Novel approach to assess the impact of the Fano factor on the sensitivity of low-mass dark matter experiments

D. Durnford,* Q. Arnaud, and G. Gerbier
Department of Physics, Engineering Physics & Astronomy, Queen’s University, Kingston, Ontario K7L 3N6, Canada

As first suggested by U. Fano in the 1940s, the statistical fluctuation of the number of pairs produced in an ionizing interaction is known to be sub-Poissonian. The dispersion is reduced by the so-called “Fano factor,” which empirically encapsulates the correlations in the process of ionization. In modeling the energy response of an ionization measurement device, the effect of the Fano factor is commonly folded into the overall energy resolution. While such an approximate treatment is appropriate when a significant number of ionization pairs are expected to be produced, the Fano factor needs to be accounted for directly at the level of pair creation when only a few are expected. To do so, one needs a discrete probability distribution of the number of pairs created \( N \) with independent control of both the expectation \( \mu \) and Fano factor \( F \). Although no distribution \( P(N|\mu, F) \) with this convenient form exists, we propose the use of the COM-Poisson distribution together with strategies for utilizing it to effectively fulfill this need. We then use this distribution to assess the impact that the Fano factor may have on the sensitivity of low-mass WIMP search experiments.

I. INTRODUCTION

Following an ionizing particle interaction of deposited energy \( E \), the mean number of pairs created is given by

\[
\mu = \frac{E}{W(E)},
\]

where \( W \) is generally both energy and particle-type dependent and represents the mean energy needed to create either one electron-ion pair in liquid and gaseous detectors or one electron-hole pair in semiconductor devices [1]. The actual number of pairs created \( N \) is subject to statistical fluctuations, which limits the achievable energy resolution of any ionization measuring device to monoen energetic. As first anticipated by Fano, the variance of these fluctuations \( \sigma_N^2 \) is lower than expected for a Poisson process by a factor \( F \), known hereafter as the “Fano factor,” which is defined as [2]

\[
F = \frac{\sigma_N^2}{\mu}.
\]

By definition \( F = 1 \) for a Poissonian process, whereas \( F < 1 \) for ionization fluctuations. Experimentally measuring the Fano factor is challenging, as one needs to both strongly suppress and precisely quantify all sources of resolution degradation that do not arise from ionization fluctuations. In spite of this, measurements of \( F \) have been carried out for a wide variety of materials including argon \( (F \approx 0.23) \), xenon \( (F \approx 0.17) \), silicon \( (F \approx 0.16) \), germanium \( (F \approx 0.12) \), and others [3–6].

Although these measurements set a fundamental upper limit on the resolution possible with these detector media, they do not provide more information about the actual probability distribution of the number of pairs created than its dispersion. While the latter can be predicted with Monte Carlo simulations of the processes involved in energy loss at a microscopic level [7] (see Supplemental Material [8]), this approach is too computationally expensive to be practical for most applications. This is true for any scenario in which one needs to simulate the measurement of a signal that is not monoen energetic. In this case, one might think to fold the effect of the Fano factor into the overall energy resolution. While such an approach is appropriate at high energies, it is not valid when \( \mu \) is small, in which case a more accurate treatment is necessary. To account for the Fano factor at the level of pair creation, one would require a discrete probability distribution \( P(N|\mu, F) \) of the number \( N \) of pairs created for any value of \( \mu \) and \( F \). Although there is no distribution with this exact convenient form, we propose the COM-Poisson distribution as a viable solution. It is a discrete distribution that allows for independent control of the mean and variance with two parameters, \( \lambda \) and \( \nu \). While these variables do not correspond to \( \mu \) and \( F \), we have developed a methodology to effectively translate \( P(N|\lambda, \nu) \) into \( P(N|\mu, F) \). We believe the COM-Poisson distribution may provide a much needed tool in the area of low-mass dark matter research. Since new, popular models favor particle masses on the order of a few GeV/c\(^2\) or less [9, 10], a growing cohort of direct detection experiments are now confronted with the issue of modeling ionization statistics at the single pair regime. This includes gaseous dark matter search experiments like NEWS-G [11], liquid noble experiments like DarkSide [12], and solid-state experiments such as SuperCDMS, Edelweiss, DAMIC, and Sensei [13–16]. While these detector technologies differ in many ways, the requirements for modeling ionization statistics are essentially the same for each, and are fulfilled by the COM-Poisson distribution.

