Intrinsic Fano factor of nuclear recoils for dark matter searches

M. Matheny , A. Roberts ,[*] A. Srinivasan, and A. N. Villano [†]

Department of Physics, University of Colorado Denver, Denver, Colorado 80217, USA

Nuclear recoils in germanium and silicon are shown to have much larger variance in electron-hole production than their electron recoil counterparts for recoil energies between 10 and 200 keV. This effect—owing primarily to deviations in the amount of energy given to the crystal lattice in response to a nuclear recoil of a given energy—has been predicted by the Lindhard model. We parametrize the variance in terms of an intrinsic nuclear recoil Fano factor that is 24.3 ± 0.2 and 26 ± 8 at around 25 keV for silicon and germanium, respectively. The variance has important effects on the expected signal shapes for experiments utilizing low-energy nuclear recoils such as direct dark matter searches and coherent neutrino-nucleus scattering measurements.

I. INTRODUCTION

One of the most intensely researched channels for direct detection of dark matter is scattering off of a nucleus in a target material [1–5]. While the ionization distribution for these recoils has never been well understood in solids, the Lindhard model [6] provides a benchmark. Experiments that simultaneously measured two deposition channels (such as ionization and heat) did not worry about the exact ionization distributions in the past because they could measure recoil energy directly [7,8] (without a model for ionization production). With the two-channel measurements (ionization and heat) the results of dark matter searches was not systematically limited due to the lack of knowledge on the ionization.

In recent years, there has been a dramatic improvement in the detection energy threshold of many experiments [9,10], due largely to improvements in the measurement resolution for ionization or heat individually. The best detectors of the new generation of low-mass dark-matter-seeking experiments have single electron-hole pair sensitivity [11,12]. These detectors have not yet been able to achieve the ionization-yield insensitivity that their higher-energy predecessors have, so the dark-matter signal depends sensitively on the ionization yield and ionization variance produced by a low-energy nuclear recoil. In fact, it is often true that dominant systematic uncertainties in dark matter limits come from the uncertainty in the ionization yield [9]. For single electron-hole devices the ionization variance also becomes a driving factor in the accuracy of

signal models for low-mass dark matter via nuclear scattering.

While much of the literature has focused on the ionization yield [13–16], there are existing published data that constrain the ionization variance either directly or indirectly [17–19]. And there are even more data still that might be used to more precisely measure the ionization variance if a resolution model was published [20].

We report here on the best such existing data to constrain the ionization variance in silicon and germanium, and provide a procedure by which such information can be extracted from a dark matter detector that measures two channels such as ionization and heat. While our constraints are limited to the recoil energy region above about 24 keV, the techniques give insight into how this information can be extracted to lower energies in the future and the basic size and trend of the ionization variance for nuclear recoils.

To analyze the silicon data we have taken note that two previous publications have reported an "excess" ionization variance beyond the expected instrumental variance. We converted that variance into a Fano factor by using $\sigma_e^2 / \bar{N} = F$, where σ_e is the excess width in the ionization measurement.

We have used a similar method for a previous germanium publication. There, we used electron recoils to constrain the instrumental resolutions, obtaining an excellent fit to the measured widths for electron recoils (see Fig. 2). This instrumental resolution does not predict the nuclear recoil widths, so we include an "excess" variance in the form of a nuclear recoil Fano factor to obtain good fits. The method is similar to that of scintillator references [21] which have accounted for instrumental resolution in order to obtain information on the intrinsic optical photon production process. The key difference being that the result in that publication shows sub-Poisson fluctuations

---

[*]Corresponding author.
amy.roberts@ucdenver.edu
[†]Corresponding author.
anthony.villano@ucdenver.edu

123009-1

in optical photon production where we see larger-than-Poisson fluctuations in electron-hole pair production for nuclear recoils.

Our results for the intrinsic nuclear recoil Fano factors of silicon and germanium (see Figs. 1 and 6) are not in line with a division of some theoretical predictions of the electron recoil Fano factor [22] by the ionization yield; this naive modification gives $F \sim 0.41$. Those predictions, however, assume an underlying phonon distribution that is Poisson. Our results are roughly in line with the Lindhard [6] predictions which do not make that assumption.

II. THE FANO FACTOR

The ionization variance for electron recoils is very succinctly characterized in terms of the Fano factor, F [23]. Given an average number of electron-hole pairs produced, $\bar{N}$, the variance in this number of pairs is given simply by

$$\sigma_N^2 = F\bar{N}. \qquad (1)$$

While this specification does not give insight into moments of the N distribution of higher order than the variance, it emphasizes that $F = 1$ corresponds to a behavior that is *qualitatively* similar to a Poisson distribution in the lowest two moments. For these reasons we find it simple and convenient to parametrize the nuclear recoil ionization variance in the same way, but with a modified intrinsic Fano factor, F_{nr} [24]. The Fano factor for electron recoils seems to be in the range [25] 0.084–0.16 [26–28], but may have a temperature and/or energy dependence [29]. It is even possible that the "true" intrinsic Fano factor has not yet been measured directly and is lower than all the above measurements [30].

In silicon there are two studies that we are aware of that measured the ionization variance in addition to the ionization yield for nuclear recoils. Both were done in the early 1990s with secondary neutron beams produced from primary proton beams via the reaction $^7\mathrm{Li}(p, n)^7\mathrm{Be}$ [17,18]. The measurement by Dougherty makes use of neutron elastic-scattering resonances present in silicon. The measurement of Gerbier et al. uses a fixed-angle secondary neutron detector and a timing coincidence to constrain the true recoil energy in the silicon scattering detector.

Both of these measurements report the "extra" ionization variance after subtracting the expected variance due to known sources of errors such as instrumental noise or angular uncertainty in the secondary neutron scatters. The extracted additional ionization variance can be compared with the total recoil energy (inferred in the Dougherty measurement and measured in the Gerbier measurement) to give what we will define as the intrinsic fractional ionization width, ξ. This fractional ionization width is defined as the ionization width (in energy units) divided by the ionization energy collected, so that $\xi = \sigma_N/\bar{N}$. With these definitions the intrinsic nuclear recoil Fano factor, F_{nr}, is given by

$$F_{\mathrm{nr}} = \bar{N}\xi^2 = \frac{E_r \bar{Q}}{\epsilon_\gamma}\xi^2, \qquad (2)$$

where E_r is the true recoil energy, $\bar{Q}$ is the average ionization yield (ratio of "collected" ionization energy to total energy; unity for electron recoils), and ϵ_γ is the average energy to produce one electron-hole pair for an electron recoil.

