

# Statistical significances and projections for proton decay experiments

Prudhvi N. Bhattiprolu<sup>1</sup>, Stephen P. Martin,<sup>2</sup> and James D. Wells<sup>1</sup>

<sup>1</sup>*Leinweber Center for Theoretical Physics, University of Michigan, Ann Arbor, Michigan 48109, USA*

<sup>2</sup>*Department of Physics, Northern Illinois University, DeKalb, Illinois 60115, USA*



(Received 28 October 2022; accepted 24 January 2023; published 14 March 2023)

We study the statistical significances for exclusion and discovery of proton decay at current and future neutrino detectors. Various counterintuitive flaws associated with frequentist and modified frequentist statistical measures of significance for multichannel counting experiments are discussed in a general context and illustrated with examples. We argue in favor of conservative Bayesian-motivated statistical measures, and as an application we employ these measures to obtain the current lower limits on proton partial lifetime at various confidence levels, based on Super-Kamiokande's data, generalizing the 90% C.L. published limits. Finally, we present projections for exclusion and discovery reaches for proton partial lifetimes in  $p \rightarrow \bar{\nu}K^+$  and  $p \rightarrow e^+\pi^0$  decay channels at Hyper-Kamiokande, DUNE, JUNO, and THEIA.

DOI: [10.1103/PhysRevD.107.055016](https://doi.org/10.1103/PhysRevD.107.055016)

## I. INTRODUCTION

In order to account for the observed matter-antimatter asymmetry in our universe, baryon number must be violated as required by the Sakharov conditions [1]. Although baryon number is a global symmetry of the (renormalizable) Standard Model (SM) Lagrangian, it may be violated by nonperturbative electroweak sphaleron effects (as yet unconfirmed by experiment) that are heavily suppressed at temperatures much lower than the electroweak scale [2,3]. The sphaleron effects, however, together with the  $CP$ -violation in the electroweak sector are not sufficient to explain the observed baryon asymmetry, and therefore provide a key motivation for theories beyond the SM with additional  $B$ -violation. Grand unified theories (GUTs), with or without supersymmetry, are well-motivated and generically predict baryon number violation, and therefore can lead to proton decay [4–43]. After integrating out the heavy fields, the nonrenormalizable operators built out of the SM fields that allow proton decay are of dimension-six or higher, with the suppression scale of order the GUT breaking scale.

In this paper, we consider proton decay in the  $p \rightarrow \bar{\nu}K^+$  and  $p \rightarrow e^+\pi^0$  decay channels that are typically predicted to be the leading modes in supersymmetric [18–43] and nonsupersymmetric [5–17] GUTs, respectively. At present, the strongest constraints on these proton partial lifetimes are from the Super-Kamioka neutrino detection experiment

(Super-Kamiokande), where the most stringent published 90% C.L. lower limits are  $5.9 \times 10^{33}$  years for the  $p \rightarrow \bar{\nu}K^+$  mode [44] and  $2.4 \times 10^{34}$  years for the  $p \rightarrow e^+\pi^0$  mode [45]. We will make projections for the exclusion and discovery reaches for these proton modes decays at future neutrino detectors at Deep Underground Neutrino Experiment (DUNE) [46], Jiangmen Underground Neutrino Observatory (JUNO) [47], Hyper-Kamiokande (the successor to Super-Kamiokande, and an order of magnitude larger) [48], and THEIA (a novel detector concept with water based liquid scintillator, 10% liquid scintillator and 90% water, that can detect and distinguish between Cerenkov and the scintillation light) [49].

In order to project the exclusion and discovery reaches, it is necessary to make choices regarding the statistical tools to be employed. Indeed, the results for such projections are only meaningful in the context of those choices. Here, we are interested in counting experiments with multiple independent channels with different signal rates and backgrounds, with uncertainties.

Our statistical analysis choices are guided by several requirements.

- (i) We aim for statistical measures that avoid reporting an exclusion or discovery when the experiment is actually not sensitive to the physics signal hypothesis under investigation. As we will discuss, pure frequentist statistics can suffer from this problem.
- (ii) We choose statistical measures such that the presence of a noninformative channel (one with a much higher background and/or a much lower signal rate than other channels) does not unduly affect the exclusion or discovery conclusion.
- (iii) We avoid statistical measures that contain the subtle flaw that they could counterintuitively imply a

---

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP<sup>3</sup>.*

greater sensitivity for an experiment if it increases its background.

Regarding this last point, in a previous paper [50], we have discussed the fact that the median expected significance for discovery or exclusion has just such a counter-intuitive flaw in the context of frequentist  $p$ -values for a single-channel counting experiment. We proposed a solution to that problem. As we will see below, this type of problem also occurs in the case of multichannel counting experiments, and can be avoided using Bayesian-motivated statistical measures.

For these reasons, Sec. II of this paper is devoted to a rather extensive discussion of the statistical issues associated with multichannel counting experiments with background and nuisance parameter uncertainties, in which we highlight some of the problems that can occur and explain our choices of statistical tools in a general context. In Sec. III we apply these statistical measures to discuss the present exclusions from Super-Kamiokande, and we project exclusion and discovery prospects for proton decay at DUNE, JUNO, Hyper-Kamiokande, and THEIA, for the proton decay modes  $p \rightarrow e^+ \pi^0$  and  $p \rightarrow \bar{\nu} K^+$ . Section IV summarizes our findings for exclusion and discovery prospects for runtimes of 10 and 20 years.

## II. STATISTICS FOR DISCOVERY AND EXCLUSION

### A. Basic definitions

In this paper we are concerned with new physics signals and backgrounds, which are both assumed to occur as random discrete events governed by Poisson statistics, possibly in multiple independent channels. In general, given data resulting from an experiment, the significance of a possible exclusion or discovery can be given in terms of a  $p$ -value, defined as the probability of obtaining a result of equal or greater incompatibility with a null hypothesis  $H_0$ . In high-energy physics, the  $p$ -value is often conventionally reported as a significance, defined by

$$Z \equiv \sqrt{2} \operatorname{erfc}^{-1}(2p), \quad (2.1)$$

which in the special case of a Gaussian distribution would coincide with the number of standard deviations.

The assumption for discovery is that the null hypothesis is a background-only hypothesis  $H_0 = H_b$ , while for exclusion the null hypothesis is a signal plus background model  $H_0 = H_{s+b}$ . Consider a test-statistic  $Q$  defined in such a way that larger  $Q$  is more signal-like and smaller  $Q$  is more background-like. In a single-channel counting experiment, for example,  $Q$  is simply the number of observed events. Then, for an experimental outcome  $Q_{\text{obs}}$ , one has the  $p$ -value for discovery:

$$p_{\text{disc}} = P(Q \geq Q_{\text{obs}} | H_b), \quad (2.2)$$

and the  $p$ -value for exclusion:

$$p_{\text{excl}} = P(Q \leq Q_{\text{obs}} | H_{s+b}). \quad (2.3)$$

In a frequentist approach, the  $p$ -value for a given data outcome is often used to provide a quantitative measure of the credence we give to  $H_0$ . However, the  $p$ -value cannot be directly interpreted as the probability that the null hypothesis is true, given the data. Nevertheless, small  $p$ -values are considered a measure of evidence against  $H_0$  in frequentist statistics. In particle physics, two popular standards for exclusion are to require that  $p_{\text{excl}} < 0.10$  or  $0.05$ , commonly referred to as 90% or 95% exclusion. For rejection of the background-only hypothesis in favor of some new model, a higher standard is almost always required, with either  $Z_{\text{disc}} > 3$  ( $p_{\text{disc}} < 0.001350$ ) for “evidence,” or  $Z_{\text{disc}} > 5$  ( $p_{\text{disc}} < 2.867 \times 10^{-7}$ ) for “discovery”.

In high energy physics experiments in the 21st century, starting with the Higgs boson searches at the LEP  $e^-e^+$  collider and for all kinds of searches for new phenomena at the Large Hadron Collider (LHC), it has become very common to use a modified frequentist statistical measure for exclusion, called the  $\text{CL}_s$  method. This is a more conservative approach to assigning exclusion significances than  $p_{\text{excl}}$ . The idea of  $\text{CL}_s$  [51–54] is to divide the usual  $p$ -value for exclusion by the  $p$ -value that would be obtained with the signal assumed absent:

$$\text{CL}_s(Q_{\text{obs}}) = \frac{P(Q \leq Q_{\text{obs}} | H_{s+b})}{P(Q \leq Q_{\text{obs}} | H_b)}. \quad (2.4)$$

A specific motivation for using  $\text{CL}_s$  rather than  $p_{\text{excl}}$  is to avoid reporting an exclusion in cases for which the experiment is actually not sensitive to the purported signal hypothesis, but the observed data has a small  $p$ -value anyway. This can occur, for example, in a counting experiment if the observed number of events is significantly smaller than the background estimate, as we will discuss in detail shortly.

Note that, by design,  $\text{CL}_s$  is not a  $p$ -value or even a probability, but rather a ratio of probabilities. Nevertheless, the exclusion is reported using  $\text{CL}_s$  in place of the exclusion  $p$ -value, so that one reports 95% (or 90%) exclusion if  $\text{CL}_s < 0.05$  (or  $0.1$ ). Because the denominator is always less than 1, the modified frequentist measure  $\text{CL}_s$  is always more conservative in reporting exclusions than the frequentist  $p$  value, in the sense that using it reduces the false exclusion rate compared to using  $p_{\text{excl}}$ . In particle physics literature,  $\text{CL}_s$  was introduced in Ref. [51] and detailed (along with its advantages, reviewed and illustrated below) in Refs. [52–54].

It is also useful to have a counterpart to the  $p_{\text{disc}}$  statistic that similarly guards against claiming discovery in situations where the experiment is not sensitive to the signal model. In Ref. [55], an approach to discovery significance

was proposed using the Bayes factor [56–58] of the null hypothesis  $H_0 = H_b$  to the alternative hypothesis  $H_1 = H_{s+b}$ . For an experiment investigating a putative signal with strength  $s$ , the Bayes factor  $B_{01}$  is (using the probabilities in place of the likelihoods, to which they are proportional):

$$B_{01} = \frac{P(Q_{\text{obs}}|H_b)}{\int_0^\infty ds' \pi(s') P(Q_{\text{obs}}|H_{s'+b})}, \quad (2.5)$$

where  $\pi(s')$  is a Bayesian prior probability distribution for the signal strength. As mentioned in [55], this expression is only meaningful in the case of a prior that is proper, i.e.,  $\int_0^\infty ds' \pi(s') = 1$ , since otherwise the arbitrary normalization of an improper prior would make the Bayes factor  $B_{01}$  also arbitrary. This precludes the use of a flat prior, for example. For a single-channel counting experiment with background mean  $b$ , that reference argues in favor of the proper prior  $\pi(s') = b/(s' + b)^2$ , referred to as the objective signal prior. However, we find it counterintuitive to use a prior for the signal that depends on the background. Instead, we choose simply  $\pi(s') = \delta(s' - s)$ , expressing certainty in the prediction of the signal model. If the signal model prediction is not perfectly well known, it is straightforward to generalize this with an appropriate  $\pi(s')$ . We therefore define the simple likelihood ratio statistic for the confidence level in the discovery,

$$\text{CL}_{\text{disc}}(Q_{\text{obs}}) = \frac{P(Q_{\text{obs}}|H_b)}{P(Q_{\text{obs}}|H_{s+b})}. \quad (2.6)$$

While various scales have been proposed (see, e.g., Jeffreys' in [58] and Kass and Raftery's in [57]) to interpret the Bayes factor as a measure of evidence in favor of or against a null hypothesis, we propose to use  $\text{CL}_{\text{disc}}$  in place of  $p$  in Eq. (2.1) to obtain a discovery significance  $Z$ , in exactly the same way that a frequentist  $p_{\text{disc}}$  would be used. As we will illustrate below, our choice gives results that are always more conservative than the significances obtained from  $p_{\text{disc}}$ . This is very similar to the way the modified frequentist measure  $\text{CL}_s$  is now commonly used in place of  $p$  in Eq. (2.1) to report an exclusion significance that is always more conservative than that of the standard frequentist method, even though  $\text{CL}_s$ , like  $\text{CL}_{\text{disc}}$ , is not a probability.

## B. Single-channel counting experiments

To illustrate the statistical methods discussed above let us consider the special case of a simple experiment that counts the number of events  $n$ , with signal and background modeled as independent Poisson processes with means  $s$  and  $b$  respectively. For a mean  $\mu$ , the Poisson probability to observe  $n$  events is

$$P(n|\mu) = \frac{\mu^n e^{-\mu}}{n!}. \quad (2.7)$$

Therefore, in the idealized case of perfectly known background, the  $p$ -value for discovery is the probability that data generated under hypothesis  $H_0 = H_b$  is equally or more signal-like than the actual observed number of events  $n$ :

$$p_{\text{disc}}(n, b) = \sum_{k=n}^{\infty} P(k|b) = \gamma(n, b)/\Gamma(n). \quad (2.8)$$

The  $p$ -value for exclusion is the probability that data generated under hypothesis  $H_0 = H_{s+b}$  is equally or more backgroundlike than the actual observed number of events  $n$ :

$$p_{\text{excl}}(n, b, s) = \sum_{k=0}^n P(k|s+b) = \Gamma(n+1, s+b)/\Gamma(n+1). \quad (2.9)$$

In these equations,  $\gamma(z, x)$  and  $\Gamma(z, x)$  are the lower and upper incomplete gamma functions, respectively, defined by

$$\gamma(z, x) = \int_0^x dt t^{z-1} e^{-t}, \quad \Gamma(z, x) = \int_x^\infty dt t^{z-1} e^{-t}, \quad (2.10)$$

so that  $\Gamma(z) = \gamma(z, x) + \Gamma(z, x)$  is the ordinary gamma function.

The  $\text{CL}_s$  statistic for exclusion in this case is

$$\text{CL}_s(n, b, s) = \frac{p_{\text{excl}}(n, b, s)}{p_{\text{excl}}(n, b, 0)} = \frac{\Gamma(n+1, s+b)}{\Gamma(n+1, b)}. \quad (2.11)$$

This is larger than  $p_{\text{excl}}(n, b, s)$  by a factor  $\Gamma(n+1)/\Gamma(n+1, b)$ .

Figure 1 illustrates the idea of the  $\text{CL}_s$  method [51–54]. In the figure,  $p_{\text{excl}}(n, b, s)$  (the shaded area under the blue histograms) is divided by  $p_{\text{excl}}(n, b, 0)$  (the shaded area under the red histograms) to give  $\text{CL}_s$ . The first panel shows the case  $b = 2.2$ ,  $s = 8.4$ , and  $n = 5$ . In situations like this, where the  $H_b$  and  $H_{s+b}$  hypothesis distributions do not have much overlap,  $p_{\text{excl}}$  and  $\text{CL}_s$  evaluate to very similar results due to the denominator of the  $\text{CL}_s$  definition being close to 1. For this particular case, one finds  $p_{\text{excl}} = 0.0475$  and  $\text{CL}_s = 0.0487$ , and by either criterion one would report a better than 95% exclusion.

The second panel of Fig. 1 illustrates the case  $b = 8.4$ ,  $s = 2.2$ , and  $n = 5$ , so that the overlap between the distributions for  $H_b$  and  $H_{s+b}$  is much larger. In cases like this with a larger overlap (i.e., the signal regions get polluted by the background) statistical conclusions based on  $p_{\text{excl}}$  alone can be too aggressive. Since we engineered this example to have the same  $b + s$  and  $n$  as for the first

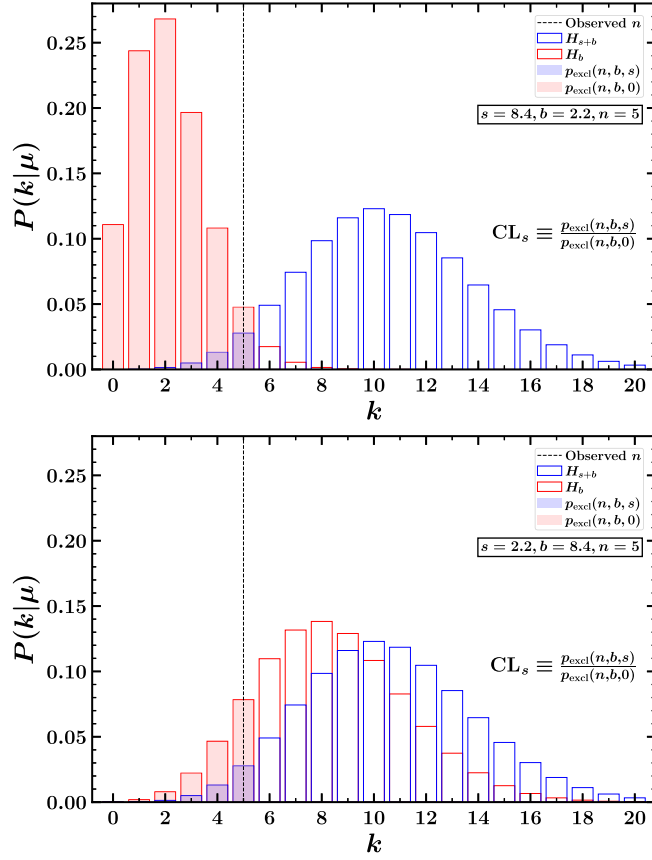


FIG. 1. Illustration of the idea of the  $CL_s$  statistic for exclusion as an improvement over  $p_{\text{excl}}$ . The Poisson distributions  $P(k|\mu)$  are generated under the hypotheses that signal and background are both present  $\mu = s + b$  (blue histograms) and that the signal is absent  $\mu = b$  (red histograms). For the observed number of events  $n$ ,  $p_{\text{excl}}(n, b, s)$  [from Eq. (2.9)] is shown by the shaded area part under the blue histogram, and  $p_{\text{excl}}(n, b, 0)$  is the shaded area part under the red histogram, while  $CL_s$  [from Eq. (2.11)] is their ratio. In the first plot, the Poisson means of the signal and background are taken to be  $s = 8.4$  and  $b = 2.2$ , respectively, while in the second plot they are  $s = 2.2$  and  $b = 8.4$ . In both plots, the observed number of events is  $n = 5$ . In the first plot, there is little overlap between the distributions from the  $H_b$  and  $H_{s+b}$  hypotheses, and  $p_{\text{excl}} = 0.0475$  and  $CL_s = 0.0487$ , so one would report better than 95% exclusion using either criterion. In the second plot, the overlap is much larger. Although  $p_{\text{excl}} = 0.0475$  is the same (since  $s + b$  and  $n$  did not change), one finds  $CL_s = 0.3022$ , and one refrains from reporting an exclusion of the hypothesis  $H_{s+b}$ .

panel, we get the same<sup>1</sup>  $p_{\text{excl}} = 0.0475$ , which taken at face value would again give a better than 95% exclusion.

<sup>1</sup>The general fact that  $p_{\text{excl}}(n, b, s)$  depends only on the sum  $s + b$ , and not on  $s$  or  $b$  separately, is a clear reason to reject it as a measure of confidence in the presence of the signal model, because it says that any exclusion for signal  $s$  and background  $b$  would imply an equally strong exclusion for the case that the signal is  $s = 0$  if the background  $b$  were increased by the numerical value of  $s$ .

However, proponents of the  $CL_s$  criteria point out that here it must be recognized that for  $b = 8.4$ , the outcome  $n \leq 5$  would have been a low-probability occurrence no matter what<sup>2</sup> the signal mean  $s$  was. Thus, the frequentist  $p_{\text{excl}}$  is really telling us more about the observed data than making a useful statement about the signal hypothesis. One finds that  $CL_s = 0.3022$ , and using this one would, sensibly and conservatively, refrain from excluding the signal hypothesis.

In fact, no matter the outcome for  $n$ , the experiment with  $b = 8.4$  simply lacks the statistical ability to exclude the  $s = 2.2$  signal model at 90% confidence, according to the  $CL_s$  statistic. This can be seen by computing it for the least signal-like outcome,  $n = 0$ , which gives  $CL_s = 0.1108$ . One possible practical interpretation of the very small  $p_{\text{excl}}$  in such cases with  $n$  significantly less than  $b$  might be that the background estimate could be wrong for reasons unknown, while another is that the background simply fluctuated low from its true mean. In any case, the intuitive interpretation of the  $CL_s$  statistic is that the quoted significance for exclusion should be reduced from the usual frequentist value, due to the large overlap between the signal + background region and the background-only region.

Indeed, if the number of events is sufficiently small, one finds that the usual frequentist  $p$ -value would correspond to an exclusion even in cases that defy sensible practical interpretation. Considering the case  $n = 0$  more generally, one finds  $p_{\text{excl}}(n = 0, s, b) = e^{-(s+b)}$ , which becomes arbitrarily small for any fixed  $s$ , if  $b$  is sufficiently large. One could use this to make an absurd claim of exclusion for a model that predicted  $s = 10^{-500}$  or even  $s = 0$  exactly, simply by observing a smaller than expected number of events, if the background is large enough. In contrast, usage of the statistic  $CL_s(n = 0, b, s) = e^{-s}$  conforms to the intuitively reasonable idea that, as an absolute prerequisite for excluding a signal hypothesis, the expected signal strength must not be too small. Specifically, only models that predict  $s > -\ln(0.05) \approx 2.996$  can be excluded at 95% confidence according to the  $CL_s$  measure, for any  $b$  and for any possible experimental outcome  $n$ . Similarly, 90% exclusion by the  $CL_s$  method requires  $s > -\ln(0.1) \approx 2.303$ .

The dependence of the exclusion significance on  $b$  is shown for fixed  $s = 4$  and  $n = 0, 1, 2, 3$  in Fig. 2. For very small  $b$ , the two statistics are nearly equal,  $p_{\text{excl}} \simeq CL_s$ . For any fixed  $n$ , in the limit of large  $b$  one has  $CL_s = e^{-s}$ , while  $p_{\text{excl}}$  becomes absurdly small in comparison, which would imply an absurdly large  $Z_{\text{excl}}$ .

<sup>2</sup>Here we are taking it as a requirement that  $s \geq 0$ , although in some situations quantum interference with the background could allow for  $s < 0$ . See, for example, the case of a digluon resonance at the LHC [59].



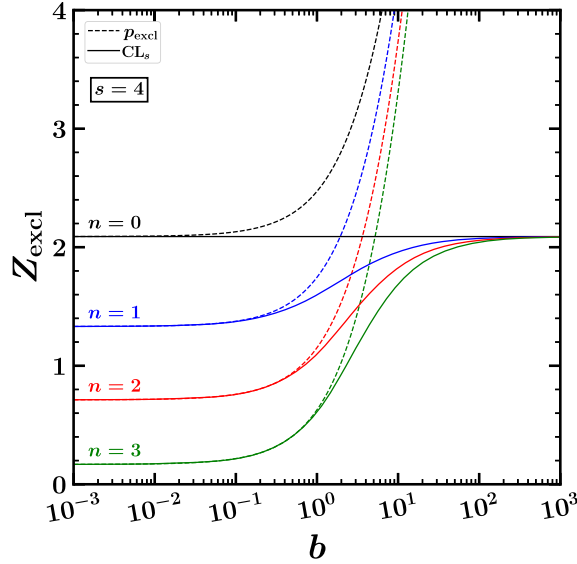


FIG. 2. Comparison of significances  $Z$  obtained using Eq. (2.1) from  $p_{\text{excl}}$  [dashed lines, from Eq. (2.9)] and  $\text{CL}_s$  [solid lines, from Eq. (2.11)], for fixed  $s = 4$  as a function of varying  $b$ , for  $n = 0, 1, 2$ , and  $3$ . For very small  $b$ , the two statistics are nearly equal,  $p_{\text{excl}} \simeq \text{CL}_s$ . In the limit of large  $b$  one has  $\text{CL}_s = e^{-s}$ , independent of  $n$ , while  $p_{\text{excl}}$  becomes absurdly small in comparison.

