Appendix A

Veto Validation and Selection Controls

A.1 Scintillator and cadmium neutron data . . . 2
A.2 Transport geometry and prototype channel state . . . 4
A.3 Detailed veto-physics validation . . . 6
A.3.1 High-precision neutron data . . . 6
A.3.2 Secondary-neutron production in lead . . . 7
A.3.3 Cadmium capture-gamma response . . . 9
A.4 Tracking-orientation systematic . . . 10
A.5 Supplementary waveform and selection diagnostics . . . 11
A.5.1 Run-disjoint classifier hierarchy . . . 13
A.6 Calibration-controlled late-window population . . . 22
A.6.1 Input products and preprocessing . . . 22
A.6.2 Score definition . . . 23
A.6.3 Template mismatch and interpretation . . . 25
A.7 Transfer to the published IAXO-D0 candidates . . . 26
A.8 TPC-selected tracks as an experimental veto control sample . . . 28
A.8.1 Experimental sample and independent denominator . . . 28
A.8.2 TPC-only track selection and line fits . . . 28
A.8.3 First evaluation of the frozen veto selection criteria . . . 30
A.8.4 Comparison of experimental and simulated veto signals . . . 31
A.8.5 Exploratory classifiers and prompt-preserving hybrids . . . 34
A.8.6 Historical independent veto and joint background selection . . . 36
A.9 Observable importance in the adopted veto . . . 39
A.9.1 Reliance at the frozen operating point . . . 41
A.9.2 Controlled timing and spatial ablation . . . 42

A.9.3 Exploratory projection of the veto observables 43
A.10 Acquisition conditions and design controls 45
A.10.1 Material ordering and quenching 45
A.10.2 Capture and reconstructed-response diagnostics 47
A.10.3 Generated exposure in the HENSA layer scan 48
A.10.4 Passive-shielding and late-window cross-checks 49
A.10.5 Closed-layer shielding: counts and statistical intervals 49

A.1 Scintillator and cadmium neutron data

The evaluated-data comparisons below supplement the design discussion in the veto-system chapter. They test constituent cross-section inputs for BC408 scintillator and cadmium; they do not by themselves validate moderation times or detector response.

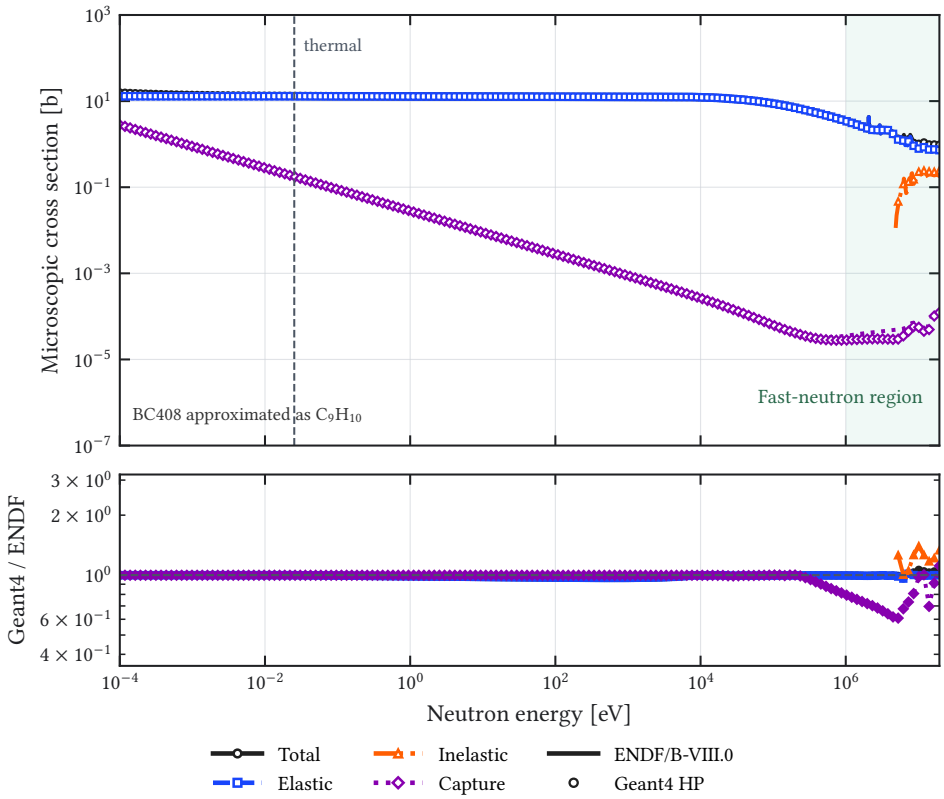


Figure A.1: Comparison of ENDF/B-VIII.0 and Geant4 high-precision neutron cross sections for BC408, approximated as C_9H_{10}. The comparison tests consistency of the constituent-weighted elastic, capture, and inelastic input data. It does not test bound-hydrogen thermal scattering, secondary energy–angle distributions, or the full moderation-time response.

The reference curves use ENDF/B-VIII.0 MF=3 pointwise cross sections. BC408 is approximated as polyvinyltoluene, C_9H_{10}, with atom-fraction-weighted carbon and hydrogen contributions; cadmium uses natural isotope abundances. The Geant4 points use G4NDL4.6 for version 11.0.3 and the same channel grouping as the natural-lead comparison in Section A.3.

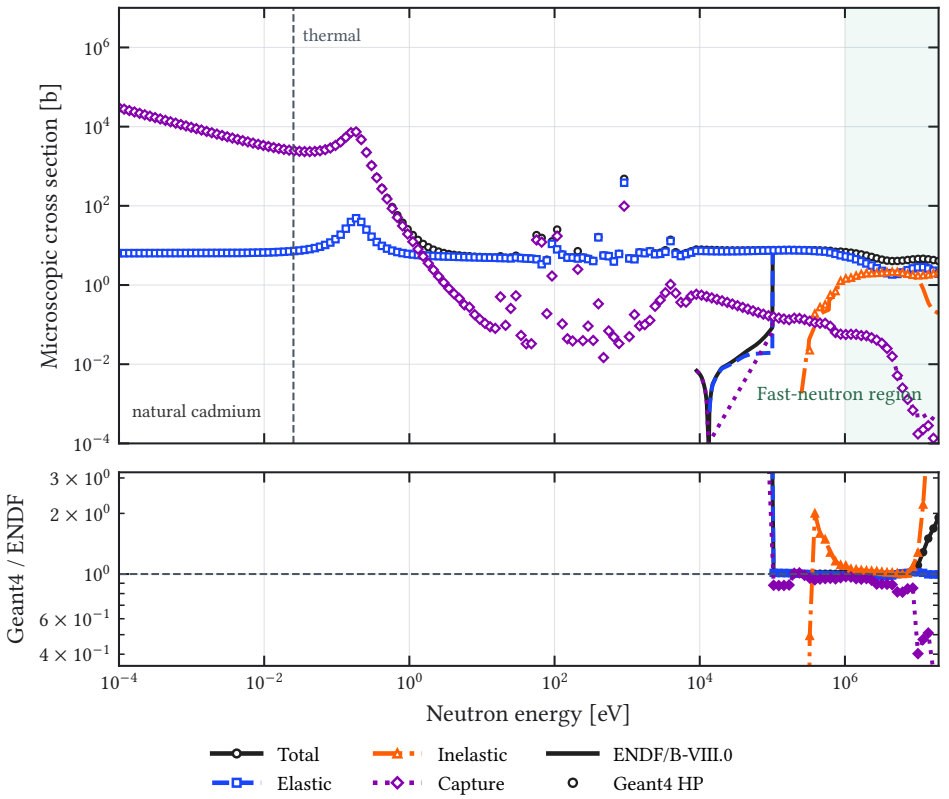


Figure A.2: Comparison of ENDF/B-VIII.0 and `Geant4` high-precision neutron cross sections for natural cadmium. The large thermal and epithermal capture cross section motivates cadmium layers next to scintillator. The emitted gamma cascade is prompt relative to capture; moderation and transport can delay the capture relative to the initiating event.

A.2 Transport geometry and prototype channel state

The number of scintillators in the transport geometry, the number of defined readout positions, and the number of operating channels describe different detector conditions. The archived IAXO-D0 readout defines 59 mapped panel positions, whereas 57 channels were operational in the published measurements. The two broken channels are retained as placeholder DAQ identifiers 6000 and 7000 for `Right_L1_N2` and `Right_L1_N3`. The paper states, "In total 57 veto panels were installed," and gives prototype bar lengths of 150 cm and 65 cm [1]. We distinguish that published count from the complete 59-position readout inventory; the readout alone does not establish an independently verified count of installed bars.

Two simulated geometry revisions are preserved. The selected 59-panel design is implemented by the IAXO-D1 `Default.gdml` and `CompleteVeto3Layers.gdml` files in the `iaxo-geometry` repository at revision `1883e466c44f`. The

corresponding IAXO-D1 generator, `VetoSystem.kt`, builds three short inner-Top panels and four long panels in each remaining Top layer. A separate revision, `079aa389f714`, contains 60 scintillators in its byte-identical `Default.gdml` and `CompleteVeto3Layers.gdml` files. These are the G1 and G3 snapshots captured on September 21, 2026, and identified in `registry.json` under `outputs/geometry-registry-20260921`. The counts in Table A.1 include only scintillator instances reachable from each GDML world volume. Neither a nominal filename nor the September snapshot assigns the same contents or revision to every historical HENSA production from May 2026.

Face	Selected GDML	September GDML	Prototype positions	Operating channels
Top	3 / 4 / 4	4 / 4 / 4	3 / 4 / 4	3 / 4 / 4
Bottom	4 / 4 / 4	4 / 4 / 4	4 / 4 / 4	4 / 4 / 4
Right	3 / 3 / 3	3 / 3 / 3	3 / 3 / 3	1 / 3 / 3
Left	3 / 3 / 3	3 / 3 / 3	3 / 3 / 3	3 / 3 / 3
Front	3 / 3 / 3	3 / 3 / 3	3 / 3 / 3	3 / 3 / 3
Back	3 / 3 / 3	3 / 3 / 3	3 / 3 / 3	3 / 3 / 3
Total	59	60	59	57

Table A.1: Three-layer panel inventories. Each face lists L1/L2/L3 using the layer names in its own geometry or readout. The simulation map also defines 60 aliases; the prototype map omits `Top_L1_N4` and retains two inactive Right-L1 positions.

The selected revision contains 38 bars of 150 cm and 21 bars of 80 cm, all with a 20×5 cm^2 cross-section. Its Top layers have lengths 80/150/150 cm, with long bars in every Bottom, Left, and Right layer and short bars in every Front and Back layer. The September revision contains 34 long and 26 short bars, with Top lengths 150/150/80 cm and Bottom lengths 80/150/150 cm in its own GDML layer order. Both revisions use 80 cm short modules, compared with the prototype's 65 cm bars.

The 60-alias simulation map and 59-position paper readout are recorded in `veto_channel_maps.json` under `code/veto-analysis`. The alias-map size alone does not determine which transport geometry was used by a production. The prototype readout was extracted as `iaxoD0Readout-paper.root` from the readout embedded in the original paper-analysis file; its extraction and the two broken-channel identifiers are documented in `agents.md`. The level-3 tensor uses the intersection of the two alias sets, calculated by `veto_level3_data.py` under `code/veto-analysis/level3`. This gives 59 coordinates by removing `Top_L1_N4` from the simulated peak representation after transport and reconstruction. It leaves both inactive prototype aliases in the tensor because they remain defined in the experimental readout. Their simulated activity can therefore distinguish the two input populations. This feature mask neither removes

scintillator or cadmium from the transport geometry nor implements the prototype's complete 57-channel operating state.

A.3 Detailed veto-physics validation

The veto chapter summarizes the nuclear-data and detector-response validation. The comparisons below separate nuclear-data checks, acquisition effects, and selection controls. They test selected transport inputs and quantify the response variation among the configurations examined. Their scope is narrower than a complete detector-response validation or a measurement of the BabyIAXO background level.

A.3.1 High-precision neutron data

Natural-lead reference curves were constructed from ENDF/B-VIII.0 evaluations by weighting ^{204}Pb, ^{206}Pb, ^{207}Pb, and ^{208}Pb by their natural abundances [2]. The corresponding `Geant4` points were extracted from the G4NDL4.6 high-precision tables used by `Geant4` 11.0.3 [3, 4]. Elastic, inelastic, capture, and $(n, 2n)$ channels were compared with the same isotope weights.

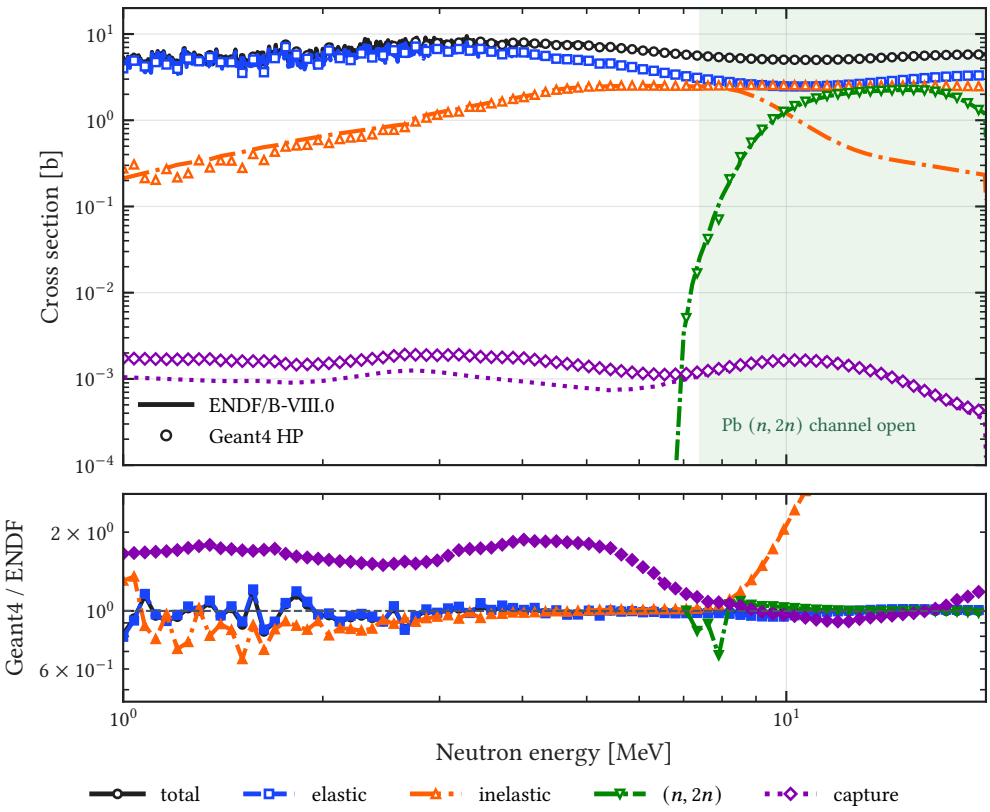


Figure A.3: Natural-lead neutron cross sections from ENDF/B-VIII.0 and the G4NDL4.6 high-precision tables used by `Geant4` 11.0.3. The comparison isolates the transport input from detector-geometry effects.

Analogous comparisons for PVT-like BC408 and natural cadmium are shown in Figures A.1 and A.2. Those comparisons check the evaluated elastic and capture inputs relevant to moderation in scintillator and capture in the cadmium sheets. Constituent-weighted cross sections do not test the thermal-scattering law $S(\alpha, \beta)$, which describes energy and momentum transfer to bound atoms. These comparisons therefore do not determine the uncertainty of the moderation-time distribution or its acceptance inside the acquisition window.

A.3.2 Secondary-neutron production in lead

Interaction probabilities do not determine the energy and angle of the emitted secondary neutrons. A thin-target benchmark was therefore performed against the EXFOR natural-lead (n, xn) double-differential data of Takahashi *et al.* at 14.1 MeV [5, 6]. The reported integral uses $15^\circ < \theta < 165^\circ$ and $0.4 < E_n < 16$ MeV; the experimental spectra are tabulated at 20°–160° angle centers within those bin edges. Each model used 2.0×10^6 incident neutrons

on a 1 mm natural-lead target.

Model	Integrated production cross section [b]	Model/EXFOR
EXFOR Takahashi *et al.*	5.875	1.000
QGSP_BIC_HP	5.358	0.912
QGSP_BERT_HP	5.404	0.920
FTFP_BERT_HP	5.392	0.918
Shielding	5.381	0.916

Table A.2: Natural-lead outgoing-neutron benchmark integrated over $15° < \theta < 165°$ and $0.4 < E_n < 16$ MeV. The four high-precision reference lists are 8–9% below the EXFOR integral and differ from one another by less than 1% in this low-energy benchmark.

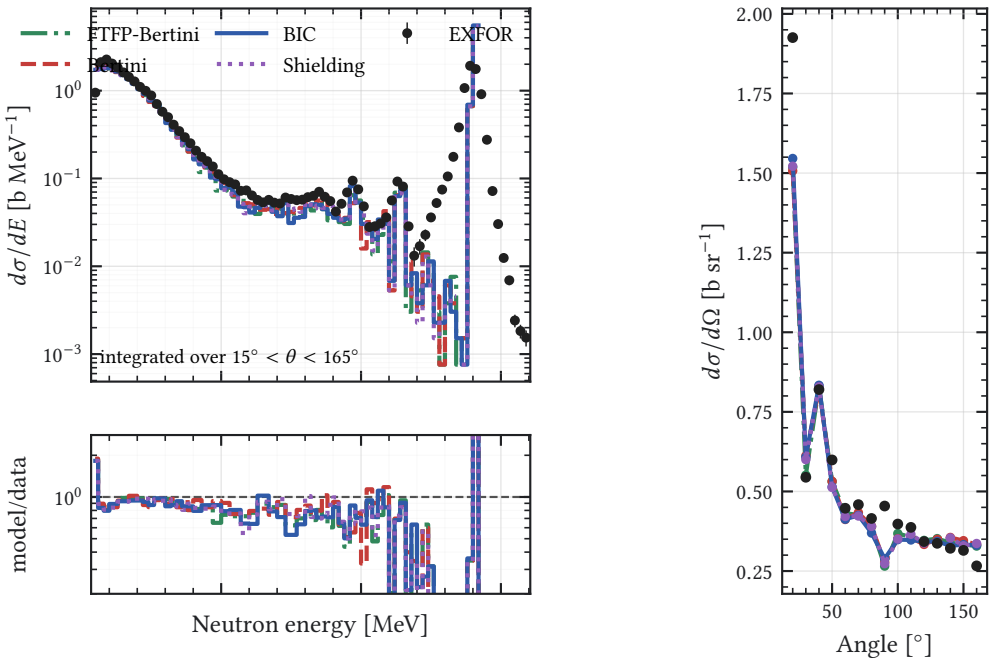


Figure A.4: Outgoing-neutron energy and angle distributions from the natural-lead thin-target benchmark. Below 20 MeV, the tested reference lists are similar because the final state is controlled primarily by the high-precision neutron data.

The comparison supports the low-energy shower interpretation but does not validate the full GeV-scale HENSA response. At higher energy, cascade-model differences must be assessed in the detector geometry and treated as a model envelope.

A.3.3 Cadmium capture-gamma response

A 1 mm natural-cadmium sheet is effectively opaque to thermal neutrons but is not a fast-neutron shield. Its role is to capture neutrons after moderation and emit a gamma cascade close to active scintillator.

Before neutron-process construction, the `REST-for-Physics` interface passes the strict-isotope and `PhotonEvaporation` options to `G4ParticleHPManager`. The default and strict-`ParticleHP` variants give 3.75 and 3.78 photons per cadmium capture and 8.69 and 8.74 MeV of emitted gamma energy per capture. The `PhotonEvaporation` variant gives 6.73 photons and 9.01 MeV per capture, while suppressing the overproduced 9.043 MeV line.

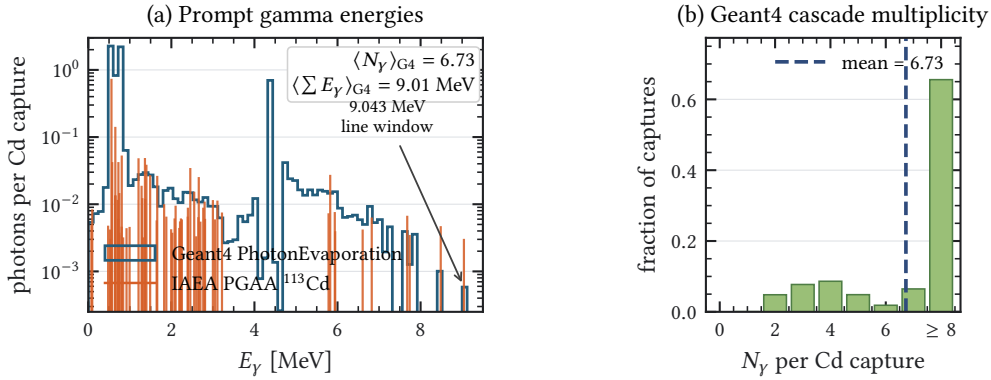


Figure A.5: Evaluated ^{113}Cd prompt-gamma lines compared with the `PhotonEvaporation` cascade in the HENSA three-layer geometry [7]. The histogram is normalized per simulated cadmium capture.

Line window	IAEA yield	Simulated yield	Simulation/IAEA
0.558 32 MeV	0.7375	0.9019	1.22
0.651 19 MeV	0.1420	0.0394	0.28
0.805 85 MeV	0.0531	0.0156	0.29
8.484 60 MeV	0.00472	0.00101	0.21
9.042 90 MeV	0.00305	0.00058	0.19

Table A.3: Selected ^{113}Cd prompt-gamma line-window yields for `PhotonEvaporation`. Individual line yields differ from the evaluated values by factors ranging from 0.19 to 1.22, including discrepancies of about a factor of five. The model is therefore used only to test aggregate cascade-response sensitivity, not for line-spectroscopy predictions.

The corresponding detector-response replay shows that the probability of reconstructing any veto peak varies less between models than rejection at fixed energy thresholds. The

reconstructed peak-energy proxy is expressed in electron-equivalent megaelectronvolts (MeV_{ee}); it is distinct from the emitted cascade energy.

Peak-level criterion	default `ParticleHP`	strict `ParticleHP`	`Photon Evaporation`
Any reconstructed veto peak	0.719 ± 0.009	0.742 ± 0.010	0.720 ± 0.012
$E_{\mathrm{VETO}}^{\mathrm{peaks}} > 4\ \mathrm{MeV_{ee}}$	0.572 ± 0.010	0.596 ± 0.011	0.523 ± 0.013
$E_{\mathrm{VETO}}^{\mathrm{peaks}} > 10\ \mathrm{MeV_{ee}}$	0.375 ± 0.010	0.383 ± 0.011	0.314 ± 0.012

Table A.4: Conditional operating points after replaying the recovered HENSA model-comparison files through the common detector-response and peak-finding chain. The uncertainties shown are statistical standard errors for the finite comparison samples.

The observed spread is an envelope of the tested cascade models, not a confidence interval on the true response. The stability of the any-peak probability does not establish the stability of an energy-sensitive classifier.

A.4 Tracking-orientation systematic

BabyIAXO changes inclination during solar tracking. To test the resulting change in directional acceptance, the three-layer detector geometry was held fixed while the incident source direction was rotated over −25° to +25°. The scan used the same source and reconstruction settings at every point.

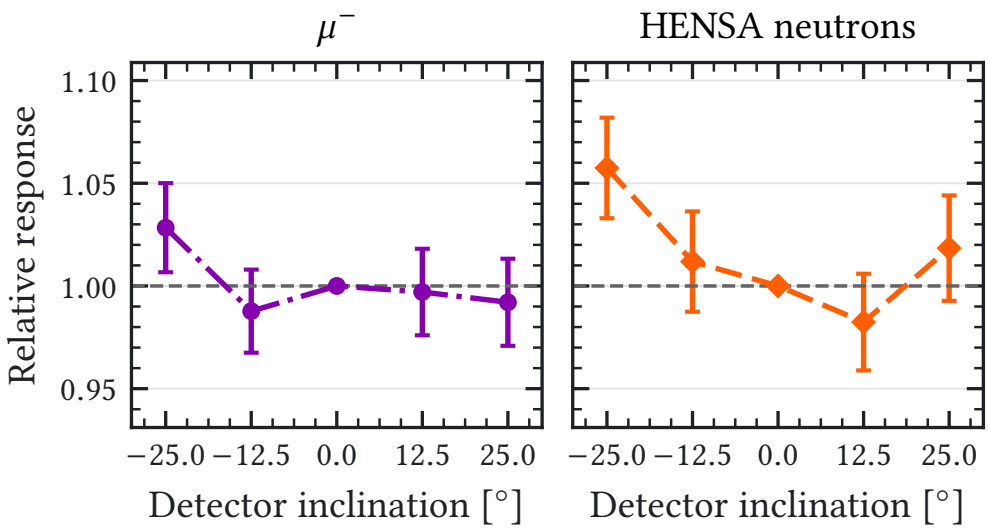


Figure A.6: Detector-inclination scan for muons and HENSA-driven neutrons. The response is the number of processed analysis entries per simulated primary, normalized independently to the horizontal geometry for each source. Error bars are one-standard-deviation counting uncertainties propagated through the ratio; the horizontal point defines the normalization and is fixed to unity.

The neutron response ranges from 0.982 ± 0.024 to 1.057 ± 0.024, and the muon response from 0.988 ± 0.020 to 1.028 ± 0.022, relative to the horizontal configuration. No monotonic dependence is resolved within this scan. The horizontal geometry is therefore an adequate reference for the tested source model, although the corrected source-angle convention must be propagated before transferring this conclusion to the final tracking response.

A.5 Supplementary waveform and selection diagnostics

The main chapter retains the simulation-derived capture-retention plot and a compact historical classifier comparison. This section records the complementary timing and accidental-selection diagnostics.

The capture-history input contains 57,799 records in cadmium-sheet regions. The retention curve is conditional on the 55,581 records with $0 < \Delta t_{\mathrm{gas}} < 2000\ \mu$s; this requirement excludes 2217 pre-trigger records and one record at or beyond 2 ms. Here Δt_{gas} is measured relative to the first simulated gas deposit and is therefore a trigger proxy rather than a reconstructed acquisition time. The median positive delay is 25.6 µs. The cumulative 69.7%, 79.7%, and 88.8% values at 50, 70, and 100 µs use those 55,581 records as their denominator. They integrate from zero delay to the stated upper boundary and must not be interpreted as the acceptance of the reconstructed 10–50 µs delayed interval. Specifically, 28.3% of the

physical records occur at $0 < \Delta t_{\mathrm{gas}} \leq 10\ \mu$s, whereas 41.4% fall in $10 < \Delta t_{\mathrm{gas}} \leq 50\ \mu$s. The truncated-exponential maximum-likelihood fit uses 39,532 records in $10 \leq \Delta t_{\mathrm{gas}} < 300\ \mu$s and gives $\tau = 46.9\ \mu$s. An independent bootstrap that resamples the 2659 capture-bearing analysis entries, keeping captures from each entry together, gives a central 68% statistical interval of 46.67–47.17 µs, consistent with the profile-likelihood result. This check accounts for correlations within an analysis entry; it does not include transport-model uncertainties or establish independence of entries sharing an unavailable primary-history identifier.

An event-mixing study time-scrambles simulated neutron veto trains relative to the Micromegas trigger. The quantitative comparison is summarized in Chapter 4; the figure below preserves the full selection trade-off. Requiring late activity in several veto groups reduces acceptance of the time-scrambled control at the cost of lower neutron acceptance.

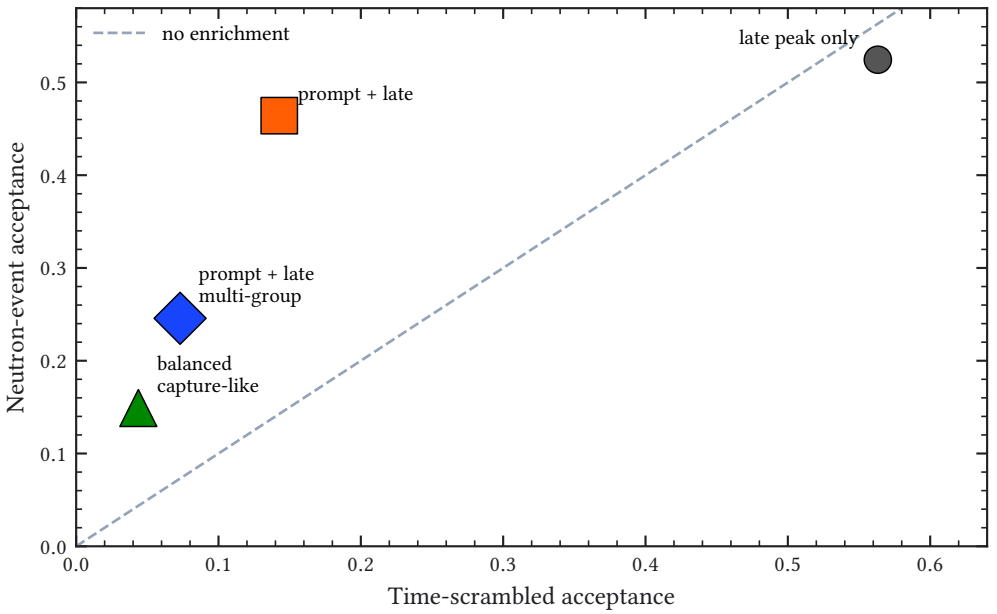


Figure A.7: Illustrative event-mixed capture-like selections using time-scrambled neutron peak trains. A late peak alone is not discriminating; prompt context and segmentation reduce acceptance of this constructed null population.

The preliminary design-study selector is `aggregate-veto-hgb-20260510`. Its five-observable aggregate comparison showed that waveform multiplicity and signal quality outperform a visible-energy threshold within the historical overlaid sample. An event passes this historical overlaid diagnostic when the neutron-like score is below 0.518913730, the threshold calibrated to nominal 90% acceptance of calibration-noise events. The comparison used one random 65/35 split of events into training and test samples, with 2555 held-out calibration-noise control events and 2555 held-out HENSA+noise events. At the nominal 90% control-acceptance point, 1950 of 2555 HENSA+noise events were rejected, giving 76.3%

with an exact two-sided 90% Clopper–Pearson interval of 74.9%–77.7%. At the nominal 95% point, 1744 of 2555 were rejected, giving 68.3% with a corresponding interval of 66.7%–69.8%. Because the threshold was selected from the same empirical receiver operating characteristic curve and the split was not group-disjoint, these values are retained only as the development benchmark that motivated the final hierarchy.

A.5.1 Run-disjoint classifier hierarchy

The historical classifier family is identified by `veto-bdt-hierarchy-20260829-r1`. It uses 139,454 calibration-control events and 20,274 aligned historical cosmic-neutron events with empirical calibration activity overlaid. Runs 1335, 1339, and 1342 form the training partition, run 1345 fixes the nominal 90% and 95% thresholds, and run 1347 is the held-out test partition. The sampled calibration event determines the partition of each overlaid neutron event. This construction prevents the same experimental noise run from entering model training and final evaluation.

The simulation and prototype readouts require separate electronic-channel maps. The maps contain 60 and 59 semantic panel aliases, respectively, with two inactive positions included in the prototype map as described in Section A.2. All 1,193,960 calibration peaks and all simulated and overlaid peaks used by the hierarchy have known channel mappings. Each mapped panel alias is identified by its face, layer, and panel index.

Two prototype-inactive positions, `Right_L1_N2` and `Right_L1_N3`, retain simulated activity. The 59-coordinate neural input intersects simulated and prototype panel names, excluding `Top_L1_N4` but retaining both inactive positions after transport; no scintillator is removed from the material geometry. Their occupancy can distinguish data from simulation, creating possible bias without demonstrating that the neural model exploited it. Symmetrically removing both positions from the adopted frozen prompt-plus-residual-BDT rule preserved nominal neutron and electron survivors; one neutron decision changed across 200 activity-overlay realizations. This archived-sample check neither validates the historical neural model nor resolves the acquisition-timing mismatch.

In the historical feature construction, `REST-for-Physics` peak times were converted with Eq. A.1, rounded to the nearest waveform bin to remove floating-point boundary residues, and clipped to the digitized range 0–511. The boundary correction below removes peaks outside that range instead of assigning them to an edge bin. The five disjoint windows are pre-control 0–189, prompt 190–210, early post-trigger 211–239, capture 240–280, and late-tail 281–511. These are legacy bin categories, distinct from the later capture-specific −10–5 µs and 10–50 µs intervals. Under Eq. 4.12, the broad prompt and capture categories span −8.2–−4.2 µs and 1.8–9.8 µs, respectively. Their historical names do not establish a hardware-trigger calibration or identify the adopted prompt window.

The level-1 feature vector contains total reconstructed veto-energy proxy, peak count, number of distinct mapped panel aliases, active face count, active layer count, and maximum

peak amplitude. The full level-2 vector contains 164 observables after adding per-window count and amplitude summaries, face and layer occupancies and entropies, opposite-face activity, and position–time cells. The reduced variants isolate broad timing summaries, capture-specific features, or position summaries; the complete feature counts and performance are retained in Table A.5. Both models use a `HistGradientBoostingClassifier` with 220 boosting iterations, learning rate 0.045, at most 12 leaves, minimum leaf size 30, L_2 regularization 0.10, balanced class weights, and fixed random seed 20260829. No Micromegas observable is used.

At the nominal 90% operating point, level 1 accepts 24,798 of 27,529 held-out calibration events and rejects 3377 of 4035 held-out neutron events. The full level-2 model accepts 24,851 calibration events and rejects 3396 neutron events. The paired 3000-replicate bootstrap gives a level-2 minus level-1 rejection difference of 0.47 percentage points, with a central 90% interval of 0.07–0.82 percentage points; the AUC difference is consistent with zero. Removing total reconstructed energy from the full model reduces the rejection to 3190 of 4035, or 79.1%, while retaining 90.1% held-out control acceptance. The reduced model retains discriminating information within the overlaid sample, but its remaining amplitude and multiplicity inputs prevent this comparison from isolating timing or segmentation.

Historical aligned response; nominal 90% control acceptance.

Classifier	Inputs	Held-out AUC	Neutron rejection [%]
Level 1: aggregate	6	0.911	$83.7^{+0.9}_{-1.0}$
Level 2: aggregate + time	43	0.913	$84.0^{+0.9}_{-1.0}$
Level 2: aggregate + capture delay	24	0.913	$84.2^{+0.9}_{-1.0}$
Level 2: aggregate + position	49	0.912	$84.3^{+0.9}_{-1.0}$
Level 2: full spatiotemporal	164	0.911	$84.2^{+0.9}_{-1.0}$
Level 3: neural hybrid	tensor + 6	0.913	$84.1^{+0.9}_{-1.0}$
Domain-sensitivity test			
Full model without total energy	163	0.896	$79.1^{+1.1}_{-1.1}$

Table A.5: Complete historical run-disjoint classifier comparison with the original aligned peak response, supplementing Table 4.9. Uncertainties are exact two-sided 90% Clopper–Pearson intervals on 4 035 held-out neutron events. The final row removes the total-energy proxy as a domain-sensitivity test. The effect of rejecting out-of-record peaks is reported separately in Appendix A.5.1.

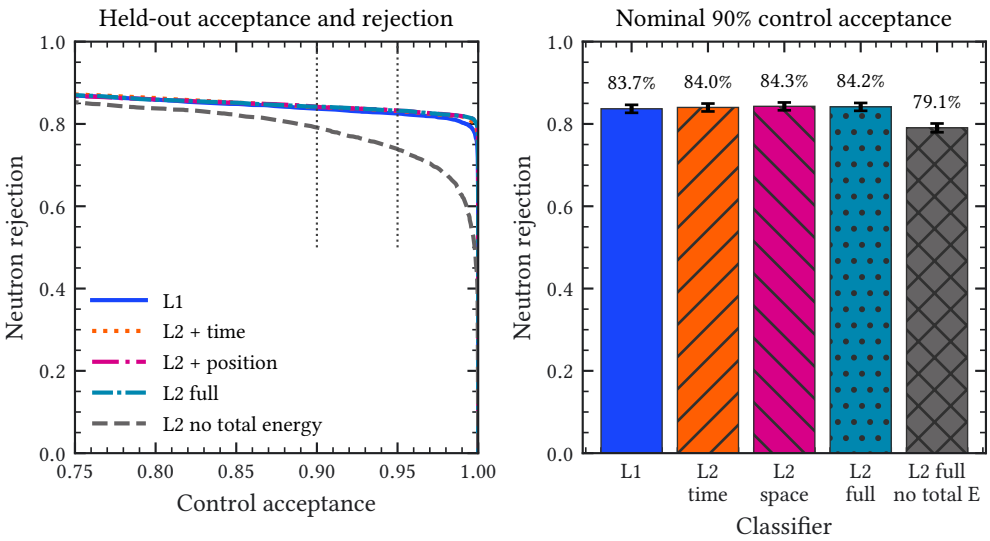


Figure A.8: Run-disjoint comparison of the aggregate and engineered spatiotemporal veto classifiers. The left panel gives the held-out acceptance–rejection curves; the right panel gives the operating point fixed at nominal 90% calibration-control acceptance on an independent run. Error bars are exact 90% confidence intervals. The dashed curve omits the total-energy scalar while retaining peak amplitudes, multiplicities, timing, and segmentation.

Across the five calibration runs, acceptance spans 88.2%–90.9% for level 1 and 84.4%–90.5% for the full level-2 model. Permutation tests identify total energy as the dominant historical input. Historical performance intervals and paired-bootstrap differences condition on the trained models, frozen thresholds, and input response; they do not include the unresolved response mismatch.

Recorded-window boundary check

The aligned neutron input contains 385 peaks outside bins 0–511, distributed over 153 of the 20,274 events; 34 affected events belong to the 4035-event held-out partition. Clipping these peaks to bin 0 creates recorded activity where the model should instead reject an unrecorded peak. The revised feature constructors round to a sample and reject out-of-range or non-finite times before computing peak counts, amplitudes, and spatial or timing features. Historical model files and thresholds are retained, allowing the effect of this correction to be compared without retraining.

Frozen model	Nominal acceptance	Historical tags	Boundary-check tags
Level 1	90%	3 377	3 377
Level 1	95%	3 328	3 328
Level 2 without total energy	90%	3 190	3 170
Level 2 without total energy	95%	2 985	2 962

Table A.6: Frozen-model sensitivity to rejecting out-of-window peaks in the same 4 035 held-out neutron events. Nominal acceptance refers to the historical calibration-control operating point. The total-energy scalar remains unchanged because the flat input does not retain calibrated energy per peak; the level-1 rows consequently test only the remaining features.

The full level-2 model also retains its historical tag count in this limited comparison, whereas the model without total energy loses 20 and 23 tags at the two operating points. These changes quantify the local feature defect, not the full acquisition-response uncertainty. The empirical bin alignment, absence of simulated peaks above $t_{\mathrm{rel}} = 41.6\,\mu\mathrm{s}$, and unavailable calibrated energy per peak require a waveform replay with a common trigger convention before producing replacement efficiency estimates. The revised training and dataset-building paths reject an affected sample when they would combine window-corrected peaks with the old unwindowed energy scalar.

Capture-delay controls

The capture-delay comparison uses the same run partitions as the classifier hierarchy. Its prompt and capture-like intervals are −10–5 µs and 10–50 µs in the aligned coordinate of Eq. 4.12, not the broad hierarchy's stored-bin categories. The 18-variable signature contains seven timing features and eleven contextual features. The timing features are prompt-presence, delayed-presence, and coincidence indicators; first delayed time; last-prompt-to-first-delayed gap; and amplitude-weighted delayed-time mean and width. Missing times are represented by zero together with the presence indicators. The contextual inputs describe prompt and delayed counts and amplitudes, panel/face/layer occupancies, delayed fractions, and panels active only after the prompt interval. TPC quantities, truth-capture labels, run identifiers, and donor indices are excluded from classifier inputs. Figure A.9 separates truth interpretation from reconstructed observables and the historical level-1 benchmark.

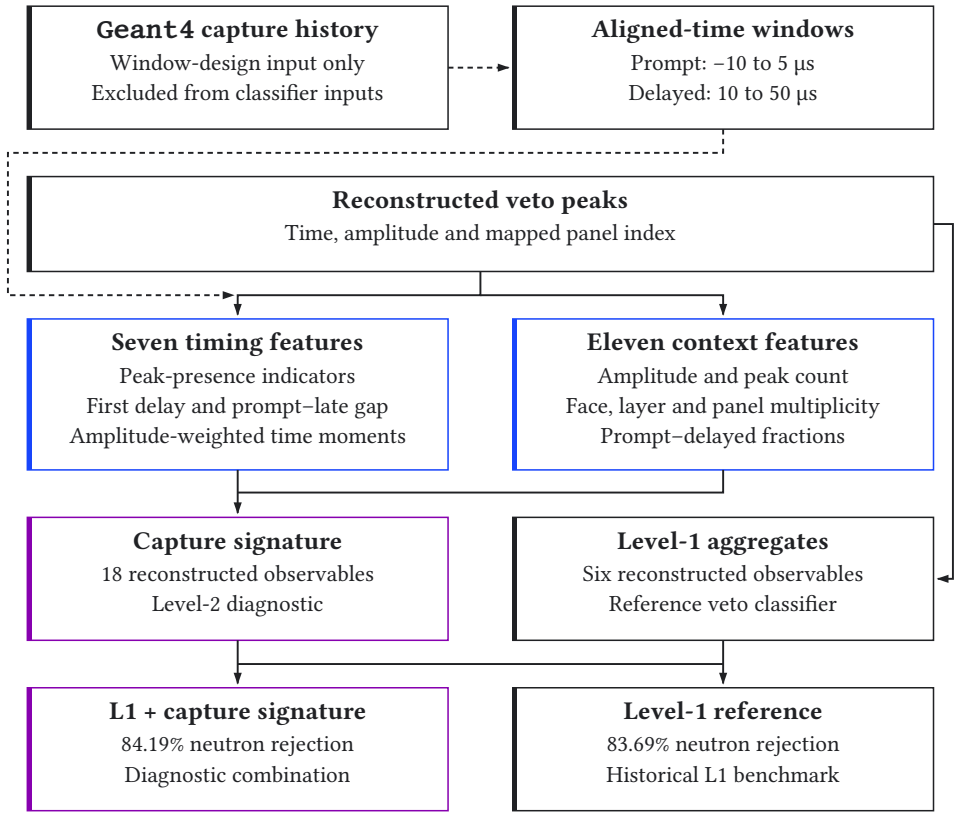


Figure A.9: Role of capture-delay information in the veto analysis. Geant4 history information is used only to define and interpret the physical timing intervals. The classifiers receive reconstructed veto observables: the six level-1 aggregates define the historical benchmark, while the capture-specific diagnostic combines aligned peak timing with delayed peak amplitude and detector segmentation.

The event in Fig. 4.33 is selected as the closest robust multivariate representative to the median of its predeclared prompt-plus-delayed category. It illustrates the mapped panel–time structure rather than defining a population frequency.

At nominal 90% control acceptance, the timing-feature model rejects 2011 of 4035 held-out neutrons, the full signature rejects 2991, and level 1 plus the signature rejects 3397. Their measured control acceptances are 86.81%, 89.84%, and 90.38%, respectively. The combined model improves rejection relative to level 1 by 0.50 percentage points, with a central 90% paired-bootstrap interval of 0.20–0.82 points; its AUC difference is compatible with zero. Its minimum acceptance across the five control runs is 86.33%, below the 88.20% minimum of level 1.