What follows is a more detailed discussion about the

* d.durnford@queensu.ca
problem of modeling ionization statistics (Sec. II), the COM-Poisson distribution (Sec. III), and strategies for using it (Sec. IV). Finally, we use the COM-Poisson distribution to assess the potential impact of the Fano factor on the sensitivity of dark matter detection experiments in Sec. V.

II. MODELING IONIZATION STATISTICS

A detector’s energy response to monoenergetic radiation is defined as the convolution of the probability distribution of ionization with the detector’s energy resolution function. At high energies when one expects a large number of pairs to be created, the detector response tends to a Gaussian due to the central limit theorem [17]. Therefore the Fano factor can approximately be accounted for by including it in the standard deviation of the overall energy response. However, this approach is not appropriate at the single-pair regime, which particle detectors can now probe as experiments push the low-energy frontier. This, together with the improved understanding/reduction of other resolution degrading factors means that the Fano factor must be accounted for directly at the level of pair creation. Doing so in a Monte Carlo simulation would ideally require a probability distribution $P(N|\mu, F)$ to model the probability of creating $N$ pairs. We considered four minimal requirements for a probability distribution to be appropriate for this task:

1. That the probability distribution is discrete.
2. That it allows for independent control of both the mean and the Fano factor.
3. That it is defined for continuous values of the mean.
4. That it is defined for values of the Fano factor within an appropriate range. Specifically, we required that the distribution be defined for any value of $F$ that is $\leq 1$ down to $\sim 0.1$.

This is not an exhaustive list of requirements for a model; one could consider other distribution shape properties such as kurtosis and skewness for example. However, in this work we concern ourselves only with these minimum requirements, as fulfilling even these is challenging. There are several well-known discrete probability distributions to consider for this purpose, as well as several potential candidates. One distribution first considered is the binomial distribution, with probability distribution function (PDF):

$$P(X = k|n,p) = \binom{n}{k} p^k (1 - p)^{n-k}$$

for $n \in \mathbb{N}_0$, $k \leq n$, $p \in [0,1]$. (3)

It is a discrete distribution (it satisfies the first requirements), with a mean of $\mu = np$ (which satisfies the third requirement). The variance is given by $np(1-p)$, and so it does allow the mean and the Fano factor to be varied. However, we can express the Fano factor for the binomial distribution as

$$F = 1 - p = 1 - \frac{\mu}{n}.$$  (4)

Thus, while the binomial distribution can be used with some values of $F$, it is not an appropriate model. Because $n$ is an integer, the distribution is not defined for continuous values of $F$ as we require, and the allowed values of $F$ vary with $\mu$ [8]. There are also many niche distributions designed specifically to satisfy the need for underdispersed or overdispersed models. Examples of these include the negative-binomial distribution, a good tool for overdispersion only [18]. Another is the generalized Poisson distribution, which can also model underdispersion, but there is a lower limit on the Fano factor achievable with this distribution, so it is also inappropriate [18, 19]. A class of distribution that can satisfy our requirements is the family of weighted Poisson distributions. This includes anything that can be written in the form [20]

$$P(X = x|\lambda, w) = \frac{e^{-\lambda} \lambda^x w_x}{W x!}$$

for $x \in \mathbb{N}_0$, $\lambda > 0$, (5)

where $w_x$ is a weight function (usually with two parameters itself) and $W$ is a normalizing constant. As a solution to the problem of modeling ionization statistics we propose the COM-Poisson distribution. It is a member of the family of weighted Poisson distributions, with weight function $w_x = (x!)^{1-\nu}$ [18]. While the distribution parameters $\lambda$ and $\nu$ do not correspond to the mean and the Fano factor, having only two parameters makes it easier to translate the COM-Poisson distribution $P(N|\lambda, \nu)$ into $P(N|\mu, F)$. Thus it is a more “user-friendly” case of a weighted Poisson distribution. Additionally, our choice of the COM-Poisson distribution benefits from studies of its properties by others [21–23], many of which we make use of (see Sec. III). It meets all of our requirements: it is discrete, allows for independent control of the mean $\mu$ and $F$, and is defined for continuous values of $\mu$ and $F$ (including $F > 1$).

