Determining the bubble nucleation efficiency of low-energy nuclear recoils in superheated C_3F_8 dark matter detectors

B. Ali,[1] I. J. Arnquist,[2] D. Baxter,[3] E. Behnke,[4] M. Bressler,[5] B. Broerman,[6] K. Clark,[6] J. I. Collar,[7] P. S. Cooper,[3] C. Cripe,[4] M. Crisler,[3,2] C. E. Dahl,[8,3] M. Das,[9] D. Durnford,[10,*] S. Fallows,[10] J. Farine,[11,12,13] R. Filgas,[1] A. García-Viltres,[14] F. Girard,[15,16] G. Giroux,[6] O. Harris,[17] E. W. Hoppe,[2] C. M. Jackson,[2] M. Jin,[18,†] C. B. Krauss,[10] V. Kumar,[9] M. Lafreniere,[16] M. Laurin,[16] I. Lawson,[15,12] A. Leblanc,[15] H. Leng,[19] I. Levine,[4] C. Licciardi,[11,12,13] S. Linden,[12] P. Mitra,[10] V. Monette,[16] C. Moore,[6] R. Neilson,[5] A. J. Noble,[6] H. Nozard,[16] S. Pal,[10] M.-C. Piro,[10,‡] A. Plante,[16] S. Priya,[19] C. Rethmeier,[10] A. E. Robinson,[16] J. Savoie,[16] O. Scallon,[16] A. Sonnenschein,[3] N. Starinski,[16] I. Štekl,[1] D. Tiwari,[16] F. Tardif,[16] E. Vázquez-Jáuregui,[14] U. Wichoski,[11,12,13] V. Zacek,[16] and J. Zhang[18,§]

(PICO Collaboration)

[1]*Institute of Experimental and Applied Physics, Czech Technical University in Prague, Prague, Cz-12800, Czech Republic*
[2]*Pacific Northwest National Laboratory, Richland, Washington 99354, USA*
[3]*Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA*
[4]*Department of Physics, Indiana University South Bend, South Bend, Indiana 46634, USA*
[5]*Department of Physics, Drexel University, Philadelphia, Pennsylvania 19104, USA*
[6]*Department of Physics, Queen's University, Kingston, K7L 3N6, Canada*
[7]*Enrico Fermi Institute, KICP and Department of Physics, University of Chicago, Chicago, Illinois 60637, USA*
[8]*Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208, USA*
[9]*High Energy Nuclear & Particle Physics Division, Saha Institute of Nuclear Physics, Kolkata, India*
[10]*Department of Physics, University of Alberta, Edmonton, T6G 2E1, Canada*
[11]*School of Biological, Chemical, and Forensic Sciences, Laurentian University, Sudbury, ON P3E 2C6, Canada*
[12]*SNOLAB, Lively, Ontario, P3Y 1N2, Canada*
[13]*Department of Physics, Carleton University, Ottawa, Ontario, K1S 5B6, Canada*
[14]*Instituto de Física, Universidad Nacional Autónoma de México, A.P. 20-364, Ciudad de México 01000, México*
[15]*Department of Physics, Laurentian University, Sudbury, P3E 2C6, Canada*
[16]*Département de Physique, Université de Montréal, Montréal, H3C 3J7, Canada*
[17]*Northeastern Illinois University, Chicago, Illinois 60625, USA*
[18]*Department of Physics and Astronomy, Northwestern University, Evanston, Illinois 60208, USA*
[19]*Materials Research Institute, Penn State, University Park, Pennsylvania 16802, USA*

The bubble nucleation efficiency of low-energy nuclear recoils in superheated liquids plays a crucial role in interpreting results from direct searches for weakly interacting massive particle (WIMP) dark matter. The PICO Collaboration presents the results of the efficiencies for bubble nucleation from carbon and fluorine recoils in superheated C_3F_8 from calibration data taken with 5 distinct neutron spectra at various thermodynamic thresholds ranging from 2.1 keV to 3.9 keV. Instead of assuming any particular functional forms for the nuclear recoil efficiency, a generalized piecewise linear model is proposed with systematic errors included as nuisance parameters to minimize model-introduced uncertainties. A Markov-Chain Monte-Carlo (MCMC) routine is applied to sample the nuclear recoil efficiency for fluorine and carbon at 2.45 keV and 3.29 keV thermodynamic thresholds simultaneously. The nucleation efficiency for fluorine was found to be $\geq 50\,\%$ for nuclear recoils of 3.3 keV (3.7 keV) at a thermodynamic Seitz threshold of 2.45 keV (3.29 keV), and for carbon the efficiency was found to be $\geq 50\,\%$ for recoils of 10.6 keV (11.1 keV) at a threshold of 2.45 keV (3.29 keV). Simulated data sets are used to calculate a p-value for the fit, confirming that the model used is compatible with the data. The fit paradigm is also assessed for potential systematic biases, which although small, are corrected for. Additional steps are performed to calculate the expected interaction rates of WIMPs in the PICO-60 detector, a requirement for calculating WIMP exclusion limits.

I. INTRODUCTION

Superheated liquids are excellent targets for direct dark matter detection experiments searching for heavy dark matter particles such as WIMPs (Weakly Interacting Massive Particles) scattering off atomic nuclei. The nuclear recoil (NR) that is predicted to result from such a scatter creates a single bubble at the interaction site in the superheated target. World-leading limits have been set on dark matter-nucleus scattering rates based on the observed absence of such bubble nucleation [1–6]. In the interpretation of these (and future) experimental results, the bubble nucleation efficiency for NRs as a function of both the thermodynamic state of the chamber and the nuclear recoil energy, is crucial to determining dark matter sensitivity. To characterize this detector response, neutrons are used as proxies for dark matter particles, generating the same NR-induced bubbles as expected from dark matter, but with few-cm mean free paths and recoil energy spectra that depend

* ddurnfor@ualberta.ca
† now at Shenzhen Artificial Intelligence and Data Science Institute
‡ mariecci@ualberta.ca
§ now at Argonne National Laboratory

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

on the incident neutron energy. This paper presents results from the PICO Collaboration's campaign over nearly a decade using a variety of neutron sources to calibrate the low-energy (keV-scale) nuclear recoil sensitivity of superheated C_3F_8.

A. Bubble nucleation by nuclear recoils

The thermodynamic basis for bubble nucleation by nuclear recoils is described by Seitz's "hot-spike" model [7], a detailed modern treatment of which is given in [8], summarized here. This model is based on two well-defined thermodynamic quantities, the critical radius r_c and the energy threshold Q_{Seitz}. r_c is the radius above which vapor bubbles in the superheated fluid will spontaneously grow, eventually becoming the macroscopic bubbles observed by experiments such as PICO. The critical radius is given by

$$r_c = \frac{2\sigma}{P_b - P_l}, \tag{1}$$

where σ is the surface tension of the fluid, P_b is the pressure of the vapor filling the bubble, which is approximately equal to the saturation pressure of the fluid at the operating temperature T (controlled experimentally), and P_l is the pressure of the superheated liquid in the bubble chamber (also controlled experimentally). In Seitz's model, nuclear recoils create this critical proto-bubble by locally heating the fluid, and the amount of heat required to create a critically-sized proto-bubble is referred to as the Seitz threshold, given by

$$\begin{aligned} Q_{\text{Seitz}} \approx 4\pi r_c^2 \left(\sigma - T\frac{\partial\sigma}{\partial T}\right) + \frac{4\pi}{3} r_c^3 \rho_b (h_b - h_l) \\ - \frac{4\pi}{3} r_c^3 (P_b - P_l). \end{aligned} \tag{2}$$

Here, ρ_b is the density of the vapor filling the bubble, and h_b and h_l are the specific enthalpies of the gaseous and liquid states. The three terms in Q_{Seitz} represent the energy needed to create the bubble surface, the energy required to vaporize fluid to fill the bubble interior, and the recapture of reversible work present in both of the first two terms. Any particle interaction injecting heat greater than Q_{Seitz} ($\sim$3 keV in the experiments considered here) into a volume small compared to r_c ($\sim$25 nm) will create a critically-sized proto-bubble, which will then grow to the macroscopic bubble detected in these experiments.

As a consequence of this idealized model, the efficiency for bubble nucleation would be a step function from 0% to 100% when the energy deposited within a critical length scale exceeds the Seitz threshold:

$$E_{\text{dep}} = \int_0^{\lambda r_c} \frac{dE}{dx} dx \geq Q_{\text{Seitz}}, \tag{3}$$

where λ is a unitless scale factor of $O(1)$, often referred to as the "Harper" parameter [9]. For low-energy nuclear recoils, this would simplify further to $E_{\text{dep}} \approx E_r$ (all recoil energy deposited within the critical scale), and a calculation of Q_{Seitz} from the temperature and pressure of the superheated target would be sufficient to determine the nuclear recoil detection threshold. Unfortunately, this is not true for at least three reasons. First, Eq. (2) does not include corrections to surface tension at a small radius of curvature, an effect described by the Tolman length [10] and covered in detail in [8]. The Tolman length itself is unknown and leads to $O(0.1)$-keV uncertainties in Q_{Seitz}. Second, Q_{Seitz} does not account for energy losses that do not contribute to local heating, such as radiative losses (e.g. fluorescence), and irreversible work (e.g. acoustic radiation), or thermal diffusion transporting heat outside critical radius. Finally, and most importantly, Eq. (3) does not reflect event-to-event variation in track structure (i.e. straggling in $\frac{dE}{dx}$) which simulations using SRIM [11] show to be significant in nuclear recoils at the energies and length scales considered here.

The effects above can both shift the nuclear recoil detection threshold and, in the case of straggling, broaden it, leading to detection efficiencies below 100% near the threshold. Indeed, past data with measured mono-energetic recoils [12–14] demonstrate that the nucleation threshold is not an ideal step function. Since there is no quantitative theory to describe these effects, past efforts have invoked ad-hoc parameterizations for the threshold function, such as an exponential [1, 15, 16], a sigmoid function [14] and a "superheated factor" [17, 18]. Efforts remain to understand and explain the processes that contribute to the resolution function that shifts and convolutes the Seitz step-threshold [14]. Until these effects are fully understood, nucleation efficiencies must be determined by performing dedicated neutron calibrations, as described in this work.

The following sections comprise a global analysis of the PICO Collaboration's neutron calibrations to date in superheated C_3F_8, the target fluid for the PICO-2L, PICO-60, PICO-40L, and PICO-500 dark matter searches. The general calibration scheme, the specific experimental setups used, and the resulting data and corresponding simulations are described in Section II. Section III describes the methods used to extract bubble nucleation efficiencies from this data, including the parameterization of the efficiency function, treatment of systematic uncertainties as nuisance parameters, and the specific Markov Chain Monte Carlo (MCMC) technique used to explore the resulting high-dimensional parameter space. Section IV describes a parametric Monte Carlo study validating the methodology created for this analysis. Appendices describe the application of this technique to directly constrain WIMP sensitivity in the PICO experiments and present an evaluation of and correction for any bias introduced in this analysis.

II. CALIBRATION PROGRAM — DATA AND SIMULATIONS

The objective of the PICO C_3F_8 nuclear recoil calibration program is to constrain the bubble nucleation efficiency functions $\varepsilon_s(E_T, E_r)$, where ε is the probability of bubble nucleation, s indicates the recoil species (carbon or fluorine), E_T is the thermodynamic threshold set by the pressure and tem-

perature of the chamber (equal to $Q_{\rm Seitz}$ as calculated in [8]), and E_r is the nuclear recoil energy. Specifically, we aim to constrain the E_r dependence of these functions at the fixed various E_T employed in the PICO dark matter searches. Constraints on $\varepsilon_s(E_T, E_r)$'s come from rate measurements in bubble chambers exposed to known neutron sources. In general, these constraints take the form of a convolution of the underlying recoil spectrum with the efficiency curve:

$$R_{\rm obs} = \sum_{s={\rm C,F}} \int_0^\infty R_s(E_r) \cdot \varepsilon_s(E_T, E_r)\, dE_r, \qquad (4)$$

where $R_{\rm obs}$ is the experimentally observed rate and the $R_s(E_r)$ are the nuclear recoil spectra (by species) due to the neutron source, determined from simulation.