Table I shows the resulting intrinsic nuclear recoil Fano factors for the silicon nuclear recoils measured in the two references we have been discussing. Even at low recoil energies, around 3 keV, the intrinsic Fano factors show that the ionization variance is such that the number of created pairs have more variance than a Poisson process with the same average number of pairs.

TABLE I. Data from the past publications constraining the intrinsic Fano factor in silicon [17,18]. The intrinsic Fano factor is calculated from Eq. (2) in the text, using the measurements of the noninstrumental widths and the ionization efficiency (yield).

| Recoil energy (keV)	Ionization efficiency (%)	Non-instrumental width (eV)	Non-instrumental width (%)	Effective fano	Reference
109.1 ± 0.7	51.4 ± 2	$\cdots$	6.1 ± 1.2	208 ± 82	Dougherty [17]
75.7 ± 0.4	45.6 ± 0.5	$\cdots$	5.3 ± 0.6	123 ± 28	Dougherty [17]
25.3 ± 0.3	35.5 ± 0.6	$\cdots$	3.6 ± 0.3	24.3 ± 4.1	Dougherty [17]
7.50 ± 0.03	26.9 ± 0.4	$\cdots$	2.8 ± 0.4	5.75 ± 1.65	Dougherty [17]
4.15 ± 0.15	22.5 ± 0.5	$\cdots$	2.2 ± 0.9	2.35 ± 1.92	Dougherty [17]
21.7 ± 0.2	40.7 ± 0.5	1000 ± 59	$\cdots$	29.80 ± 17.11	Gerbier [18]
19.5 ± 0.2	38.7 ± 0.7	1101 ± 108	$\cdots$	42.27 ± 8.33	Gerbier [18]
13.5 ± 0.3	33.6 ± 0.7	601 ± 42	$\cdots$	20.96 ± 2.96	Gerbier [18]
8.6 ± 0.1	31.1 ± 0.5	348 ± 13	$\cdots$	11.91 ± 0.91	Gerbier [18]
4.7 ± 0.1	26.6 ± 0.8	185 ± 36	$\cdots$	7.20 ± 2.81	Gerbier [18]
4.15 ± 0.1	27.4 ± 0.8	166 ± 39	$\cdots$	6.38 ± 3.00	Gerbier [18]
3.9 ± 0.1	22.9 ± 2.0	241 ± 66	$\cdots$	17.11 ± 9.49	Gerbier [18]
3.3 ± 0.1	25.9 ± 1.6	131 ± 55	$\cdots$	5.28 ± 4.45	Gerbier [18]
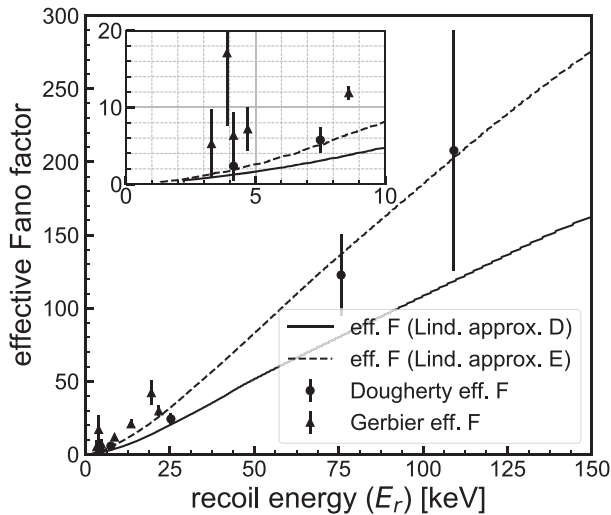
FIG. 1. The measurements of Dougherty [17] and Gerbier [18] converted into the intrinsic Fano factor for nuclear recoils. We also show the predictions of Lindhard [6] in the so-called Approximation D (solid) and Approximation E (dashed) curves (see text for descriptions). The inset shows a zoom of the low-energy region below 10 keV in recoil energy.

The model of Lindhard *et al.*, articulated in the early papers [6,31–33], contains predictions for the variance in the production of electron-hole pairs in a solid medium in addition to the *average* ionization (ionization yield). We compare this theoretical ionization variance with "extra" ionization variance extracted by Dougherty and Gerbier as a possible explanation.

Figure 1 shows the ionization variance results of the previous measurements by Dougherty and Gerbier cast in terms of the intrinsic nuclear recoil Fano factor. Lindhard's predictions—shown in terms of the intrinsic Fano factor—are also shown on the plot. The predictions shown employ two different approximations used in the Lindhard work: the approximate separability between electronic and nuclear energy deposits (referred to as Approximation D); and the additional assumption of forward-scattering dominance in nuclear collisions (referred to as Approximation E). Approximation E produces a lower ionization yield and a larger ionization variance. The lower ionization yield for Approximation E is expected because in that case nuclear collisions are assumed to transfer small amounts of energy and therefore contribute a smaller fraction of their energy to ionization. Furthermore, the larger ionization variance is a consequence of the approximate proportionality between F_{nr} and $\sqrt{\epsilon}$, where ϵ is the average energy needed to create a single electron-hole pair [22]—a quantity that increases with decreasing yield.

Despite clear evidence for a very large ionization production variance for nuclear recoils, and the importance of this variance for low-mass dark matter searches, studies of this effect are scant. Dark matter collaborations like SuperCDMS and EDELWEISS have excellent sensitivity

to this effect because of their direct measurements of ionization yield. In the next sections we argue that the large ionization variance expected in moderate-energy nuclear recoils produces larger-than-expected measured ionization yield widths in cryogenic semiconductor detectors, and that this fact can be used to measure the ionization variance for silicon or germanium.

III. PREVIOUS GERMANIUM IONIZATION YIELD MEASUREMENT

While the previously discussed measurements of the ionization variance in silicon came in the early 1990s, other technologies that came later had excellent means to probe the ionization variance in germanium. Two such similar technologies came out of the cryogenic dark matter searches of EDELWEISS and SuperCDMS [19,20].

EDELWEISS [19] was possibly the first to note in published work that the nuclear recoil band in cryogenic ionization/phonon devices is expected to be significantly *narrower* than the electron recoil band if only the effects of sensor resolution are included. Recently, the narrowness of the nuclear recoil band when using empirical resolution functions has also been noted in the SuperCDMS detectors [34]. The width of the nuclear recoil band is directly related to the variance of the ionization yield (or what EDELWEISS and some others call the "quenching"). In this work we use Q to denote the random variable corresponding to the *measured* ionization yield for an event, and $\bar{Q}$ to denote the average of that quantity, equivalent to the $\langle Q \rangle$ of EDELWEISS.

