m_β, defined by[15]

$$m_\beta^2 = m_1^2 c_{12}^2 c_{13}^2 + m_2^2 s_{12}^2 c_{13}^2 + m_3^2 s_{13}^2. \qquad \text{(A18)}$$

Constraints on m_β have mainly been derived from tritium decay, by both MAINZ [36] and Troitsk [37]; however also future proposals focus on tritium[16] (such as Project 8 [39], which aims at detecting cyclotron radiation from relativistic tritium ions). The most prominent on-going effort is provided by KATRIN [40], which has recently started data taking and whose sensitivity on m_β could possibly be of $\mathcal{O}(1\text{ eV})$, with quoted values being dependent on the statistics used (e.g. Ref. [41] predicts a sensitivity of 0.20 eV based on frequentist methods and a sensitivity of 0.17 eV when using Bayesian statistics with a flat prior m_β^2).

The heat maps for our posterior probability densities for all combinations used of orderings and cosmological limits are displayed in Fig. 10. The underlying data are equivalent to those shown in the main text. As can be seen, the generic tendencies are basically the same as for the Σ–$|m_{ee}|$ plot. However, even if the ultimate sensitivity limit of KATRIN is reached, it will not add a significant constraint on the

[15]Note that, in fact, there is no mass that could be associated with an electron neutrino, as this is a not a mass eigenstate. Thus, the m_β defined in Eq. (A18) is really nothing more than an effective quantity, however, an observable one.

[16]Alternative isotopes, such as rhenium as used by MARE [38], do not seem to be competitive.

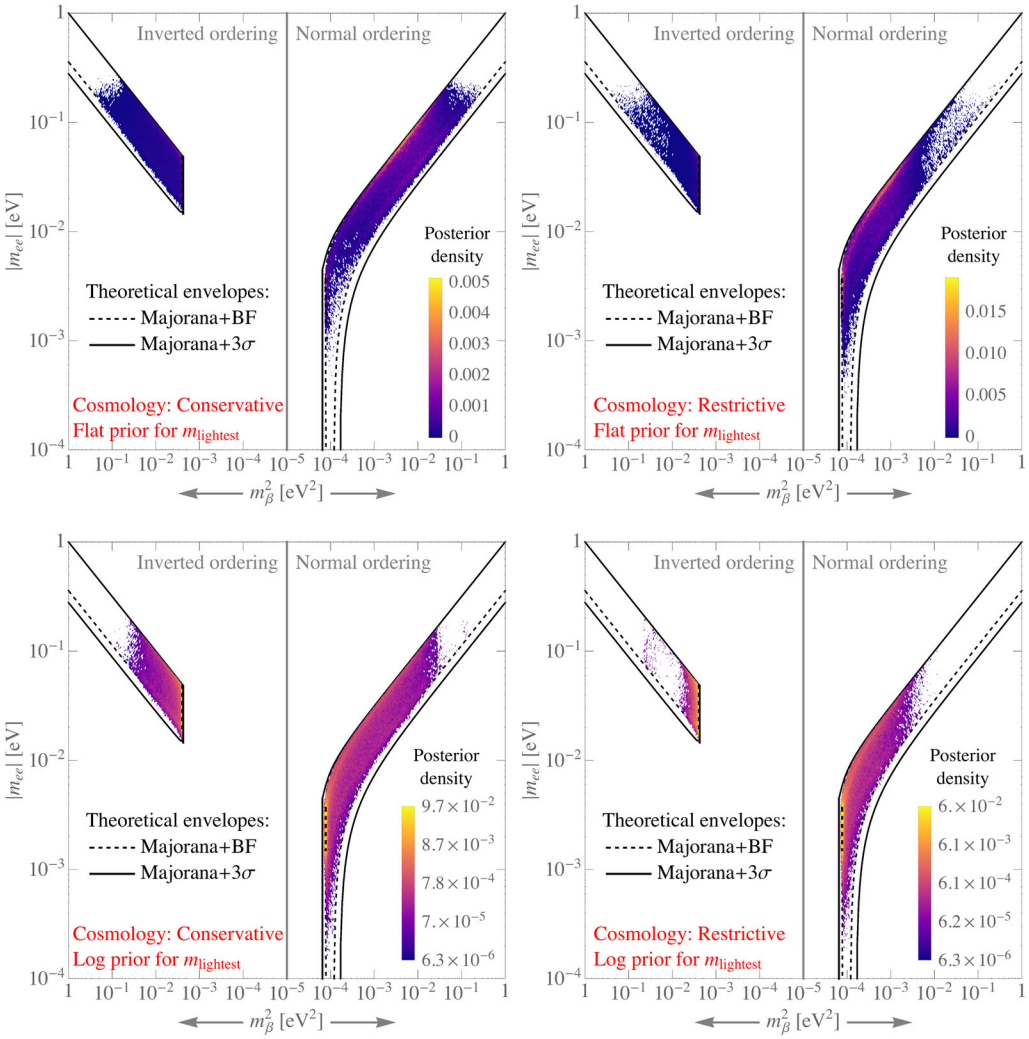

FIG. 10. Heat map of posterior probability density for both combinations of cosmological data sets and prior on m_{lightest}. The upper panels depict the flat prior while the lower panels show the log prior. Note that the color-coding changes from linear to logarithmic. As for the plots in the main text, we used a random subset of the posterior sample with a size of 2×10^6 for each mass ordering, binned into a grid of 400×400 pixels.

parameter space for the scenarios investigated—and thus the current bounds do not provide any significant constraints. This is our justification for neglecting data from single beta decay experiments.