III. THE COM-POISSON DISTRIBUTION

The Conway-Maxwell-Poisson distribution (COM-Poisson) is a two-variable generalization of the Poisson distribution, first proposed by Conway and Maxwell for application to queuing systems [24]. In more recent years, it has garnered attention for its utility in modeling underdispersed and overdispersed data [21]. It has found use in marketing, biology, transportation, and a variety of other applications [22]. The COM-Poisson probability distribution function for a random variable $X$ is defined as [22]

$$P(X = x|\lambda, \nu) = \frac{\lambda^x}{(x!)^\nu} Z(\lambda, \nu)$$

for $x \in \mathbb{N}_0$, $\lambda > 0$, $\nu \geq 0$, (6)

where $Z$ is a normalizing constant.
where $Z(\lambda, \nu)$ is a normalizing constant:

$$Z(\lambda, \nu) = \sum_{s=0}^{\infty} \frac{\lambda^s}{(s!)^\nu}. \quad (7)$$

The parameter $\nu$ controls the dispersion of the distribution. In particular, having $\nu > 1$ will result in underdispersion, and $\nu < 1$ overdispersion. In the special case of $\nu = 1$ the COM-Poisson distribution reduces to the regular Poisson distribution, and $\lambda$ simply becomes the expectation value. The COM-Poisson distribution also reduces to the Bernoulli distribution in the limit $\nu \to \infty$, as well as the geometric distribution when $\nu = 0 \,[21]$. However, this representation does not easily lend itself to computation, so the mean $\mu$ and variance $\sigma^2_N$ can instead be expressed as infinite sums by substituting Eq. (7) into Eq. (9):

$$\mu(\lambda, \nu) = \sum_{s=0}^{\infty} \frac{s\lambda^s}{(s!)^\nu} Z(\lambda, \nu),$$

$$\sigma^2_N(\lambda, \nu) = \sum_{s=0}^{\infty} \frac{s^2\lambda^s}{(s!)^\nu} Z(\lambda, \nu) - \mu^2. \quad (10)$$

As with the normalizing constant $Z$, these sums converge relatively quickly and are easy to compute to arbitrary precision. One useful property of the COM-Poisson distribution is that values of the PDF can be calculated recursively [21] using

$$P(X = x - 1) \frac{P(X = x)}{P(X = x)} = x^\nu \frac{\lambda^x}{\lambda}. \quad (11)$$

IV. USING THE COM-POISSON DISTRIBUTION

To model ionization statistics, one needs to be able to independently specify the mean $\mu$ and Fano factor $F$ of the discrete probability distribution function of the number $N$ of pairs created. Because the distribution parameters of COM-Poisson do not correspond to these variables, a method to obtain $\lambda$ and $\nu$ yielding the desired $\mu$ and $F$ is needed. In other words, one needs a mapping between $(\lambda, \nu)$ and $(\mu, F)$ parameter space to reexpress the COM-Poisson distribution $P(N|\lambda, \nu)$ as $P(N|\mu, F)$. What follows is a discussion of strategies to do so in different regions of $(\mu, F)$ parameter space. The three regimes addressed in the subsections below are identified in Fig. 1.

![FIG. 1. Strategies for using the COM-Poisson distribution shown in $(\mu, F)$ parameter space (top) and $(\mu, \sigma^2_N)$ parameter space (bottom). The blue shaded region represents parameter space where the optimization algorithm and look-up tables will be used (see Sec. IV C). The red shaded region represents the regime where the asymptotic expressions will be used (see Sec. IV A). In the overlap region, the asymptotic expressions are comparably accurate to the optimization algorithm, so points in the look-up tables in this area are calculated with the more accurate of the two. These regions are bounded by the limits of $F$ that we are concerned with (orange dashed lines), namely $0.1 \leq F \leq 1$, and by the Bernoulli modes (purple bumps).]
A. Asymptotic regime

Work has been done to develop an asymptotic expression for $Z$, which to first order can be expressed as [21]

$$Z_{\text{asymp}}(\lambda, \nu) = \frac{e^{\nu^{1/\nu}}}{\lambda^{(\nu-1)/(2\nu)}(2\pi)^{(\nu-1)/2}\sqrt{\nu}} \left(1 + O\left(\lambda^{-1/\nu}\right)\right). \tag{12}$$

This asymptotic expression is nominally accurate when $\nu < 1$ or when $\lambda > 10^{\nu}$ [21]. Beyond making the computation of $Z$ easier, this expression has greater applicability in the context of this work. It can be used to approximate the mean and variance of the distribution for $\lambda$ and $\nu$ by substituting this expression into Eq. (9). From this we have [21, 23]

$$\mu \approx \lambda^{1/\nu} - \frac{\nu - 1}{2\nu}, \quad \sigma_{N}^{2} \approx \frac{1}{\nu} \lambda^{1/\nu}. \tag{13}$$