At a given recoil energy the width of the quenching measurement was estimated in the 2004 EDELWEISS work [19] by [35]

$$(\sigma_{\mathrm{NR}}^{\mathrm{EDW}})^2 = \frac{1}{E_r^2}\left(\left(1+\frac{V}{\epsilon_\gamma}\bar{Q}\right)^2\sigma_I^2 + \left(1+\frac{V}{\epsilon_\gamma}\right)^2\bar{Q}^2\sigma_H^2\right), \quad (3)$$

where σ_I^2 is the variance in the ionization signal in energy units and σ_H^2 is the variance in the heat signal in energy units. Since the quenching factor (less than unity for nuclear recoils) decreases each term in the equation, it is easy to see the variance in the event-by-event measured quenching should be significantly less for nuclear recoils than for electron recoils. In fact, this is not the case. EDELWEISS measures the variance in the nuclear recoils to be comparable to that of the electron recoils [19]. We have reproduced the EDELWEISS analysis by first computing the expected ionization yield width for electron recoils and then doing a simple fit to constrain how much larger the nuclear recoil width that EDELWEISS measures is from the prediction in Eq. (3).

For the electron recoils the average ionization yield $\bar{Q}$ is taken to be unity. EDELWEISS parametrized the energy-dependent sensor resolutions by the following functional forms [19]:
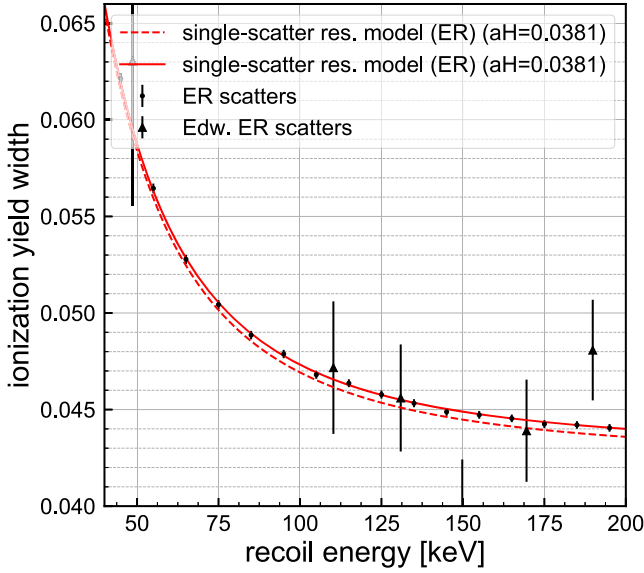
FIG. 2. The measured ionization width for electron recoils. The triangular data points are from the EDELWEISS [19] measurement on detector GGA3, and the circular data points are our simulation of that measurement with the "tuned" resolutions (see text). The solid curve is our exact model (σ_{ER}) for the ionization width given the appropriate resolution, and the dashed curve is the zeroth-order model (σ_{ER}^{EDW}) used by EDELWEISS.

$$\sigma_I(E) = \sqrt{(\sigma_I^0)^2 + (a_I E)^2},$$
$$\sigma_H(E) = \sqrt{(\sigma_H^0)^2 + (a_H E)^2}, \tag{4}$$

where a_I and a_H are adjustable parameters used to "tune" the ionization yield width. Using these resolutions we can compute the *exact* ionization yield width as a function of energy [36]. The resolutions can also be used to calculate the nuclear recoil width given an intrinsic Fano factor for nuclear recoils. The calculations are outlined in the Appendix.

Figure 2 shows the energy dependence of our exact expression for the electron recoil width (see the Appendix) with the EDELWEISS approximations [Eq. (3)]. Looking at the electron recoil band of the EDELWEISS GGA3 detector, we tune the parameter a_H to be 0.0381 in full width at half-maximum (FWHM) units [37]. This value gives the best fit to the EDELWEISS electron recoil yield widths as a function of energy, as shown in Fig. 2. Also in the figure we have displayed the widths resulting from a simulation of these distributions given the tuned sensor resolutions and our best expression for the electron recoil yield width given in the Appendix. We see that the EDELWEISS approximation to the yield width [Eq. (3)] is lower by an amount that seems unimportant for this analysis, given the precision of the electron recoil yield width data, but the exact expression (see the Appendix) matches the more precise simulation well. When we use this exact numerical PYTHON routine for the *nuclear-recoil*

yield each call takes around 1 min. Since our fitting routines need to call this function thousands of times we use an approximation that is higher order than the EDELWEISS approximation (so is more accurate, but not exact) but has a smaller computation time, making it usable for our purposes (see Sec. IV and the Appendix).

IV. ESTABLISHED GERMANIUM IONIZATION YIELD WIDTH

In the EDELWEISS publication [19] it is clear that the ionization yield width of nuclear-recoil events is systematically larger than expected. It is our goal to use a fitting technique to quantify precisely how much larger the measured ionization yield width for the EDELWEISS GGA3 detector is than expected [see Eq. (3) for the expectation] as a function of recoil energy.

It has been noted that the expected ionization yield width for nuclear recoils given in Eq. (3) is derived from a lowest-order "moment expansion" of the definition of the ionization yield random variable, Q. While this approximation is not bad for the electron-recoil ionization yield (see Fig. 2), it is not as accurate for the nuclear recoil version because of the smallness of $\bar{Q}$. For that reason a moment expansion out to order $1/E_r^6$—denoted by σ_{NR}^{SAI}—is used in our fitting for both the electron and nuclear recoil ionization yield width functions (see the Appendix for details).

In the EDELWEISS publication [19] the following functional form for the ionization yield is used because it fits the mean of the ionization data well:

$$\bar{Q} = A E_r^B. \tag{5}$$

We adopt this form of the average ionization yield in order to extract the "additional" ionization yield width. EDELWEISS has extracted this additional yield width by assuming a constant, called C, needs to be added in quadrature to the result of Eq. (3) and using the measured ionization yield widths to fit for the value of that parameter. We execute a similar fit, using the EDELWEISS measured points for the detector GGA3, and the corresponding resolutions but with a slightly more flexible function that allows C to be a function of the recoil energy: $C(E_r) = C_0 + m \cdot E_r$, with C_0 and m parameters. In our fit the more exact curve for the expectation of the ionization yield width [derived from Eq. (A2)] is used. We use a Markov chain Monte Carlo (MCMC) technique [38] to be sure to populate the full posterior distribution in the parameter space and account properly for parameter correlations. To incorporate systematics the fit is taken over a six-dimensional space: C_0, m, A, B, a_H, and η. The last variable is a fractional multiplier applied to the detector voltage to account for possible measurement deviations in that detector setting.