A single constraint in Eq. (4) is insufficient to deduce the underlying $\varepsilon_{\rm C}$ and $\varepsilon_{\rm F}$, as the convolution cannot indicate which recoils are nucleating bubbles — i.e. threshold shift is indistinguishable from a soft threshold. A suite of measurements with sources producing different recoil spectra, however, is capable of making this distinction. The ability to reconstruct multiple-scattering events in calibration experiments is also key. Each experiment produces not a single rate measurement but a set of measurements giving the 1-bubble event rate, 2-bubble event rate, etc., with each multiplicity effectively probing a different underlying recoil spectrum. In this sense, Eq. (4) is a simplification applying only to the single-bubble rate in a chamber much smaller than the neutron mean free path, which is not the case in the measurements presented here. Finally, the E_T dependence of the efficiency curves can be explored by varying the thermodynamic state of a calibration chamber. However, unless assumptions are made relating to the E_T and E_r dependence of the efficiency function, this does little to constrain E_r dependence at the E_T of interest.

A. Neutron interactions in C_3F_8

Neutrons interact predominantly by elastic scattering off of ^{12}C and ^{19}F nuclei (with $< 1\%$ ^{13}C) in the C_3F_8 target. The energy spectrum of the recoiling nuclei is described by:

$$E_r = \frac{2A}{[A+1]^2}(1-\cos\theta)E_n \qquad (5)$$

where A is the atomic mass of the recoiling nucleus, θ is the neutron scattering angle in the center of mass frame, and E_n is the incident neutron energy. By fixing $\theta = \pi$, the maximum recoil energies are $E_{r,C_{max}} = 0.28E_n$ and $E_{r,F_{max}} = 0.19E_n$ for carbon and fluorine respectively. When the incoming neutrons are mono-energetic and the scattering is isotropic, this produces the recoil spectra and rate-over-threshold shown in Fig. 1.

Actual neutron-carbon and neutron-fluorine cross-sections are shown in Fig. 2, illustrating three resonances in the neutron-fluorine scattering cross-section in the energy scale of interest. By calibrating with mono-energetic neutrons at energies on and off these resonances, it is possible to vary

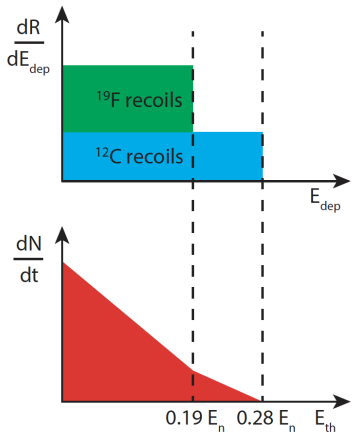

FIG. 1. Top: Cartoon nuclear recoil energy spectra in the case of isotropic (s-wave) elastic scattering. Bottom: Corresponding count rate measured by a threshold detector, as a function of threshold.

the relative rates of carbon and fluorine scattering, allowing separate calibration of $\varepsilon_{\rm F}$ and $\varepsilon_{\rm C}$. It should be noted that on-resonance scattering is *not* isotropic, so at these energies, the fluorine recoil spectra deviate from the idealized box-shaped spectrum in Fig. 1 [19].

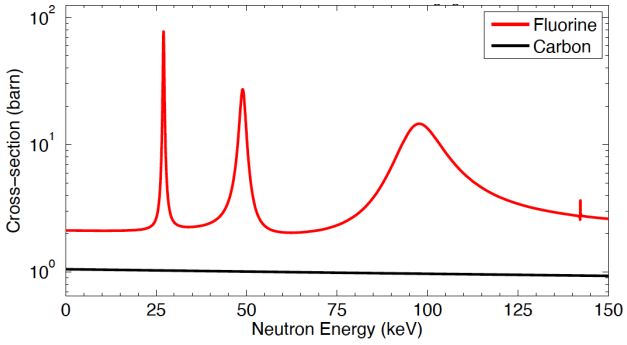

FIG. 2. Neutron-carbon and neutron-fluorine elastic scattering cross sections as a function of neutron energy, taken from [20], for the neutron energy range used during detector calibration.

B. Experimental setup

1. *Detectors*

Two detectors have been employed in PICO C_3F_8 nuclear recoil calibration program: the PICO-2L detector at SNOLAB, described in [2], and a portable version, the PICO-0.1L test bubble chamber, described in [8]. Designed to accomplish calibrations that are difficult in larger chambers, PICO-0.1L consists of a 100-mL, centimeter-thick high-pressure-rated quartz jar attached to a hydraulically-driven bellows and immersed in a chilled water bath. The quartz vessel is filled with 27 $\pm$ 1 gram of C_3F_8 (19.4 mL of fluid at a density of

1.39 g/mL at 12°C and 30 psia), as measured by a scale, where the uncertainty is due to losses in the fill lines and the resolution of the scale readout. The remaining space in the vessel is filled with a water buffer, both to create an incompressible volume for pressure control and to avoid any contact between the superheated C_3F_8 target and the stainless steel surfaces and seals in the bellows system. Bubble nucleation events are recorded by two piezoelectric acoustic sensors mounted on the jar and by two cameras (150 frames/s) located at 90 degrees from each other. Two LED panels alternate to illuminate the chamber, each in synchronization with one of the cameras. The PICO-0.1L vessel is shown in Fig. 3, both alone and with the attached imaging and pressure- and temperature-control systems.

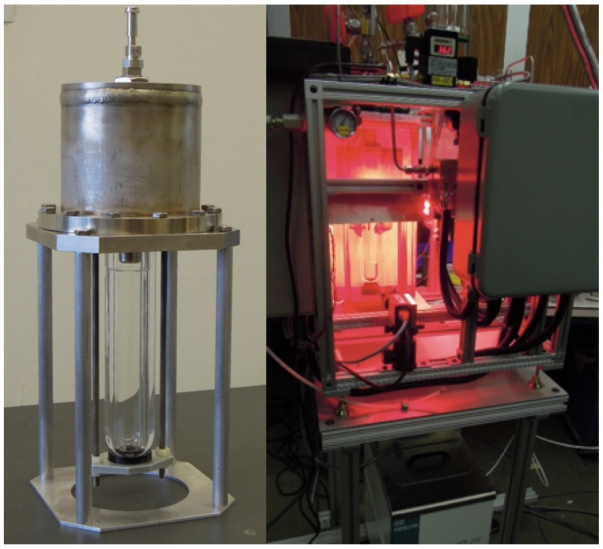

FIG. 3. PICO-0.1L detector. Left: The quartz vessel mounted to its hydraulic "top hat" containing the bellows used for pressure control of the chamber. Right: The complete setup of the chamber inside its thermal bath equipped with two piezoelectric sensors, two cameras and two LED panels, and CompactRIO-based pressure- and temperature-control system.

2. *Neutron sources and neutron monitor*

PICO-0.1L was deployed from 2013 to 2019 at the Van de Graaff Tandem accelerator facility at Université de Montréal, where data were taken both with mono-energetic neutron beams produced by the accelerator and with a SbBe photoneutron source. PICO-2L operated at SNOLAB, with periodic neutron calibrations performed using an AmBe source during the dark matter search.

a. Mono-energetic neutron beam Monoenergetic neutrons were produced with the Tandem Van de Graaff accelerator via the reaction ^{51}V(p,n)^{51}Cr [21, 22]. The energies of the neutrons produced by this reaction are the proton energy minus 1564 keV, the Q-value of the reaction so that a mono-energetic proton beam and thin target produce a mono-energetic neutron beam. The energy resolution of the neutron beam is significantly enhanced when the proton energy is at one of the many sharp resonances shown in Fig. 4. The data considered here were taken at resonances VII, VIII, and XI, corresponding to neutron energies of 50, 61, and 97 keV, respectively. These three ^{51}V(p,n)^{51}Cr resonances were chosen to align on and between the neutron scattering resonances shown in Fig. 2, with the 50- and 97-keV neutrons on-resonance and the 61-keV neutrons off-resonance for scattering on fluorine.

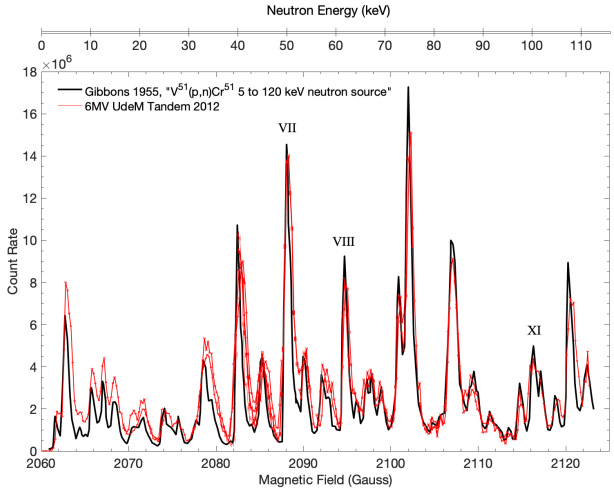

FIG. 4. Neutron yield of the ^{51}V(p,n)^{51}Cr reaction as a function of the magnetic field and corresponding neutron energy. The black curve is taken from [21], and is consistent measurements done with the Tandem at the Université de Montréal [22] in red. The resonances VII, VIII, and XI chosen for the analysis are also indicated.

Figure 5 illustrates the experimental setup for neutron beam data. The PICO-0.1L test chamber sits directly downstream of the ^{51}V target, with a penetration through the chilled water bath giving neutrons a direct path to the quartz vessel. The details of this penetration changed between the 2013 and 2014 datasets, as reflected in the analysis in Section III. Two ^{3}He neutron counters monitor the beam flux, one immediately below the ^{51}V target and one below the PICO-0.1L water bath. These counters provide the neutron flux normalization described in Sec. II D and provide live feedback to ensure that the proton beam stays at the peak of the selected ^{51}V(p,n)^{51}Cr resonance during data taking.

b. Mono-energetic SbBe photoneutrons To study the detector response at lower recoil energies with the PICO-0.1L test chamber, additional calibration runs were performed using a SbBe photoneutron source, a two-component source that contains a ^{124}Sb gamma source and a ^{9}Be conversion target, producing neutrons via the ^{9}Be(γ,n) reaction [23, 24]. There are two significant gamma rays emitted by ^{124}Sb above the 1665-keV Q-value of this reaction: 1691 keV and 2091 keV, with branching ratios of 48.4% and 5.7%, producing 24- and 378-keV neutrons, respectively. A challenge with this calibration is that the source emits $O(10^6)$ higher gamma flux than neutron flux. While PICO bubble chambers are in general gamma-blind [8], the detector does start to become

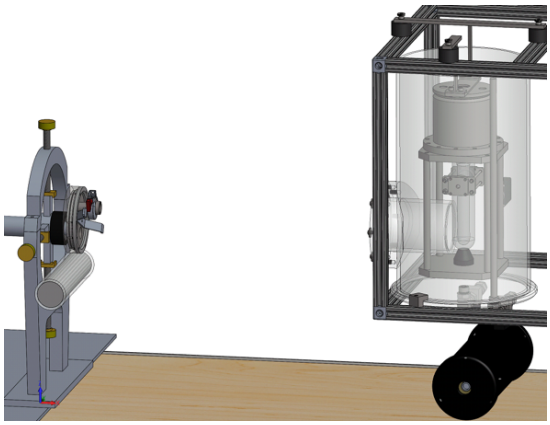

FIG. 5. Experimental setup of the test beam calibration at the Université de Montréal. On the left a ^{3}He counter (horizontal cylinder) is placed underneath the vanadium target located at the end of the proton beam. At the right, the test chamber is facing the beamline with another ^{3}He counter located underneath as a second neutron flux monitor.

gamma sensitive at Seitz thresholds below 3 keV, increasing the single-bubble event rate. This gamma response is measured by removing the beryllium disk and is considered in the analysis as a background to the SbBe response of the detector. Throughout the measurements, two 1-inch thick lead disks were inserted in front of the SbBe source to attenuate the gamma flux, but the gamma-induced single-bubble rate was still deemed too high for useful analysis, and therefore only multi-bubble events are considered from the SbBe exposure. The setup for the calibration with the SbBe source (without lead attenuators) is shown in Fig. 6.