Matching the saved scores event by event resolves the neutron comparison into 3357 shared tags, 40 new tags, 20 lost tags, and 618 entries accepted by both models (Table A.7). The newly tagged entries have more delayed activity than the lost-tag group, but these outcome-selected cohorts do not isolate the contribution of an individual feature or identify capture products. The reconstructed activity includes the calibration overlay. Among the 27 529 held-out calibration controls, 818 acquire a tag and 901 lose one, increasing aggregate retention from 24 798/27 529 to 24 881/27 529. This aggregate change does not remove the run dependence described above.

Neutron cohort	Events	Delayed peaks	Delayed amplitude	New delayed panels
Tagged by both	3 357	3	994	2
Newly tagged by L1+S	40	3	688.5	2.5
Level-1 tags lost with L1+S	20	1	86.5	1
Accepted by both	618	1	179.5	1

Table A.7: Historical neutron tags with and without the capture signature at the frozen nominal 90% control-acceptance thresholds. The last three columns are cohort medians, including zero delayed activity; amplitude is in stored analysis units. New delayed panels are active in the 10–50 µs interval and absent from the −10–5 µs prompt interval.

The event-mixing control applies 250 circular shifts to each of 879 neutron peak trains within −60–50 µs. The complete train is shifted together, preserving its amplitudes and spatial structure while removing alignment with the trigger. Central 90% ranges of mixed-control acceptance are 54.2–58.9% for a late peak, 12.5–16.2% for prompt plus late activity, 6.0–8.5% with two delayed groups, and 3.4–5.5% with the delayed-energy requirement. These are properties of a constructed control, rather than measured accidental-coincidence probabilities. Table A.8 retains the exact mean acceptances and their ratios.

Delayed peaks: 10–50 µs.

Observable requirement	Neutron acceptance [%]	Time-scrambled acceptance [%]	Ratio
At least one delayed peak	52.4	56.3	0.93
Prompt activity followed by a delayed peak	46.4	14.3	3.25
Prompt activity and delayed peaks in at least two veto groups	24.6	7.3	3.36
Prompt activity and delayed peaks in at least two veto groups, with summed delayed peak-energy proxy of at least 10^4 analysis units	14.9	4.4	3.41

Table A.8: Observable capture-delay signatures in 879 simulated neutron events. The time-scrambled column is the mean over 250 circularly shifted realizations of neutron-derived peak trains. The ratio divides neutron acceptance by this mean; it is not a measured background rejection or neutron purity. A late peak alone is not neutron-specific; the discriminating information is the trigger-correlated prompt–delayed sequence supplemented by segmentation and the summed delayed veto peak-energy proxy.

The separate capture-history comparison contains 2706 selected entries, including 2659 with cadmium captures. Of these capture-bearing entries, 644 have no reconstructed veto peak (24.2%); the same events are 23.8% of the complete selected sample. This conditional sample is distinct from both the 879-event mixing study and the 4035-event classifier test. The capture CDF and its fit are reported above and in Fig. 4.18; the main-text reconstructed-signature summary does not superimpose hardware or aligned-window boundaries on gas-relative truth times.

Level-3 neural-model controls

The level-3 neural comparison represents each event by a $59 \times 512 \times 2$ panel–time tensor containing transformed summed amplitude and occupancy. The 59 coordinates are the common readout aliases described in Section A.2, rather than a reduction of the physical transport geometry to 59 panels. The simulation-only panel is masked; run, raw channel, and overlay identifiers, TPC quantities, and `Geant4` truth are excluded. A shared temporal convolution encodes each panel; two graph blocks separately connect same-face neighbors, adjacent layers, and opposite faces before attentive pooling. Four predeclared variants compare a scalar multilayer perceptron, tensor-only temporal convolution, tensor-plus-graph model, and a neural hybrid with the six level-1 aggregates. The neural hybrid contains 53 330 trainable parameters and was trained for 20 epochs with the same run-disjoint partition. This neural hybrid is distinct from the adopted prompt-plus-residual-BDT rule at nominal 88% residual control acceptance.

At nominal 90% control acceptance, neutron rejection is 83.82% for the scalar network,

71.15% for temporal convolution, 72.64% with graph blocks, and 84.11% for the neural hybrid. The neural hybrid changes rejection relative to level 1 by 0.42 percentage points [0.10, 0.74] under the central 90% paired bootstrap, while its AUC change is compatible with zero and its five-run acceptance spans 83.38–90.61%. The minimum occurs in run 1339 and fails the predeclared run-stability criterion; it is below the full level-2 and level-1 minima of 84.4% and 88.2%. At the frozen threshold, shifts up to five bins and 10% amplitude rescaling change rejection by less than one percentage point. Masking one representative panel per face–layer cell changes rejection by at most 0.37 percentage points and can reduce control acceptance by 1.25 percentage points. Figure A.10 retains these performance and perturbation comparisons; its perturbation envelopes are not confidence intervals.

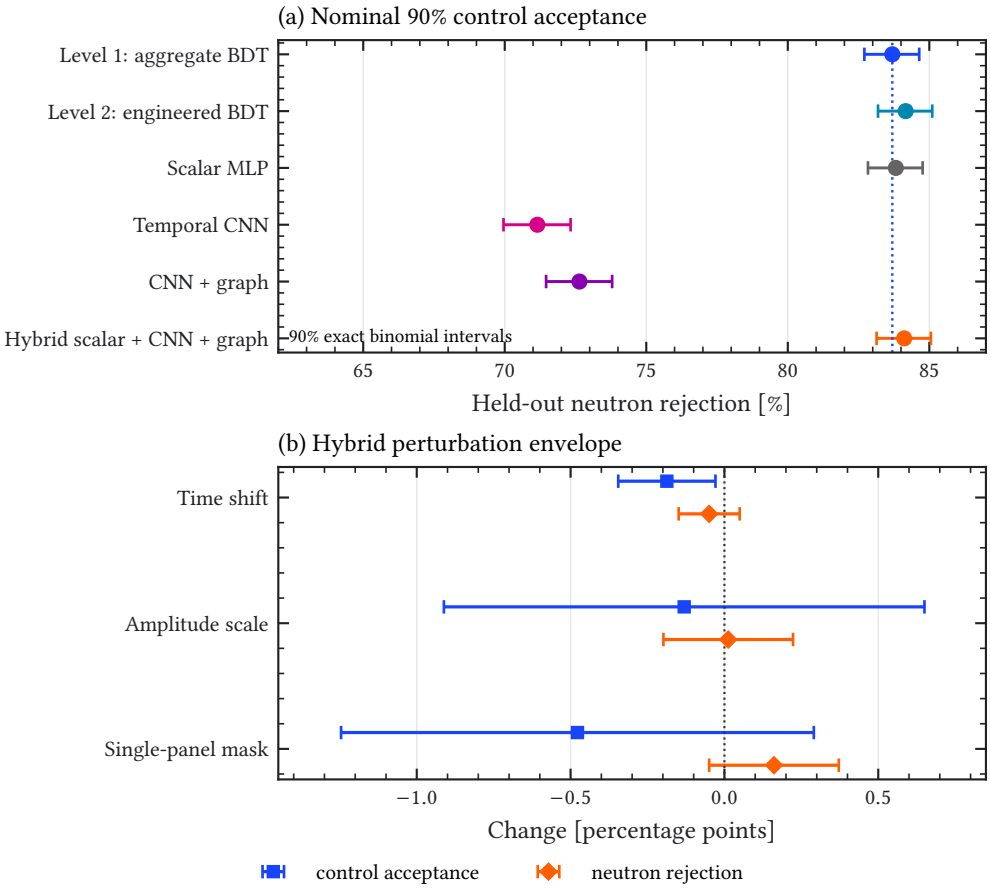


Figure A.10: Run-disjoint level-3 comparison at the independently frozen nominal 90% control-acceptance threshold (a) and perturbation envelope of the hybrid neural model (b). The tensor-only temporal and graph branches underperform the aggregate models; the hybrid recovers the aggregate performance but does not materially exceed it. Counting intervals are exact two-sided 90% intervals; perturbation intervals span all declared time shifts, amplitude scalings, or representative single-panel masks.

Historical TPC-candidate overlay comparison

The historical candidate comparison applies level 1 to all 249 TPC-baseline neutron candidates in Table B.16, a broader population than the independent TPC-BDT-plus-veto evaluation (Section 4.6.5). Three causal-history activation events are excluded without veto credit before historical classifier scoring; each of the remaining 246 candidates receives 200 independent calibration overlays. Mean historical level-1 rejection of this non-activation candidate population is 91.14%, leaving 21.8 candidates (central 90% overlay range: 19–25); the reproducible first realization leaves 22. This is the historical level-1 classifier result, rather

than rejection by the fixed prompt-amplitude path of the adopted veto. Full level 2 leaves 20.6 on average, 1.2 fewer than level 1. The neural hybrid leaves 21.0 (central 90% overlay range: 18–24), with 20 in the first realization. Its mean exceeds level 2 by 0.4 and is only 0.8 below level 1. Repeated overlays measure sensitivity to the sampled activity; they do not add independent neutron exposure or supply the independently evaluated background prediction.

Current delayed-activation accounting uses the acquisition containing the decay without credit from the production shower (Section 5.2.2). The earlier three-event ancestry exclusion defines neither a complete activation inventory nor an activation-specific upper bound. Detailed architecture, candidate identities, and overlay draws remain in the independent veto-validation document.

A.6 Calibration-controlled late-window population

The exploratory prototype reanalysis selects background events that resemble the historical cosmic-neutron-plus-noise template more closely than calibration or prompt-muon controls. This April 2026 template uses the analytic cosmic-neutron source and predates the HENSA layer-scan production. It is not used as a neutron-fraction measurement. The term *selected control event* below means only that the event passes the frozen score requirement.

A.6.1 Input products and preprocessing

Two experimental background products are used for different diagnostics. The merged paper-analysis file contains 995,271 entries and supports direct vector-branch and channel-correlation studies. The separate flattened score input contains 777,752 events; the score loader reads every entry without an X-ray energy, position, or topology selection. The calibration score input contains 139,454 events. No row-level mapping between these products has been recovered, so the difference cannot be assigned to rejected or incomplete records. Neither denominator is interchangeable with the published 1,305,996-event acquisition or its 257 Micromegas-selected candidates.

The simulation templates contain 21,569 analyzed muon entries and 20,274 analyzed neutron entries after merging the waveform-level production outputs. Calibration-triggered veto peak trains are concatenated with each simulated event, retaining correlations among the measured peaks in a donor event. Because this operation follows peak finding, it does not model merged pulses, baseline changes, threshold migration, or saturation when simulated and measured pulses overlap. `REST-for-Physics` peak times are mapped to the experimental bin convention through

$$t_{\mathrm{bin}} = \frac{t_{\mathrm{REST}}}{0.2\ \mu\mathrm{s}} + 231. \tag{A.1}$$

The prompt alignment selects 21,062/21,569 muon entries in the unoverlaid template and

902,484/995,271 entries in the merged experimental background product. The separate flat score input has a 91.13% prompt-tag fraction.

A.6.2 Score definition

Three time regions are defined in the mapped bin coordinate of Eq. A.1: the prompt region $190 \le t_{\mathrm{bin}} \le 210$, the near-capture region $240 \le t_{\mathrm{bin}} \le 280$, and the inclusive late region $211 \le t_{\mathrm{bin}} \le 511$. The implementation historically calls the near-capture region the "neutron window." The late region overlaps it and is an in-waveform observable; neither region is the long-lived delayed-activation channel defined from event history in Section 5.2.2. The near- and late-window sums are sums of all reconstructed peak amplitudes, in ADC counts; their multiplicities count only peaks above 200 ADC counts. Neither amplitude sum is a calibrated energy, and the native Micromegas observable has not been converted to the publication's keV scale for this comparison. For non-negative amplitude or energy-like inputs, the transformation is $L(y) = \log_{10}[\max(y, 0) + 1]$. The event vector is

$$\vec{x} = \left(L(A_{\mathrm{near}}), \min(N_{\mathrm{near}}, 10), L(A_{\mathrm{late}}), \min(N_{\mathrm{late}}, 20), N_{\mathrm{unique}}, L(E_{\mathrm{TPC}})\right), \tag{A.2}$$

where A_{near} and A_{late} are the amplitude sums, N_{unique} is clipped at 59 channels, and the near/late multiplicities count peaks above 200 ADC counts. The one-dimensional histograms use 53 uniform edges over $[0, 5.2]$, unit-width edges over $[-0.5, 10.5]$, 55 uniform edges over $[0, 5.4]$, unit-width edges over $[-0.5, 20.5]$, unit-width edges over $[-0.5, 60.5]$, and 66 uniform edges over $[0, 6.5]$, in the order of Eq. A.2. Each class likelihood is the product of the six separately normalized one-dimensional histograms after adding 10^{-6} to every bin,

$$p_k(\vec{x}) = \prod_{j=1}^{6} h_{k,j}(x_j), \qquad p_{\mathrm{ref}}(\vec{x}) = \frac{p_{\mathrm{cal}}(\vec{x}) + p_{\mu+\mathrm{noise}}(\vec{x})}{2}. \tag{A.3}$$

The resulting product ignores correlations, including the overlap between the near and late windows, and is used only as an ordering statistic. The score is

$$S_{\mathrm{n}} = \log p_{\mathrm{neutron+noise}}(\vec{x}) - \log p_{\mathrm{ref}}(\vec{x}). \tag{A.4}$$

Prompt-muon events are removed with $N_{\mathrm{prompt}} \ge 2$ for peaks above 200 ADC counts in bins $190 \le t_{\mathrm{bin}} \le 210$. Events with at least 100 veto peaks or 50 unique veto channels are excluded as burst-like. The frozen control threshold is the upper 1% tail of the clean calibration score distribution,

$$S_{\mathrm{n}} \ge 4.668479. \tag{A.5}$$

The displayed threshold is rounded; the replay retains the full-precision 99th calibration percentile. It is not a universal physical threshold. The calibration events used to set it

are also used to fit the calibration histograms; the 1% selected calibration tail is therefore a construction condition rather than an independent validation.

Figure A.11 preserves the historical full-input score distributions. The common-cohort comparison with a veto-only score is reported in Table 4.11 of Chapter 4. The original six-observable score is denoted S_6 in that comparison. For the veto-only score S_5, the Micromegas-energy factor is omitted from each class likelihood before recomputing the reference mixture in Eq. A.3. The remaining five histograms, input events, prompt requirement, and burst exclusions are unchanged. The veto-only score has its own 99th clean-calibration percentile, 5.610744 after rounding, and selects 3154 background and 1041 neutron-plus-noise events. This tests dependence on the Micromegas observable without imposing an energy-matched comparison or a new region of interest. For the calibration-run test in Fig. 4.38, we refitted the calibration histogram for each of the five veto observables with one run omitted and set the threshold from the clean events in the other four runs. The omitted run was then scored with that reference. Muon and neutron histograms were kept fixed, including their original pooled calibration-noise overlays; the test is therefore a check of run dependence rather than a fully independent response validation.

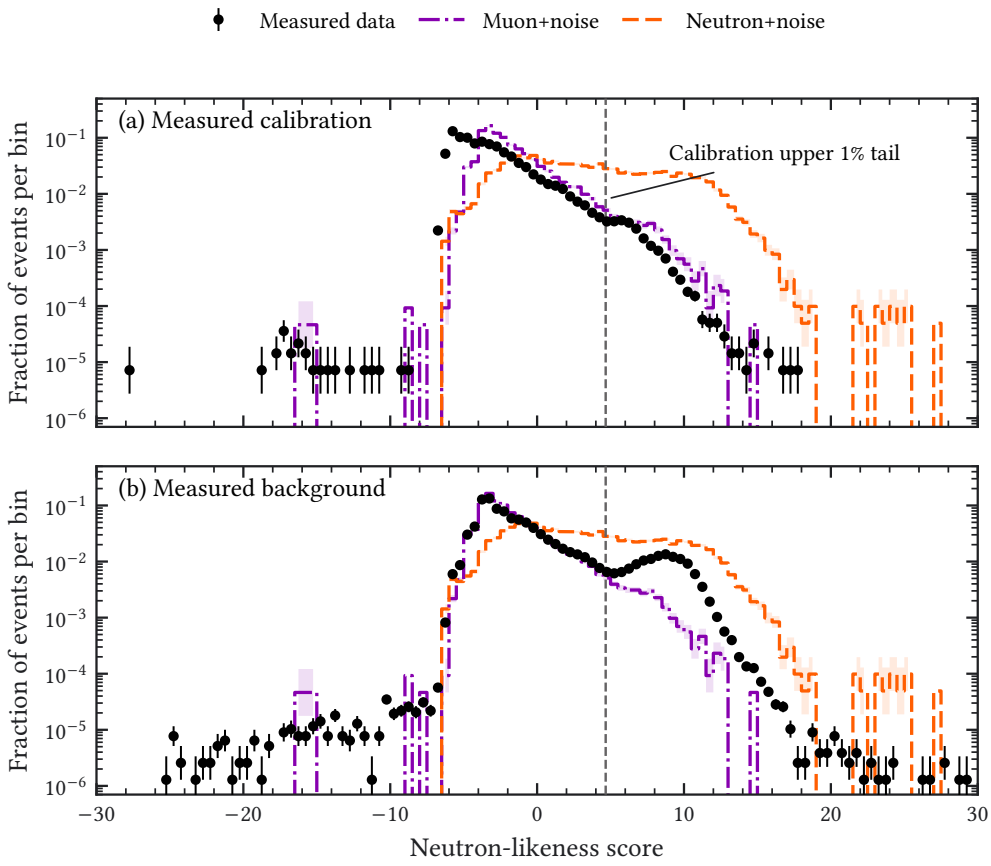


Figure A.11: Late-window score distributions for the full calibration sample (a) and experimental-background sample (b), with the same simulated muon-plus-noise and neutron-plus-noise distributions in both panels. Each distribution is normalized by its event count in the displayed score range, using bins of width 0.5; no events lie outside this range. Black markers and colored bands show central 68.27% Wilson counting intervals. Empty measured bins are omitted on the logarithmic scale. The 99th-percentile calibration threshold is defined after prompt and burst removal and retains the upper 1% tail of that clean calibration population. A large score ranks an event closer to the neutron-plus-noise template but does not establish its physical origin.

A.6.3 Template mismatch and interpretation

The score selects 3877 experimental events, or 0.50% of the full flat input and 5.88% of the clean prompt-suppressed subset. The selection suppresses the prompt-muon population and, by construction, accepts only the upper score tail of the calibration control. It does not establish that the residual background events have a non-accidental origin: the Micromegas energy enters the score, and the control and background samples differ in energy distribution, operating conditions, and acquisition history. The historical neutron template also does not

reproduce the selected population quantitatively.

Median observable	Selected data	Selected neutron+noise
Veto peaks	47	9
Veto peaks above 200 ADC counts	10	5
Unique veto channels	12	8
Unique channels above 200 ADC counts	4	5
Near-capture-window peak count	1	1
Late-window peak count	7	4
Late-window amplitude sum	6 822	2 707
Late-window amplitude sum above 200 ADC counts	2 772	2 241.5
Micromegas energy observable	3 025	10 566
Micromegas hit multiplicity	6	51
xy topology variable	2.31	14.19
z topology variable	2.70	7.59

Table A.9: Median properties of the selected experimental control population and the selected neutron+noise template. Veto peaks and unique channels include all reconstructed peaks; near- and late-window counts require amplitudes above 200 ADC counts. The late-window amplitude sum is in ADC counts, while the Micromegas energy observable retains its native analysis scale. The total veto-energy scalar is omitted because the measured and simulated export products use different definitions for that field.

The selected count remains a control-population observable. The mismatch in Table A.9 and the absence of independently calibrated neutron templates prevent a neutron-fraction measurement.

A.7 Transfer to the published IAXO-D0 candidates

The comparison in Section 4.7.2 preserves the original 56 run, event, and timestamp identities, their Micromegas energies, and the original advanced-veto decisions. We joined 19 events to the archived multi-peak reconstruction and recovered 37 from the original acquisition files. All timestamps agree exactly. Three background controls reproduce peak channel, bin, ADC amplitude, physical time, and calibrated-energy fields exactly; two consecutive calibration records also reproduce the archived vectors with the same processing objects. The saved model reproduces 143 archived reference scores with zero numerical difference.

The waveform chain removes unknown channels and TPC channels 4489 and 4550, then applies the archived TPC and VETO baseline corrections and common-noise corrections in their original order. Veto peaks use the robust full-record baseline, a 50 ADC threshold above it, and a ten-bin peak separation. Only channels present in the frozen experimental map enter the features; the added reconstructed identifiers 4728 and 4643 are excluded to preserve

the training representation. No additional time shift or amplitude rescaling is applied. A channel-indexing defect in the archived common-noise correction can alter VETO samples during the TPC stage. The comparison preserves that behavior; its correction requires a separate reconstruction and classifier assessment.

Additional selection	Rejected	Rejection [%] 90% interval
Published advanced veto	7	12.5 [6.0, 22.2]
Residual BDT, adopted threshold	3	5.4 [1.5, 13.3]
Additional 500 ADC prompt tag + adopted BDT	5	8.9 [3.6, 17.9]
Residual BDT, calibration-adjusted threshold	7	12.5 [6.0, 22.2]

Table A.10: Additional rejection on the same 56 published post-prompt candidates. Intervals are exact two-sided 90% binomial intervals. The last row changes only the score threshold, using separate Xe–Ne calibration runs with a nominal 97% retention target; it does not change the adopted background-model selection.

Evaluation runs	Controls	Adopted BDT retention [%]	Adjusted BDT retention [%]
1696 + 1698	8 344	99.90	97.97
1748	1 000	99.30	94.80
1830	630	99.37	95.40
1859 + 1867	4 464	99.91	98.92

Table A.11: Conditional calibration retention after the original Micromegas, energy-region, and prompt-veto selections. The adopted threshold is 0.9382; the adjusted threshold is 0.7545, fixed using runs 1663 and 1666. Neither threshold uses the displayed evaluation controls for setting its value. The 500 ADC prompt criterion is excluded here to isolate the residual BDT.

For the second and third periods, we also selected 1000 unique physical calibration identities per development run and an independent evaluation sample by a fixed hash of the original identity, before scoring. All 630 eligible identities in run 1830 are retained in that evaluation sample. Repeated historical rows are not counted as additional physical events; their alternative weighting leaves the second-period threshold unchanged. Setting 97% retention in runs 1746 and 1818 gives thresholds 0.7294 and 0.7052, respectively, but only 940/1000 and 591/630 events survive in their reserved controls. The exact 90% intervals are 92.6%–95.2% and 92.0%–95.3%, demonstrating a run dependence beyond the threshold-setting target.

A.8 TPC-selected tracks as an experimental veto control sample

A simulation-derived veto selection must retain its intended behavior when applied to measured signals. In particular, improved simulated neutron rejection is useful only if prompt muon tagging and calibration acceptance remain adequate in experimental data. The study described here selects extended tracks using only the time projection chamber (TPC), evaluates the previously frozen veto classifiers, and investigates alternatives using the discrepancies found in that comparison. The first evaluation and the subsequent exploratory classifier development are distinguished explicitly.

A.8.1 Experimental sample and independent denominator

The source background data contain 995 271 entries from runs 1336, 1340, 1341, 1343, 1346, and 1348. Runs 1336 and 1340 were used for TPC-selection development; runs 1341, 1343, 1346, and 1348 form the evaluation group. An initial extraction requiring at least eight reconstructed hits in each strip projection and $\max(\sigma_X, \sigma_Y) \geq 3\,\mathrm{mm}$ retained 162 273 entries. This extraction accessed only TPC quantities and event identities. The original detector-hit objects were recovered from 725 per-subrun files and matched to the merged analysis using run, subrun, event, subevent, timestamp, and, where repeated identities occurred, exact reconstructed TPC charge. The XZ and YZ hit counts and summed charge were checked against the stored scalar observables. No veto quantity was used to resolve an event identity.

The prototype acquisition description states that the Micromegas electronics trigger readout of the veto channels [1]. This supports an independent TPC-selected denominator, although the online trigger configuration of each individual run was not recovered in this study. Valid readouts containing no reconstructed veto peak remain in the denominator. Missing readouts and nonfinite or unmapped peak records are treated as unavailable responses, rather than as measured zero signals. The broad extraction contains one unavailable veto readout; all 351 candidates in the principal evaluation sample have valid readout. Across the broad extraction, 106 valid zero-peak records were retained.

A.8.2 TPC-only track selection and line fits

The nominal scalar selection requires at least 12 hits in each readout projection, $\max(\sigma_X, \sigma_Y) \geq 4\,\mathrm{mm}$, $\min(\sigma_X, \sigma_Y) \geq 1\,\mathrm{mm}$, and positive finite charge in both projections. No conversion of the TPC charge into an absolute deposited energy is assumed. The track-shape refinement uses a charge-weighted orthogonal line fit independently in XZ and YZ. For hit coordinates $\mathbf{r}_i = (u_i, z_i)$, with $u = X$ or Y, and positive reconstructed charges q_i, the centroid and covariance are

$$\bar{\mathbf{r}} = \frac{\sum_i q_i \mathbf{r}_i}{\sum_i q_i}, \qquad C = \frac{\sum_i q_i (\mathbf{r}_i - \bar{\mathbf{r}})(\mathbf{r}_i - \bar{\mathbf{r}})^{\mathsf{T}}}{\sum_i q_i}. \tag{A.6}$$

The eigenvector of the larger eigenvalue defines the fitted line; the square root of the smaller eigenvalue is the transverse RMS. This construction remains defined for a track parallel to either coordinate axis and avoids choosing one measured coordinate as an error-free regression variable. It is a geometric fit; no fit probability is assigned because per-hit position uncertainties are not modeled here.

The charge-weighted 5–95% projected span must be at least 15 mm in one projection and 8 mm in both. Each transverse RMS must be at most 1.5 mm, and the minor-to-major variance ratio must not exceed 0.08. The largest gap within each central projected span, including the interpolated quantile boundaries, must not exceed 30% of that span. The two central Z intervals must overlap by at least 50% of their union. The main sample excludes events with a flagged saturated TPC channel. This selection yields 468 candidates over all six runs, including 351 in the evaluation runs; 583 additional evaluation candidates pass the same geometry selection criteria but have TPC saturation. These selections favor extended, approximately straight charged-particle tracks and particular detector-crossing geometries. Their muon purity has not been measured.

The scalar loose and tight alternatives use respectively 8 and 20 hits per projection, maximum spatial widths of 3 mm and 5 mm, and minimum widths of 0.5 mm and 1.5 mm. The corresponding straight-track alternatives use maximum-projection spans of 12 mm and 20 mm, transverse RMS limits of 2 mm and 1 mm, variance-ratio limits of 0.15 and 0.04, and gap fractions of 0.4 and 0.2. The minimum span in both projections and the Z-overlap requirement remain unchanged. These alternatives were declared before selected veto responses were opened. A missing-boundary error in the original gap calculation was identified in the TPC-only development gallery and corrected before the first veto evaluation; both selection versions and the final source hash are preserved. Inspection of TPC marginal distributions across the full source and later reuse of evaluation results are nevertheless reasons to describe the overall study as exploratory.

Figure 4.27 in the main text shows a representative selected event and its two orthogonal fits. The event is chosen by minimum robust distance to the median of the selected TPC shape observables, without inspecting its veto response. Only a common offset is subtracted from Z; neither absolute drift position nor the direction of travel is inferred. The two projected fits are not presented as a unique three-dimensional hit association.

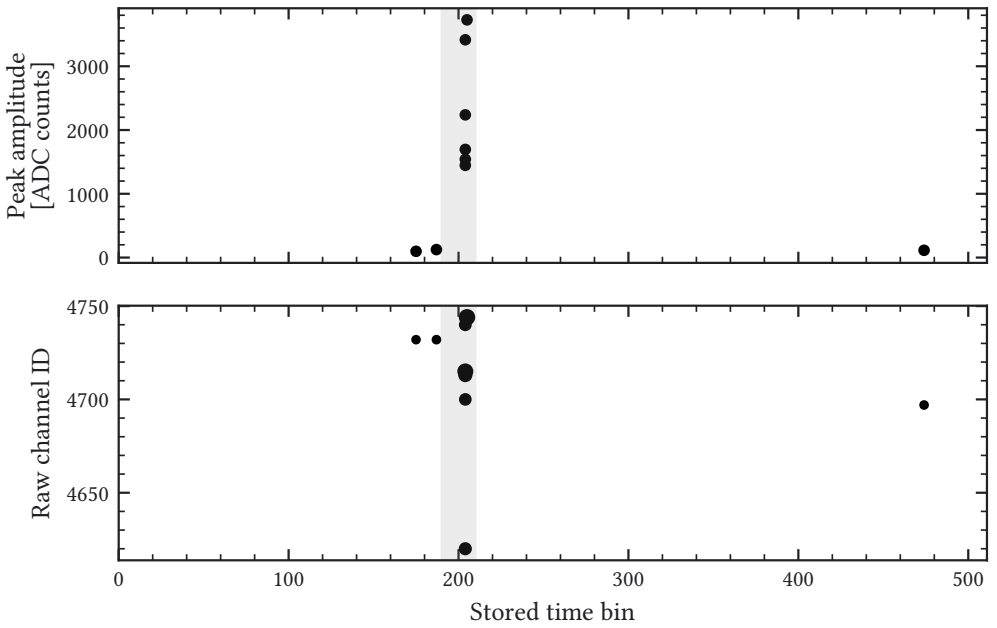


Figure A.12: Stored veto peaks accompanying the TPC-selected event in Figure 4.27. The upper panel gives peak amplitude and the lower panel raw channel identity, with marker area increasing with amplitude. The shaded region is the fixed prompt interval, bins 190–210. These are reconstructed peak records, not digitized waveforms; no amplitude rescaling or additional timing alignment is applied.

A.8.3 First evaluation of the frozen veto selection criteria

The prompt reference rejects an event when the sum of nonnegative veto peak amplitudes in stored bins 190–210 reaches 500 ADC counts. The multiplicity reference requires at least two peaks above 200 ADC counts in the same interval. These are simple prompt-muon-like selections in the present reprocessing; reproducing their exact relation to the published waveform selection would require the original implementation and configuration. The BDT thresholds are the previously saved nominal 90% calibration-acceptance operating points. For the experimental track population, the measured quantity is

$$f_{\mathrm{tag}} = \frac{N(\text{TPC selection} \cap \text{valid veto} \cap \text{veto tag})}{N(\text{TPC selection} \cap \text{valid veto})}. \tag{A.7}$$

It is conditional on the TPC selection and is not an absolute efficiency for all incident muons. Table A.12 gives exact two-sided 90% binomial intervals; these cover counting statistics conditional on the model and selection, not purity, detector modeling, or correlated run systematics. Calibration retention is measured separately on all 27 529 evaluation-control events in run 1347, without the TPC track selection.

Frozen selection	Tagged	Tagging [%] 90% interval	Calibration retained [%]
Prompt amplitude	339/351	96.6 [94.5, 98.0]	87.8
Prompt multiplicity	333/351	94.9 [92.5, 96.7]	88.9
Historical neutron level 1	202/351	57.5 [53.0, 62.0]	90.1
Neutron level 2, amplitude	89/351	25.4 [21.6, 29.5]	90.1
Neutron level 2, counts/timing	132/351	37.6 [33.3, 42.1]	87.5
Muon level 2, amplitude	273/351	77.8 [73.8, 81.4]	89.3
Prompt OR neutron counts/timing	340/351	96.9 [94.9, 98.2]	79.9

Table A.12: Performance with frozen selection criteria on unsaturated TPC-selected experimental tracks. The historical level-1 model retains the mixed-definition energy proxy and is included only as a diagnostic. The last row combines the original fixed selection criteria, without adjusting their joint calibration acceptance.

The prompt selection criterion tags most selected tracks, whereas the simulation-derived neutron BDTs miss a substantial fraction. Of the 339 prompt-tagged candidates, 251 fail the revised neutron level-2 amplitude BDT. Their median prompt amplitude is 11 963 ADC counts, compared with 9676.5 ADC counts for the 88 prompt-tagged candidates that also pass that BDT. The failure therefore cannot be summarized as insufficient prompt signal amplitude. Adding the frozen neutron timing BDT to the prompt selection criterion recovers one additional candidate but lowers calibration retention by approximately eight percentage points. That combination does not demonstrate a favorable trade-off on this population.

The prompt counts by evaluation run are 25/25, 98/101, 107/113, and 109/112 for runs 1341, 1343, 1346, and 1348, respectively. Among the 214 nominal candidates without flagged veto saturation, prompt and neutron level-2 amplitude tagging are 95.3% and 25.2%; among the 137 with flagged veto saturation, they are 98.5% and 25.5%. Excluding the high-multiplicity tail by requiring at most 30 veto peaks leaves 268 candidates, with 95.9% prompt tagging and 21.6% neutron-BDT tagging. These post hoc diagnostics show that neither saturation nor the multiplicity tail alone explains the observed BDT deficit. They are not alternative samples selected to optimize the reported efficiency.

Equally wide off-time windows, bins 40–60, 90–110, and 140–160, exceed the same summed-amplitude threshold for 26/351, 30/351, and 26/351 candidates, respectively. The resulting 7.4–8.5% fractions demonstrate nonzero off-time activity. No accidental subtraction is performed: correlated activity and the timing distribution need not be stationary across the record.

A.8.4 Comparison of experimental and simulated veto signals

A direct response comparison uses the nominal scalar TPC selection in both data and simulated muons. The available flattened simulation does not retain the hit coordinates required for

the straight-track refinement or an equivalent measured saturation flag. Consequently, the common scalar comparison includes all experimental TPC saturation states and uses 53 764 valid experimental events and 2058 simulated muons from the saved evaluation partition. This is a common observable selection, not proof that the geometrical or particle populations are identical. The historical 20k source hit records were not recovered during this analysis, so no matched straight-track simulation efficiency is quoted. One calibration peak train is overlaid on each simulated event, with its donor confined to the same calibration partition. No extra train is added to measured background events.

The median maximum peak amplitude is 3473 ADC counts in data and 2352 ADC counts in simulation; median prompt amplitude sums are 11 445 and 9139 ADC counts, respectively. The 90th percentile of the prompt sum is closer, 16 061 versus 14 883, but agreement of this one quantile does not establish agreement of the distributions. Approximately 49.2% of experimental candidates have a maximum peak amplitude of at least 3500 ADC counts, compared with 18.2% of the simulated muons. The data contain a strong population near the digitizer ceiling, while the simulation has a different high-amplitude shape. A gain or saturation correction cannot be inferred from their ratio without controlling geometry, particle composition, and electronics response.

The multiplicity discrepancy is larger: 23.46% of experimental candidates have more than 30 peaks, compared with 0.097% of the simulated sample (2/2058). The respective 90th-percentile peak counts are 94 and 15. The same experimental tail is visible in the stricter unsaturated track selection, where 83/351 candidates exceed 30 peaks. It is therefore inappropriate to claim that the full measured and simulated veto records match.

This comparison concerns stored peak records, not full waveforms. Empirical peak concatenation cannot reproduce pulse overlap, peak merging, afterpulses, gain drift, baseline changes, or saturation and subsequent peak refinding. The present differences do not identify which of these mechanisms dominates. They motivate separate checks of channel gains and saturation, prompt-time support, and run-dependent non-prompt activity before assigning a detector-response correction.

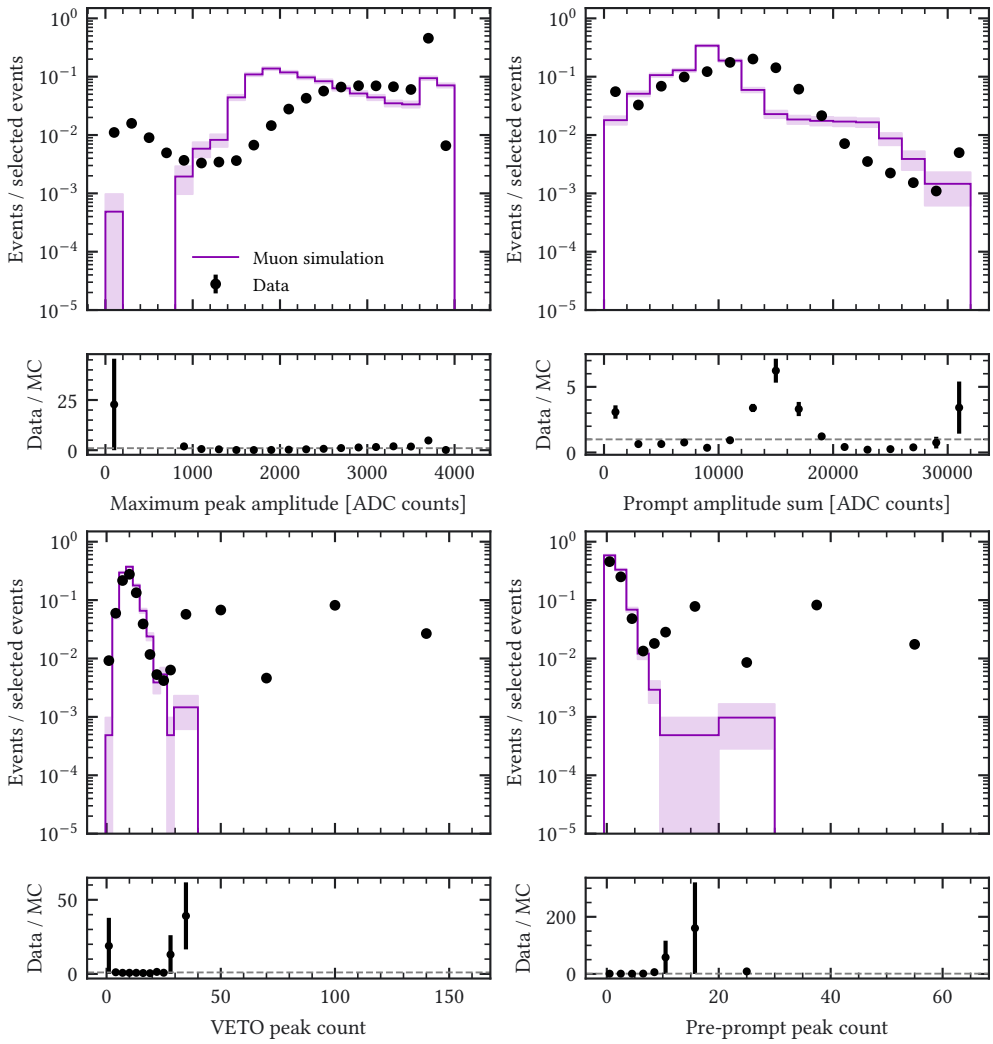


Figure A.13: Veto response for the common scalar TPC selection: experimental candidates (black points) and simulated muons with one overlaid calibration event (purple steps). Each histogram is normalized by its full selected-event count; tails beyond the displayed range are folded into the terminal bins. Bin widths are nonuniform in the multiplicity panels. The uncertainties are counting errors conditional on the normalization, and the simulation bands show finite-sample uncertainty. Ratios omit bins without simulation support. Amplitudes are digitizer observables, without a new physical energy calibration or fitted scale correction.

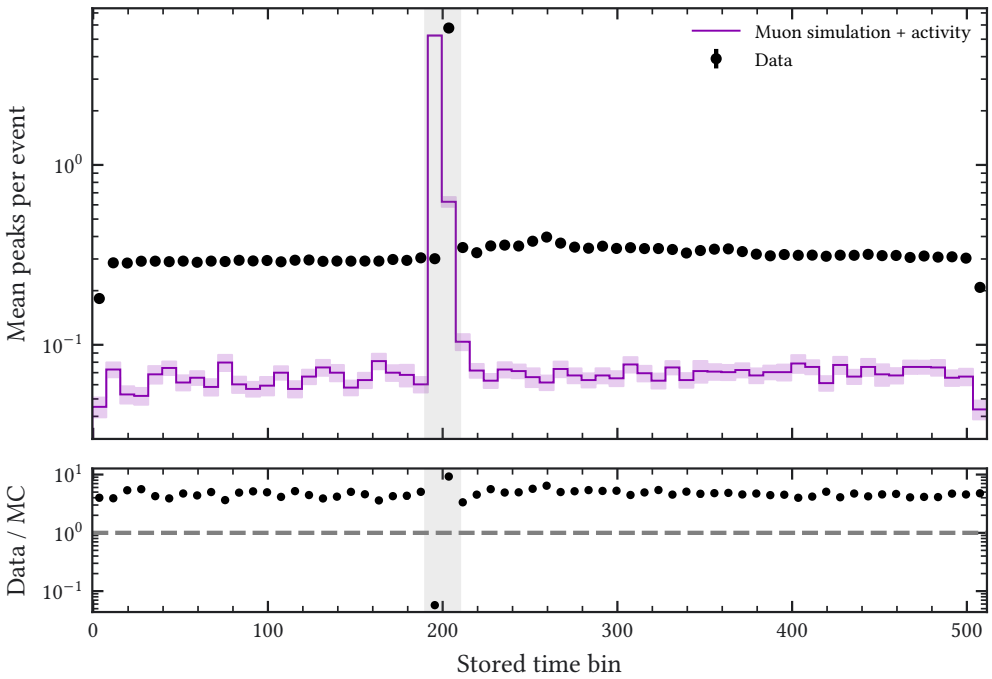


Figure A.14: Mean number of reconstructed veto peaks per event in eight-bin time intervals, for the same common scalar TPC samples as Figure A.13. Uncertainties are standard errors calculated from the event-level bin counts, retaining correlations among peaks within an event. The shaded band identifies the frozen prompt interval. Neither sample is renormalized to the prompt peak, and no additional time shift is fitted. The lower panel gives data divided by simulation where the simulation has nonzero support.

A.8.5 Exploratory classifiers and prompt-preserving hybrids

The first transfer comparison motivated compact neutron-only and mixed neutron–muon BDTs, a mixed monotonic BDT, and a full mixed BDT. They use the existing calibration partition and exclude the inconsistent total-energy scalar. The compact vector contains prompt amplitude and multiplicity, three post-prompt multiplicities, panel, face and layer occupancies, and maximum amplitude. Fits share 220 boosting iterations, learning rate 0.045, at most 12 leaves, minimum leaf size 30, and L_2 regularization 0.1. Training assigns equal total weights to calibration and simulated particles, with equal neutron and muon weights in mixed models. No experimental selected track supplies a training label or threshold. Because the background evaluation runs had already been inspected, these comparisons are exploratory.

The monotonic model constrains the score to increase with prompt amplitude and multiplicity at fixed remaining inputs. Thresholds are set on run 1345; the prompt-reference-matched target is its 88.1914% acceptance of the fixed 500-ADC criterion. Table A.13 gives the achieved acceptance on run 1347. The mixed models improve track tagging relative to

neutron-only training but remain below the measured prompt reference.

Selection	TPC tracks tagged [%]	Simulated neutrons tagged [%]	Calibration retained [%]
Prompt 500 ADC	96.6	61.0	87.8
Neutron compact	33.9	75.1	86.5
Mixed compact	70.1	72.9	87.0
Mixed monotonic	73.8	73.0	87.0
Mixed full	62.1	78.4	87.9

Table A.13: Exploratory single-BDT candidates at the prompt-reference-matched calibration target. Denominators are 351 experimental tracks, 4035 simulated neutron-plus-activity events, and 27 529 calibration controls. The measured calibration acceptance differs slightly between models despite matching their threshold-setting targets.