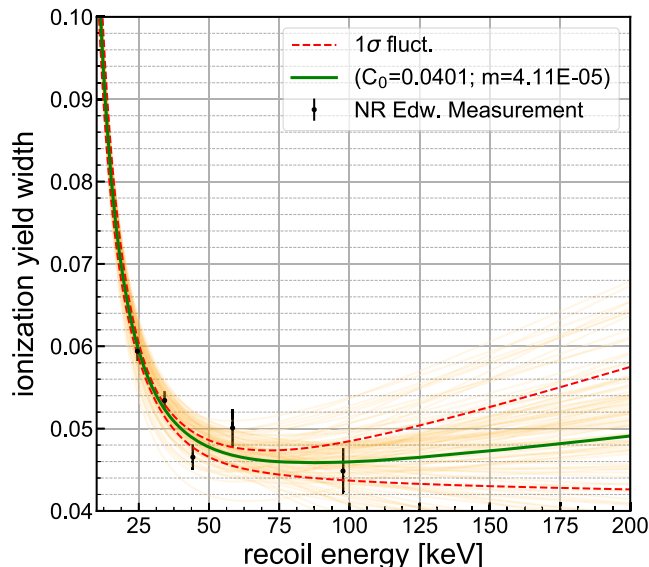The result of the fit is shown in Fig. 3 with the maximum likelihood curve for the extracted nuclear recoil yield

FIG. 3. Our fit to the nuclear recoil ionization width using the MCMC procedure. The solid curve is the maximum-likelihood fit to the C function, and the dashed lines are the assessed 1σ statistical uncertainty bounds. The data points are the EDELWEISS [19] measured values for detector GGA3, and the transparent curves are a sampling of 100 realizations of $C(E_r)$ using parameters pulled from the posterior parameter distributions.

bandwidth and several randomly sampled curves from the correct posterior distribution. For the full reproducible code for this fit see the public data release [39].

The nuclear recoil bandwidth is well-reproduced by using the fitted $C(E_r)$ added in quadrature to the base-level estimate. The base-level estimate is given in the Appendix and is symbolically referred to as $\sigma_{\text{NR}}^{\text{SAI}}$. Given the flexibility of our exact model for the ionization yield distribution, we proceeded to use this $C(E_r)$ and its associated error to obtain the variance on N as a function of recoil energy, parametrized by the intrinsic nuclear recoil Fano factor, F_{nr}.

V. MULTIPLE SCATTERING CORRECTION

For the EDELWEISS data the nuclear recoil ionization yield information is generated via scattering of neutrons from a ^{252}Cf source. Because of the use of neutrons, multiple scattering is an obvious effect that will increase the measured ionization yield width. The EDELWEISS study [19] accounted for this effect using a Monte Carlo simulation and concluded:

Although multiple interactions tend to lower $\langle Q \rangle$; this effect remains weak, and the Q distribution associated with single interactions events is only slightly narrower and completely included in the wider band.

the "wider band" being the band that encompasses the measured nuclear recoils. In other words, multiple scattering cannot account for the full observed ionization yield width.

We have resimulated the effect of multiple scattering in a detector that matches the EDELWEISS GGA3 germanium detector (approximately cylindrical with 70 mm diameter and 20 mm thick). For this we have used a GEANT4 [40,41] simulation where the geometry—aside from the germanium detector—was not made identical to the EDELWEISS setup, but where generic elements such as typical cryostat materials and polyethylene shielding were included. Our specific geometry (from inside to outside) included the following: the germanium detector; an electronics "tower" made mostly of copper with small amounts of insulating carbon; an "inner vacuum chamber" wall made of stainless steel; liquid helium; a stainless steel Dewar with vacuum jacket; and a rectangular polyethylene shield and supporting structure (aluminum). The source is located between the Dewar and polyethylene shield, 66 cm below the detector at a radial distance of 35 cm from the cylindrical axis of the Dewar and germanium detector.

Our simulation uses GEANT 4.10.1.p02 and the so-called "Shielding" physics list [42]. The main attribute of this physics list in the context of our analysis is the high-precision neutron-scattering library for neutron energies below 20 MeV. The use of this "NEUTRONHP" library [43,44] gives more precise realizations of the nuclear recoils because of the implementation of the detailed low-energy neutron interaction library G4NDL. A small drawback of the library is that it sacrifices strict energy-momentum conservation on an event-by-event basis, but that is not an important deterrent for this study since the recoil spectrum is more correct.

While the simulation setup does not match the EDELWEISS geometry, we point out that the geometry will principally affect the energy distribution of the neutron flux near the detector. The yield width is insensitive to that distribution because all of our scattering neutrons lie above 20 keV where the elastic scattering cross section is away from the resonance region, relatively flat, and well known [45]. Therefore the distribution of multiple scatters *within* the detector—which does affect the Q distribution—will not depend strongly on the energy distribution of the neutron flux or the geometry, but rather if the germanium elastic cross sections used are close to reality. The elastic cross sections used in our version of "NEUTRONHP" are in an energy region that has been well measured and match other evaluations such as the JENDL 5.0 evaluation [45].

We use these simulated data by applying the ionization yield model used in Sec. III. More precisely, we "tune" the sensor resolutions in the same way as produced the best match to the electron recoil bandwidth, take the ionization yield to be $\bar{Q} = 0.16E_r^{0.18}$, and take the intrinsic nuclear recoil Fano factor to be zero.