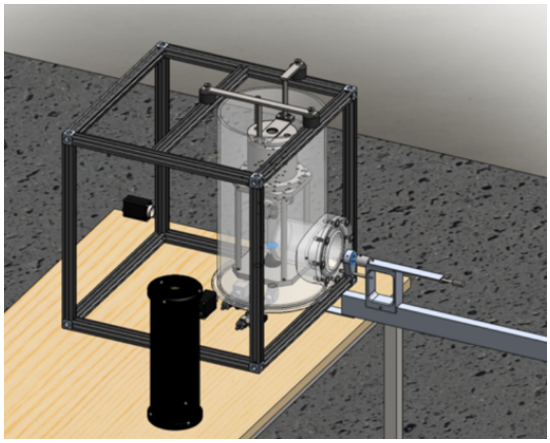

FIG. 6. Experimental setup of the SbBe calibration at the Université de Montréal. The PICO-0.1L test chamber is aligned with a graduated source rail to measure the SbBe source position. A ^{3}He counter is placed beside the bubble chamber to verify the neutron flux at the chamber.

c. AmBe neutrons Neutron calibration with an AmBe source was carried out at SNOLAB with the PICO-2L detector filled with C_3F_8. In contrast to the mono-energetic neutron sources above, this broad-spectrum source of $O(1)$-MeV neutrons generates nuclear recoils far above the detector threshold. The threshold-constraining power of this measurement comes from the high-statistics, low-background measurement of high-multiplicity events (up to 6 bubbles and beyond) captured in the 2-liter target. Moreover, the different detector, source, and style of constraint provided by this dataset serve as an important check on the systematic uncertainties associated with each of the neutron calibrations performed.

C. Simulations of neutron scattering

For each neutron calibration data set, a Geant4 [25] (SbBe and AmBe data) and/or MCNP-Polimi [26] (beam and AmBe data) Monte Carlo simulation is performed to provide the recoil energy spectra for single- and multiple-scatter events within the detector. These are used, in conjunction with hypothesized bubble nucleation efficiency functions, to calculate the expected event rate by bubble multiplicity for each calibration experiment. Neutron scattering cross-sections for both transport through the geometry and signal production in the detector are taken from Evaluated Neutron Data File ENDF/B-VII [20] for both MCNP and Geant4 [27], with corrections from [19] for the differential scattering cross sections at the fluorine resonances, which ENDF incorrectly treats as s-wave (isotropic).

Each of the three neutron sources considered requires a slightly different treatment for particle generation. For the beam simulations, the energy-angle relation for neutrons coming from the ^{51}V(p,n)^{51}Cr reaction is included directly in the MCNP-Polimi neutron source definition, based on data in [21]. A similar energy-angle relation exists for the SbBe photoneutron source, but because the Sb gammas are emitted isotropically into a large Be target, a simple neutron source definition capturing this relation is not possible. Furthermore, this relation cannot be simulated natively in MCNP-Polimi — while MCNP-Polimi does simulate the ^{9}Be(γ,n) reaction, it does so by treating the incoming photon as a zero-momentum particle, and therefore misses the angular dependence of the resulting neutron energy. To address this shortcoming, a photon-only simulation of the Sb source and Be target was run and neutron production was inserted manually to generate a list of neutrons with a given initial position, direction, and energy. This list was then used as the input for a neutron simulation in Geant4. The AmBe exposure in PICO-2L was simulated with both MCNP-Polimi and Geant4 to test different source spectra included in the respective simulation packages. No significant discrepancies in recoil spectra were found, and the MCNP-Polimi simulation was selected for this analysis. Examples of the resulting simulated recoil energy spectra are shown in Fig. 7, which shows the general dominance of fluorine (carbon) scatters at low (high) recoil energies, as was depicted in Fig. 1.

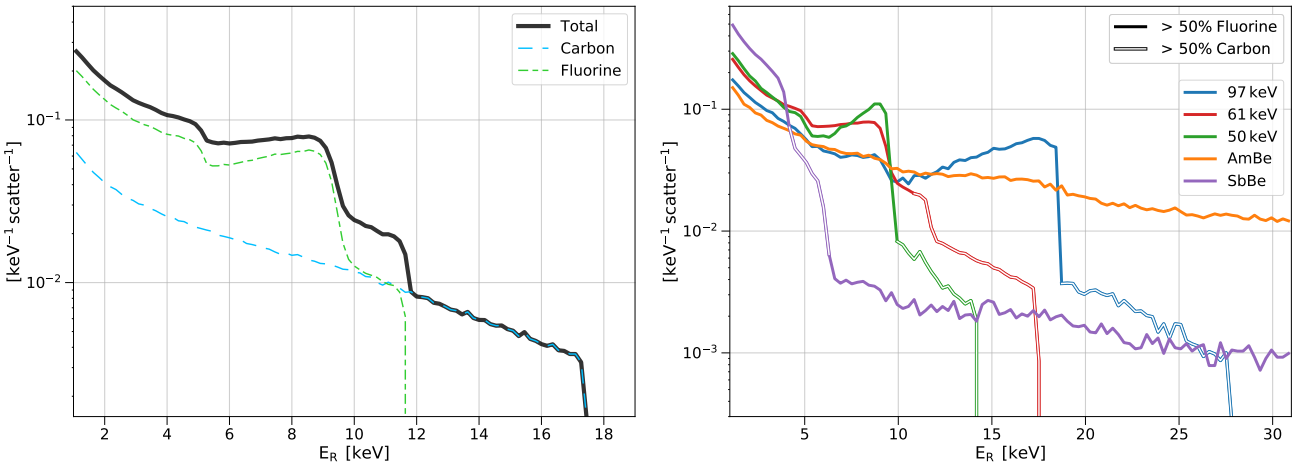

FIG. 7. Left: simulated recoil energy spectrum for the 61 keV beam experiment (2013) showing the separate contributions of scatters off of fluorine and carbon. Right: simulated recoil energy spectra for the 97 keV (2013), 61 keV (2013), and 50 keV (2014) beam experiments, as well as the SbBe and AmBe experiments. The dominance of scattering off of fluorine or carbon atoms is indicated by solid or hollow lines respectively. The spectra in the left and right panels are normalized to the total number of scatters in each simulation off of either target species.

D. Neutron flux measurements

Constraints from both beam and SbBe data in PICO-0.1L are more powerful when the neutron flux (as well as energy) is well known. This is not true for the PICO-2L AmBe data, where constraints derive largely from ratios of rates at different multiplicities. This section describes the ancillary measurements used to anchor the neutron flux from these sources.

1. *Neutron beam flux*

The relative flux from the neutron beam is continuously monitored by the ^{3}He counter sitting immediately below the ^{51}V target. Translating this to an absolute neutron flux at the PICO-0.1L chamber is challenging in simulation, requiring precise knowledge of the materials and geometry surrounding the ^{51}V target and ^{3}He counter. It is also difficult to verify via in-situ measurement — the second ^{3}He counter near the chamber is insufficient for this purpose, due primarily to uncertainties in the composition and geometry of materials near that counter, including the PICO-0.1L water bath and support structure. For this reason, two independent measurements were performed to verify the absolute neutron beam flux.

a. ^{3}He-only measurement The first flux measurement was performed with the two ^{3}He counters only, with the PICO-0.1L chamber and all unnecessary structures removed. The target-side ^{3}He counter remained in its usual position below the ^{51}V target, and the chamber-side counter was suspended downstream of the target as shown in Fig. 8. All geometry details, including the beam pipe, were entered into an MCNP [26] Monte Carlo to simulate the neutron capture rates in the two counters. The comparison between the measured neutron capture ratio of both counters with Monte Carlo simulations at each beam energy is shown in Table I. At all three energies, the agreement is reasonable, and while there is a trend to slightly over-predict the near-target to downstream ratio, the source of this bias is unclear and no correction is attempted. Instead, these uncertainties are included as nuisance parameters in Section III B.

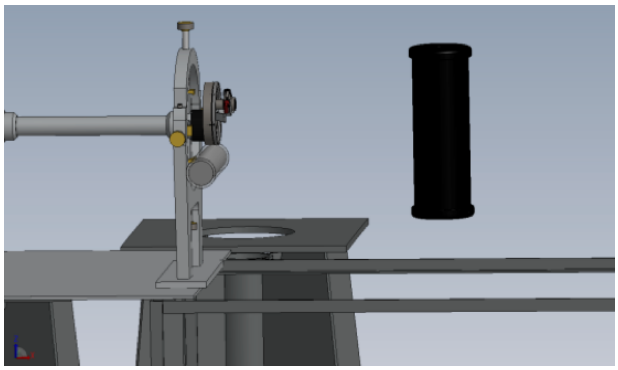

FIG. 8. Experimental setup at the Université de Montréal for measuring the neutron flux. Two ^{3}He counters were used: one directly below the ^{51}V target and one suspended directly downstream of the target.

b. Target activation (^{51}Cr) measurement The second flux calibration technique is to measure the ^{51}Cr activity in the ^{51}V target before and after exposure to the proton beam, directly measuring the number of neutrons produced via the ^{51}V(p,n)^{51}Cr reaction. The 320-keV gammas produced by the ^{51}Cr electron-capture decay are measured with a Ge-detector, calibrated using a ^{133}Ba source (303- and 356-keV gammas) with precisely known activity (+/-3%) and geometry similar to the ^{51}V target disc. An exponential decay fit on the gamma activity was performed with a fixed 27.7-day half-life to determine the total quantity of ^{51}Cr produced (see Fig. 9). Table II presents the results obtained at 50-keV neutron energy (the only beam energy where this measurement was performed).

Energy	Measured Ratio	Simulated Ratio	Meas / Sim
50 keV	2.28 ± 0.07	2.26 ± 0.08	1.01 ± 0.05
61 keV	2.02 ± 0.07	2.26 ± 0.08	0.89 ± 0.04
97 keV	2.07 ± 0.10	2.21 ± 0.07	0.93 ± 0.05

TABLE I. Measured neutron capture rate ratio of the two ^{3}He counters in the setup shown in Fig. 8 (near-target rate / downstream rate), compared to Monte Carlo simulation. Uncertainties in the measurement are statistical uncertainties from the number of captures recorded in the two ^{3}He counters. Uncertainties on the simulated ratio are highly correlated between the different energies and include uncertainties on the ^{3}He content of the counters and the density/makeup of the neutron moderator surrounding the counters.

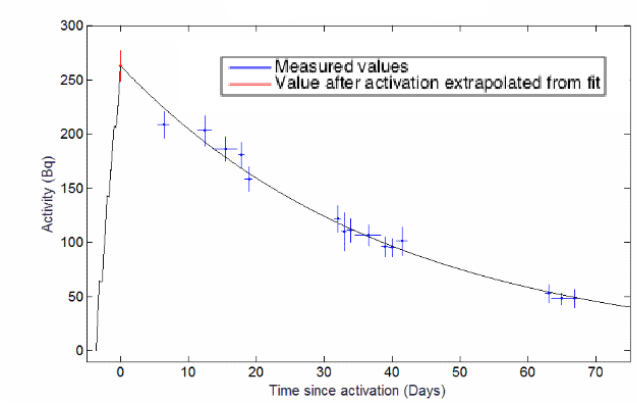

FIG. 9. ^{51}Cr activity in the ^{51}V target after exposure to the proton beam. These data were taken at a proton beam energy corresponding to 50-keV neutron production, with a fresh ^{51}V target (no ^{51}Cr activity prior to exposure). A single-parameter fit with fixed ^{51}Cr half-life gives the activation rate of the target when exposed to the beam.

This check is not directly used elsewhere in the analysis, but the agreement seen in Table II validates the use of the target-side ^{3}He counter and corresponding MCNP simulation to fix the absolute neutron beam flux.

Calculation method	Neutrons produced ($\times 10^8$)
^{3}He plus MCNP simulation	9.18±0.52
^{51}Cr activity	9.52±0.51
Ratio	1.04±0.08

TABLE II. Total neutrons produced during a sample 50-keV neutron beam run, measured via the ^{3}He counter (with MCNP simulation to convert the ^{3}He capture rate to neutron production rate) and via ^{51}Cr activity in the ^{51}V target.