The hybrid retains every prompt tag and applies a frozen neutron BDT to prompt survivors,

$$V_{\text{hybrid}} = P \vee (s \geq c). \tag{A.8}$$

Its threshold is fixed using prompt-surviving run-1345 controls, with the acceptance budget defined relative to all controls in that run. Maintaining prompt-tagged tracks follows from this logical OR and does not demonstrate learned muon recognition. The prompt-only reference is recalibrated at each target to distinguish additional neutron-sensitive information from a lower amplitude threshold.

Selection	TPC tracks tagged [%]	Simulated neutrons tagged [%]	Calibration retained [%]
Prompt 500 ADC	96.6	61.0	87.8
Prompt only, 88% target	96.9	62.6	87.6
Hybrid counts/timing, 88%	96.6	63.9	87.6
Hybrid level 1 amplitude, 88%	96.6	64.7	87.6
Hybrid level 2 amplitude, 88%	96.6	68.0	87.7
Prompt only, 85% target	96.9	68.9	84.5
Hybrid level 2 amplitude, 85%	96.9	77.3	84.7
Hybrid level 2 amplitude, 80%	96.9	82.0	79.8

Table A.14: Prompt-preserving hybrid selections and prompt-only controls at several total calibration-acceptance targets. All hybrid BDTs were trained on prompt survivors. The rows use the same three evaluation denominators as Table A.13. Targets are fixed on run 1345; the final column reports actual run-1347 retention. The level-2 amplitude model excludes the inconsistent total-energy scalar.

The level-2 amplitude-aware hybrid at the 88% target preserves exactly the 339 prompt-tagged experimental tracks and tags 2742/4035 simulated neutron events, compared with

2462/4035 for the fixed prompt reference. Its simulated neutron gain is 6.94 percentage points, with a central 90% paired-bootstrap interval of 6.29–7.61 percentage points, while calibration retention changes from 87.80% to 87.67%. The hybrid rejects 36 additional calibration controls out of 27 529. Relative to the prompt-only selection criterion recalibrated to the same 88% target, the neutron gain remains 5.33 percentage points [4.73, 5.92], with slightly higher measured calibration retention; the recalibrated prompt selection criterion tags one extra experimental track. The hybrid therefore contributes neutron-sensitive information beyond a simple reduction of the prompt threshold in this comparison.

At the 85% target, the amplitude hybrid and matched prompt-only criterion both tag 340/351 tracks, while neutron rejection is 77.30% and 68.92% and measured control retention is 84.71% and 84.48%. The paired neutron difference is 8.38 percentage points [7.61, 9.17]. The 80% target gives 82.03% neutron rejection at 79.80% control retention and remains an alternative operating point. Paired intervals hold models, thresholds, and donor records fixed; they omit detector-response and run uncertainties.

An exploratory six-bin (1.2 µs) shift of physical simulated peaks was motivated by development-sample prompt medians of bin 204 in data and bin 198 in simulation. Measured peaks and activity donors were unchanged. Retraining after this shift changes full mixed-model track tagging from 62.1% to 67.2%, and monotonic-model tagging from 73.8% to 72.6%. Neither recovers the 96.6% prompt reference, so this correction is not adopted. The final hybrid uses the original time mapping, no amplitude rescaling, and the threshold specified in Section A.8.6. The selected tracks have unmeasured muon purity, and simulated neutron rejection remains conditional on the peak-overlay response.

A.8.6 Historical independent veto and joint background selection

This historical evaluation retains its original independent samples, response and timing convention. Its yields are separate from the normalized partial inventory now reported in Section 5.4.2; the two samples are not pooled.

Source	Generated primaries N_{gen}	Stored TPC trigger	TPC baseline $C_V C_E C_M C_F$	+ TPC BDT	+ prompt veto	+ neutron veto
n (total)	2 793 974 491	63 617	141	29	14	12
$e^{\pm}$	521 175 516	192	2	1	1	1
μ^-	2 819 074 149	1 550 944	45	7	7	7
γ	7 974 310 398	1 383	0	0	0	0
p	237 013 803	12 195	36	3	2	1

Table A.15: Cumulative selection on independent cosmic samples. Stored REST events have sensitive-gas energy deposition before response and reconstruction; the veto columns are sequential. The neutron baseline contains 140 interaction acquisitions and one activation acquisition rejected by the BDT. Stored totals preserve source accounting; activation is evaluated separately in Section 5.2.2. The historical muon source contains μ^- only, with the response and timing model of Chapter 4. Gamma zeros are observed counts. Proton reconstruction is complete, including a fixed-exposure replacement of one corrupted file. Development samples are excluded.

The adopted nominal 88% level-2 hybrid retains the transfer limitations established above and rejects events satisfying

$$P = \left[\sum_{190 \le b_i \le 210} A_i \ge 500 \right], \qquad V = P \vee \left[s_{\mathrm{res}} \ge 0.9382199287372652 \right], \tag{A.9}$$

where A_i is the nonnegative reconstructed peak amplitude in ADC counts and s_{res} is the score of the frozen neutron BDT trained on prompt survivors. Its 163 veto observables exclude the inconsistent total-energy scalar and all TPC variables. Weights, feature order, channel maps and threshold are archived together; the 85% and 80% targets remain alternatives.

Conventional simulated REST times map to $b = t_{\mathrm{REST}}/(0.2\,\mu\mathrm{s}) + 231$; measured peaks already use stored bins. Coordinates are rounded to the nearest sample; peaks outside 0–511 are discarded rather than moved to the boundary. The two readouts use separate channel-to-panel maps before combination, without the exploratory six-bin correction or amplitude rescaling. Results therefore remain conditional on this response and timing transfer.

Prompt-background acceptance is $C_V C_E C_M C_F C_{\mathrm{BDT}}(1 - V)$, without threshold reoptimization on TPC survivors. The 88% veto retention refers to the original controls, not the C_{BDT} subset; multiplying it by the 80% TPC target does not give joint signal acceptance. Neither the measured 351-track control nor the 4,035-neutron classifier evaluation supplies multiplicative background weights.

Table A.15 gives the complete independent neutron and electron exposures, from 250 and 180 files, respectively. Neutron baseline events are joined to physical veto peaks by campaign, file, entry and event identifier; electron identities and projection energies are matched before peak extraction. Each source retains all generated primaries, including those without a stored

event, in its yield denominator.

Each parent receives one complete measured veto record from calibration run 1347, disjoint from training and threshold setting. The SHA256 digest of a registered salt, draw index, campaign, file and parent identifier selects the donor independently of row order and veto decision. Draw zero is fixed as nominal before scoring. Donors preserve within-record channel, timing and amplitude correlations; peak concatenation omits waveform merging and additional saturation. The conditional control model's finite-bank and run-condition uncertainties are outside the counting limits.

The corrected neutron sequence in the table ends at 12 rather than 13: the stored trigger and reconstructed position identify one formerly exempted candidate as a prompt recoil, rejected by its 647 ADC prompt sum under the unchanged rule. Its later decay does not determine that acquisition's epoch; production activity still receives no coincidence credit at a later decay. The genuinely delayed baseline candidate fails the TPC BDT. The electron survivor has veto peaks but does not satisfy the rejection rule; its survival is not missing veto data. Original and corrected assignments are preserved separately in the companion event atlas.

Across 200 activity draws, the neutron count has mean 10.335 and central empirical 90% range 8–12, using inverse-empirical-CDF quantiles, with 12 nominally. The electron survives 180 draws, giving range 0–1. Physical peaks alone give the same nominal counts (12 neutrons, one electron), although identities can differ between draws. These repetitions test donor sensitivity, not independent particle exposure or experimental efficiency.

For k survivors from N generated primaries, $Y = k/N$. Selected parents have observed multiplicity one; the MC standard error includes zero-response parents,

$$\sigma_Y = \sqrt{\frac{k - k^2/N}{N(N-1)}}. \tag{A.10}$$

The individual one-sided 90% bound uses the Poisson working model at realized exposure,

$$Y_{90} = \frac{\chi^2_{2(k+1)}(0.90)}{2N}. \tag{A.11}$$

This historical counting convention is neither a distribution-free multiplicity guarantee nor a stopping-uniform interval. The prompt/delayed partition covers stored acquisitions, not all later radioactive decays, and supplies no complete activation rate or activation-specific bound.

Conditional on the response and nominal draw, neutron and electron/positron yields are $(4.29 \pm 1.24) \times 10^{-9}$ and $(1.92 \pm 1.92) \times 10^{-9}$ per respective primary. Their individual 90% upper bounds are 6.37×10^{-9} (outward rounded) and 7.47×10^{-9}.

Component	Survivors	Yield ± MC error per generated primary	Upper limit 90%
n	12	$(4.29 \pm 1.24) \times 10^{-9}$	6.37×10^{-9}
$e^{\pm}$	1	$(1.92 \pm 1.92) \times 10^{-9}$	7.47×10^{-9}
μ^-	7	$(2.48 \pm 0.939) \times 10^{-9}$	4.18×10^{-9}
γ	0	0	2.9×10^{-10}
p	1	$(4.22 \pm 4.22) \times 10^{-9}$	1.642×10^{-8}

Table A.16: Final yields after the TPC BDT and hybrid veto in independent samples. Errors and individual one-sided 90% counting limits are conditional on the fixed response and nominal activity draw (Section 4.6.5). The neutron acquisition-epoch correction is documented in the companion event atlas; delayed activation is incomplete. The historical μ^- benchmark contains negative muons only. Zero gamma survivors give a positive counting-model limit.

For physical primary rate Q_j in the simulated incident phase space,

$$B_j = \frac{Q_j Y_j}{(5\,\mathrm{keV})(\pi\,\mathrm{cm}^2)}. \tag{A.12}$$

The same scaling converts bounds for fixed Q_j. Incident-primary rates were not validated for this historical sample, so its yields are neither absolute backgrounds nor added across species; missing files or normalizations do not imply zero background. Muon and gamma samples use the same rule on the table's complete exposures (496 and 201 files). Muons are negative only, without charge-mixture correction; gamma has no baseline survivors and retains a counting limit. Section B.7 documents the proton evaluation. The independent selection supplement preserves the models, full-precision thresholds, exposure ledgers, event matches, donors and corrected assignments.

A.9 Observable importance in the adopted veto

The importance diagnostic uses the frozen residual neutron BDT adopted in Section 4.6.4. The adopted weights and their archived source model have identical checksums and the same ordered set of 163 veto observables. Recomputed scores for all 24 171 prompt-surviving held-out calibration controls and 1573 neutron-plus-activity events agree with the saved scores, with a maximum absolute difference of approximately 5.55×10^{-17}. The diagnostic leaves the production model, thresholds, and unpermuted candidate decisions unchanged and reuses the existing held-out data rather than providing a new blind evaluation.

Table 4.6 defines the five permutation groups, and Fig. 4.30 compares their effects with the leading individual observables. The joint permutations use a balanced sample of 1573 prompt-surviving neutron-plus-activity events and 1573 calibration controls, with ten shuffles per

group. Each shuffle preserves dependencies within that group while breaking its correlations with the remaining inputs. The baseline area under the receiver operating characteristic curve (AUC) is 0.776 395. Permuting the 68 amplitude-derived inputs reduces the AUC by 0.293 947 5, with a shuffle standard deviation of 0.011 888 7. This group includes amplitudes and their fractions across time windows, faces, and layers, spatial amplitude distributions, and amplitude-weighted timing. Figure A.15 expands the much smaller effects of the other four groups.

The group effects are not additive because correlated inputs can substitute for one another, and the shuffle standard deviations are not confidence intervals. The small effect of permuting the timing group does not show that neutron-capture timing is uninformative: window-dependent amplitudes and occupancies retain timing information in other groups. These AUC changes describe the classifier's dependence on its inputs; they do not measure rejection at the frozen 88% calibration-retention target or establish an experimental neutron efficiency.

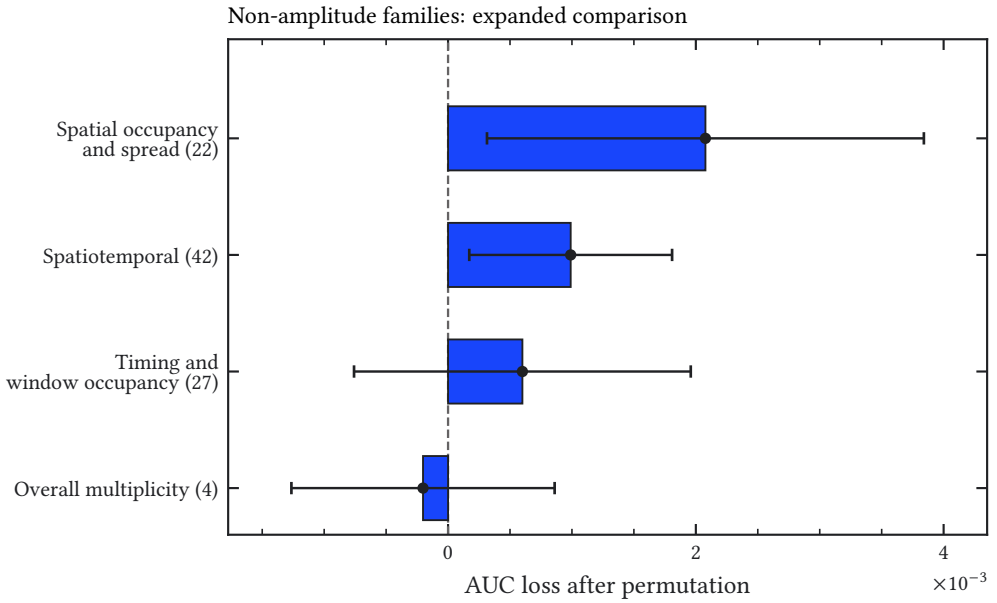


Figure A.15: Expanded view of the four non-amplitude permutation groups in the adopted residual neutron BDT: spatial occupancy and spread (22 inputs), spatiotemporal (42), timing and window occupancy (27), and overall multiplicity (4). Markers show the mean AUC decrease after joint permutation; error bars are standard deviations across ten shuffles on the balanced prompt-surviving sample. Signed negative values are retained and indicate that a shuffled sample can have a larger AUC than the original sample. The expanded axis has a separate scale from the full family comparison in Fig. 4.30.

A.9.1 Reliance at the frozen operating point

We also evaluated changes in rejection and control retention at the adopted residual threshold $s_{\mathrm{thr}} = 0.9382199287372652$, keeping the original prompt decisions fixed. The unpermuted scores reproduce the saved decisions for all 24 171 prompt-surviving calibration controls and 1573 neutron-plus-activity entries. The complete hybrid retains 24 135 of 27 529 controls and rejects 2742 of 4035 neutron entries; these full original populations define the percentage-point changes below.

For each family, one common permutation moves all its inputs together across the complete prompt-surviving mixture. Ten shared permutations pair the family comparisons. This mixture differs from the balanced AUC sample in Fig. 4.30; its results supplement rather than replace that diagnostic.

Permuted family	Neutron rejection change [percentage points]	Control retention change [percentage points]
Amplitude-derived	−6.44 ± 0.10	−0.947 ± 0.016
Overall multiplicity	+0.104 ± 0.027	−0.003 ± 0.003
Timing and window occupancy	+0.146 ± 0.041	−0.004 ± 0.006
Spatial occupancy and spread	−0.315 ± 0.091	+0.006 ± 0.005
Spatiotemporal	+0.050 ± 0.019	0.000 ± 0.000

Table A.17: Mean changes after family permutation at the frozen hybrid threshold, with the standard deviation across ten shuffles. Positive values indicate more neutron rejection or more calibration retention, respectively; the deviations are not confidence intervals.

The amplitude-derived family dominates this diagnostic as well as the AUC comparison. Maximum peak amplitude alone lowers rejection by 3.74 percentage points, whereas summed prompt amplitude and inner-layer prompt amplitude lower it by 2.91 and 2.53 points, respectively. Their effects are not additive, and the ordering need not match the AUC ranking. Permutation breaks relations among feature families and depends on the input mixture; it does not represent a measured gain change or establish a calibrated detector uncertainty.

To separate recipient-sample variation from shuffle variability, we made 2048 resampling draws grouped by the calibration donor's run, subrun, and event identifier, including repeated identities and keeping each class's denominator with its records. All donor mappings are complete. For the amplitude-derived family, the central 68.27% resampling ranges are $[-6.84, -6.05]$ percentage points for neutron rejection and $[-0.975, -0.918]$ for control retention. These ranges are conditional on the ten fixed shuffle outcomes; they exclude new activity assignments, banks or runs, training, threshold, and detector-response uncertainty. Unique archived neutron source entries do not establish physical-parent independence, so this is an activity-donor sampling diagnostic rather than a physical-parent bootstrap or a

new independent validation.

A.9.2 Controlled timing and spatial ablation

The comparison in Section 4.6.4 partitions the ordered 163 inputs into the three disjoint blocks defined in Table 4.7. Block G contains maximum peak amplitude, peak count, unique panel count, active face count, and active layer count. Block N contains the 27 timing and window-occupancy inputs, 22 spatial inputs, and 42 spatiotemporal count inputs in Table 4.6. Block A contains the other 67 amplitude-derived inputs, including amplitude-weighted timing. The global nonnegative amplitude sum is excluded from G, unlike the six-aggregate reference in Fig. 4.31; resolved sums in A retain additional amplitude-scale information. None of these input restrictions decorrelates occupancy from energy or defines independent physical mechanisms.

Every fit uses the same 77 923 prompt-surviving training rows: 73 257 measured controls and 4666 neutron-plus-activity entries from the original training partitions. Runs 1335, 1339, and 1342 train the models; run 1345 sets thresholds and run 1347 evaluates them. The original 500-ADC prompt mask remains fixed in all partitions. The fits use scikit-learn 1.8.0, 220 boosting iterations, learning rate 0.045, at most 12 leaves, minimum leaf size 30, L_2 regularization 0.10, balanced class weights, no early stopping, and seed 20260908. The full fit reproduces all 25 744 archived post-prompt evaluation scores within 5.55×10^{-17}, with identical decisions and the exact adopted threshold. The production model and its policy are preserved.

Prompt retention on run 1345 is $24\,586/27\,878 = 88.1914\%$. The nominal 88% total-retention target therefore requires 99.7830% conditional retention among its prompt survivors. Each threshold is fixed by the higher empirical score quantile, with the original next-representable-value convention for ties. Every model retains 24 533 of 27 878 threshold-setting controls, or 88.0013%. Table A.18 reports achieved evaluation counts relative to the complete original populations.

Inputs	Count	Controls retained	Neutrons rejected
G	5	24 092	2 529
$G + N$	96	24 115	2 621
$G + A$	72	24 135	2 744
$G + N + A$	163	24 135	2 742

Table A.18: Controlled input ablation on 27 529 original calibration controls and 4 035 neutron-plus-activity entries, using the blocks G, N, and A defined in Table 4.7. Counts include the unchanged prompt decisions; thresholds use run 1345 alone at nominal 88% total retention.

Paired uncertainty uses 4096 recipient donor-cluster resampling draws, grouping records

by calibration run, subrun, and event identifier and carrying the class denominators with their records. All mappings are complete; repeated donor identities remain grouped across model comparisons. The plotted ranges contain the central 68.27% of these draws, while the main-text ranges contain the central 90%. The fits, thresholds, prompt rule, and original activity assignments are fixed. These ranges exclude training variation, threshold estimation, new runs or donors, and detector-response uncertainty; source-entry uniqueness does not establish independent physical parents. The reused evaluation data make this an exploratory representation comparison rather than a new independent validation.

A.9.3 Exploratory projection of the veto observables

We projected the 163 inputs of the adopted veto BDT onto two dimensions to inspect the response patterns of the different event populations. Events are sampled before applying either the prompt veto or the residual BDT rejection. Uniform Manifold Approximation and Projection (UMAP) constructs this view from neighborhoods in the input space [8]. Before fitting, amplitudes, counts, and occupancies are transformed with $\log(1 + x)$, and each observable is centered and scaled to unit variance over the displayed sample. Fractions, entropies, and stored-bin times remain linear before standardization. Population labels, TPC observables, event identifiers, and BDT scores are excluded from the projection inputs.

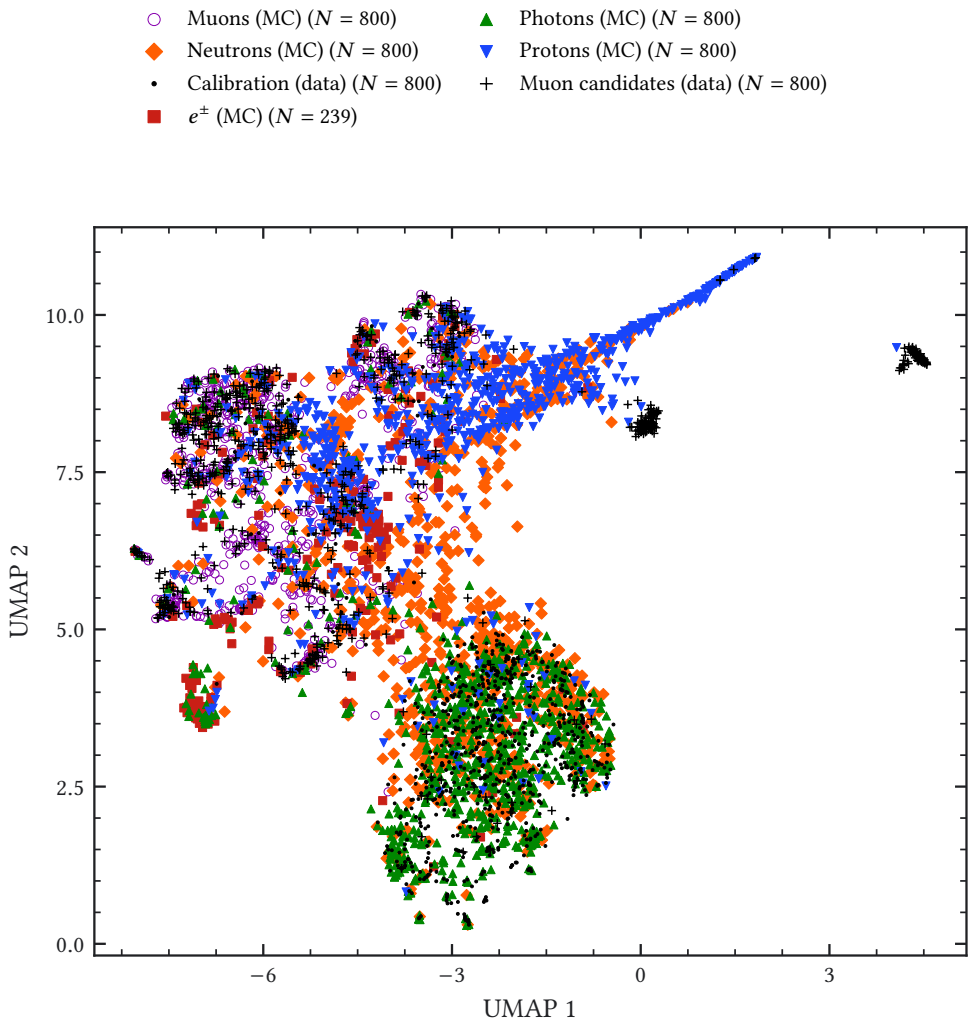


Figure A.16: UMAP projection of the 163 veto observables for 5039 events. Each cohort contributes up to 800 events; all 239 electron/positron events are retained. Simulated particles include measured calibration activity, containing unrelated physical activity and electronic noise. Measured muon candidates have an additional TPC selection and unmeasured purity. UMAP uses 30 neighbors and the parameter `min_dist`=0.15. Colors and markers identify populations after fitting. Sample sizes and visual densities do not represent relative particle fluxes.

Figure A.16 shows overlapping populations with different production and selection conditions. The measured muon candidates and the simulated muons in this overview do not have a common TPC selection. The geometry also depends on the relative weights of the input families: after giving the five families equal total marginal variance, only 37.5% of

the 15 nearest UMAP neighbors are retained, evaluated within a fixed subset of 1000 events. The apparent groups therefore provide an exploratory view of the inputs and do not establish physical separation or veto rejection efficiency.

A.10 Acquisition conditions and design controls

The selected HENSA design study uses 200 ns sampling, 1014 ns shaping, and a `REST-for-Physics` alignment parameter of 60 µs. Its simulation readout map contains 60 panel aliases. Section A.2 distinguishes the selected 59-panel GDML design from the separate 60-panel snapshots preserved in September; map size alone does not identify a production's transport geometry. Its processed-entry population has no additional low-energy TPC selection. The experimental IAXO-D0 paper readout defines 59 panel positions, of which 57 supplied operating veto channels in the published data. Its published record lasts approximately 100 µs, with the Micromegas trigger approximately 30 µs after its start [1]. Hardware timing, the reconstruction alignment parameter, and the stored-bin coordinate are separate conventions. The experimental exposure is 52.1 d; its central background levels and calibration acceptance follow the publication, with counting intervals recomputed from the stated counts.

A.10.1 Material ordering and quenching

The simplified sandwich comparisons precede light attenuation, waveform shaping, and peak finding. Figure 4.13 compares lead placement at fixed total lead thickness. Here we retain the separate active-material comparison and the dependence on scintillator quenching. These are saved-energy diagnostics; they do not establish the response of an assembled veto.

Figure A.17 fixes the ordering to 20 cm Pb, 10 cm scintillator, and 20 cm Pb. The cadmium variant has a 5 mm sheet on each face of the scintillator, totaling 10 mm, and is distinct from the detector's 1 mm sheets. The configured source is uniform in $\ln E$ from 1 MeV to 100 GeV; its integrated threshold response consequently reflects the sampled energy mixture. The three variants use a common Birks coefficient of 0.126 mm MeV^{-1}, without material-specific light-yield calibration or optical collection. The aggregate summaries do not identify the response branch used in every input file, so their nominal electron-equivalent energy scale is retained as a comparison proxy. The Wilson intervals treat saved entries as independent trials; physical-parent independence cannot be re-established from these aggregate summaries.

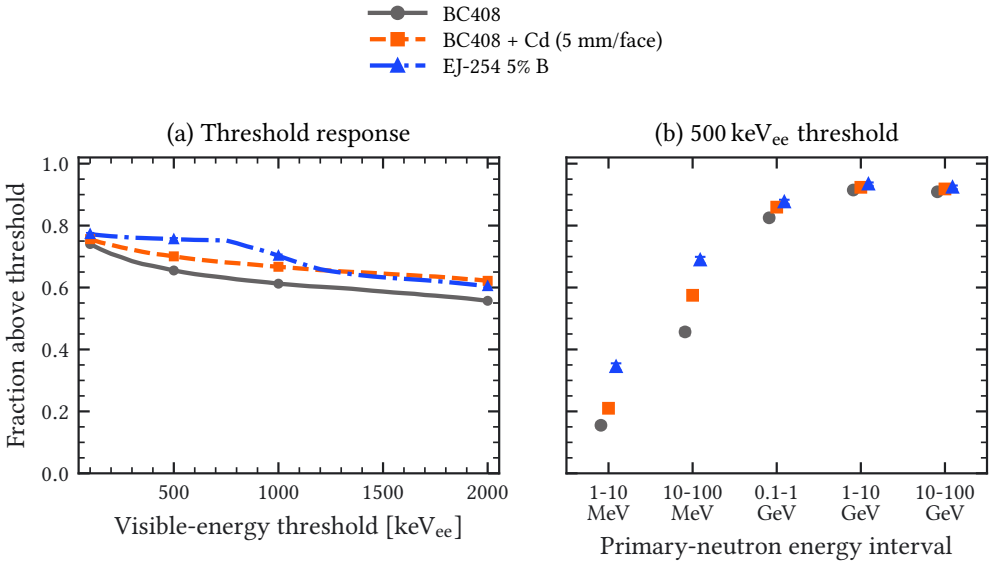


Figure A.17: Active-material dependence in a fixed Pb–scintillator–Pb sandwich. Panel (a): fraction of scan entries above the nominal visible-energy threshold for the sampled source. Panel (b): fractions in primary-energy intervals at 500 keV$_{ee}$. Error bars show central 68.27% Wilson intervals.

At 500 keV$_{ee}$, the integrated fractions are 65.50%, 70.06%, and 75.57% for BC408, cadmium-assisted BC408, and boron-loaded EJ-254, respectively. These differences describe energy-dependent material sensitivity within the shared response approximation; they do not select a detector material without its optical response and operating constraints. Figure A.18 gives the energy-weighted visible fraction across the wider set of orderings.

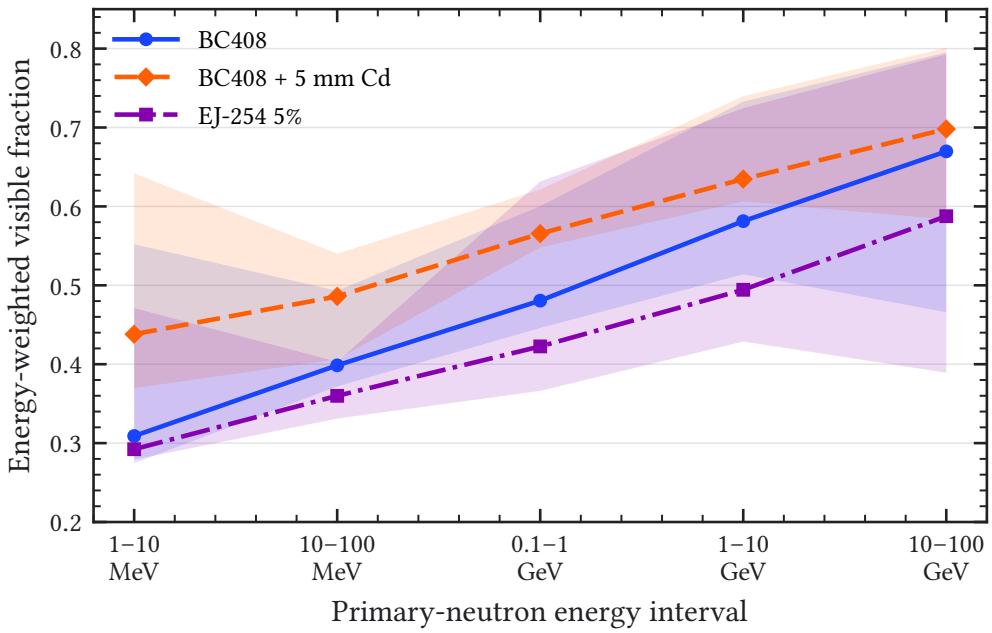


Figure A.18: Energy-weighted visible fraction, $\sum E_{\mathrm{vis}} / \sum E_{\mathrm{dep}}$, after the Birks correction in seven simplified sandwich orderings per scintillator and energy interval. Each point includes 6483–10 106 selected events. Bands are the full configuration-to-configuration ranges, rather than statistical intervals.

A.10.2 Capture and reconstructed-response diagnostics

Figure A.19 retains the complete capture diagnostic for the 2706-entry response sample used in Figure 4.17. It is separate from the 3864-history full-chain population in Section 4.3.6. Material fractions from these samples describe their own saved populations and should not be combined.

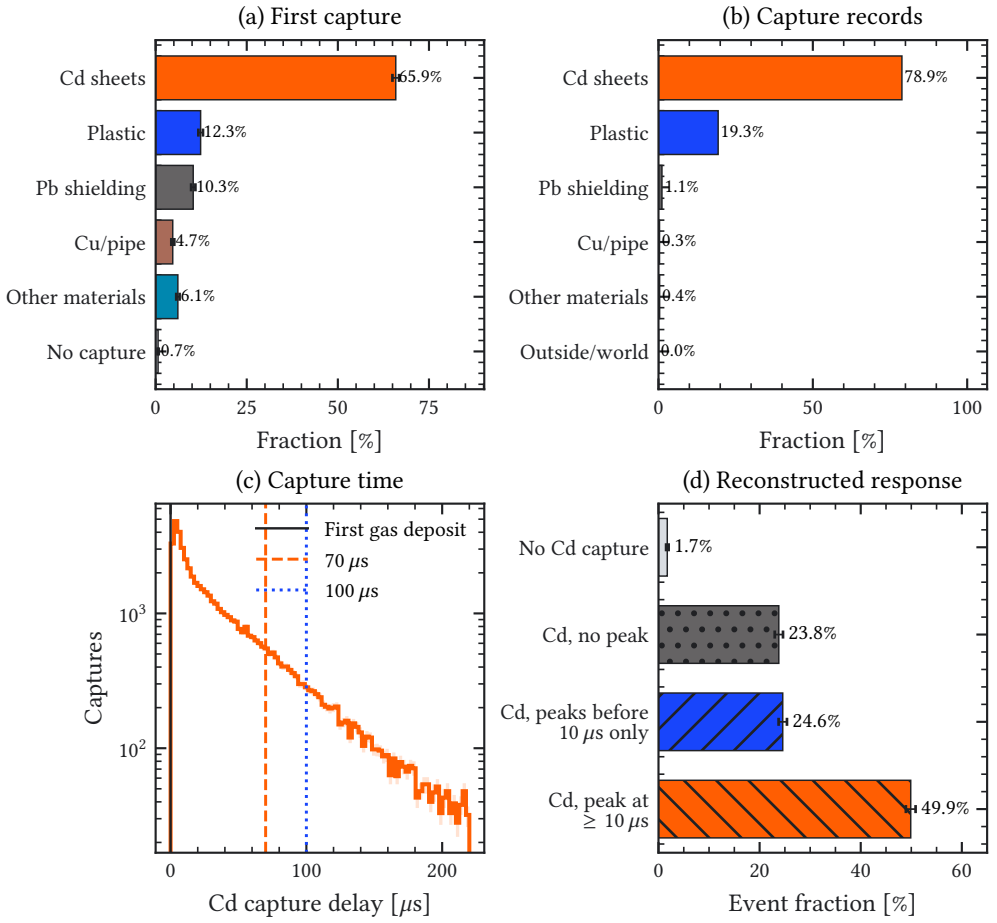


Figure A.19: Capture diagnostics for 2706 processed entries, without an additional TPC selection. Panels (a) and (b): first-capture material per entry and the material of all 73 295 capture records, respectively. Panel (c): cadmium-capture delay from the first simulated gas deposit for the 2453 entries whose first gas-deposit time is below 1 µs in the stored transport-time coordinate; the band shows a central 68.27% whole-event bootstrap interval. Panel (d): capture-to-peak correspondence, with late peaks defined by reconstructed time $t \geq 10$ µs. Entry fractions in panels (a) and (d) have central 68.27% Wilson intervals; panel (b) counts correlated capture records.

A.10.3 Generated exposure in the HENSA layer scan

The generated-primary counts in Table A.19 give a physical normalization for the layer comparison. Entry probability counts processed analysis entries per generated primary; tagging is conditional on those entries. The nearly constant entry probability of the cadmium configurations separates geometric acceptance from the gain in reconstructed veto tagging.

Configuration	Primaries	Entry probability	Any tag
1 layer + Cd	8.941×10^8	2.209×10^{-5}	0.465
2 layers + Cd	4.779×10^8	2.247×10^{-5}	0.675
3 layers + Cd	3.320×10^8	2.199×10^{-5}	0.741
4 layers + Cd	2.764×10^8	2.258×10^{-5}	0.771
3 layers + Gd	2.278×10^8	2.262×10^{-5}	0.725
3 layers + steel	2.455×10^8	2.452×10^{-5}	0.420

Table A.19: Generated exposure and reconstructed tagging for the HENSA outdoor-neutron design scan. The denominator includes all generated primaries; the tagging fraction uses the processed-entry population without the final TPC baseline.

A.10.4 Passive-shielding and late-window cross-checks

The lead-thickness photon and neutron summaries in Chapter 4 show the components that determine the passive design. The remaining species do not change that conclusion: the muon contribution varies weakly with lead thickness, while proton cascades show the same need for an active response as neutron cascades. The full per-particle projections and experimental late-window event rasters are preserved in the independent veto-validation document. For the late-window comparison, Table A.9 retains the quantitative mismatch in TPC energy, hit multiplicity, and veto activity.

A.10.5 Closed-layer shielding: counts and statistical intervals

The closed-layer comparison uses B^E/B^E_{Pb}, with B^E defined in Equation 4.3. Valid reconstructed events must have calibrated maximum-track energies in 2–7 keV; the counting area is the full 36 cm^2 readout. Track-multiplicity, fiducial, BDT, and active-veto requirements are not applied.

The shielding contains a total lead thickness of 20 cm. We varied the total borated polyethylene (BPE) thickness M, the inner lead thickness p, and the number of BPE layers n, as summarized in Table A.20. From the cavity outward, the sequence is $\mathrm{Pb}(p) + [\mathrm{BPE}(M/n), \mathrm{Pb}((20-p)/n)]^n$, with thicknesses in centimeters. At $p = 20$ cm, the adjacent BPE layers form one continuous layer, so this configuration is counted once for each M. The inventory therefore contains $4 \times (4 \times 3 + 1) = 52$ distinct BPE configurations, 52 matched-air configurations with the same layer dimensions, and one full-lead control.

Parameter	Values
Total BPE thickness M	2.5 cm, 5 cm, 7.5 cm and 10 cm
Inner lead thickness p	0 cm, 5 cm, 10 cm, 15 cm and 20 cm
BPE layers n	1, 2, 3; one geometry at $p = 20$ cm

Table A.20: Parameters of the closed-layer scan. The matched-air configurations replace BPE by air while preserving the corresponding geometry and total lead thickness.

For configuration i, let N_i be the number of generated neutrons and Y_{ij} equal one if neutron j produces at least one energy-selected subevent, and zero otherwise. The selected-parent count $K_i = \sum_{j=1}^{N_i} Y_{ij}$ counts each neutron once. In these samples, every selected neutron produces one energy-selected subevent, so K_i also equals the energy-selected subevent count. The estimate $\widehat{\epsilon}_i = K_i/N_i$ includes every generated neutron, including those that produce no saved or reconstructed event. Neither the stored-event count nor the preceding selected population replaces N_i.

Treating the independent generated neutrons as Bernoulli trials, we use central 68.27% Clopper–Pearson intervals for the acceptance [9]. Writing $Q(q; a, b)$ for the q quantile of a beta distribution with parameters a, b, and $\alpha = 1 - 0.6827$, the limits for $0 < K_i < N_i$ are

$$\epsilon_{i,L} = Q(\alpha/2; K_i, N_i - K_i + 1), \qquad \epsilon_{i,U} = Q(1 - \alpha/2; K_i + 1, N_i - K_i). \tag{A.13}$$

The lower endpoint is zero for $K_i = 0$, and the upper endpoint is one for $K_i = N_i$. The one-sided 90% upper limit is $\epsilon_{i,90} = Q(0.90; K_i + 1, N_i - K_i)$, with endpoint one when $K_i = N_i$. For zero selected parents, this reduces to $\epsilon_{i,90} = 1 - 0.1^{1/N_i}$.

The background ratio uses the observed energy-selected full-lead acceptance, $\widehat{\epsilon}_{\mathrm{Pb}} = 1990/143\,800\,000 \simeq 1.3839 \times 10^{-5}$, through

$$R_i = \frac{B_i^E}{B_{\mathrm{Pb}}^E} = \frac{\widehat{\epsilon}_i}{\widehat{\epsilon}_{\mathrm{Pb}}}, \qquad [R_{i,L}, R_{i,U}] = \frac{[\epsilon_{i,L}, \epsilon_{i,U}]}{\widehat{\epsilon}_{\mathrm{Pb}}}, \qquad R_{i,90} = \frac{\epsilon_{i,90}}{\widehat{\epsilon}_{\mathrm{Pb}}}. \tag{A.14}$$

Thus each configuration interval and upper-limit marker is conditional on the observed Pb normalization. The shaded Pb interval separately shows its own binomial limits divided by $\widehat{\epsilon}_{\mathrm{Pb}}$; that uncertainty is not included in the configuration error bars. A confidence interval for the ratio of two unknown acceptances would require a joint construction. The displayed intervals are pointwise Monte Carlo intervals under the adopted source, transport, and detector-response models. They exclude model uncertainty and do not provide simultaneous coverage for a minimum selected across the scan.

Table A.21 gives the counts for the lowest observed energy-selected yield at each BPE thickness, the layer-number comparison at $M = 10$ cm, and the air configuration matched to the lowest BPE background. The full comparison contains 3 551 600 000 generated neutrons from the completed samples with the same scientific settings. The smaller baseline counts

describe a later selection and are not the numerators of the energy-selected comparison.

Material	M	p	n	$N/10^6$	Energy	Baseline
Pb control	0	20	0	143.8	1 990	26
BPE	2.5	15	1	52.0	663	7
BPE	5	5	1	52.0	592	5
BPE	7.5	5	1	52.0	551	7
BPE	10	5	1	51.8	499	6
BPE	10	5	2	51.8	553	7
BPE	10	5	3	51.8	558	6
Matched air	10	5	1	13.6	166	3

Table A.21: Generated neutrons and selected counts for representative closed-layer configurations. The thicknesses M and p are in centimeters; N includes all generated neutrons. Energy-selected counts are unique selected parents, whereas the last column gives the subevents passing the full baseline selection before BDT and active-veto rejection.

The ^{6}Li-loaded LiF/PMMA comparison is a separate population. Its 23 configurations comprise the Pb control, four BPE references, and 18 LiF/PMMA variants. The BPE references use $p = 5$ cm: $M = 7.5$ cm with $n = 1$, and $M = 10$ cm with $n = 1, 2, 3$. Films of 1 mm, 2 mm and 5 mm replace an equal BPE thickness at the TPC-facing face of the innermost BPE shell or of each BPE shell. These placements coincide for $n = 1$; the thickness is specified per treated face, while the lead, cavity, and outer shielding envelope remain fixed. The displayed overnight population contains 3672 completed samples of 400 000 neutrons, or 1 468 800 000 generated neutrons in total. Each configuration contains between 63 600 000 and 64 000 000 generated neutrons. Pilot and earlier lithium production samples are excluded from this comparison. The successive selected totals are 64 978 Valid events, 15 705 after energy selection, 9295 after track-multiplicity selection, and 199 after the full baseline. Its background ratio uses its own Pb control, with $K_{\mathrm{Pb}} = 796$ energy-selected events and $N_{\mathrm{Pb}} = 63\,600\,000$, and the same statistical construction above.

Appendix B

Background-Source Validation

B.1 Reference geometries . . . 54
B.2 Detector-material activities and selected response . . . 55
B.2.1 Parent-correlated estimates and limits . . . 56
B.2.2 Source-resolved selection . . . 57
B.3 Lead activity, copper protection, and source depth . . . 62
B.3.1 Full-detector copper control . . . 62
B.3.2 Depth and numerical transport checks . . . 63
B.4 Radiogenic neutrons in the lead shield . . . 66
B.4.1 Enlarged interface source . . . 66
B.4.2 Transport and selection . . . 67
B.5 Activation normalization and decay response . . . 68
B.5.1 Angular normalization and source closure . . . 68
B.5.2 Nuclear states, feeding, and response trials . . . 69
B.5.3 Finite irradiation and equilibrium . . . 70
B.5.4 Production inventory and selected-decay response . . . 71
B.6 Environmental-source and transport checks . . . 72
B.6.1 Paired NaI reference and calibration . . . 74
B.6.2 Boundary normalization and complete reference exposure . . . 75
B.6.3 Weighted transport and independent confirmation . . . 76
B.6.4 Shield-facing gamma confirmation . . . 77
B.6.5 Native residual-neutron production . . . 78
B.7 Historical cosmic production coverage and interaction mechanisms . . . 79
B.8 Normalization and response of the partial cosmic inventory . . . 81

The background predictions in Chapter 5 depend on the generated radiation field, the radioactive inventory, and the reconstructed detector response. This appendix gives the checks that support those predictions: the geometries and activity assumptions, the accounting of

generated particles and selected events, and the treatment of correlated decay histories. The results use the final time projection chamber (TPC) baseline and the stated boosted decision tree (BDT) unless an alternative is identified explicitly. Source and detector-model assumptions remain separate from Monte Carlo (MC) counting uncertainty.