Thus there are closed form expressions for the mean and the Fano factor wherever $Z_{\text{asymp}}$ is accurate. We can now treat these as a system of equations and solve for the distribution parameters as functions of $\mu$ and $F$:

$$\nu(\mu, F) \approx \frac{2\mu + 1 + \sqrt{4\mu^{2} + 4\mu + 1 - 8\mu F}}{4\mu F},$$

$$\lambda(\mu, F, \nu) \approx (\mu F)^{\nu}. \tag{14}$$

These expressions provide a way to calculate $\lambda$ and $\nu$ directly. For the sake of simplicity, we choose to utilize them above $\mu = 20$ (the red shaded region in Fig. 1) where we have verified that they are accurate to 0.01% or better for both $\mu$ and $F$. The asymptotic approximation for $Z$ solves several other problems as well. As mentioned in Sec. III, $Z$ converges quickly in the case of underdispersion ($\nu > 1$), but not so in the case of overdispersion [22], so it may be necessary to use Eq. (12) for $\nu < 1$. Another computational challenge with $Z$ arises for underdispersion when $\lambda$ becomes very large, at which point $Z$ itself becomes so large that it is not storable as a normal double-precision value. In this case, it is necessary to calculate log $Z$ directly using Eq. (12).

B. Bernoulli modes

Because the COM-Poisson distribution is discrete, there is a fundamental lower limit to the variance achievable with it, which is a function of the mean. To see how this is the case, consider a situation in which one has integer data with a mean of 0.5 and a variance of 0.25. This necessarily means that there are an equal number of counts of 0 and 1. In fact, it is impossible for the data to have a smaller variance, as this would require more counts to be either 0 or 1, which would shift the mean of the data. A slightly smaller or larger mean allows for a smaller variance, but there is still a minimum variance given by $\mu(1-\mu)$. This is the regime where discrete probability distributions reduce to the Bernoulli distribution, which the COM-Poisson distribution does when $\nu \to \infty$.

This also applies to larger values of the mean, creating an unending series of “Bernoulli modes” shown in Fig. 1 defining the minimum possible variance of discrete data, which are given by

$$\sigma_{N_{\text{min}}}^{2} = -(\mu - \mu_{i})(\mu - \mu_{i} - 1) \tag{15}$$

with $\mu_{i} \in \mathbb{N}$ such that $\mu_{i} - 1 < \mu < \mu_{i}$. This means that some $(\mu, F)$ parameter space is fundamentally inaccessible by the COM-Poisson distribution or any other discrete distribution. As a consequence of this, the Fano factor of any material necessarily tends to 1 when $\mu \ll 1$. Using the COM-Poisson distribution near the Bernoulli modes is also difficult as $\nu \to \infty$. To avoid this issue, we transition to using the Bernoulli distribution when within a distance of 0.1% in $(\mu, F)$ parameter space of the Bernoulli modes.

C. Optimization algorithm

At this point, we have addressed a large portion of $(\mu, F)$ parameter space. At low values of $F$, the Bernoulli modes provide a lower bound on the accessible parameter space. We know empirically that $F \leq 1$ and that $F \geq 0.1$ for most detectors. At $\mu \gtrsim 20$ we have asymptotic expressions for $\lambda$ and $\nu$ as functions of $\mu$ and $F$ that allow us to control the COM-Poisson distribution directly. At even higher values of $\mu$ there is also the option to use a Gaussian distribution to model ionization statistics to save computation time.

However, there is still a large region of parameter space yet to be addressed between the Bernoulli modes and the asymptotic regime, shown in blue in Fig. 1. Thus another strategy to determine the distribution parameters $\lambda$ and $\nu$ for a desired $\mu$ and $F$ is necessary. The approach we employed was to use an optimization algorithm. We show an illustrative example of this method in Fig. 2. For given values of $\lambda_{i}$ and $\nu_{i}$, we calculated the relative error between the desired mean and variance $(\mu$ and $\sigma_{N}^{2})$ and obtained mean and variance $(\mu_{i}$ and $\sigma_{N_{i}}^{2})$ calculated with Eq. (10). In both cases there is a valley of $(\lambda, \nu)$ parameter space giving the desired mean and variance. To reduce this problem to a single scalar minimization problem, and to obtain the unique $\lambda$ and $\nu$ that give us the mean and variance we want, we define the following quantity $X$:

$$X(\lambda, \nu|\mu, \sigma_{N}^{2}) = \left( w_{1} \left| \frac{\mu - \mu_{i}}{\mu} \right| + w_{2} \left| \frac{\sigma_{N}^{2} - \sigma_{N_{i}}^{2}}{\sigma_{N}^{2}} \right| \right)^{p}. \tag{16}$$