2. *SbBe neutron flux*

Rather than base the neutron flux from the SbBe source on the underlying ^{124}Sb activity and simulated ^{9}Be(γ,n) rate, both of which contain significant uncertainty, the neutron flux from the SbBe source was measured directly, using the same ^{3}He counter that was used as the downstream counter for the beam neutron flux normalization. Multiple measurements with the SbBe source and ^{3}He counter in various relative positions and orientations give a total neutron yield of 209 ± 22 neutrons per second (corrected for the ^{124}Sb half-life in individual datasets), where the uncertainty reflects the measurement-to-measurement variation in the SbBe neutron yield.

E. Data quality and preparation

Each calibration dataset is reduced to a small set of parameters for subsequent analysis. These parameters include the thermodynamic threshold of the bubble chamber ($Q_{\rm Seitz}$), total livetime at that threshold, total source exposure at that threshold (i.e. integrated beam intensity during the livetime), number of events observed (separated by bubble multiplicity), and background bubble rate (also separated by bubble multiplicity). Background-subtracted event rates as a function of threshold are shown in Fig. 10.

The above data reduction is performed largely by hand; camera images are visually inspected (hand-scanned) to verify bubble counts, a process that is taken to be 100% efficient. A series of quality cuts are also applied to eliminate regions in time and/or regions in the detector where chamber performance is compromised. These cuts include the removal of runs where pressure- or temperature-control was not functioning as intended; the removal of the first 10 (30) seconds in every PICO-0.1L (PICO-2L) expansion, to allow time for the chamber to reach equilibrium; a fiducial cut in PICO-2L as described in [2], applied to single-bubble events only; and a fiducial cut in 2013 PICO-0.1L data, removing both single- and mutliple-bubble events where a bubble appears in the top ∼cm of the target (which in that run became excessively foamy). The fiducial cuts in PICO-2L and PICO-0.1L are applied to simulated data as well as to real data.

III. CONSTRAINTS ON BUBBLE NUCLEATION EFFICIENCY FOR NUCLEAR RECOILS

The objective of this analysis is to constrain $\varepsilon_{\rm C}(E_T, E_r)$ and $\varepsilon_{\rm F}(E_T, E_r)$ at E_T of 2.45 keV and 3.29 keV, two operating conditions of the PICO-60 dark matter detector [5]. This is accomplished through a maximum likelihood analysis, treating each rate measurement as an independent Poisson variable. We express this likelihood as

$$\log\mathscr{L} = \sum_i \sum_j \left[-\nu_{i,j}(\{x_{s,p,T}\},\{s_k\}) + k_{i,j} + k_{i,j}\log\left(\frac{\nu_{i,j}(\{x_{s,p,T}\},\{s_k\})}{k_{i,j}}\right)\right] - \sum_k \frac{s_k^2}{2}, \quad (6)$$

where $\nu_{i,j}$ are the expected numbers of events (including background rate) with bubble multiplicity j for experiment i, and $k_{i,j}$ are the corresponding observed number of events. One "experiment" refers to a combination of detector, threshold,

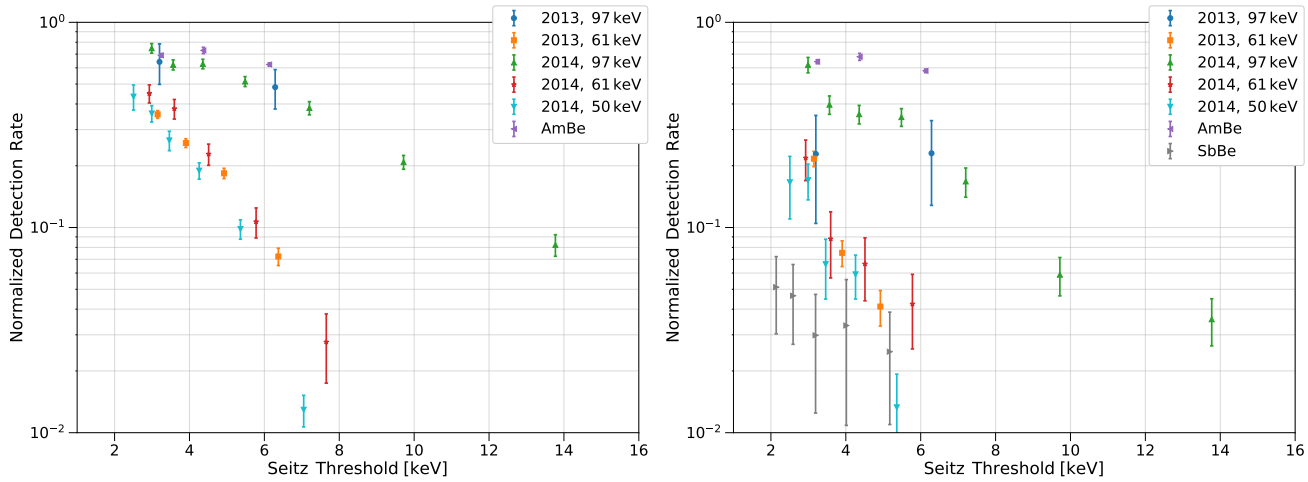

FIG. 10. Background-subtracted measured rates of events with one or more bubble (left) and with two or more bubbles (right), normalized to the rates predicted by simulation when assuming a step-function bubble nucleation threshold at 1 keV. Error bars show statistical uncertainty only. Data at Seitz thresholds greater than 3.6 keV are not included in the full analysis, but the trends with Seitz threshold illustrate the relative sensitivity of each dataset to the underlying nuclear recoil detection threshold.

Dataset	Detector	Thresholds (keV)	Multiplicity
97 keV Beam	PICO-0.1-2013	3.2	1,2,3+
61 keV Beam	PICO-0.1-2013	3.1	1,2,3+
97 keV Beam	PICO-0.1-2014	3.0, 3.6	1,2,3+
61 keV Beam	PICO-0.1-2014	2.9, 3.6	1,2,3+
50 keV Beam	PICO-0.1-2014	2.5, 3.5	1,2,3+
SbBe	PICO-0.1	2.1, 2.6, 3.2	2,3+
AmBe	PICO-2L	3.2	1,2,3,4,5,6,7+

TABLE III. Datasets contributing to the maximum likelihood analysis defined in Eq. (6). Each row corresponds to a single combination of detector and neutron source, with corresponding simulated nuclear recoil spectra. Different bubble multiplicities are considered for each setup, with the '+' indicating that the final multiplicity bin includes events with higher bubble multiplicity as well.

and neutron source. Table III lists the 12 experiments included in this analysis and the bubble multiplicities considered in each. The $\nu_{i,j}$ depend both on the hypothesized efficiency curves, parameterized by $\{x_{s,p,T}\}$, and on a set of nuisance parameters $\{s_k\}$, each of which represents the number of standard deviations that a given source of systematic uncertainty deviates from its nominal value. The total likelihood is frequently cast as an effective chi-square statistic, defined as

$$\chi^2 = -2 \times \log \mathscr{L}. \tag{7}$$

Section III A describes the adopted parameterization of the efficiency curves, defining the $x_{s,p,T}$'s. Section III B defines nuisance parameters considered and the magnitudes of the corresponding systematic uncertainties. Section III C describes the Markov Chain Monte Carlo (MCMC) technique for exploring the resulting 34-dimensional parameter space, and the results of that exploration are given in Section III D.

A. Parameterization of the efficiency functions

To avoid artificially constraining the shape of the bubble nucleation efficiency curve, the curve is modeled as a piecewise-linear function, as shown in Fig. 11. Each "knot" in the function is held at fixed nucleation probability but allowed to translate in recoil energy. By increasing the number of knots in the function, any shape efficiency curve can be approximated (at the cost of a higher-dimensional parameter space to explore). The $\{x_{s,p,T}\}$ are the knot locations on the recoil energy axis for a set of these piecewise-linear functions, where the index s indicates recoil species ($s \in \{\mathrm{C}, \mathrm{F}\}$), the index p indicates nucleation probability ($p \in \{0, 0.2, 0.5, 0.8, 1\}$), and the index T indicates the Seitz threshold ($T \in \{2.45\ \mathrm{keV}, 3.29\ \mathrm{keV}\}$), giving a total of 20 free parameters setting the efficiency curves in this analysis.

Several physics-driven constraints are imposed on the $\{x_{s,p,T}\}$. First, the efficiency curves are taken to be monotonic in both recoil energy and threshold:

$$\frac{\partial \varepsilon_s}{\partial E_r} \geq 0 \quad \rightarrow \quad x_{s,p_{i+1},T} \geq x_{s,p_i,T}, \tag{8}$$

$$\frac{\partial \varepsilon_s}{\partial E_T} \leq 0 \quad \rightarrow \quad x_{s,p,T_{i+1}} \geq x_{s,p,T_i}. \tag{9}$$

Second, the bubble nucleation efficiency for carbon recoils is assumed to be lower than the bubble nucleation efficiency for fluorine recoils at the same energy:

$$\varepsilon_{\mathrm{C}}(E_T, E_r) \leq \varepsilon_{\mathrm{F}}(E_T, E_r) \quad \rightarrow \quad x_{\mathrm{C},p,T} \geq x_{\mathrm{F},p,T}. \tag{10}$$

This constraint is based on nuclear recoil stopping models and on simulations using SRIM [11] of nuclear recoil cascades in C_3F_8, both of which indicate that carbon recoils deposit their energy over a much greater distance than fluorine recoils of the same energy. Finally, no bubble nucleation is allowed for

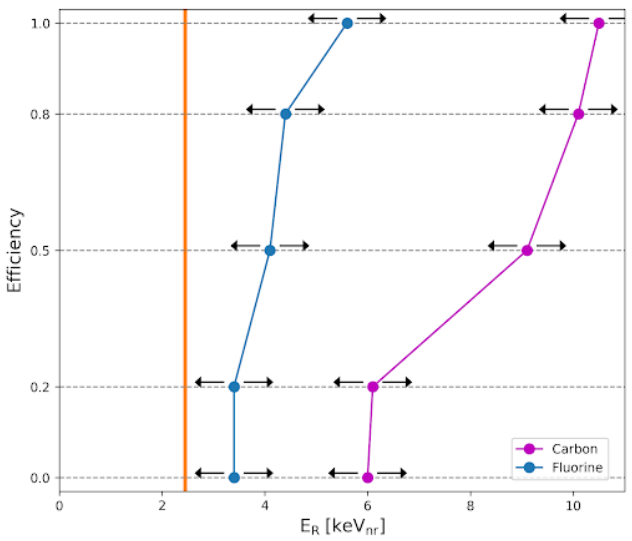

FIG. 11. Cartoon showing the piecewise linear model used in this analysis. Each blue (purple) dot represents a movable set-point of fluorine (carbon) at efficiencies of 0, 0.2, 0.5, 0.8, and 1. In this analysis, the carbon threshold is required to lie to the right of the fluorine threshold, which in turn must lie to the right of the Seitz threshold (orange vertical line).

recoils with energies below the Seitz threshold:

$$x_{s,0,T} \geq T. \tag{11}$$

The majority of the datasets shown in Table III are taken at thresholds near but not precisely at the thresholds used in PICO-60, so some E_T dependence must be assumed when using calibration data to constrain the $x_{s,p,T}$. This is done by extrapolating from the nearest-neighbor threshold fencepost. That is, for calibration data taken at threshold E_T, the efficiency curves applied to the calibration data are given by

$$x_{s,p}(E_T) = \frac{E_T}{\widehat{T}} x_{s,p,\widehat{T}}, \tag{12}$$

where $\widehat{T}$ is the nearest fencepost (2.45 keV or 3.29 keV) to E_T. This is done to diminish the influence of high-threshold calibration data on the low-threshold efficiency curve, and vice versa.

B. Systematic uncertainties

Two types of systematic uncertainty are considered in this analysis: uncertainties on the thresholds at which calibration data are taken, and uncertainties on the neutron exposure in each dataset. In both cases, uncertainties are treated as multipliers to the nominal threshold or neutron exposure, where the multiplier is sampled from a log-normal distribution with mean of one. Each source of uncertainty effects some subset of the datasets in Table III, though the magnitude of a given systematic effect may vary with dataset (e.g. a miscalibrated pressure transducer is a single source of systematic uncertainty, but has a larger impact at high threshold than it does at low threshold).