B.1 Reference geometries

Table B.1 identifies the geometries used in the source studies. The files belong to the `iaxo/iaxo-geometry` repository at revision `079aa389f714`; their complete hashes and the derived telescope controls are preserved with the supplementary study records. All five files match that revision, and G1–G4 also match the independently preserved study inputs. This correspondence identifies the reference geometry; an older production is assigned that geometry only when its saved input establishes the match.

ID	Configuration	Use in this work
G1	IAXO-D1 reference	Full-detector telescope comparison.
G2	BabyIAXO with telescope pipe	Recovered telescope source and pipe geometry.
G3	IAXO-D1, three veto layers	Neutron reference; byte-identical to G1 at this revision.
G4	IAXO-D1 chamber and pipe	Reduced calibration-response geometry.
G5	IAXO-D1, closed pipe	Environmental-gamma reference with a closed pipe.

Table B.1: Geometry references. The archived geometry files are respectively `Default.gdml`, `WithTelescopePipe.gdml`, `CompleteVeto3Layers.gdml`, `ChamberAndPipe.gdml`, and `DefaultClosedPipe.gdml`.

The telescope controls displace the source disk by −100 mm, 0 mm, and 100 mm, first in G2 and then in the combined G1+G2 geometry. Each derivative has a distinct saved geometry, so source placement and detector geometry can be compared without changing the incident energy distribution. These controls test geometrical acceptance; they do not supply the missing telescope activities.

Control	Parent geometry	Disk offset [mm]
C1	G2, BabyIAXO telescope	−100
C2	G2, BabyIAXO telescope	0
C3	G2, BabyIAXO telescope	100
C4	G1+G2, IAXO-D1 with telescope	−100
C5	G1+G2, IAXO-D1 with telescope	0
C6	G1+G2, IAXO-D1 with telescope	100

Table B.2: Generated telescope-control geometries. Offsets are relative to the nominal source disk along its axis; all six derivatives have distinct archived geometry hashes. The combined controls retain the telescope-side pipe and the IAXO-D1 detector.

B.2 Detector-material activities and selected response

Material activity and detector acceptance enter the prediction at different stages. The adopted activities are mainly screening-limit scenarios; the readout ^{40}K values are reference values from the collaboration inventory [10]. Individual assay uncertainties and confidence levels are not retained in the numerical inventory. We therefore treat these activities as fixed normalization scenarios, rather than measured central values or joint confidence bounds on background. Missing inputs remain unknown.

The total activity is the specific activity multiplied by the mass of the transported source. The current Kapton readout measures 134 mm × 134 mm × 0.5 mm and has mass 0.012 748 76 kg; the earlier 0.001 022 4 kg normalization is superseded. The patterned copper cathode includes the annulus, support bars, and central ring, with their overlaps counted once. The Mylar source contains only the polyethylene terephthalate (PET) film of the aluminized window; its mass excludes the aluminum coating, whose activity is unavailable.

Material / source	^{238}U	^{232}Th	^{40}K	^{60}Co
Copper structures[1]	6.2×10^{-5}	2×10^{-5}	—	5.25×10^{-4}
Copper pipe	6.2×10^{-5}	2×10^{-5}	—	4.59×10^{-4}
Copper readout	0.26	0.14	0.57	0.031
Kapton readout	0.11	0.031	0.077	0.016
PTFE structures	6.2×10^{-5}	4.1×10^{-5}	—	—
Mylar film (PET)	0.29	0.008	0.033	0.001 8

Table B.3: Adopted specific activities in Bq kg^{-1}. The ^{40}K readout entries are reference values; the remaining supplied entries are activity-limit scenarios with unspecified individual assay confidence levels. Dashes denote unavailable inputs. Masses and source volumes are given in Table 5.11.

The additional ^{235}U specific activities are 0.093 Bq kg^{-1} for copper readout, 0.046 Bq kg^{-1} for Kapton, and 0.0029 Bq kg^{-1} for PET. For PET, the ^{226}Ra, ^{228}Th, and ^{137}Cs scenarios are respectively 0.0059 Bq kg^{-1}, 0.0036 Bq kg^{-1}, and 0.002 Bq kg^{-1}. The corresponding PTFE inputs are unavailable. Full chains starting at ^{238}U and ^{226}Ra, or at ^{232}Th and ^{228}Th, are overlapping hypotheses for the same material. They cannot be added without dividing the activity into disjoint chain segments. Likewise, gas and deposited radon daughters require a consistent retention model before combination with the bulk inventory.

Standalone transport samples of 2×10^8 decays per material–isotope configuration show that electrons escape mainly from shallow depths, whereas photons can emerge from deeper material. Those escape spectra precede detector response and provide no additional topology-rejection factor. The configured 1 mm range threshold, 1 keV conversion floor, and 0.05 mm gas step limit also have different roles: none alone establishes convergence at the 2 keV analysis threshold. Selected yields with tighter region-specific secondary-production thresholds would be needed to assign a numerical convergence uncertainty.

B.2.1 Parent-correlated estimates and limits

A complete decay chain can supply several reconstructed events. The independent simulation unit is therefore the generated parent, including parents with no stored event. If K_j is the selected multiplicity of parent j, define $S = \sum_j K_j$, $Q = \sum_j K_j^2$, and N as the complete generated-parent count. The response per becquerel is

$$R = \frac{S}{N \Delta E A_{\mathrm{fid}}}, \qquad \Delta E = 5\,\mathrm{keV}, \qquad A_{\mathrm{fid}} = \pi\ \mathrm{cm}^2. \tag{B.1}$$

[1]The chamber, backplate, local copper box, and patterned cathode share these specific activities.

MC errors retain the grouping by parent. The moment-matched scaled-Poisson approximation [11] gives

$$n_{\mathrm{eff}} = \frac{S^2}{Q}, \qquad s = \frac{Q}{S}, \qquad R_{90}^{\dagger} = \frac{s}{2N\Delta E A_{\mathrm{fid}}} F^{-1}_{\chi^2_{2(n_{\mathrm{eff}}+1)}}(0.90). \tag{B.2}$$

Continuation to noninteger n_{eff} is a working approximation for the multiplicity distribution. It becomes the ordinary Poisson construction when each selected parent contributes one event. Populated samples use the parent-correlated normal construction where specified in their numerical receipt. For an empty sample, the $2.3026/(N\Delta E A_{\mathrm{fid}})$ limit assumes rare single-event counting; it does not bound an unseen multiplicity tail. The dagger marks these counting assumptions, and the intervals are not uniform over all possible stopping times of recovered productions.

An activity-normalized subtotal includes only supplied activities for non-overlapping sources. For m source terms, each upper bound receives tail probability $0.1/m$ before the activity-weighted limits are added. This yields the stated family confidence level within the component counting models, while preserving their physical limitations. It does not assign a confidence level to the activity scenarios.

The same method is used for the telescope and electronics cards. Telescope weights include the photon yield per decay and sampled solid angle; the sampling unit is a generated photon history. For cards, the coefficient is $w_{i,c} = A_{i,c}/(N_{i,c}\Delta E A_{\mathrm{fid}})$, and $B_i = \sum_c w_{i,c} S_{i,c}$. Independent card variances are added with these unequal exposure weights. Four-card isotope bounds allocate $0.1/4$ per card, and the 24-term card-only subtotal allocates $0.1/24$, including zero-survivor strata. The historical energy-binned card selection is superseded by the common-baseline evaluation and supplies no additional rejection factor.

B.2.2 Source-resolved selection

Table B.5 preserves the generated exposure and the two selection endpoints for every simulated material source. Baseline counts apply reconstruction validity, the energy window, track multiplicity, and fiducial containment cumulatively. The BDT column adds the fixed simulation threshold; no veto is applied. Sources with unavailable activities remain useful response calculations but do not enter the known-activity subtotal. Figure B.1 shows how that supplied-activity subset changes with the BDT.

DRAFT — 2026-10-25 22:19 CET | Git d7910367

Table B.4: Detector-material background by component and isotope, before veto rejection. Backgrounds are in counts $\mathrm{keV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. A dash denotes missing activity; ★ marks a chain alternative excluded from subtotals. The † convention follows Section 5.1.1.

Isotope	Activity A_{ic} [Bq]	Baseline Estimate	+ BDT Estimate ± MC error	90% limit After BDT
Copper chamber				
^{238}U	1.63×10^{-4}	1.04×10^{-10}	0	$\leq 2.40 \times 10^{-10\dagger}$
^{232}Th	5.26×10^{-5}	0	0	$\leq 7.72 \times 10^{-11\dagger}$
^{40}K	—	—	—	—
^{60}Co	1.38×10^{-3}	0	0	$\leq 2.03 \times 10^{-9\dagger}$
Subtotal		1.04×10^{-10}	0	$\leq 3.46 \times 10^{-9\dagger}$
Copper backplate				
^{238}U	1.5×10^{-4}	9.53×10^{-11}	0	$\leq 2.20 \times 10^{-10\dagger}$
^{232}Th	4.83×10^{-5}	0	0	$\leq 7.08 \times 10^{-11\dagger}$
^{40}K	—	—	—	—
^{60}Co	1.27×10^{-3}	0	0	$\leq 1.86 \times 10^{-9\dagger}$
Subtotal		9.53×10^{-11}	0	$\leq 3.18 \times 10^{-9\dagger}$
Local copper box				
^{238}U	2.68×10^{-3}	0	0	$\leq 3.93 \times 10^{-9\dagger}$
^{232}Th	8.64×10^{-4}	0	0	$\leq 1.27 \times 10^{-9\dagger}$
^{40}K	—	—	—	—
^{60}Co	2.27×10^{-2}	3.37×10^{-10}	$(9.19 \pm 5.31) \times 10^{-11}$	$\leq 2.05 \times 10^{-10\dagger}$
Subtotal		3.37×10^{-10}	$(9.19 \pm 5.31) \times 10^{-11}$	$\leq 7.93 \times 10^{-9\dagger}$
Copper detector pipe				
^{238}U	1.26×10^{-3}	0	0	$\leq 1.85 \times 10^{-9\dagger}$
^{232}Th	4.06×10^{-4}	0	0	$\leq 5.96 \times 10^{-10\dagger}$
^{40}K	—	—	—	—
^{60}Co	9.33×10^{-3}	1.77×10^{-10}	$(2.52 \pm 1.78) \times 10^{-11}$	$\leq 6.72 \times 10^{-11\dagger}$
Subtotal		1.77×10^{-10}	$(2.52 \pm 1.78) \times 10^{-11}$	$\leq 3.70 \times 10^{-9\dagger}$
Patterned copper cathode				
^{238}U	3.42×10^{-6}	1.54×10^{-11}	$(2.14 \pm 0.96) \times 10^{-12}$	$\leq 3.97 \times 10^{-12\dagger}$
^{232}Th	1.1×10^{-6}	1.86×10^{-12}	$(9.29 \pm 2.94) \times 10^{-13}$	$\leq 1.44 \times 10^{-12\dagger}$
^{40}K	—	—	—	—
^{60}Co	2.89×10^{-5}	1.84×10^{-11}	0	$\leq 4.25 \times 10^{-11\dagger}$
Subtotal		3.57×10^{-11}	$(3.06 \pm 1.00) \times 10^{-12}$	$\leq 6.92 \times 10^{-11\dagger}$
Copper readout				
^{238}U	1.68×10^{-3}	4.88×10^{-8}	$(7.44 \pm 0.39) \times 10^{-9}$	$\leq 7.94 \times 10^{-9\dagger}$
^{235}U	6×10^{-4}	2.53×10^{-8}	$(8.10 \pm 0.73) \times 10^{-9}$	$\leq 9.04 \times 10^{-9\dagger}$
^{232}Th	9.03×10^{-4}	1.63×10^{-8}	$(3.12 \pm 0.29) \times 10^{-9}$	$\leq 3.50 \times 10^{-9\dagger}$
^{40}K	3.68×10^{-3}	7.02×10^{-9}	0	$\leq 5.40 \times 10^{-9\dagger}$
^{60}Co	2×10^{-4}	7.64×10^{-10}	$(1.27 \pm 1.27) \times 10^{-10}$	$\leq 4.96 \times 10^{-10\dagger}$
Subtotal		9.82×10^{-8}	$(1.88 \pm 0.09) \times 10^{-8}$	$\leq 3.15 \times 10^{-8\dagger}$
Kapton readout				
^{238}U	1.4×10^{-3}	2.75×10^{-9}	$(1.94 \pm 0.56) \times 10^{-9}$	$\leq 2.88 \times 10^{-9\dagger}$
^{235}U	5.86×10^{-4}	2.31×10^{-9}	$(1.08 \pm 0.29) \times 10^{-9}$	$\leq 1.56 \times 10^{-9\dagger}$
^{232}Th	3.95×10^{-4}	5.77×10^{-10}	$(2.88 \pm 1.09) \times 10^{-10}$	$\leq 4.85 \times 10^{-10\dagger}$
^{40}K	9.82×10^{-4}	0	0	$\leq 1.44 \times 10^{-9\dagger}$
^{60}Co	2.04×10^{-4}	0	0	$\leq 3.00 \times 10^{-10\dagger}$
Subtotal		5.64×10^{-9}	$(3.31 \pm 0.64) \times 10^{-9}$	$\leq 8.88 \times 10^{-9\dagger}$
PTFE chamber liner				
^{238}U	1.3×10^{-6}	5.58×10^{-12}	$(2.79 \pm 0.99) \times 10^{-12}$	$\leq 4.53 \times 10^{-12\dagger}$
^{226}Ra★	—	—	—	—
^{235}U	—	—	—	—
^{232}Th	8.59×10^{-7}	4.09×10^{-12}	$(1.66 \pm 0.43) \times 10^{-12}$	$\leq 2.36 \times 10^{-12\dagger}$
^{228}Th★	—	—	—	—
^{40}K	—	—	—	—
^{60}Co	—	—	—	—
^{137}Cs	—	—	—	—
Subtotal		9.66×10^{-12}	$(4.44 \pm 1.07) \times 10^{-12}$	$\leq 7.59 \times 10^{-12\dagger}$
PTFE cathode support				
^{238}U	8.36×10^{-6}	5.33×10^{-12}	$(5.33 \pm 5.33) \times 10^{-12}$	$\leq 2.08 \times 10^{-11\dagger}$
^{226}Ra★	—	—	—	—
^{235}U	—	—	—	—
^{232}Th	5.53×10^{-6}	0	0	$\leq 8.11 \times 10^{-12\dagger}$
^{228}Th★	—	—	—	—
^{40}K	—	—	—	—
^{60}Co	—	—	—	—
^{137}Cs	—	—	—	—
Subtotal		5.33×10^{-12}	$(5.33 \pm 5.33) \times 10^{-12}$	$\leq 3.59 \times 10^{-11\dagger}$
Mylar film (PET)				
^{238}U	1.14×10^{-6}	4.38×10^{-9}	$(1.05 \pm 0.08) \times 10^{-9}$	$\leq 1.15 \times 10^{-9\dagger}$
^{226}Ra★	2.31×10^{-8}	5.08×10^{-11}	$(1.39 \pm 0.10) \times 10^{-11}$	$\leq 1.52 \times 10^{-11\dagger}$
^{235}U	1.14×10^{-8}	5.45×10^{-11}	$(1.99 \pm 0.10) \times 10^{-11}$	$\leq 2.11 \times 10^{-11\dagger}$
^{232}Th	3.14×10^{-8}	7.37×10^{-11}	$(2.38 \pm 0.14) \times 10^{-11}$	$\leq 2.57 \times 10^{-11\dagger}$
^{228}Th★	1.41×10^{-8}	1.46×10^{-11}	$(3.18 \pm 0.32) \times 10^{-12}$	$\leq 3.65 \times 10^{-12\dagger}$
^{40}K	1.29×10^{-7}	2.18×10^{-11}	$(3.21 \pm 0.51) \times 10^{-12}$	$\leq 3.98 \times 10^{-12\dagger}$
^{60}Co	7.05×10^{-9}	4.83×10^{-12}	$(9.01 \pm 0.48) \times 10^{-13}$	$\leq 9.63 \times 10^{-13\dagger}$
^{137}Cs	7.84×10^{-9}	4.06×10^{-12}	$(6.64 \pm 0.58) \times 10^{-13}$	$\leq 7.38 \times 10^{-13\dagger}$
Subtotal		4.54×10^{-9}	$(1.10 \pm 0.08) \times 10^{-9}$	$\leq 1.27 \times 10^{-9\dagger}$

Known-activity subtotal Baseline: 1.09×10^{-7} + BDT: $(2.33 \pm 0.11) \times 10^{-8}$ 90% limit: $\leq 8.16 \times 10^{-8\dagger}$

Source	Parents	Baseline	BDT
Copper backplate ^{238}U	100 000	1	0
Copper chamber ^{238}U	100 000	1	0
Copper readout ^{238}U	5 196 566	2 377	362
Patterned copper cathode ^{238}U	509 459	36	5
Copper detector pipe ^{238}U	100 000	0	0
Local copper box ^{238}U	100 000	0	0
Copper backplate ^{232}Th	100 000	0	0
Copper chamber ^{232}Th	100 000	0	0
Copper readout ^{232}Th	2 172 831	617	118
Patterned copper cathode ^{232}Th	755 532	20	10
Copper detector pipe ^{232}Th	100 000	0	0
Local copper box ^{232}Th	100 000	0	0
Copper backplate ^{40}K	100 000	0	0
Copper chamber ^{40}K	100 000	0	0
Copper readout ^{40}K	100 000	3	0
Patterned copper cathode ^{40}K	100 000	0	0
Copper detector pipe ^{40}K	100 000	0	0
Local copper box ^{40}K	100 000	0	0
Copper backplate ^{60}Co	100 000	0	0
Copper chamber ^{60}Co	100 000	0	0
Copper readout ^{60}Co	100 000	6	1
Patterned copper cathode ^{60}Co	100 000	1	0
Copper detector pipe ^{60}Co	47 100 000	14	2
Local copper box ^{60}Co	47 100 000	11	3
Kapton readout ^{238}U	551 585	17	12
Kapton readout ^{235}U	484 446	30	14
Kapton readout ^{232}Th	610 761	14	7
Kapton readout ^{40}K	100 000	0	0
Kapton readout ^{60}Co	100 000	0	0
Mylar film (PET) ^{238}U	13 108	794	190
Mylar film (PET) ^{235}U	15 525	1 169	426
Mylar film (PET) ^{232}Th	24 194	893	289
Mylar film (PET) ^{40}K	100 000	265	39
Mylar film (PET) ^{60}Co	175 994	1 894	353
Mylar film (PET) ^{226}Ra	20 505	707	193

Continued on next page

Table B.5 continued

Source	Parents	Baseline	BDT
Mylar film (PET) ^{228}Th	27 082	440	96
Mylar film (PET) ^{137}Cs	100 000	813	133
PTFE chamber liner ^{238}U	237 225	16	8
PTFE cathode support ^{238}U	100 000	1	1
PTFE chamber liner ^{235}U	441 754	95	38
PTFE cathode support ^{235}U	100 000	2	1
PTFE chamber liner ^{232}Th	494 819	37	15
PTFE cathode support ^{232}Th	100 000	0	0
PTFE chamber liner ^{40}K	100 000	0	0
PTFE cathode support ^{40}K	100 000	0	0
PTFE chamber liner ^{60}Co	100 000	0	0
PTFE cathode support ^{60}Co	100 000	1	0
PTFE chamber liner ^{226}Ra	392 451	23	16
PTFE cathode support ^{226}Ra	100 000	0	0
PTFE chamber liner ^{228}Th	583 808	27	16
PTFE cathode support ^{228}Th	100 000	3	1
PTFE chamber liner ^{137}Cs	100 000	0	0
PTFE cathode support ^{137}Cs	100 000	0	0
Copper readout ^{235}U	580 000	384	123

Table B.5: Generated-parent exposure and selected endpoints for every detector-material source. The baseline applies the final validity, energy, track, and fiducial requirements; BDT adds the fixed topology threshold. Counts are selected events, so full-chain multiplicity can exceed one per parent. All generated parents, including zero-yield histories, remain in the exposure. Alternative nested full chains are not additive. No veto is applied.

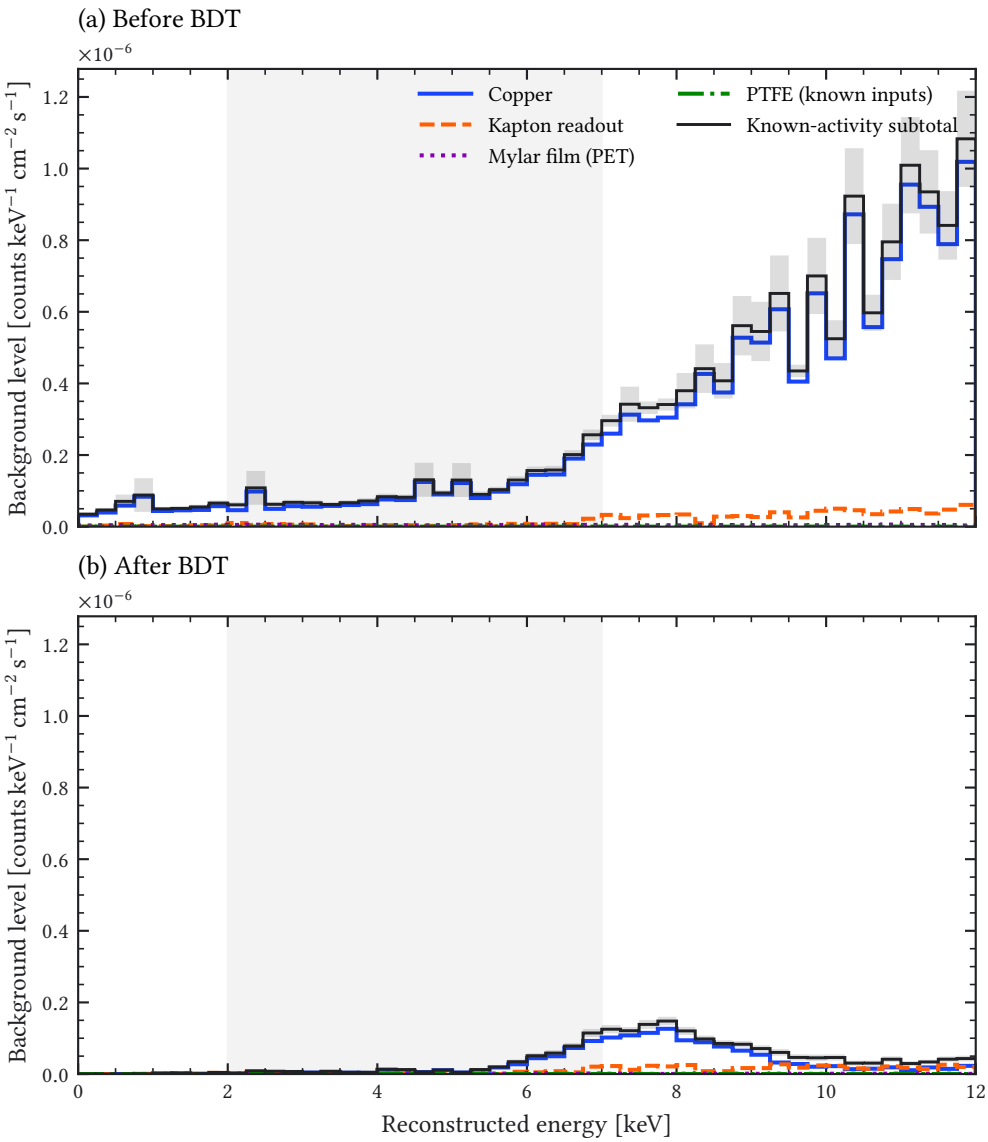


Figure B.1: Reconstructed detector-material spectra before and after the simulation BDT. Both panels require valid reconstruction, track multiplicity, and fiducial containment; the shaded interval is 2–7 keV. The known-activity subset is that of Table B.4; its band gives one parent-correlated MC standard error in populated bins. Empty bins are zero estimates, not upper limits. No veto is applied.

B.3 Lead activity, copper protection, and source depth

The electromagnetic inner-lead source and the full-shield neutron inventory have different masses. The inner shell is a 220 mm × 220 mm × 360 mm box minus a 200 mm × 200 mm × 350 mm cavity displaced by 5 mm along z. At 11.35 g cm^{-3}, its mass is 38.8624 kg. The full lead shield has mass 2088.4 kg and retains its own neutron normalization. The adopted full-shield U/Th/K activities are 0.33 mBq kg^{-1}, 0.10 mBq kg^{-1}, and 1.2 mBq kg^{-1}.

For the same inner-shell ^{210}Pb histories, changing the specific activity from $a_0 = 80$ Bq kg^{-1} to a scales the background, its MC error, and its conditional upper limit by a/a_0. The two scenarios are fully correlated because they reuse the events. The archaeological benchmark, 7.15×10^{-4} Bq kg^{-1}, is the 90% assay ceiling reported for a particular Roman-lead sample [12]; it corresponds to 27.8 mBq in this shell. It is not a measured activity of the proposed shield, and the conditional MC limit is not a joint assay-and-response interval. The outer lead and independent U/Th/K neutron scenarios are not scaled with this ratio.

B.3.1 Full-detector copper control

The copper box has 10 mm walls along x, 25 mm walls along y, a 10 mm negative-z end wall, and an open positive-z end. Replacing only this material by air tests protection against the inner-lead source while preserving the source and transport settings. The nominal sample contains 329 028 353 371 generated parents and the air control 29 688 781 279; each uses its own complete exposure. The control starts from 1455 stored events, gives 1452 reconstructed rows, and retains 1448, 239, 187, and seven events through the validity, energy, track, and fiducial stages. Three omitted raw events were recovered before reconstruction, so missing response is not counted as rejection. The nominal sample has 26 baseline and ten BDT survivors; the air control has seven baseline survivors and none after the BDT.

For geometry g, the activity-normalized rate is

$$B_g = \frac{aM_{\mathrm{inner}}}{\Delta E A_{\mathrm{fid}}} \frac{\sum_i Y_{ig}}{N_g}, \tag{B.3}$$

where Y_{ig} counts selected events of parent i. All selected multiplicities are one in these samples. The joint conditional Poisson rate-ratio construction gives $B_{\mathrm{air}}/B_{\mathrm{Cu}} = 3.0$, with a two-sided 90% interval of 1.3–6.3 before the BDT. After the BDT the interval is 0–3.87, so no copper benefit is resolved at that selection stage. Dividing the two separate rate limits would not produce this ratio interval. The copper's own activity remains a separate source.

Geometry	Baseline	Baseline + BDT	90% limit
Copper box	1.56×10^{-8}	$(6.02 \pm 1.90) \times 10^{-9}$	$\leq 9.27 \times 10^{-9\dagger}$
Air control	4.67×10^{-8}	0	$\leq 1.54 \times 10^{-8\dagger}$

Table B.6: Inner-shell ^{210}Pb background in counts $\mathrm{keV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, with the same source activity and no veto. The BDT errors are MC standard errors; the last column gives individual conditional limits under the counting assumptions of Section 5.1.1.

B.3.2 Depth and numerical transport checks

The selected source positions are recovered from the transport, rather than reconstructed track centers. For the open shell, $d = \max(|x| - 100, |y| - 100, -55.5 - z)$, with coordinates in millimeters, labels nested box layers rather than Euclidean distance to a corner. One BDT survivor originates at $d = 9.08\,\mathrm{mm}$. The selected events therefore do not justify an infinitesimal source surface, and a source restricted to the first 10 mm cannot bound the deeper-lead contribution.

The independent slab calculation samples ^{210}Pb uniformly in lead, then transports its radiation through copper to a vacuum scoring plane. It records first outgoing photon crossings with their parent identities, depths, energies, and directions. All photons from one parent are grouped for MC errors; charged particles are transported but do not enter the quoted photon-flux ratios. Each of the seven configurations contains 5×10^7 fixed-count parents in ten independently seeded shards.

Pb [mm]	Cu [mm]	Production threshold [μm]	Photons	Flux ± MC error
50	0	10	23 736	$(2.16 \pm 0.01) \times 10^{-3}$
50	2	10	14 123	$(1.28 \pm 0.01) \times 10^{-3}$
50	5	10	10 201	$(9.26 \pm 0.09) \times 10^{-4}$
50	10	1	6 775	$(6.15 \pm 0.07) \times 10^{-4}$
50	10	10	6 836	$(6.21 \pm 0.08) \times 10^{-4}$
50	25	10	2 166	$(1.97 \pm 0.04) \times 10^{-4}$
100	10	10	3 396	$(6.17 \pm 0.11) \times 10^{-4}$

Table B.7: Planar photon transport for ordinary lead at $80\,\mathrm{Bq\,kg}^{-1}$. Fluxes are in photons $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, with 5×10^7 parents per configuration. These are first-crossing photon observables, not selected TPC backgrounds.

With 10 mm copper, doubling the lead depth gives a flux ratio 0.994 ± 0.021, and reducing the secondary-production threshold from 10 µm to 1 µm gives 0.991 ± 0.017. No shift is resolved at this MC precision; these checks do not bound transport-model systematics. Replacing the first 10, 20, and 30 mm by the archaeological activity benchmark leaves respectively $(7.5 \pm 0.3)\%$, $(1.7 \pm 0.2)\%$, and $(0.5 \pm 0.1)\%$ of the ordinary-lead photon flux. A 10 mm layer

buried behind 10 mm ordinary lead leaves (94.25 ± 0.28)%; the same replacement at the outer face leaves (99.90 ± 0.04)%. The 90% and 95% source-depth quantiles are 8.6 mm and 12.2 mm, with 95% parent-bootstrap intervals of 8.3–8.9 mm and 11.7–13.0 mm.

These placement scenarios reuse the same histories. For $R = \sum_i w_i Y_i^\gamma / \sum_i Y_i^\gamma$, where w_i is the depth-dependent activity ratio, their MC variance retains the covariance:

$$\widehat{\mathrm{Var}}(R) = \frac{N}{N-1} \frac{\sum_i (Y_i^\gamma)^2 (w_i - R)^2}{(\sum_i Y_i^\gamma)^2}. \tag{B.4}$$

Depth intervals use 1000 Poisson bootstrap resamples of parent clusters. Zero-yield parents enter the absolute exposure, but not the conditional depth distribution. Figure B.3 shows these depth and energy dependences. The finite detector box, openings, acceptance, and TPC selection are absent from the slab; its flux reduction cannot be substituted for the full-detector rate reduction.

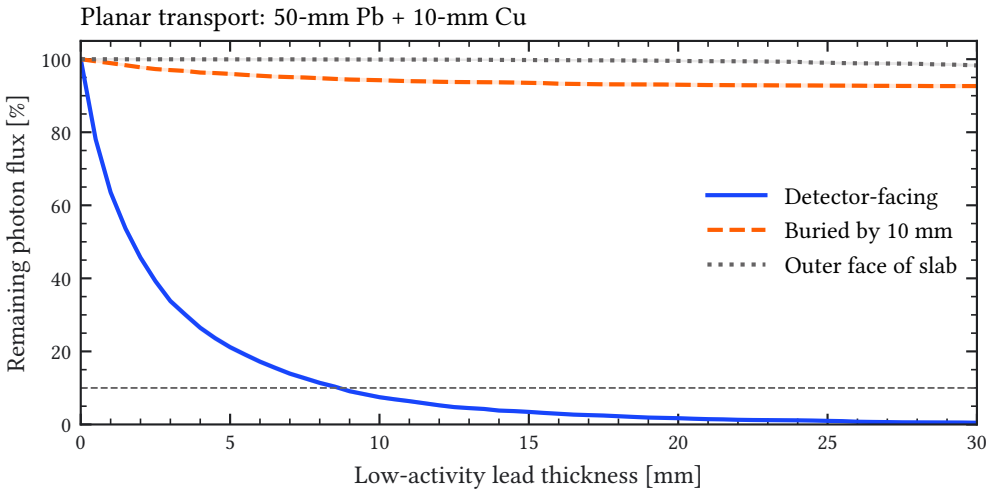


Figure B.2: Remaining photon flux as low-activity lead replaces ordinary lead at different depths in the planar 50 mm Pb, 10 mm Cu geometry. Fluxes are normalized to ordinary lead in the same geometry, before TPC response and selection. Bands show MC standard errors with correlations between activity scenarios retained.

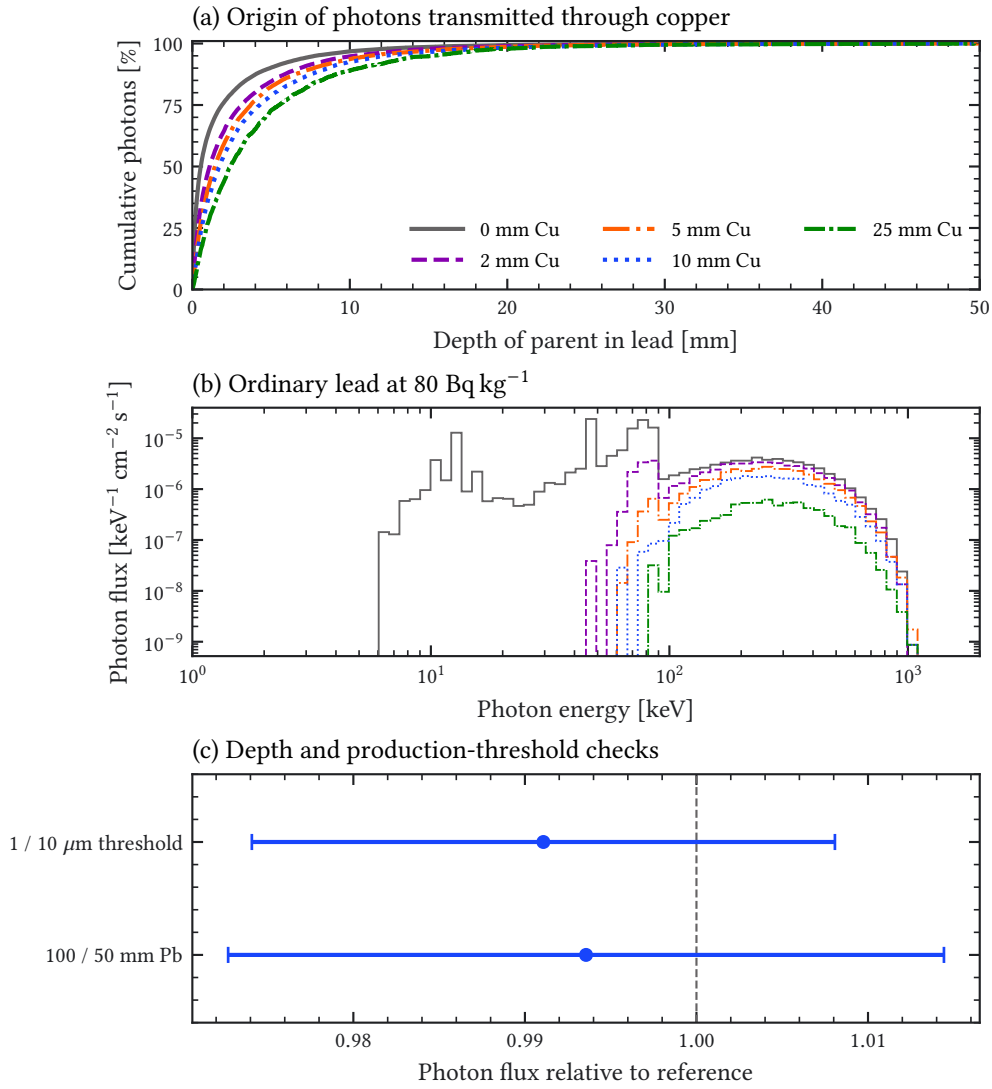


Figure B.3: Planar inner-shield checks. The upper panel gives the cumulative depths of parents producing escaping photons; the middle panel gives their energy spectra for ordinary lead, with a common line pattern for each copper thickness. The lower panel compares deeper lead and tighter secondary-production thresholds with the 50 mm lead, 10 mm copper reference. Errors are MC standard errors. These photon observables precede TPC response and selection.

As an independent scale check, the normal-incidence uncollided transmission is $T(E) = \exp[-(\mu/\rho)(E)\rho t]$, using the NIST total attenuation coefficients [13]. For $t = 1\,\mathrm{cm}$, copper gives $T(80\,\mathrm{keV}) = 1.07 \times 10^{-3}$, whereas lead gives $T(1\,\mathrm{MeV}) = 0.447$. These interaction-free transmissions exclude scattered radiation, fluorescence, bremsstrahlung, and openings; they

are not selected-event probabilities.

B.4 Radiogenic neutrons in the lead shield

The radiogenic-neutron calculation separates nuclear emission from detector response. Its bulk lead reference uses 2088.4 kg, 0.33 mBq kg^{-1} of ^{238}U, and 0.10 mBq kg^{-1} of ^{232}Th. The total parent activities are 689 mBq and 209 mBq; the assumed natural-uranium ratio $A_{235}/A_{238} = 0.046$ gives 31.7 mBq of ^{235}U. The early and late uranium segments are separately in equilibrium at that parent activity. The adopted full chains contain eight, seven, and six alpha decays per ^{238}U, ^{235}U, and ^{232}Th parent, including minor branches in the numerical line lists.

For an alpha line of energy E_ℓ and intensity w_ℓ, the thick-target yield is [14]

$$Y_n = \sum_\ell w_\ell \sum_i \frac{N_A f_i}{M_i} \int_0^{E_\ell} \frac{\sigma_i(E)}{S(E)}\,\mathrm{d}E, \tag{B.5}$$

where f_i is the isotope mass fraction, M_i the molar mass, and $S(E)$ the mass stopping power. Natural abundances are converted consistently to atoms per gram; applying the differential cross sections supplies the neutron spectrum. The inputs are the NeuCBOT snapshot `572afbfbf1b6`, SRIM stopping powers, and TALYS-slim `v0.0.1`. The integration uses a 10 keV alpha grid and endpoint-aware trapezoids, with spectra integrated into 100 keV neutron bins. The larger of the binary-channel and spectrum-integrated normalizations is retained when they differ. Missing or empty reaction and decay inputs are rejected; recovered ^{234}U, ^{230}Th, and ^{218}Rn line lists retain the evaluated source records. Checks against published copper and PTFE yield ranges test the units and scale [15], rather than experimentally validating the nuclear model.

The natural-lead (α, n) thresholds exceed 12 MeV, while every adopted alpha line lies below 9 MeV. Direct production on pure lead is therefore kinematically closed in this model, rather than a zero-count MC result. Alphas escaping into copper, air, or impurities require separate treatment.

B.4.1 Enlarged interface source

The interface calculation deliberately exceeds the actual source within a straight-line continuous slowing-down model. For uniform lead activity, the outgoing alpha intensity is bounded in that model by

$$I_\ell(\boldsymbol{x}, \Omega) \le \frac{a_\ell \rho_{\mathrm{Pb}} R_{\mathrm{Pb}}(E_\ell)}{4\pi}, \qquad R_{\mathrm{Pb}}(E_\ell) = \int_0^{E_\ell} \frac{\mathrm{d}E}{\rho_{\mathrm{Pb}} S_{\mathrm{Pb}}(E)}. \tag{B.6}$$

Here a_ℓ includes the line intensity in Bq g^{-1}. Allowing a whole alpha range in every direction ignores occlusion and energy loss before escape. For target material m, its neutron-energy-bin kernel $K_{m,b}$ gives the enlarged source

$$Q^{\mathrm{env}}_{m,b} = V\rho_m \sum_\ell a_\ell \rho_{\mathrm{Pb}} R_{\mathrm{Pb}}(E_\ell) \max_{0\le E\le E_\ell} K_{m,b}(E). \tag{B.7}$$

The maximum includes the endpoint cell and bounds the discretized spectrum under isotropic neutron emission. Both target kernels are assigned throughout the complete 135 cm × 135 cm × 300 cm world, including fictitious copper or air positions; the transported material geometry remains unchanged. Alpha straggling and neutron anisotropy remain model assumptions outside this bound.

For this screening alone, ^{210}Pb at 80 Bq kg^{-1} is extended throughout the lead with ^{210}Po in equilibrium. The equilibrium uranium-chain ^{210}Po is subtracted before adding the excess chain, preventing double counting. This is broader than the electromagnetic inner-shell scenario and does not bound unsupported surface contamination. The enlarged copper and air source rates are 1.63×10^{-4} s^{-1} and 3.15×10^{-4} s^{-1}; they are screening normalizations, rather than laboratory neutron-emission predictions. The physical 43.1872 kg copper-liner bulk reference gives 3.04×10^{-10} s^{-1} at its separate copper activities. Unspecified impurities, seams, and exposed polymers remain outside the inventory.

B.4.2 Transport and selection

Each interface or copper-bulk source contains 400 000 independent emitted neutrons in four fixed-count jobs. The independent smoke tests are excluded from these exposures. The response uses the conventional REST detector response with Ar/isobutane 98/2, 1.4 atm, 293.15 K, 131.666 V cm^{-1}, and $W = 26.238$ eV. The existing Lindhard process precedes diffusion and waveform reconstruction; this is not an independent recoil calibration. All transported subevents are included, and the selected response is events per emitted neutron, rather than the probability of at least one event. No veto rejection or irradiation history is added.

Source	Neutrons	Reco.	Energy	Tracks	Baseline
Pb alpha to Cu	400 000	0	0	0	0
Pb alpha to air	400 000	0	0	0	0
Copper-liner bulk	400 000	38	5	4	0
^{238}U fission	1 000 000	22	4	1	0

Table B.8: Radiogenic-neutron selection endpoints. Energy and Tracks include all preceding baseline requirements; the baseline adds fiducial containment. All 38 copper events are valid, whereas 21 of the 22 fission events are valid. None passes the baseline, so no additional BDT rejection is measured. The exposure counts emitted neutrons, rather than radioactive parents or whole fissions.

For source s, $B_s = Q_s \sum_j n_{s,j}/(N_s \Delta E A_{\mathrm{fid}})$, with parent-correlated moments and the dagger convention above. The fission background has the individual one-sided 90% counting-model limit $B_{90}^{\dagger} \leq 1.11 \times 10^{-13}$ counts keV^{-1} cm^{-2} s^{-1}, already at the baseline stage. The four-source family limit uses four one-sided 97.5% component bounds before addition, as reported in Table 5.20. This allocates the family tail probability within the counting models, without assigning a confidence level to contamination or nuclear-data assumptions. Neutron histories overlapping an inclusive radioactive-chain sample must be removed before this separate calculation is added.

B.5 Activation normalization and decay response

Production is normalized to all generated primaries, including those without gas deposits, separately from isolated-decay response. IAXO-D1 uses the archived HENSA outdoor spectrum above 1 MeV and its bin-integrated fluence convention, including moderation and capture but excluding lower-energy incidence. One neutron can produce prompt and later acquisitions; cosmic acquisitions alone do not determine accumulated activity.