Nonobservation of a significant excess above background expectations can be used to constrain new physics. In particular, for a single-channel counting experiment, the minimum signal needed to claim an exclusion at a

given confidence level  $1 - \alpha$ , equivalent to significance  $Z = \sqrt{2} \text{erfc}^{-1}(2\alpha)$ , for a perfectly known background mean  $b$ , is obtained [60,61] by solving for  $s$  in either

$$\alpha = \frac{\Gamma(n+1, s+b)}{\Gamma(n+1)} \quad (p_{\text{excl}} \text{ method}) \quad (2.12)$$

in the standard frequentist approach, or

$$\alpha = \frac{\Gamma(n+1, s+b)}{\Gamma(n+1, b)} \quad (\text{CL}_s \text{ method}) \quad (2.13)$$

in the modified frequentist approach. Figure 3 shows the 90% C.L. ( $\alpha = 0.1$ , left panel) and 95% C.L. ( $\alpha = 0.05$ , right panel) upper limits on signal as functions of the background mean, for a fixed number of observed events  $n = 0, 1, 2$ , using the  $p_{\text{excl}}$  (red lines) and  $\text{CL}_s$  (blue lines) criteria. Also shown in the figure are the 90% C.L. and 95% C.L. upper limits on  $s$  that are obtained using the Feldman-Cousins (FC) method based on an ordering principle introduced in Ref. [62]. The upper limits obtained by the FC method for a fixed  $n$  do not always decrease with increasing  $b$ ; instead they have a sawtooth pattern, as can be seen from the dotted lines in the figure. This behavior is because of the discreteness of Poisson distributions. The solid black lines in Fig. 3 show the results obtained by the FC method after requiring them to be nonincreasing as a function of background mean.

It is clear from the figure that the upper limits on  $s$  obtained using the standard frequentist  $p_{\text{excl}}$  approach are

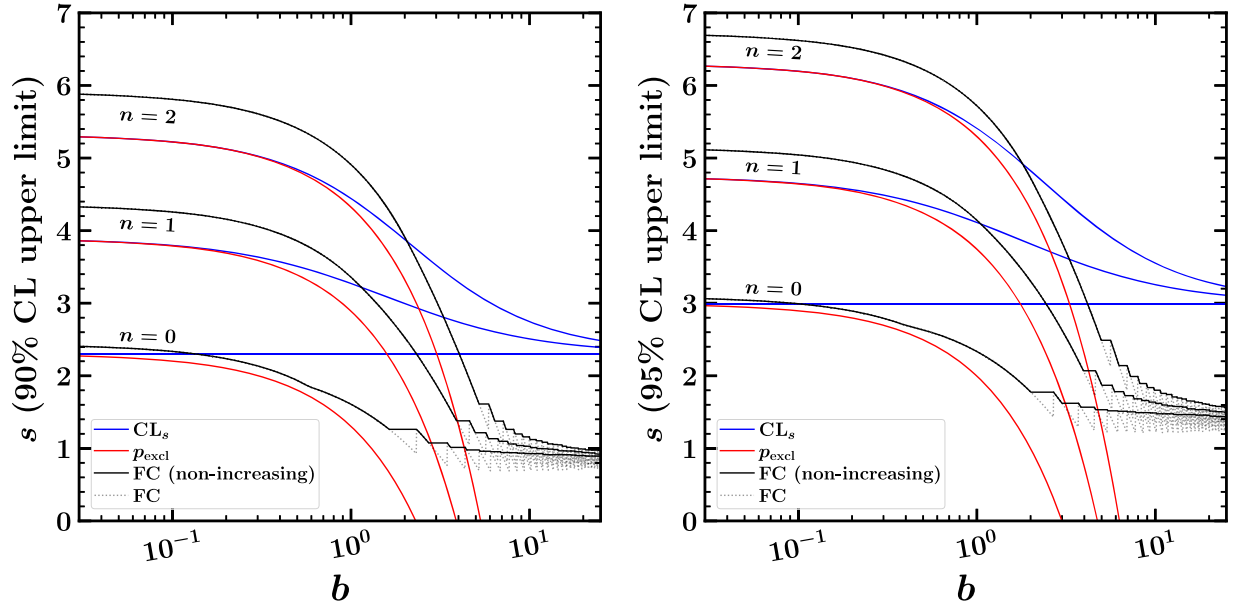


FIG. 3. The 90% C.L. (left panel) and 95% C.L. (right panel) upper limits on signal as functions of the background mean  $b$ , for a fixed number of observed events  $n = 0, 1, 2$ , using the  $\text{CL}_s$  technique [blue lines, from Eq. (2.13)], standard frequentist  $p$ -value approach [red lines, from Eq. (2.12)], and Feldman-Cousins method [solid black lines, from Ref. [62]]. The dotted black lines show the results obtained by the Feldman-Cousins method before requiring them to be nonincreasing as a function of background mean.

the least conservative, and can even go negative in the case where the number of observed events  $n$  is small compared to the expected background mean. For a fixed  $n$ , despite the upper limits given by the  $\text{CL}_s$  and FC methods being very different from each other, we note that they are both almost flat at very small backgrounds and then decrease slowly (or stay constant) as a function of background, always remaining positive. For small  $b$  the FC upper limits are more conservative, and for large  $b$ , the  $\text{CL}_s$  upper limits are more conservative. The other striking difference between these two upper limits is that, for  $n = 0$ , the FC upper limits decrease with  $b$ , but the  $\text{CL}_s$  upper limits are independent of  $b$ . In particular, at a chosen confidence level  $1 - \alpha$ , for  $n = 0$  the  $\text{CL}_s$  upper limit on  $s$  is  $-\ln(\alpha)$ . The same result also holds for any  $n$  in the limit that the background is extremely large. At 90% (95%) C.L., the upper limit given by  $\text{CL}_s$  for  $n = 0$ , or for any  $n$  as  $b \rightarrow \infty$ , is around 2.303 (2.996). On the other hand, the upper limit given by the FC method decreases as a function of  $b$  and approaches a constant value at large  $b$ . For example, for  $n = 0$ , the 90% (95%) C.L. upper limit given by the FC method, after requiring it to be nonincreasing as a function of  $b$ , is approximately 0.8 (1.34) at large  $b$ .

It is important for the following that the result for  $\text{CL}_s(n, b, s)$  in the case of a single Poisson channel in Eq. (2.11) can also be obtained [63] as a Bayesian credible interval, using a flat prior for the signal and likelihoods  $\mathcal{L}(s|n, b) \propto P(n|s + b)$ :

$$\begin{aligned} \text{CL}_{\text{excl}}(n, b, s) &= \frac{\int_s^\infty ds' \mathcal{L}(s'|n, b)}{\int_0^\infty ds' \mathcal{L}(s'|n, b)} \\ &= \frac{\int_s^\infty ds' e^{-(s'+b)} (s' + b)^n}{\int_0^\infty ds' e^{-(s'+b)} (s' + b)^n}. \end{aligned} \quad (2.14)$$

Performing the integrations,  $\text{CL}_{\text{excl}}(n, b, s)$  as defined by Eq. (2.14) is precisely equal to  $\text{CL}_s(n, b, s)$  as defined by Eq. (2.11).<sup>3</sup> However, despite the numerical equivalence, the interpretation is quite different, since the ratio of frequentist  $p$ -values is not directly a Bayesian confidence interval. Moreover, the equivalence between  $\text{CL}_s$  and  $\text{CL}_{\text{excl}}$  is only approximate in more complicated generalizations. Looking ahead to the case of experiments which collect counts in multiple independent channels governed by Poisson statistics, and which may have nuisance parameters including uncertainties in the backgrounds, we will argue for a generalization based straightforwardly on the Bayesian version  $\text{CL}_{\text{excl}}$  as given in Eq. (2.14)

<sup>3</sup>If the signal mean is instead allowed to be negative with  $s + b \geq 0$  (see previous footnote), then  $\text{CL}_{\text{excl}}(n, b, s)$  can be defined as  $\text{CL}_{\text{excl}}(n, b, s) = \frac{\int_{s-b}^\infty ds' e^{-(s'+b)} (s' + b)^n}{\int_0^\infty ds' e^{-(s'+b)} (s' + b)^n}$ . After performing the integrations,  $\text{CL}_{\text{excl}}(n, b, s)$  is now precisely equal to  $p_{\text{excl}}(n, b, s)$  as defined in Eq. (2.9).

rather than  $\text{CL}_s$  given in Eq. (2.4) or its specialization Eq. (2.11).

For a single-channel counting experiment, the discovery confidence level statistic defined in Eq. (2.6) becomes

$$\text{CL}_{\text{disc}}(n, b, s) = \frac{P(n|b)}{P(n|b + s)} = \frac{e^s}{(1 + s/b)^n}, \quad (2.15)$$

which can be used in place of  $p$  in Eq. (2.1) to obtain a discovery significance. (If the result is greater than 1, then clearly no discovery claim should be contemplated.) Note that unlike  $p_{\text{disc}}(n, b)$ , the result for  $\text{CL}_{\text{disc}}(n, b, s)$  depends on the strength of the signal whose discovery is under investigation. It is always more conservative than  $p_{\text{disc}}(n, b)$  in claiming discovery, just as  $\text{CL}_s$  is more conservative than  $p_{\text{excl}}$  in claiming exclusion. For example, in the extreme case  $s = 0$ , one has  $\text{CL}_{\text{disc}}(n, b, s = 0) = 1$  for any  $b$  and  $n$ , so one would never claim discovery using that criteria. In contrast, the frequentist statistic  $p_{\text{disc}}(n, b)$  can be arbitrarily small, implying an arbitrarily large discovery significance  $Z$ , even in situations where the physics provides absolutely no possible source for a signal.<sup>4</sup> As we will see below,  $\text{CL}_{\text{disc}}$  also generalizes more straightforwardly to cases that have multiple independent channels governed by Poisson statistics, and which may have nuisance parameters including uncertainties in the backgrounds.

Figure 4 compares the discovery significance obtained from  $p_{\text{disc}}$  and  $\text{CL}_{\text{disc}}$  as a function of  $s$  for fixed  $n$ , with different curves for different values of  $b$ . Note that the discovery significance obtained from  $\text{CL}_{\text{disc}}$ , which is always more conservative than that of  $p_{\text{disc}}$ , is maximized at  $s = n - b$ .

Given the number of observed events  $n$  and an expected background mean, the standard  $p$ -value for discovery  $p_{\text{disc}}$  does not depend on the signal. So, for a perfectly known background mean  $b$ , we can compute the number of events needed for discovery at a significance  $Z$  by solving for  $n$  from [see Eqs. (2.1) and (2.8)]

$$\frac{1}{2} \text{erfc}\left(\frac{Z}{\sqrt{2}}\right) = \frac{\gamma(n, b)}{\Gamma(n)} \quad (p_{\text{disc}} \text{ method}). \quad (2.16)$$

On the other hand,  $\text{CL}_{\text{disc}}$  depends also on the signal, in which case the number of events needed for discovery for a known background  $b$  and signal mean  $s$  at a given significance  $Z$  can be obtained by solving for  $n$  from [see Eqs. (2.1) and (2.15)]

<sup>4</sup>For example, imagine a search for a new fundamental particle of mass 1 TeV, conducted by dropping a bag of hammers from the top of a tall building, with a somewhat noisy detector surrounding the impact point on the sidewalk. For this experiment, theoretical modeling confidently predicts  $s = 0$ , so one should reasonably refrain from announcing discovery even if one estimated  $b = 0.01$  and observed  $n = 3$ .

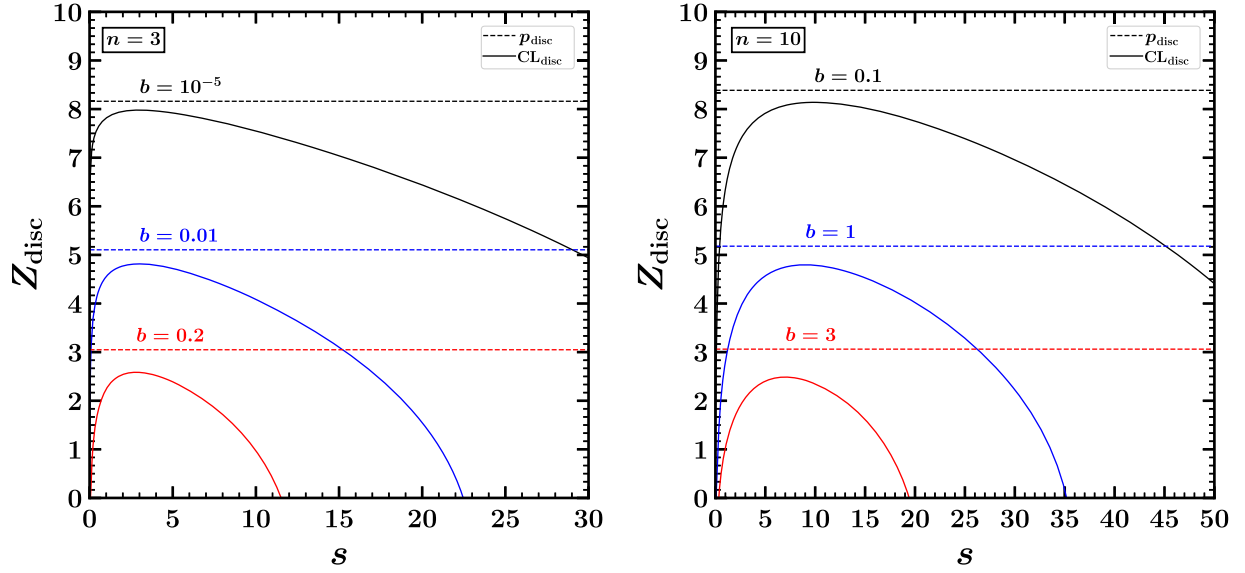


FIG. 4. Comparison of discovery significances obtained using Eq. (2.1) from  $p_{\text{disc}}$  [dashed lines, from Eq. (2.8)] and  $\text{CL}_{\text{disc}}$  [solid lines, from Eq. (2.15)] as a function of  $s$  for  $n=3$  (left panel) and  $n=10$  (right panel), for various choices of  $b$ .

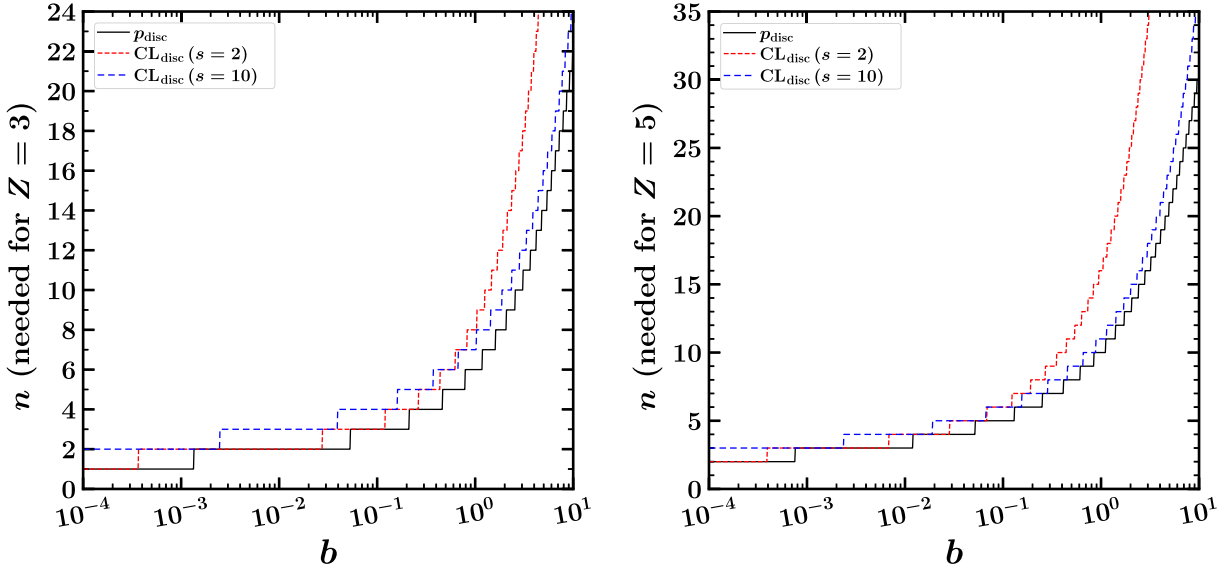


FIG. 5. The observed number of events  $n$  needed for  $Z=3$  evidence (left panel) and  $Z=5$  discovery (right panel) as functions of the expected background mean  $b$ . The solid black lines show the result obtained from Eq. (2.16) using the standard frequentist approach based on  $p_{\text{disc}}$ , which is independent of the signal mean  $s$ . The dashed red and blue lines show the results obtained from Eq. (2.17) using  $\text{CL}_{\text{disc}}$  for the cases of signal mean  $s=2$  and 10, respectively.

$$\frac{1}{2} \operatorname{erfc}\left(\frac{Z}{\sqrt{2}}\right) = \frac{e^s}{(1+s/b)^n} \quad (\text{CL}_{\text{disc}} \text{ method}). \quad (2.17)$$

Figure 5 shows the observed number of events required for  $Z=3$  evidence (left panel) and  $Z=5$  discovery (right panel) given by the  $p_{\text{disc}}$  approach (solid black lines), and the  $\text{CL}_{\text{disc}}$  approach for two choices of the signal mean  $s=2$  (dashed red lines) and 10 (dashed blue lines) as functions of  $b$ . It is clear from the figure that, for a given

background mean, the observed number of events needed for discovery given by the  $\text{CL}_{\text{disc}}$  approach are at least as large as the result given by the  $p_{\text{disc}}$  criterion, and often much larger when the background is not very small.

We now turn to the question of projecting expectations for exclusion and discovery at ongoing and future experiments. In simulations or assessments of a proposed experiment, one considers the statistics of pseudodata generated under an alternative hypothesis  $H_1$ . For assessments of prospects for exclusion the alternative hypothesis is that the

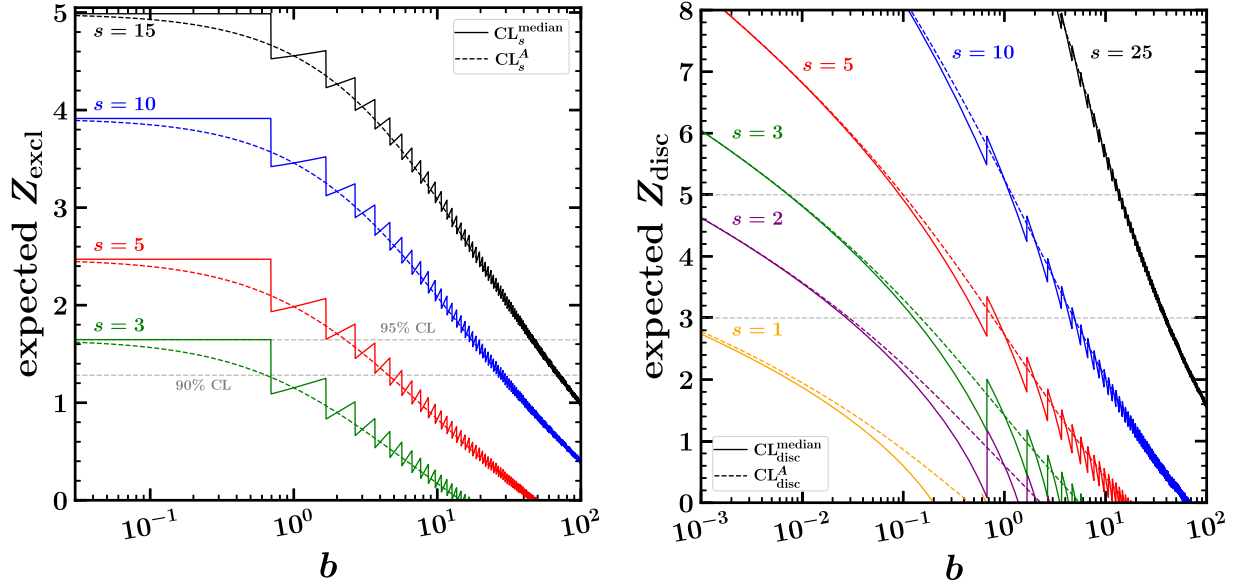


FIG. 6. Median (solid lines) and exact Asimov (dashed lines) expected significances obtained using Eq. (2.1) from  $\text{CL}_s = \text{CL}_{\text{excl}}$  [from Eqs. (2.11) and (2.19)] for exclusion (left panel) and  $\text{CL}_{\text{disc}}$  [from Eqs. (2.15) and (2.21)] for discovery (right panel), as a function of the background mean  $b$  for various values of signal mean  $s$ , for a single-channel Poisson counting experiment. Due to the discrete nature of Poisson statistics, the median expected significances suffer from a sawtooth behavior. On the other hand, the exact Asimov expected significances behave sensibly as they decrease monotonically with  $b$ .

signal source is absent,  $H_1 = H_b$ , while for discovery, the pseudodata is generated assuming that both signal and background are present,  $H_1 = H_{s+b}$ .

A common way to project an expected result is to set the number of events  $n$  equal to the median expected value under the hypothesis  $H_1$ . However, due to the discrete nature of Poisson statistics events, the median expected outcome has the striking flaw that it can predict smaller significances if an experiment takes more data or reduces its background. This counterintuitive feature of the median expected significance was pointed out and studied in detail in Refs. [64,65], and in [50] where it was referred to as the “sawtooth problem”. It occurs for the median expected  $\text{CL}_s$  and  $\text{CL}_{\text{disc}}$  as well. The sawtooth behavior of the median expected  $\text{CL}_s$  and  $\text{CL}_{\text{disc}}$  as a function of the background mean  $b$ , for various values of signal mean  $s$ , is evident from Fig. 6. For comparison, Fig. 6 also show the significances obtained from the exact Asimov expected  $\text{CL}_s$  and  $\text{CL}_{\text{disc}}$  (dashed lines), detailed below, that are smooth and sensible.

Therefore, in Ref. [50], we proposed instead to use an exact Asimov approach for projecting sensitivities of planned experiments, where the observed number of events  $n$  is replaced by its mean expected value  $\langle n_{\text{excl}} \rangle = b$  for exclusion and  $\langle n_{\text{disc}} \rangle = s + b$  for discovery. From Eqs. (2.9) and (2.11) we thus obtain for the expected exclusion in the case of a single-channel counting experiment with signal and background means  $s$  and  $b$ :

$$p_{\text{excl}}^A = \frac{\Gamma(b+1, s+b)}{\Gamma(b+1)}, \quad (2.18)$$

$$\text{CL}_s^A = \text{CL}_{\text{excl}}^A = \frac{\Gamma(b+1, s+b)}{\Gamma(b+1, b)}, \quad (2.19)$$

Similarly, for the expected discovery significance, we obtain from Eqs. (2.8) and (2.15):

$$p_{\text{disc}}^A = \frac{\gamma(s+b, b)}{\Gamma(s+b)}, \quad (2.20)$$

$$\text{CL}_{\text{disc}}^A = \frac{e^s}{(1+s/b)^{s+b}}. \quad (2.21)$$

Figure 7 compares the exact Asimov expected significances obtained from frequentist (dashed lines) and modified frequentist  $\text{CL}_s$ /Bayesian  $\text{CL}_{\text{disc}}$  (solid lines) confidence levels, for both exclusion (left panel) and discovery (right panel) cases. This illustrates the more general fact that  $\text{CL}_s$  and  $\text{CL}_{\text{disc}}$  are more conservative than  $p_{\text{excl}}$  and  $p_{\text{disc}}$ , respectively.

In order to project expected exclusions based on the  $p_{\text{excl}}$  or  $\text{CL}_s$  approaches, we set Eq. (2.18) or (2.19) equal to the desired  $\alpha = 0.10$  or  $0.05$ , and then solve for  $s$ . We also consider projections based on the FC method, in two different ways. One is the Feldman-Cousins experimental sensitivity, advocated within Ref. [62], that is defined as the arithmetic mean of the upper limits obtained by the FC method at a chosen confidence level<sup>5</sup>  $s_{\text{FC}}^{\text{UL}}(n, b)$  in a large

<sup>5</sup>These upper limits on signal are defined in Ref. [62], and are shown as a function of background  $b$  for  $n = 0, 1, 2$  with solid black lines in Fig. 3.