The simulated ionization yield distributions in Fig. 4 show that the single-scatter contribution has a clearly higher average yield than the distribution that includes all scatters. However, the width of the distribution is only
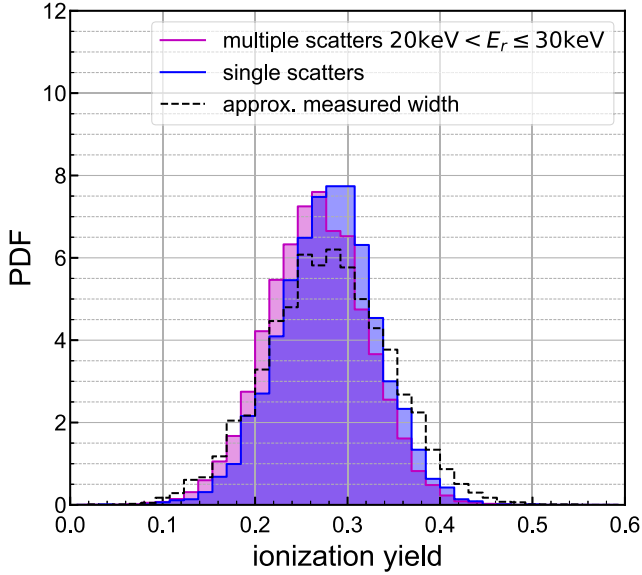
FIG. 4. Simulated ionization yield histograms of the single-scatter distribution (blue) and all-scatter distribution (magenta) for the energy range 20–30 keV. The black dashed histograms are single scatters with an "extra" resolution applied to bring the width of the distribution in line with what was measured by EDELWEISS as discussed in their publication [19].
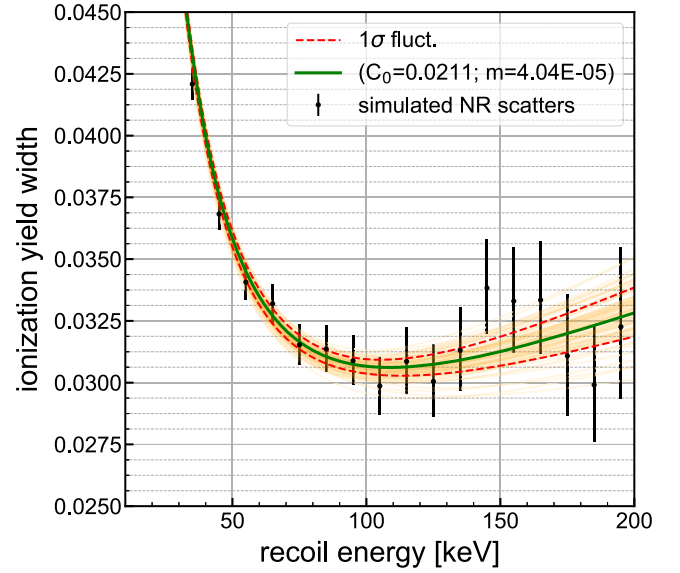

FIG. 5. A fit to the simulated multiple-scatter ionization width using the GEANT4 recoil data, and our yield model. The points are the ionization yield widths of the simulated data, and the solid line is the maximum likelihood fit to the C_m function. The dashed lines are the assessed 1σ statistical uncertainty bounds. The transparent curves are a sampling of 100 realizations of $C_m(E_r)$ using parameters pulled from the posterior parameter distributions.

modestly wider over the energy range shown (20–30 keV). The empirical distribution (Fig. 4 black dashed histogram) is clearly significantly wider than our simulation with multiple scatters included—a feature that gets more significant with increasing energy.

We have systematically fit the distribution widths from the simulation as a function of energy and compared them with the single-scatter width predictions discussed previously. Figure 5 shows the ionization widths that result for a full *simulated* ^{252}Cf dataset with multiple scattering included. Of course, the resulting ionization widths are larger than would result from a nuclear recoil sample consisting only of single scatters. Since our ionization yield model only makes predictions for single scatters, we compare the multiple scatters to that prediction to see how much wider the ionization yield distribution becomes. As in the previous section we fit a function $C_m(E_r) = C_{0m} + m_m E_r$ that describes the quadrature addition necessary to bring the single-scatter prediction in line with the simulated multiple-scatter results. In this case we do not let A, B, a_H, or η vary but set them equal to their best fit values from the MCMC in Sec. IV. The varying fit parameters are C_{0m} and m_m.

It is clear from the fit displayed in Fig. 5 that the quadrature addition needed to describe the effect of multiple scattering is observable but significantly *less* than what is required to describe the EDELWEISS ionization yield width data. This multiple-scatter correction to the yield widths will be used in Sec. VI to extract the required

additional correction needed to describe the EDELWEISS data. We argue that this additional correction is related to unaccounted uncertainty in the fundamental ionization production by nuclear recoils, and can be described by an intrinsic nuclear recoil Fano factor.

VI. EXTRACTING THE GERMANIUM INTRINSIC FANO FACTOR

We posit that the reason the measured ionization variance on EDELWEISS' GGA3 detector is larger than the expected when including multiple scattering (see Sec. V) is an unaccounted intrinsic ionization variance in the nuclear scattering process. We quantify this additional variance by taking the quadrature subtraction of the corrections extracted in Secs. III and V. The result is a correction, $C'(E_r)$, that is equal to the intrinsic ionization variance. Equation (6) shows the relationship of the intrinsic variance to the previous corrections,

$$C'(E_r) = \sqrt{C(E_r)^2 - C_m(E_r)^2}. \qquad (6)$$

Our intrinsic ionization variance is then converted into a Fano factor for nuclear recoils, F_{nr}, as advocated in Sec. II. The conversion to the nuclear recoil Fano factor is made by assuming the intrinsic variance is produced by simply increasing the nuclear recoil Fano factor from $F_{\text{nr}} = 0$ to some finite (positive) value within the framework of the
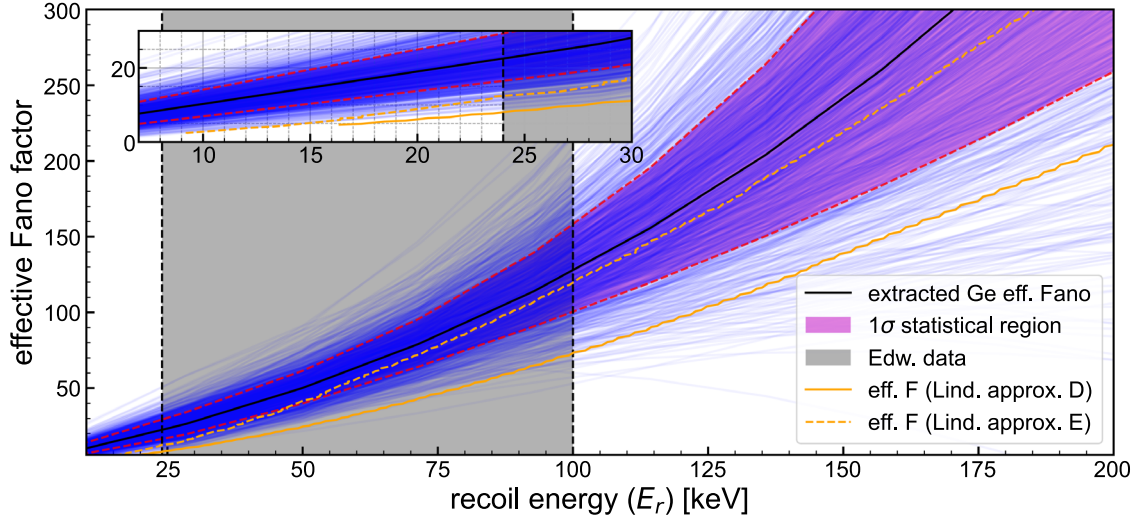
FIG. 6. The extracted nuclear recoil intrinsic Fano factor for germanium using the EDELWEISS [19] data. The black line is the best fit, the magenta shaded region denotes the 1σ statistical uncertainty region, and the inset shows a zoom of the region between 7 and 30 keV. The blue transparent curves are 1000 samples drawn from the MCMC posterior distribution—these give a sense for the presence of outlying behaviors. The gray shaded region is the region in which there are EDELWEISS yield width data; in that region the results are *not* an extrapolation. The Lindhard prediction for the nuclear recoil Fano factor are given in orange for Approximations D (solid lines) and E (dashed lines). To produce those predictions we used a Lindhard model for the ionization yield with parameter $k = 0.157$ [46].