B.5.1 Angular normalization and source closure

The source reader samples archived bin masses H_{ab} as $H_{ab}/\cos\theta_b$, where θ_b is the zenith-bin center. All 100 saved source histograms reproduce this transformation. The restricted proposal fluence is 1.0380904×10^{-2} cm^{-2} s^{-1}, on a sphere of radius 1777.9904 mm and projected area 99 313.5978 cm^2. Its proposal rate is $Q_{\mathrm{prop}} = 1030.964\,95$ s^{-1}. Each product and complete descendant chain inherits $c_k = \cos\theta_{b(k)}$ from its incident neutron, giving the equilibrium activity contribution

$$A_{\mathrm{eq}} = \frac{Q_{\mathrm{prop}}}{N_{\mathrm{gen}}} \sum_k c_k. \tag{B.8}$$

The target incident rate is 687.270 68 s^{-1}. Per-history weighting restores the scalar energy and angular bin masses while retaining their response correlations. The downward target distribution $\sin\theta\cos^2\theta$ is assumed, rather than measured by HENSA.

All 1 400 000 primaries have unique within-file identities. Positions match the source sphere to 1.9×10^{-8} relative precision; the mean squared impact radius, $0.499899R^2$, agrees with a uniform disk. Two independent 700 000-primary blocks give weighted central-square crossing probabilities 0.238398 and 0.238310, versus 0.238731 expected: deviations of 0.58 and 0.74 standard errors. These checks test reference conversion and geometry, excluding validation of the DESY field or missing low-energy incidence.

B.5.2 Nuclear states, feeding, and response trials

The installed data are `G4ENSDFSTATE2.3` and `RadioactiveDecay5.6` [4]. State identifiers separate ground states and isomers, including ^{60}Co and the 0.806 s ^{207}Pb isomer with a stable ground state. Saved daughters verify complete chains: states above the 100 µs mean-lifetime subevent boundary are counted separately; shorter cascades remain in the isolated parent response.

Reaction ancestry excludes unchanged target ground states, elastic recoils and electromagnetic conversion products; unresolved targets remain conservative inventory entries, and named tritons denote ^{3}H. Parents censored by the 10^{27} ns decay-time limit are excluded from completed trials and replaced by fixed-count trials at a raised limit. Missing terminal chains are completed once, inheriting ancestor lifetimes and angular weights. Ten unsupported ground-state double-beta modes retain an efficiency-unity envelope; the corrected ^{10}C lookup and chain repairs preserve physics data and selection, excluding incomplete responses.

States are sampled uniformly within logical production volumes, with equal quadrature weights for at most six approximately equally spaced physical copies; no activation-density map is measured. Livermore electromagnetic and radioactive-decay transport includes parent and prompt daughters, suppressing separately inventoried long-lived daughters and omitting secondary hadronic production. At most one event per modeled trial is stored; identities are reconciled through reconstruction, baseline and BDT. The independent cosmic evaluation's common baseline and frozen BDT (Section 5.1.2) are used without retuning. Predeclared efficiency denominators include untriggered decays; no prompt or neutron-veto rejection is applied, although decay radiation can reach a scintillator.

Fixed generated and selected counts give each independent state–placement cell an exact one-sided binomial response bound. A common tail allocation gives their weighted sum simultaneous 90% response-only coverage conditional on observed production, feeding and spatial quadrature; unmapped volumes and unsupported modes retain efficiency-unity allowances. Production sampling, absent channels, nuclear data, incident spectrum, spatial modeling and response transfer are excluded. No sensitive-gas product was observed; this

does not establish zero gas activation. The 249 entries without a recorded volume have air-component targets outside the imported GDML world; unresolved coordinate and retention assumptions prevent discarding or assigning them to detector materials.

B.5.3 Finite irradiation and equilibrium

For isotope/state i in volume v, production, daughter feeding and removal give

$$\frac{dN_{i,v}}{dt} = P_{i,v}(t) + \sum_j b_{j\to i}\lambda_j N_{j,v} - (\lambda_i + \kappa_{i,v})N_{i,v}, \qquad \lambda_i = \frac{\ln 2}{T_{1/2,i}}. \tag{B.9}$$

Here P is direct production, b daughter feeding and κ removal, approximately q/V for well-mixed gas of volume V flowing at rate q. For a retained isotope without feeding and with zero initial inventory, constant irradiation gives $A_i = P_i(1 - e^{-\lambda_i t_{\mathrm{irr}}})$; source-off cooldown adds a factor $e^{-\lambda_i t_{\mathrm{cool}}}$. Acquisition stops do not end irradiation; pre-existing activity enters through $N_i(0)$. The reference assumes complete retention and zero initial induced activity.

For constant irradiation, S_i sums the independent exponential waiting times of state i and its radioactive ancestors. The branch activity is its stationary production weight times $\Pr(S_i \le t)$, evaluated by extended-precision convolution rather than an isolated $1 - e^{-\lambda_i t}$ factor that omits feeding delays. Single- and two-state analytic checks pass, and the sums reconcile all 188 974 nuclear-state entries. Table B.9 gives the resulting timescales.

State / inventory	1 year	3 years	5 years	10 years	Equilibrium
			Activity (Bq)		
^{60}Co ground	7.63×10^{-3}	2.02×10^{-2}	2.98×10^{-2}	4.53×10^{-2}	6.31×10^{-2}
^{60m}Co	2.91×10^{-2}	2.91×10^{-2}	2.91×10^{-2}	2.91×10^{-2}	3.04×10^{-2}
^{58}Co ground	7.31×10^{-2}	7.52×10^{-2}	7.52×10^{-2}	7.52×10^{-2}	7.52×10^{-2}
^{58m}Co	3.55×10^{-2}	3.55×10^{-2}	3.55×10^{-2}	3.55×10^{-2}	3.55×10^{-2}
^{3}H	1.28×10^{-1}	3.64×10^{-1}	5.75×10^{-1}	1.01×10^{0}	2.35×10^{0}
^{14}C	1.51×10^{-5}	4.54×10^{-5}	7.56×10^{-5}	1.51×10^{-4}	1.24×10^{-1}
^{36}Cl	2.53×10^{-5}	7.60×10^{-5}	1.27×10^{-4}	2.53×10^{-4}	1.10×10^{1}
^{64}Cu	1.55×10^{0}	1.55×10^{0}	1.55×10^{0}	1.55×10^{0}	1.55×10^{0}
All states	5.89×10^{1}	5.99×10^{1}	6.04×10^{1}	6.12×10^{1}	9.33×10^{1}

Table B.9: Retained activity under uninterrupted irradiation by the existing HENSA reference field. Initially induced inventories are zero. Daughter activities include all recorded ancestor lifetimes. Equilibrium is the formal limit with fixed target composition and no removal; long-lived isotopes need not approach it on operational timescales.

Formal equilibrium assumes unlimited irradiation and fixed targets; tritium reaches only about one quarter of saturation after five years. Gas flow, inter-volume transport and pre-existing contamination require removal terms or initial conditions; assayed activity must not

be added again as cosmogenic activity. The response-coverage envelope is not an activation upper limit in the chapter budget. Addition to prompt neutrons requires compatible fields and disjoint acquisitions.

B.5.4 Production inventory and selected-decay response

The following tables summarize the reference inventory and response; its selected background comes from ground-state cobalt in copper.

An independent sample also selects one ^{62}Cu and one ^{64}Cu decay in complete cells. The equilibrium subtotal over complete state–placement cells is dominated by these copper decays and equals $(2.57 \pm 1.81) \times 10^{-9}$ counts keV^{-1} cm^{-2} s^{-1}; its nominal response MC standard error assumes fixed activities and excludes unquantified selection bias. Fit, lifetime and navigation diagnostics leave other cells unqualified; failed responses are not zero observations. This subset diagnostic cannot replace the reference estimate or establish a complete activation rate.

Isotope	Main region	Direct	Chain-fed	Total	5 years	Equilibrium
		Produced states (counts)			Activity (Bq)	
^{36}Cl	Veto wrapping	22 254	0	22 254	1.27×10^{-4}	1.10×10^{1}
^{203}Pb	Lead shielding	6 135	9 870	16 005	7.87×10^{0}	7.87×10^{0}
^{204}Pb	Lead shielding	9 071	4 450	13 521	1.44×10^{0}	6.64×10^{0}
^{205}Pb	Lead shielding	6 590	6 239	12 829	1.62×10^{0}	6.31×10^{0}
^{202}Pb	Lead shielding	5 530	2 819	8 349	1.47×10^{0}	4.08×10^{0}
^{202}Tl	Lead shielding	1 551	6 367	7 918	1.28×10^{0}	3.86×10^{0}
^{201}Pb	Lead shielding	3 941	3 245	7 186	3.51×10^{0}	3.51×10^{0}
^{200}Tl	Lead shielding	1 451	4 741	6 192	3.03×10^{0}	3.03×10^{0}
^{201}Tl	Lead shielding	1 423	4 449	5 872	2.86×10^{0}	2.86×10^{0}
^{3}H	Veto scintillator	4 477	134	4 611	5.75×10^{-1}	2.35×10^{0}
^{199}Pb	Lead shielding	2 036	1 881	3 917	1.94×10^{0}	1.94×10^{0}
^{199}Tl	Lead shielding	1 379	2 375	3 754	1.84×10^{0}	1.84×10^{0}
^{197}Tl	Lead shielding	1 020	2 583	3 603	1.75×10^{0}	1.75×10^{0}
^{198}Tl	Lead shielding	1 145	2 397	3 542	1.74×10^{0}	1.74×10^{0}
^{200}Pb	Lead shielding	3 361	67	3 428	1.69×10^{0}	1.69×10^{0}
^{64}Cu	Cu components	3 156	0	3 156	1.55×10^{0}	1.55×10^{0}
^{197}Hg	Lead shielding	399	2 291	2 690	1.31×10^{0}	1.31×10^{0}
^{60}Co	Cu components	130	63	193	5.90×10^{-2}	9.35×10^{-2}
^{58}Co	Cu components	100	117	217	1.11×10^{-1}	1.11×10^{-1}
^{10}C	Veto scintillator	214	2	216	1.13×10^{-1}	1.13×10^{-1}
Remaining isotopes		34 006	25 515	59 521	2.47×10^{1}	2.96×10^{1}
All long-lived states		109 369	79 605	188 974	6.04×10^{1}	9.33×10^{1}

Table B.10: Activation production and retained activity for the existing HENSA reference spectrum, from 1 400 000 incident neutrons. Count columns are raw simulated nuclei; Bq columns include the incident-rate and angular corrections and daughter buildup. Isotope totals sum distinct nuclear states and all regions; main region gives the largest weighted production contribution. The remaining isotopes contribute 31.8% of equilibrium activity. Equilibrium is the formal saturation limit, not five years of exposure.

The total activity reaches 60.4 Bq after five years and 93.3 Bq at formal equilibrium. To obtain the low-energy background, Table B.11 combines these activities with the response

from 106 790 930 isolated-decay trials for 1446 observed state–volume pairs.

Isotope	Main region	Generated decays	Stored TPC trigger	TPC baseline	+ BDT	Background at equilibrium counts keV^{-1} cm^{-2} s^{-1}	Upper bound
^{60}Co	Cu components	36 099 768	18 517	18	7	1.42×10^{-10}	$\leq 2.10 \times 10^{-6}$
^{58}Co	Cu components	44 811 697	6 504	13	2	8.84×10^{-11}	$\leq 1.48 \times 10^{-6}$
^{10}C	Veto scintillator	64 800	0	0	0	0	$\leq 4.18 \times 10^{-5}$
^{64}Cu	Cu components	347 425	197	0	0	0	$\leq 4.49 \times 10^{-5}$
^{3}H	Veto scintillator	540 000	0	0	0	0	$\leq 1.56 \times 10^{-4}$
^{115}In	Capture layers	46 800	0	0	0	0	$\leq 4.20 \times 10^{-4}$
^{195}Hg	Lead shielding	2 700	0	0	0	0	$\leq 3.58 \times 10^{-4}$
^{203}Pb	Lead shielding	44 430	0	0	0	0	$\leq 3.57 \times 10^{-4}$
^{206}Pb	Lead shielding	1 800	0	0	0	0	$\leq 3.13 \times 10^{-4}$
^{36}Cl	Veto wrapping	339 342	0	0	0	0	$\leq 2.72 \times 10^{-4}$
Remaining states		24 492 168	1 194	0	0	0	$\leq 6.68 \times 10^{-3}$
Modeled inventory		106 790 930	26 412	31	9	2.31×10^{-10}	$\leq 8.64 \times 10^{-3}$
Unmodeled inventory		—	—	—	—	—	$\leq 1.84 \times 10^{-2}$
Combined envelope		—	—	—	—	—	$\leq 2.71 \times 10^{-2}$

Table B.11: Equilibrium activation background after the common TPC baseline and BDT, without veto rejection. Modeled bounds are simultaneous 90% response-only limits conditional on observed production and spatial assumptions. The unmodeled row assumes selection efficiency at most unity; the combined envelope includes both terms. A zero estimate does not establish zero background.

The modeled equilibrium background is 2.31×10^{-10} counts keV^{-1} cm^{-2} s^{-1}, from cobalt decays. After five and ten years it is 1.57×10^{-10} and 1.92×10^{-10} in the same units, respectively 68% and 83% of the equilibrium estimate. These are instantaneous rates, not exposure averages. The five-year combined conditional envelope is $\leq 1.52 \times 10^{-2}$ in the same units. The modeled value is a partial estimate, not an established total. Large bounds arise from sparse zero-survivor state–placement samples: 3815 of 3821 cells have no selected event, and the shared tail allocation gives a zero-count efficiency bound near $10.55/N$. Millions of ground-state cobalt trials do not tighten its sparsely sampled isomer response. The 589 unmodeled production entries retain an efficiency-unity allowance, not a predicted rate. Production sampling, unobserved production channels and nuclear/source/spatial-model uncertainties are outside these bounds; negligible activation is not demonstrated.

B.6 Environmental-source and transport checks

Auxiliary concrete-emission study. An auxiliary `Geant4` calculation generated ^{40}K, ^{232}Th, ^{235}U and ^{238}U decays in a 1 m-thick, transversely extended concrete slab. Its approximately cosine-weighted escaping radiation illustrates material emission; the detector calculations use NaI and HENSA reference fields.

The original gamma continuum and nine lines used the NaI fit without a shielded control; their transfer to the paired source is given below. The earlier indoor and outdoor

neutron calculations used positive HENSA-minus-CRY residuals below 10 MeV. HENSA bin fluences were converted from hours to seconds, and rebinned CRY was normalized in the measured 20 MeV–10 GeV tail before subtraction. Clipping negative differences does not isolate terrestrial emission: moderated cosmic and room-return neutrons can remain. These alternative fields are not added to each other or to an inclusive cosmic source.

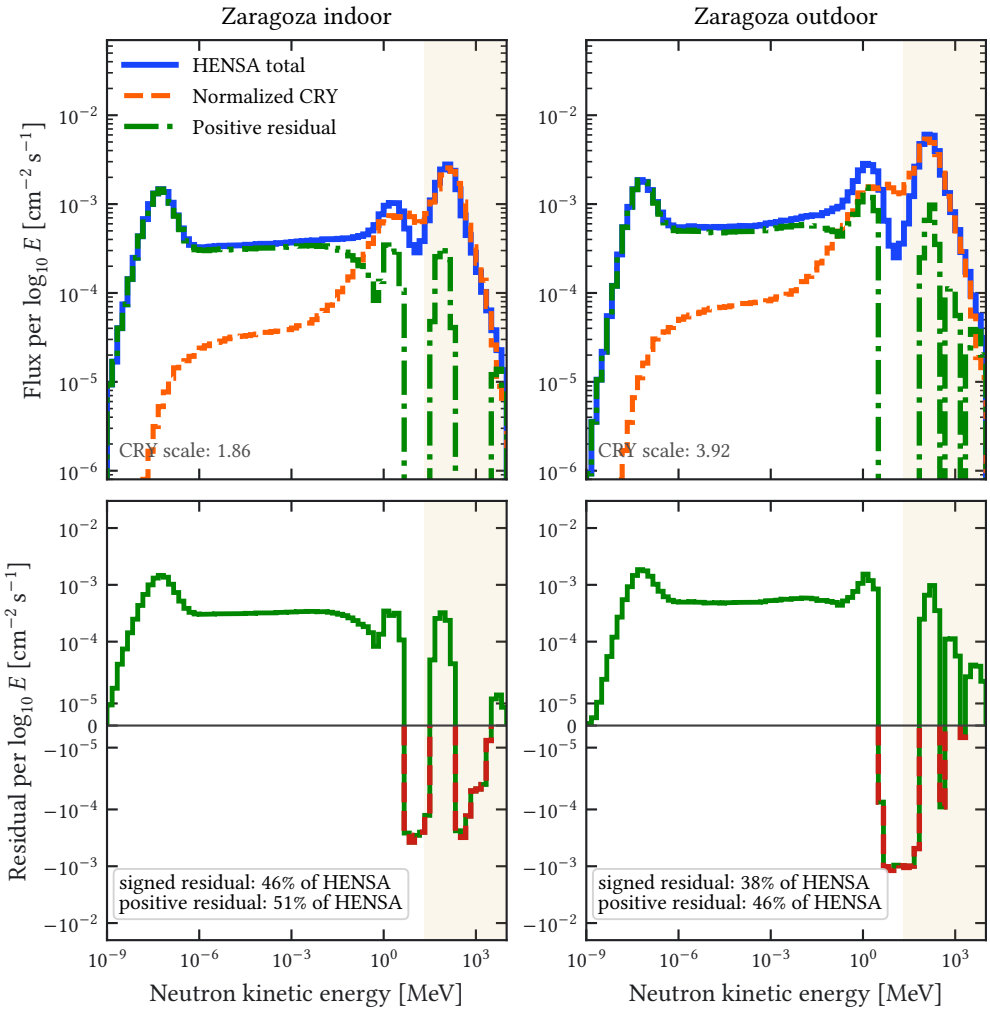


Figure B.4: Earlier HENSA-minus-CRY source construction for the Zaragoza reference fields. Upper panels compare measured, scaled cosmic and positive residual spectra; the shaded interval fixes the cosmic normalization. Lower panels retain the signed difference. The retained detector calculations sampled only the positive residual below 10 MeV.

B.6.1 Paired NaI reference and calibration

The joint fit uses 92 h without Pb and 240 h with lead bricks. A separate two-hour control acquisition with a different slow-channel threshold is retained only as a sensitivity case. Raw acquisition sums reproduce the ROOT count histograms; defective merged MCA files, which combine accumulated time with one acquisition's counts, are excluded. For the existing bare-NaI photon response $S(\boldsymbol{\phi})$ and a shared nonnegative residual $b(E)$, piecewise constant over approximately 20 keV, the original counts follow

$$n_i \sim \mathrm{Pois}\left[t_0\left(S_{0i}(\boldsymbol{\phi}) + b_{0i}\right)\right], \tag{B.10}$$

$$m_j \sim \mathrm{Pois}\left[t_{\mathrm{Pb}}\left(\lambda S_{\mathrm{Pb},j}(\boldsymbol{\phi}) + b_{\mathrm{Pb},j}\right)\right]. \tag{B.11}$$

Both response and residual are integrated over each measurement's own calibrated bins, without interpolating counts. The nominal $\lambda = 0$ assigns the control to the shared residual; other fixed values assume scalar response transfer, not measured Pb transmission. Incomplete coverage, scattering and temporal variation prevent identifying the residual with intrinsic activity alone. Independent arbitrary leakage and residual spectra would leave the external field unidentified.

Source-calibration peaks do not transfer accurately to the background runs. We therefore align their ^{40}K and ^{208}Tl peaks in sample, without independent validation of linearity. The control's ^{40}K centroid drifts by about 23 keV between its first and last six acquisitions. Table B.12 compares calibration and control choices at zero assumed leakage. Eighty bootstrap repetitions resample both original channel spectra and response MC cells, refit the calibration and preserve source-parameter correlations. Their intervals exclude gain drift, peak-model discrepancy, leakage and Pb-induced response changes. The large joint deviance, approximately 7.86×10^4, limits the source to a conditional model.

Calibration / control choice	**Fluence [cm^{-2} s^{-1}]**
Nominal	2.474
Wider peak windows	2.517
First control half	2.463
Last control half	2.480
All control acquisitions	2.474
Previous energy axes	2.442
November energy axes	2.526
Source-only calibration	2.642

Table B.12: Conditional scalar photon fluence at zero assumed control leakage. Entries are alternative fits, not systematic confidence bounds.

B.6.2 Boundary normalization and complete reference exposure

For an isotropic scalar differential fluence $\Phi(E)$, the inward current is $\Phi(E)/4$. Uniform surface positions and cosine incidence, $p(\theta) = 2\sin\theta\cos\theta$, give the source rate

$$R = \frac{S}{4}\int \Phi(E)\,\mathrm{d}E, \qquad R(r) = \pi r^2 \int \Phi(E)\,\mathrm{d}E \tag{B.12}$$

for a general surface of area S or a sphere of radius r, respectively. A radial inward beam would describe a different field. Reducing a source sphere from 5 m to 1.8 m changes its rate by 0.1296 and encloses the reference world, whose circumscribed radius is 1777.99 mm. The change is equivalent only if the radiation delivered to the geometry is preserved; laboratory structures absent between the boundaries do not acquire a transport model through an area rescaling.

The complete fitted-gamma reference contains 11 853 754 133 photons and 31 reconstructed events: 21 continuum, four ^{40}K, one 1.764 MeV ^{214}Bi, and five 2.615 MeV ^{208}Tl events. Eight pass the energy stage, three the track stage, and none the fiducial requirement. The complete generated exposure includes histories with no saved detector interaction. Each component retains its original proposal rate R_k and completed count N_k. Let $\rho_k(E)$ be the paired-source to original-source ratio, using differential fluences for the continuum and the individual line fluences for discrete components. For the rare independent single-event counting model, a conservative combined zero-count bound is

$$B_{90} = -\ln(0.1)\max_k \frac{R_k \sup_E \rho_k(E)}{N_k \Delta E A_{\mathrm{fid}}} = 9.40150\times10^{-5}\ \mathrm{counts\,keV^{-1}\,cm^{-2}\,s^{-1}}. \tag{B.13}$$

The quoted value rounds upwards to 9.41×10^{-5}; the earlier-source value was 3.27294×10^{-5}. The least-sensitive exposure is the 0.609 MeV ^{214}Bi line, whose fluence increases by a factor of 2.87249; this is not an observed dominant background. The continuum supremum, 1.38444, is evaluated at all original bin boundaries and fitted log-linear knots, rather than inferred from surviving events. The construction uses the maximum possible source weight without adding component-wise 90% limits; it excludes source uncertainty and retains the original counting and stopping assumptions. The separate enclosing-sphere neutron references contain 37 887 020 indoor and 36 303 029 outdoor primaries, with five and one reconstructed events, respectively, and no baseline survivor. Their individual conditional limits are those of Table B.13. No additional BDT rejection is measured in any of these zero-baseline samples. An audit found that the earlier histogram reader used the spacing of the first two bin centers as its within-bin sampling width for every selected bin. For logarithmic neutron bins this differs from sampling over the actual bin boundaries. The native enclosing-box calculation below uses the actual boundaries while retaining the integrated bin probabilities; the two source implementations are not pooled.

Sphere reference	Generated	Reco.	90% upper limit
Fitted gammas	11 853 754 133	31	$\leq 9.41 \times 10^{-5\dagger}$
Indoor neutrons	37 887 020	5	$\leq 1.10 \times 10^{-5\dagger}$
Outdoor neutrons	36 303 029	1	$\leq 1.85 \times 10^{-5\dagger}$

Table B.13: Complete enclosing-sphere environmental references with zero baseline survivors. Limits are in counts $\mathrm{keV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, under the rare-count working models of Section 5.1.1, with no veto. The source rates are respectively $1.943 \times 10^6\ \mathrm{s}^{-1}$, $2.834 \times 10^3\ \mathrm{s}^{-1}$, and $4.564 \times 10^3\ \mathrm{s}^{-1}$; the gamma bound transfers the individual exposures to the paired NaI source using their maximum possible weights. The neutron fields are alternatives. The large positive limits indicate limited MC sensitivity, rather than a predicted rate at that level.

B.6.3 Weighted transport and independent confirmation

Direct exposure sufficient for a 10^{-9} counts $\mathrm{keV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ zero-count limit is computationally expensive for these source rates. We therefore tested complete-history importance sampling while preserving the physical gas deposits and native response. Positions facing the beam pipe and their angular complement, together with impact-parameter intervals, partition the complete incident field. For a region defined by $c = \cos\theta$ and impact parameter b, its fraction is

$$f_j = \frac{c_{\max} - c_{\min}}{2} \frac{b_{\max}^2 - b_{\min}^2}{r^2}. \tag{B.14}$$

All region fractions sum to one; neither off-axis scattering nor secondary photons are omitted. Physical energies are reconstructed before the history likelihood W_{ij} is applied. If K_{ij} counts selected physical subevents of original primary i, the estimator is

$$\widehat{B} = \frac{R_{\mathrm{in}}}{\Delta E A_{\mathrm{fid}}} \sum_j \frac{f_j}{N_j} \sum_{i=1}^{N_j} W_{ij} K_{ij}. \tag{B.15}$$

All generated primaries, including those without gas deposits, enter N_j; descendants remain grouped by original primary for the variance. Weighting gas energy before selection, or treating split descendants as independent physical acquisitions, would change the result.

Source-current controls, populated analog comparisons, physical energy-window controls, and paired native-response replays check the implementation. One paired check reproduces all 77 scalar TPC observables for each of two neutron events with matched response seeds. Those events contain no quenched hit, so that comparison does not validate the nuclear-quenching branch. The configured native and standalone gamma/neutron particle-process inventories also match; agreement of configuration does not establish agreement of rare selected tails. Failed controls and unfinished transport jobs remain excluded from completed exposure.

The independent confirmation comprises eight replicas of all twelve regions for each field, totaling 24 192 000 primaries. The development sample is excluded. All 288 jobs complete transport and native reconstruction; transport consumes 6.23 core-hours and the summed worker time including response is 6.67 hours. The selected populations are preserved in Table B.14.

Source	Generated	Reco.	Baseline	BDT
Gammas	17 280 000	2 739	114	17
Indoor neutrons	3 456 000	51	1	1
Outdoor neutrons	3 456 000	70	1	0

Table B.14: Complete independent near-boundary confirmation. Counts before likelihood weighting are biased-sample populations, rather than physical efficiencies. All regions and physical gas inputs are accounted for; the two neutron fields are alternatives.

For the earlier source without the shielded NaI reference, the post-BDT gamma estimate is 1.22×10^{-10} counts $\mathrm{keV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, with parent-cluster and between-replica standard errors of approximately the same size. One parent supplies 99.9989% of the weighted result, despite a gamma survivor in every replica. The indoor neutron survivor has a very small weight, while the outdoor sample has none. Neither these observations nor empirical error bars bound unobserved weight tails. No numerical 90% background limit is assigned, and the 10^{-9} sensitivity objective remains unverified.

B.6.4 Shield-facing gamma confirmation

This separate calculation assigns incident photons to the $2\,\mathrm{m}^2$ outer-lead envelope, excluding the frontal opening and rays whose backward continuation crosses the virtual pipe boundary. For the earlier source at scalar fluence $2.743\,63\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, its cosine-incidence proposal rate is $13\,718.16\,\mathrm{s}^{-1}$. It is a local isotropic-boundary model; a common reciprocal first-entry assignment with the full pipe source and equivalence to the enclosing laboratory sphere are not established. Two angular regions have physical fractions 0.0151901 and 0.9848099, independent of their deliberately balanced sampling. Whole-history likelihoods are applied after conventional REST reconstruction, and no veto rejection is credited.

The completed comparison uses 32 000 000 primaries per proposal. Lead-only biased transport gives three baseline survivors and none after the BDT; adding the gas-interaction proposal gives 136 baseline and 25 BDT survivors. The latter estimate is $(4.12 \pm 3.74) \times 10^{-10}$ counts $\mathrm{keV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, with one parent supplying 90.46%. The independent between-replica error is 3.73×10^{-10} in the same units. The proposals estimate the same source and are not additive.

An energy-resolved confirmation retains 467 completed jobs and 934 000 000 primaries from the original 1 000 000 000-primary allocation. Every energy–angle stratum uses its actual

completed denominator and flux fraction; the incomplete allocation is not described as a completed campaign. The post-BDT estimate is $(1.61 \pm 1.22) \times 10^{-9}$ counts keV^{-1} cm^{-2} s^{-1}, with a between-job error of 1.21×10^{-9}. The 45 complete ten-member replica sets give $(1.64 \pm 1.23) \times 10^{-9}$. One verified ^{208}Tl history supplies 74.94% of the estimate; the ^{208}Tl and high-energy continuum strata supply about 83.1% and 15.9% of the observed rate. This concentration limits the precision of the weighted result. No numerical 90% limit is inferred from a Gaussian endpoint, an effective sample size, or a zero-count stratum.

The saved incident energy of each contributing physical parent permits reweighting to the paired NaI source used in Chapter 5. Discrete lines use their individual fluence ratios; continuum events use the fitted differential fluence divided by the original within-bin proposal density. All completed primary denominators, history weights and event selections remain fixed. The earlier baseline and BDT estimates are reproduced before changing source weights. For the nominal paired source, the estimates become $(4.51 \pm 2.46) \times 10^{-9}$ and $(1.42 \pm 1.15) \times 10^{-9}$ in units of counts keV^{-1} cm^{-2} s^{-1}, respectively; these errors describe empirical physical-parent transport statistics. The conditional source-only 68.27% intervals from the correlated NaI bootstrap samples are $[4.445, 4.581] \times 10^{-9}$ and $[1.404, 1.429] \times 10^{-9}$, with correlation 0.411, at fixed detector transports and zero control leakage. The largest parent supplies 81.01% of the reweighted BDT estimate, so the source correction does not resolve the rare-history uncertainty or establish an upper limit. The resolved pipe sample selects cap photons whose initial straight-line continuation intersects the entrance window; the closed cap and secondary transport remain. The reweighted baseline and post-BDT estimates are $(5.23 \pm 0.36) \times 10^{-8}$ and $(5.85 \pm 1.20) \times 10^{-9}$ in the same units; 22 parents remain after the measured-activity overlay, giving $(5.36 \pm 1.15) \times 10^{-9}$. At fixed transports, the conditional source-only 68.27% interval on the last result is $[5.298, 5.413] \times 10^{-9}$. The conservative direct-region 90% bound multiplies the original single-event Poisson mean bound for 22 survivors by the original per-event normalization and the global source-ratio supremum 2.87249. It gives 2.22×10^{-8}, excluding source uncertainty and retaining the original counting assumptions. For the four zero-survivor complementary regions, the same maximum-weight construction gives a separate joint bound of 1.07×10^{-4}, dominated by the lateral region; it excludes the direct region.

B.6.5 Native residual-neutron production

The earlier calculation samples the indoor and outdoor residual-bin boundaries, with scalar fluences 3.6087867×10^{-3} and 5.8106865×10^{-3} cm^{-2} s^{-1}. Area-weighted faces, uniform positions and inward $p(\mu) = 2\mu$ on the 135 cm × 135 cm × 300 cm box ($S = 198\,450$ cm^2) give rates 179.040 93 s^{-1} and 288.282 69 s^{-1}. Transport retains the moderator, veto, shielding, air and secondary photons. The outdoor source ends at 3.162 MeV; these limits do not bound the full outdoor HENSA field or the separate terrestrial contribution.

The conventional TPC response and BDT use matched veto timing and measured

accidentals, without optical-photon or Garfield transport. Delayed events receive veto credit only with demonstrated coincidence. Table B.15 includes every completed independent primary, even without gas deposits: $T_{\mathrm{eq}} = N_{\mathrm{gen}}/R_{\mathrm{gen}}$, $B_{90} = \ln(10)/(T_{\mathrm{eq}}\Delta E A_{\mathrm{fid}})$. Unfinished blocks provide no exposure; these conditional bounds lack stopping-uniform coverage. Box and sphere calculations remain separate.

Field	Generated	Reco./Valid	Energy	Tracks	Fid./BDT	90% limit
Indoor	96 000 000	11/11	3	1	0/0	$\leq 2.74 \times 10^{-7\dagger}$
Outdoor	179 200 000	27/25	7	6	0/0	$\leq 2.36 \times 10^{-7\dagger}$

Table B.15: Cumulative counts and conditional limits for the earlier alternative residual sources. Paired columns give reconstructed/valid and fiducial/BDT counts. All TPC BDT, prompt-veto and neutron-veto estimates are zero; no additional rejection is measured. Limits are in counts $\mathrm{keV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, with $\Delta E = 5\,\mathrm{keV}$ and $A_{\mathrm{fid}} = \pi\,\mathrm{cm}^2$. †: Section 5.1.1.

B.7 Historical cosmic production coverage and interaction mechanisms

This section preserves the earlier source inventory and independent evaluation. Their original exposure and response are separate from the normalized partial calculation in Section 5.4.2.

The complete-family TPC inventory differs from the independent samples used to evaluate the final classifier and veto. Table B.16 gives the native cumulative baseline counts for the broader inventory. Only proton reconstruction remains partial there; its complete registered primary count cannot be used as the exposure of the displayed reconstructed subset. The historical zero-muon upper bound from incomplete response is superseded. The recovered negative-muon family contains 1 093 326 reconstructed entries and 52 baseline candidates, whereas the independent classifier sample has 45 baseline candidates. The independent gamma sample has none. Missing reconstruction is never counted as physical rejection.

Source	Primaries	Reco.	Valid	Energy	Tracks	Fiducial
	N_{gen}	N_0	C_V	$+ C_E$	$+ C_M$	$+ C_F$
$e^{\pm}$	673 481 974	246	245	82	20	2
γ	16 250 001 302	2 724	2 715	857	229	2
μ^-	3 174 660 130	1 093 326	1 088 160	275 432	185 819	52
n	4 774 215 260	108 833	107 000	23 382	13 916	249
p	343 475 361	14 586	14 323	3 871	2 259	44

Table B.16: Broader cosmic inventory before BDT or veto selection. All baseline requirements are cumulative and N_{gen} includes primaries without stored events. Proton reconstruction is partial, so its inventory denominator is not the exposure of the reconstructed subset. Development groups are included here and excluded from the independent final evaluation.

The complete neutron family yields $249/(4.774215260 \times 10^9) = 5.22 \times 10^{-8}$ baseline events per generated neutron, with a central 90% Poisson interval $[4.68, 5.79] \times 10^{-8}$. The matched ancestry partition is 246 prompt/non-delayed histories and three delayed histories. The independent sample instead starts from 141 baseline acquisitions, so its result is not obtained by applying an extra selection to the 249-event population. Table B.17 preserves that independent neutron exposure and the corresponding electron sample. Baseline identities match the final classifier/veto input exactly, and independent implementations reproduce every cumulative count.

Source	**Primaries** N_{gen}	**Reco.** N_0	**Valid** C_V	**Energy** $+ C_E$	**Tracks** $+ C_M$	**Fiducial** $+ C_F$
$e^{\pm}$	521 175 516	192	191	65	14	2
n	2 793 974 491	63 398	62 329	13 547	8 102	141

Table B.17: Native cumulative TPC selection on the fixed independent neutron and electron/positron samples. The baseline populations are 141 and two, respectively. Their generated-primary denominators remain fixed through the classifier and veto evaluation; the complete muon/gamma extension is reported in Appendix A.8.6.

The independent proton evaluation is complete. Three failed response files were repaired; one irrecoverably truncated transport file was replaced by independently seeded batches with the same 2 819 078-primary exposure and unchanged source, geometry, physics, and response. The original file is excluded once, preserving the 237 013 803-primary denominator for the final evaluation. The broader partly reconstructed proton inventory is not substituted for this complete sample.

Truth histories explain why a cosmic primary and the particle ionizing the gas can differ. We classify the track carrying the largest gas deposit and follow its ancestry, independently of the reconstructed selection. For the separate processed diagnostic samples, secondary electrons or positrons dominate 98.8% of gamma events and 96.8% of electron/positron events. In the proton sample, electromagnetic descendants are 43.5% of events, secondary protons or elastic recoils 26.8%, primary protons 15.9%, charged cascade particles 9.6%, and neutron or nuclear-fragment descendants 4.2%. Those fractions refer to events with a nonzero Micromegas signal and do not measure the acceptance of the final background selection.

Delayed activation is distinguished from a prompt event lacking veto activity by acquisition timing, rather than radioactive ancestry alone. In the independent neutron sample, the baseline contains 140 interaction acquisitions and one genuinely delayed acquisition; the latter fails the TPC BDT. A prompt recoil cannot inherit the later decay's veto exemption. The dedicated production-and-decay calculation above supplies the activation result; the zero in this restricted cosmic delayed subset is not a complete activation limit. Historical 15 mm hit-centroid selections and their event galleries remain supplementary study records, since

their populations and normalization area differ from the final 10 mm track-center baseline.

B.8 Normalization and response of the partial cosmic inventory

The normalized source estimates in Chapter 5 use a partial inventory of completed, qualified products. Unfinished or unqualified products are unknown outcomes, rather than zero observations. Missing muon exposure and incomplete neutron and veto accounting prevent a final joint cosmic prediction. The following definitions specify the source weights and response conditions; historical comparisons retain their original samples and are not pooled with this calculation. The incident proposal rate is $Q_j = \pi R^2 \Phi_j$, using the projected generation disk and scalar fluence over the generated energy and angular range. The CRY horizontal-current bin masses are converted to scalar fluence by division by the zenith-bin midpoint cosine once, without additional bin-width or azimuthal factors. These individual-particle sources omit correlated air-shower arrivals and unmodeled hadron species.

The Guan parameterization supplies the total muon proposal for each charge sign [16]. We assign the physical charge mixture through the CRY joint energy and zenith distributions [17],

$$w_{\mu^+}(E,\theta) = f_+(E,\theta) = \frac{H_+(E,\theta)}{H_+(E,\theta) + H_-(E,\theta)}, \qquad w_{\mu^-} = 1 - f_+. \tag{B.16}$$

An empty joint cell uses the fraction pooled over zenith in the same energy row. Both signs use the total Guan proposal rate with complementary weights, avoiding double counting of the total field. The incident zenith is measured from the downward source axis. For electrons and positrons, the generator chooses charge with probabilities p_σ obtained from the original histogram entries, whereas the physical scalar-fluence fractions are q_σ. The weight $w_{e^\sigma} = q_\sigma / p_\sigma$ corrects this mixture while preserving each conditional energy–angle distribution.

Let N_j be the number of generated physical parents in an admitted source sample and K_{jp} the number of selected acquisitions from parent p. All descendants of one parent share its incident-source weight. With $D = \Delta E A_{\mathrm{fid}}$, define

$$S_j = \sum_p w_{jp} K_{jp}, \qquad Q_{2,j} = \sum_p (w_{jp} K_{jp})^2, \qquad \widehat{B}_j = \frac{Q_j S_j}{N_j D}. \tag{B.17}$$

Zero-contribution parents remain in N_j. The fixed-N_j Monte Carlo variance is

$$\widehat{\mathrm{Var}}(\widehat{B}_j) = \left(\frac{Q_j}{D}\right)^2 \frac{Q_{2,j} - S_j^2/N_j}{N_j(N_j - 1)}. \tag{B.18}$$

Thus, correlated acquisitions are grouped before their contribution is squared.

The individual conditional limits for the available samples use the integer number M_j of parents with at least one selected acquisition, rather than the fractional weighted sum. Under the rare-parent Poisson model, $\mu_{90}(M)$ satisfies $P[\mathrm{Poisson}(\mu_{90}) \leq M] = 0.1$. Assuming at most one selected acquisition per parent gives the contribution bound b_j: unity for muon fractions and unweighted sources, or the larger electron/positron charge-mixture weight. Observed multiplicity above one withholds this limit; observed single acquisitions do not establish the physical bound. For independent source samples in a reconciled inventory, one planned joint conservative envelope is

$$U_{90}^{\mathrm{joint}} = \max_j \left(\frac{Q_j b_j}{N_j D} \right) \mu_{90}\left(\sum_j M_j \right). \tag{B.19}$$

Adding separately calculated 90% limits does not define this joint construction. This expression does not supply a complete-inventory limit from the available partial sample. Its nominal counting model does not cover preferential completion, response-fit selection or missing source exposure.

The transport gas is Ar–isobutane (2%) at 1.4 bar; the calibrated response retains 1% isobutane at 1.4 atm and 155 V cm^{-1}. The prediction conditions on this transfer and the fixed calibration and selections. For retained neutron histories, the auxiliary veto calculation must reproduce the archived TPC observables exactly. A failed or lost auxiliary result preserves the admitted TPC population and exposure without veto rejection credit.

Physical veto coincidence requires complete gas and sensitive-scintillator histories in a common acquisition time frame. One complete measured calibration record supplies unrelated veto activity to each physical parent, preserving its channel and timing correlations across acquisitions. The native selector `.*scintillatorVolume.*` excludes wrapping and light guides. Every positive sensitive deposit is checked, independently of raw window membership or veto tag. Radioactive or incomplete ancestry with an uncertified time frame prevents coincidence credit, even when the gas is prompt. Native offsets and track durations alone do not certify that frame; exported peak times are never shifted. Delayed activation additionally requires the causal nucleus's production time plus maximum gas drift time to precede the TPC window, with its decay inside, in a verified common frame. No radioactive timing certificate is assumed here. A gas history whose prompt/delayed scope is unresolved remains unclassified and receives no veto credit. Confirmed delayed decays are separated from the prompt prediction; unclassified contributions may overlap activation and are not added to a separate activation estimate.

The CRY neutron field and full outdoor HENSA spectrum are alternative incident-field hypotheses. HENSA includes terrestrial and atmospheric neutrons [18], with an assumed angular distribution, and is not added to either the CRY neutron contribution or the separately modeled environmental residual. Statistical errors and limits condition on the source models, admitted response and nominal measured-activity draw, excluding source-histogram statistics,

shower correlations, response transfer and unquantified fit-selection bias. The separate TPC and veto calibration acceptances do not establish joint X-ray acceptance and do not divide the predicted background levels.

Appendix C

Background Normalization and Selection Validation

C.1 Reconstruction and reference samples . . . 86
C.2 Detector-response assumptions . . . 87
C.3 Source normalization and statistical intervals . . . 89
C.3.1 Event weights and successive selection stages . . . 93
C.4 Historical selections and compatibility of source results . . . 94
C.5 Historical learned-selector validation . . . 95
C.5.1 Frozen candidate and comparator definitions . . . 95
C.5.2 Bounded 2025 repair-study audit . . . 98
C.5.3 Incident-photon acceptance of the archived reference . . . 99
C.6 Calibration and paired-response validation . . . 100
C.6.1 Two-width calibration illustration . . . 100
C.6.2 Paired response of archived neutron events . . . 103
C.7 Cosmic veto activity and accidental occupancy . . . 104
C.8 Gas and radon response accounting . . . 105
C.8.1 Completed cylindrical-wall response . . . 105
C.8.2 Historical gas and radon selection sequence . . . 106
C.8.3 Parent-tag bounds and background limits . . . 107
C.8.4 Argon and krypton counting model . . . 109
C.8.5 Activity inventory and absolute normalization . . . 109
C.9 Energy-resolved radioactive inventory . . . 110

This appendix retains the response assumptions, source normalizations and validation checks needed to interpret Chapter 5. The final TPC baseline is defined in Appendix F. Historical selections remain identified where their numbers have not been recomputed; no learned rejection is inferred from a failed transfer test.