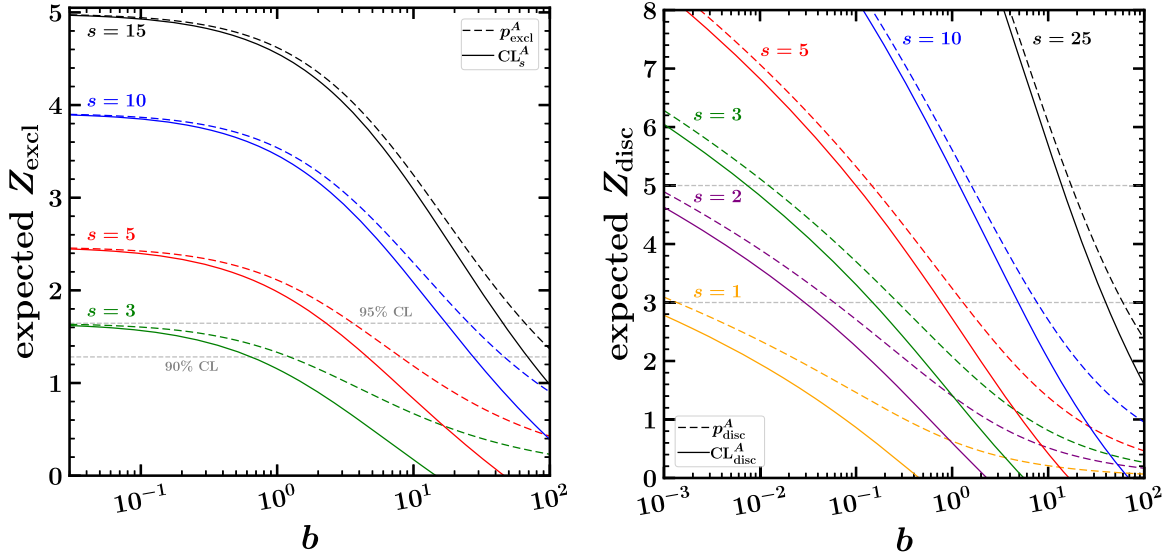


FIG. 7. The exact Asimov expected significances obtained from frequentist  $p$ -values (dashed lines) and modified frequentist  $\text{CL}_s$ /Bayesian  $\text{CL}_{\text{disc}}$  confidence levels (solid lines), converted to significances  $Z$  using Eq. (2.1), for a single-channel Poisson counting experiment. Results are presented as functions of the background mean  $b$  for various values of signal mean  $s$ . The term “exact Asimov” means that we set the number of events equal to the mean expected according to the hypothesis  $H_1$ , so  $n = b$  for exclusion and  $n = s + b$  for discovery. The left panel compares  $p_{\text{excl}}^A$  to  $\text{CL}_s^A$  for exclusion, from Eqs. (2.18) and (2.19). The right panel compares  $p_{\text{disc}}^A$  to  $\text{CL}_{\text{disc}}^A$  for discovery, from Eqs. (2.20) and (2.21).

number of pseudoexperiments with data generated under background-only hypothesis<sup>6</sup>:

$$\text{FC sensitivity} = \sum_{n=0}^{\infty} P(n|b) s_{\text{FC}}^{\text{UL}}(n, b). \quad (2.22)$$

The other way is to simply compute the upper limit on signal given by the FC method with the observed number of events taken to be the nearest integer to the expected background mean  $n = \text{round}(b)$ .<sup>7</sup> We consider the latter for future reference, as it was alluded to in Ref. [67] while projecting exclusion sensitivity for proton decay in  $p \rightarrow \bar{\nu} K^+$  channel at DUNE.

In Fig. 8, we compare the expected 90% C.L. (left panels) and 95% C.L. (right panels) upper limits on the signal mean  $s$ , obtained using the exact Asimov  $\text{CL}_s$  (blue lines) and  $p_{\text{excl}}$  (red lines), FC experimental sensitivity<sup>8</sup>

<sup>6</sup>An implementation of the Feldman-Cousins method to evaluate the upper limits and the experimental sensitivity, advocated within Ref. [62], is made available with the `zstats v2.0` package [66].

<sup>7</sup>When rounding half-integral values of  $b = 0.5, 1.5, 2.5, 3.5, \dots$ , we follow the IEEE 754 standard of taking the nearest even integer such that  $\text{round}(b) = 0, 2, 2, 4, \dots$

<sup>8</sup>In evaluating the FC sensitivity, we used the upper limits obtained by the Feldman-Cousins method for a fixed  $n$  before requiring them to be nonincreasing as a function of background mean. This does not make much difference as the FC upper limit differs from its nonincreasing (with  $b$ ) version only when the number of observed events are few compared to the expected background mean  $b$ , for which the probability of occurrence is small and will rapidly fall off for even smaller  $n$ .

(green lines), and FC upper limit with  $n = \text{round}(b)$  (black lines). We note the following from the figure. First, unlike the case with the observed upper limits (i.e., fixed  $n$ ), the  $p_{\text{excl}}$  method gives sensible positive expected upper limits with the exact Asimov approach for all  $b$ , but still is less conservative than the  $\text{CL}_s$  and FC sensitivity results. Second, the upper limit given by the FC method with  $n = \text{round}(b)$  suffers from a sawtooth problem and is therefore counterintuitive and flawed as a method of comparing experimental prospects for different scenarios, as it implies that an experiment could become more sensitive if it had larger background. Finally, the FC sensitivity and the upper limits given by exact Asimov  $\text{CL}_s$  are both sensible as they increase monotonically with  $b$ , and are also comparable at small backgrounds. At large backgrounds, however, the FC sensitivity is slightly more conservative. We also note that  $\text{CL}_s$  upper limits are much easier to evaluate than the FC upper limits.

We now turn to the issue of prospects for discovery, using the exact Asimov criterion. The signal mean needed for an expected discovery at a significance  $Z$  is given by the solution for  $s$  in setting Eq. (2.20) for  $p_{\text{disc}}$ , or (2.21) for  $\text{CL}_{\text{disc}}$ , equal to  $\frac{1}{2} \text{erfc}(\frac{Z}{\sqrt{2}})$  for the desired  $Z$ . Figure 9 compares the signals  $s$  needed for an expected  $Z = 3$  evidence or  $Z = 5$  discovery, as a function of background mean  $b$ , based on  $p_{\text{disc}}^A$  and  $\text{CL}_{\text{disc}}^A$ . We note that as expected the results from  $\text{CL}_{\text{disc}}^A$  are more conservative than those obtained from  $p_{\text{disc}}^A$ .

For very small  $b$ , note that for  $Z = 3$  the  $s$  needed in Fig. 9 is actually less than 1. Here, it is important to note

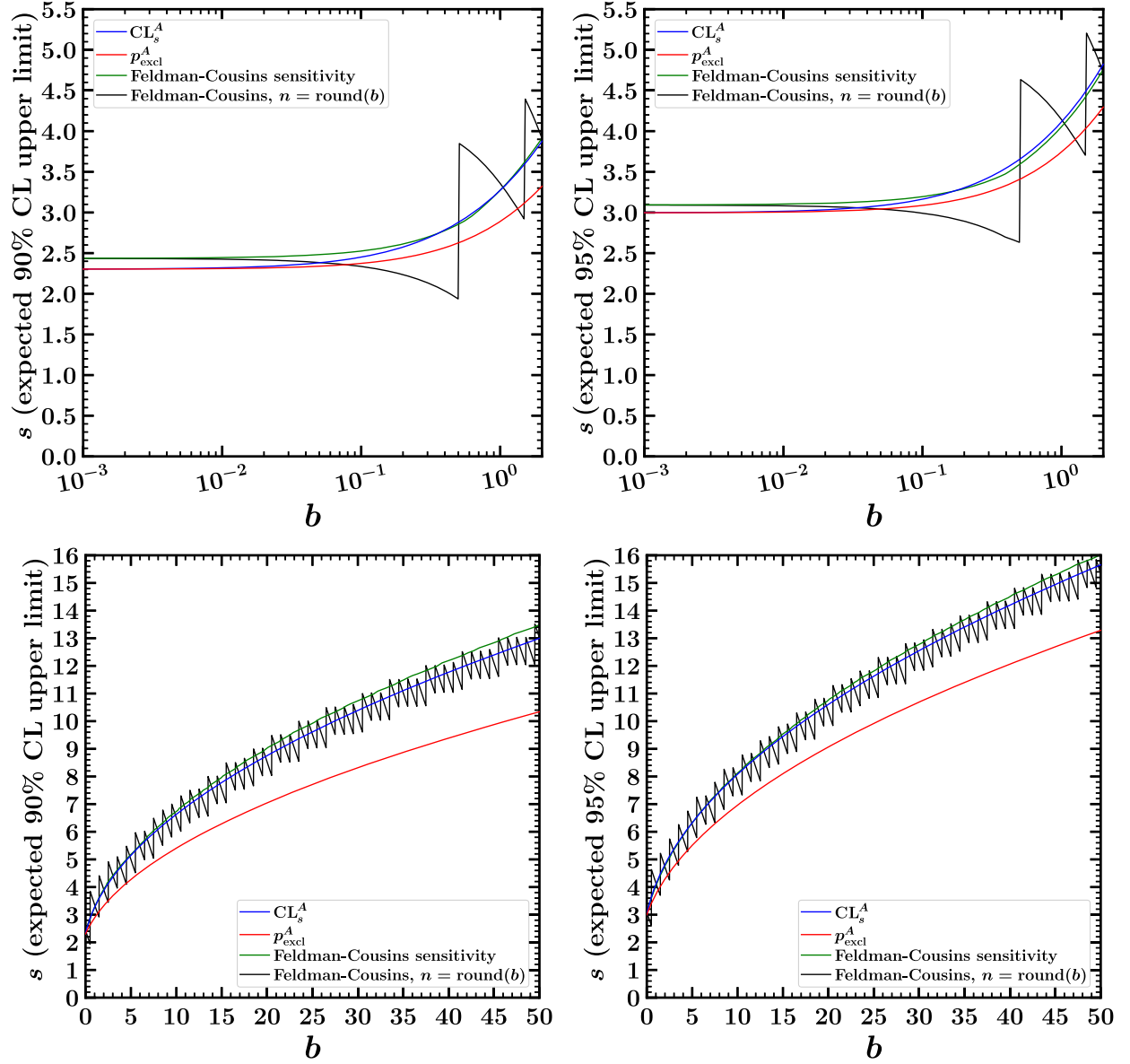


FIG. 8. The expected 90% C.L. (left panels) and 95% C.L. (right panels) upper limit on signal as a function of the background mean, using the exact Asimov modified frequentist  $\text{CL}_s$  [blue lines, from Eq. (2.19)] and standard frequentist  $p$ -value [red lines, from Eq. (2.18)], the Feldman-Cousins experimental sensitivity [green lines, from Ref. [62] and Eq. (2.22)], and the Feldman-Cousins method from Ref. [62] with  $n = \text{round}(b)$  [black lines]. The top and bottom panels show the same information but with logarithmic and linear scales, respectively, for  $b$ .

that the discovery statistics  $p_{\text{disc}}$  and  $\text{CL}_{\text{disc}}$  are not well defined in the strict background-free limit  $b \rightarrow 0$ . Specifically,

$$p_{\text{disc}}(n, 0) = \begin{cases} 0 & \text{if } n \neq 0 \\ 1 & \text{if } n = 0, \end{cases} \quad (2.23)$$

$$\text{CL}_{\text{disc}}(n, 0, s) = \begin{cases} 0 & \text{if } n \neq 0, s \neq 0 \\ 1 & \text{otherwise.} \end{cases} \quad (2.24)$$

Since  $\langle n_{\text{disc}} \rangle = s$  for  $b = 0$ , the above implies that the exact Asimov expected discovery significances are both infinite,  $Z(p_{\text{disc}}^A) = Z(\text{CL}_{\text{disc}}^A) = \infty$ , for any nonzero  $s$  (however small). However, as a practical matter, it is clearly unreasonable to suggest an expectation of a discovery if the mean expected number of signal events is much less than 1. Therefore, in order to be conservative, in cases with an extremely small background we can impose an additional requirement that  $P(n \geq 1)$  should be greater than some fixed value in order to claim an expected discovery.

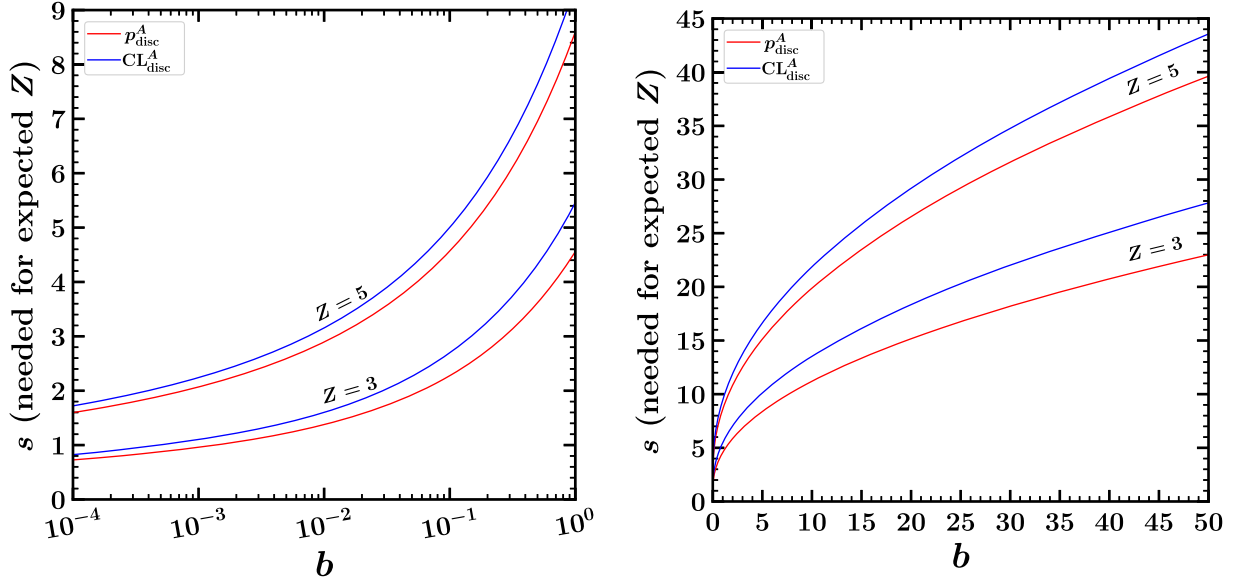


FIG. 9. The signal needed for an expected  $Z = 3$  evidence (lower curves) or  $Z = 5$  discovery (higher curves), as a function of background mean  $b$ , using the exact Asimov  $p_{\text{disc}}$  [red lines, from Eq. (2.20)] and  $\text{CL}_{\text{disc}}^A$  [blue lines, from Eq. (2.21)].

Figure 10 shows the probability of observing at least one event,

$$P(n \geq 1) = \sum_{n=1}^{\infty} P(n|s) = 1 - e^{-s}, \quad (2.25)$$

as a function of signal mean  $s$ . For example, if we require  $P(n \geq 1) > (50\%, 63.2\%, 95\%)$  then the signal mean has to be  $s > (0.693, 1.0, 2.996)$  respectively. Requiring  $s > \ln 2 \approx 0.693$  guarantees the median number of events

is at least 1, and  $s > 1$  guarantees the expected mean number of events  $\langle n_{\text{disc}} \rangle > 1$ .

### C. Exclusion for multichannel counting experiments

Consider a counting experiment with  $N$  independent channels. For each channel  $i = 1, \dots, N$ , the background and possible signal are assumed to be governed by Poisson distributions with means  $b_i$  and  $s_i$ . For future convenience, we define

$$s = \sum_{i=1}^N s_i, \quad (2.26)$$

$$r_i = s_i/s, \quad (2.27)$$

so that  $s$  is the total mean expected signal in all channels, and the  $r_i$  are the expected fractions of the total signal events for each channel.

Given an observation  $\{n_i\}$ , the  $p$ -value for exclusion is<sup>9</sup>

$$p_{\text{excl}}(\vec{n}, \vec{b}, \vec{s}) = \sum_{\{k_i\}} \prod_{i=1}^N P(k_i | s_i + b_i), \quad (2.28)$$

where the sums over non-negative integer numbers of events  $\{k_i\}$  are restricted according to the condition that

$$Q(\vec{k}) \leq Q(\vec{n}), \quad (2.29)$$

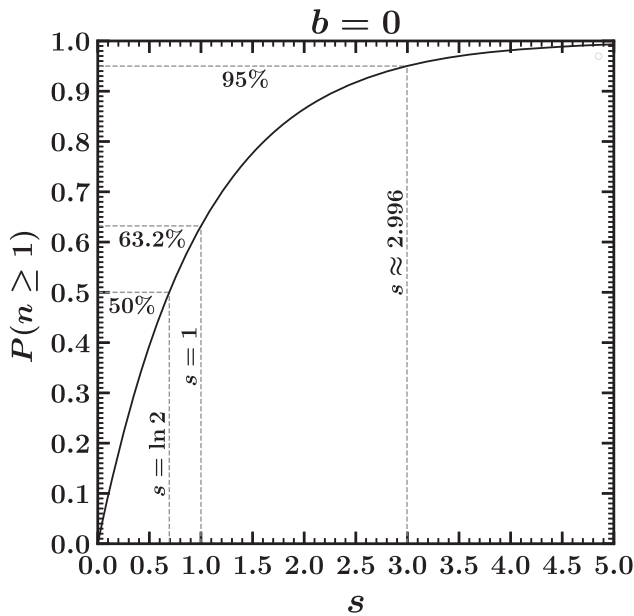


FIG. 10. Probability of observing at least one event obtained from Eq. (2.25), as a function of the signal mean  $s$ , in the case with no background  $b = 0$ .

<sup>9</sup>In the following we use  $\vec{n}$  as the argument of a function to denote the dependence on the full set  $\{n_i\}$ . This applies similarly for  $\vec{b}$  and  $\vec{s}$  to represent the dependences on  $\{b_i\}$  and  $\{s_i\}$ .

where  $Q$  is an appropriately chosen test-statistic with the property that larger  $Q$  is more signal-like.

We can also compute:

$$p_{\text{excl}}(\vec{n}, \vec{b}, 0) = \sum_{\{k_i\}} \prod_{i=1}^N P(k_i | b_i), \quad (2.30)$$

with the same restrictions on  $k_i$  as in Eq. (2.29). Then we have

$$\text{CL}_s(\vec{n}, \vec{b}, \vec{s}) = \frac{p_{\text{excl}}(\vec{n}, \vec{b}, \vec{s})}{p_{\text{excl}}(\vec{n}, \vec{b}, 0)}, \quad (2.31)$$

which is interpreted as the confidence level in the hypothesis that the signal is present.

For the single channel case, the obvious choice for  $Q$  is the observed number of events, but in the multichannel case one can consider different choices for  $Q$ . A simple and good choice<sup>10</sup> of test-statistic  $Q$  is the likelihood ratio,

$$q(\vec{n}, \vec{b}, \vec{s}) = \prod_{i=1}^N \frac{P(n_i | s_i + b_i)}{P(n_i | b_i)}, \quad (2.32)$$

which simplifies to

$$q = \prod_{i=1}^N e^{-s_i} \left(1 + \frac{s_i}{b_i}\right)^{n_i}. \quad (2.33)$$

It is more convenient to use instead

$$Q = \ln(q) = -s + \sum_{i=1}^N n_i \ln(1 + s_i/b_i), \quad (2.34)$$

which gives exactly the same results for  $p_{\text{excl}}$  and  $\text{CL}_s$  as  $Q = q$ , since  $\ln(q)$  increases monotonically with  $q$ . The contribution  $-s$  is an irrelevant constant (independent of the data  $\{n_i\}$ ), so the use of  $Q = \ln(q)$  amounts to taking the sum of the individual  $n_i$ 's, but weighting each of the channels by the factor  $w_i = \ln(1 + s_i/b_i)$ . This means that, using Eq. (2.34) in Eq. (2.29), the restriction on the  $\{k_i\}$  appearing in the sums in Eqs. (2.28) and (2.30) becomes:

$$\sum_{i=1}^N (n_i - k_i) \ln(1 + s_i/b_i) \geq 0. \quad (2.35)$$

In contrast, the Bayesian way is to define, as a generalization of Eq. (2.14):

<sup>10</sup>There are other choices, including the profile likelihood ratio, but these are more complicated and end up giving very similar (and often identical) results.

$$\text{CL}_{\text{excl}}(\vec{n}, \vec{b}, \vec{s}) = \frac{\int_s^\infty ds' \prod_{i=1}^N P(n_i | r_i s' + b_i)}{\int_0^\infty ds' \prod_{i=1}^N P(n_i | r_i s' + b_i)}. \quad (2.36)$$

Unlike in the special case of a single channel,  $\text{CL}_{\text{excl}}(\vec{n}, \vec{b}, \vec{s})$  defined in this way is not exactly equal to  $\text{CL}_s(\vec{n}, \vec{b}, \vec{s})$  defined by Eq. (2.31). Therefore, we will now study some simple test cases to illustrate the differences.

First, let us consider what happens when there are two channels, one of which (the “bad,” or noninformative channel) has a much lower signal and higher background than the other (the “good channel”). As a specific numerical case, suppose:

$$b_1 = 2, \quad s_1 = 7, \quad n_1 = 2, \quad (\text{good channel}), \quad (2.37)$$

$$b_2 = 10, \quad s_2 = 0.01, \quad n_2 = \text{varying}, \quad (\text{bad channel}). \quad (2.38)$$

In this case, because the bad channel 2 has a tiny expected signal  $s_2$  and a large background  $b_2$ , one intuitively expects it to provide essentially no information about the correctness of the signal hypothesis, no matter what  $n_2$  is observed. Considering only the good channel 1, we obtain

$$p_{\text{excl}} = 0.006232, \quad Z_{\text{excl}} = 2.4987 \quad (\text{channel 1 alone}), \quad (2.39)$$

$$\text{CL}_s = \text{CL}_{\text{excl}} = 0.009210,$$

$$Z_{\text{excl}} = 2.3571, \quad (\text{channel 1 alone}). \quad (2.40)$$

However, combining both channels using the formulas (2.28), (2.31), and (2.36) above, we have the results shown in the left panel of Fig. 11. Counterintuitively, adding another channel with a larger background and almost no expected signal has increased our confidence in the exclusion as measured by either the frequentist  $p_{\text{excl}}$  or the modified frequentist  $\text{CL}_s$  measures, when  $n_2$  is small. In contrast,  $\text{CL}_{\text{excl}}$  behaves as intuitively expected; the result obtained including both channels is numerically almost independent of  $n_2$  and almost identical to the result obtained only from channel 1.

To understand the origin of this counterintuitive effect for  $p_{\text{excl}}$  and  $\text{CL}_s$ , let us consider which integers  $k_1, k_2$  contribute to the sums in Eqs. (2.28) and (2.30). In general,  $k_1 = 0$  and 1 each contribute for a very large range of  $k_2$ , so that very nearly we have a factor  $\sum_{k_2=0}^\infty P(k_2 | s_2 + b_2) \approx 1$  for channel 2 in Eq. (2.28). However, for  $k_1 = n_1 = 2$ , we only get a factor of  $\sum_{k_2=0}^{n_2} P(k_2 | s_2 + b_2) < 1$  contributing to the  $p$ -values. The problem boils down to this fact: for the contributions with  $k_1 = n_1$ , only a subset of the  $k_2$  values contribute, even though any result for  $k_2$  should give us essentially no information about the presence of the (tiny) signal. This explains why the counterintuitive problem



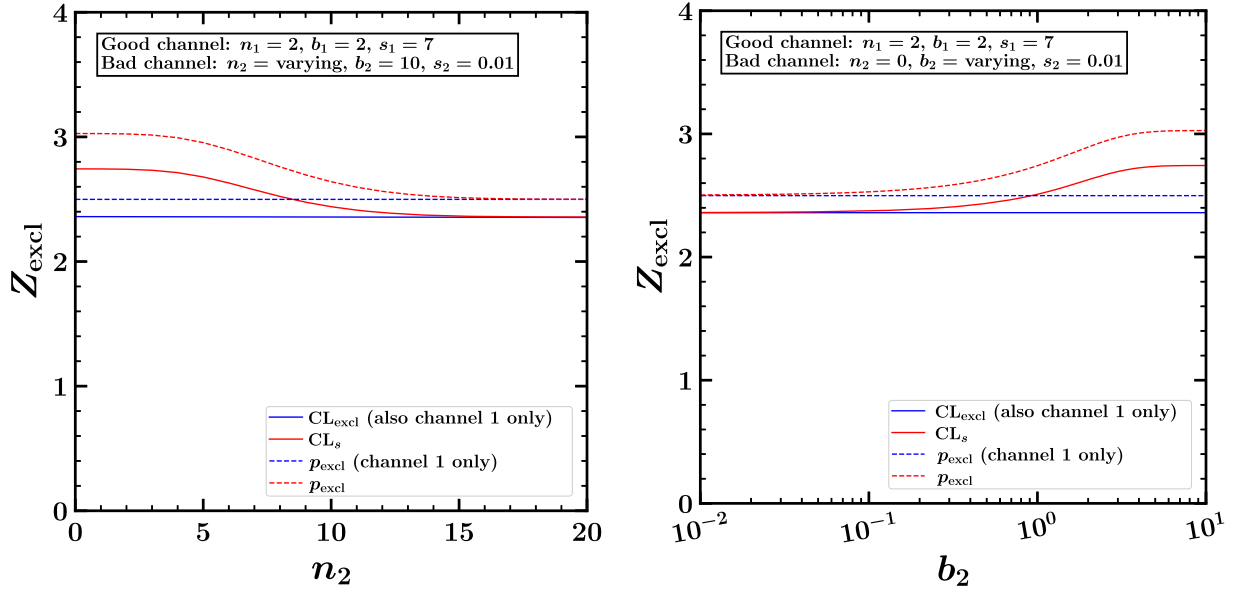


FIG. 11. Comparison of exclusion significances  $Z$  in the case of a counting experiment with a good channel and a bad channel. The solid lines are the modified frequentist  $\text{CL}_s$  [solid red line, from Eqs. (2.28)–(2.31) and (2.35)] and  $\text{CL}_{\text{excl}}$  [solid blue line, from Eq. (2.36)]. In this example,  $\text{CL}_{\text{excl}}$  is visually indistinguishable from the result obtained from channel 1 only, conforming with the fact that channel 2 contains essentially no information about the signal. Also shown are the results for  $p_{\text{excl}}$  obtained from considering channel 1 only [dashed blue line, from Eq. (2.9)] and from both channels [dashed red line, from Eqs. (2.28) and (2.35)].

disappears for reasonably large  $n_2$ , where we see from the left panel of Fig. 11 that  $\text{CL}_{\text{excl}} \approx \text{CL}_s$  and  $p_{\text{excl}}$  agree with their counterparts from channel 1 only.