model given in Eq. (A2) by setting the variance on the independent random variable N taken to be $\sigma_N = \sqrt{F_{\mathrm{nr}}\bar{N}}$. The actual value of $F_{\mathrm{nr}}(E_r)$ is then simply given by

$$\sigma_{\mathrm{NR}}^{\mathrm{SAI}}(E_r; F_{\mathrm{nr}}) = \sqrt{C'(E_r)^2 + \sigma_{\mathrm{NR}}^{\mathrm{SAI}}(E_r; F_{\mathrm{nr}} = 0)^2}. \quad (7)$$

Figure 6 shows the extracted intrinsic Fano factor, F_{nr} as a function of the recoil energy. The estimate for the uncertainties on the resulting $F_{\mathrm{nr}}(E_r)$ were obtained from the MCMC posterior distribution of all of the parameters $(A, B, a_H, \eta, C_0, m)$ in the original fit and the posterior distribution of the C_{0m} and m_m parameters in the multiples fit. A single realization of $F_{\mathrm{nr}}(E_r)$ is obtained by using a sample of the original and multiples fit and then subtracting them in quadrature to get $C'(E_r)$. Each sample of C' is turned into a sample of F_{nr} through Eq. (7). The maximum likelihood parameters are taken as the central value for F_{nr}, and we obtain the approximate 1σ deviations by taking the standard deviation of all samples at each energy—these are plotted as the magenta band in Fig. 6.

These uncertainties include the systematic uncertainty on the result with contributions from several parameters which, while nominally fixed, are not known with certainty. They are in order of decreasing importance: multiple scattering; a finite-binning uncertainty on the EDELWEISS ionization yield data; a possible departure of the quantity V/ϵ_γ from the nominal $4/3$ value (fit parameter η); charge trapping (fit parameters A and B); and

the functional form of the average ionization yield. The uncertainties are obtained by directly estimating the contribution (in the case of the finite binning) or including nuisance parameters in the six-parameter MCMC [38] fit to the EDELWEISS ionization yield width data for GGA3 for the extraction of $C(E_r)$. For each of the parameters representing the systematic uncertainties, a prior was chosen that was reflective of the state of knowledge on the parameters. The total uncertainty is estimated in Fig. 6, and the factional impact of each of the uncertainties or corrections is given in Table II.

TABLE II. The uncertainties and correction sizes for the extraction of the nuclear recoil intrinsic Fano factor for germanium using the EDELWEISS [19] data. The first column lists uncertainties with a (U) and corrections with a (C). The third column are the parameters in the fit related to that category (if any). The last column lists the fit parameters that have relevant correlations with that particular category.

Classification	Size (%)	Parameters	Relevant correlations
Statistical (U)	40–80	C_0, m	None
Multiple scattering (U)	< 6	None	None
Finite binning (U)	5	⋯	⋯
V/ϵ_γ (U)	< 20	η	a_H
Charge trapping (U)	< 20	A, B	C_0, m
Yield variation (U)	< 20	A, B	C_0, m
Multiple scattering (C)	60–70	None	None

VII. CONCLUSIONS

We have espoused the preference for quantifying the inherent uncertainty on the number of electron-hole pairs produced as an intrinsic Fano factor, F_{nr}, for nuclear recoils. We have also presented constraints on such a parameter from previous measurements on silicon and germanium–two important target materials for precision low-mass dark matter searches [47]. In the latter case we extracted meaningful F_{nr} measurements by a technique that can be adapted to low-threshold detectors measuring ionization and heat, but that did not require a specialized neutron scattering setup. We have used the Lindhard predictions as a guide, with the hope that future experiments will be able to distinguish between approximations in that work and/or inspire the development of a more accurate framework.

Our results indicate that the intrinsic nuclear recoil Fano factor is larger than expected for both silicon and germanium–24.3 ± 0.2 and 26 ± 8, respectively, at 25 keV recoil energy. The expectation in some literature is based on the assumption that the number of phonons created is a Poisson random variable [22]. In that case the electron recoil Fano factor is around 0.13 for germanium [48] and the intrinsic nuclear recoil Fano factor should be larger by about a factor of $1/\sqrt{\bar{Q}}$—still far lower than our suggested values. In the authors' view this would seem to indicate that for nuclear recoils the number of created phonons is *not* Poisson distributed and has a distribution that is significantly wider than naively expected; this wider distribution could then be imprinted on the electron-hole pairs in a way similar to the derivation in [22]. The authors do not see any reason why the number of phonons produced should have a Poisson distribution; in fact, the Lindhard references explicitly compute an ionization variance that is out of line with that assumption [6]. The Lindhard predictions for the intrinsic nuclear recoil Fano factor are shown in Fig. 6 for germanium and have an F_{nr} at least as large as 8 at 25 keV recoil energy. Those intrinsic nuclear recoil Fano predictions are not inconsistent with our measurement above around 50 keV but appear to be systematically lower than our measurement below 50 keV—perhaps due to an ionization yield that decreases more sharply toward lower recoil energy than the Lindhard theory suggests.

Based on our ionization yield model, which can describe EDELWEISS data well, the variance induced by the intrinsic Fano factor is correlated in its effect on ionization and heat resolutions. Roughly speaking, this means that the widening of "nuclear recoil bands" in low-threshold dark matter searches with discrimination capabilities (such as SuperCDMS [20] and EDELWEISS [49]) may be smaller than one would naively expect.