C.1 Reconstruction and reference samples

Simulation and laboratory data meet at the raw-waveform reconstruction stage. Simulation converts transport deposits to charge and light, projects them onto channels, and produces digitized waveforms. Laboratory records supply measured waveforms with their readout metadata. Both then undergo channel masking, baseline and common-noise correction, peak finding, hit reconstruction and projected clustering. The final selection uses reconstructed observables; transport truth remains a diagnostic.

Stage	Physical task
Transport conversion	Convert energy deposits to visible gas and scintillator signals while preserving particle and parent history.
Detector response	Charge diffusion, gain and resolution; light attenuation and propagation; subsystem shaping, sampling and digitization.
Waveform conditioning	Readout mapping, channel masking, baseline and common-noise correction for measured and simulated signals.
Signal reconstruction	Find channel peaks, reconstruct hits, and form XZ/YZ clusters with the native `REST-for-Physics` procedures.
Event observables	Maximum-track energy and center, cluster multiplicities, charge widths, projection asymmetries and veto activity.

Table C.1: Common physical reconstruction stages. The complete process-name inventory is preserved with the separate configuration record; sharing the chain does not establish equivalence of the detector responses.

The ^{55}Fe sample anchors energy and charge-shape comparisons near 5.9 keV. Uniform-energy photons test incident acceptance only when every generated photon, including non-interactions and reconstruction failures, is counted. Source-specific background samples measure survival under their stated normalizations. Measured calibration veto records provide unrelated activity for accidental signal-loss studies; they are not physical correlations with simulated photons.

Argon fluorescence can produce two separate charge clusters after absorption of a ^{55}Fe photon. Figure C.1 illustrates why one cluster per projection is a physical selection, even for an X-ray calibration source.

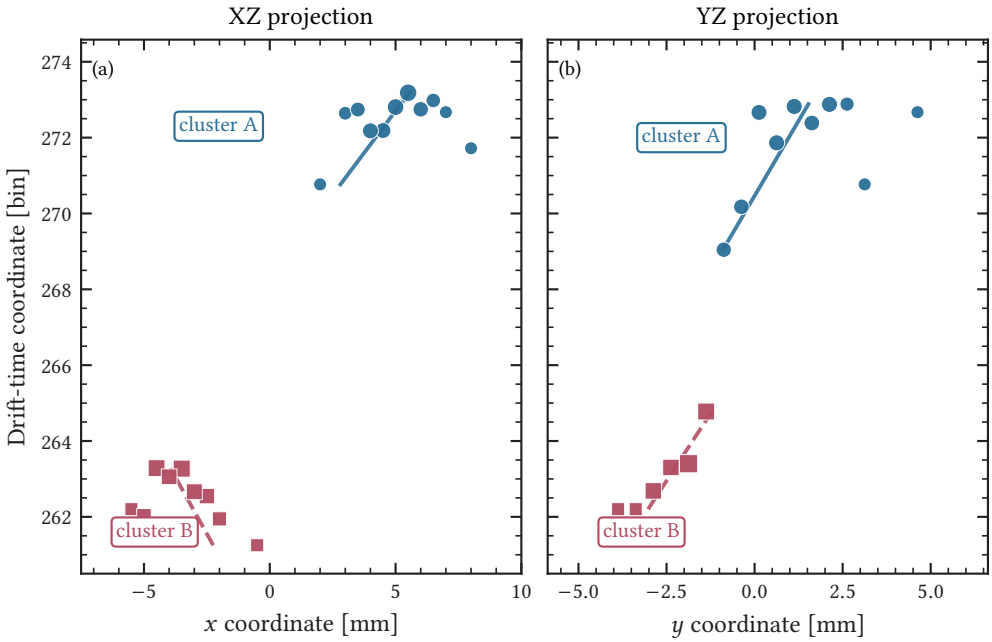


Figure C.1: Simulated ^{55}Fe event with two reconstructed charge clusters. Photoelectric absorption of the 5.90 keV photon is followed by argon-fluorescence reabsorption at a second site. The event fails the one-track requirement in both projections; it is not a Compton scatter.

C.2 Detector-response assumptions

The gas and veto visible-energy corrections are applied before channel projection and digitization. Section 2.1.3 defines nuclear-recoil ionization and the Lindhard approximation. For complete contained recoil histories, Equation 2.6 assigns electron-equivalent energy to the local deposits before charge transport. Archived background samples retain their recorded response settings and have not been recomputed with this contained-recoil model.

For scintillator deposits, the first-order Birks approximation is

$$E_{\mathrm{vis}} \simeq \frac{E_{\mathrm{dep}}}{1 + k_{\mathrm{B}} E_{\mathrm{dep}}/\Delta x}, \tag{C.1}$$

with a configurable nominal value $k_{\mathrm{B}} = 0.126\ \mathrm{mm\,MeV^{-1}}$ [19–22]. Here Δx should be the physical path associated with the deposit. The expression gives a light-yield proxy in the low-stopping-power normalization; comparison with the measured muon-equivalent energy scale also requires applying the same calibration convention to each response variant. The historical process instead estimates it from the shorter distance to a neighboring stored point in the same volume, using a 0.5 mm fallback when a suitable distance is unavailable. This

chord can belong to the following segment and need not equal the true `Geant4` step length. Photon-track local deposits are left unquenched, which also makes the result sensitive to the secondary-production thresholds used for unresolved secondary electrons.

A deterministic two-step check illustrates the effect without assigning a correction to the production samples. Deposits of 2 MeV and 0.2 MeV over true steps of 1 mm and 10 mm give 1.797 MeV visible energy with the stated Birks constant; the neighboring-point estimate gives 2.150 MeV, a 19.7% difference in this constructed example. Splitting a segment at constant stopping power leaves the true-step calculation unchanged. These algebraic checks identify a segmentation dependence; they do not replace a transport test with recorded true steps, initial recoil energies, and nonionizing losses.

The nominal Birks constant is not a dedicated calibration of the BabyIAXO bars: for comparison, a BC408 measurement gives $k_B = (0.155 \pm 0.005)\ \mathrm{mm\,MeV^{-1}}$ [22]. The historical neutron diagnostic reduces the inclusive energy-weighted gas and veto signals by approximately 0.22% and 25.7%, respectively. These inclusive ratios do not bound the effect on low-energy recoil candidates or panel-threshold efficiency. Campaign-specific response versions must be recovered before applying a revised model to an archived background result.

For a veto hit at distance d from the effective readout end, the response applies

$$E_{\mathrm{att}} = E_{\mathrm{vis}} \exp\left(-\frac{d}{\lambda_{\mathrm{att}}}\right), \qquad t_{\mathrm{att}} = t + \frac{d}{v_{\mathrm{eff}}}. \tag{C.2}$$

The current analysis configuration uses an effective attenuation length of 215 cm; this is consistent with the approximately factor-two response decrease over a 150 cm prototype bar, for which $\lambda_{\mathrm{att}} \simeq 150/\ln 2 = 216$ cm [1]. The manufacturer's material value of approximately 400 cm is a different quantity; archived campaigns retain their own recorded response parameters. This effective treatment absorbs reflections, surface finish, optical coupling, and channel-to-channel gain into calibrated parameters rather than tracking optical photons explicitly.

TPC diffusion uses gas metadata derived with `Garfield++/Magboltz`. A hit of energy E is converted to an effective primary-electron population $N_e \simeq E/W$, with optional Poisson or Fano fluctuations, and the electron positions are broadened according to

$$\sigma_T = D_T\sqrt{z_d}, \qquad \sigma_L = D_L\sqrt{z_d}. \tag{C.3}$$

An empirical detector-hit smearing is applied separately from diffusion to reproduce the measured calibration width and residual gain, avalanche, electronics, and calibration effects. The subsequent raw-signal conversion uses 512-bin waveforms, subsystem-specific sampling and shaping, the ^{55}Fe peak as the Micromegas energy anchor, and the measured through-going-muon response as the module-dependent veto anchor.

C.3 Source normalization and statistical intervals

One common reporting policy. Section 5.1.1 gives the reader-facing distinction between a background estimate, a standard error and an upper limit. The desired quantity for a radioactive source is

$$\frac{B}{A_s} = \frac{\mathbb{E}[K]}{\Delta E\, A_{\mathrm{fid}}}, \qquad \widehat{\mathbb{E}[K]} = \frac{1}{N}\sum_{i=1}^{N} K_i,$$

where K_i counts *all* selected detector events from generated history i, including histories with $K_i = 0$. For incident-particle sources, the corresponding incident rate replaces the decay activity with the source-generation normalization included explicitly. The normalization, selection and physical scope must be the same for the central estimate and its limit.

Choose the counting model before choosing the formula. If the source and event-building implementation establish $K_i \in \{0, 1\}$, a binomial construction at fixed N bounds the mean response directly; a Poisson model is a rare-event approximation, not a cure for correlation. For data-dependent stopping, an appropriate sequential construction is required to claim coverage uniformly over stopping times; an interval conditional on the realized exposure must not be presented as such a guarantee. The parent-tag construction in Appendix C.8.3 addresses that stopping issue. If several accepted descendants are possible, the calculation must instead model or bound their number while preserving the parent grouping. If a physical and implementation audit establishes $0 \le K_i \le M$, bounds on the tail probabilities can be combined using

$$\mathbb{E}[K] = \sum_{j=1}^{M} \Pr(K \ge j).$$

Assigning tail-error probabilities in advance with $\sum_j \alpha_j \le 0.1$ makes the sum of valid one-sided $1 - \alpha_j$ tail bounds a conservative 90% bound on the mean, without requiring the tails to be independent. The value of M must cover the retained transport physics and event builder; the largest multiplicity seen in a finite sample is not such a proof.

Model-based limits and the † convention. A useful limit does not require a distribution-free proof covering every possible transport history. Throughout the current background tables, † identifies a limit whose nominal coverage relies on an approximate counting model; the number itself is reported in the same units as the background estimate. For the argon and krypton samples, we adopt a rare-event Poisson working model for the total accepted count n at the realized exposure of N generated parents. The ordinary primary decay remains in one subevent and all selected events in these samples have distinct parents. This supports using the model as an approximation, but does not prove the absence of a rare correlated

secondary contribution. The one-sided response limit is [23]

$$R^{\dagger}_{1-\alpha} = \frac{\mu_{1-\alpha}(n)}{N\,\Delta E\,A_{\mathrm{fid}}}, \qquad \mu_{1-\alpha}(n) = \frac{1}{2}F^{-1}_{\chi^2_{2(n+1)}}(1-\alpha).$$

Here F^{-1} is the chi-square quantile; $\alpha = 0.1$ gives the one-sided 90% limit, not the upper endpoint of a central 90% interval. For zero selected events, $\mu_{90}(0) = -\ln(0.1) = 2.3026$, so the limit remains positive. The Poisson construction has coverage under its stipulated model; using it for these simulated histories is the approximation. It does not establish coverage for arbitrary unobserved multiplicity tails or arbitrary data-dependent stopping. All simulated descendants remain in the numerator: no transport process is disabled or discarded. For sufficiently populated multi-event samples, † instead identifies the parent-correlated normal construction specified with that source. The symbol also follows a limit into an absolute-background table when activity normalization is applied; it is not removed by multiplication by an activity.

Few survivors, no survivors, and missing results. A verified single-event or bounded-history construction remains usable with few or zero selected events; there is no universal requirement for 100 survivors. The 100-tagged-history safeguard used for some parent-correlated normal approximations in this work is only a practical reporting rule. It neither proves coverage above that threshold nor prohibits a justified small-sample method below it. Where used, the normal approximation assumes a sufficiently regular finite-variance parent response and is marked †. An empirical resampling method cannot discover an unobserved high-multiplicity tail, and an all-zero resample does not establish zero background [23].

A table entry *not established* means that neither a suitable working model nor a stronger background upper-limit construction has yet been adopted for that source. The estimate and its descriptive Monte Carlo standard error may still be reported when the sample and normalization are complete. By contrast, *incomplete reconstruction* means that the numerator is not fully evaluated, and *normalization unavailable* means that a per-primary yield or per-Bq response cannot yet be converted to an absolute background. These cases must never be encoded as zero survivors from the full generated exposure. The remedy is identified with the source: recover missing files, establish the multiplicity/counting model, determine the activity or flux, or increase statistics when precision is the actual limitation. No remedy may silently remove daughter decays or secondary activation from the stated physics scope.

Limits are individual per source unless simultaneous coverage is explicitly constructed. Monte Carlo-only limits are conditional on the stated response, selection and source model; source-activity and detector-response uncertainties require separate propagation. Historical tables using other interval conventions remain identified as historical, not silently reinterpreted as results of this policy.

Historical interval conventions and Poisson working models. For an equal-weight source stratum modeled by a Poisson count, the central 90% Garwood confidence interval is propagated through the same activity, exposure, energy-window, and fiducial-area normalization as the nominal background level [9, 24]. Some historical tables instead quote rows with $n = 0$, 1, or 2 survivors as one-sided 90% confidence bounds,

$$N_{90}(n) = \frac{1}{2}\chi^2_{0.90;\,2(n+1)} = \begin{cases} 2.3026, & n = 0, \\ 3.8897, & n = 1, \\ 5.3223, & n = 2, \end{cases} \tag{C.4}$$

in equivalent selected events. The bound is converted to a background level with the same physical normalization as the corresponding sample. This count-dependent choice is a table-reporting convention, not a single interval procedure with guaranteed 90% coverage over all possible observations. For a budget decision or a combined upper bound, a one-sided construction is fixed in advance and applied at every count, or a unified interval procedure with verified coverage is used.

For a source assembled from differently normalized strata, let c_i convert the selected mean count μ_i in stratum i to its background level, so that $B = \sum_i c_i\mu_i$. For fixed, independent Poisson samples, a combined interval can be constructed from the joint likelihood $\prod_i \mathrm{Pois}(n_i \mid \mu_i)$, with source and response uncertainties included as nuisance parameters [23]. When an explicitly conservative simultaneous upper bound is required, tail probabilities α_i may instead be assigned in advance with $\sum_i \alpha_i \le 0.1$. Summing the corresponding normalized one-sided $1-\alpha_i$ bounds then gives at least 90% simultaneous coverage by the union bound, provided the marginal constructions are valid. Simply adding marginal 90% endpoints does not establish that statement.

A Garwood interval on the unweighted total count is not exact for an unequal-weight mixture. Correlated descendants, shower particles, and repeated response realizations are grouped by their independent parent history; their multiplicity does not supply independent Poisson trials. For such samples, uncertainty must follow the per-history response estimator or a validated sampling construction. Finite Monte Carlo uncertainty is reported separately from source activity, angular-spectrum, and detector-response systematics. Where a legacy table does not preserve the required strata, its interval is marked as approximate and the row remains auxiliary. Mutually exclusive source scenarios are also kept separate: atmospheric and low-radioactivity argon are alternatives, as are a full HENSA neutron field and a decomposition that uses CRY plus the HENSA-minus-CRY residual. An aggregate neutron row is not combined with its own prompt and delayed subchannels.

For an equal-weight rare-event endpoint with at most one count per independent history,

zero selected events give the Poisson-approximation one-sided bound

$$B_{90} = \frac{-\ln 0.1}{T_{\mathrm{eq}}\, A_{\mathrm{fid}}\, \Delta E}. \tag{C.5}$$

This expression is useful for planning exposure only after the source rate and selection denominator have been validated. For exactly N independent generated histories with Bernoulli acceptance, the corresponding exact bound is $p_{90} = 1 - 0.1^{1/N}$, with $-\ln(0.1)/N$ its large-N limit. With $A_{\mathrm{fid}} = \pi$ cm^2 and $\Delta E = 5$ keV, a zero-count bound of 10^{-8} counts keV^{-1} cm^{-2} s^{-1} requires approximately 170 d of equivalent exposure. It is not a compute-time estimate, and it cannot be applied unchanged to an all-zero sample with unrestricted descendant multiplicity or unequal weights. The evaluation exposure or a statistically valid stopping rule is fixed before the final sample is inspected; selection optimization uses separate histories.

Fixed-volume neutron comparison. For the shielding comparison in Section 4.2.5, N includes all generated neutrons in completed, reconstructed blocks, including primaries with no saved detector event; pending or failed blocks are excluded from both exposure and selected counts. Table C.2 records the checkpoint of 13 October 2026 at 18:10 UTC. With k accepted TPC baseline events and the Poisson mean limit $\mu_{90}(k)$ defined above, the conversion is

$$\widehat{b} = \frac{k\lambda}{N A_{\mathrm{fid}} \Delta E}, \qquad b_{90}^{\dagger} = \frac{\lambda \mu_{90}(k)}{N A_{\mathrm{fid}} \Delta E}, \tag{C.6}$$

where $\lambda = 1336.87$ s^{-1}, $A_{\mathrm{fid}} = \pi$ cm^2, and $\Delta E = 5$ keV. The plotted B^{TPC} corresponds to $\widehat{b}$. For B^E, Equation 4.3 uses the energy-selected count, the same exposure $T = N/\lambda$, and the full readout area $A_E = 36$ cm^2. These events satisfy valid reconstruction and the same calibrated maximum-track energy requirement, before track-multiplicity and fiducial selection. At both stages, every contributing primary in this sample produces one selected event; this observation does not establish Bernoulli support for all possible histories. The selections are nested, and their different counting areas prevent interpreting B^{TPC}/B^E as a selection efficiency. Coverage for the computing-time stopping rule and exclusion of failed blocks has not been validated. The central 68.27% Garwood intervals and one-sided 90% limits therefore use a conditional Poisson working model, with flux, geometry, materials, and detector response fixed and their systematic uncertainties excluded. The limits apply individually to each configuration and checkpoint; they are neither simultaneous monitoring bounds nor limits on reduction factors.

Shielding	Generated	Energy	TPC
200 mm Pb	1 556 629 693	4 015	51
150 mm Pb + 50 mm BPE	1 701 880 633	4 009	44
150 mm Pb + 49 mm BPE + 1 mm LiF/PMMA	1 685 528 143	3 922	36

Table C.2: Generated neutrons and accepted counts for Fig. 4.5. Energy and TPC denote the 2–7 keV energy selection and complete TPC baseline, respectively; no BDT or active-veto rejection is applied.

C.3.1 Event weights and successive selection stages

For source i, the selected physical rate is

$$R_i^{\text{selected}} = \sum_{j \in i} w_{ij} I_j(C_i^{\text{selected}}), \tag{C.7}$$

where w_{ij} carries activity or incident-flux normalization, exposure and any source phase-space weights, and I_j is the event-selection indicator. For an equal-weight source generated directly from its physical distribution, this becomes $R_i^{\text{raw}} N_i(C_i^{\text{selected}})/N_i^{\text{generated}}$. The delayed channel receives no rejection from the initiating prompt pulse. For one mutually exclusive source hypothesis $\mathcal{H}$,

$$R_{\text{tot}}^{\text{selected}} = \sum_{i \in \mathcal{H}} R_i^{\text{selected}}. \tag{C.8}$$

An aggregate neutron row cannot coexist with its prompt and delayed subchannels; alternative argon activities likewise remain separate scenarios.

For baseline B, additional topology T and veto survival V,

$$P(B \cap T \cap V) = P(B)P(T \mid B)P(V \mid B \cap T). \tag{C.9}$$

This identity assumes no independence of TPC topology and veto activity. All factors refer to the same specified population and use consistent weights. For uncorrelated signal, the veto factor is

$$\varepsilon_{\text{sig}}^{\text{veto}} = 1 - P_{\text{acc}}(C_{\text{veto}}), \tag{C.10}$$

with accidental probability determined by compatible measured activity or an appropriate overlay. An unevaluated topology or veto stage contributes no additional rejection credit; it does not establish measured perfect acceptance.

C.4 Historical selections and compatibility of source results

The final baseline is defined in Appendix F. Historical numerical results retain the selection actually evaluated; adopting the final definition does not recompute a saved source table. The archived conservative reference used the same 2–7 keV window, one reconstructed track in each projection, and a maximum-track center within 10 mm, but less explicit reconstruction-validity tests. Its identifier is `background-analysis-v2-conservative-reference`. The following table retains the physical distinctions needed to compare it with the final baseline.

Quantity	Archived convention
Detector	IAXO-D1, argon with 1% isobutane at 1.4 bar; BabyIAXO Xe–Ne and telescope illumination require a separate response.
Energy and position	Calibrated maximum-track energy, with closed 2–7 keV boundaries; maximum-track center with open $r < 10$ mm boundary.
Validity	Projected maximum-track flags followed by one track per view; explicit final positive-energy and finite-observable tests must be reevaluated.
Topology and veto	No learned topology rejection in the archived conservative reference; source-appropriate veto survival is a separate stage.
Normalization	36 cm^2 before the fiducial stage, π cm^2 afterward, and $\Delta E = 5$ keV.
Intervals	Historical mixed convention: central 90% Garwood intervals above two survivors; one-sided 90% bounds for zero, one or two.

Table C.3: Conventions needed to interpret historical IAXO-D1 results. Compatibility with this archived reference does not establish reevaluation with the final baseline. The historical interval convention is not one procedure with guaranteed coverage at every count.

Earlier 15 mm readout-energy or all-hit-centroid selections are distinct from the final maximum-track-center definition. The response identifier `fe55-shape1014-ld165-td063-bin234-v1` does not uniquely identify a selection: archived studies used transparent intervals, different classifiers, energy-dependent requirements, and different fiducial observables. A common source sum therefore requires the same evaluated selection, normalization, response assumptions and mutually exclusive source inventory. Unavailable activities and unevaluated source components remain open entries; they are not zero contributions.

C.5 Historical learned-selector validation

C.5.1 Frozen candidate and comparator definitions

The frozen `background-analysis-v1` selector uses the primary random seed 1337; two additional seeds are retained only as robustness checks. It is implemented with the `HistGradientBoostingClassifier` in `scikit-learn` [25]. For reconstructed topology vector $\mathbf{x}$, the ordering score is

$$s_{\mathrm{BDT}}(\mathbf{x}) = P(^{55}\mathrm{Fe\text{-}like} \mid \mathbf{x}), \tag{C.11}$$

as returned by the classifier probability estimator. The value is not interpreted as an absolute physical probability because it depends on the training mixture and class weighting. The frozen topology selection criterion is $s_{\mathrm{BDT}} \geq s_0$, with s_0 fixed on a signal partition disjoint from model fitting to retain 80% of simulated ^{55}Fe. During grouped cosmic cross-fitting, no production file is scored by a model trained on that file.

The transparent histogram comparator is a binned signal-to-background log-likelihood ratio. For class $c \in \{S, B\}$, feature j, and bin b,

$$\widehat{p}_{cjb} = \frac{n_{cjb} + \alpha}{N_{cj} + M\alpha}, \qquad s_{\mathrm{LO}}(\mathbf{x}) = \sum_j \ln \frac{\widehat{p}_{Sj\,b_j(x_j)}}{\widehat{p}_{Bj\,b_j(x_j)}}. \tag{C.12}$$

The preserved comparison uses $M = 80$ equal-width bins between the pooled finite 0.1 and 99.9 percentiles, edge assignment outside that range, and pseudocount $\alpha = 0.5$. A larger value is more signal-like, and the comparison threshold is set to 80% validation-^{55}Fe acceptance. This definition is related to, but not identical to, the earlier `REST-for-Physics`-native `TRestDataSetOdds` workflow, which constructs calibration-only marginal densities and uses the opposite score convention.

Feature group and derived analysis inputs	Physical interpretation and role

Hit multiplicity
`n_hits_x`, `n_hits_y`
`n_hits_min`, `n_hits_max`
`abs_n_hits_balance`
	Primary separation. Strip-view hit counts expressed directly and through symmetric summaries. Extended or poorly matched tracks tend to activate more strips or exhibit a large view imbalance.

Charge-cloud size
`sigma_x_mm`, `sigma_y_mm`
`sigma_z_mm`
`sigma_xy_max_mm`
`sigma_xy_mean_mm`
	Primary separation. Widths of the reconstructed charge distribution in the two strip projections and drift coordinate, together with deterministic transverse summaries. Compact X-ray conversions should be narrow.

Projection shape and symmetry
`abs_sigma_balance`
`abs_skew_xy`, `abs_skew_z`
	Supporting separation. Disagreement between transverse widths and non-Gaussian asymmetry of the reconstructed charge cloud test whether the two views describe one compact conversion.

Charge balance
`max_energy_fraction`
`abs_energy_balance`
	Supporting consistency, with one inactive field. Projection-energy balance tests agreement between views. The dominant-track fraction is retained in the frozen contract because it can identify fragmentation before the one-track gate, but it equals unity for every finite event in the evaluated candidate-v1 samples and supplies no separation here.

Table C.4: Reconstructed feature groups used by the frozen candidate-v1 BDT topology selector. The four rows are also the physically coherent families permuted jointly in the archived joint-family permutation study; deterministic summaries are kept with their primitive width or hit inputs. The names in the first column are fields in the flattened `TRestAnalysisTree` analysis export, rather than literal ROOT branch names. The calibrated maximum-track energy defines the 2–7 keV analysis window but is not used as a classifier input.

The earlier event-random measured comparison retained 9 of 5102 background candidates at 80% calibration acceptance, compared with 10 for the binned log-odds method and 16 for transparent manual selection criteria. It used a different calibration-centered preselection and allowed events from the same operating period in both partitions. These development results do not establish run-to-run transfer. The independent checks in Tables C.5 and C.6 determine the interpretation of the selector.

Experimental evaluation	Independent result and interpretation
Earlier event-random development split, BDT at 80% reference	**Calibration:** 80.00%. **Background:** 9/5 102 = 0.176%. **Interpretation:** Useful selector-development result, but events from the same operating period can occur in both partitions and the preselection differs from the candidate-v1 contract.
Untouched August–September run block, calibration-anchored BDT, primary seed	**Calibration:** 34 814/43 726 = 79.62%, 90% C.I. 79.30–79.94%. **Background:** 21/97 = 21.65%, 90% C.I. 14.99–29.66%. **Interpretation:** Threshold fitted only on a disjoint July validation block. Features are robustly referenced to the nearest operational ^{55}Fe calibration run.
Untouched August–September run block, fixed simulation-derived intervals	**Calibration:** 18 908/43 726 = 43.24%. **Background:** 8/97 = 8.25%. **Interpretation:** Transparent reference; lower background acceptance is accompanied by substantially lower and unmatched signal acceptance.

Table C.5: Run-block closure of the experimental topology study. The July 2024 runs are divided chronologically into model-training and threshold-calibration blocks; all August–September runs remain untouched until the final test. Exact binomial intervals are quoted for the primary BDT result. The earlier event-random row is shown for context but is not directly numerically comparable because its compact-tree sample and energy preselection differ.

`background-analysis-v1` check	**Result**	**Consequence**
Cosmic train/evaluation production-group overlap	0 groups	Leakage condition satisfied.
Untouched simulated ^{55}Fe test	9 224/11 499 = 80.22%	Simulated 80% efficiency gate satisfied.
Measured R02756 ^{55}Fe domain check	8 889/64 162 = 13.85%	Simulation-to-data efficiency gate failed.
Finite-feature fiducial cosmic candidates	61/294 = 20.75% pass topology	Cross-fitted diagnostic; not a validated background acceptance.
Aggregate HENSA after topology and veto	8 events, $(1.10^{+0.88}_{-0.55}) \times 10^{-7}$	Replaces the training-contaminated diagnostic 5-event value, but cannot form the prompt/delayed budget split without history labels.

Table C.6: Grouped cross-fit and domain-validation outcome for the frozen candidate selector. Production file plus campaign defines the cosmic grouping, and no group is scored by a model trained on that group. The strong loss of measured ^{55}Fe acceptance means that all resulting cosmic levels are diagnostic and are marked as not validated for the final budget.

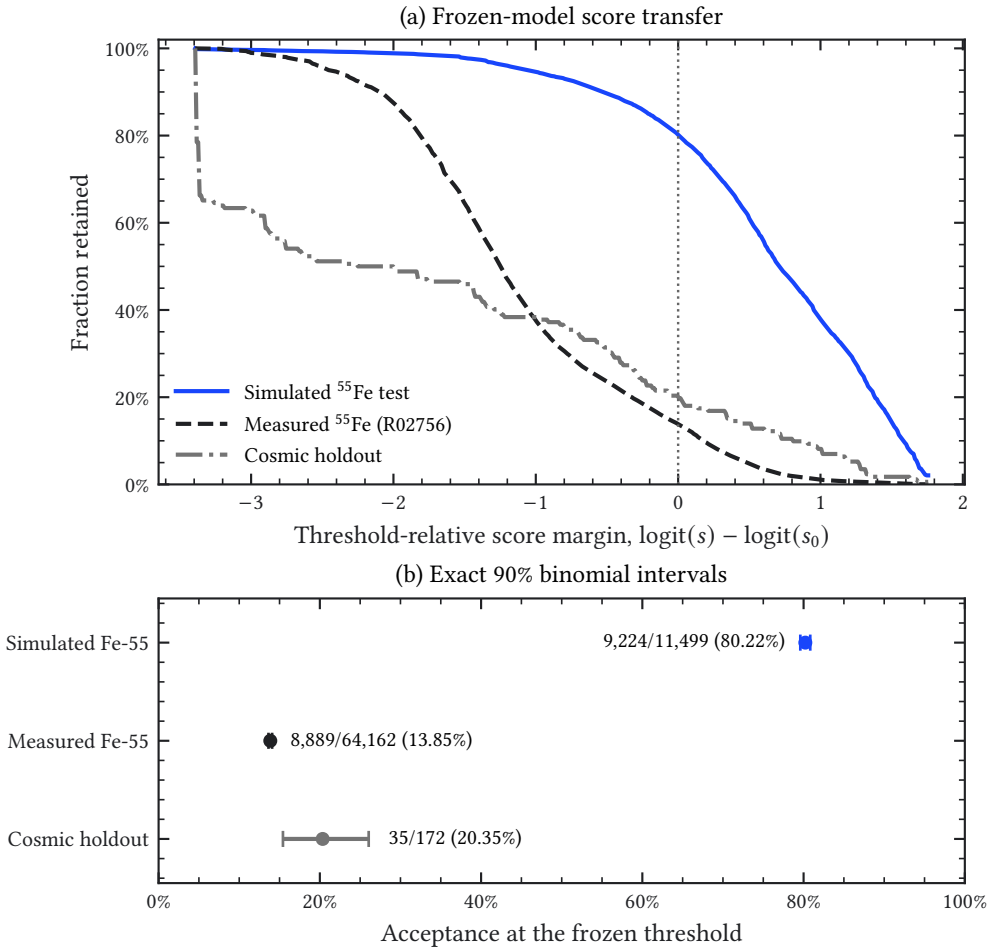


Figure C.2: Historical simulation-trained BDT transfer test. Panel (a) shows the fraction above each threshold-relative score; zero marks the unchanged selection criterion. Panel (b) compares acceptances at that selection criterion, with 90% counting intervals. Simulated ^{55}Fe acceptance is 80.22%, but measured R02756 acceptance is 13.85%. This tests the earlier conventional response, not the refined `Garfield` model or separately calibrated thresholds.

C.5.2 Bounded 2025 repair-study audit

The bounded repair study was defined only after candidate-v1 failed. R02905 provided a stable calibration reference at a peak span of approximately 1.3%. R02997 was retained for threshold setting with a declared 9.9% gain-drift flag, and the shorter R02998 calibration run was retained for non-blind validation with an approximately 4.5% peak span. The originally proposed background runs R03020–R03022 were excluded before model development because they contained one event, only approximately 10 s of data, and an anomalous high-rate or

incomplete analysis, respectively. The non-overlapping R02993 and R02994 files supplied the 83 non-blind background candidates reported in Table C.7.

The long R03015 background run and R03018 calibration run were reserved as the final blind block. R03018 carries predeclared gain-drift and spatial-support flags, but neither candidate scores nor acceptances were inspected. The failed non-blind result records `selected_candidate=null` and `blind_input_opened=false`; no retuning followed. Thus, the unopened block remains available for a new, predeclared selector that accounts for detector conditions.

Predeclared candidate	Independent non-blind validation	Promotion result
Calibration-anchored logistic, 15 features	**Signal (R02998):** 2 668/3 338 = 79.93%, 90% C.I. 78.75–81.06%. **Background (R02993+R02994):** 50/83 = 60.24%, 90% C.I. 50.63–69.29%.	Signal gate passed; background upper bound exceeded the predeclared 25% limit.
Signal-only robust L1, 15 features	**Signal (R02998):** 2 536/3 338 = 75.97%, 90% C.I. 74.73–77.19%. **Background (R02993+R02994):** 47/83 = 56.63%, 90% C.I. 47.01–65.88%.	Failed the 75% signal lower-bound gate and the background gate.
Transparent five-feature box	**Signal (R02998):** 2 405/3 338 = 72.05%, 90% C.I. 70.74–73.33%. **Background (R02993+R02994):** 50/83 = 60.24%, 90% C.I. 50.63–69.29%.	Failed the signal point, signal lower-bound, run-range, and background gates.

Table C.7: Predeclared non-blind validation of the bounded 2025 argon topology-candidate family. The thresholds were fixed on R02997; the table contains only the independent R02998 signal and R02993/R02994 background results. No candidate passed all gates.

No candidate passed all predeclared gates, and R03018/R03015 remains unopened for selector evaluation. The conservative reference therefore receives no learned topology-rejection credit from either historical study.

C.5.3 Incident-photon acceptance of the archived reference

The complete `saveAllEvents` uniform-photon production retains non-interactions and reconstruction failures. The following efficiency is therefore per incident photon in the specified point-source, 8°-cone illumination, rather than conditional on saved gas events. It applies the historical conservative selection and excludes learned topology, optics, veto and accidental-live-time factors.

True incident energy [keV]	Generated	Pass conservative reference	Incident-reference efficiency [%]
2.0–2.5	41 885	7 333	17.507 [17.203, 17.816]
2.5–3.0	41 726	8 815	21.126 [20.798, 21.457]
3.0–3.5	41 801	10 519	25.164 [24.815, 25.516]
3.5–4.0	41 561	12 113	29.145 [28.778, 29.514]
4.0–4.5	41 497	12 401	29.884 [29.514, 30.256]
4.5–5.0	41 531	12 463	30.009 [29.639, 30.381]
5.0–5.5	41 912	12 241	29.206 [28.841, 29.574]
5.5–6.0	41 754	10 920	26.153 [25.799, 26.509]
6.0–6.5	41 712	9 001	21.579 [21.248, 21.913]
6.5–7.0	41 712	6 338	15.195 [14.906, 15.487]

Table C.8: Micromegas-chain efficiency versus true incident energy for the historical IAXO-D1 reference selection, pending recomputation with the final TPC baseline. The numerator requires reconstructed energy in 2–7 keV, one valid track in each projection, and a maximum-track center within 10 mm. The denominator contains every generated photon in the point-source, 8°-cone illumination. Bracketed ranges are exact two-sided 90% Clopper–Pearson intervals. No learned topology, optics, veto, or accidental-live-time factor is included.

Across the flat generated 2–7 keV spectrum, 102144 of 417091 photons pass: 24.49%, with a two-sided 90% interval of 24.38–24.60%. The response peaks at 30.01% in the 4.5–5.0 keV bin. These historical values require reevaluation with the explicit final validity criteria; a BabyIAXO axion efficiency additionally requires Xe–Ne response, telescope illumination and separately established acquisition/veto losses.

C.6 Calibration and paired-response validation

C.6.1 Two-width calibration illustration

The calibration illustration in Figure C.3 applies the final baseline to preserved R02756 samples without changing response parameters or energy calibration. Measured development and checking samples are even and odd blocks of twelve entry-rank blocks from canonical processed file 1; overlapping file 0 is excluded. The REST detector response uses even and odd rows of a saved 2000-event sample. The Garfield detector response uses even and odd collected-electron-bank indices, with the accepted v8 response settings. These partitions separate threshold setting from checking, but the response study had already inspected the checking samples, so this is retrospective validation rather than a blind detector-response test.

For $\boldsymbol{x} = (\overline{\sigma}_{xy}, \sigma_z)$, the common score is

$$d^2 = (\boldsymbol{x} - \boldsymbol{\mu})^{\mathsf{T}} \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}), \qquad d^2 \le t.$$

The center and covariance are determined only from REST detector-response development events:

$$\boldsymbol{\mu} = (0.86219, 0.56099)\ \text{mm}, \qquad \Sigma = \begin{pmatrix} 0.063763 & 0.035628 \\ 0.035628 & 0.042484 \end{pmatrix}\ \text{mm}^2.$$

The nearest-rank 80th percentile, including ties, gives $t_{\text{data}} = 16.04436$, $t_{\text{REST}} = 2.22340$ and $t_{\text{Garfield}} = 16.64575$. Only the ellipse size changes; no background sample enters calibration.

Checking populations after the baseline contain 13298 measured, 635 REST and 1374 Garfield events. At their separately calibrated thresholds, 10614, 489 and 1056 events pass. At the unchanged REST threshold, only 2754 measured and 305 Garfield events pass, compared with the same 489 REST events. Finite calibration-sample uncertainty is additional to the conditional checking-count intervals in Table C.9. The Garfield development export contains 998 reconstructed rows from 1000 requested bank events; the missing rows are not counted as reconstructed candidates.

Figure C.3 shows the accepted ellipse interiors and the six checking-background candidates in the measured-data panel. The ellipse center, orientation, and relative axis lengths are fixed; only its size changes between the common and separately calibrated thresholds. This illustrative selection is not used for the background predictions.

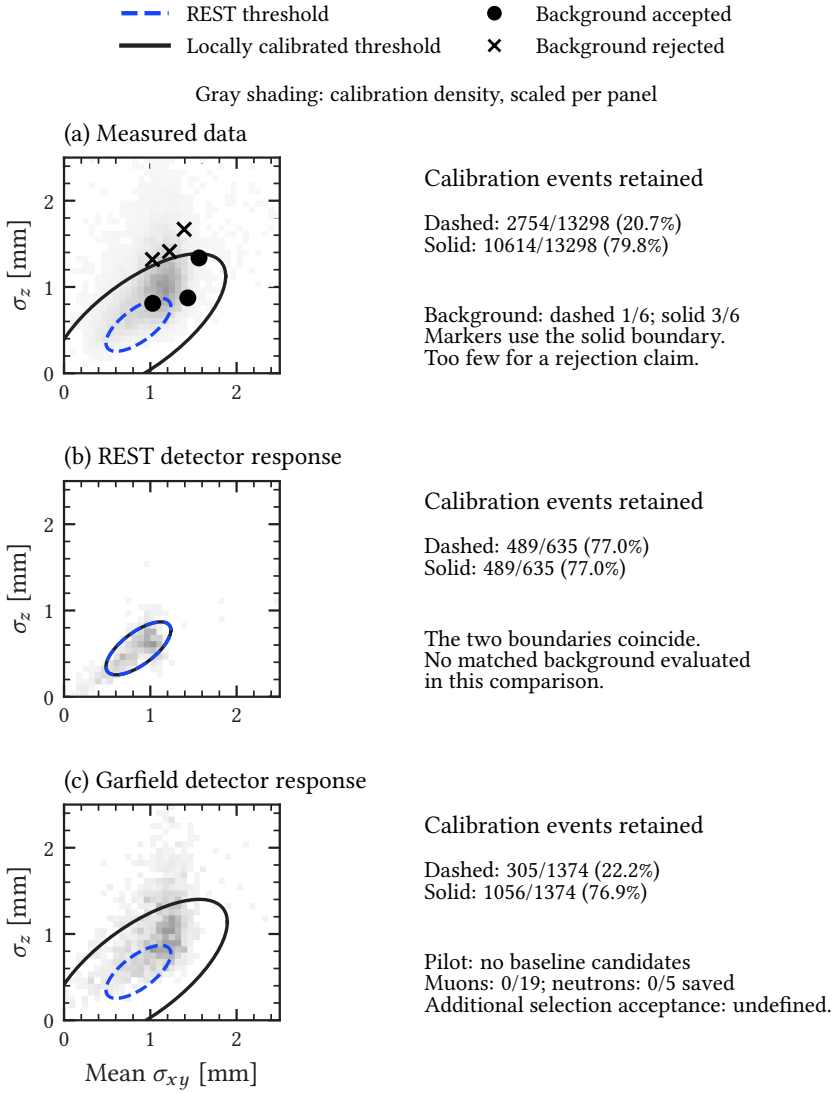


Figure C.3: Illustrative two-width selection after the TPC baseline. Dashed ellipses use the REST threshold; solid ellipses use separately calibrated thresholds. Interiors are accepted. Calibration counts include events outside the plotting range; the six background candidates are shown in (a).

Checking calibration sample	Acceptance: common size [%]	Acceptance: calibrated size [%]
Measured R02756	20.7 [20.0, 21.4]	79.8 [79.1, 80.5]
REST detector response	77.0 [73.6, 80.1]	77.0 [73.6, 80.1]
Garfield detector response	22.2 [20.1, 24.5]	76.9 [74.6, 79.0]

Table C.9: Checking-sample acceptance of the illustrative ellipse, conditional on the baseline. Sizes target 80% development acceptance and are not adjusted on these checking events. Brackets are 95% Wilson counting intervals at the fixed score and sizes; these are calibration acceptances, not background rejection.

The energy-bin record retains the dependence hidden by a calibration-spectrum average; all ratios below are accepted/baseline checking counts at the separately calibrated thresholds.

Energy [keV]	Measured	Conventional	`Garfield`-refined
2–3	177/534	6/14	19/60
3–4	196/409	10/20	28/42
4–5	545/792	68/91	71/104
5–6	4750/5820	225/292	412/540
6–7	4946/5743	180/218	526/628

Table C.10: Accepted/baseline checking counts for each separately calibrated two-width threshold. Bins include their lower edge; only the last includes its upper edge. Low-energy simulation bins are sparse. These counts expose the energy dependence hidden by a single calibration-spectrum average.

Measured acceptance at the separately calibrated threshold rises from 33.1% at 2–3 keV to 86.1% at 6–7 keV. Across six measured checking blocks, acceptance ranges from 78.2% to 81.9%. This does not establish stability between runs. The later saved-waveform checking-background sample contains six baseline candidates: five in R02721 and one in R02761, with none in R02739 or R02766. The unchanged data-calibrated threshold accepts three, with a conditional 95% Wilson interval of 18.8–81.2%; the REST threshold accepts one. The nine development-background candidates are excluded from this check. These sparse results do not determine a transferable rejection factor. Earlier overlapping waveform controls and scalar exports are retained in the separate analysis record, rather than pooled with this sample.

C.6.2 Paired response of archived neutron events

A completed response comparison reuses the same 3864 stored histories from 64 neutron files before reconstructed-energy or topology selection. Event, subevent, file and source-entry identities match between the archived REST analysis and the Garfield replay. Both apply the final baseline and the unchanged simulation-trained BDT, with their previously established

energy scales.

Stage	REST	Garfield
Stored histories	3 864	3 864
Reconstructed rows	3 864	3 618
Valid	3 802	3 575
Energy	1 658	353
One track per view	94	152
Baseline	11	2
Baseline + fixed BDT	0	0

Table C.11: Paired neutron-response diagnostic, with cumulative selection counts after reconstruction. Five Garfield histories have unsupported multiple acquisition windows and remain unresolved. Its last entry is therefore an observed count, not complete zero survival of all 3864 histories. No veto requirement is applied.