To show another facet of this disturbing effect, in the right panel of Fig. 11 we use the same data except that  $n_2 = 0$  is fixed and  $b_2$  is varying. Again, we see that despite channel 2 containing essentially no information about the signal, the modified frequentist  $\text{CL}_s$  including both channels depends on  $b_2$ , while  $\text{CL}_{\text{excl}}$  is almost exactly flat, conforming to intuitive expectation.

Another study case is shown in the first panel of Fig. 12, with:

$$n_1 = 1, \quad b_1 = 1, \quad s_1 = 4, \quad (2.41)$$

$$n_2 = 1, \quad b_2 = 1, \quad s_2 = \text{varying}. \quad (2.42)$$

The variation of exclusion significances as a function of  $s_2$  is shown in the first panel of Fig. 12. For  $s_2 = s_1 = 4$  exactly, the results satisfy  $\text{CL}_s = \text{CL}_{\text{excl}}$  and agree precisely with the result that would be obtained for a single combined channel with  $n = 2, b = 2, s = 8$ . However, the  $Z$  value for  $\text{CL}_s$  has a small discontinuity at exactly  $s_2 = 4$ , such that for all other values of  $s_2$ ,  $\text{CL}_s$  has a higher exclusion significance  $Z$  than  $\text{CL}_{\text{excl}}$ . Numerically:

$$\text{CL}_s = 0.004093, \quad (Z = 2.644), \quad (\text{for } s_2 = 4) \quad (2.43)$$

$$\text{CL}_s = 0.003616, \quad (Z = 2.686), \quad (\text{for } s_2 = 4 \pm \epsilon), \quad (2.44)$$

for  $\epsilon$  arbitrarily small but nonzero. This discontinuity can be traced to the fact that for  $s_2 = 4$  exactly, the weights satisfy  $w_1 = w_2$  exactly for the two channels, which affects which integers are summed over due to Eq. (2.35). There are also discontinuities in  $\text{CL}_s$  at  $s_2 = \sqrt{5} - 1 \approx 1.23607$ , where  $w_1 = 2w_2$ , and at  $s_2 = 5^{1/3} - 1 \approx 0.709976$ , where  $w_1 = 3w_2$ , etc.

For another case study, consider:

$$n_1 = 2, \quad b_1 = 2.4, \quad s_1 = 8.5, \quad (2.45)$$

$$n_2 = 4, \quad b_2 = 2.3, \quad s_2 = \text{varying}. \quad (2.46)$$

The results are depicted in the second panel of Fig. 12, and show more pronounced discontinuities in both frequentist  $p_{\text{excl}}$  and the modified frequentist  $\text{CL}_s$ . In contrast, the Bayesian result  $\text{CL}_{\text{excl}}$  is smooth as we vary  $s_2$ , and gives more conservative exclusion significances.

Let us now consider the question of projecting expected exclusion significances for future experiments. In the multichannel case, one can define Asimov results for  $p_{\text{excl}}$  and  $\text{CL}_s$  by replacing each  $n_i$  in Eqs. (2.28) and (2.31) by the mean expected result  $b_i$  in the restriction Eq. (2.35). However, in the multichannel case, the resulting sets of  $\{k_i\}$  that contribute to the sums will depend discontinuously on the  $\{s_i\}$  and  $\{b_i\}$ , leading to the same sort of sawtooth problems that occurs in the median expected significance. In particular, an increase in the backgrounds often leads, counterintuitively, to a larger

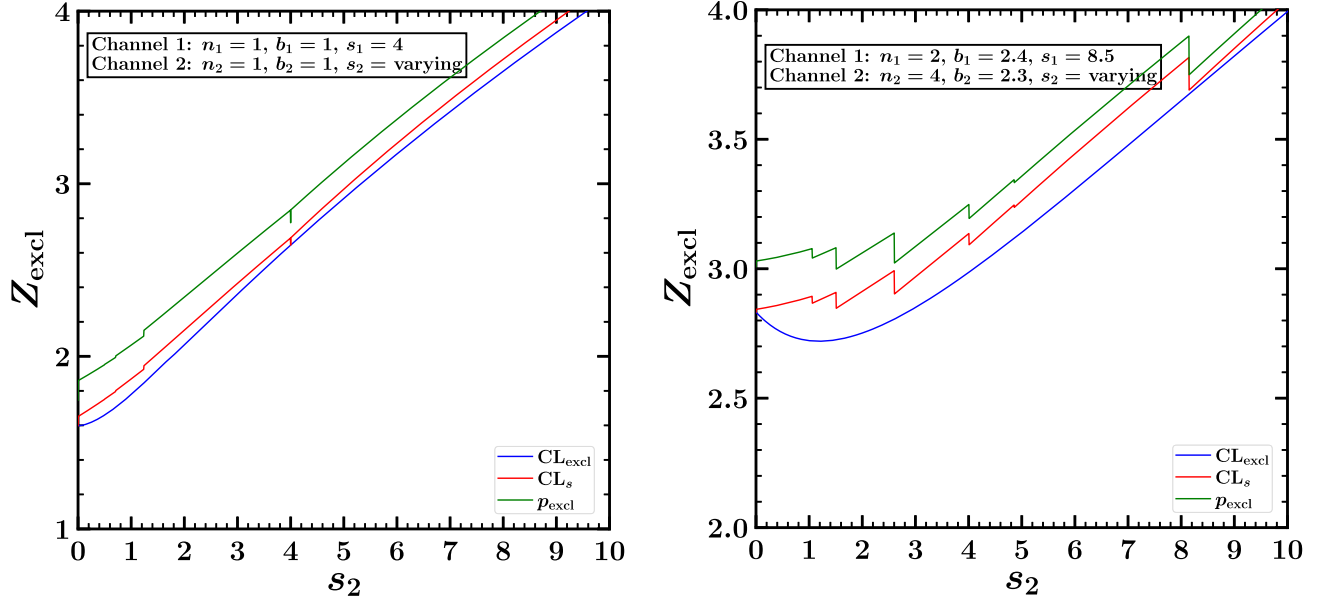


FIG. 12. Comparison of exclusion significances  $Z$  obtained from  $\text{CL}_s$  [red line, from Eqs. (2.28)–(2.31) and (2.35)] and  $\text{CL}_{\text{excl}}$  [blue line, from Eq. (2.36)] and  $p_{\text{excl}}$  [green line, from Eqs. (2.28) and (2.35)], for the test cases of Eqs. (2.41) and (2.42) [left panel] and (2.45) and (2.46) [right panel]. The results for  $p_{\text{excl}}$  and  $\text{CL}_s$  exhibit discontinuities as  $s_2$  is varied, due to abrupt changes in which outcomes  $k_1$  and  $k_2$  are summed over. The Bayesian version  $\text{CL}_{\text{excl}}$  does not have such discontinuities.

expected significance. (This problem did not occur in the single-channel case, because the sum  $\sum_{k=0}^n$  was evaluated in closed form in terms of incomplete  $\Gamma$  functions, after which the argument  $n$  could be interpreted as a continuous real number rather than an integer.) In contrast, if one uses  $\text{CL}_{\text{excl}}(\vec{n}, \vec{b}, \vec{s})$ , then the exact Asimov

method is perfectly straightforward and continuous, since it does not involve sums over integers subject to restrictions. Thus one can simply replace  $n_i$  by  $b_i$  in Eq. (2.36) to obtain the exact Asimov result. The Asimov results for  $p_{\text{excl}}$ ,  $\text{CL}_s$ , and  $\text{CL}_{\text{excl}}$  are compared in Fig. 13 for two test cases, showing the sawtooth behavior of the first two and

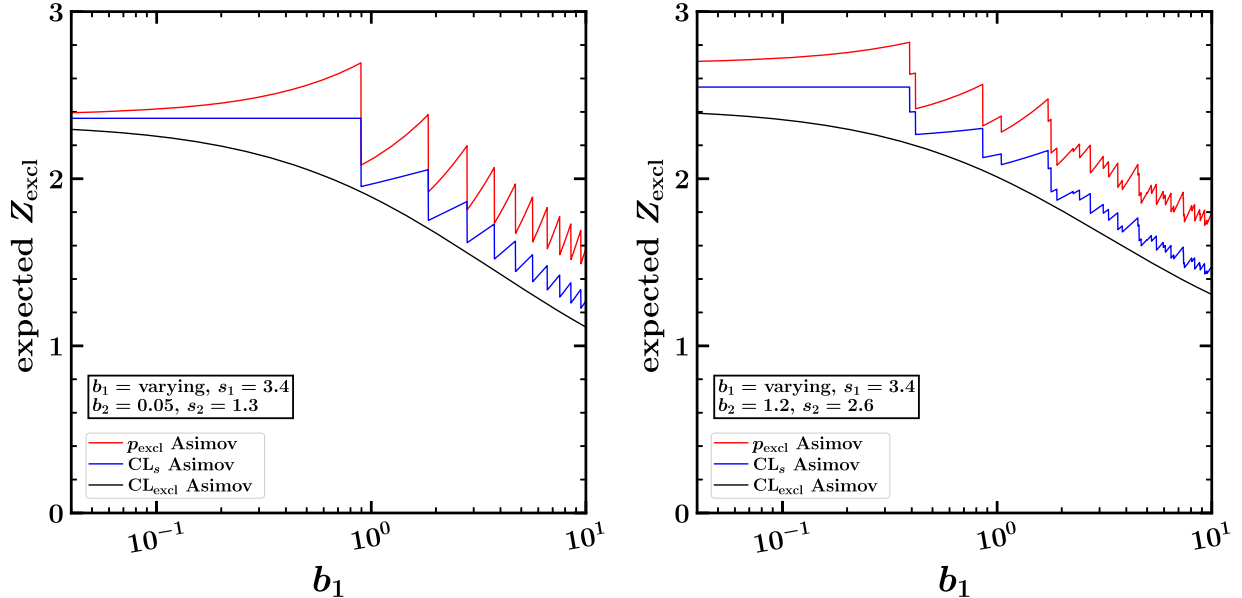


FIG. 13. Comparison of the Asimov expected exclusion significances  $Z$  obtained from  $p_{\text{excl}}$  [red lines, from Eq. (2.28)],  $\text{CL}_s$  [blue lines, from Eq. (2.31)], and  $\text{CL}_{\text{excl}}$  [black lines, from Eq. (2.36)], for two test cases with two independent channels, as labeled. The Asimov results are obtained by setting  $n_i = b_i$  for each channel. Due to the noncontinuous effect of the restriction of Eq. (2.35), the Asimov  $p_{\text{excl}}$  and  $\text{CL}_s$  have a counterintuitive nonmonotonic behavior as the first channel background mean  $b_1$  is varied, while the Asimov  $\text{CL}_{\text{excl}}$  is monotonic in the expected way.

the smooth, monotonic (and more conservative) behavior of the latter.

In view of the preceding discussion, we propose  $\text{CL}_{\text{excl}}$  in Eq. (2.36) as the preferred statistic for exclusion for multichannel counting experiments. Unlike  $p_{\text{excl}}$  and  $\text{CL}_s$  (with which it coincides in the single-channel case), it does not suffer from the problem of being affected significantly by the presence of a bad channel, and does not have discontinuities when signal and background means are changed infinitesimally. The exact Asimov result is straightforward to obtain and behaves continuously and monotonically in the expected way with respect to changes in the background. Furthermore, the introduction of background uncertainties and probability distributions for nuisance parameters is more straightforward, avoiding discontinuities in the integrand, as we will see below.

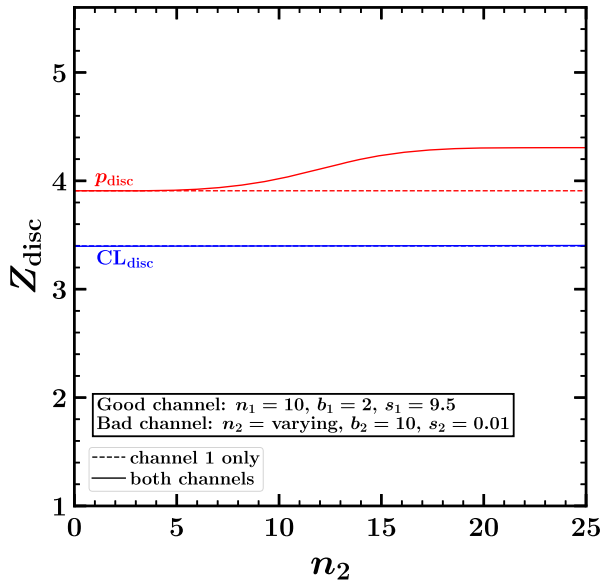
#### D. Discovery for multichannel counting experiments

For the discovery case, the frequentist  $p$ -value is defined by

$$p_{\text{disc}}(\vec{n}, \vec{b}, \vec{s}) = \sum_{\{k_i\}} \prod_{i=1}^N P(k_i | b_i). \quad (2.47)$$

The sum over  $\{k_i\}$  is restricted by the condition that the test-statistic  $\ln(q)$  defined by Eq. (2.34) is not smaller for  $\{k_i\}$  than for the observed data  $\{n_i\}$ , so:

$$\sum_{i=1}^N (n_i - k_i) \ln(1 + s_i/b_i) \leq 0. \quad (2.48)$$



Unlike the single-channel special case,  $p_{\text{disc}}$  depends on the signal strengths  $s_i$  when there is more than one channel because of this restriction. Note that the inequality has the opposite sense compared to the exclusion case, Eq. (2.35).

A more conservative, and simpler, alternative to  $p_{\text{disc}}(\vec{n}, \vec{b}, \vec{s})$  is the generalization of Eq. (2.15),

$$\text{CL}_{\text{disc}}(\vec{n}, \vec{b}, \vec{s}) = \prod_{i=1}^N \frac{P(n_i | b_i)}{P(n_i | s_i + b_i)}. \quad (2.49)$$

In order to compare these criteria for discovery, we first consider a case with one good channel and one bad channel, starting from the following values:

$$b_1 = 2, \quad s_1 = 9.5, \quad n_1 = 10, \quad (\text{good channel}), \quad (2.50)$$

$$b_2 = 10, \quad s_2 = 0.01, \quad n_2 = 10, \quad (\text{bad channel}). \quad (2.51)$$

In Fig. 14, we show the results for the discovery significance  $Z$  obtained from  $p_{\text{disc}}$  and  $\text{CL}_{\text{disc}}$ , considering variations in both  $n_2$  and  $b_2$  as the other quantities are held fixed, and compare to the same results using only channel 1. As in the exclusion case, we note that  $p_{\text{disc}}$  is affected in a nontrivial way by the presence of the bad channel, contrary to intuitive expectations. The step function discontinuities in  $p_{\text{disc}}$  are not a numerical artifact, but occur at values of  $b_2$  such that the ratio of weights  $w_1/w_2 = \ln(1 + s_1/b_1)/\ln(1 + s_2/b_2)$  is a rational number, so that

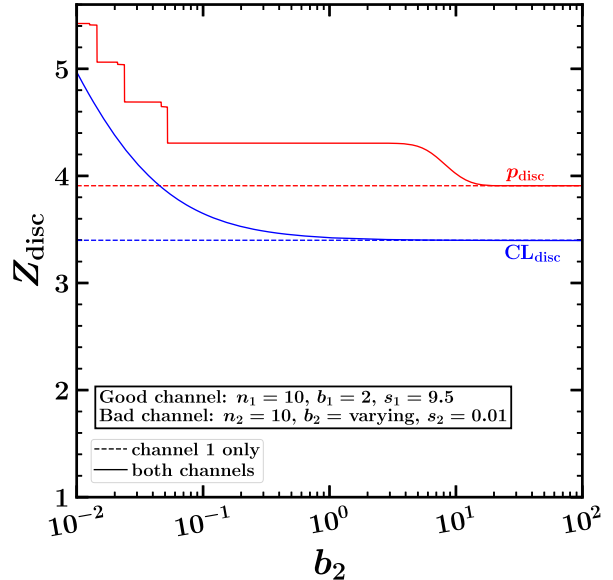


FIG. 14. Comparison of discovery significance  $Z$  in the case of a counting experiment with a good channel and a bad channel. The solid lines are obtained from  $p_{\text{disc}}$  [red lines, from Eqs. (2.47) and (2.48)] and  $\text{CL}_{\text{disc}}$  [blue lines, from Eq. (2.49)]. The dashed lines are obtained in the same way, but considering only the data from channel 1. In this example,  $\text{CL}_{\text{disc}}$  is more resistant to the effects of the noninformative channel, except in the case that  $b_2$  is very small. The step function discontinuities in  $p_{\text{disc}}$  in the right panel are not numerical artifacts, but occur at values of  $b_2$  such that the ratio of weights  $w_1/w_2 = \ln(1 + s_1/b_1)/\ln(1 + s_2/b_2)$  is rational.

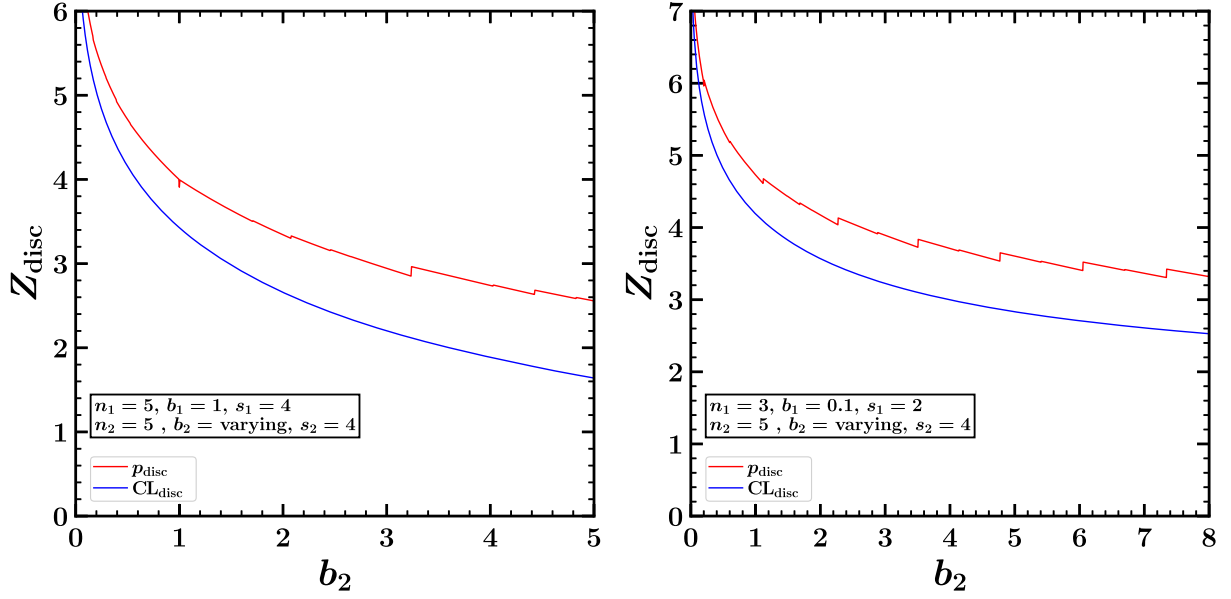


FIG. 15. Comparison of significances  $Z$  for discovery, obtained using  $p_{\text{disc}}$  [red lines, from Eqs. (2.47) and (2.48)] and  $\text{CL}_{\text{disc}}$  [blue lines, from Eq. (2.49)], for two 2-channel test cases with data as labeled.

the integer number of terms appearing in the  $\sum_{\{k_i\}}$  in Eq. (2.47) changes discontinuously.

In contrast,  $\text{CL}_{\text{disc}}$  is seen to be much less affected by the presence of the bad channel. The reason for this is that for any channel  $i$  with very small  $s_i$ , the numerator and denominator factors for that channel will cancel in the limit  $s_i/b_i \rightarrow 0$  in Eq. (2.49). The exception (in the right panel of Fig. 14) occurs in the case that  $b_2$  is also small, in which

case  $n_2 = 10$  is a surprising outcome for both the background-only and background + signal hypotheses.

Further comparisons between the significances obtained from  $p_{\text{disc}}$  and  $\text{CL}_{\text{disc}}$  for two test cases are shown in Fig. 15. The results obtained from  $p_{\text{disc}}$  have numerous discontinuities, which are small numerically but have the disturbing property of being nonmonotonic as the background  $b_2$  is varied. The results from  $\text{CL}_{\text{disc}}$  are reliably

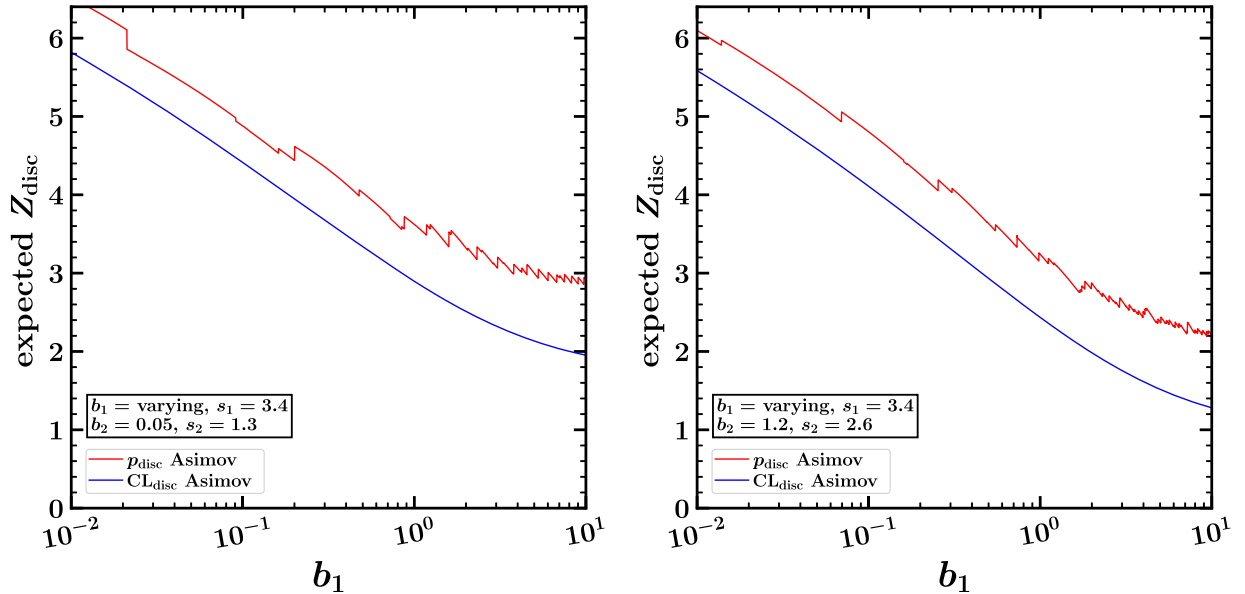


FIG. 16. Comparison of the Asimov expected discovery significances  $Z$  obtained from  $p_{\text{disc}}$  [red lines, from Eqs. (2.47) and (2.48)] and  $\text{CL}_{\text{disc}}$  [blue lines, from Eq. (2.49)], for two test cases with two independent channels, as labeled. The Asimov results are obtained by setting  $n_i = b_i + s_i$  for each channel. Due to the noncontinuous effect of the restriction of Eq. (2.48), the Asimov  $p_{\text{disc}}$  has a counterintuitive nonmonotonic behavior as the first channel background mean  $b_1$  is varied, while the Asimov  $\text{CL}_{\text{disc}}$  is monotonic in the expected way, and more conservative.



more conservative, as we have already noted, and do not suffer discontinuities because there is no restricted sum over integers in its definition.