There is a lot of existing data that might be exploited using our technique but it is often true that precise resolution data are not published. If the sensor resolution is carefully extracted, then our technique might serve to extract F_{nr} more precisely for both silicon and germanium in the low-energy region. Such information is invaluable to low-mass nuclear recoil dark matter searches in silicon and germanium that employ detectors *without* nuclear recoil discrimination capabilities.

APPENDIX: CALCULATION OF σ_{NR} AND CROSS-CHECKS

1. Calculation of σ_{NR}

Both the EDELWEISS and SuperCDMS detectors can be correctly modeled by assuming the measurements of the ionization and heat depend on three (approximately) independent random variables: the number of electron-hole pairs created in a detectable interaction, N; the variation (noise fluctuations) in the ionization sensor, δI; and the variation in the heat detection δH. The distributions of δI and δH have zero mean and are approximately normally distributed with an energy-dependent standard deviation given by the ionization and heat sensor resolutions. The typical measured quantities in these experiments are specific combinations of those random variables defined thusly

$$\tilde{E}_r \equiv E_r + \left(1 + \frac{V}{\epsilon_\gamma}\right)\delta H - \frac{V}{\epsilon_\gamma}\delta I,$$

$$Q \equiv \frac{\epsilon_\gamma N + \delta I}{\tilde{E}_r}. \tag{A1}$$

The variable $\tilde{E}_r$ is the *measured* recoil energy, Q is the measured ionization efficiency (yield), E_r is the *true* recoil energy, and V is the voltage across the cylindrical detector. With this model if the sensor resolutions are published (or otherwise known), the only remaining things needed to predict exact distributions for all the measured quantities are the true recoil energy distribution (which can be simulated) and the distribution of the random variable N. The latter is directly related to the Fano factor or the intrinsic nuclear recoil Fano factor. Since N is rather high for recoil energies above $\sim$10 keV, the distribution is taken to be approximately normal, with the mean given by the average ionization yield at the particular recoil energy $[\bar{Q}(E_r)]$ and the width being given by the intrinsic Fano factor, F_{nr}.

We have done the exact calculation simply by recognizing the joint conditional probability distribution for $\tilde{E}_r$ and Q must have the following form:

$$P(Q, \tilde{E}_r | \delta H, \delta I, N, E_r) = \delta\left(\tilde{E}_r - \left[E_r + \left(1 + \frac{V}{\epsilon_\gamma}\right)\delta H - \left(\frac{V}{\epsilon_\gamma}\right)\delta I\right]\right)\delta\left(Q - \left[\frac{\epsilon_\gamma N + \delta I}{E_r + (1 + V/\epsilon_\gamma)\delta H - (V/\epsilon_\gamma)\delta I}\right]\right). \tag{A2}$$

Equation (A2) will correctly give the ionization yield (Q) distribution at a single measured energy or over a range of measured energies. The distribution may be close to normal for a wide range of parameters but not exactly normal. The distribution is especially far from normal when the heat or ionization has a large enough variance so that the measured recoil energy becomes consistent with zero. The ionization yield standard deviation with this "exact" calculation is referred to as σ_{NR}.

The procedure outlined above involves integrals that are difficult to accomplish analytically. For that reason slower numerical techniques are used and the computation time makes it difficult to use (around 1 min for one calculation at one energy and parameter-value point). In this work, as discussed in Secs. III, V, and VI, the fitting requires many evaluations of the function and so it must be approximated.

Part of the problem is not only the functional dependence on E_r, but the functional dependence on our nuisance parameters A, B, a_H, and η. In the general case—nuclear recoils with average yield modeled by the A and B parameters—we compute the "moment" expansion of Q in Eq. (A1) to order $1/E_r^6$. We refer to this expression as $\sigma_{\mathrm{NR}}^{\mathrm{SA}}$. For electron recoils, the agreement is quite good if we simply take this expansion with $A = 1$ and $B = 0$ (see Fig. 2). The expansion to lower order ($1/E_r^2$) is the expression used by EDELWEISS–$\sigma_{\mathrm{NR}}^{\mathrm{EDW}}$ [see Eq. (3)].

For nuclear recoils, the agreement is not as good, so we add a correction based on the preferred values of the nuisance parameters from our fit to the EDELWEISS data. Taking $A_0 = 0.149$, $B_0 = 0.178$, $a_{H0} = 0.038$, and $\eta_0 = 1.000$ we can use the exact function to create a static correction for use in the nuclear recoil case. This is the approximation we use to describe our nuclear recoil ionization yield widths in our fitting procedure of Sec. IV:

$$
\begin{aligned}
\sigma_{\mathrm{NR}}^{\mathrm{SAI}}(E_r, A, B, \eta, F_{\mathrm{nr}} = 0)^2 &= \sigma_{\mathrm{NR}}^{\mathrm{SA}}(E_r, A, B, \eta, F_{\mathrm{nr}} = 0)^2 \\
&+ [\sigma_{\mathrm{NR}}(E_r, A_0, B_0, \eta_0, F_{\mathrm{nr}} = 0)^2 - \sigma_{\mathrm{NR}}^{\mathrm{SA}}(E_r, A_0, B_0, \eta_0, F_{\mathrm{nr}} = 0)^2].
\end{aligned} \tag{A3}
$$

The form shown in Eq. (A3) is much faster to compute than the exact version, but gives ionization yield widths that differ from the exact model by at most 7% over our parameter space (E_r plus nuisance parameters).

2. Cross-check with electron recoil Fano factor

One excellent check for consistency of our method is to fit the electron recoil ionization yield band and extract the electron recoil Fano factor, F. We cannot accomplish this with the real EDELWEISS data [19] that we have used for the majority of this paper because the data are not precise enough (about 10% relative uncertainty) and the Fano contribution to the ionization yield variance is expected only to be around 0.1%. This is in contrast to the nuclear recoil Fano contribution that we have measured to be at least 10%.

Instead what we have done is simulate electron recoil band data in a similar way as was done in Fig. 2, the high-precision simulated data points. We selected a Fano factor of $F = 0.15$, adjusted the instrumental resolution so that the Fano contribution was about 1% in yield variance, and simulated the data with about 0.5% relative uncertainty on each data point. With this fit–using the same MCMC fitting method we used for nuclear recoils in this work—we extracted a Fano factor of $F = 0.13 \pm 0.08$, consistent with the Fano factor we set for the simulation ($F = 0.15$).

[1] Q. Arnaud *et al.*, Astrophys. Space Phys. Rev. **97**, 54 (2018).