Among 3859 jointly supported histories, eleven pass only the REST baseline, two pass only the Garfield baseline, and none pass both. A comparison restricted to REST survivors would therefore miss both entering events. The original transport gas contains 2% isobutane, whereas the refined response is supported by 1% calibration data; the archived response settings and visible-deposit ionization approximation also differ. The selection migration cannot be attributed to Garfield alone, and the observed zero BDT counts do not establish equal rejection or an absolute limit.

C.7 Cosmic veto activity and accidental occupancy

The common-geometry waveform comparison is presented in Section 5.4.3. The historical event-level reference below retains the rates needed to interpret the point-trigger approximation. Peak multiplicity, energy-sharing and channel-occupancy grids are preserved in the separate diagnostic record; they define no additional rejection factor.

The earlier cosmic-only productions provide an event-level reference for these supplementary diagnostics. A visible trigger is a simulated primary event with at least one reconstructed `rawPeaksVETO` peak after detector-response processing; several peaks from the same primary count as one trigger. These productions use an earlier detector configuration and are not combined with the common-geometry waveform samples of Section 5.4.3. In particular, their gas density and one pipe-cap placement differ from the updated reference geometry. Their rates should therefore be read as configuration-dependent estimates, not an independent validation of the experimental activity.

For independent Poisson point triggers, the occupancy of a window Δt is

$$P_{\text{cosmic veto}}(\Delta t) = 1 - \exp(-R_{\text{veto}}\Delta t)\,, \tag{C.13}$$

where R_{veto} is the visible event rate, not the individual-peak rate. Finite correlated peak trains require their timing to be included explicitly, as in the cumulative comparison in the main text.

Source	Triggers	Time [s]	Rate [Hz]	$P_{100\,\mu\mathrm{s}}$ [%]	$P_{1\,\mathrm{ms}}$ [%]
Muons	97 443	266.21	3.66×10^{2}	3.59	30.7
$e^{\pm}$	65 908	1372.31	4.80×10^{1}	0.479	4.69
γ	36 094	601.18	6.00×10^{1}	0.599	5.83
n	38 415	949.75	4.04×10^{1}	0.404	3.96
p	68 166	24328.40	2.80×10^{0}	0.0280	0.280

Table C.12: Cosmic-induced visible veto-trigger rates from the dedicated veto-noise productions. A visible veto trigger is defined as an event with at least one reconstructed `rawPeaksVETO` peak after detector-response processing. The muon exposure includes the correction for the duplicated azimuth factor in the legacy Guan metadata, identified by its stored flux; its generation area is unchanged. The last two columns give the point-trigger Poisson occupancy for windows of 100 µs and 1 ms; extended correlated peak trains require a timing-overlap calculation.

C.8 Gas and radon response accounting

The gas and surface-contamination results separate generated parents from stored detector subevents and retain correlations within each radioactive history. Source location and gas conditions remain assumptions of the response; no migration of charged daughters or implantation-depth distribution is inferred from a cathode-volume generator. The completed gas/cathode and cylindrical-wall results below support the source tables in Chapter 5.

C.8.1 Completed cylindrical-wall response

The corrected source surfaces in Figure 5.5 were each evaluated with 50 predetermined shards of 10 000 generated parents. All 100 transports and reconstructions passed the original-source, exposure, event-identity and file-integrity checks. The full radioactive histories and secondary transport are retained. No endcap or implantation-depth distribution is added to either cylindrical-sidewall hypothesis.

Stage	Gas-facing PTFE	Copper behind PTFE
Generated parents	500 000	500 000
Stored events	178 714	7 045
Valid, C_V	177 585	7 005
Energy, $C_V C_E$	21 392	1 066
One/view, $C_V C_E C_M$	16 751	833
Baseline, $C_V C_E C_M C_F$	25	4
Baseline + BDT	11	1

Table C.13: Complete corrected-wall selection counts. Each source has 50 predetermined shards of 10000 generated parents. Selection stages are cumulative; the fiducial radius is 1 cm. The last row adds the common simulation BDT, with no veto requirement.

Wall source	Baseline Estimate	Baseline + BDT Estimate ± MC error counts $\mathrm{keV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{Bq}^{-1}$	90% limit After BDT
Gas-facing PTFE	3.18×10^{-6}	$(1.40 \pm 0.42) \times 10^{-6}$	$\leq 2.12 \times 10^{-6\dagger}$
Copper behind PTFE	5.09×10^{-7}	$(1.27 \pm 1.27) \times 10^{-7}$	$\leq 4.96 \times 10^{-7\dagger}$

Table C.14: Wall responses per Bq for the same full transport and conventional response as the gas/radon study. The error retains parent-history correlations. †: conditional one-sided 90% Poisson limits following Section 5.1.1.

For comparison with the other deposited-activity scenarios, the absolute-rate inventory adopts an explicitly illustrative surface density of 1 mBq m^{-2}, not a measurement. The sidewall areas are $2\pi(0.050)(0.030)$ and $2\pi(0.051)(0.030)$ m^2, giving 9.425×10^{-6} and 9.613×10^{-6} Bq, respectively. The 10 nm navigation offset is not included in the macroscopic normalization area. Multiplying either response in Table C.14 by its activity gives the corresponding illustrative background; other activity assumptions scale these estimates linearly. The two locations are alternatives, and neither is included in the quoted known-inventory subtotal. All selected events in these samples come from distinct parents; the limits adopt the rare-event Poisson approximation without changing any transport physics.

C.8.2 Historical gas and radon selection sequence

The historical counts below use the conventional argon detector response with the archived diffusion, shaping and sampling settings. Their TPC selection required 2–7 keV maximum-track energy, projected validity, one track per projection and a maximum-track center with $r < 10$ mm. They remain distinct from the explicit final-baseline evaluation in Tables 5.5 and 5.6. The final historical column combines transparent TPC intervals with a veto classifier; it is neither a TPC BDT result nor an adopted veto efficiency. The source geometry

`NoVetoesGasNotSplit.gdml` does not establish a physical scintillator prediction for these samples.

Source	Stored	Processed	Energy stage	Historical TPC selection	+ intervals + veto BDT
^{39}Ar gas	100 000	62 051	13 164	357	28
^{85}Kr gas	100 000	62 160	13 861	377	29
^{222}Rn gas	100 000	59 856	357	15	0
^{218}Po cathode	100 187	99 950	5 500	697	14
^{210}Pb cathode	100 145	99 901	10 990	3 009	68

Table C.15: Unchanged historical gas/radon counts, not final-baseline results. The energy stage includes the historical valid-track checks and 2–7 keV window. The last column combines additional TPC intervals with a veto classifier; it is neither a TPC BDT result nor an adopted veto efficiency. Stored entries are source-filtered detector events, possibly including several subevents per decay chain, not generated parent decays.

The retained source hypotheses separate initial gas ^{222}Rn, ^{218}Po cathode chains stopped at ^{210}Pb, and separate ^{210}Pb chains. Cathode sampling occurs in the aluminum-window volume; ion migration and implantation depth are not modeled. The completed gas/cathode evaluation includes all 640 predetermined files for five sources, with generated-parent counts reconciled against completed logs and metadata, including parents producing no stored subevent. Missing or failed files are never counted as zero-response trials.

For N generated parents and k_i selected subevents from parent i, including $k_i = 0$, the mean yield and parent-correlated standard error are

$$\bar{k} = \frac{\sum_i k_i}{N}, \qquad \mathrm{SE}(\bar{k}) = \sqrt{\frac{\sum_i k_i^2 - (\sum_i k_i)^2/N}{N(N-1)}}.$$

Source activity and detector-response uncertainties are separate.

C.8.3 Parent-tag bounds and background limits

Surface-chain jobs stop at 500 stored entries per job, with a wall-time safeguard and completion of the current parent after a soft abort. The realized parent count is therefore random. A fixed-N binomial interval alone does not establish coverage under this stopping rule. Parents from all predetermined files are instead concatenated in a fixed order and treated with one predeclared sequential construction [26].

Let $Z_i = \mathbf{1}(k_i > 0)$ and $p = \Pr(k_i > 0)$. For a proposed lower null boundary $p_0 < 1$, define

$$E_N(p_0) = \sum_{j=1}^{64} w_j \prod_{i=1}^{N} \left[1 + \rho_j \frac{p_0 - Z_i}{1 - p_0}\right].$$

The ρ_j are 64 Gauss–Legendre nodes mapped to $(0, 1)$ and w_j are their positive weights normalized to unity. This finite mixture defines the construction. For independent histories with constant conditional probability $p \geq p_0$, the factors are nonnegative with conditional expectation at most one. The mixture is therefore a supermartingale starting at one, with probability at most 10% of ever reaching ten. The upper bound U_{90} solves $E_N(U_{90}) = 10$ and is one if every parent is tagged. It remains valid at the stated stopping times under the complete-history and source-stability assumptions. The same construction is used for all five sources; its coverage is individual, not simultaneous across the table.

	90% upper bound on $\Pr(k_i > 0)$	
Source	**Baseline**	**Baseline + BDT**
^{39}Ar gas	$\leq 4.17 \times 10^{-3}$	$\leq 4.19 \times 10^{-4}$
^{85}Kr gas	$\leq 4.39 \times 10^{-3}$	$\leq 5.34 \times 10^{-4}$
^{222}Rn gas	$\leq 2.74 \times 10^{-4}$	$\leq 5.86 \times 10^{-5}$
^{218}Po cathode	$\leq 1.96 \times 10^{-2}$	$\leq 4.99 \times 10^{-3}$
^{210}Pb cathode	$\leq 6.55 \times 10^{-2}$	$\leq 2.65 \times 10^{-2}$

Table C.16: Dimensionless parent-tag probability bounds before and after the calibrated simulation BDT. These are statistical diagnostics, not background-rate upper limits when a parent can produce several selected events. The BDT is held fixed during the background evaluation.

A tag bound concerns p, whereas the background response is $R = B/A_s = \mathbb{E}[k]/(\Delta E A_{\text{fid}})$. If a justified transport/event-building bound $k \leq M$ existed, it would give $R_{90} \leq MU_{90}/(5\pi)$. Observed multiplicities do not prove such a bound. In particular, the three nominal decays of the lead chain do not establish $M = 3$ for full transport with secondary nuclear processes and unrestricted subevent stacking. An unobserved multiplicity tail remains a limit on conversion from tags to rates.

Gas ^{222}Rn uses `fullChain=off`, killing decay-produced daughters before their later waiting stage; its event-building configuration allows at most one reconstructed subevent per parent. For this verified single-event case, $R_{90} = U_{90}/(5\pi)$. For populated multi-event samples, the reported † bound is instead

$$R_{90}^{\dagger} = \widehat{R} + 1.2816 \frac{\mathrm{SE}(\bar{k})}{\Delta E A_{\text{fid}}}.$$

This normal approximation assumes stable finite-variance multiplicity and regular large-

sample behavior under the stored-entry stopping rule; it has no finite-sample, stopping-uniform guarantee. The practical reporting requirement is at least 100 tagged parents, not a theorem of coverage. Cathode ^{218}Po and ^{210}Pb have 692 and 2979 tagged parents before the simulation BDT, and 159 and 1168 afterward. No normal approximation is used for empty or sparse samples. Post-BDT counting results are conditional on the independently fixed calibration model and threshold and exclude transfer and model-choice uncertainty.

C.8.4 Argon and krypton counting model

The recorded ^{85}Kr production uses a 100 µs subevent mean-lifetime threshold. The approximately 514 keV ^{85}Rb excited state has a 1.015 µs half-life, or approximately 1.46 µs mean lifetime [27], so its ordinary primary chain stays within one subevent. The comparison uses mean lifetime rather than a sampled delay. The direct ^{39}Ar decay likewise ends at stable ^{39}K. Neither observation proves single-event support for all enabled secondary transport; the archived nuclear/process audit has not established a hard multiplicity cap or found evidence for a numerically significant secondary tail.

Each gas sample contains 100000 generated parents; the 26 argon and 35 krypton post-BDT survivors have distinct parents. The tables adopt the rare-event Poisson working model of Appendix C.3, marked †, without disabling secondary transport. For krypton the one-sided 90% mean-count limit is 43.8715 events, giving $R^{\dagger}_{90} = 2.793 \times 10^{-5}$ counts keV^{-1} cm^{-2} s^{-1} Bq^{-1}. At the reference activity the absolute bound is 5.2215×10^{-12} counts keV^{-1} cm^{-2} s^{-1} before outward table rounding. Parent-tag limits remain separate diagnostics.

C.8.5 Activity inventory and absolute normalization

The archived gas cylinder, after subtracting readout and window solids, has volume 234.8966 cm^3. Its GDML density is 2.27661 kg m^{-3} and argon mass fraction 0.9711 for the 2% isobutane mixture at 1.4 bar, giving 0.519313 g of argon. The atmospheric activity $1.01 \pm 0.02\,(\text{stat}) \pm 0.08\,(\text{syst})$ Bq kg^{-1} from Ref. [28] yields $A_{39} = 0.5245 \pm 0.0428$ mBq under independent Gaussian treatment of those uncertainty components. This is a source assumption, not an assay of the operating detector. For $B = A_{39}R$, the MC and activity standard-error terms combine in quadrature. A one-sided joint bound allocates 5% failure probability to each factor:

$$B_{90} = R_{95}\left(\widehat{A}_{39} + \Phi^{-1}(0.95)\sigma_A\right).$$

The union bound gives at most 10% probability that either factor fails; independence is not required for that step. The argon response uses the Poisson working model, so the absolute bound retains †. Multiplying two 90% endpoints would not provide the same joint coverage.

Other activities use the fixed inputs in Table 5.8: argon mass for krypton, net gas volume for radon, and $\pi(0.015\text{ m})^2$ projected area for cathode contamination. Their absolute limits

are conditional on those specified activities. The underground-argon scenario scales the atmospheric result by the fixed factor 1400, with no uncertainty assigned to that reduction. These alternatives do not define a measured summed intrinsic background.

Transport uses argon with 2% isobutane at 1.4 bar, while the preserved response uses 1% at nominal 1.4 atm. The response remains conditional on that gas transfer. The original-log checksums, parent reconciliation, calibration and independent selection checks are preserved with the gas/radon analysis record.

C.9 Energy-resolved radioactive inventory

Figure 5.14 is built from native reconstructed events, not digitized curves or a rescaling of previously drawn histograms. It uses the same source exposures, calibrated energy scale, activity assignments and non-overlapping radioactive inventory as Table 5.25. The groups are atmospheric ^{39}Ar, detector materials with assigned activities, inner-lead ^{210}Pb and U/Th/K, front-end electronics, and the remaining gas and cathode scenarios. Alternative wall-contamination hypotheses, unknown activities, residual-neutron fields and absolute cosmic/environmental spectra are not inserted into the sum. The conventional REST detector response is retained; this figure is not the `Garfield` response study.

The first panel applies $C_V C_M C_F$ without C_E; the second applies the complete baseline and the adopted TPC BDT only within 2–7 keV. No BDT acceptance is extrapolated outside that interval, and no veto rejection is credited. The first bin spans 0.10–0.25 keV, followed by 0.25 keV bins to 10 keV. Every bin is divided by its actual width and by $A_{\mathrm{fid}} = \pi\,\mathrm{cm}^2$. No smoothing is applied.

For source s of activity A_s, simulated exposure N_s and bin width δE_b, let k_{isb} be the number of accepted events in bin b from parent i. The rate and its Monte Carlo variance are

$$B_{sb} = \frac{A_s}{N_s \delta E_b A_{\mathrm{fid}}} \sum_i k_{isb}, \tag{C.14}$$

$$\mathrm{Var}(B_{sb}) = \left(\frac{A_s}{N_s \delta E_b A_{\mathrm{fid}}}\right)^2 \frac{N_s}{N_s - 1} \left[\sum_i k_{isb}^2 - \frac{(\sum_i k_{isb})^2}{N_s}\right]. \tag{C.15}$$

Parent identities include the source file, and implicit zero-response primaries remain in N_s. Independent source variances add for each displayed bin. The gray band is a per-bin standard error, not a confidence band or a detector-response uncertainty envelope; empty bins do not establish a zero true rate. Bin-to-bin parent correlations must be retained for integrated uncertainties: the rate integral is checked against the source tables, whereas its uncertainty is taken from the existing parent-cluster inventory calculation, not obtained by summing the plotted bin variances.

The preserved component-spectrum analysis includes source and component bin tables, source-wise 2–7 keV closure checks, activities, input/output hashes, and the plotting script.

The selected spectrum integrates to the same known-inventory subtotal as the independently tabulated budget. The extension below the calibration region describes the model response only, not a validated 0.1 keV experimental threshold.

Appendix D

Detector Services and Calibration Controls

D.1 Primary charge and avalanche fluctuations 113
D.2 Prototype gas and acquisition services 114
D.3 Ultraviolet photoelectron calibration 116

D.1 Primary charge and avalanche fluctuations

The resolution model in Chapter 2 separates fluctuations in primary ionization, avalanche gain, and readout noise. For electron-equivalent deposited energy E_{ee},

$$\langle N\rangle = E_{\mathrm{ee}}/W, \qquad \mathrm{Var}(N) = F\langle N\rangle, \tag{D.1}$$

where W is the mean ionization energy and F is the Fano factor. Recoil quenching enters the electron-equivalent energy before applying this relation. For negligible attachment, the Townsend coefficient gives the mean avalanche gain $G = \exp[\int_0^L \alpha(x)\,\mathrm{d}x]$. If the individual avalanches are independent and their relative gain variance is $b = \sigma_G^2/G^2$, the variance of the amplified charge is $G^2\langle N\rangle(F + b)$. The approximately Gaussian relative FWHM is therefore $2\sqrt{2\ln 2}\sqrt{(F + b)/\langle N\rangle}$, before adding independent electronic noise [29]. An event-wide gain fluctuation is a distinct contribution because it correlates the avalanches in one event.

The simplified ^{55}Fe primary-charge study uses a 6 cm × 6 cm × 3 cm volume, with photons directed from the center of one face toward the gas. For argon with 2% isobutane at 1.4 bar and 135 V cm^{-1}, the fitted primary-charge centroid and FWHM are 222.29 and 15.05 electrons (6.77%). For equal-partial-pressure xenon–neon with 2.3% isobutane at 1.05 bar and 110 V cm^{-1}, they are 205.35 and 14.43 electrons (7.03%). Both simulations use 20 °C and exclude amplification and readout fluctuations. These approximately 7% widths are intrinsic

ionization references, rather than predictions of the complete detector resolution.

An earlier R02756 sample of 80 669 measured events compares six energy estimators with a two-line ^{55}Fe fit over 4.75–6.75 keV. The fitted K_α FWHM/centroid is 26.6% for readout energy, 26.7% for peak-neighborhood charge, 26.6% for summed channel pulse heights, 28.1% for charge above threshold, 29.4% for full-window charge, and 50.1% for the largest channel pulse height. The last estimator loses information when charge is distributed across strips. The comparison with 900 000 simulated entries anchors each main peak at 5.9 keV and fits an additional Gaussian energy broadening to the data. That fitted width is separate from the event-level gain fluctuations and waveform validation in Section 3.6; the complete historical spectra are preserved in the independent detector-calibration document.

D.2 Prototype gas and acquisition services

The IAXO-D0/D1 gas line reduces bottle pressure and regulates detector flow and pressure. Evacuation assists gas changes, and relief valves protect the thin entrance window, whose aluminized Mylar and copper support were designed for pressure differences up to 1.5 bar. Open-loop operation exhausts the used mixture; closed-loop operation recirculates and purifies it, reducing consumption of expensive xenon–neon mixtures [30]. Pressure, flow, oxygen, and moisture must be monitored during recirculation. Figure D.1 summarizes these functional paths; the complete engineering drawing is preserved separately.

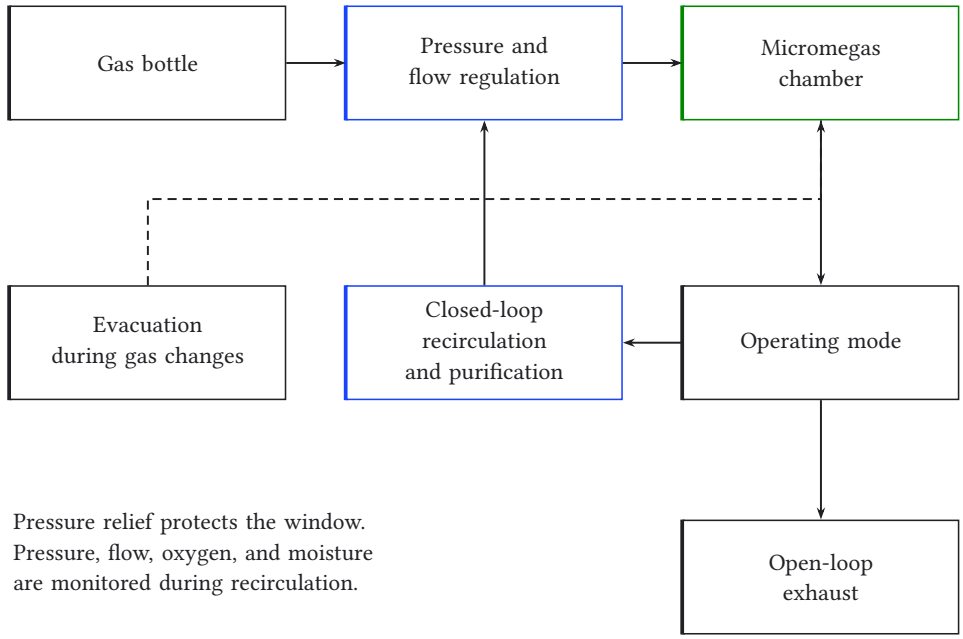


Figure D.1: Functional gas paths of the IAXO-D0/D1 services. Solid arrows identify supply, exhaust, and recirculation paths; the dashed branch identifies evacuation during gas changes. Valve-level implementation and the full engineering drawing are retained in the separate detector-services document.

Two high-voltage channels bias each Micromegas detector, separately setting the cathode drift field and mesh amplification field. Controlled ramping, filtering, current monitoring, and trip handling limit discharge and noise effects. The `hvps` library described in Section 3.8.3 provides serial control of the CAEN supplies; its `Node-RED` interface integrates monitoring, recovery from short trips, alerts, and shutdown after repeated trips [31]. Veto photomultipliers use separate high-voltage channels, with gain and timing monitored alongside detector conditions.

The commissioned IAXO-D0 front end uses four AGET chips [32], each with 64 channels and a 512-sample buffer; 60 channels per chip serve the 240 strips. The hardware supports sampling up to 100 MHz, shaping between 50 ns and 1 µs, selectable polarity, four charge ranges, and self-triggering. Feminos provides configuration, readout timing, and Gigabit Ethernet transfer, with a Trigger Clock Module for synchronous multi-board acquisition [33]. The run-specific register configuration determines the physical settings. The acquisition software assembles UDP frames into event waveforms and run metadata, as described in Section 3.8.2. These specifications describe the IAXO-D0 reference hardware; the IAXO-D1 and final BabyIAXO designs use different electronics.

D.3 Ultraviolet photoelectron calibration

Section 2.5.3 presents the CEA Saclay setup, the arrival-time difference method, and the comparison with `Magboltz`. The original presentation [34] also records the gas-analyzer readbacks, pulse-amplitude fluctuations, and exploratory empirical corrections. Figure 2.19 uses the curves before those corrections; the study does not provide an independent BabyIAXO background measurement.

Appendix E

Detector-Response Validation and Technical Controls

E.1 TPC response specification . . . 118
E.1.1 Ionization, drift and local amplification . . . 118
E.1.2 Charge collection and waveform formation . . . 119
E.1.3 Native reconstruction settings . . . 120
E.2 R02729 comparison before additional gain broadening . . . 121
E.3 Additional event-gain calibration . . . 123
E.4 Comparison across operating conditions . . . 124
E.5 Timing comparison and admitted event populations . . . 126
E.6 Historical response and reconstruction controls . . . 128
E.7 Optical and waveform controls for the veto . . . 129
E.7.1 Muon-reference scale and run dependence . . . 131
E.7.2 Single-module neutron and sideband controls . . . 132
E.8 Conditional full-geometry neutron response . . . 133
E.8.1 Geometry, clocks and optical interpolation . . . 133
E.8.2 Acceptance-aware muon anchor . . . 135
E.8.3 TPC population and coincidence reference . . . 135
E.8.4 Frozen selections and paired rejection . . . 136
E.8.5 Response uncertainty and limits of the prediction . . . 138
E.9 Extending the comparison with the common TPC baseline . . . 139
E.9.1 Experimental background: more events, but few baseline survivors . 139
E.9.2 Gas-matched deposited-energy calibration controls . . . 140

This appendix gives the response settings and complementary checks needed to interpret Chapter 3 and the background model. The current individual-electron TPC response, the earlier avalanche-library response and the veto optical calculation are distinguished

throughout. Calibration, numerical checks and comparisons with separate samples have different scopes; none changes the response or efficiencies of the reference background productions. The common TPC baseline is defined in Section F.1. Historical energy-only selections retain their original populations.

E.1 TPC response specification

The response in Section 3.4 connects gas ionization to channel waveforms before the common native reconstruction. Table E.1 gives the reference conditions. The conventional response acts on `Geant4` energy deposits; the current `Garfield++` route stops supported particles at their first gas entry and assigns the gas cascade to `Heed` [35, 36]. The historical avalanche-library response is summarized in Section E.6.

Quantity	Reference value
Gas and pressure	Ar/isobutane 99/1; 1.4 bar
Temperature; Penning probability	Assumed 293.15 K; zero
Bulk field; distance step	155 V cm^{-1}; 0.01 cm
Recorded W; Fano factor	26.37 eV; 0.1709
Bulk drift speed	23.623 847 8 mm µs^{-1}
D_L; D_T	$0.0352816\sqrt{\text{cm}}$; $0.0759793\sqrt{\text{cm}}$
Bulk handover; mesh underside	$z = -14.6$ mm; $z = -14.75$ mm
Anode; reconstruction reference	$z = -14.8$ mm; $z = -15$ mm

Table E.1: Reference gas and transport settings. Coefficients come from the archived bulk manifest; the canonical configuration specifies the mixture. The older manifest does not independently record its fraction. Forward and reconstructed drift speeds are distinct.

The frozen conventional R02756 chain converts deposited energy to primary charge with mean $1000E_{\text{dep}}/W$, for keV deposits and an eV W-value. Its configuration enables Fano fluctuations with precedence over Poisson excitation; the inspected older source does not establish the exact remote sampler. With unit-electron weights, the hit field called energy becomes charge-like. Configured $D_L = 0.165\sqrt{\text{cm}}$ is an effective fitted broadening, while $D_T = 0.0626333\sqrt{\text{cm}}$ and field is 153.33 V cm^{-1}. The event smearing setting $R = 215$ acts on this charge-like sum and is not a measured 215% energy resolution. Channel formation uses 40 ns bins and a 1014 ns shaper; ADC conversion is separate from reconstructed keV calibration. These effective settings are not independent measurements of gas or electronics.

E.1.1 Ionization, drift and local amplification

Gas-entry records preserve parent and track identities, species, energy, position, direction and time in cm, ns and eV. Entries advance 10 nm along the direction to resolve

boundary ambiguity. `TransportPhoton` includes relaxation and photon reabsorption; `TransportDeltaElectron` and `NewTrack` handle supported charged particles. The resulting conduction electrons receive no second E/W conversion or Fano draw. The interface does not accept neutrons, recoil nuclei or deposited-energy lists. A deposit-based extension needs a recoil-specific charge yield and electron-birth model, with each ionizing segment assigned to one calculation. The Fano factor specifies count variance, not a complete integer distribution.

Individual electrons drift with `AvalancheMC` and `Magboltz` coefficients in a uniform bulk field. The square sensor spans $x, y = \pm 30$ mm and $-14.8 < z < 14.996$ mm. Bulk transport ends 150 µm above the mesh underside; primaries born below that interface start locally at their birth positions and times. Local injection assumes 0.1 eV towards the anode because macroscopic drift provides no microscopic instantaneous velocity. Separate ionization and drift random streams preserve the gas cascade in numerical controls. Failed drift, zero-charge avalanches and absent source events remain distinct.

The local field resolves a 3×3 hole motif over a nominal diamond pad. The pad diagonal is 500 µm, with repetition period 353.553 µm; internal hole spacing is 100 µm. Aperture diameter, foil thickness and amplification gap are 40 µm, 5 µm and 50 µm. The provisional registration rotates the periodic axes by 45° around $(x_0, y_0) = (0, -125)$ μm, with zero motif phase. The Laplace solution uses periodic lateral boundaries, mesh at −335 V, anode at zero and the bulk-field Neumann condition at local $z = 500$ µm. Its $216 \times 216 \times 109$ grid has 1.6368 µm lateral and 0.625 µm minimum vertical spacing. These are ideal-metal fields, without dielectric charging or channel weighting fields.

`AvalancheMicroscopic` follows each injection in that field with collision data initialized up to 200 eV. Finite-flight conductor intersection and preserved post-collision directions prevent missed foil contacts and transport-chunk direction loss. Collection requires the correct terminal status and proximity within 0.1 µm of the anode. The `SetCollisionSteps(100)` setting is a numerical API parameter. Grid sensitivity remains measurable: changing lateral resolution from 144×144 to 216×216 changes collected charge per injected primary by +5.846% (paired 95% interval [1.463, 10.229]%, 32 event groups). Absolute gain is therefore not demonstrated to be grid-converged.

E.1.2 Charge collection and waveform formation

The nominal readout contains 29 163 ownership polygons, including 122 border triangles, connected to 240 DAQ channels. Endpoint containment decides ownership in global XY without an extra rotation. Same-channel overlaps form a union, shared edges receive one deterministic assignment, conflicting interiors are rejected and uncovered charge is recorded. This represents nominal electrical ownership rather than measured copper clearance. Every collected electron retains unit charge and its individual time $t_i = t_{\mathrm{bulk},p(i)} + t_{\mathrm{micro},i}$. No centroid approximation or library-mean normalization replaces this multiplication. Collected gain is

divided by all injected primaries, including zero-charge injections.

At sample times $t_j = 40\,\mathrm{ns}\; j$, the waveform is

$$s_c(t_j) = n_c(t_j) + ag \sum_{i \to c} T\left[\frac{t_j - t_{\mathrm{trig}} - \delta_c - (t_i - t_{\mathrm{origin}})}{\lambda}\right]. \tag{E.1}$$

The unit-peak template T is tabulated in ns, a converts electrons to ADC, g is an optional mean-one event gain, δ_c a channel delay and λ a dimensionless stretch. Recentering on the mean arrival time preserves relative times while replacing absolute timing with sampled trigger placement. Individual floating-point arrivals are evaluated directly at the ADC samples; superposition precedes signed 16-bit rounding and overflow checks. The measured strong-pulse median template can include gas and illumination spread. Synchronous baseline-block resampling, with mean block length 32 samples, retains short-time and channel correlation. Strip Ramo currents, ion motion and a complete AGET circuit are not calculated. A 32-event control with 7263 injected primaries and 414 602 anode electrons reproduces all 3 948 544 integer samples endpoint by endpoint at unit gain without noise or clipping; this tests accounting rather than amplitude calibration.

E.1.3 Native reconstruction settings

Data and simulation enter at the saved pedestal/common-mode-corrected waveform boundary. The run metadata still request residual 75-sample baseline correction and per-time-bin median common-noise subtraction, separately for TPC and veto channels. The latter process is skipped with fewer than 200 input channels. For R02756 and R02729, baseline samples 20:200 overlap the pulse-search interval 180:500. Pulse candidates require 20 consecutive samples above $3.2\sigma_b$ and within-pulse standard deviation above $3.0\sigma_b$. The latter requirement is not an integrated-charge threshold. The 512 samples span 20.48 µs.

A successful per-channel Gaussian fit supplies amplitude A_c and peak time μ_c, with reconstructed position $z_c = z_{\mathrm{readout}} + n_z v_{\mathrm{rec}} \mu_c$. Here μ_c is in microseconds and $v_{\mathrm{rec}} = 23.709\,955\,583\,\mathrm{mm\,\mu s^{-1}}$. The hit-time field is zero; timing resides in Z. Connected XZ and YZ components link hits at distance strictly below 3.5 mm; the largest amplitude component is selected independently in each view. This does not establish a uniquely paired three-dimensional track. Energy is the calibrated sum $c_E(E_X + E_Y)$ of fitted main-track amplitudes, not the waveform integral or all-event deposited energy. Centroids and widths use amplitude-weighted population moments. In particular,

$$\sigma_z^2 = \frac{E_X \sigma_{zX}^2 + E_Y \sigma_{zY}^2}{E_X + E_Y} + \frac{E_X E_Y}{(E_X + E_Y)^2}(\bar{z}_X - \bar{z}_Y)^2, \tag{E.2}$$

so inter-view timing also broadens the measured longitudinal observable. Track-count plots show $\max(N_X, N_Y)$, whereas baseline multiplicity requires $N_X = N_Y = 1$. Invalid and no-

signal events retain their appropriate input populations.

E.2 R02729 comparison before additional gain broadening

The full comparison in Figure E.1 uses 10 000 measured checking waveforms and 15 000 fresh gas-entry histories, selected before response calculation. R02729 contains over 62 000 saved events at Ar/isobutane 99/1, with about 1% peak variation across twelve time blocks. The simulation follows Heed, drift, direct microscopic amplification and individual collection before the same native REST reconstruction. Only the amplitude conversion is calibrated to the measured peak; widths, positions and acceptance are not fitted. Measured noise and microscopic avalanche fluctuations are included, but no additional event gain or spatial gain map is applied. The checking identities exclude amplitude-fit events; the inspected acquisition is not a blind validation. There are 9956 valid measured events and 12 574 valid simulated events, before energy, multiplicity or fiducial selection. Every histogram uses all valid events, including out-of-range values; no-signal and invalid histories remain in the separate input counts. The track-count panel shows $\max(N_X, N_Y)$.

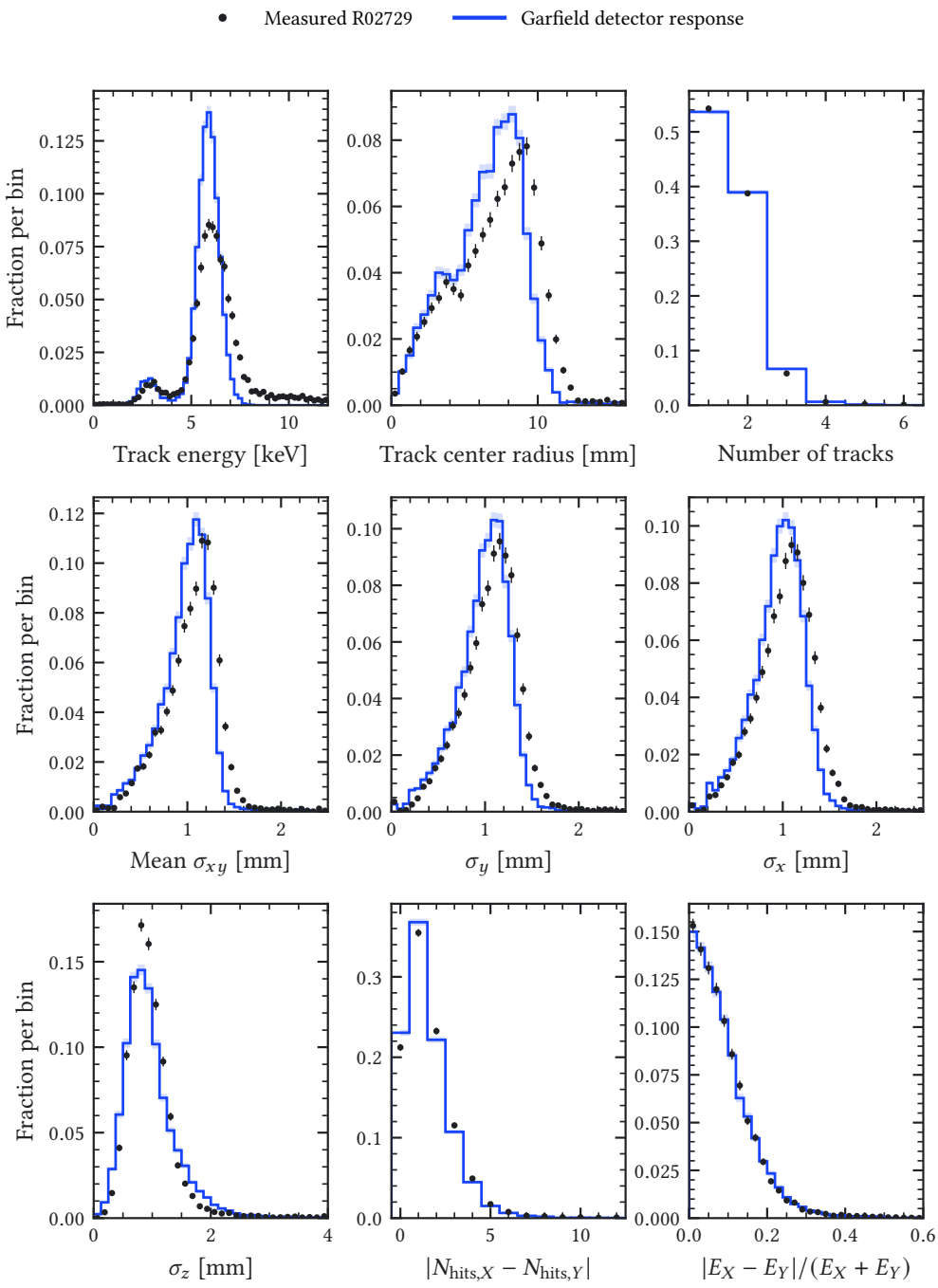


Figure E.1: R02729 response before additional gain broadening: 9956 valid measured events (black) and 12 574 simulated events (blue). Histograms show full-sample bin fractions with 68.27% Wilson counting intervals.

Projected multiplicities and charge balances are close to data, but transverse widths are smaller, positions more central and the high-energy tail narrower. The central longitudinal width alone does not validate the combined selection: the baseline retains 49.8% of valid simulated events and 39.4% of measured events (Table 3.6). These fractions do not include the probability that an emitted photon produces a recorded event.

E.3 Additional event-gain calibration

The extra mean-one event gain in Equation 3.7 acts before integer ADC conversion and native reconstruction, multiplying the signal in both views while leaving separately added noise unchanged. The mean conversion stays at 0.849994 ADC counts per collected electron and 0.000 558 764 keV per native summed-amplitude unit. Gas transport, microscopic histories, collection, pulse shape and noise stay fixed.

The measured development and checking samples contain 10 000 events each from six alternating time blocks, with no common acquisition identity. There are 9961 valid measured development events. The simulated development sample has 2000 fresh gas-entry histories, separate from the 15 000 checking histories; no interaction or validity requirement is imposed when these samples are formed. The source lines near 5.9 keV and 6.5 keV have relative weights 24.4 and 2.85.

A Gaussian fit in 4.4–7.5 keV, with 31 equal bins, uses valid development events before baseline or veto selection. The objective is $|\log(\sigma_{\mathrm{MC}}/\sigma_{\mathrm{data}})|$. A free constant was excluded because it absorbed shoulders and destabilized the width. The measured fit gives $\mu = 6.1088(96)$ keV and $\sigma = 0.7356(88)$ keV from fit covariance. These peak-core widths include unresolved source lines and response tails.

After a $0 \leq s_g \leq 0.30$ scan, local refinement uses the first upward width crossing in $0.08 \leq s_g \leq 0.16$ and allows at most eight new trials. Each trial repeats waveform formation and reconstruction with common random seeds. Failed, bounded or unbracketed fits stop the replay without widening the bracket. Four new trials select $s_g = 0.0917$, or gain RMS 9.19%. The simulated development sample has 1665 valid events, including 1468 in the fit window, with $\sigma = 0.7397(204)$ keV and $\mu = 5.8418(219)$ keV. Width agreement is the fitted target; the center remains lower.

Checking waveforms reuse every retained microscopic history without resampling, including zero-charge and invalid events. A zero-spread replay verifies identities and reconstructed columns. Three hundred independent whole-event bootstrap pairs, including invalid events before fitting, give an approximate standard deviation 0.0079 on s_g. This is conditional on the estimator and window and excludes original amplitude calibration and physical-model uncertainty. The checking discrepancy motivated the study, making it a retrospective consistency check.

Checking sample	Valid	μ [keV]	σ [keV]
Measured calibration	9 956	6.146	0.751 ± 0.009
REST detector response	11 792	6.060	0.964 ± 0.015
`Garfield` detector response	12 574	5.895	0.533 ± 0.004
`Garfield` response + gain variation	12 574	5.850	0.771 ± 0.008

Table E.2: Checking peak-core fits in the fixed 4.4–7.5 keV window before baseline selection. Width uncertainties are fit-covariance errors, excluding calibration and model uncertainty; these are not isolated intrinsic resolutions.

With 0.1 keV bins and a Gaussian smoothing width of 0.1 keV, modes are 5.95 keV in data, 6.05 keV in REST and 5.85 keV in both `Garfield` samples. The corrected simulation has 0.76% of valid events above 8 keV, compared with 9.91% measured; above 10 keV the data have 503 events and both `Garfield` samples zero. In that measured tail, 42% have one track per view and 45% lie inside 10 mm, compared with 56% and 90% in the fit window. These differences establish a remaining distribution discrepancy without identifying its cause.

E.4 Comparison across operating conditions

The initial amplitude-calibrated response completes 40 of 44 acquisitions: 28 whole development runs and 12 whole checking runs across ten physical families. Three amplitude fits failed and one acquisition was excluded. Two electronics groups share pulse shapes and channel delays learned from whole development runs; each acquisition supplies residual noise, trigger placement and reconstruction settings. Only amplitude conversion is fitted within a run. The later R02729 gain variation is absent.

Each condition has 512 simulated histories split into 320 development and 192 checking histories before fitting; measured checking samples start with 240 events per acquisition. Within-run checking blocks are distinct from the blocks used for amplitude, noise and trigger calibration. Reused banks within families and common source histories across families correlate the comparisons. Previously inspected acquisitions make this a consistency check.

Figure E.2 compares valid checking events before baseline selection. The 12 whole checking runs have median simulation/data width ratios 0.977 transversely and 0.943 longitudinally. Tails and correlations need not agree when central widths do.