For the purpose of projecting discovery prospects in future experiments, one can again define the Asimov values of  $p_{\text{disc}}$  and  $\text{CL}_{\text{disc}}$  by replacing  $n_i$  with  $b_i + s_i$  in Eqs. (2.47) and (2.49) respectively. These are compared for two test cases in Fig. 16. In the case of  $p_{\text{disc}}$ , the constraint put on the sum by Eq. (2.48) leads to a nonmonotonic sawtooth behavior, although much less pronounced than in the exclusion case in Fig. 13.

For the reasons just discussed, and because of the ease of generalization to the case of background uncertainties as discussed in the next section, we propose to use  $\text{CL}_{\text{disc}}$  as the figure of merit for the significance of a possible discovery, and for projecting the discovery reach of future experiments.

### E. Background uncertainty and other nuisance parameters

In the real world, the background level is never perfectly known. Furthermore, the background and signal may depend on other nuisance parameter(s), to be called  $\nu$  below. These can be dealt with in a Bayesian approach by assuming probability densities  $f(b)$  and  $g(\nu)$ , subject to the normalization conditions  $\int_0^\infty db f(b) = 1$  and  $\int d\nu g(\nu) = 1$ .

For example, following [50], we can model the background uncertainty in terms of an on-off problem [68–73], where  $m$  is the number of Poisson events in a signal-off (background-only) region, and the ratio of background means in the signal-off and signal-on regions is called  $\tau$ . In terms of  $m$  and  $\tau$ , the point estimate for the background and its variance are

$$\hat{b} = m/\tau, \quad \Delta_b = \sqrt{m}/\tau, \quad (2.52)$$

or equivalently

$$\tau = \hat{b}/\Delta_b^2, \quad m = \hat{b}^2/\Delta_b^2, \quad (2.53)$$

so that the probability density of  $b$  is

$$f(b) = f(b|\hat{b}, \Delta_b) = \tau^{m+1} b^m e^{-\tau b} / m!, \quad (2.54)$$

the posterior probability distribution for  $b$  obtained by using Bayes' theorem with Poisson likelihood for background in the signal-off region  $P(m|\tau b)$  and flat prior for  $b$ . Note that this probability distribution can be used as a model even in situations where the estimates of the background and its uncertainty come partly or completely from theory rather than some signal-off region data.

In the case of Eq. (2.54), the probability for observing  $n$  events in the signal-on region is obtained by averaging over  $b$  [71–75] to obtain

$$\Delta P(n, \hat{b}, \Delta_b, s) = \int_0^\infty db f(b|\hat{b}, \Delta_b) P(n|s+b), \quad (2.55)$$

We can then extend the definitions of frequentist  $p$ -values and to the uncertain background case by simply replacing the Poisson probability  $P(n|s+b)$  with  $\Delta P(n, \hat{b}, \Delta_b, s)$  [51]:

$$p_{\text{excl}}(n, \hat{b}, \Delta_b, s) = \sum_{k=0}^n \Delta P(k, \hat{b}, \Delta_b, s). \quad (2.56)$$

$$p_{\text{disc}}(n, \hat{b}, \Delta_b) = \sum_{k=n}^\infty \Delta P(k, \hat{b}, \Delta_b, 0), \quad (2.57)$$

Explicit formulas for  $\Delta P(n, \hat{b}, \Delta_b, s)$ ,  $p_{\text{excl}}(n, \hat{b}, \Delta_b, s)$ , and  $p_{\text{disc}}(n, \hat{b}, \Delta_b)$  can be found in Eqs. (12)–(15) of Ref. [50]. Besides these, we note the simple formula:

$$p_{\text{excl}}(n, \hat{b}, \Delta_b, 0) = \frac{B(1/(1 + \Delta_b^2/\hat{b}), m+1, n+1)}{B(m+1, n+1)}. \quad (2.58)$$

Similarly, the confidence levels discussed in the previous sections can be obtained in the uncertain background case as

$$\text{CL}_s(n, \hat{b}, \Delta_b, s) = \frac{p_{\text{excl}}(n, \hat{b}, \Delta_b, s)}{p_{\text{excl}}(n, \hat{b}, \Delta_b, 0)}, \quad (2.59)$$

$$\begin{aligned} \text{CL}_{\text{excl}}(n, \hat{b}, \Delta_b, s) &= \frac{\int_s^\infty ds' \Delta P(n, \hat{b}, \Delta_b, s')}{\int_0^\infty ds' \Delta P(n, \hat{b}, \Delta_b, s')} \\ &= \frac{p_{\text{excl}}(n, \hat{b}, \Delta_b, s)}{p_{\text{excl}}(n, \hat{b}, \Delta_b, 0)}, \end{aligned} \quad (2.60)$$

$$\text{CL}_{\text{disc}}(n, \hat{b}, \Delta_b, s) = \frac{\Delta P(n, \hat{b}, \Delta_b, 0)}{\Delta P(n, \hat{b}, \Delta_b, s)}. \quad (2.61)$$

Note that we retain the property  $\text{CL}_{\text{excl}} = \text{CL}_s$  in the single-channel case with nonzero background uncertainty.

The exact Asimov expectations for  $p_{\text{excl}}$ ,  $\text{CL}_s = \text{CL}_{\text{excl}}$ , and  $p_{\text{disc}}$ ,  $\text{CL}_{\text{disc}}$  in the uncertain background case are obtained by replacing  $n$  in the preceding equations by its expected mean in each case:

$$\langle n_{\text{excl}} \rangle = \sum_{n=0}^\infty n \Delta P(n, \hat{b}, \Delta_b, 0) = \hat{b} + \Delta_b^2/\hat{b}, \quad (2.62)$$

$$\langle n_{\text{disc}} \rangle = \sum_{n=0}^\infty n \Delta P(n, \hat{b}, \Delta_b, s) = s + \hat{b} + \Delta_b^2/\hat{b}, \quad (2.63)$$

for exclusion and discovery, respectively.

More generally, for any probability distributions  $f(\vec{b})$  and  $g(\nu)$  for the background and other nuisance parameters, one can marginalize (integrate) over  $b_i$  and  $\nu$ . In the case of exclusion, Eq. (2.28) generalizes to

$$p_{\text{excl}} = \int d\nu g(\nu) \int d\vec{b} f(\vec{b}) \sum_{\{k_i\}} \prod_{i=1}^N P(k_i | s_i + b_i), \quad (2.64)$$

and similarly for Eq. (2.30), which then gives  $\text{CL}_s$ . However, note that the sum  $\sum_{\{k_i\}}$  is subject to the restriction Eq. (2.35), so that even the numbers of terms in the sum depends in a discontinuous way on  $\nu$  and  $b_i$  as we integrate over them in the multichannel case. Reference [76] contains a discussion of various ways to account for the uncertainties in the background and nuisance parameters in the frequentist methods. As argued above, we prefer instead to generalize Eq. (2.36), resulting in:

$$\text{CL}_{\text{excl}} = \frac{1}{D} \int d\nu g(\nu) \int d\vec{b} f(\vec{b}) \times \int_s^\infty ds' \prod_{i=1}^N P(n_i | r_i s' + b_i). \quad (2.65)$$

Here we have used a shorthand notation to be used several times below, such that the normalization factor  $D$  is equal to the expression that follows it with  $s = 0$ .

Similarly, in the case of discovery in the presence of background uncertainties and nuisance parameters, we can generalize Eq. (2.47) to obtain

$$p_{\text{disc}} = \int d\nu g(\nu) \int d\vec{b} f(\vec{b}) \sum_{\{k_i\}} \prod_{i=1}^N P(k_i | b_i), \quad (2.66)$$

this time subject to the constraint Eq. (2.48) on the terms in the sum. However, as argued above, we prefer to use the more conservative

$$\text{CL}_{\text{disc}} = \frac{\int d\nu g(\nu) \int d\vec{b} f(\vec{b}) \prod_{i=1}^N P(n_i | b_i)}{\int d\nu g(\nu) \int d\vec{b} f(\vec{b}) \prod_{i=1}^N P(n_i | s_i + b_i)}. \quad (2.67)$$

To obtain the Asimov results, one can substitute in the mean expected values for  $n_i$ , namely

$$\langle n_{i,\text{excl}} \rangle = \int d\nu g(\nu) \int_0^\infty d\vec{b} f(\vec{b}) \sum_{n_i=0}^\infty n_i P(n_i | b_i), \quad (2.68)$$

$$\langle n_{i,\text{disc}} \rangle = \int d\nu g(\nu) \int_0^\infty d\vec{b} f(\vec{b}) \sum_{n_i=0}^\infty n_i P(n_i | s_i + b_i). \quad (2.69)$$

### III. APPLICATION TO PROTON DECAY

In this section, we will first consider the application of the Bayesian statistic  $\text{CL}_{\text{excl}}$  to estimate the current lower limits on proton partial lifetimes in  $p \rightarrow \bar{\nu} K^+$  and  $p \rightarrow e^+ \pi^0$  modes, based on Super-Kamiokande's data, at various confidence levels generalizing the 90% C.L. published limits. We will then consider the prospects for exclusion or discovery of these proton decay modes for several planned future neutrino experiments: DUNE [46], JUNO [47], Hyper-Kamiokande [48], and THEIA [49]. We do this by applying the Bayesian approach of using  $\text{CL}_{\text{excl}}$  and  $\text{CL}_{\text{disc}}$  with the exact Asimov criterion of replacing the observed counts by their respective expected means.

As discussed above, the Bayesian approaches  $\text{CL}_{\text{excl}}$  for exclusion and  $\text{CL}_{\text{disc}}$  for discovery are ideal methods to obtain these limits and projections, as they: (1) guard against claiming exclusion (or discovery) when an experiment is actually not sensitive to the signal model, and therefore are more conservative than the frequentist  $p_{\text{excl}}$  and  $p_{\text{disc}}$ ; (2) are well-behaved in multichannel counting experiments in the sense that, unlike the (modified) frequentist approach,  $\text{CL}_{\text{excl}}$  and  $\text{CL}_{\text{disc}}$  are not overly affected by the presence of noninformative channels and do not have any discontinuities as the signal and background means are varied; and (3) are easily able to include uncertainties in the backgrounds and the signal selection efficiencies, especially for multichannel counting experiments.

The estimates for the backgrounds and the signal selection efficiencies in a specific proton decay mode have been obtained by the DUNE, JUNO, and THEIA collaborations by modeling the experiments as single-channel counting experiments, whereas Hyper-Kamiokande searches for proton decay are modeled as multichannel counting experiments based on the signal regions and search strategies used at Super-Kamiokande. Before we present our results, we first review the methods we employ to obtain the limits/projections for proton partial lifetimes at single-channel and multichannel counting experiments, based on the methods elucidated in Sec. II.

The number of decays in a specific decay channel at an experiment with  $N_0$  initial number of protons for a runtime of  $\Delta t$  is given by

$$\Delta N = N_0 \Gamma \Delta t, \quad (3.1)$$

where the proton partial width  $\Gamma$  is extremely small. [More generally,  $\Delta N = N_0(1 - e^{-\Gamma \Delta t})$ .] Therefore the signal can be computed as

$$s = \epsilon(\Delta N) = \epsilon N_0 \Gamma \Delta t, \quad (3.2)$$

where  $0 \leq \epsilon \leq 1$  is the signal selection efficiency. In terms of the number of protons per kiloton of detector material  $N_p$  and the exposure  $\lambda$  ( $=$  runtime  $\times$  number of

kilotons of detector material) of the experiment in units of kiloton-years, we can reexpress Eq. (3.2) as

$$s = \Gamma N_p \epsilon \lambda. \quad (3.3)$$

The present exclusion limit at confidence level  $1 - \alpha$  for the proton partial lifetime is then provided by [77]

$$\tau_p = 1/\Gamma = N_p \epsilon \lambda / s, \quad (3.4)$$

where  $s$  is the number of signal events that gives  $\text{CL}_{\text{excl}}$  equal to  $\alpha$ . For a future experiment, the exclusion reach for the proton partial lifetime at confidence level  $1 - \alpha$  is given by the same formula Eq. (3.4), where  $s$  is now the signal that makes the exact Asimov  $\text{CL}_{\text{excl}}^A$  equal to  $\alpha$ . The discovery reach for a given significance  $Z$  is likewise obtained from Eq. (3.4) using the  $s$  that provides for  $\text{CL}_{\text{disc}}^A = \frac{1}{2} \text{erfc}(Z/\sqrt{2})$ .

Equation (3.4) holds for an experiment with a single search channel with known background  $b$  and signal selection efficiency  $\epsilon$ . For the more general case of an experiment with one or more independent search channels with possibly uncertain backgrounds and signal efficiencies, we employ a Bayesian approach to obtain the limit/reach for proton partial lifetime, as discussed above. First, for the exclusion case, given the number of observed events  $n_i$  in each search channel labeled  $i$ , the upper limit on proton partial width at a confidence level  $1 - \alpha$  is obtained by solving for  $\Gamma$  in (see Eq. (2.65), and Ref. [78]):

$$\alpha = \frac{1}{D} \int_{\Gamma} d\Gamma' \prod_{i=1}^N \int_0^1 d\epsilon_i g(\epsilon_i) \int_0^{\infty} db_i f(b_i) P(n_i | s'_i + b_i). \quad (3.5)$$

Here,  $D$  is a normalization factor, defined to equal the expression that follows it evaluated at  $\Gamma = 0$ , and in each search channel labeled by  $i$ , the signal rate is

$$s'_i = N_p \epsilon_i \lambda_i \Gamma', \quad (3.6)$$

and  $g(\epsilon_i)$  and  $f(b_i)$  are the probability distributions for the signal efficiency  $\epsilon_i$  and the background  $b_i$ . These distributions can take different forms to parameterize our lack of perfect knowledge of the efficiency and background, such that  $\int_0^1 d\epsilon_i g(\epsilon_i) = 1$  and  $\int_0^{\infty} db_i f(b_i) = 1$ . For example, the probability distribution of true signal selection efficiency  $\epsilon_i$  might be taken to be a truncated Gaussian distribution with central value  $\hat{\epsilon}_i$  and variance  $\Delta_{\epsilon_i}$ , as in the Super-Kamiokande search analyses in Refs. [44,45]:

$$g(\epsilon_i | \hat{\epsilon}_i, \Delta_{\epsilon_i}) = \sqrt{\frac{2}{\pi}} \frac{\exp\left[-\frac{(\epsilon_i - \hat{\epsilon}_i)^2}{2\Delta_{\epsilon_i}^2}\right]}{\Delta_{\epsilon_i} \left( \text{erf}\left(\frac{1 - \hat{\epsilon}_i}{\sqrt{2}\Delta_{\epsilon_i}}\right) + \text{erf}\left(\frac{\hat{\epsilon}_i}{\sqrt{2}\Delta_{\epsilon_i}}\right) \right)}. \quad (3.7)$$

The probability distribution of true background  $b_i$  in the  $i$ th search channel  $f(b_i | \hat{b}_i, \Delta_{b_i})$  can be taken to be given by Eq. (2.54) as in the on-off problem, in terms of quantities  $m_i$  and  $\tau_i$ , related to the central value  $\hat{b}_i$  and variance  $\Delta_{b_i}$  by Eq. (2.52). Equation (3.5) assumes that the search channels are independent.

If the background and the signal selection efficiencies are perfectly known, i.e.,  $f(b_i | \hat{b}_i, \Delta_{b_i}) = \delta(b_i - \hat{b}_i)$  and  $g(\epsilon_i | \hat{\epsilon}_i, \Delta_{\epsilon_i}) = \delta(\epsilon_i - \hat{\epsilon}_i)$ , then we get

$$\alpha = \frac{1}{D} \int_{\Gamma} d\Gamma' \prod_i P(n_i | s'_i + b_i), \quad (3.8)$$

with  $s'_i = N_p \epsilon_i \lambda_i \Gamma'$  after identifying  $\hat{b}_i = b_i$  and  $\hat{\epsilon}_i = \epsilon_i$ . This corresponds to Eq. (2.36). Specializing further to a single search channel (dropping the subscript  $i$ ), this reduces to Eq. (2.14) with  $s' = N_p \epsilon \lambda \Gamma'$ .

For projecting the exclusion reach for partial lifetime at future experiments, we make use of the exact Asimov method by replacing the number of events  $n_i$  in each search channel by their respective expected means,

$$\langle b_i \rangle = \int_0^{\infty} db_i f(b_i) b_i, \quad (3.9)$$

for example  $\langle b_i \rangle = (m_i + 1)/\tau_i = \hat{b}_i + \Delta_{b_i}^2/b_i$  if the on-off problem treatment is used for the background. The expected confidence level  $1 - \alpha$  upper limit on partial width  $\Gamma$  is then solved from Eq. (3.5) with  $n_i$  replaced by  $\langle b_i \rangle$ :

$$\alpha = \frac{1}{D} \int_{\Gamma} d\Gamma' \prod_i \int_0^1 d\epsilon_i g(\epsilon_i) \times \int_0^{\infty} db_i f(b_i) P(\langle b_i \rangle | s'_i + b_i). \quad (3.10)$$

Equation (3.10) gives the Asimov expected lower limit on the partial lifetime via  $\tau_p = 1/\Gamma$ .

For the expected discovery reach for proton partial widths at future experiments, we use a method based on the exact Asimov evaluation of the statistic  $\text{CL}_{\text{disc}}$ . In particular, we solve for  $\Gamma$  from [see Eq. (2.67)]

$$\frac{1}{2} \text{erfc}\left(\frac{Z}{\sqrt{2}}\right) = \frac{\prod_i \int_0^{\infty} db_i f(b_i) P(\langle n_i \rangle | b_i)}{\prod_i \int_0^1 d\epsilon_i g(\epsilon_i) \int_0^{\infty} db_i f(b_i) P(\langle n_i \rangle | s_i + b_i)}, \quad (3.11)$$

where  $s_i = N_p \lambda_i \epsilon_i \Gamma$  and  $\langle n_i \rangle = \langle s_i \rangle + \langle b_i \rangle$ , with  $\langle b_i \rangle$  as given in Eq. (3.9), and

$$\langle s_i \rangle = \Gamma N_p \lambda_i \int_0^1 d\epsilon_i g(\epsilon_i) \epsilon_i. \quad (3.12)$$

TABLE I. Super-Kamiokande's data for  $p \rightarrow \bar{\nu}K^+$  and  $p \rightarrow e^+\pi^0$  decay modes, taken from Refs. [44,45], respectively. In each decay mode, the exposures  $\lambda_i$  in kton-years, total backgrounds  $\hat{b}_i \pm \Delta_{b_i}$ , signal efficiencies  $\hat{\epsilon}_i \pm \Delta_{\epsilon_i}$ , and the observed number of counts  $n_i$  are listed.  $\langle s_i^{90\text{C.L.}} \rangle$  are the expected signal events, defined in Eq. (3.12), for proton partial lifetime set equal to its 90% C.L. lower limit. The last column gives a brief description of each of the channels referring to the detector period (SK I-IV) and the name of the search method used in Refs. [44,45].

Decay mode	$\lambda_i$	$\hat{b}_i \pm \Delta_{b_i}$	$\hat{\epsilon}_i \pm \Delta_{\epsilon_i}$ [%]	$n_i$	$\langle s_i^{90\text{C.L.}} \rangle$	Comment
$p \rightarrow \bar{\nu}K^+$	91.7	$0.08 \pm 0.02$	$7.9 \pm 0.1$	0	0.37	SK-I, prompt $\gamma$
		$0.18 \pm 0.04$	$7.8 \pm 0.1$	0	0.36	SK-I, $\pi^+\pi^0$
		$193.21 \pm 3.58$	$33.9 \pm 0.3$	177	1.57	SK-I, $p_\mu$ spectrum
	49.2	$0.14 \pm 0.03$	$6.3 \pm 0.1$	0	0.16	SK-II, prompt $\gamma$
		$0.17 \pm 0.03$	$6.7 \pm 0.1$	0	0.17	SK-II, $\pi^+\pi^0$
		$94.27 \pm 1.72$	$30.6 \pm 0.3$	78	0.76	SK-II, $p_\mu$ spectrum
	31.9	$0.03 \pm 0.01$	$7.7 \pm 0.1$	0	0.12	SK-III, prompt $\gamma$
		$0.09 \pm 0.01$	$7.9 \pm 0.1$	0	0.13	SK-III, $\pi^+\pi^0$
		$69.00 \pm 1.28$	$32.6 \pm 0.3$	85	0.53	SK-III, $p_\mu$ spectrum
	87.3	$0.13 \pm 0.03$	$9.1 \pm 0.1$	0	0.4	SK-IV, prompt $\gamma$
		$0.18 \pm 0.03$	$10.0 \pm 0.1$	0	0.44	SK-IV, $\pi^+\pi^0$
		$223.14 \pm 4.10$	$37.6 \pm 0.3$	226	1.66	SK-IV, $p_\mu$ spec.
$p \rightarrow e^+\pi^0$	111.4	$0.01 \pm 0.01$	$18.3 \pm 1.7$	0	0.28	SK-I, lower
		$0.15 \pm 0.06$	$20.0 \pm 3.3$	0	0.3	SK-I, upper
	59.4	$0.01 \pm 0.01$	$16.6 \pm 1.7$	0	0.13	SK-II, lower
		$0.11 \pm 0.04$	$19.4 \pm 3.0$	0	0.16	SK-II, upper
	38.6	0.01	$18.7 \pm 1.7$	0	0.1	SK-III, lower
		$0.07 \pm 0.03$	$20.3 \pm 3.3$	0	0.11	SK-III, upper
	241.3	0.01	$18.2 \pm 1.5$	0	0.6	SK-IV, lower
		$0.25 \pm 0.11$	$19.2 \pm 3.1$	0	0.63	SK-IV, upper

This gives the expected discovery reach for partial lifetime using  $\tau_p = 1/\Gamma$  corresponding to a chosen significance  $Z$ .

Based on Super-Kamiokande's data, taken from Refs. [44,45], which we quote for completeness in Table I, we now compute the upper limit on proton partial widths in the  $p \rightarrow \bar{\nu}K^+$  and  $p \rightarrow e^+\pi^0$  decay modes that are excluded at various confidence levels (e.g., 95%, 90%, 68%, 50% C.L.) using Eq. (3.5), which can then be translated into corresponding lower limits on the proton partial lifetime. Super-Kamiokande uses a water Cerenkov detector with a fiducial mass of 22.5 ktons, and the analysis for  $p \rightarrow e^+\pi^0$  in Ref. [45] also includes data from an enlarged fiducial mass of 27.2 ktons. While Super-Kamiokande can probe for proton decay in both  $p \rightarrow \bar{\nu}K^+$  and  $p \rightarrow e^+\pi^0$  decay modes, it is less sensitive to the former decay mode, because the  $K^+$  is produced below its Cerenkov threshold in water and is only identified from its decay constituents. Figure 17 shows our own computed estimates of the current confidence levels for the exclusion of proton decay at Super-Kamiokande in  $p \rightarrow \bar{\nu}K^+$  (left panel) and  $p \rightarrow e^+\pi^0$  (right panel) channels as a function of proton partial lifetime in the respective decay channels. This generalizes the results presented by the Super-Kamiokande collaboration, which gave results only for 90% C.L. exclusions.