[2] A. H. Abdelhameed *et al.* (CRESST Collaboration), Phys. Rev. D **100**, 102002 (2019).

[3] G. Angloher *et al.*, Eur. Phys. J. C **74**, 3184 (2014).

[4] E. Aprile *et al.* (XENON Collaboration 7), Phys. Rev. Lett. **121**, 111302 (2018).

[5] R. Agnese *et al.* (SuperCDMS Collaboration), Phys. Rev. Lett. **120**, 061802 (2018).

[6] J. Lindhard, V. Nielsen, M. Scharff, and P. Thomsen, Kgl. Dan. Vidensk. Selsk. Mat. Fys. Medd. **33**, 1 (1963).

[7] E. Armengaud *et al.* (EDELWEISS Collaboration), Phys. Rev. D **99**, 082003 (2019).

[8] R. Agnese *et al.* (SuperCDMS Collaboration), Phys. Rev. D **92**, 072003 (2015).

[9] R. Agnese *et al.* (SuperCDMS Collaboration), Phys. Rev. D **99**, 062001 (2019).

[10] A. Aguilar-Arevalo *et al.* (DAMIC Collaboration), Phys. Rev. Lett. **125**, 241803 (2020).

[11] J. Tiffenberg, M. Sofo-Haro, A. Drlica-Wagner, R. Essig, Y. Guardincerri, S. Holland, T. Volansky, and T.-T. Yu, Phys. Rev. Lett. **119**, 131802 (2017).

[12] R. K. Romani *et al.*, Appl. Phys. Lett. **112,** 043501 (2018).

[13] A. N. Villano, M. Fritts, N. Mast, S. Brown, P. Cushman, K. Harris, and V. Mandic, Phys. Rev. D **105,** 083014 (2022).

[14] B. J. Scholz, A. E. Chavarria, J. I. Collar, P. Privitera, and A. E. Robinson, Phys. Rev. D **94,** 122003 (2016).

[15] A. E. Chavarria *et al.*, Phys. Rev. D **94,** 082007 (2016).

[16] P. Sorensen, Phys. Rev. D **91,** 083509 (2015).

[17] B. L. Dougherty, Phys. Rev. A **45,** 2104 (1992).

[18] G. Gerbier *et al.*, Phys. Rev. D **42,** 3211 (1990).

[19] O. Martineau *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **530,** 426 (2004).

[20] R. Agnese *et al.* (SuperCDMS Collaboration), Phys. Rev. Lett. **112,** 241302 (2014).

[21] A. Bousselham, H. H. Barrett, V. Bora, and K. Shah, Nucl. Instrum. Methods Phys. Res., Sect. A **620,** 359 (2010).

[22] D.-M. Mei *et al.*, J. Phys. G **47,** 105106 (2020).

[23] U. Fano, Phys. Rev. **70,** 44 (1946).

[24] Note that the convenience is in comparison to the electron recoil case and the qualitative similarity to the Poisson distribution, and the underlying causes of the variance are different in the nuclear recoil case, arising from the inherent randomness in the energy deposition process in terms of the fraction of energy given to the lattice. The electron recoils have approximately zero energy given to the lattice.

[25] The measurement temperatures are different for these measurements, being 90 K, 77 K, and 87 K for the lowest-to-highest measurements.

[26] T. Yamaya, R. Asano, H. Endo, and K. Umeda, Nucl. Instrum. Methods **159,** 181 (1979).

[27] J. Eberhardt, Nucl. Instrum. Methods **80,** 291 (1970).

[28] J. Palms, P. Venugopala Rao, and R. Wood, Nucl. Instrum. Methods **76,** 59 (1969).

[29] F. Perotti and C. Fiorini, Nucl. Instrum. Methods Phys. Res., Sect. A **423,** 356 (1999).

[30] H. R. Zulliger and D. W. Aitken, IEEE Trans. Nucl. Sci. **17,** 187 (1970).

[31] J. Lindhard, V. Nielsen, and M. Scharff, *Approximation Method in Classical Scattering by Screened Coulomb Fields: Notes on Atomic Collisions, 1*, Matematisk-fysiske meddelelser (Munksgaard, Copenhagen, Denmark, 1968).

[32] J. Lindhard, M. Scharff, and H. Schioett, Kgl. Dan. Vidensk. Selsk. Mat. Fys. Medd. **33,** 1 (1963).

[33] J. Lindhard and M. Scharff, Phys. Rev. **124,** 128 (1961).

[34] A. Kennedy, SuperCDMS prototype detector design and testing, Ph.D. thesis, University of Minnesota, 2017.

[35] Actually because of the definition of quenching as E_I/E_r the expected width on this quantity needs to be derived more carefully than the standard $\sqrt{(df/dA)^2\sigma_A^2 + (df/dB)^2\sigma_B^2}$.

[36] As noted before the electron recoils also have a Fano factor but its effect was included in the measurement of the EDELWEISS resolution parameters.

[37] Equation (4) has the same form when written in terms of the instrumental FWHM values, except that the constants will have to be multiplied by a factor; due to convention a_H is reported as if this equation is the FWHM version.

[38] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, Publ. Astron. Soc. Pac. **125,** 306 (2013).

[39] A. N. Villano and A. Roberts, Intrinsic Fano factor of nuclear recoils for dark matter searches (2022), 10.17605/OSF.IO/5F9EP.

[40] J. Allison *et al.*, IEEE Trans. Nucl. Sci. **53,** 270 (2006).

[41] S. Agostinelli *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **506,** 250 (2003).

[42] S. Agostinelli *et al.* (Geant4 Collaboration), Geant4 Guide for Physics Lists, https://geant4.web.cern.ch/index.php/.

[43] S. Agostinelli *et al.* (Geant4 Collaboration), *Geant4 Physics Reference Manual, Chapter 39: Low Energy Neutron Interactions*, https://geant4.web.cern.ch/index.php/.

[44] L. Thulliez, C. Jouanne, and E. Dumonteil, Nucl. Instrum. Methods Phys. Res., Sect. A **1027,** 166187 (2022).

[45] Iwamoto, Osamu *et al.*, EPJ Web Conf. **239,** 09002 (2020).

[46] M. F. Albakry *et al.*, Phys. Rev. D **105,** 122002 (2022).

[47] R. Agnese *et al.* (SuperCDMS Collaboration), Phys. Rev. D **95,** 082002 (2017).

[48] R. C. Alig, S. Bloom, and C. W. Struck, Phys. Rev. B **22,** 5565 (1980).

[49] L. Hehn *et al.*, Eur. Phys. J. C **76,** 548 (2016).