This is the weighted average of the relative error in mean and variance for a given $\lambda$ and $\nu$. This quantity is then minimized to find the desired values of $\lambda$ and $\nu$. The weights $w_{1}$ and $w_{2}$ and exponent $p$ were tuned to
FIG. 2. Illustrative example of the minimization algorithm for $\mu = 0.5$ and $F = 0.6$. First (left) the relative error in the mean $\delta(\mu)$ is shown as a function of $\lambda$ and $\nu$, and then (middle) the relative error in the variance $\delta(\sigma_N^2)$. Both plots clearly show a “valley” where the desired mean and variance are obtainable. Then (right) the weighted average of the relative error in mean and variance (the quantity $X$) is plotted as a function of $\lambda$ and $\nu$, effectively overlapping the two “valleys.” Thus, the values of $\lambda$ and $\nu$ giving the desired $\mu$ and $F$ can be found by minimizing $X$.

V. APPLICATIONS

We now make use of this tool to assess how ionization statistics affects the sensitivity of particle detectors to low energy events. In order to remain as general as possible, we first study the impact of the Fano factor on an experiment’s detection efficiency. This approach has the advantage of not needing to make any assumptions about the spectral shape of the signal being observed. In Fig. 4 we show detection efficiency curves obtained from Monte Carlo simulations where for each given expected
number of pairs $\mu$, the actual number of pairs created is randomly drawn $10^6$ times from the COM-Poisson distribution with a fixed value of $F$. The detection efficiency is then calculated as the proportion of events reconstructed above threshold and is shown as a function of energy in units of $\mu$. To assess how energy threshold and resolution (e.g., baseline fluctuations) may change the impact of the Fano factor, Fig. 4 shows detection efficiency curves computed for three different scenarios of threshold $E_{th}$ and detector resolution, which is modeled with a Gaussian distribution of standard deviation $\sigma$. The three scenarios considered are $E_{th} = 1 e^-/\sigma = 0.25 e^-$ (top panel), $E_{th} = 4 e^-/\sigma = 1 e^-$ (middle panel), and $E_{th} = 1 e^-/\sigma = 0.25 e^-$ (bottom panel). For each, the ratio with respect to $F = 1$ is shown in the subpanel.

From these results we can observe that in all scenarios, lower values of $F$ increase sensitivity to events with $\mu > E_{th}$, as this results in a reduced probability of these being reconstructed below threshold. For events with $\mu < E_{th}$, one expects the effect to be reversed due to low values of $F$ decreasing the probability of an event being reconstructed above threshold, and this is indeed observed in the $4 e^-/\sigma$ threshold scenarios (lower panels). However, the trend is nontrivially different in the $1 e^-/\sigma$-threshold scenario in which all the efficiency curves quickly tend to overlap in the low energy regime. This should not be misinterpreted as the Fano factor having no impact on the detection efficiency, but rather that when $\mu \ll 1$ the Fano factor is naturally forced to converge to a threshold scenario in which the Fano factor over the whole energy range. Ultimately, an extremely small probability would make the detection efficiency almost insensitive to the Fano factor as resolution effects would dominate over fluctuations in pair creation. Finally, one can conclude from the lower panels of Fig. 4 that the impact of $F$ can be most extreme for the detection of events below a high energy threshold.

From the above, we can infer that the ramifications of the Fano factor will depend on how crucial subthreshold event detection is to an experiment. For direct dark matter detection experiments searching for low-mass weakly interacting massive particles (WIMPs), this is all the more true. They are faced with the challenge of measuring the minute recoil energies of target nuclei following a WIMP elastic scattering interaction. The theoretical recoil energy spectrum of these events is approximately an exponential distribution with a slope that increases and a maximum energy cutoff that decreases as the WIMP mass decreases [27, 28]. For very low WIMP masses, an experiment’s sensitivity may derive primarily, if not entirely, from subthreshold events. An in-depth study of how the Fano factor may affect specific, existing experiments is out of the scope of this paper. Indeed, this could strongly depend on numerous factors which vary from one experiment to another including the target atomic number $A$, quenching for that target (ionization yield), and energy resolution. Rather, we wish to demonstrate the importance of accounting for the Fano factor when deriving sensitivity to low-mass WIMPs and to show that the COM-Poisson distribution proves to be a useful tool to do so. To that end, we considered a hypothetical WIMP search experiment with a target arbitrarily chosen to be neon and the quenching parametrization used by [11]. We also incorporate the most generic detector energy resolution by modeling it with a Gaussian distribution. We use the WIMP recoil energy spectrum given by [27] with a local dark matter density of $\rho_x = 0.3 \text{GeV/cm}^3$ and standard halo parameters.