From the data in Table I, we estimated the current lower limits on proton partial lifetimes to be

$$\tau_p/\text{Br}(p \rightarrow \bar{\nu}K^+) > \begin{cases} 5.1 \times 10^{33} \text{ years at 95\% C.L.,} \\ 6.6 \times 10^{33} \text{ years at 90\% C.L.,} \\ 1.3 \times 10^{34} \text{ years at 68\% C.L.,} \\ 2.2 \times 10^{34} \text{ years at 50\% C.L.,} \end{cases} \quad (3.13)$$

and

$$\tau_p/\text{Br}(p \rightarrow e^+\pi^0) > \begin{cases} 1.9 \times 10^{34} \text{ years at 95\% C.L.,} \\ 2.4 \times 10^{34} \text{ years at 90\% C.L.,} \\ 4.9 \times 10^{34} \text{ years at 68\% C.L.,} \\ 8.1 \times 10^{34} \text{ years at 50\% C.L.} \end{cases} \quad (3.14)$$

In comparison, the published 90% C.L. exclusion limit on proton partial lifetimes from the Super-Kamiokande collaboration are



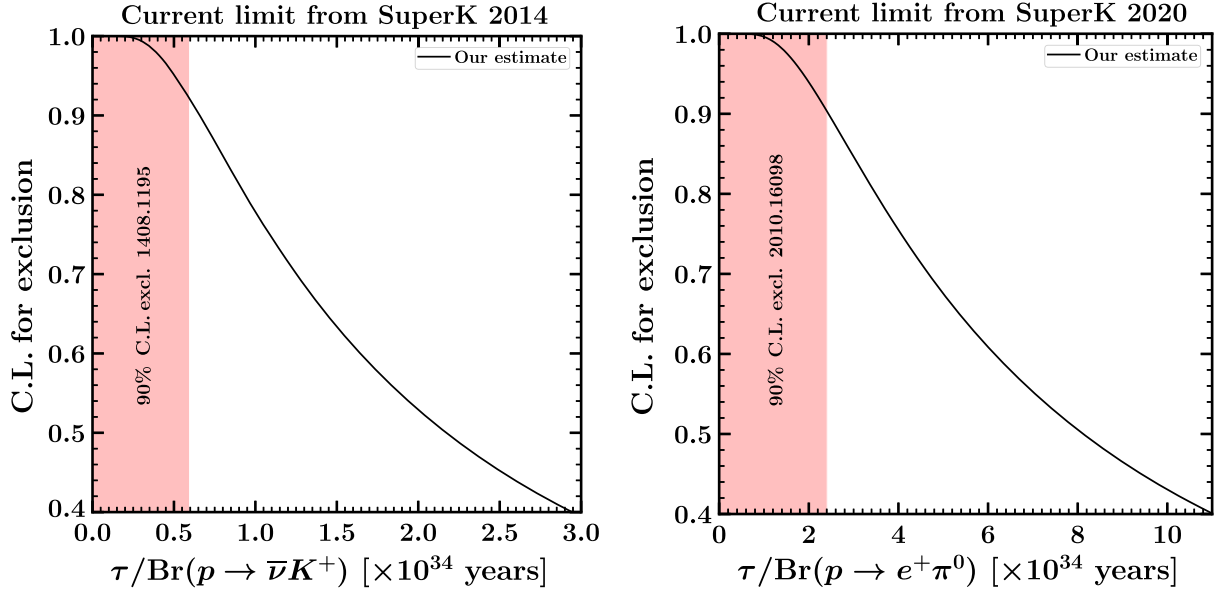


FIG. 17. The current confidence level for the exclusion of proton decay in  $p \rightarrow \bar{\nu}K^+$  (left panel) and  $p \rightarrow e^+\pi^0$  (right panel) channels, as a function of the respective proton partial lifetimes. Our confidence level estimates (solid black lines) are obtained using Eq. (3.5) based on Super-Kamiokande's data through 2014 [44] (left panel) and 2020 [45] (right panel), summarized in Table I. The red shaded regions correspond to Super-Kamiokande's published exclusions on proton partial lifetimes at 90% C.L., from [44] and [45].

$$\begin{aligned} \tau_p/\text{Br}(p \rightarrow \bar{\nu}K^+) \\ > 5.9 \times 10^{33} \text{ years at 90\% C.L.}, \end{aligned} \quad (3.15)$$

$$\begin{aligned} \tau_p/\text{Br}(p \rightarrow e^+\pi^0) \\ > 2.4 \times 10^{34} \text{ years at 90\% C.L.} \end{aligned} \quad (3.16)$$

shown as the shaded red regions in Fig. 17. We see that in the case of  $p \rightarrow \bar{\nu}K^+$ , our<sup>11</sup> estimate for the 90% C.L. limit is slightly stronger ( $6.6 \times 10^{33}$  years rather than  $5.9 \times 10^{33}$  years) than the journal published limit in Ref. [44]. In this paper, we only consider the limits from data published in journal articles. In the case of  $p \rightarrow \bar{\nu}K^+$ , there is more data from the continuation of run SK-IV, which was not used for the published limit in Ref. [44]. It is therefore quite possible that a future limit, based on data already taken, will be stronger. In the case of  $p \rightarrow e^+\pi^0$ , our estimate for the 90% C.L. limit agrees perfectly with the Super-Kamiokande published limit in Ref. [45].

<sup>11</sup>Besides using the probability distribution for true background as in the on-off problem [Eq. (2.54)], we have considered various other distributions such as a Gaussian, and a convolution of Gaussian and Poisson (only for search channels with extremely low backgrounds) as done in Refs. [44,45], but there was no noticeable change in our results. In Super-Kamiokande's analysis for  $p \rightarrow \bar{\nu}K^+$  decay mode in Ref. [44], the search channels with large backgrounds that are referred to as “ $p_\mu$  spectrum” in Table I were further divided into subchannels, but due to insufficient data made available, we are not able to include that subdivision.

We now discuss projections for exclusion and discovery of proton decay at possible future neutrino detectors DUNE, JUNO, Hyper-Kamiokande, and THEIA. Both DUNE and JUNO will be primarily searching for proton decay in  $p \rightarrow \bar{\nu}K^+$  decay mode. For these searches, DUNE uses its far detector with a total of 40 kiloton (kton) fiducial mass of liquid argon [46] and can track and reconstruct charged kaons with high efficiency, and JUNO uses its central detector with a 20 kton fiducial mass of a liquid scintillator [47]. On the other hand, Hyper-Kamiokande [48] uses a water Cerenkov detector with 186 ktons of fiducial mass and is sensitive to both  $p \rightarrow \bar{\nu}K^+$  and  $p \rightarrow e^+\pi^0$  decay modes among others. As was the case with Super-Kamiokande, Hyper-Kamiokande will be more sensitive to the  $p \rightarrow e^+\pi^0$  mode, compared to the  $p \rightarrow \bar{\nu}K^+$  mode, due to much better reconstruction of the Cerenkov rings of the positron and the electromagnetic showers emanating from  $\pi^0 \rightarrow \gamma\gamma$ . THEIA is a new detector concept with water-based liquid scintillator (10% liquid scintillator and 90% water) that will be able to detect and distinguish between the Cerenkov and the scintillation light [49]. Here, we project sensitivities for both THEIA-25 and THEIA-100 with fiducial masses 17 and 80 ktons, respectively, that were considered in Ref. [49]. Due to the ability to detect scintillation signals from charged particles such as  $K^+$  produced below its Cerenkov threshold, and Cerenkov signals, the THEIA detector aims to have enhanced sensitivity to the  $p \rightarrow \bar{\nu}K^+$  mode [49] while also being able to probe the  $p \rightarrow e^+\pi^0$  mode [79]. The numbers of protons per kiloton of detector material are

$$N_p = \begin{cases} 2.71 \times 10^{32} & (\text{DUNE}), \\ 3.38 \times 10^{32} & (\text{JUNO}), \\ 3.34 \times 10^{32} & (\text{Hyper-Kamiokande}), \\ 3.35 \times 10^{32} & (\text{THEIA}). \end{cases} \quad (3.17)$$

For the purposes of projecting sensitivities for THEIA and JUNO, we took the liquid scintillator in both detectors to have  $6.75 \times 10^{33}$  protons per 20 kilotons based on Ref. [47].

Figure 18 shows the runtimes at DUNE that are required for an expected 90% C.L. exclusion (first panel) and  $Z = 3$

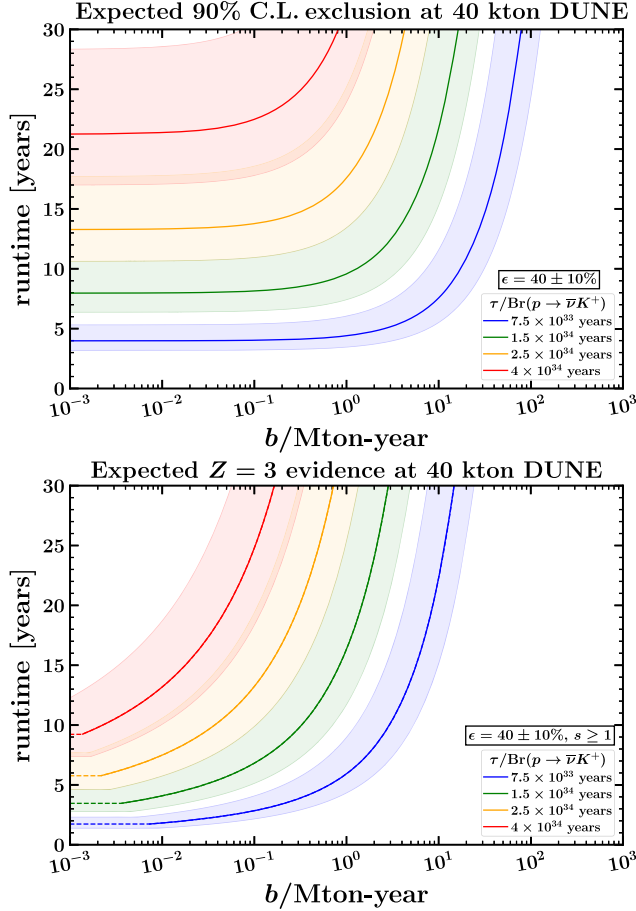


FIG. 18. The required runtimes at DUNE (with 40 kiloton fiducial mass of liquid argon) for an expected 90% C.L. exclusion (top panel) and  $Z = 3$  evidence (bottom panel) as a function of background rate per megaton-year of exposure, for various proton partial lifetimes in the  $p \rightarrow \bar{\nu}K^+$  channel, as labeled. The runtimes are computed using Eq. (3.18) where the signal needed for 90% C.L. exclusion ( $Z = 3$  evidence) is obtained from setting Eq. (2.19) [Eq. (2.21)] equal to 0.1 (0.00135). We also require  $s \geq 1$  in the bottom panel, which yields the horizontal dashed lines for very small  $b$  in the lower left corner. The solid lines (and dashed lines in the bottom panel) assume the signal selection efficiency  $\epsilon$  to be 40%, and the shaded bands encompassing them correspond to varying  $\epsilon$  by  $\pm 10\%$ .

evidence (second panel), in the  $p \rightarrow \bar{\nu}K^+$  decay mode, as a function of the background rate per megaton-year of exposure. The colored lines and bands correspond to various choices of proton partial lifetimes. For the purposes of illustration, we chose a signal selection efficiency  $\epsilon = 40 \pm 10\%$  that is plausible, based on various signal selection efficiencies that are considered in Refs. [67,80–84]. The solid lines in the figure assume  $\epsilon = 40\%$ , and the shaded bands surrounding them vary  $\epsilon$  by  $\pm 10\%$ . The required runtimes  $\Delta t$  in the figure are obtained using Eq. (3.4), which gives

$$\Delta t = \frac{s\tau_p}{N_p N_{\text{kton}} \epsilon}, \quad (3.18)$$

where  $N_{\text{kton}}$  is the number of kilotons of detector material, and  $s$  is the upper limit on signal for 90% C.L. exclusion obtained from setting  $\text{CL}_{\text{excl}}^A$  [as in Eq. (2.19)] equal to 0.1, or the signal needed for  $Z = 3$  evidence obtained from setting  $\text{CL}_{\text{disc}}^A$  [as in Eq. (2.21)] equal to 0.00135. As discussed at the end of Sec. II B, the zero background limit for the discovery case is not well defined, in a sense that at  $b = 0$ , any nonzero signal, albeit arbitrarily small, would yield an infinite significance. Therefore, to be conservative, we require that the mean expected number of signal events  $s$  is at least 1 in order to have an expected discovery. The dashed lines for very small  $b/\text{Mton-year}$  in the lower left corner of the bottom panel (for discovery case) of Fig. 18 correspond to this additional requirement that  $s \geq 1$ . It is clear from the figure that if the estimated background per megaton-year of exposure at DUNE increases, the required runtime increases more steeply for discovery than for exclusion.

In Fig. 19, we show the expected 90% C.L. exclusion reach (first panel) and the expected  $Z = 3$  evidence reach (second panel) for proton partial lifetime in  $p \rightarrow \bar{\nu}K^+$  decay channel at DUNE as a function of the runtime in years. The three colored lines/bands correspond to various assumed background rates per megaton-year of exposure taken from Refs. [46,67,80–84]. The signal selection efficiency is again taken to be  $\epsilon = 40\%$  (solid colored lines)  $\pm 10\%$  (shaded bands). The signals computed from setting Eq. (2.19) equal to 0.1, and Eq. (2.21) equal to 0.00135, are plugged into Eq. (3.4) to obtain the expected 90% C.L. exclusion, and  $Z = 3$  evidence, reaches for proton partial lifetime, respectively. The black dashed curves correspond to a very optimistic scenario with  $b = 0$  and  $\epsilon = 46\%$  [67], and using the requirement  $s = 1$  in the discovery case (bottom panel). Also shown in Fig. 19 and other figures below are horizontal lines at our previously mentioned estimates of the current 95%, 90%, 68%, and 50% C.L. exclusion limit based on Super-Kamiokande’s data from 2014 [44].

The usual standard for discovery in particle physics is a significance of  $Z = 5$ . Therefore, we show in Fig. 20 the

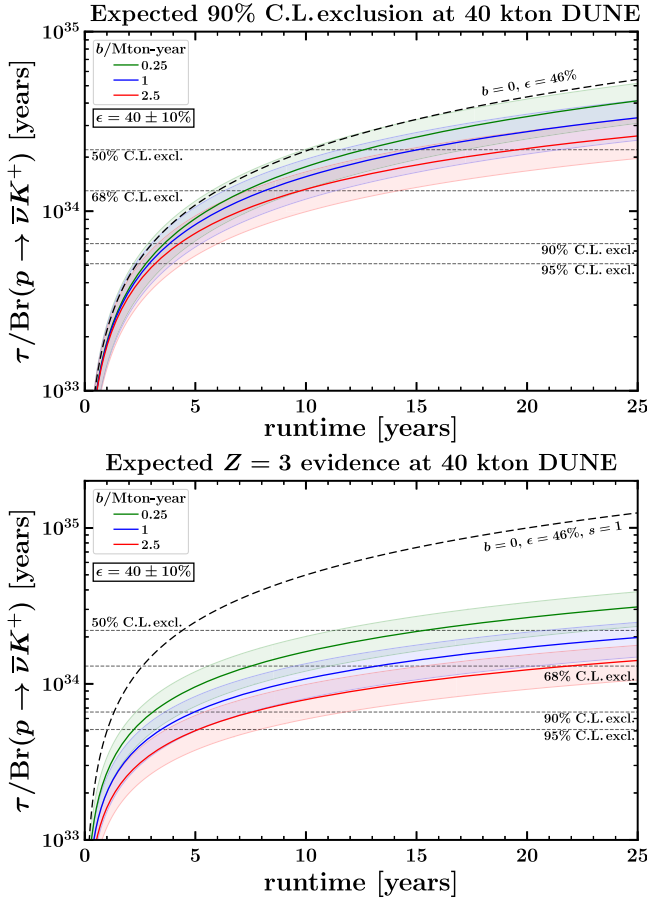


FIG. 19. Proton partial lifetimes in  $p \rightarrow \bar{\nu}K^+$  channel that are expected to be excluded at 90% C.L. [top panel, from Eqs. (2.19) and (3.4)] or discovered at  $Z = 3$  significance [bottom panel, from Eqs. (2.1), (2.21), and (3.4)] at 40 kton DUNE, as a function of runtime for various background rates per megaton-year of exposure, as labeled. The signal selection efficiency  $\epsilon$  is taken to be 40% (solid lines)  $\pm 10\%$  (shaded bands). The long dashed black line in each panel shows the idealized optimistic case of no background and  $\epsilon = 46\%$  [67], with the expected mean number of events required to be  $s = 1$  in the second panel. Our estimates of the current 95%, 90%, 68%, and 50% C.L. exclusion limit on proton partial lifetime, based on Super-Kamiokande's data from 2014 [44], are shown as horizontal dashed lines.

expected reach for  $Z = 5$  in the  $p \rightarrow \bar{\nu}K^+$  channel at 40 kton DUNE, as a function of the runtime. We note that even after 25 years, the discovery reach in this channel with nominal background rates remains below the value of  $\tau_p(p \rightarrow \bar{\nu}K^+)$  that we estimate to be excluded at 50% C.L. by the Super-Kamiokande data already published in 2014. Of course, a 50% C.L. exclusion is far from definitive, but this indicates the challenge being faced. This could change if the background can be reduced to near 0, as indicated by the dashed line, while maintaining a high efficiency for the signal.

As noted above in Sec. II B, if the mean expected number of signal events is  $s = 1$ , and one makes the optimistic

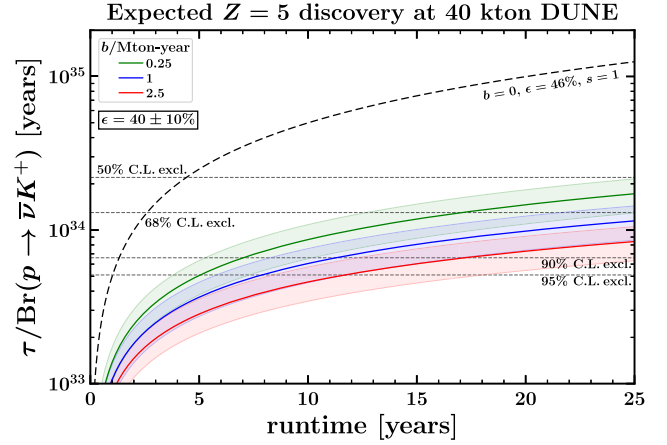


FIG. 20. Proton partial lifetimes in the  $p \rightarrow \bar{\nu}K^+$  channel that are expected to be discovered with a significance  $Z = 5$  at 40 kton DUNE, as a function of the runtime, for various background rates per megaton-year of exposure, as labeled. The results are obtained from Eqs. (2.1), (2.21), and (3.4). The horizontal dashed lines shown are our estimates of the current 95%, 90%, 68%, and 50% C.L. exclusion limit on proton partial lifetime, based on Super-Kamiokande's data from 2014 [44].

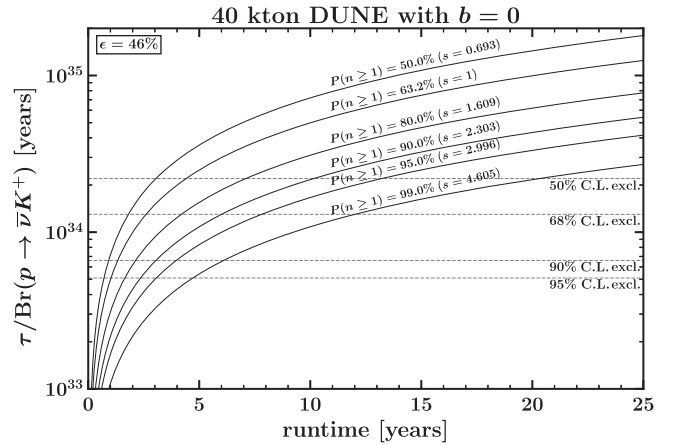


FIG. 21. Proton partial lifetimes in the  $p \rightarrow \bar{\nu}K^+$  channel that give different probabilities of observing at least one event from Eq. (2.25), which in turn correspond to different values of the expected signal, as labeled. The signal selection efficiency  $\epsilon$  is taken to be 46% based on [67]. The horizontal dashed lines shown are our estimates of the current 95%, 90%, 68%, and 50% C.L. exclusion limit on proton partial lifetime, based on Super-Kamiokande's data from 2014 [44].

assumption that the background is completely negligible ( $b = 0$ ), then the probability of obtaining at least one event is about 63.2%. Figure 21 shows the value of  $\tau/\text{Br}(p \rightarrow \bar{\nu}K^+)$ , as a function of the runtime, that would give various other probabilities of obtaining at least one event, again with the very optimistic assumption of absolutely no background  $b = 0$  and  $\epsilon = 46\%$  [67]. Each of these choices for  $P(n \geq 1)$  is equivalent to a requirement on the signal  $s$ , as labeled in the figure.

Reference [82] also provided a preliminary estimate for the background and signal efficiency for proton decay search in  $p \rightarrow e^+\pi^0$  mode at DUNE. Although DUNE is most sensitive to  $p \rightarrow \bar{\nu}K^+$  mode, for completeness, we will also show our expected reach estimates for proton partial lifetime in  $p \rightarrow e^+\pi^0$  mode at DUNE after 10 years and 20 years of runtime in our summary plots in Figs. 27 and 28 in the Outlook section below.

We now turn to projections for JUNO with 20 ktons of a liquid scintillator. We again obtain the upper limit on the signal using Eq. (2.19) for exclusion reach, and the signal needed for discovery using Eq. (2.21) for discovery reach, then applying Eq. (3.4). Figure 22 shows the proton lifetime in  $p \rightarrow \bar{\nu}K^+$  decay channel that is expected to be excluded at 90% or 95% C.L. (top panel) or discovered at  $Z = 3$  or  $Z = 5$  significance (bottom panel) or discovered at  $Z = 3$  or  $Z = 5$  significance (bottom panel)

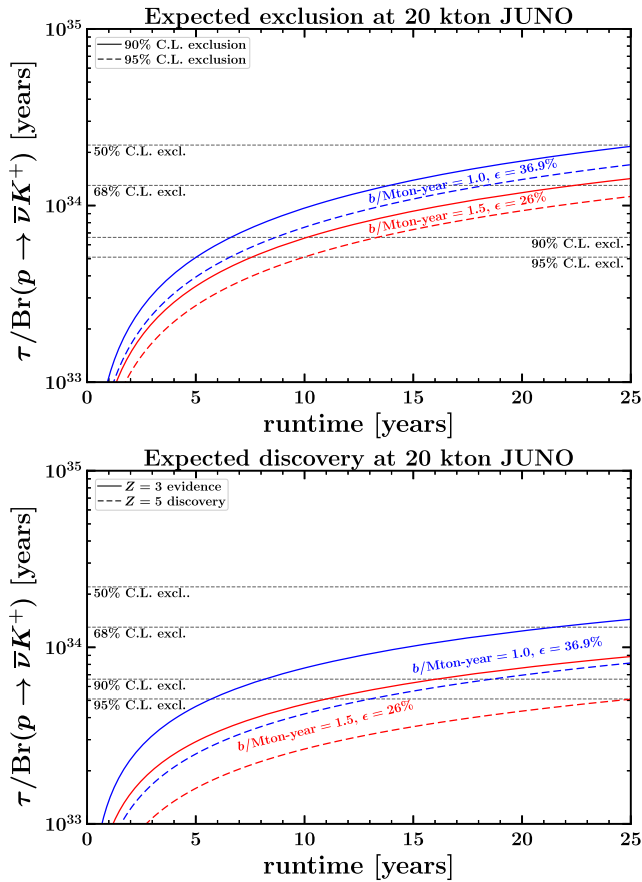


FIG. 22. Proton partial lifetime in the  $p \rightarrow \bar{\nu}K^+$  channel that is expected to be excluded at 90% or 95% C.L. [top panel, from Eqs. (2.19) and (3.4)] or discovered at  $Z = 3$  or  $Z = 5$  significance [bottom panel, from Eqs. (2.1), (2.21), and (3.4)] at 20 ktun JUNO, as a function of runtime, for two different estimated [85,86] combinations of background rates per year and signal selection efficiencies ( $b/\text{year}$ ,  $\epsilon$ ), as labeled. Our estimates of the current 95%, 90%, 68%, and 50% C.L. exclusion limit on proton partial lifetime, based on Super-Kamiokande’s data from 2014 [44], are shown as horizontal dashed lines.

at JUNO, as a function of the runtime. The two curves correspond to two different estimates of the background accumulated per megaton-year of exposure and the signal selection efficiency as labeled, taken from Ref. [85] ( $b/\text{Mton-year} = 1.0$ ,  $\epsilon = 36.9\%$ ; upper curve) and Ref. [86]<sup>12</sup> ( $b/\text{Mton-year} = 1.5$ ,  $\epsilon = 26\%$ ; lower curve). For comparison, our estimates of the current 95%, 90%, 68%, and 50% C.L. exclusion limit on proton partial lifetime, based on Super-Kamiokande’s data from 2014 [44], are shown as horizontal dashed lines.