Nuisance parameter	% uncertainty (multiple entries indicate E_T dependence)
PT uncertainty: beam 2013	97 keV: 8.0 61 keV: 8.0
PT uncertainty: beam 2014	97 keV: 7.0 / 9.9 61 keV: 1.7 / 2.5 50 keV: 7.0 / 14
PT uncertainty: SbBe	6.5 / 7.0 / 7.5
PT uncertainty: Ambe	8.0
Fencepost Thresholds	3.0

TABLE IV. Five nuisance parameters describing uncertainty in calibration thresholds, and their one-sigma amplitudes. Each of the four calibration setups has an associated nuisance parameter capturing systematic pressure/temperature uncertainty. The final row captures uncertainty in the location of the two reference threshold fenceposts, which are defined by the PICO-60 operating conditions.

1. *Uncertainties on calibration thresholds*

The five nuisance parameters capturing systematic uncertainty on the thresholds at which calibration data are taken are shown in Table IV. The four pressure and temperature (PT) uncertainties represent the errors caused by mis-calibration of a chamber's pressure and temperature transducers (typically taken to be 1 psi and 0.1°C), as well as variations in pressure and temperature during the course of a run. These variations are modeled as coherent across a given experimental setup to minimize the number of nuisance parameters included in the analysis. Treating the PT variations in this way is conservative, in that it decreases the χ^2 cost associated with coherent fluctuations in threshold.

The final nuisance parameter impacting threshold is a global parameter representing uncertainty translating thresholds in our calibration chambers to thresholds in PICO-60. Put another way, the two threshold fenceposts given in Section III A are *defined* to be the PICO-60 operating condition, and this nuisance parameter captures all uncertainty on the threshold energy that condition corresponds to. This includes fundamental uncertainties in the Seitz threshold calculation (e.g. uncertainty in the Tolman length) as well as systematic mis-calibration of the pressure and temperature transducers in PICO-60.

2. *Uncertainties on neutron exposure*

Exposure uncertainties come in two parts: geometric uncertainties impacting neutron transport to the bubble chamber target, and uncertainties on the source strength. For SbBe and AmBe data, these uncertainties are combined into a single nuisance parameter, but for neutron beam data they are separated, allowing for a common geometry uncertainty in

Nuisance parameter	% uncertainty
PICO-0.1L Geometry 2013	7.5
PICO-0.1L Geometry 2014	3.1
Source: Beam, 2013, 97 keV	8.3
Source: Beam, 2013, 61 keV	4.7
Source: Beam, 2014, 97 keV	5.9
Source: Beam, 2014, 61 keV	5.0
Source: Beam, 2014, 50 keV	5.4
Source + Geometry: SbBe	10.3
Source + Geometry: AmBe	26

TABLE V. Nine nuisance parameters describing uncertainties in neutron exposure and their one-sigma amplitudes. The first two rows describe uncertainties in chamber geometry, impacting all 2013 and 2014 neutron beam data, respectively.

each setup with separate source-strength nuisance parameters at each beam energy (see Table V).

The dominant contribution for geometric uncertainty in PICO-0.1L comes from the length of the path through the water bath (i.e. the few-mm gap between the end of the neutron conduit and the quartz vessel) that neutrons must traverse to reach the target fluid. The width of this water gap in the 2013 setup is measured to be 2 ± 1 mm using camera images of the chamber. The water bath and neutron conduit were rebuilt in 2014 with significantly reduced uncertainty on the gap distance, resulting in the exposure uncertainties listed in line 1 and line 2 in Table V. Geometric uncertainties in PICO-0.1L in the SbBe setup are negligible relative to the source strength uncertainty. Uncertainties on source strength, both for the neutron beam flux and the SbBe photoneutron source, are given by the source strength measurements described in Section II D. Geometric uncertainty in PICO-2L is estimated based on the different neutron flux seen at the target fluid in the Geant4 and MCNP simulations described in Section II B. This uncertainty is much larger than uncertainty on the AmBe source strength, but has little impact on the final analysis as this nuisance parameter turns out to be well constrained by the global calibration dataset.

C. Mapping the log-likelihood function

Constraining the nuclear recoil bubble nucleation efficiency curves is achieved by mapping the log-likelihood given by Eq. (6) over the 34-dimensional parameter space using the *emcee* MCMC library [28], a python MCMC toolkit designed especially for high-dimensional spaces. This tool employs multiple inter-dependent "walkers" to explore the parameter space, evaluating the likelihood at multiple points in each step in the Markov Chain and creating a proposal for the entire ensemble of walkers at each step.

Even for a tool such as *emcee*, is is impractical to employ the MCMC in the traditional fashion, where the full 34-dimensional parameter space would be sampled with density proportional to the likelihood. Instead, we employ a novel iterative MCMC approach, dubbed "fast burn-in", designed for this analysis. Rather than attempt to fill in the 34-dimensional volume, the "fast burn-in" aims to map the envelope of the likelihood function along multiple 1-dimensional projections of the full parameter space. The shape of each 1-D envelope can then be used to constrain the projected parameter.

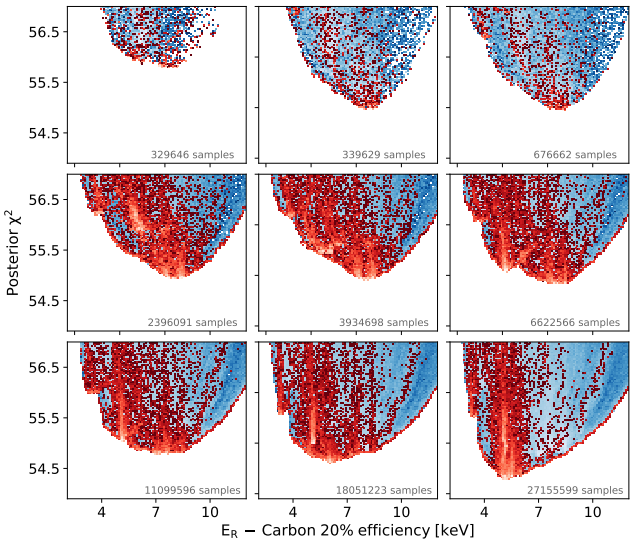

FIG. 12. The sampled χ^2 (or $-2\log\mathscr{L}$) distribution projected into $x_{\mathrm{C},20\%,2.45\mathrm{keV}}$ — the carbon recoil energy with 20% nucleation efficiency at a thermodynamic threshold of 2.45 keV — showing the progressive sampling of the distribution over many epochs (blue) with the events selected to start the next epoch shown in red. The cumulative number of points evaluated is indicated in each frame.

To accomplish the fast burn-in, the MCMC is run in a series of fixed-length "epochs". At the end of each epoch, the global set of likelihood evaluations thus far is examined, and points that fall on the boundary of the likelihood function in at least one of the 1-D projections (i.e. points that yield the maximum likelihood observed so far at a given value of the parameter of interest) are selected as the starting positions for the next epoch's walker ensemble. In this case, the 1-D projections are made on each of the 34 parameters of the likelihood function: the $\{x_{s,p,T}\}$ and $\{s_k\}$. That is, at the end of each epoch, all evaluations made thus far are projected onto the $(x, \log\mathscr{L})$ plane for each $x \in (\{x_{s,p,T}\} \cup \{s_k\})$. Each projection is divided into M bins in the projected dimension, and in each bin the point giving the highest $\log\mathscr{L}$ is identified. This yields up to $34 \times M$ points (selected points in each projection are not necessarily unique), which are used as the initial walker position set for the next epoch. This process is illustrated in Fig. 12.

This process can be further tuned by changing the proposal step scale of the MCMC (a unitless parameter defined in *emcee*), the number of MCMC steps per epoch (k), and the number of bins for each parameter at the end of the epoch (M). The likelihood function was first mapped with a rapidly traversing "exploratory" phase with a proposal step scale of 2, $k = 5$, and $M = 100$. Then, a more thorough exploration was performed with a step scale of 1.2, $k = 10$, $M = 500$.

For both phases of the MCMC exploration, the convergence criteria used were based on both the maximum likelihood

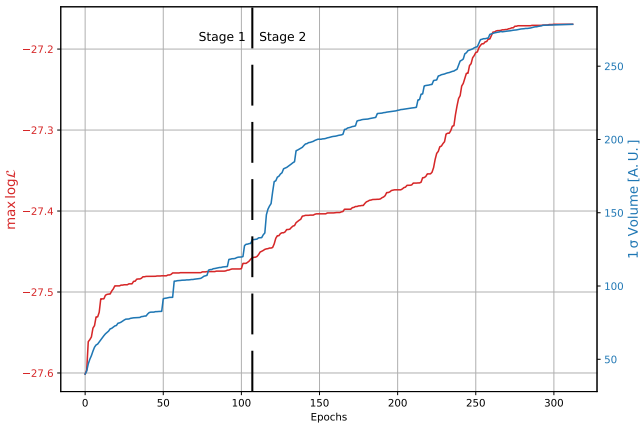

FIG. 13. Progression of the maximum log-likelihood and "1 σ" volume of the likelihood function.

reached and the "volume" of the likelihood function within "1 σ" of the current best-fit after each epoch. Specifically, the volume considered is the 34-dimensional volume subtended by all MCMC samples within 0.5 of the current maximum log-likelihood value, serving as a measure of the stability of the boundary of the likelihood function. Both these criteria proved useful, as often the MCMC's progress would halt temporarily for one quantity but not the other. The progression is shown in Fig. 13. The convergence criteria, checked after each epoch, was that there be 25 consecutive epochs with less than a 0.1 % change in log-likelihood or 1 σ volume. As an additional measure, it was required that the first (resp. second) phase reach at least 150 (resp. 300) epochs. While it cannot be known with absolute certainty if this criterion is sufficient to guarantee a converged fit to the data, fits of simulated datasets often satisfied the convergence criteria long before the mandated 150 and 300 epochs for the two stages of the fit, suggesting that this is indeed sufficiently stringent.

The computational resources required to carry out this procedure can be substantial. Fortunately, because the fast burn-in method results in large numbers (usually hundreds to thousands) of MCMC walkers running at once, the process can be greatly sped up by parallelization. For this analysis, the "Cedar" cluster operated by Compute Canada was used, using at least 8 cores in parallel. Even so, typical run-times for the fast burn-in code for this analysis are 4-6 weeks (5 to 8 thousand CPU hours).

D. Results

The best-fit nucleation efficiency curves and their 1σ error bands are shown in Fig. 14. The error bands were approximated to be the bounds of the parameter space where $\log\mathscr{L} \geq \max\{\log\mathscr{L}\} - \frac{1}{2}$ (implicitly profiling over all other parameters) for the purpose of convenient visualization. These results are consistent with those used in [4, 5]; the WIMP exclusion limits in [5] follow a dedicated fast burn-in on WIMP sensitivity, as described in Appendix A. The goodness-of-fit for these

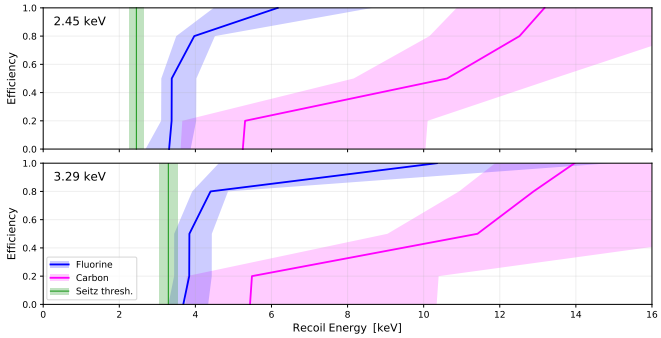

FIG. 14. Best-fit and 1σ error bands for the nucleation efficiency curves of fluorine (blue) and carbon (magenta), at the thresholds used in the PICO-60 WIMP search [5]. Error bands indicate the 1σ range of each knot position, so that every efficiency curve within 1σ of the best-fit falls in the shaded area (but not every curve falling in the shaded area is within 1σ of the best-fit). The corresponding Seitz thresholds are shown as a vertical green line, with the green band indicating the systematic threshold uncertainties described in Table IV.