4. Conclusion

The probability density heat maps and $0\nu\beta\beta$ results are based on currently available data. They should be updated as new data becomes available.

[1] I. Esteban, M. C. Gonzalez-Garcia, M. Maltoni, I. Martinez-Soler, and T. Schwetz, J. High Energy Phys. 01 (2017) 087.
[2] A. Gando *et al.* (KamLAND-Zen Collaboration), Phys. Rev. Lett. **117**, 082503 (2016); **117**, 109903 (2016).
[3] J. B. Albert *et al.* (EXO-200 Collaboration), Nature (London) **510**, 229 (2014).
[4] M. Agostini *et al.*, Nature (London) **544**, 47 (2017).
[5] M. Gerbino, M. Lattanzi, O. Mena, and K. Freese, arXiv: 1611.07847.
[6] F. Capozzi, E. Di Valentino, E. Lisi, A. Marrone, A. Melchiorri, and A. Palazzo, Phys. Rev. D **95**, 096014 (2017).
[7] F. Simpson, R. Jimenez, C. Pena-Garay, and L. Verde, J. Cosmol. Astropart. Phys. 06 (2017) 029.

[8] T. Schwetz, K. Freese, M. Gerbino, E. Giusarma, S. Hannestad, M. Lattanzi, O. Mena, and S. Vagnozzi, arXiv:1703.04585.
[9] M. Schervish, Am. Stat. **50**, 203 (1996).
[10] J. Neyman, Synthese **36**, 97 (1977).
[11] S. Davidson, G. Isidori, and A. Strumia, Phys. Lett. B **646**, 100 (2007).
[12] A. Caldwell, D. Kollár, and K. Kröninger, Comput. Phys. Commun. **180**, 2197 (2009).
[13] M. Lindner, A. Merle, and W. Rodejohann, Phys. Rev. D **73**, 053005 (2006).
[14] A. Merle and W. Rodejohann, Phys. Rev. D **73**, 073012 (2006).
[15] W. Maneschg, A. Merle, and W. Rodejohann, Europhys. Lett. **85**, 51002 (2009).
[16] T. Hahn, Comput. Phys. Commun. **168**, 78 (2005).
[17] M. Agostini, G. Benato, and J. Detwiler, Phys. Rev. D **96**, 053001 (2017).
[18] M. Doi, T. Kotani, and E. Takasugi, Prog. Theor. Phys. Suppl. **83**, 1 (1985).
[19] J. Suhonen and O. Civitarese, Phys. Rep. **300**, 123 (1998).
[20] W. Rodejohann, Int. J. Mod. Phys. E **20**, 1833 (2011).
[21] J. Menendez, A. Poves, E. Caurier, and F. Nowacki, Nucl. Phys. **A818**, 139 (2009).
[22] F. Simkovic, V. Rodin, A. Faessler, and P. Vogel, Phys. Rev. C **87**, 045501 (2013).
[23] A. Meroni, S. T. Petcov, and F. Simkovic, J. High Energy Phys. 02 (2015) 025.
[24] P. S. Bhupal Dev, S. Goswami, M. Mitra, and W. Rodejohann, Phys. Rev. D **88**, 091301 (2013).
[25] T. R. Rodriguez and G. Martinez-Pinedo, Phys. Rev. Lett. **105**, 252503 (2010).
[26] J. Barea, J. Kotila, and F. Iachello, Phys. Rev. C **87**, 014315 (2013).
[27] J. Suhonen and O. Civitarese, Nucl. Phys. **A847**, 207 (2010).
[28] M. T. Mustonen and J. Engel, Phys. Rev. C **87**, 064302 (2013).
[29] M. Agostini *et al.*, Nature (London) **544**, 47 (2017).
[30] A. Gando *et al.* (KamLAND-Zen Collaboration), Phys. Rev. Lett. **117**, 082503 (2016); **117**, 109903 (2016).
[31] J. B. Albert *et al.* (EXO-200 Collaboration), Nature (London) **510**, 229 (2014).
[32] S. F. King, A. Merle, and A. J. Stuart, J. High Energy Phys. 12 (2013) 005.
[33] J. Gehrlein, A. Merle, and M. Spinrath, Phys. Rev. D **94**, 093003 (2016).
[34] J. Gehrlein, A. Merle, and M. Spinrath, J. High Energy Phys. 09 (2015) 066.
[35] M. Agostini, A. Merle, and K. Zuber, Eur. Phys. J. C **76**, 176 (2016).
[36] C. Kraus *et al.*, Eur. Phys. J. C **40**, 447 (2005).
[37] V. M. Lobashev *et al.*, Phys. Lett. B **460**, 227 (1999).
[38] E. Ferri *et al.*, Phys. Procedia **61**, 227 (2015).
[39] B. Monreal and J. A. Formaggio, Phys. Rev. D **80**, 051301 (2009).
[40] A. Osipowicz *et al.* (KATRIN Collaboration), arXiv:hep-ex/0109033.
[41] O. Host, O. Lahav, F. B. Abdalla, and K. Eitel, Phys. Rev. D **76**, 113005 (2007).