We show in Fig. 5 the background-free sensitivities derived for different values of the Fano factor in the same three threshold/resolution scenarios depicted in Fig. 4. These curves correspond to the 90% confidence level upper limits on the spin-independent WIMP-nucleon elastic scattering cross section ($\sigma_s$) an experiment would report in the absence of any signal, as a function of WIMP mass $M_x$. These were calculated assuming an arbitrary exposure of $1 \text{kg \cdot day}$, although background-free limits simply scale with exposure. In the first two cases, $F$ only has a significant impact at intermediate WIMP masses. At high masses, WIMP sensitivity primarily comes from events with energies far above threshold where the detection efficiency of the experiment is 100% regardless of $F$ as long as no upper analysis threshold is set. At low WIMP masses, sensitivity is dominated by Bernoulli-regime events with $\mu < 1$ where $F$ is naturally bounded. However, in the low threshold case (top), a smaller value of $F$ conveys a greater sensitivity (albeit only slightly) to $0.5 - 10 \text{GeV/cm}^2$ WIMPs, as the detection efficiency is always higher for lower $F$ (see Fig. 4 top). In the scenario of high threshold but poor energy resolution (middle), the effect of $F$ at intermediate masses is reversed and far greater in magnitude. This can be understood by considering that in the $0.3 - 2 \text{GeV/cm}^2$ mass range, most WIMP scattering events are below threshold, and so a larger value of $F$ can dramatically increase sensitivity (see Fig. 4 middle).

Finally, the bottom panel of Fig. 5 depicts an extreme scenario, combining a high threshold with good resolution. As with the previous scenarios, $F$ has essentially no effect at high mass. However, in this case the limits do not converge at low WIMP masses because the experiment is simply not sensitive to Bernoulli-regime events at all. For this reason the impact of the Fano factor at low masses is extreme, with small values of $F$ inducing a multiple order of magnitude reduction in sensitivity. Ultimately, at the lowest WIMP masses an experiment may not be sensitive at all depending on the value of $F$. 
FIG. 4. Detection efficiency of an experiment with a $1\,e^{-}$ (top) and a $4\,e^{-}$ (middle and bottom) energy threshold, using the COM-Poisson distribution to model ionization statistics and with $\sigma = 0.25\,e^{-}$ (top and bottom) and $\sigma = 1\,e^{-}$ (middle) Gaussian resolution, with given values of $F$ (where possible, see Sec. IV B). The ratio of the curves with respect to $F = 1$ are shown in the subpanels.

FIG. 5. Spin-independent WIMP-nucleon scattering cross section exclusion curves for a hypothetical experiment with a neon target, assuming a $1\,e^{-}$ (top) and a $4\,e^{-}$ (middle and bottom) energy threshold. The energy resolution of the experiment is simulated with the COM-Poisson distribution and Gaussian resolution with $\sigma = 0.25\,e^{-}$ (top and bottom) and $\sigma = 1\,e^{-}$ (middle). The ratios of the curves with respect to $F = 1$ are plotted in the subpanels where appropriate.
VI. DISCUSSION AND CONCLUSIONS

In this work, we have proposed a novel approach and developed the strategies and tools to account for the Fano factor at the level of pair creation with the COM-Poisson distribution. By using it to assess the impact that the Fano factor may have on the sensitivity of low-mass WIMP search experiments, we have demonstrated both the need for modeling ionization statistics and the usefulness of the COM-Poisson distribution to do so. Although there is no physical motivation for the choice of COM-Poisson other than its apparent suitability for this application, we would like to stress that there is no comparable alternative to the proposed approach at the time of this publication. Additional rationale for using the COM-Poisson distribution can be found in [8]. To encourage and facilitate the use of this tool by others, we have provided free access to the look-up tables discussed in Sec. IV C and the code to use them at [26]. We have also provided code to produce the detection efficiency curves shown in Fig. 4 as an illustrative example. The authors are open to providing assistance in using the tools developed and discussed in this work.

ACKNOWLEDGEMENTS

This research was undertaken, in part, thanks to funding from the Canada Excellence Research Chairs Program. The present work was performed within the NEWS-G Collaboration, and benefited from the input and feedback of its members.