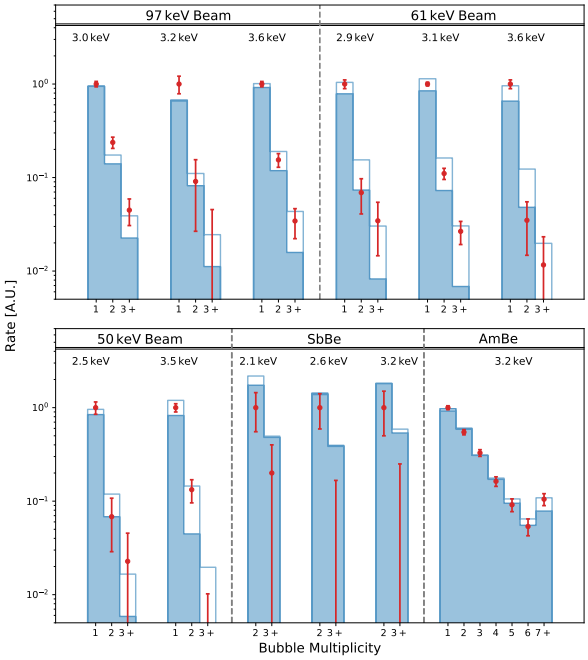

FIG. 15. Comparison between the best-fit bubble rate multiplicity and experimental data. The red points are the experimental data points, while the blue histograms are the result of the MCNP or Geant4 simulations convolved with the best-fit bubble efficiency curves found with the MCMC (with best-fit nuisance parameter values applied). The empty blue bars represent the $\pm 1\sigma$ error window of the fit.

results can be seen in Fig. 15 (showing predicted vs observed bubble rates in calibration data) and Fig. 16 (showing posterior constraints on nuisance parameters). Notably, while the fluorine curve turns on near the Seitz threshold at both fenceposts, all nucleation efficiency curves deviate significantly from the Seitz threshold at high nucleation probability and

have non-trivial shapes that are not readily comparable to standard functional forms such as a sigmoid or exponential function. It is also worth noting that the carbon bubble nucleation threshold is significantly higher than the fluorine threshold, as expected — the by-hand constraint enforcing that relation has almost no impact on the result. The fluorine efficiency curves do tend to have smaller uncertainty bands than carbon, primarily because a greater proportion of the bubbles produced in the calibration experiments, particularly at low recoil energy, are created from recoils with fluorine.

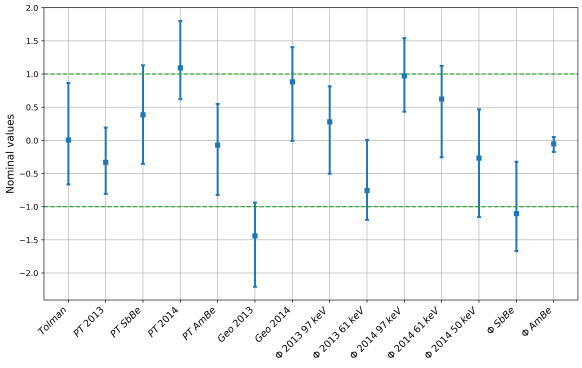

FIG. 16. Posterior constraints on nuisance parameters. Each nuisance parameter on this plot has a Gaussian prior with mean of 0 and standard deviation of 1. 11 of 14 best-fit values fall within the 1σ prior band, and in all cases the 1σ posterior error bars overlap with the 1σ prior band. The right-most point, representing the neutron flux from the AmBe source in PICO-2L, is highly constrained by the data, largely due to the high-bubble-multiplicity resolved in that dataset.

IV. PARAMETRIC MONTE CARLO STUDY

An important test of this methodology that can be performed is to generate and fit simulated datasets. Doing so provides numerous benefits, including the ability to check if there are any inherent biases in the model, validating the convergence criteria described in Section III C, and providing context for the final posterior χ^2 value obtained for the fit of the data. To generate the simulated data, the best-fit model depicted in Fig. 14 is used (hence "parametric"). For each experiment, threshold, and multiplicity, the expected number of counts is calculated from the model. Then a Poisson-distributed random number is inserted into the dataset. Examples of this simulated data are shown in Fig. 17.

These simulated datasets are then fit with the same procedure described in Section III C. The results of these 25 fits compared to the original best-fit are shown in Fig. 18. To ensure that the convergence criteria defined for the original fit were sufficient, these fits were continued until they met that criteria, plus an additional 50 epochs in stage 2 of the fit to catch any MCMCs potentially caught in local minima. No jumps in the maximum log-likelihood or $1\,\sigma$ volume as previously defined were observed in the last 50 epochs, suggesting that these criteria were sufficient.

Another powerful result of this study is the ability to interpret the posterior χ^2 values obtained as proper goodness-of-fit statistics. This is non-trivial due to the strong correlations between many of the fit parameters, increasing the effective number of degrees of freedom far beyond the nominal expectation of $\#\text{points} - \#\text{parameters} = 17$. The distribution of final χ^2 values for the simulated datasets is shown in Fig. 19. By fitting this distribution with the χ^2 PDF, an effective number of degrees of freedom of 46 is obtained. In that case, the value of χ^2 obtained for the real data of 54.2 is reasonable, being $< 1\,\sigma$ away from the expectation, with a p value of $p = 0.19$.

Because the nucleation efficiency model presented in this paper is ad hoc and un-tested on other data, it is prudent to wonder if the model and fitting method produce unbiased results. The parametrically simulated data described in this section can also be used to assess this by comparing the fit results to the input values. One can see in Fig. 18 that in some cases there are small, but persistent systematic offsets between the true value of a parameter and the 25 fits, such as at an efficiency of 0 for carbon with a threshold of 3.29 keV. A full characterization of possible model biases (and correction for them) is presented in Appendix B.

V. SUMMARY

A set of mono-energetic and broad-spectrum neutron sources have been used to calibrate the nuclear recoil response of C_3F_8 in PICO detectors. A flexible, many-parameter functional form for nucleation efficiency is assumed for data-driven analysis. A modified MCMC approach was used for exploration of the high-dimensional likelihood space to simultaneously estimate nucleation efficiency at 3.29-keV and 2.45-keV thermodynamic thresholds, for both carbon and fluorine recoils. All of these nucleation efficiency curves deviate from their corresponding Seitz threshold. The 50 % nucleation efficiency point for fluorine recoils was found to be 3.3 keV (3.7 keV) at a thermodynamic Seitz threshold of 2.45 keV (3.29 keV), and for carbon the efficiency was found to be 50 % for recoils of 10.6 keV (11.1 keV) at a threshold of 2.45 keV (3.29 keV). The fact that the fluorine efficiency curves at both thresholds are nearly the same suggests that there is not enough data at the lower threshold of 2.45 keV to fully constraint this result. Also, the relative lack of calibration data for carbon alone resulted in wider uncertainty bands for those efficiency curves.

Beyond these initial results, an extensive study of simulated datasets was carried out. This validated the convergence criteria used for the MCMC exploration of the real calibration data. Additionally, the χ^2 results for the simulated datasets were used to calculate a p-value for the fit of the real data of 0.19, confirming that the model is adequately flexible to describe the data and that the MCMC procedure resulted in a reasonable fit. Finally, these results were also used to characterize small systematic biases in this fit paradigm, providing bias-corrected nucleation efficiency curves. This correction is small, and therefore does not invalidate previous results ob-

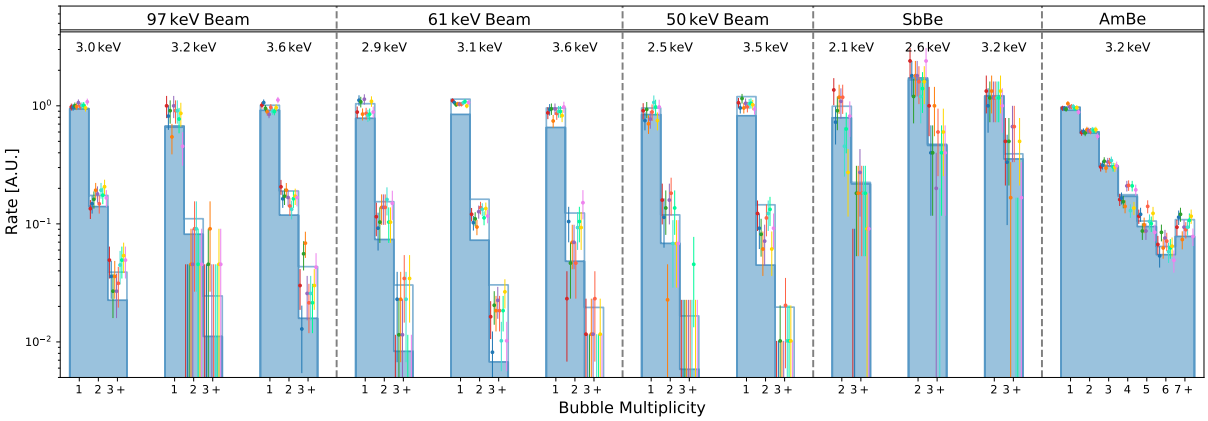

FIG. 17. 10 parametrically simulated datasets (multi-colored points) based on the best-fit to the data (blue histogram with empty bars for $\pm 1\,\sigma$ bands).

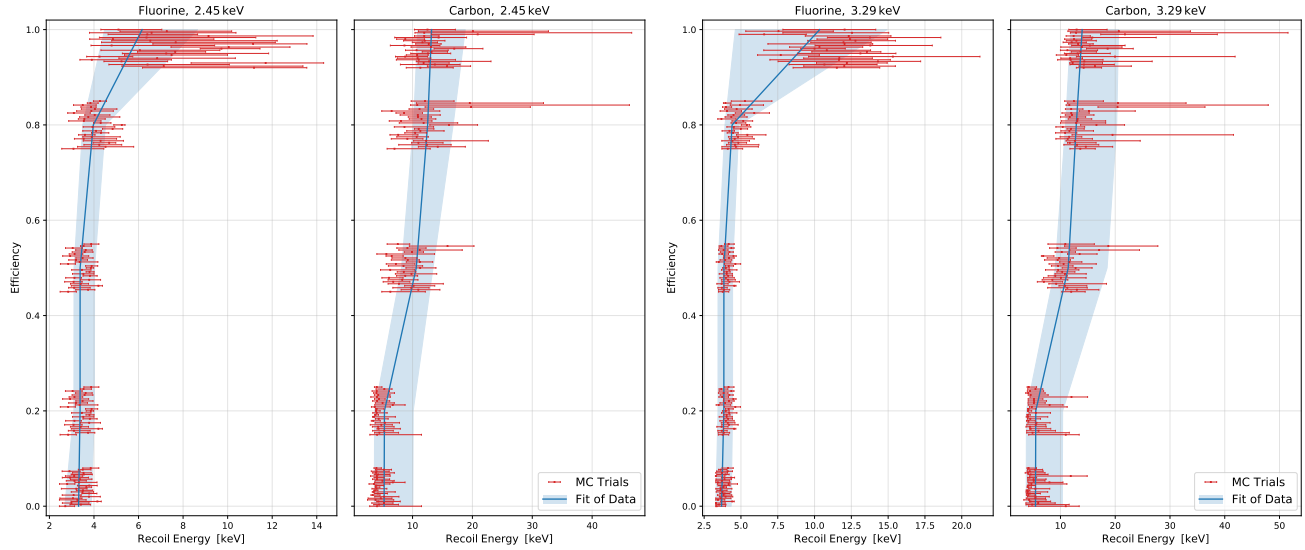

FIG. 18. Fits of 25 simulated datasets (red) compared to the best-fit and $1\,\sigma$ band of the actual data (light blue) for the thresholds of 2.45 keV (left) and 3.29 keV (right).

tained with this data.