For projected exclusion sensitivities, both DUNE [67] and JUNO [47] experiments made use of the Feldman-Cousins (FC) method [62] to obtain the upper limit on the signal assuming a fixed number of observed events, e.g.,  $n = 0$ . This approach can be problematic for projections because the FC upper limits at a fixed  $n$  decrease with  $b$  (as can be seen from Fig. 3), and for projections it can imply that the expected sensitivity of the experiment gets better if the background increases. Also considered in Ref. [67] is the usage of the FC method with  $n = b$ . For integer values of  $b$ , the FC upper limit with  $n = b$  sensibly increases as the background increases. But for noninteger  $b$ ,  $n$  is still an integer, and the FC upper limit with  $n = \text{round}(b)$  does not always increase with  $b$ , as shown above in Fig. 8. As a result, the projected sensitivity does not always decrease with  $b$ . This is why we chose to use the  $\text{CL}_{\text{excl}}$  ( $= \text{CL}_s$  for single-channel counting experiments) upper limit with the exact Asimov approximation given by Eq. (2.19) for DUNE and JUNO. While the FC sensitivity of Ref. [62] from Eq. (2.22) also gives sensible projections for exclusion, we note that it is computationally more intense to evaluate (and gives only slightly more conservative results) than the exact Asimov expected  $\text{CL}_{\text{excl}}$  upper limits.

We next turn to projections for Hyper-Kamiokande. Figures 23 and 24 show our estimates for the proton partial lifetimes in  $p \rightarrow \bar{\nu}K^+$  and  $p \rightarrow e^+\pi^0$  decay channels, respectively, that are expected to be excluded at 90% or 95% C.L. (top panels) or discovered at  $Z = 3$  or  $Z = 5$  significance (bottom panels), as a function of runtime at Hyper-Kamiokande. In order to obtain the exclusion and discovery reaches for  $\tau_p$ , the upper limit on partial width and the partial width needed for discovery are solved from Eqs. (3.10) and (3.11), respectively. These equations are used to combine the independent search channels in each decay mode, based on the background means and the signal selection efficiencies, along with their uncertainties, given in Ref. [48] and summarized in our Table II. Figures 23 and 24 also show our previously discussed estimates of the

<sup>12</sup>In order to obtain the expected reaches for proton partial lifetime at JUNO using Eq. (3.4), we redefined the signal efficiencies by multiplying the signal efficiencies given in Ref. [86] with the branching ratio of about 84.5% of the  $K^+$  decays that is included in JUNO’s analysis.



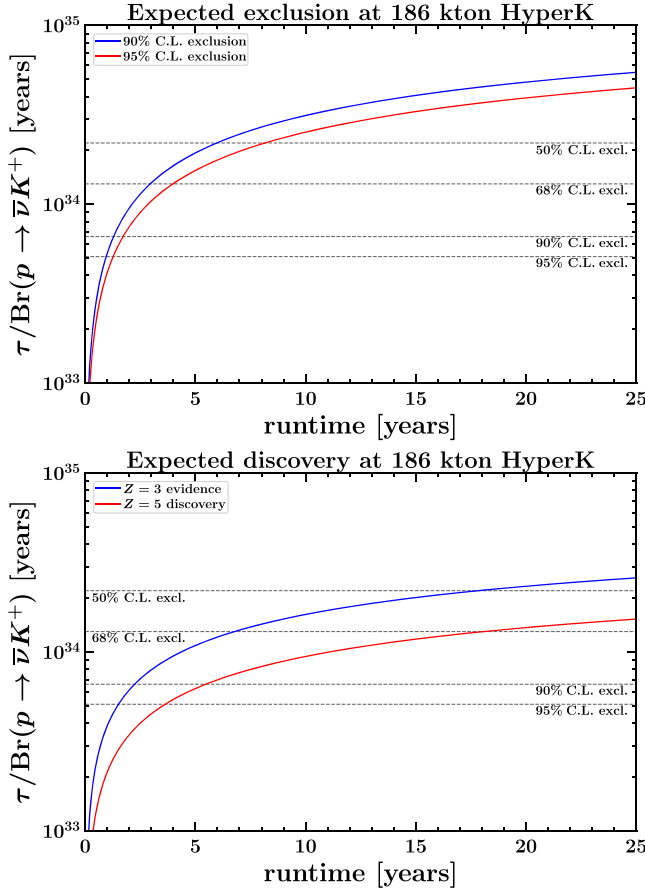


FIG. 23. Proton partial lifetime in the  $p \rightarrow \bar{\nu}K^+$  channel that is expected to be excluded at 90% or 95% C.L. [top panel; from Eq. (3.10)] or discovered at  $Z = 3$  or  $Z = 5$  significance [bottom panel; from Eq. (3.11)] at Hyper-Kamiokande with 186 kilotons of water, as a function of runtime, with the uncertainties in background and signal selection efficiency listed in Table II, taken from Ref. [48]. Our estimates of the current 95%, 90%, 68%, and 50% C.L. exclusion limit on proton partial lifetime, based on Super-Kamiokande's data from 2014 [44], are shown as horizontal dashed lines.

current exclusion limits at 95%, 90%, 68%, 50% C.L. in  $p \rightarrow \bar{\nu}K^+$  and  $p \rightarrow e^+\pi^0$  decay modes based on the data from Refs. [44] and [45], respectively.

Finally, we turn to projections for THEIA. In Figs. 25 and 26, we show the expected reaches, as a function of runtime, for proton partial lifetime in  $p \rightarrow \bar{\nu}K^+$  and  $p \rightarrow e^+\pi^0$  decay modes, respectively, for 90% or 95% C.L. exclusion (top panels) and discovery at  $Z = 3$  or  $Z = 5$  significance (bottom panels). The lower (red) lines show the results for THEIA-25 with 17 kt of fiducial mass of water based liquid scintillator, while the upper (blue) lines are for THEIA-100 with 80 kt of fiducial mass. The expected reach for proton partial lifetime is computed using Eq. (3.4), where the expected signal for 90% C.L. exclusion ( $Z = 3$  evidence) is obtained from setting Eq. (2.19) [Eq. (2.21)] to 0.1 (0.00135). The estimates for the background rate per megaton-year of exposure

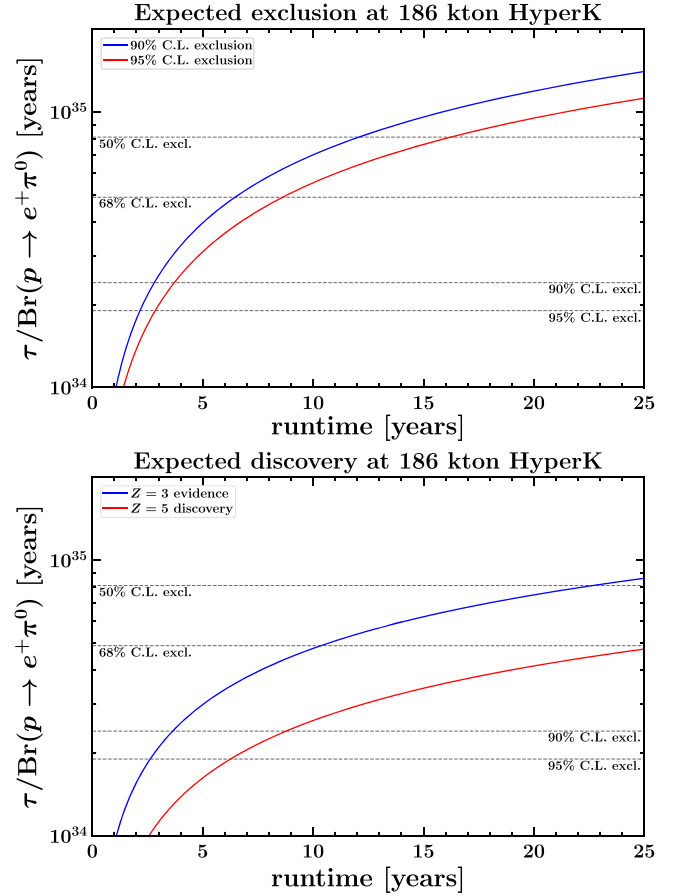


FIG. 24. Proton partial lifetime in the  $p \rightarrow e^+\pi^0$  channel that is expected to be excluded at 90% or 95% C.L. [top panel; from Eq. (3.10)] or discovered at  $Z = 3$  or  $Z = 5$  significance [bottom panel; from Eq. (3.11)] at Hyper-Kamiokande with 186 kilotons of water, as a function of runtime, with the uncertainties in background and signal selection efficiency listed in Table II, taken from Ref. [48]. Our estimates of the current 95%, 90%, 68%, and 50% C.L. exclusion limit on proton partial lifetime, based on Super-Kamiokande's data from 2020 [45], are shown as horizontal dashed lines.

and the signal selection efficiency for the decays modes  $p \rightarrow \bar{\nu}K^+$  and  $p \rightarrow e^+\pi^0$  are taken from Refs. [49] and [79], respectively. As before, we also show our estimates for the current lower limits at various confidence levels based on the data from Super-Kamiokande [44,45].

#### IV. OUTLOOK

We summarize our projections for future proton decay searches in the final states  $e^+\pi^0$  and  $\bar{\nu}K^+$  at DUNE, JUNO, and Hyper-Kamiokande in Fig. 27 for exclusion (assuming the signal is indeed absent), and in Fig. 28 for discovery (assuming the signal is actually present). And in Fig. 29 we summarize our projections at THEIA for 95% C.L. exclusion and  $Z = 5$  discovery for various fiducial masses  $N_{\text{kton}} = (10, 25, 50, 100)$  kton. In each

TABLE II. Estimated backgrounds  $\hat{b}_i \pm \Delta_{b_i}$  per megaton-year of exposure and signal efficiencies  $\hat{\epsilon}_i \pm \Delta_{\epsilon_i}$  at Hyper-Kamiokande, taken from Ref. [48], for  $p \rightarrow \bar{\nu}K^+$  and  $p \rightarrow e^+\pi^0$  decay modes. The last column gives a brief description of each of the channels referring to the name of the search method used in Ref. [48]. Exposure in each channel for a 186 kton Hyper-Kamiokande is given by  $\lambda_i = 0.186 \text{ Mton} \times \text{runtime in years}$ .

Decay mode	$\hat{b}_i \pm \Delta_{b_i}$ [/Mton-year]	$\hat{\epsilon}_i \pm \Delta_{\epsilon_i}$ [%]	Comment
$p \rightarrow \bar{\nu}K^+$	$0.9 \pm 0.2$	$12.7 \pm 2.4$	prompt $\gamma$
	$0.7 \pm 0.2$	$10.8 \pm 1.1$	$\pi^+\pi^0$
	1916	31	$p_\mu$ spectrum
$p \rightarrow e^+\pi^0$	$0.06 \pm 0.02$	$18.7 \pm 1.2$	$0 < p_{\text{tot}} < 100 \text{ MeV}/c$
	$0.62 \pm 0.20$	$19.4 \pm 2.9$	$100 < p_{\text{tot}} < 250 \text{ MeV}/c$

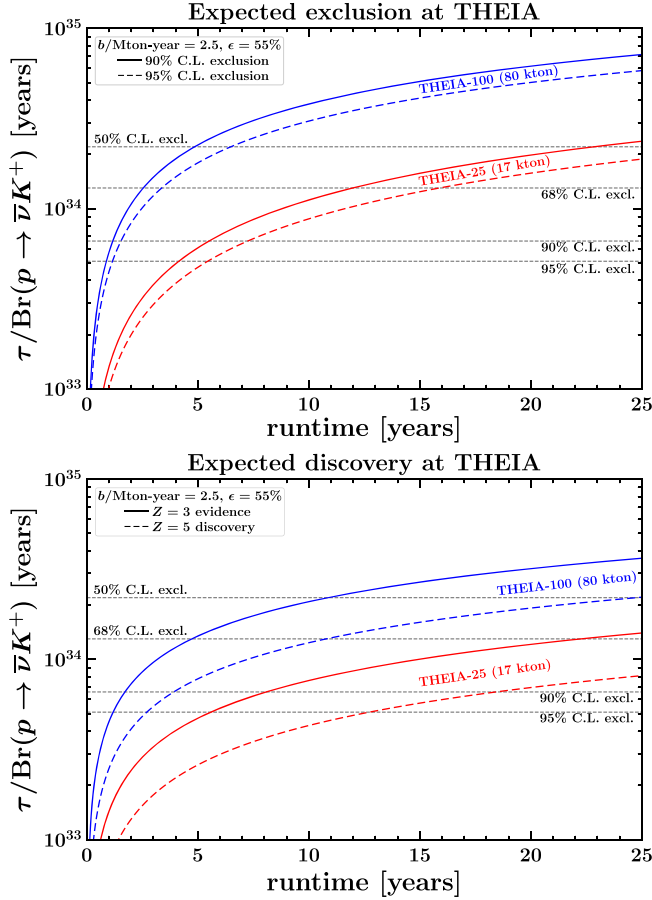


FIG. 25. Expected 90% or 95% C.L. exclusion reaches [top panel; from Eqs. (2.19) and (3.4)] and  $Z = 3$  or  $Z = 5$  discovery reaches [bottom panel; from Eqs. (2.1), (2.21), and (3.4)] for proton partial lifetime in  $p \rightarrow \bar{\nu}K^+$  with THEIA-25 (red lines) and THEIA-100 (blue lines) with 17 and 80 ktons of water based liquid scintillator, respectively, as a function of runtime. The estimates for the background (per megaton-year of exposure) and the signal efficiencies are taken from Ref. [49]. Our estimates of the current 95%, 90%, 68%, and 50% C.L. exclusion limit on proton partial lifetime, based on Super-Kamiokande's data from 2014 [44], are shown as horizontal dashed lines.

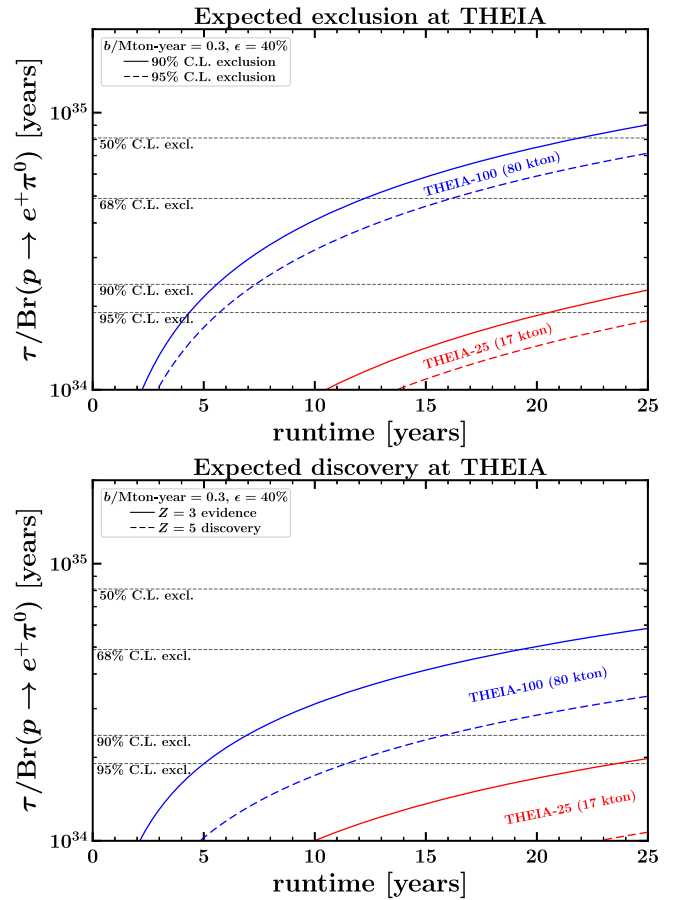


FIG. 26. Expected 90% or 95% C.L. exclusion reaches [top panel; from Eqs. (2.19) and (3.4)] and  $Z = 3$  or  $Z = 5$  discovery reaches [bottom panel; from Eqs. (2.1), (2.21), and (3.4)] for proton partial lifetime in  $p \rightarrow e^+\pi^0$  with THEIA-25 (red lines) and THEIA-100 (blue lines) with 17 and 80 ktons of water based liquid scintillator, respectively, as a function of runtime. The estimates for the background (per megaton-year of exposure) and the signal efficiencies are taken from Ref. [79]. Our estimates of the current 95%, 90%, 68%, and 50% C.L. exclusion limit on proton partial lifetime, based on Super-Kamiokande's data from 2020 [45], are shown as horizontal dashed lines.

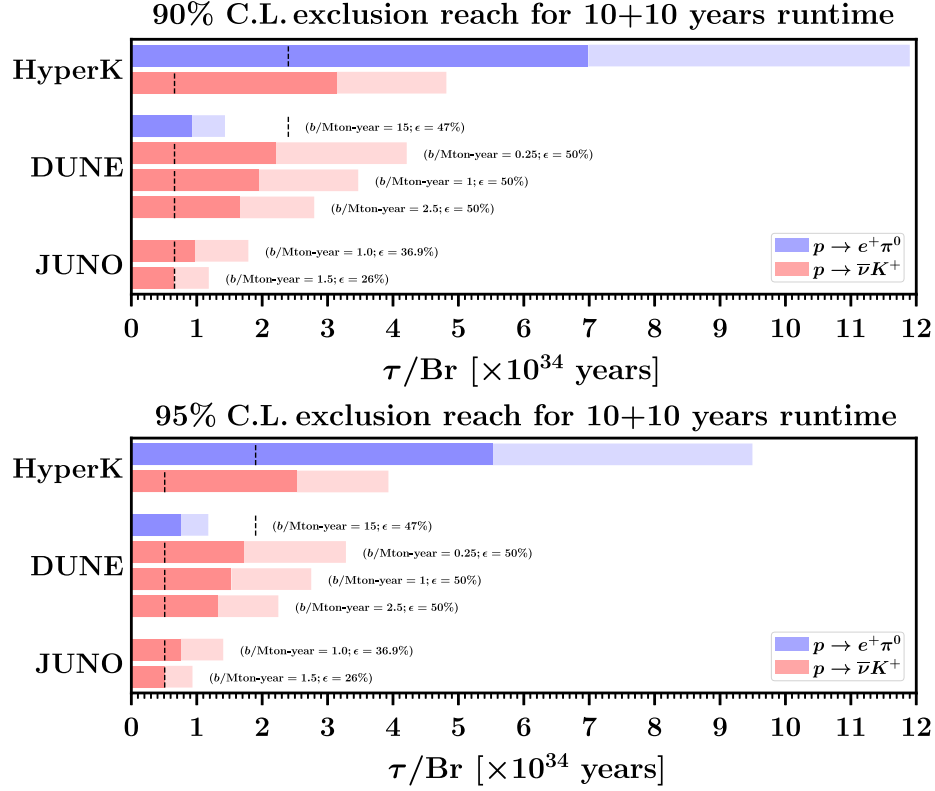


FIG. 27. Expected exclusion reaches at 90% C.L. (top panel) and 95% C.L. (bottom panel) for proton partial lifetime in  $p \rightarrow e^+\pi^0$  (blue bars) and  $p \rightarrow \bar{\nu}K^+$  (red bars) decay channels at JUNO, DUNE, and Hyper-Kamiokande after 10 years (darker shading) and 20 years (lighter shading) of runtime. The assumed backgrounds and signal efficiencies for JUNO and DUNE are labeled in the plots, and for Hyper-Kamiokande, the corresponding information is given in Table II, quoted from Ref. [48]. These results are based on preliminary estimates of the backgrounds and signal efficiencies, which are likely to change as the experiments progress, and therefore should be viewed with some caution as comparisons. The vertical dashed lines are our estimates of the current 90% C.L. (top panel) and 95% C.L. (bottom panel) lower limits based on Super-Kamiokande's data from 2014 [44] and 2020 [45].

case, we show results for 10 years and 20 years of runtime. The assumed backgrounds and signal efficiencies for DUNE,<sup>13</sup> JUNO, and THEIA in each proton decay mode are labeled in the plots, while the corresponding information for the multichannel Hyper-Kamiokande searches was given in Table II above, quoted from Ref. [48]. The vertical dashed lines correspond to our estimate of the current 90% C.L. (Fig. 28, top panel of Fig. 27, and bottom panel of Fig. 29) or 95% C.L. (bottom panel of Fig. 27 and top panel of Fig. 29) lower limit on proton partial lifetime in the respective decay channels, based on the published Super-Kamiokande data [44,45].

As noted above, our projections here are based on the exact Asimov evaluation of the Bayesian statistics  $\text{CL}_{\text{excl}}$  and  $\text{CL}_{\text{disc}}$ . Our results are somewhat more conservative than previous projections appearing in Refs. [48] and the

Snowmass report [79], which we have generalized to include 90% C.L. exclusion and  $Z = 3$  evidence reach estimates as a function of runtime (for various estimates of backgrounds and signal efficiencies, notably for DUNE and JUNO) as well as estimates for 95% C.L. exclusion and  $Z = 5$  discovery. In the cases of single-channel searches for DUNE, JUNO, and THEIA, we have also investigated the use of the exact Asimov frequentist  $p$ -value measures  $p_{\text{excl}}$  and  $p_{\text{disc}}$ . These results are not shown in the figures; we find that they are only slightly less conservative than the estimates shown.

The two panels of Fig. 27 show the projected exclusion reaches at 90% and 95% confidence level, while the two panels of Fig. 28 give the projected reaches for  $Z = 3$  evidence and  $Z = 5$  discovery at DUNE, JUNO, and Hyper-Kamiokande. And the top (bottom) panel of Fig. 29 shows the projected 95% C.L. exclusion ( $Z = 5$  discovery) reaches at THEIA with various fiducial masses of the detector material. As expected, for each planned experiment the reaches for exclusion are substantially higher than the corresponding reaches for a possible discovery. We note that the prospects for a definitive  $Z = 5$  discovery are

<sup>13</sup>For projections at DUNE in  $p \rightarrow \bar{\nu}K^+$  channel, we are using the optimistic choices based on Ref. [80]. More pessimistic choices from Refs. [81–84] will of course lead to lower reach estimates.

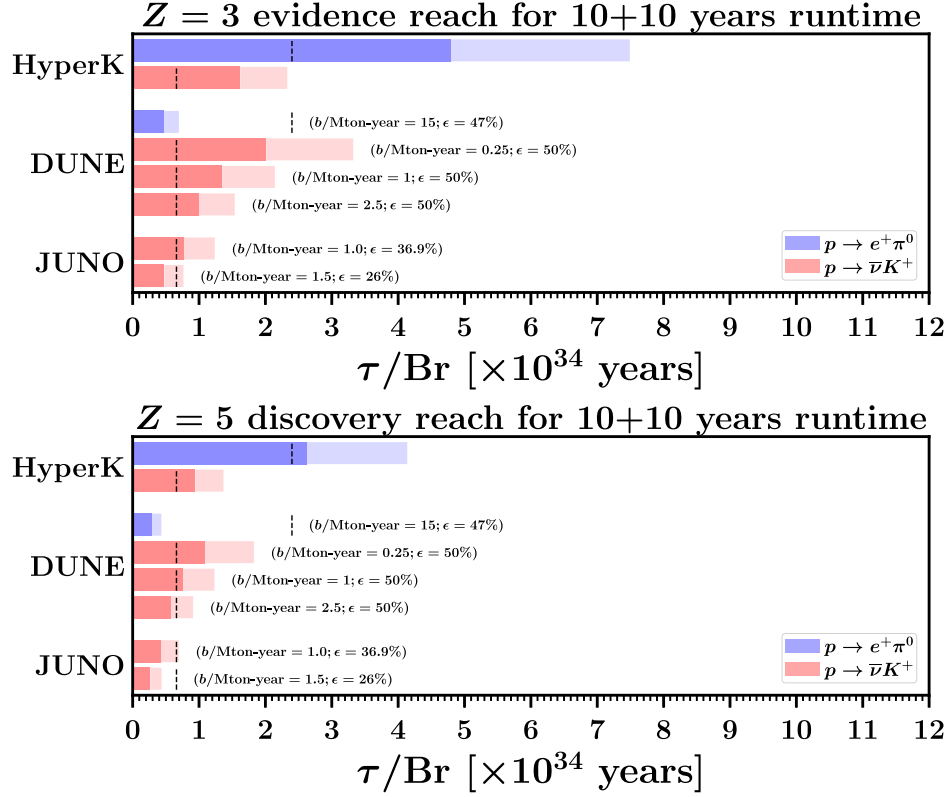


FIG. 28. Expected reaches for  $Z = 3$  evidence (top panel) and  $Z = 5$  discovery (bottom panel) for proton partial lifetime in  $p \rightarrow e^+\pi^0$  (blue bars) and  $p \rightarrow \bar{\nu}K^+$  (red bars) decay channels, at JUNO, DUNE, and Hyper-Kamiokande after 10 years (darker shading) and 20 years (lighter shading) of runtime. The assumed backgrounds and signal efficiencies for JUNO and DUNE are labeled in the plots, and for Hyper-Kamiokande, the corresponding information is given in Table II, quoted from Ref. [48]. These results are based on preliminary estimates of the backgrounds and signal efficiencies, which are likely to change as the experiments progress, and therefore should be viewed with some caution as comparisons. The vertical dashed lines are our estimates of the current 90% C.L. lower limits based on Super-Kamiokande's data from 2014 [44] and 2020 [45].

particularly modest after one takes into account the limits already obtained by Super-Kamiokande.