APPENDIX: DETAILS OF OPTIMIZATION ALGORITHM AND LOOK-UP TABLES

The optimization algorithm described in Sec. IV C uses a combination of a built-in minimizer of ROOT as well as a grid search. This is necessary because the ROOT minimizer was often unable to find the minima of $X$ on its own (see Sec. IV C). First, for a desired $\mu$ and $F$, a box in $(\log_{10} \lambda, \nu)$ parameter space is defined based on the asymptotic approximations for $\lambda$ and $\nu$ [Eq. (14)]. Even at low values of $\mu$ and $F$ where these expressions are not very accurate, they can still serve as a reasonable initial starting point. Then, a “grid search” is performed in which the value of $X$ is computed (see Sec. IV C) for every point on a $100 \times 100$ grid of points in that box, and the grid point with the lowest value of $X$ is found. Next, another smaller box is set based on this point, and the minimum grid square is used as an initial guess for the ROOT Minuit2 algorithm. Specifically, a bound minimization is carried out with the “combined” method within the defined box (again, with $\lambda$ on a $\log_{10}$ scale). With this new minima (giving values of $\lambda$ and $\nu$), the $\mu$ and $F$ obtained are calculated with Eq. (10).

If these values yield the desired $\mu$ and $F$ within acceptable tolerance (0.1% in this case), then the algorithm terminates. If this is not the case, the process can be repeated iteratively within smaller and smaller sections of parameter space with randomly perturbed initial guesses and bounds for the box. These extra measures are rarely necessary, as the first efforts of the algorithm are usually sufficient. However, near the Bernoulli modes the values of $\lambda$ and $\nu$ can vary wildly (see Fig. 3), so difficulties do occur.

![Histograms of the error in $\mu$ and $F$ obtained with the look-up tables and asymptotic approximations for $10^6$ points randomly drawn in $(\mu, F)$ parameter space within $0 < \mu < 100$ and $0.1 < F < 1$.](image)

The version of our look-up tables used for this paper was made with $10,000$ logarithmically spaced values of $\mu$ from 0.001 to 20, and 1,000 linearly spaced values of $F$ from 0.1 to 1. Points that fall within 0.1% of Bernoulli modes are excluded. It should be noted that in this scheme, a large proportion of the points are skipped because of this criteria. However, it was important to maintain a greater density of points at lower values of $\mu$ where $\lambda$ and $\nu$ vary more rapidly, as well as maintaining a regular grid to make interpolation possible. The contents of the look-up table are shown in Fig. 3, where one can see that the values of $\lambda$ and $\nu$ vary considerably and have emergent features at low values of $F$, defying at-
tempts to conveniently parametrize them. As mentioned in Sec. IV C, ultimately this table should work not just for values of $\mu$ and $F$ directly contained in the table, but for any values through interpolation. To do this, simple code to perform a bilinear interpolation of $\log_{10} \lambda/\nu$ was written.

To assess this and ultimately the accuracy of the look-up tables and the asymptotic approximations a test was devised in which random points in $\mu/F$ were chosen within $0 < \mu < 100$ and $0.1 < F < 1$. The appropriate method [either look-up table with interpolation or Eq. (14)] were then used to calculate $\lambda$ and $\nu$ corresponding to the desired values of $\mu$ and $F$, and the error of the obtained vs desired $\mu$ and $F$ was calculated with Eq. (10). The results of this test for $10^6$ points is shown in Fig. 6. The maximum error obtained in $\mu$ or $F$ for any point was less than or equal to 0.1%, as desired, and often much smaller than this. More information about the accuracy of the look-up tables (and code to test this) can be found at [26]. In the future, incremental improvements may be made to the look-up tables or interpolation code, but we consider this to be sufficient validation of our COM-Poisson code.