ACKNOWLEDGEMENTS

The PICO collaboration wishes to thank SNOLAB and its staff for support through underground space, logistical and technical services. SNOLAB operations are supported by the Canada Foundation for Innovation and the Province of Ontario Ministry of Research and Innovation, with underground access provided by Vale at the Creighton mine site. We wish to acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canada Foundation for Innovation (CFI) for funding, and the Arthur B. McDonald Canadian Astroparticle Physics Research Institute. We acknowledge that this work is supported by the National Science Foundation (NSF) (Grant 0919526, 1506337, 1242637, and 1205987), by the U.S. Department of Energy (DOE) Office of Science, Office of High Energy Physics (grants No. DE-SC0017815 and DE-SC0012161), by the DOE Office of Science Graduate Student Research (SCGSR) award, by the Department of Atomic Energy (DAE), Government of India, under the Centre for AstroParticle Physics II project (CAPP-II) at the Saha Institute of Nuclear Physics (SINP), and Institutional support of IEAP CTU (DKRVO). This work is also supported by the German-Mexican research collaboration grant SP 778/4-1 (DFG) and 278017 (CONACYT), the project CONACYT CB-2017-2018/A1-S-8960, DGAPA UNAM grant PAPIIT-IN108020, and Fundación Marcos Moshinsky. This work is partially supported by the Kavli Institute for Cosmological Physics at the University of Chicago through NSF grants 1125897 and 1806722, and an endowment from the Kavli Foundation and its founder Fred Kavli. We also wish to

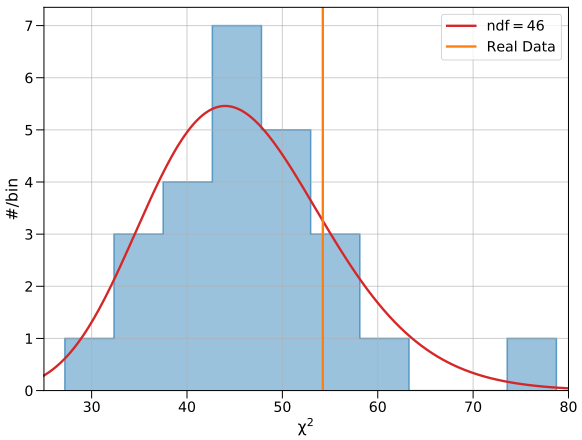

FIG. 19. χ^2 data for the 25 simulated datasets (blue histogram), the best-fit χ^2 distribution to that data (red curve), and the value of χ^2 obtained for the real data (orange line). Based on the best-fit χ^2 distribution, the real data give a p-value of 0.19.

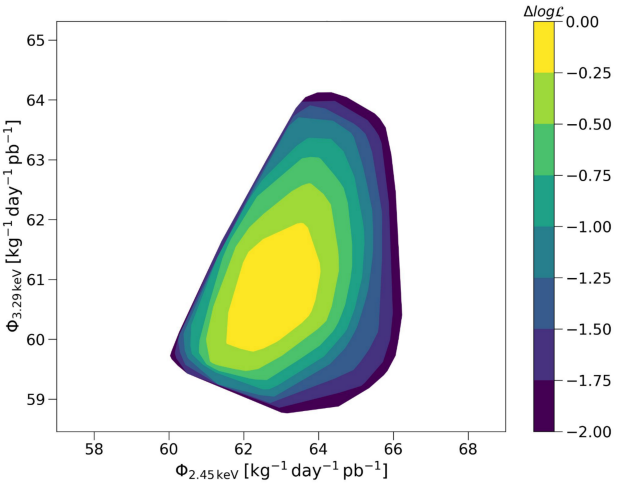

FIG. 20. Posterior likelihood plot for sensitivity to 50-GeV WIMPs with SDp coupling.

acknowledge the support from Fermi National Accelerator Laboratory under Contract No. DE-AC02-07CH11359, and from Pacific Northwest National Laboratory, which is operated by Battelle for the U.S. Department of Energy under Contract No. DE-AC05-76RL01830. We also thank Compute Canada (`www.computecanada.ca`) and the Centre for Advanced Computing, ACENET, Calcul Québec, Compute Ontario, and WestGrid for computational support. The work of M. Bressler is supported by the Department of Energy Office of Science Graduate Instrumentation Research Award (GIRA). The work of D. Durnford is supported by the NSERC Canada Graduate Scholarships - Doctoral program (CGSD). IUSB wishes to acknowledge the work of D. Marizata.

APPENDIX A: ANALYSIS OF WIMP SENSITIVITY

Early PICO dark matter searches using C_3F_8 took conservative 1-σ nucleation efficiency curves to produce WIMP exclusion limits [2, 4]. However, with a profile likelihood ratio analysis, the nucleation efficiency likelihood can be directly incorporated into the WIMP limit calculation, eliminating a redundant conservative step and thus improving the final WIMP limit at most WIMP masses. To that end, the log-likelihood function for the nuclear recoil calibration data described in Eq. (6) is recast as a likelihood function for the expected number of WIMP interactions for a given dark matter particle mass M_χ and coupling $\mathscr{O}$, spin-independent (SI) or spin-dependent on proton (SDp). For the recent PICO-60 results [5], physics data taken at thermodynamic thresholds of 3.29 keV and 2.45 keV is combined form to a WIMP exclusion limit, so correlated uncertainties in the bubble nucleation efficiencies at these two thresholds must also be mapped out. To achieve this, the fast burn-in MCMC procedure of section III C is used to explore the likelihood band projected onto a set WIMP event rates $\{\Phi_{M_\chi,T,\mathscr{O}}\}$,

$$\Phi_{M_\chi,T,\mathscr{O}} = \sum_{s=\mathrm{C,F}} \int_0^\infty R_{s,M_\chi,\mathscr{O}}(E_r)\cdot\varepsilon_s(E_T,E_r)\,dE_r, \qquad (13)$$

where $R_{s,M_\chi,\mathscr{O}}$ is the differential recoil spectrum (normalized by total WIMP-nucleon cross-section) for WIMPs with mass M_χ and coupling $\mathscr{O}$ on recoil species s. For this analysis we take $M_\chi \in \{3\text{ GeV}, 10\text{ GeV}, 19\text{ GeV}, 50\text{ GeV}\}$, $T \in \{2.45\text{ keV}, 3.29\text{ keV}\}$, and $\mathscr{O} \in \{\text{SI, SDp}\}$, calculating recoil spectra for a standard halo model as described in [29] with $\rho = 0.3\,\mathrm{GeV/cm^3}$, $v_{\mathrm{esc}} = 544\,\mathrm{km/s}$, $v_{\mathrm{Earth}} = 232\,\mathrm{km/s}$, and $v_0 = 220\,\mathrm{km/s}$. Calculations are performed following [29] — the python package *dmdd* was used specifically for fluorine/SDp coupling [30].

In addition to the 16 $\{\Phi_{M_\chi,T,\mathscr{O}}\}$ listed above, another 16 burn-in parameters are constructed from linear combinations of Φ's at different threshold,

$$\Phi_{M_\chi,\pm,\mathscr{O}} = \Phi_{M_\chi,2.45\mathrm{keV},\mathscr{O}} \pm \Phi_{M_\chi,3.29\mathrm{keV},\mathscr{O}}, \qquad (14)$$

to explore correlated uncertainties at different thresholds. The reseeding of new epochs is also changed slightly in this phase, binning in posterior likelihood (40 bins from the best-fit to $\Delta\log\mathscr{L} = -1$) rather than in the projected dimension, and taking the high and low extreme values of the projected parameter in each likelihood bin to seed the next epoch. The MCMC results can then be used to plot the 2D posterior likelihood for a given WIMP mass and coupling, as in Fig. 20. A frequentist WIMP exclusion curve can incorporate these results by treating the normalized WIMP rate Φ as a nuisance parameter with χ^2 cost defined by this analysis, as was done in [5].

APPENDIX B: MODEL BIAS

The fit results of 25 parametrically simulated data sets shown in Fig. 18 are all in reasonable agreement with the input parameters used to generate the Monte Carlo data (i.e. at

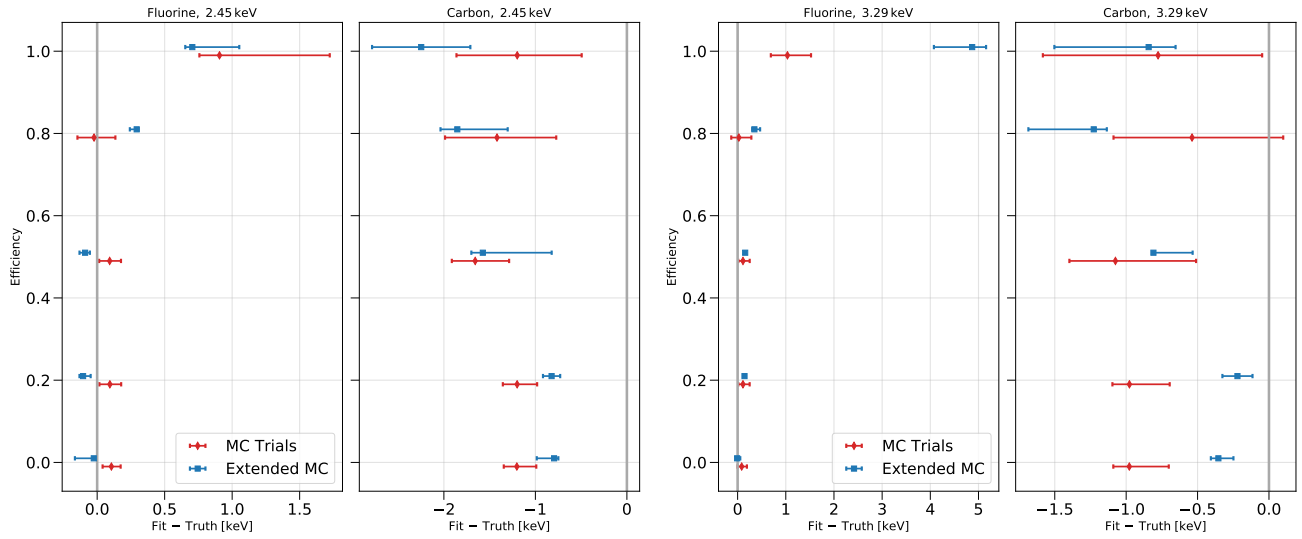

FIG. 21. Average residuals of 25 Monte Carlo dataset fits (red) and 25 extended Monte Carlo dataset fits (blue), for fluorine and carbon at the 2.45 keV (left) and 3.29 keV (right) threshold fenceposts.

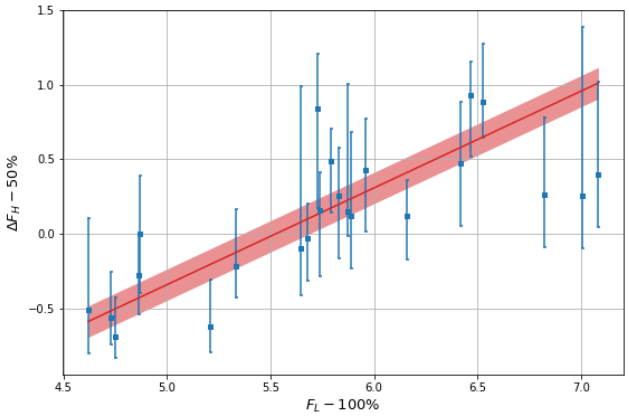

FIG. 22. Bias of the 50% efficiency point of fluorine at 3.29 keV vs. the true value of the fluorine 100% efficiency point at 2.45 keV thermodynamic threshold. The Monte Carlo trial results are shown in blue, and a linear fit to the points with $1\,\sigma$ error band are shown in red.

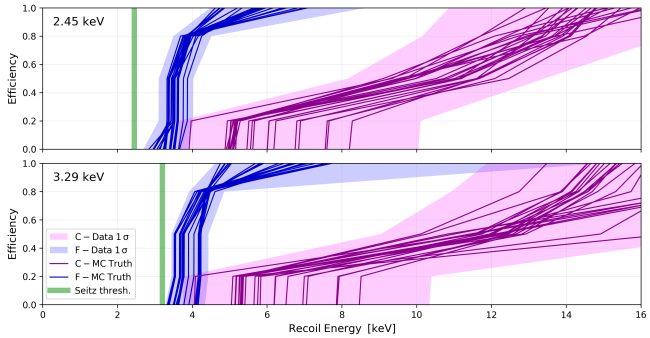

FIG. 23. Monte Carlo inputs for 25 "extended" parametrically simulated datasets, compared to the $1\,\sigma$ error bands from the fit to the real data.

no point do the fits completely lie outside the $1\,\sigma$ band of the true model). However, it is also apparent from Fig. 18 that in some cases, there are persistent (albeit small) offsets between the average fit results and the input parameters, such as for fluorine at an efficiency of 1 for either threshold. To provide a more clear estimation of the magnitude of the bias, the results for the 25 simulated datasets are averaged by combining the marginal log-likelihood functions in each parameter [31]. This is shown in Fig. 21, and confirms the conclusions that can be drawn from Fig. 18 as well — that the fluorine 100% point is consistently over-estimated, and the carbon efficiency curves under-estimated. The average residual for each parameter can be used to define a "bias function" B_θ, with which the estimate of a parameter θ is given by:

$$E[\hat{\theta}] = \theta_{\text{True}} + B_\theta(\theta_{\text{True}}). \tag{15}$$

This can be used to amend reported uncertainties to account for the bias in the methodology, or even correct for it. Conservatively, one could expand the reported error bands for the data fit by subtracting the upper or lower (as appropriate) $1\,\sigma$ limit of B_θ from each parameter. Less conservatively, one could shift the uncertainty band on both sides by the appropriate bias limit, and shift the reported best-fit value by the best-fit value of B_θ.