The results shown in Figs. 27–29 are preliminary estimates, as the presently available background and signal efficiency estimates vary significantly in their reliability, and more robust estimates will become available only when the experiments are closer to collecting data. For the same reason, the results should be viewed with some caution as a direct comparison of the different experiments, which are at very different stages of planning and development.

Proton decay experiments prior to Super-Kamiokande have ruled out the simplest variations of minimal  $SU(5)$  GUT [5], and Super-Kamiokande has seemingly ruled out the minimal supersymmetric  $SU(5)$  GUT [18–21] with sfermion masses less than around the TeV scale. However there are many other well-motivated GUT models that predict proton partial lifetimes well beyond the current lower limits (see summary tables in Refs. [77,79] and references therein).

For example, nonsupersymmetric GUTs such as some minimally extended  $SU(5)$  models [7,8] and minimal  $SO(10)$  model [14] predict  $p \rightarrow e^+\pi^0$  to be the dominant

decay mode with partial lifetimes of order  $10^{32}$ – $10^{36}$  years and  $\lesssim 5 \times 10^{35}$  years, respectively. Supersymmetric  $SU(5)$  GUTs predict the proton partial lifetime for the leading mode  $p \rightarrow \bar{\nu}K^+$  to be  $3 \times 10^{34} - 2 \times 10^{35}$  years in minimal supergravity framework (MSUGRA) and  $3 \times 10^{34}$ – $10^{36}$  years in supergravity models with nonuniversal gaugino masses (NUSUGRA), as discussed in Ref. [26] in light of the observed Higgs mass. Reference [22] revisited the minimal supersymmetric  $SU(5)$  GUT and obtained  $\tau_p/\text{Br}(p \rightarrow \bar{\nu}K^+) \lesssim (2-6) \times 10^{34}$  years assuming universality of the soft supersymmetry breaking parameters at the GUT scale with sfermion masses less than around  $\mathcal{O}(10)$  TeV. There are also supersymmetric GUTs such as the split  $SU(5)$  supersymmetry [31] and flipped  $SU(5)$  supersymmetric GUTs [28–30], where the dominant decay mode can be  $p \rightarrow e^+\pi^0$  with lifetimes of order  $10^{35}$ – $10^{37}$  years.

From our estimates of the reaches summarized in Figs. 27–29, we can see that DUNE, JUNO, Hyper-Kamiokande, and THEIA can probe a significant fraction of the parameter space of various presently viable supersymmetric and nonsupersymmetric GUTs and could eventually lead the way to a more complete theory.



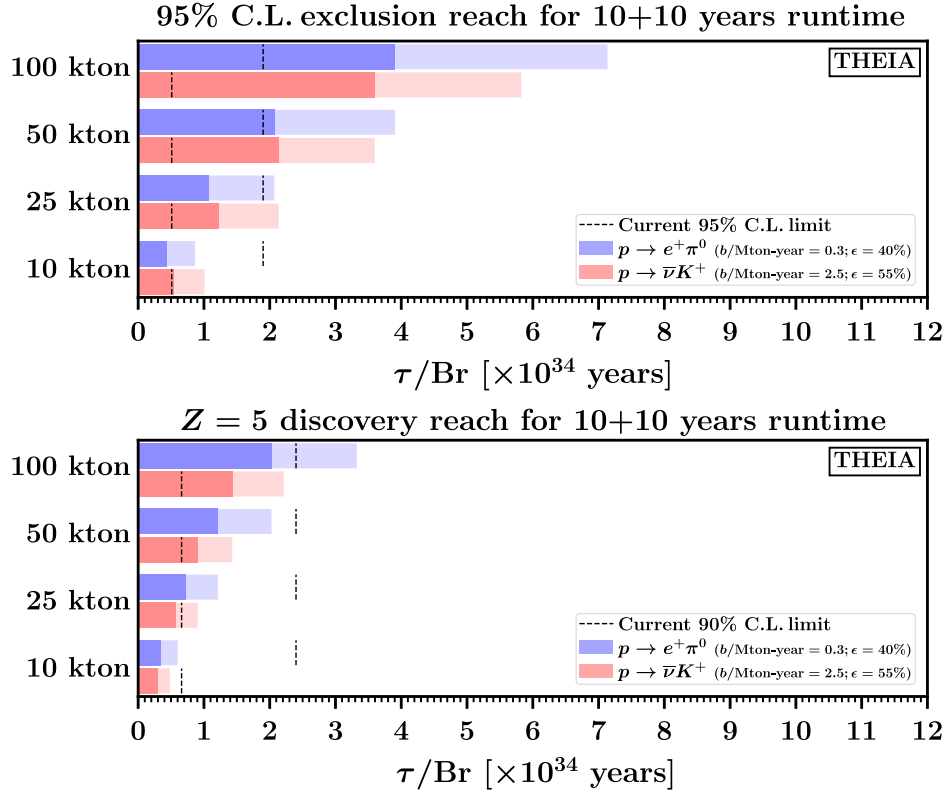


FIG. 29. Expected reaches at THEIA for 95% C.L. exclusion (top panel) and  $Z = 5$  discovery (bottom panel) for proton partial lifetime in  $p \rightarrow e^+ \pi^0$  (blue bars) and  $p \rightarrow \bar{\nu} K^+$  (red bars) decay channels with various fiducial masses, as labeled, after 10 years (darker shading) and 20 years (lighter shading) of runtime. The assumed background rates and signal efficiencies for THEIA are labeled in the plots. These results are based on preliminary estimates of the backgrounds and signal efficiencies, which are likely to change as the experiment progresses. The vertical dashed lines are our estimates of the current 95% C.L. (top panel) and 90% C.L. (bottom panel) lower limits based on Super-Kamiokande's data from 2014 [44] and 2020 [45].

The existing code repository Zstats [66] is updated with various statistical measures of significance for counting experiments with multiple independent search channels as investigated in this paper. The updates include the significances based on our proposed Bayesian-motivated measures  $CL_{\text{disc}}$  and  $CL_{\text{excl}}$ , and their application to study the statistical significances for proton decay at current and future neutrino detectors. To demonstrate the usage of the code, the repository also contains some code snippets in a PYTHON notebook that generate the data in each of the figures in this paper.

### ACKNOWLEDGMENTS

We thank Jennifer Raaf and Edward Kearns for helpful discussions. P.N.B. thanks Tracy Slatyer for helpful discussions on Bayes factors and Zirui Wang for useful

discussions on statistical combinations using RooStats package [87]. P.N.B. and J.D.W. thank Jianming Qian for useful discussions. P.N.B. and J.D.W. also thank John Thiels for his help and support in using the Great Lakes cluster at University of Michigan. This research was supported in part through computational resources and services provided by Advanced Research Computing (ARC), a division of Information and Technology Services (ITS) at the University of Michigan, Ann Arbor. This research was also supported in part through computational resources and services provided by NICADD compute cluster at Northern Illinois University. This work is supported in part by the National Science Foundation under Grant No. 2013340, and by the Department of Energy under Grant No. DE-SC0007859.

- [1] A. D. Sakharov, Violation of  $CP$  invariance,  $C$  asymmetry, and baryon asymmetry of the universe, *Pis'ma Zh. Eksp. Teor. Fiz.* **5**, 32 (1967).
- [2] G. 't Hooft, Symmetry Breaking Through Bell-Jackiw Anomalies, *Phys. Rev. Lett.* **37**, 8 (1976).
- [3] V. A. Kuzmin, V. A. Rubakov, and M. E. Shaposhnikov, On the anomalous electroweak baryon number nonconservation in the early universe, *Phys. Lett.* **155B**, 36 (1985).
- [4] J. C. Pati and A. Salam, Is Baryon Number Conserved?, *Phys. Rev. Lett.* **31**, 661 (1973).
- [5] H. Georgi and S. L. Glashow, Unity of All Elementary Particle Forces, *Phys. Rev. Lett.* **32**, 438 (1974).
- [6] P. Langacker, Grand unified theories and proton decay, *Phys. Rep.* **72**, 185 (1981).
- [7] I. Doršner and S. Saad, Towards minimal  $SU(5)$ , *Phys. Rev. D* **101**, 015009 (2020).
- [8] P. Fileviez Perez and C. Murgui, Renormalizable  $SU(5)$  unification, *Phys. Rev. D* **94**, 075014 (2016).
- [9] B. Bajc and G. Senjanovic, Seesaw at LHC, *J. High Energy Phys.* **08** (2007) 014.
- [10] I. Doršner and P. Fileviez Perez, Unification without supersymmetry: Neutrino mass, proton decay and light leptiquarks, *Nucl. Phys.* **B723**, 53 (2005).
- [11] S. Saad, Origin of a two-loop neutrino mass from  $SU(5)$  grand unification, *Phys. Rev. D* **99**, 115016 (2019).
- [12] C. Arbeláez Rodríguez, H. Kolešová, and M. Malinský, Witten's mechanism in the flipped  $SU(5)$  unification, *Phys. Rev. D* **89**, 055003 (2014).
- [13] G. Altarelli and D. Meloni, A non supersymmetric  $SO(10)$  grand unified model for all the physics below  $M_{GUT}$ , *J. High Energy Phys.* **08** (2013) 021.
- [14] K. S. Babu and S. Khan, Minimal nonsupersymmetric  $SO(10)$  model: Gauge coupling unification, proton decay, and fermion masses, *Phys. Rev. D* **92**, 075018 (2015).
- [15] S. Bertolini, L. Di Luzio, and M. Malinsky, Seesaw scale in the minimal renormalizable  $SO(10)$  grand unification, *Phys. Rev. D* **85**, 095014 (2012).
- [16] A. Preda, G. Senjanovic, and M. Zantedeschi,  $SO(10)$ : A case for hadron colliders, *arXiv:2201.02785*.
- [17] J. Chakraborty, R. Maji, and S. F. King, Unification, proton decay and topological defects in non-SUSY GUTs with thresholds, *Phys. Rev. D* **99**, 095008 (2019).
- [18] S. Dimopoulos and H. Georgi, Softly broken supersymmetry and  $SU(5)$ , *Nucl. Phys.* **B193**, 150 (1981).
- [19] S. Dimopoulos, S. Raby, and F. Wilczek, Proton decay in supersymmetric models, *Phys. Lett.* **112B**, 133 (1982).
- [20] N. Sakai and T. Yanagida, Proton decay in a class of supersymmetric grand unified models, *Nucl. Phys.* **B197**, 533 (1982).
- [21] J. Hisano, H. Murayama, and T. Yanagida, Nucleon decay in the minimal supersymmetric  $SU(5)$  grand unification, *Nucl. Phys.* **B402**, 46 (1993).
- [22] J. Ellis, J. L. Evans, N. Nagata, K. A. Olive, and L. Velasco-Sevilla, Supersymmetric proton decay revisited, *Eur. Phys. J. C* **80**, 332 (2020).
- [23] K. S. Babu, B. Bajc, and Z. Tavartkiladze, Realistic fermion masses and nucleon decay rates in SUSY  $SU(5)$  with vector-like matter, *Phys. Rev. D* **86**, 075005 (2012).
- [24] P. Nath, A. H. Chamseddine, and R. L. Arnowitt, Nucleon decay in supergravity unified theories, *Phys. Rev. D* **32**, 2348 (1985).
- [25] P. Nath and R. L. Arnowitt, Grand unification and  $B$  and  $L$  conservation, *arXiv:hep-ph/9808465*.
- [26] M. Liu and P. Nath, Higgs boson mass, proton decay, naturalness, and constraints of the LHC and Planck data, *Phys. Rev. D* **87**, 095012 (2013).
- [27] J. C. Pati, Probing grand unification through neutrino oscillations, leptogenesis, and proton decay, *Int. J. Mod. Phys. A* **18**, 4135 (2003).
- [28] J. R. Ellis, D. V. Nanopoulos, and J. Walker, Flipping  $SU(5)$  out of trouble, *Phys. Lett. B* **550**, 99 (2002).
- [29] J. Ellis, J. L. Evans, N. Nagata, D. V. Nanopoulos, and K. A. Olive, Flipped  $SU(5)$  GUT phenomenology: Proton decay and  $g_\mu - 2$ , *Eur. Phys. J. C* **81**, 1109 (2021).
- [30] J. Ellis, M. A. G. Garcia, N. Nagata, D. V. Nanopoulos, and K. A. Olive, Proton decay: Flipped vs unflipped  $SU(5)$ , *J. High Energy Phys.* **05** (2020) 021.
- [31] N. Arkani-Hamed, S. Dimopoulos, G. F. Giudice, and A. Romanino, Aspects of split supersymmetry, *Nucl. Phys.* **B709**, 3 (2005).
- [32] A. Hebecker and J. March-Russell, Proton decay signatures of orbifold GUTs, *Phys. Lett. B* **539**, 119 (2002).
- [33] M. L. Alciati, F. Feruglio, Y. Lin, and A. Varagnolo, Proton lifetime from  $SU(5)$  unification in extra dimensions, *J. High Energy Phys.* **03** (2005) 054.
- [34] K. S. Babu, B. Bajc, and S. Saad, Resurrecting minimal Yukawa sector of SUSY  $SO(10)$ , *J. High Energy Phys.* **10** (2018) 135.
- [35] K. S. Babu, T. Fukuyama, S. Khan, and S. Saad, Peccei-Quinn symmetry and nucleon decay in renormalizable SUSY  $SO(10)$ , *J. High Energy Phys.* **06** (2019) 045.
- [36] R. N. Mohapatra and M. Sevrerson, Leptonic  $CP$  violation and proton decay in SUSY  $SO(10)$ , *J. High Energy Phys.* **09** (2018) 119.
- [37] P. S. Bhupal Dev and R. N. Mohapatra, Electroweak symmetry breaking and proton decay in  $SO(10)$  SUSY-GUT with TeV  $W(R)$ , *Phys. Rev. D* **82**, 035014 (2010).
- [38] Q. Shafi and Z. Tavartkiladze, Flavor problem, proton decay and neutrino oscillations in SUSY models with anomalous  $U(1)$ , *Phys. Lett. B* **473**, 272 (2000).
- [39] V. Lucas and S. Raby, Nucleon decay in a realistic  $SO(10)$  SUSY GUT, *Phys. Rev. D* **55**, 6986 (1997).
- [40] K. S. Babu, J. C. Pati, and F. Wilczek, Suggested new modes in supersymmetric proton decay, *Phys. Lett. B* **423**, 337 (1998).
- [41] K. S. Babu, J. C. Pati, and F. Wilczek, Fermion masses, neutrino oscillations, and proton decay in the light of Super-Kamiokande, *Nucl. Phys.* **B566**, 33 (2000).
- [42] J. C. Pati, Discovery of proton decay: A must for theory, a challenge for experiment, *AIP Conf. Proc.* **533**, 37 (2000).
- [43] P. S. Bhupal Dev, B. Dutta, R. N. Mohapatra, and M. Sevrerson,  $\theta_{13}$  and proton decay in a minimal  $SO(10) \times S_4$  model of flavor, *Phys. Rev. D* **86**, 035002 (2012).
- [44] K. Abe *et al.* (Super-Kamiokande Collaboration), Search for proton decay via  $p \rightarrow \nu K^+$  using 260 kiloton  $\cdot$  year data of Super-Kamiokande, *Phys. Rev. D* **90**, 072005 (2014).
- [45] A. Takenaka *et al.* (Super-Kamiokande Collaboration), Search for proton decay via  $p \rightarrow e^+ \pi^0$  and  $p \rightarrow \mu^+ \pi^0$  with

- an enlarged fiducial volume in Super-Kamiokande I-IV, *Phys. Rev. D* **102**, 112011 (2020).
- [46] R. Acciarri *et al.* (DUNE Collaboration), Long-baseline neutrino facility (LBNF) and deep underground neutrino experiment (DUNE): Conceptual design report, volume 2: The physics program for DUNE at LBNF, [arXiv:1512.06148](#).
- [47] F. An *et al.* (JUNO Collaboration), Neutrino physics with JUNO, *J. Phys. G* **43**, 030401 (2016).
- [48] K. Abe *et al.* (Hyper-Kamiokande Collaboration), Hyper-kamiokande design report, [arXiv:1805.04163](#).
- [49] M. Askins *et al.* (Theia Collaboration), THEIA: An advanced optical neutrino detector, *Eur. Phys. J. C* **80**, 416 (2020).
- [50] P. N. Bhattiprolu, S. P. Martin, and J. D. Wells, Criteria for projected discovery and exclusion sensitivities of counting experiments, *Eur. Phys. J. C* **81**, 123 (2021).
- [51] G. Zech, Upper limits in experiments with background or measurement errors, *Nucl. Instrum. Methods Phys. Res., Sect. A* **277**, 608 (1989).
- [52] T. Junk, Confidence level computation for combining searches with small statistics, *Nucl. Instrum. Methods Phys. Res., Sect. A* **434**, 435 (1999).
- [53] A. L. Read, Modified frequentist analysis of search results (The  $CL_s$  method), Report No. CERN-OPEN-2000-205.
- [54] A. L. Read, Presentation of search results: The  $CL_s$  technique, *J. Phys. G* **28**, 2693 (2002).
- [55] J. O. Berger, A comparison of testing methodologies, in *Proceedings of the PHYSTAT-LHC Workshop on Statistical Issues for LHC Physics*, CERN 2008-001, 8-19, and The Bayesian Approach to Discovery, contribution to PHYSTAT (CERN, Geneva, Switzerland, 2011).
- [56] G. Cowan, Bayesian statistical methods for HEP, *Terascale Statistics School* (DESY, Hamburg, Germany, 2008).
- [57] R. E. Kass and A. E. Raftery, Bayes factors, *J. Am. Stat. Assoc.* **90**, 773 (1995).
- [58] H. Jeffreys, *The Theory of Probability* (Oxford University Press, Oxford, 1998); Some tests of significance, treated by the theory of probability, *Math. Proc. Cambridge Philos. Soc.* **31**, 203 (1935).
- [59] P. N. Bhattiprolu and S. P. Martin, Signal-background interference for digluon resonances at the Large Hadron Collider, *Phys. Rev. D* **102**, 015016 (2020).
- [60] W. M. Yao *et al.* (Particle Data Group), Review of particle physics, *J. Phys. G* **33**, 1 (2006).
- [61] P. A. Zyla *et al.* (Particle Data Group), Review of particle physics, *Prog. Theor. Exp. Phys.* **2020**, 083C01 (2020).
- [62] G. J. Feldman and R. D. Cousins, A unified approach to the classical statistical analysis of small signals, *Phys. Rev. D* **57**, 3873 (1998).
- [63] O. Helene, Upper limit of peak area, *Nucl. Instrum. Methods* **212**, 319 (1983).
- [64] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, Asymptotic formulae for likelihood-based tests of new physics, *Eur. Phys. J. C* **71**, 1554 (2011).
- [65] G. Cowan, Two developments in tests for discovery: Use of weighted Monte Carlo events and an improved measure, Progress on Statistical Issues in Searches, SLAC, 2012.
- [66] P. N. Bhattiprolu, S. P. Martin, and J. D. Wells, Zstats v2.0 package, <https://github.com/prudhvibhattiprolu/Zstats/> (2022).
- [67] C. Alt, Sensitivity study for proton decay via  $p \rightarrow \bar{\nu} K^+$  using a 10 kiloton dual phase liquid argon time projection chamber at the Deep Underground Neutrino Experiment, [10.3929/ethz-b-000462924](#).
- [68] T. P. Li and Y. Q. Ma, Analysis methods for results in gamma-ray astronomy, *Astrophys. J.* **272**, 317 (1983).
- [69] N. Gehrels, Confidence limits for small numbers of events in astrophysical data, *Astrophys. J.* **303**, 336 (1986).
- [70] S. N. Zhang and D. Ramsden, Statistical data analysis for gamma-ray astronomy, *Exp. Astron.* **1**, 145 (1990).
- [71] D. E. Alexandreas *et al.*, Point source search techniques in ultra high energy gamma ray astronomy, *Nucl. Instrum. Methods Phys. Res., Sect. A* **328**, 570 (1993).
- [72] J. T. Linnemann, Measures of significance in HEP and astrophysics, in *Proceedings of PhyStat 2003: Statistical Problems in Particle Physics, Astrophysics, and Cosmology* (SLAC, Stanford, California USA, 2003); *eConf C030908*, MOBT001 (2003).
- [73] R. D. Cousins, J. T. Linnemann, and J. Tucker, Evaluation of three methods for calculating statistical significance when incorporating a systematic uncertainty into a test of the background-only hypothesis for a Poisson process, *Nucl. Instrum. Methods Phys. Res., Sect. A* **595**, 480 (2008).
- [74] T. J. Haines, An experimental search for evidence of a nucleon decay signal above neutrino backgrounds, Ph.D. thesis, UC Irvine [Report No. UMI-87-03928, 1986].
- [75] R. D. Cousins and V. L. Highland, Incorporating systematic uncertainties into an upper limit, *Nucl. Instrum. Methods Phys. Res., Sect. A* **320**, 331 (1992).
- [76] ATLAS, CMS and LHC Higgs Combination Group, Procedure for the LHC Higgs boson search combination in Summer 2011, Report No. CMS-NOTE-2011-005.
- [77] A. Bueno, Z. Dai, Y. Ge, M. Laffranchi, A. J. Melgarejo, A. Mereaglia, S. Navas, and A. Rubbia, Nucleon decay searches with large liquid argon TPC detectors at shallow depths: Atmospheric neutrinos and cosmogenic backgrounds, *J. High Energy Phys.* **04** (2007) 041.
- [78] K. Kobayashi *et al.* (Super-Kamiokande Collaboration), Search for nucleon decay via modes favored by supersymmetric grand unification models in Super-Kamiokande-I, *Phys. Rev. D* **72**, 052007 (2005).
- [79] P. S. B. Dev, L. W. Koerner, S. Saad, S. Antusch, M. Askins, K. S. Babu, J. L. Barrow, J. Chakraborty, A. de Gouvêa, Z. Djuric *et al.*, Searches for baryon number violation in neutrino experiments: A white paper, [arXiv:2203.08771](#).
- [80] C. Alt, B. Radics, and A. Rubbia, Neural-network-driven proton decay sensitivity in the  $p \rightarrow \bar{\nu} K^+$  channel using large liquid argon time projection chambers, *J. High Energy Phys.* **04** (2021) 243.
- [81] C. Alt (DUNE Collaboration), Nucleon decay search with DUNE, *Proc. Sci. ICHEP2020* (2021) 226.
- [82] B. Abi *et al.* (DUNE Collaboration), Deep underground neutrino experiment (DUNE), far detector technical design report, volume II: DUNE physics, [arXiv:2002.03005](#).
- [83] B. Abi *et al.* (DUNE Collaboration), Prospects for beyond the standard model physics searches at the deep underground neutrino experiment, *Eur. Phys. J. C* **81**, 322 (2021).

- [84] A. Abed Abud *et al.* (DUNE Collaboration), Snowmass neutrino frontier: DUNE physics summary, [arXiv:2203.06100](#).
- [85] A. Abusleme *et al.* (JUNO Collaboration), JUNO sensitivity on proton decay  $p \rightarrow \bar{\nu} K^+$  searches, [arXiv:2212.08502](#).
- [86] A. Abusleme *et al.* (JUNO Collaboration), JUNO physics and detector, [Prog. Part. Nucl. Phys.](#) **123**, 103927 (2022).
- [87] L. Moneta, K. Belasco, K. S. Cranmer, S. Kreiss, A. Lazzaro, D. Piparo, G. Schott, W. Verkerke, and M. Wolf, The RooStats project, Proc. Sci. ACAT2010 (**2010**) 057 [[arXiv:1009.1003](#)].