[1] F. Sauli, *Gaseous Radiation Detectors: Fundamentals and Applications* (Cambridge University Press, Cambridge, England, 2014).
[2] U. Fano, “Ionization yield of radiations. II. The fluctuations of the number of ions,” *Phys. Rev.* 72, 26 (1947).
[3] A. Hashiba, K. Masuda, T. Doke, T. Takahashi, and Y. Fujita, “Fano factor in gaseous argon measured by the proportional scintillation method,” *Nucl. Instrum. Methods Phys. Res., Sect. A* 227, 305 (1984).
[4] A.J.P.L. Policarpo, M.A.F. Alves, M. Salete, S.C.P. Leite, and M.C.M. dos Santos, “Detection of soft X-rays with a xenon proportional scintillation counter,” *Nucl. Instrum. Methods Phys. Res., Sect. A* 118, 221 (1974).
[5] A. Owens, G.W. Fraser, and K.J. McCarthy, “On the experimental determination of the Fano factor in Si at soft X-ray wavelengths,” *Nucl. Instrum. Methods Phys. Res., Sect. A* 491, 437 (2002).
[6] B.G. Lowe, “Measurements of Fano factors in silicon and germanium in the low-energy X-ray region,” *Nucl. Instrum. Methods Phys. Res., Sect. A* 399, 354 (1997).
[7] B. Grosswendt, “Statistical fluctuations of the ionisation yield of low-energy electrons in He, Ne and Ar,” *J. Phys. B* 17, 1391 (1984).
[8] See Supplemental Material at [http://link.aps.org/supplemental/10.1103/PhysRevD.98.103013](http://link.aps.org/supplemental/10.1103/PhysRevD.98.103013) for a comparison of the COM-Poisson distribution with simulated ionization distributions.
[9] R. Essig *et al.*, “Dark sectors and new, light, weakly-coupled particles,” arXiv:1311.0029.
[10] K.M. Zurek, “Asymmetric dark matter: Theories, signatures, and constraints,” *Phys. Rep.* 537, 91 (2014).
[11] Q. Arnaud *et al.* (NEWS-G Collaboration), “First results from the NEWS-G direct dark matter search experiment at the LSM,” *Astropart. Phys.* 97, 54 (2018).
[12] P. Agnes *et al.* (DarkSide Collaboration), “Low-Mass Dark Matter Search with the DarkSide-50 Experiment,” *Phys. Rev. Lett.* 121, 081307 (2018).
[13] R. Agnese *et al.* (SuperCDMS Collaboration), “First Dark Matter Constraints from a SuperCDMS Single-Charge Sensitive Detector,” *Phys. Rev. Lett.* 121, 051301 (2018).
[14] Q. Arnaud *et al.* (EDELWEISS Collaboration), “Optimizing edelweiss detectors for low-mass wimp searches,” *Phys. Rev. D* 97, 022003 (2018).
[15] J.R.T. de Mello Neto *et al.* (DAMIC Collaboration), “The DAMIC dark matter experiment,” *Proc. Sci. ICRC2015*, 1221 (2016), 1510.02126.
[16] J. Tiffenberg, M. Sofo-Haro, A. Drlica-Wagner, R. Essig, Y. Guardincerri, S. Holland, T. Volansky, and T.-T. Yu, “Single-Electron and Single-Photon Sensitivity with a Silicon Skipper CCD,” *Phys. Rev. Lett.* 119, 131802 (2017).
[17] G. Cowan, *Statistical Data Analysis*, 1st ed. (Clarendon Press, Oxford, 1998).
[18] E.L. Plan, “Modeling and simulation of count data,” *CPT: Pharmacoetrics Syst. Pharmecol.* 3, 129 (2014).
[19] P.C. Consul and G.C. Jain, “A generalization of the poisson distribution,” *Technometrics* 15, 791 (1973).
[20] J. Del Castillo and M. Pérez-Casany, “Weighted Poisson distributions for overdispersion and underdispersion situations,” *Ann. Inst. Stat. Math.* 50, 567 (1998).
[21] G. Shmueli, T.P. Minka, J.B. Kadane, S. Borle, and P. Boatwright, “A useful distribution for fitting discrete data: Revival of the Conway-Maxwell-Poisson distribution,” *J. R. Stat. Soc. C* 54, 127 (2004).
[22] K.F. Sellers, S. Borle, and G. Shmueli, “The COM-Poisson model for count data: A survey of methods and applications,” *Appl. Stoch. Models Bus. Ind.* 28, 104 (2011).
[23] T. Minka, G. Shmueli, J.B. Kadane, S. Borle, and P. Boatwright, “Computing with the COM-Poisson distribution.” (2003).
[24] R.W. Conway and W.L. Maxwell, “A queuing model with state dependent service rates,” *J. Ind. Eng.* 132 (1962).
[25] C. Chanialidis, L. Evers, T. Neocleous, and A. Nobile, “Efficient Bayesian inference for COM-Poisson regression models,” *Stat. Comput.* 28, 595 (2018).
[26] [https://news-g.org/com-poisson-code](https://news-g.org/com-poisson-code).
[27] J.D. Lewin and P.F. Smith, “Review of mathematics, numerical factors, and corrections for dark matter experiments based on elastic nuclear recoil,” *Astropart. Phys.* 6, 87 (1996).
[28] R.W. Schnee, “Introduction to dark matter experiments,” *Proc. TASI09*, 775-829 (2011).