However, this nominal method of application assumes that the bias function is "flat", i.e. that it is a constant not dependent on the true value of any parameters. Specifically, the presumption is that $B(\theta_{\text{True}}) = B(\hat{\theta})$ for the real data. Since the Monte Carlo data described was generated from a single point in the parameter space, this assumption remains untested, with the worst-case scenario being that the bias function varies rapidly as a function of multiple parameters (the bias in θ_i

FIG. 24. Spearman correlation coefficient for the bias of every parameter of interest (y-axis) vs. the simulation input value (x-axis). The colour-scale is indicative of this coefficient as well for easier visualization, and instances of a statistically significant correlation are boxed. Note that in this variable abbreviation, "C" or "F" indicate the target atom, "L" or "H" indicate the thermodynamic threshold fencepost as 2.45 keV or 3.29 keV respectively ("Low" or "High"), and the numeric value is the efficiency fencepost.

could depend on the true value of θ_j).

To rectify this, ideally, datasets would be generated with each parameter varied independently to be $\pm 1\sigma$ from the best-fit values, spanning the entire parameter space around the estimated best-fit, with multiple trials in each case. However, this is not computationally practical. A more practical strategy is to instead vary all parameters around the best-fit point at once, by simulating datasets from efficiency curves drawn randomly from within the one sigma contour of the original fit to the data. 25 such "extended" parametric Monte Carlo datasets were generated from the input efficiency curves shown in Fig. 23. These datasets are fit in the usual way, and their average residuals are shown in Fig. 21. Compared to the average results of the original Monte Carlo study, the biases observed are roughly consistent in magnitude and direction.

However, using the average residual to calculate the bias in each parameter as a function of itself alone ignores the possibility of dependence on other parameter true values. It is thus essential to examine the bias of each parameter of interest as a function of the true value of every parameter of interest (yielding 400 combinations). Indeed, while there are many cases where the bias function does not vary with the input parameter, there are instances where the bias function is not constant. One such example is shown in Fig. 22.

One test that can succinctly summarize this complicated relationship is the correlation coefficient of the bias vs. true pa-

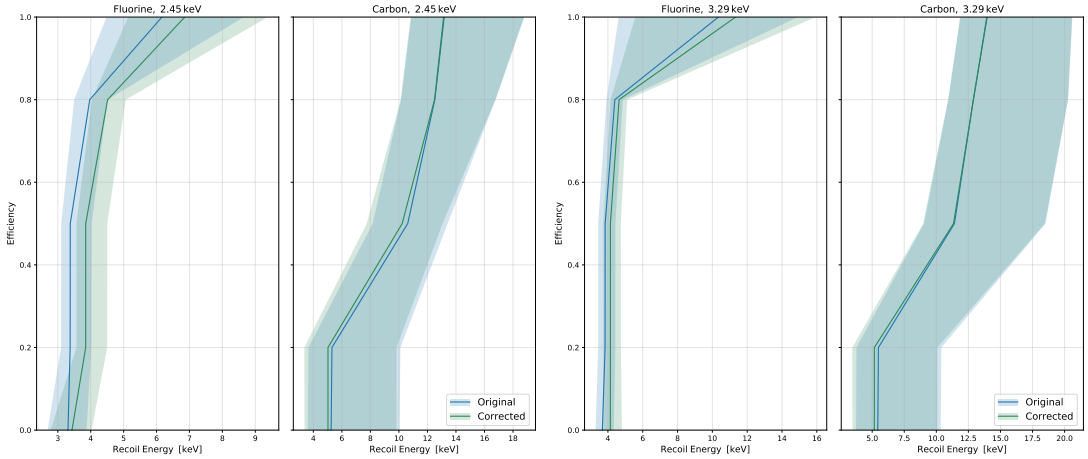

FIG. 25. Original (blue) and bias-corrected (green) nucleation efficiency curves and $1\,\sigma$ error bands for fluorine and carbon at thermodynamic thresholds of 2.45 keV (left) and 3.29 keV (right).

rameter value for every combination of the 20 parameters of interest. Specifically, the Spearman correlation coefficient ρ is appropriate in this case, as it is a non-parametric test [32]. For a sample size of 25, the null hypothesis that no correlations exists can be rejected at the 90% confidence level if the $|\rho| > 0.324$ [32]. This test statistic for every combination of the parameters of interest is shown in Fig. 24.

These results are then used to compute the bias function of every parameter, a 20-dimensional scalar function. A first or zeroth-order polynomial is used as a model for each parameter combination with or without a statistically significant correlation respectively. Before this, however, an outlier rejection procedure is applied to remove strenuous data points. For every parameter combination and every data point, a fit with a first-order polynomial is performed on all the data except the point in question. If this point is a $> 2\,\sigma$ outlier of that fit, it is rejected. For the final fit, only uncertainty in the y-intercept is included to prevent the uncertainty of each bias function from vanishing at the central region of parameter space (due to uncertainty in the slope of the first-order fits). An example of one such linear fit and resulting error band is shown in Fig. 22. This yields 400 constraints on the true values of the 20 parameters-of-interest in the form:

$$\hat{\theta}_j = \theta_j + B_{ji}(\theta_i) \tag{16}$$

relating one estimated parameter value $\hat{\theta}_j$ to the true values of the same parameter and another θ_i, with one of the 400 bias functions. B_{ji} is the bias in parameter j as a function of parameter i, and is normally distributed according to the first order polynomial fits described above. This is an over-determined, non-linear system of equations. To find an optimal solution, the following likelihood function was constructed and maximized:

$$\log \mathscr{L}\left(\{\hat{\theta}\}\,|\,\{\theta\}\right) = \sum_i \sum_j \log\left[P_{\mathrm{Norm}}\left(\hat{\theta}_j = \theta_j + B_{ji}(\theta_i)\right)\right]. \tag{17}$$

This fit was performed for the original best-fit parameter values to obtain the corrected best-fit nucleation efficiency curves. The new result is compared to the original efficiency curves in Fig. 25. It was not computationally practical to apply this method to all the MCMC samples explored in the original fit to build new $1\,\sigma$ error bands. Therefore, the approximate new results presented in Fig. 25 are produced by shifting the original error bands by the offset of the corrected best-fit curve and expanded by the statistical uncertainty in the fit of the corrected result. Fortunately, the bias-corrected result is not significantly discrepant with the original best-fit (within $1\,\sigma$ agreement), so this new analysis does not cast doubt on the WIMP sensitivity example shown in Appendix A (which is not bias-corrected) or previously published results using the same PICO calibration data [4, 5]. Indeed, the close overlap between the original and bias-corrected results justifies ignoring the fit biases in analyses with this data.

In addition to the analysis described above, further studies were carried out to investigate the validity of this fit paradigm. Specifically, non-parametric datasets were simulated and fit. Pathological cases to test the extreme limits of the model were also investigated. These studies did not lead to any definite or quantitative results, and so are not included here.

[1] E. Behnke *et al.* (COUPP), Phys. Rev. D **86**, 052001 (2012), [Erratum: Phys.Rev.D 90, 079902 (2014)], arXiv:1204.3094 [astro-ph.CO].

[2] C. Amole *et al.* (PICO), Phys. Rev. Lett. **114**, 231302 (2015), arXiv:1503.00008 [astro-ph.CO].

[3] C. Amole *et al.* (PICO), Phys. Rev. D **93**, 052014 (2016),

arXiv:1510.07754 [hep-ex].
[4] C. Amole *et al.* (PICO), Phys. Rev. Lett. **118**, 251301 (2017), arXiv:1702.07666 [astro-ph.CO].
[5] C. Amole *et al.* (PICO), Phys. Rev. D **100**, 022001 (2019), arXiv:1902.04031 [astro-ph.CO].
[6] E. Behnke *et al.*, Astropart. Phys. **90**, 85 (2017), arXiv:1611.01499 [hep-ex].
[7] F. Seitz, The Physics of Fluids **1**, 2 (1958).
[8] C. Amole *et al.* (PICO), Phys. Rev. D **100**, 082006 (2019), arXiv:1905.12522 [physics.ins-det].
[9] M. J. Harper, Nuclear Science and Engineering **114**, 118 (1993), https://doi.org/10.13182/NSE93-A24023.
[10] R. C. Tolman, The Journal of Chemical Physics **17**, 333 (1949).
[11] J. F. Ziegler, M. D. Ziegler, and J. P. Biersack, Nuclear Instruments and Methods in Physics Research B **268**, 1818 (2010).
[12] F. d'Errico, Radiation Protection Dosimetry **84**, 55 (1999), https://academic.oup.com/rpd/article-pdf/84/1-4/55/9924398/55.pdf.
[13] F. d'Errico, Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms **184**, 229 (2001), advanced Topics in Solid State Dosimetry.
[14] F. Tardif, *Direct Detection of Dark Matter with the PICO Experiment and the PICO-0.1 Calibration Chamber*, Master's thesis, Université de Montréal (2018).
[15] S. Archambault *et al.* (PICASSO), New J. Phys. **13**, 043006 (2011), arXiv:1011.4553 [physics.ins-det].
[16] E. Behnke *et al.* (COUPP), Phys. Rev. D **88**, 021101 (2013), arXiv:1304.6001 [physics.ins-det].
[17] Y.-C. Lo and R. E. Apfel, Phys. Rev. A **38**, 5260 (1988).
[18] M. Felizardo *et al.*, Phys. Rev. Lett. **105**, 211301 (2010), arXiv:1003.2987 [astro-ph.CO].
[19] A. E. Robinson, Phys. Rev. C **89**, 032801 (2014), arXiv:1401.7900 [nucl-ex].
[20] D. A. Brown *et al.*, Nucl. Data Sheets **148**, 1 (2018).
[21] J. H. Gibbons, R. L. Macklin, and H. W. Schmitt, Phys. Rev. **100**, 167 (1955).
[22] M. Lafreniere, *Mesures d'étalonnage aux neutrons et caractérisation par étude Monte Carlo de la réponse des détecteurs à gouttelettes surchauffées conçus pour la recherche et la détection directe du neutralino (la matière sombre) menant aux résultats finaux de l'expérience PICASSO*, Ph.D. thesis, Université de Montréal (2018).
[23] J. I. Collar, Phys. Rev. Lett. **110**, 211101 (2013), arXiv:1303.2686 [physics.ins-det].
[24] A. E. Robinson, Phys. Rev. C **94**, 024613 (2016), arXiv:1602.05911 [nucl-ex].
[25] S. Agostinelli *et al.* (GEANT4), Nucl. Instrum. Meth. A **506**, 250 (2003).
[26] S. A. Pozzi, E. Padovani, and M. Marseguerra, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **513**, 550 (2003).
[27] E. Mendoza and D. Cano-Ott, *Update of the Evaluated Neutron Cross Section Libraries for the Geant4 Code*, Tech. Rep. (International Atomic Energy Agency (IAEA), 2018).
[28] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, Publ. Astron. Soc. Pac. **125**, 306 (2013), arXiv:1202.3665 [astro-ph.IM].
[29] J. Lewin and P. Smith, Astroparticle Physics **6**, 87 (1996).
[30] V. Gluscevic, M. I. Gresham, S. D. McDermott, A. H. G. Peter, and K. M. Zurek, JCAP **12**, 057 (2015), arXiv:1506.04454 [hep-ph].
[31] R. Barlow, in *Statistical Problems in Particle Physics, Astrophysics and Cosmology* (2004) pp. 56–59, arXiv:physics/0406120.
[32] W. Daniel, *Applied Nonparametric Statistics*, Duxbury advanced series in statistics and decision sciences (PWS-KENT Pub., 1990).