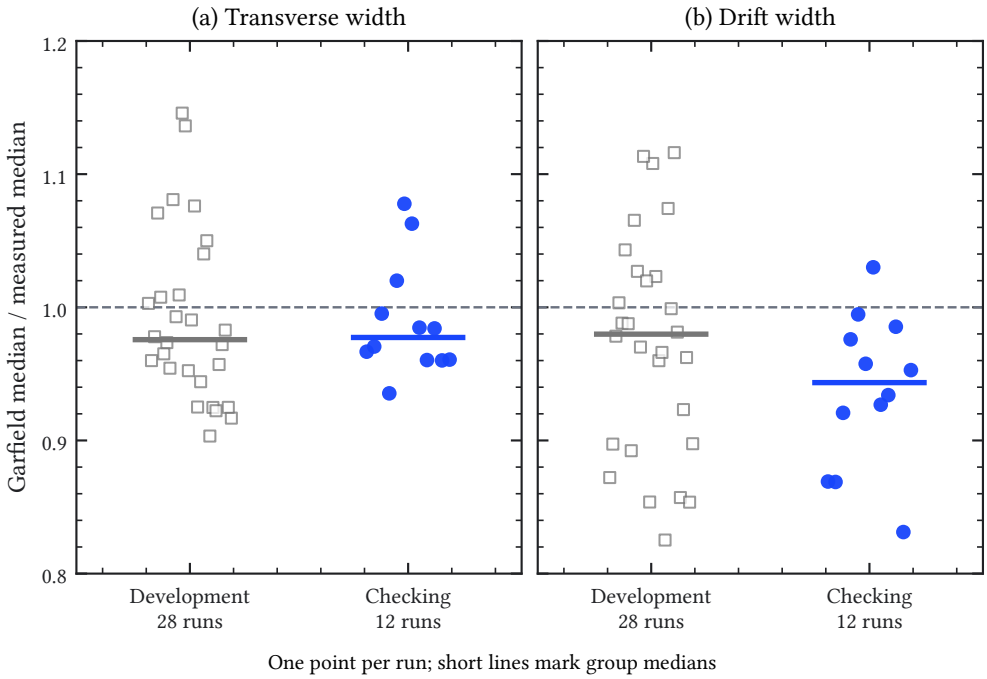


Figure E.2: Ratios of reconstructed width medians for the 40 completed runs before selection. Transverse width is the event mean of X and Y widths; drift width is the pooled Z width. Dashed lines mark equality and short lines the equal-weight run medians. The spread is run variation, with correlated simulation histories.

Simulation accepts more baseline events in 38 of 40 runs; among the 12 checking runs the median absolute discrepancy is 16.1 percentage points. Figure E.3 uses inclusive 2–7 keV, one track in each view and strict radius below 10 mm. The energy and radius discrepancies therefore remain material to selection efficiency.

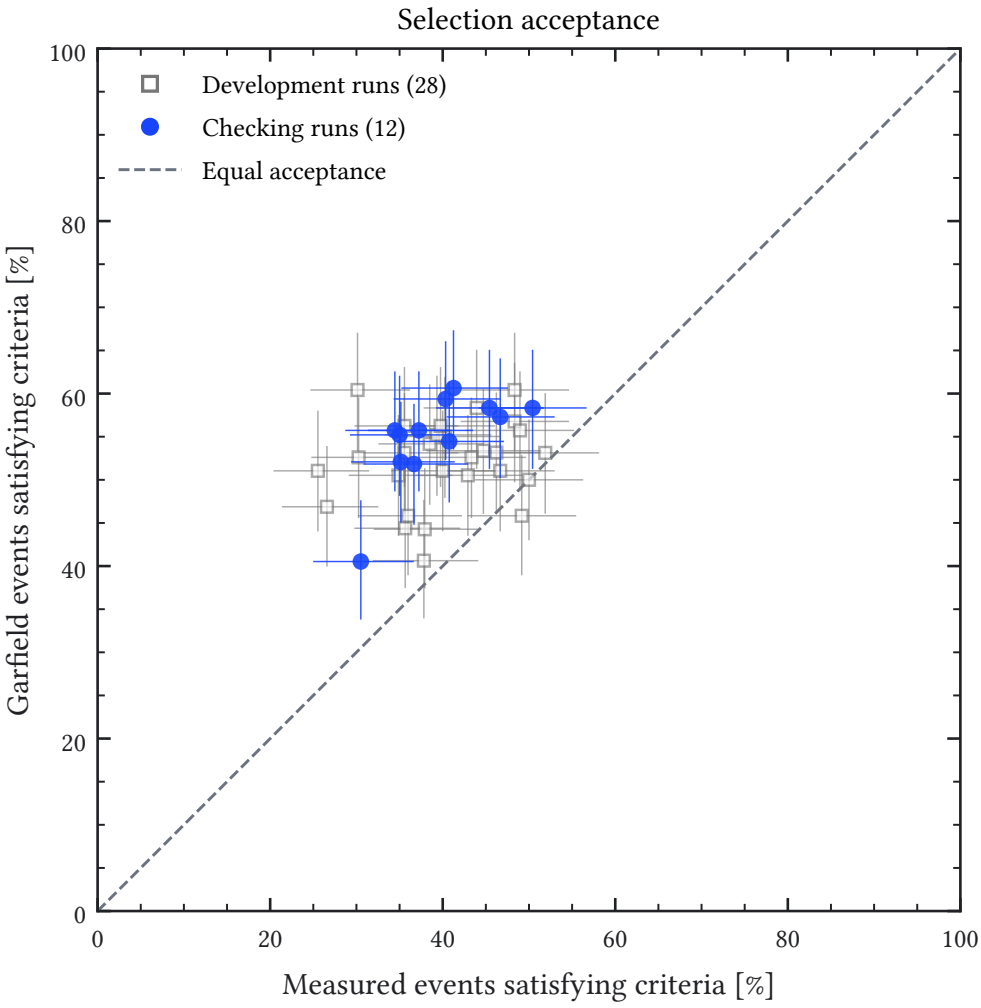


Figure E.3: Baseline acceptance among valid reconstructed checking events. Inputs contain 240 measured events and 192 simulated histories per run; reconstruction can reduce these populations. Bars are 95% Wilson counting intervals, excluding calibration, model and shared-history uncertainty. Failed fits and the excluded acquisition are omitted.

E.5 Timing comparison and admitted event populations

The timing comparison in Section 3.6.6 ends at 240 signed 16-bit channel waveforms, with 512 samples at 40 ns. Native reconstruction, veto response and diagnostic-only processes are excluded. Complete continuous batches include response initialization, input preparation and output; source generation, compilation, interpreter imports and subsequent output

verification are outside the case timer.

Calibration photons at 5.9 and 6.5 keV follow the detector axis from the nominal source position; gas-crossing controls use 3 GeV muons. These controls do not reproduce full source angular or energy distributions. The background list has 150 000 muons drawn from the Guan distributions and transported through the full shielding and veto geometry [16]. Both responses replay that fixed list; the shorter `Garfield` controls use the first particles of the respective REST lists. Ar/isobutane 99/1 at 1.4 bar is used in both calculations. Geometry copies change only the gas material.

The admitted physical region is the central 60 mm × 60 mm box within the 50 mm-radius gas cylinder. `Garfield` advances entries along straight trajectories to this box, omitting ionization before it and gas-to-surroundings transport feedback. REST gas deposits are restricted to the same region after readout rotation. Equal admitted area does not ensure identical microscopic paths or channels. Every generated particle remains in the source count; no-signal particles receive no artificial noise-only waveform.

Source	Response	Generated	Gas entry	Waveforms
Calibration photons	REST	500 000	495 279	383 879
	`Garfield`	512	508	427
Gas-crossing muons	REST	40 000	40 000	40 000
	`Garfield`	128	128	128
Muon background	REST	150 000	151	44
	`Garfield`	150 000	151	46

Table E.3: Distinct populations in the timing comparison. Each generated event starts with one particle. Gas entry includes secondaries and counts generated events with at least one gas-boundary particle. The waveform format is common; the charge-producing populations can differ.

The current microscopic model retains its calibrated amplitude and extra gain. REST retains its forward parameters, 12-bit amplitude reduction, zero internal noise and two 75-sample moving baseline corrections. Both use the same measured noise at final ADC conversion, with their respective analytic or empirical pulse descriptions. No parameter is fitted to a timing outcome. Neutrons are guarded by continued `Geant4` tracking: gas deposition or new gas secondaries cause explicit failure of the unsupported `Garfield` route. Three background parents have gas-reaching neutrons; 16 `Garfield` and seven REST gas-starting neutron steps produce neither local deposits nor new gas secondaries. The successful guard does not implement a general neutron charge response.

One physics worker is pinned to the same core of an AMD Ryzen AI 9 HX PRO 370, with one numerical-library thread. Wall times include waits and container launch and shutdown under concurrent host activity. The stage profile instead sums disjoint executable, shell-setup and coordinator CPU times, excluding container-service processes. Other processing includes

event bookkeeping and output rather than fixed initialization alone. Sequential waveform blocks reproduce small-batch channel arrays exactly, including per-event noise and ADC conversion; every output block is checked after timing.

The step timer assigns each step to its starting volume and includes copying step information, while excluding sensitive-hit recording and the user stepping action. For background, outside-gas stepping costs 544.37 s with REST and 540.84 s with `Garfield`, within complete CPU times 962.99 and 1123.53 s. Time before and after the transport event loop is 48.51 and 53.33 s, including initialization and final output. With profiling disabled, event-loop times are 729.68 and 707.69 s, compared with 911.90 and 899.85 s with profiling: observed increases of about 25% and 27%. No estimated overhead is subtracted. The controls preserve gas-entry and unsupported-particle records; REST also reproduces every waveform sample. The `Garfield` control repeats transport only, verifying its microscopic inputs.

E.6 Historical response and reconstruction controls

The historical response used `Heed` ionization and stochastic bulk drift, followed by an independently sampled bank of 2048 microscopic avalanches. Each arrival acquired the sampled charge, centroid displacement and mean delay; charge was normalized by the library mean before the fitted ADC conversion. Centroids were assigned to mapped pads within 0.251 mm. This approximation precedes the direct endpoint collection specified above. Its fitted R02756 comparison used 32 439 measured, 1000 conventional and 3842 library-response valid-track checking events. The nine marginal distributions entered its calibration, so their agreement cannot independently validate the gas transport.

The frozen cosmic transfer removed only the calibration-source illumination scale and retained an energy conversion of 5.775911241×10^{-4} keV/ADC and effective positive headroom of 4070 ADC. Two million generated muons yielded 1482 retained gas-boundary particles, 641 events with collected charge and 636 valid X/Y reconstructions. In R02760 the measured and simulated median main-track energies were 11.03 keV and 14.99 keV, with median track counts four and two. The chamber-only geometry and inclusive muon-rich background do not establish an absolute laboratory rate or a pure-muon efficiency.

Two reconstruction controls remain relevant to later comparisons. An exact DAQ-membership check with freshly initialized strip indices prevents an unmapped channel from inheriting the preceding channel's indices. Separately, the installed Gaussian hit-time fitter placed a right-hand crossing in its lower fit limit. A local opt-in correction, applied to both data and simulation, reduced paired noise-model changes in mean longitudinal width from −0.1502 to −0.0075 mm in R02704 and from −0.0650 to −0.0044 mm in R02756. These are intervention effects, not residual errors relative to data. The correction changes the measured reference as well; it is not substituted into historical response results without a matched recalibration. The complete historical fitting campaigns and workflow plates are preserved in

the separate development supplement.

E.7 Optical and waveform controls for the veto

The single-module calculation transports scintillation and Cherenkov photons through a 1500 mm × 200 mm × 50 mm plastic bar, optical coupling, two-piece Lucite guide, wrapping, window and photocathode. Reflection, refraction and absorption supply attenuation directly. A BC408-like proxy represents the aged NE-110 plastic [1]: the nominal inputs are 10 000 photons per visible MeV, refractive index 1.58, decay time 2.1 ns, Birks coefficient 0.126 mm MeV^{-1}, reflector reflectivity 0.97 and constant quantum efficiency 0.22. The 0.1 mm coupling layer and optical constants are assumptions, not measurements of the installed modules. Step-level `Geant4` Birks quenching is applied once [20]. A muon light scale cannot separately identify yield, optical losses, quantum efficiency and PMT gain, or validate neutron-recoil light production.

The measured inputs contain 8000, 3371 and 8394 saved events from R02760, R02761 and R02766, with 731 427 veto waveform rows. Thirty-seven channels occur in every event. Joint-channel background windows retain event identities, missing-channel masks, coherent noise and fortuitous activity; sparse channels are not labeled dead. The waveform comparison targets the saved, already corrected signals. For positive peak A, the amplitude-dependent negative lobe is modeled by

$$v_c(t;A) = Ah_c(t) - B_c \tanh\left[\frac{Au_c(t)}{B_c}\right], \tag{E.3}$$

where h_c and u_c have disjoint support. The shapes and floor B_c are fitted on R02760 development blocks. For simulated signals, the nonlinear term acts after summing photoelectron impulses; it describes an effective saved-waveform response. Figure E.4 shows why a linearly scaled template is insufficient.

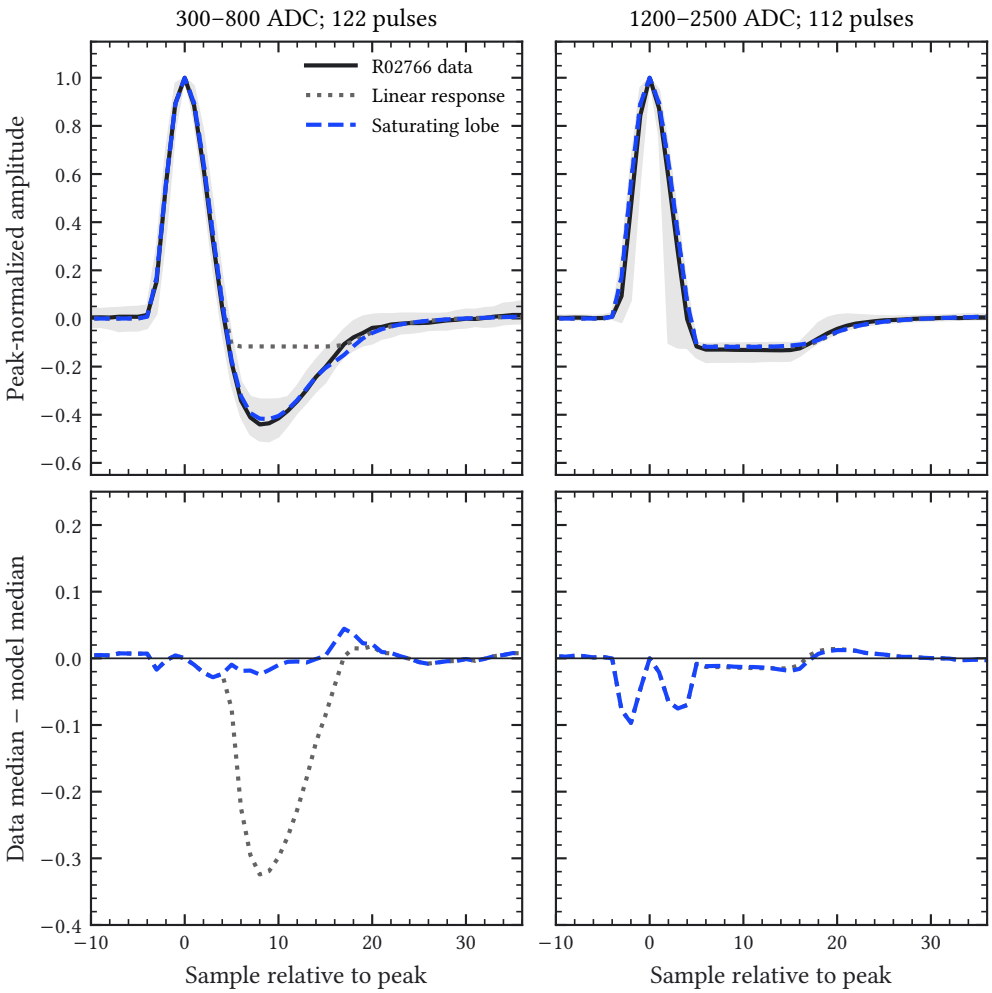


Figure E.4: Channel-4705 pulse shapes in R02766 for two amplitude populations. Black curves show measured medians; dotted and dashed curves show linear and saturating-lobe models. Shading is the central 16–84% population interval. Measured amplitude and peak time are supplied to the test, so it does not calibrate the energy scale. Lower panels show median residuals.

With the model fixed, R02761 contains 5681 evaluation pulses in 2453 events. Median event-mean normalized waveform RMSE decreases from 0.05465 to 0.04336, a 20.6% reduction; the paired bootstrap 95% improvement interval is $[0.00734, 0.00883]$. All 28 channels with at least 25 evaluation pulses satisfy the non-worsening guard, although rare large residuals persist. R02766 informed development and is not a blind sample. An independent tag requires at least three of Top L1, Top L3, Bottom L1 and Bottom L3 above 600 ADC, excluding the middle probe layers. Top-L2 responses above 300 ADC occur in 582/866 R02760 and 1649/2138 R02766 tagged events; equal pre/post windows contain 2/1 and 0/0 tags, respectively. This

supports prompt correlation without establishing muon purity or geometric efficiency.

E.7.1 Muon-reference scale and run dependence

A central normal 4 GeV muon through 50 mm of plastic provides a reference template, with 200 simulated events giving median deposit 9.724 MeV and 605.5 detected photoelectrons. The tabulated minimum-ionization loss, 2.019 MeV cm^{-1}, gives approximately 10.1 MeV across 5 cm [37]; this mean is not substituted for the fluctuating simulated deposit distribution. The reference trajectory does not describe every tagged crossing.

The conditional likelihood above 300 ADC combines the empirical photoelectron spectrum, median-one lognormal broadening and an exponential low-amplitude nuisance. Amplitudes at least 3500 ADC enter one right-censored bin. Even temporal blocks determine parameters; odd blocks check the spectrum, while all observed probes remain in the separate activity population. For R02760 channel 4705, 144 development and 134 checking pulses give 3.25 ADC per photoelectron (sampling-only bootstrap 95% interval 3.06–3.45). The reference conversion is 202 ADC/MeV, or 4.95 keV/ADC, with sampling-only interval 191–216 ADC/MeV. The lognormal width is 0.174, template fraction 0.848 and checking CDF distance 0.0755. The underpredicted upper tail prevents treating this anchor as a complete state model. Spectral broadening is not automatically reintroduced as intrinsic gain noise.

Normal-incidence 1–10 GeV scenarios give 202–203 ADC/MeV, whereas 40° crossings give 150–160 ADC/MeV. These geometry/optical scenarios exceed the sampling uncertainty and are not a confidence band. Channel 4705 exceeds 300 ADC in 278/866, 37/813 and 65/2138 tagged events in R02760, R02761 and R02766. Matching the four strongest tag panels leaves only eleven common strata: 405 R02760, 704 R02761 and 1866 R02766 probes. On this support the R02760 fraction is 3.70%; standardized R02761 and R02766 fractions are 1.37% and 1.60%. Most of the inclusive change therefore involves different tag populations; neither comparison identifies a hardware cause. High-amplitude or threshold-dominated cases are not assigned a gain merely because a flexible fit converges.

The lower-layer channel 4620 has a resolved population in all three runs. Its 16–22% censored fraction requires an explicitly exploratory allowance up to 35%, preserving the censored likelihood and the other screening requirements. Table E.4 gives this normal-trajectory anchor. The acceptance-aware alternative in Section E.8.2 uses conditional panel paths and retains unidentified channels as response hypotheses.

Run	Fit/check pulses	ADC/MeV [95%]	CDF distance
R02760	233/158	207 [192–224]	0.054
R02761	281/241	236 [227–247]	0.098
R02766	687/653	252 [242–262]	0.082

Table E.4: Channel-4620 normal-trajectory reference scales. Intervals include training-event and finite-reference-MC sampling, excluding trajectory, optical and selection uncertainty. CDF distances use fixed fits on odd blocks, including censoring.

E.7.2 Single-module neutron and sideband controls

Measured sidebands use two genuine 100-sample windows, samples 20–119 and 380–479, with baseline from samples 20–99 and coherent joint-channel records. Pre/post windows are evaluated separately and are not independent events or pure Poisson accidentals. At the intended 200 ns sampling, tested 5–90-sample gates span 1–18 µs relative to the incident neutron, not a TPC coincidence clock. The three-consecutive-sample score $S_i = \min(V_{i-1}, V_i, V_{i+1})/0.89$ uses the fitted channel-4705 pulse ratio; it suppresses isolated excursions without classifying neutrons. Thresholds below 300 ADC extrapolate below pulse training and are excluded. No boundary samples are synthesized.

R02760 even blocks and nominal simulated neutrons select working points under separate upper pointwise 95% Wilson sideband bounds. The objective averages signal-only tagging at 2 and 5 MeV with equal weights, which is a benchmark rather than a neutron spectrum. At a 5% budget the three-sample rule selects 300 ADC and five samples, tagging 2.00%; the largest-sample rule selects 1200 ADC and 90 samples, tagging 0.10%. At 1%, the former tags only one of 4000 benchmark neutrons and the latter has no nonzero-sensitivity candidate. Odd-block pre/post activity at the selected 5% point is 2.25/2.43%, 2.49/2.08% and 1.48/1.86% in R02760, R02761 and R02766. These probabilities are not measured dead time.

At 5 MeV, 79/2000 incident neutrons pass on signal alone. Overlay gives 133 vetoes, including 55 already caused by the measured background, leaving 78 additional vetoes: 3.90% with a pointwise finite-trial 95% interval 3.14–4.84%. Finite-bank reuse adds uncertainty absent from this interval. A combined veto fraction therefore cannot be labeled neutron detection.

Muon-anchored material scenarios vary k_B over 0.063, 0.126 and 0.189 mm MeV^{-1}, with or without 1 mm cadmium sheets on each face. Each new combination has 200 central reference muons and 500 neutrons per energy at 0.1, 1, 2 and 5 MeV; reused nominal points have 2000 neutrons. Refitting the same muon probes gives 3.17–3.32 ADC/photoelectron, close to nominal 3.25, without fitting neutron outcomes. Nevertheless, at 5 MeV the long-gate cadmium tag fractions are 12.8%, 4.10% and 0.40% for increasing k_B. At nominal k_B and 0.1 MeV, cadmium gives 71/2000 tags versus 0/500 without cadmium, whose two-sided 95% Wilson upper endpoint is 0.76%. These 18 µs-gate controls demonstrate model sensitivity; they

are distinct from the selected 1 µs working point and do not establish full-detector rejection.

E.8 Conditional full-geometry neutron response

The full-chain calculation connects saved shielded-detector neutron histories to a frozen TPC response and a muon-anchored optical veto. It predicts rejection conditional on reconstructed TPC events, rather than an absolute neutron-background rate. The predeclared 64-file HENSA production supplies 3864 saved source histories, with even files used for development and odd files for checking. All histories are kept without a truth-energy preselection. Saved histories are not all generated primaries and provide no exposure normalization. The separate run-259 diagnostic is excluded.

The archive uses REST 2.4.3, `CompleteVeto3Layers.gdml` (Appendix B.1), Livermore electromagnetic physics, `QGSP_BIC_HP`, high-precision neutron elastic scattering and radioactive decay. Gas and scintillator step limits are 0.05 and 0.5 mm. The broad recorded source selector, 1–100 000 MeV, does not establish its populated spectrum or absolute normalization. An identity-checked export contains 720 959 positive gas and 13 704 647 positive scintillator deposits, with audited energy sums and containment. Gas hits already carry Lindhard quenching and scintillator hits historical Birks quenching at $k_B = 0.126\,\mathrm{mm\,MeV^{-1}}$. The nominal response consumes that visible energy once. The removed raw intermediate is not recoverable independent unquenched truth.

E.8.1 Geometry, clocks and optical interpolation

The production defines 60 panels, whereas the laboratory records have 59 persisted positions. Forty-eight semantic panel assignments agree in center, length and PMT end; other mappings remain transfer assumptions and one production panel has no DAQ counterpart. Passive rotations satisfy $\boldsymbol{x}_{\mathrm{local}} = R(\boldsymbol{x}_{\mathrm{world}} - \boldsymbol{x}_{\mathrm{center}})$, so active directions require R^{T}. Geometry half-dimensions are in cm and exported world positions in mm. The $60 \times 60 \times 30$ mm gas box is rotated by 45°; local and world readout limits cannot be interchanged.

Raw hit times are in microseconds. Radioactive daughter clocks can reset, and the stored event-wide offset is unsafe for unrelated pending tracks: 4 850 447 offsets differ from recovered epochs. The exporter follows parent identities and advances epochs only at radioactive-decay reset edges, recovering 1356 such edges. Subtracting epoch and local time separately preserves prompt precision relative to the first gas deposit without folding remote activation into capture light.

Visible gas deposits generate one rounded, non-negative event-level Gaussian Fano count with $W = 26.37$ eV and $F = 0.1709$, allocated multinomially over deposit energies. Electron births use the recorded local positions and times. This point-deposit approximation does not repeat `Heed` absorption. Bulk drift, the 2048-avalanche library, strip mapping, pulse/noise response, gain dispersion, radial response, 4070 ADC headroom and energy conversion

$5.775911241 \times 10^{-4}$ keV/ADC remain frozen. Only the Fe55 illumination correction is absent. Native reconstruction uses the original R02756 metadata and historical ROOT 6.28/04 runtime; no corrected Gaussian estimator is substituted.

The veto optical kernel transports photons explicitly through bar, coupling, guide, wrapping and photocathode with `Geant4` 11.3.2. It covers 800 and 1500 mm bars, absorption lengths 1000, 1750 and 3800 mm, and 3×3×9 nodes with 40 000 photons per node. Seventy-two independent off-grid checks use 80 000 photons each. The standardized collection residuals have RMS 0.915 and maximum magnitude 2.55 (Figure E.5). These checks validate interpolation inside each assumed optical model. A twelve-panel replay from one source cascade gives residual RMS 1.26 and maximum 2.39 including shared-node uncertainty; its panels are not independent neutron validations.

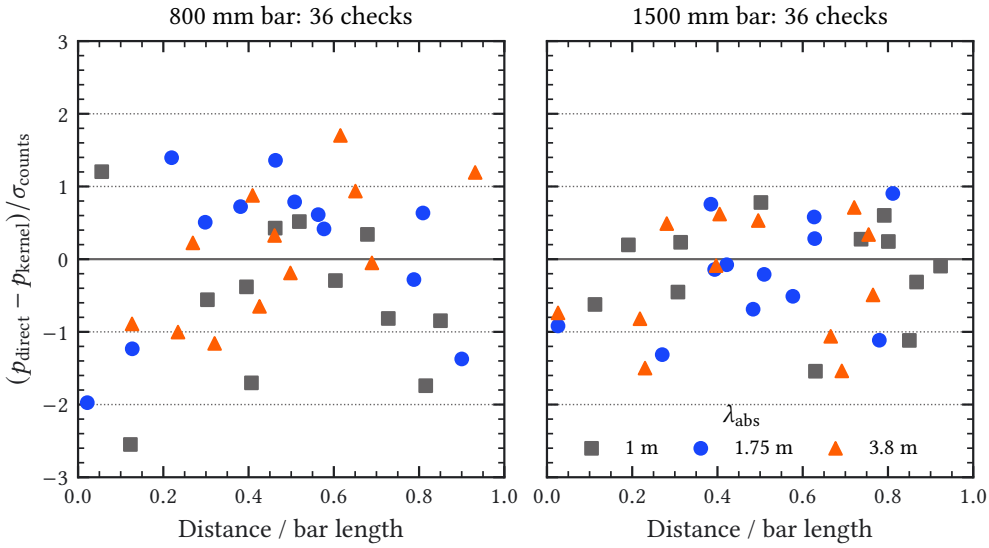


Figure E.5: Independent optical interpolation checks for two bar lengths and three absorption hypotheses. Residuals are divided by finite-count uncertainty. The 72 direct-photon histories test the kernel within fixed material and coupling assumptions; they are not experimental attenuation measurements.

Photons feed photoelectron charge and the effective bipolar pulse response. One joint measured record is added once; no separate accidental-muon population is superposed. Only the 37 commonly present channels supply instrumented rejection. Six require a transferred generic pulse template. Energy in other panels and absent measured channels retains explicit missingness.

E.8.2 Acceptance-aware muon anchor

Straight-ray intersections in the persisted laboratory geometry condition the muon reference on at least three independent tag groups. The nominal $\cos^2\theta$ intensity includes the projected-area cosine when sampling a horizontal plane. A 15 mm minimum tag path approximates its uncalibrated threshold; 5 and 30 mm, angular and source-position alternatives test this assumption. The reference visible-energy distribution scales with exact path relative to 50 mm and the optical kernel averages nine segment positions. Independent 20° and 40° controls reproduce median photoelectron counts within 1.63% and 0.76% (CDF distances 0.093 and 0.052). Median probe paths are 53.8 and 54.2 mm for channels 4705 and 4620.

Run	DAQ	ADC/PE	Sampling 95%	Check CDF
R02760	4705	2.776	2.655–2.961	0.070
R02760	4620	2.860	2.689–3.052	0.055
R02761	4620	3.241	3.075–3.444	0.106
R02766	4620	3.437	3.334–3.552	0.079

Table E.5: Conditional muon scales under the nominal optical/trajectory model. Even blocks fit the censored likelihood and odd blocks check fixed parameters. Intervals resample training pulses and the conditioned reference, excluding shared optical, angular and mapping uncertainty. Checking data were inspected previously.

These gains replace the normal-trajectory scales only in this scenario. Every absorption hypothesis is reanchored to the same probes. R02761's fitted mixture is weakly identified at its allowed upper fraction. Fine tag-pattern closure is poorer than group acceptance: only 202/866, 267/813 and 774/2138 measured tags fit one straight ray in the respective runs. Possible shower, pulse, selection and geometry contributions are unresolved; a left/right reflection scan is not adopted as a measured cable map. Only the four listed run/channel gains are individually supported. A shared-4620 anchor for other channels remains an exploratory hypothesis.

E.8.3 TPC population and coincidence reference

Five histories contain gas deposits separated by more than 10 µs, up to approximately 222 years, and are unsupported as single-trigger waveforms. The remaining 3859 histories span at most 4.404 µs. Missing native output has undefined energy rather than zero energy. Table E.6 keeps these populations separate.

Population	Histories/events
Saved source histories	3 864
Supported single-trigger histories	3 859
Native reconstruction rows	3 618
Valid X/Y reconstructions	3 575
Valid reconstructed 2–7 keV events	353
Even-file development / odd-file checking	183 / 170
Separate visible-gas-energy 2–7 keV diagnostic	749

Table E.6: Conditional full-chain neutron accounting. Visible-energy selection is an alternative diagnostic, not a stage of reconstructed selection. These counts do not provide incident-neutron efficiency or an absolute rate.

The unsupported histories contain ten gas clusters; two clusters have visible energy in the ROI but unknown reconstructed acceptance. Twenty-seven histories lie below the single-electron W scale, where the Gaussian Fano approximation is unsupported; none yields a valid ROI event. The ideal response boundary omits 3.42% of supported-sample visible energy in gas skins, which is recorded without projection back onto the readout. Production gas is Ar/isobutane 98/2 and frozen response gas 99/1. This material transfer, nuclear-recoil charge yield and acquisition triggering remain limitations despite verified diffusion and energy accounting.

For each event, $\bar{t}_i$ is the charge-weighted avalanche arrival relative to its first gas deposit, before waveform centering. The independent frozen muon bank supplies median $\bar{t}_\mu = 596.5865$ ns. An arrival-referenced coincidence hypothesis assigns

$$t_{ij}^{\mathrm{gate}} = t_{ij} - \bar{t}_i + \bar{t}_\mu. \tag{E.4}$$

At 200 ns per veto sample, this aligns the model near measured prompt bin 202. First-ionization timing is a separate scenario; neither is a measured hardware-trigger model.

E.8.4 Frozen selections and paired rejection

The three working points use 183 development ROI events and R02760 even blocks; 170 odd-file ROI events check the frozen choices. These valid, energy-selected events do not impose the final TPC multiplicity or fiducial baseline. The predeclared 288-candidate grid maximizes paired net tagging under upper pointwise 95% Wilson activity budgets in each development sideband. Those budgets do not guarantee actual-trigger X-ray acceptance or measured dead time.

Budget	Pulse score	k	m	Gate [bins]
1%	Three-sample minimum/0.89	15	2	$[-10, 11)$
5%	Five-sample matched amplitude	20	1	$[-5, 6)$
10%	Largest sample	30	1	$[-5, 76)$

Table E.7: Frozen selections. Channel thresholds are $\max(300, k\sigma_{\mathrm{train},j})$ ADC, with noise learned from development blocks only. Multiplicity $m = 2$ requires different panels within five samples. Gates are half-open relative to the peak; one sample is 200 ns.

Let b and c be veto decisions on the measured background alone and on its waveform sum with the neutron signal. The signal or summed waveform receives the nominal 3800 ADC positive ceiling once; the measured background is not clipped again. Genuine incremental rejection is

$$\Delta = \frac{N(c \wedge \neg b) - N(b \wedge \neg c)}{N_{\mathrm{ROI}}}. \tag{E.5}$$

The negative term retains decisions removed by bipolar superposition. Table E.8 compares neutron survival and actual-trigger Fe55-enriched control acceptance.

Budget	S_s	S_c	Δ [95%]	A_{2756}	A_{2754}
1%	70.0	67.6	32.4 [26.3–38.3]	98.94	99.10
5%	48.8	48.2	49.4 [40.8–58.4]	86.05	90.90
10%	37.6	32.9	55.9 [48.5–63.6]	81.61	86.03

Table E.8: All entries in percent. S_s and S_c are signal-only and signal-plus-background survival for the same 170 checking neutron ROI events. Intervals resample whole source files. A_{2756} and A_{2754} are acceptances in 473 and 1560 odd-block actual-trigger Fe55-enriched controls, not pure axion efficiencies.

The conservative point retains 468/473 and 1546/1560 controls, with separate pointwise 95% Wilson intervals 97.55–99.55% and 98.50–99.46%. Higher budgets lose considerably more controls than the sideband budget suggests. No run-specific retuning removes this discrepancy. New neutron-induced veto counts are 55, 84 and 95, while 0, 4 and 19 records are already vetoed by background alone. No lost veto is observed; zero losses in 170 trials have a two-sided 95% Wilson upper endpoint 2.21%, not a known zero probability. A bootstrap of observed zero counts cannot produce an unseen loss.

Two thousand whole-source-file replicas give the paired intervals above; whole-temporal-block resampling pools counts before calculating measured rates. Two additional 64-replica exercises hold MC fixed and either draw new records from the fixed odd bank or first resample its six temporal blocks. The resulting central 95% net-increment scatter is 29.4–32.4%, 46.2–51.2% and 53.5–61.2% for fixed-bank draws, and 29.2–32.9%, 46.2–50.6% and 53.9–60.6% after block resampling. These are distinct conditional experiments, not independent errors to sum

in quadrature. Six blocks and 64 replicas do not establish precise tail coverage. Previously inspected cross-run checks are not blind data; reserved-run outcomes were opened after freezing, although preparation had inspected coverage and metadata.

Of 1895 odd-file saved histories, 1740 yield valid reconstruction and 170 enter the unchanged ROI. Two are unsupported and 153 supported histories lack a valid result. Twelve ROI events have more than 15 keV of saved visible gas energy. This migration makes reconstructed energy essential to the denominator; saved gas energy cannot replace it or normalize to unsaved primaries.

E.8.5 Response uncertainty and limits of the prediction

Named response alternatives retain the same 170 ROI identities and paired background draws, with unchanged selection criteria. They vary unanchored gains by factors one half and two, absorption length to 1.0 or 3.8 m with matched muon anchors, angular intensity to $\cos\theta$ or $\cos^3\theta$, timing by ±0.8 µs and historical Birks coefficient by ±50% with anchors fixed. The latter is not a muon-preserving recalibration. The historical Birks estimator uses neighboring-hit distances or a 0.5 mm fallback rather than charged-step lengths. Its stable variation,

$$E_{\mathrm{vis}}(k_B') = \frac{E_{\mathrm{vis}}(k_B)}{1 + (k_B'/k_B - 1)q}, \qquad q = \frac{k_B E_{\mathrm{vis}}(k_B)}{L_{\mathrm{est}}}, \tag{E.6}$$

does not identify a measured recoil-light law.

Unanchored gains dominate the tested alternatives. At the conservative point, half and double gains change signal-only survival from 70.0% to 85.9% and 48.2%. The scenario diagnostic normalizes the paired net increment to background-accepted trials, rather than all ROI trials. Its net tagging changes from nominal 32.4% to 15.3% and 51.8%; reanchored absorption alternatives give 30.6–32.9% and fixed-anchor Birks alternatives 28.2–35.3%. These ranges are hypotheses, not calibrated confidence intervals. Removing positive headroom changes no nominal selected decision.

The historical estimator also quenches local neutral-track deposits using a neutron chord, which is not a charged-recoil stopping range. A separate `G4EmSaturation` neutral-deposit alternative changes the three signal-only veto counts by +1, −3 and zero, with finite photon-sampling variation. This does not validate either recoil-light treatment. The deposit handoff omits separate Cherenkov production; its normal-muon scintillation-only kernel gives 95.64% of the direct optical mean, including finite-library effects. Original charged-step extent, non-ionizing energy loss and local recoil identities are incompletely recoverable.

Two unsupported odd-file histories contain four gas clusters. Allowing zero to four additional ROI triggers with arbitrary veto outcomes gives combined-survival bounds 66.09–68.39%, 47.13–49.43% and 32.18–34.48% under one trigger per cluster. These accounting extrema are not confidence intervals and exclude noise-only triggers, gas-composition transfer and other unsupported response mechanisms. Muon-scale and X-ray-acceptance checks leave

recoil quenching, unidentified PMTs, geometry transfer, coincidence timing and absolute flux unresolved. The result is a conditional neutron-rejection prediction with explicit measured signal-control costs.

E.9 Extending the comparison with the common TPC baseline

The next comparison asks whether the same detector-response model can describe calibration and background after the final common TPC baseline, rather than only their inclusive observable distributions. The completed checks reported here extend the experimental sample and test a full-geometry simulation with the same nominal gas in particle transport and charge transport. They do not yet provide a new background-rejection factor or replace the results in Chapter 5.

E.9.1 Experimental background: more events, but few baseline survivors

Seven background runs were reconstructed from all entries in the selected saved-waveform products, using the same native reconstruction and exact TPC channel lookup. Their nominal conditions are Ar/isobutane 99/1 at 1.4 bar, with mesh and cathode voltage magnitudes of 335 and 800 V. The energy conversion from R02756 is transferred provisionally; agreement of nominal settings alone does not establish identical gain or trigger response.

Table E.9 applies the baseline of Section F.1 cumulatively: valid reconstruction, $2 \leq E \leq 7$ keV, one track in each readout projection, and the strict fiducial requirement $\sqrt{x^2 + y^2} < 10$ mm. Every saved event has a native reconstruction row, including those that fail validity. An independent calculation from the reconstructed branches reproduces all seven selection sequences.

Run	Role	Saved	Valid	Energy	One X/Y	Fiducial
R02720	D	3 655	2 823	1 007	105	1
R02721	C	26 400	21 039	6 499	864	5
R02733	D	3 518	2 440	812	102	1
R02739	C	2 326	1 430	512	36	0
R02760	D	30 802	29 872	7 641	1 038	7
R02761	C	3 371	3 366	825	104	1
R02766	C	8 394	8 238	2 099	311	0
Total		78466	69 208	19 395	2 560	15

Table E.9: Cumulative common-baseline counts in the extended experimental background sample. D denotes development and C denotes checking, with whole runs kept separate. These are saved-product counts, not exposure-normalized rates; the checking runs have been inspected historically and are not a new blind sample.

Only 15 events survive: nine in development and six in checking. The larger input sample therefore remains insufficient to establish strong additional topology rejection. Runs with smaller valid-reconstruction fractions have not been discarded to improve the agreement. Before interpreting pooled efficiencies, the transferred energy scale and the saved-event and trigger definitions must be checked run by run. R03015 and R03018 remain reserved.

A separate noise check samples 2048 uniformly spaced saved events per run, without energy or topology selection, across all 240 accepted TPC channels. Median channel noise RMS ranges from 5.00 to 5.82 ADC, with adjacent-sample correlations of 0.882–0.909. This supports retaining correlated noise in the response, but does not independently measure the original hardware pedestal: the stored waveforms may already have undergone baseline subtraction.

E.9.2 Gas-matched deposited-energy calibration controls

The full-geometry control and `Garfield` response both use Ar/isobutane 99/1 at 1.4 bar and assumed 293.15 K. Deposited energy generates primary electrons before drift, the avalanche library, the frozen electronics response and native reconstruction. The deposit is not transported again with `Heed` and recoil quenching acts once. Point-deposit ionization and recoil charge yield remain approximations.

Two additional ^{55}Fe controls each contain 1000 saved histories from the centered source at $z = 100$ mm. They use independent random seeds and the same detector response without refitting. The second lowers gamma, electron and positron secondary-production range thresholds from 1 to 0.01 mm, the minimum production energy from 1 to 0.1 keV, and the maximum gas step from 0.05 to 0.01 mm. Neither control applies the earlier empirical Fe55 illumination scale of 1.12.

Sample	Selected	E [keV]	Mean σ_{xy} [mm]	σ_z [mm]
Measured R02756	13 298	5.886	1.091	0.984
Earlier response	1 374	5.914	1.130	0.974
Nominal transport	372	5.948	1.100	0.959
Finer transport	390	5.933	1.110	0.997

Table E.10: Calibration medians after the final common baseline. The measured R02756 checking sample and the earlier `Garfield`-refined sample provide references for the new deposited-energy controls. All samples use native REST reconstruction. These conditional medians do not establish agreement before selection or of the complete distributions.

The widths after selection are encouraging, but the energy comparison illustrates why a selected peak alone is insufficient. For valid simulated events with true visible gas energy between 5.7 and 6.7 keV, before the reconstructed-energy, multiplicity and fiducial selection criteria, the nominal and finer controls give median reconstructed energies of 6.349 and 6.378 keV. Their median true visible energies are both about 5.901 keV: the reconstructed medians are 7.6–8.1% higher. The truth-energy window is a simulation diagnostic, not an experimental selection.

The selected medians therefore do not establish preselection energy agreement. No gain correction or finer transport setting is adopted from these controls. The few independent background survivors also prevent a precise additional topology-rejection measurement; survival in mixed laboratory background is not neutron-specific rejection.

Appendix F

Definition and Validation of the TPC Baseline Selection

F.1 Observable definitions and exact predicates . . . 143
F.2 Meaning of reconstruction validity and multiplicity . . . 144
F.3 Selection populations and efficiency denominators . . . 145
F.4 Individual criteria and correlations . . . 146
F.5 Samples, calibrations and reconstruction checks . . . 146
F.6 Checking-sample results . . . 147
F.7 Interpretation and preserved analysis record . . . 149

This appendix gives the operational definition and checking-sample results for the TPC baseline introduced in Section 3.5.3. The criteria were fixed before the evaluation reported here: usable maximum-track observables, reconstructed energy in the closed 2–7 keV interval, exactly one cluster per strip projection, and a maximum-track center within a radius of 10 mm. The radius is 1 cm, with normalization area $A_{\mathrm{fid}} = \pi$ cm^2. No veto or additional learned topology requirement enters the baseline. The diagram in Figure 3.15 remains the visual summary.

F.1 Observable definitions and exact predicates

All available TPC channels are reconstructed before event selection. A central waveform or hit mask would remove possible additional clusters and change the energy and multiplicity being tested. Hardware channel availability, waveform thresholds and reconstruction settings therefore accompany each sample rather than being inferred from its selected population.

The branch suffixes below share the prefix `tckAna2D_`. The largest-energy XZ and YZ clusters have native energies E_X and E_Y, stored as `MaxTrack_XZ_EnergyX`

and MaxTrack_YZ_EnergyY. The inspected native combined estimator MaxTrack_XZ_YZ_Energy is $E_\Sigma = E_X + E_Y$. The calibrated quantity is

$$E_{\mathrm{rec}} = c_r E_\Sigma, \tag{F.1}$$

where c_r is the accepted run or simulation calibration, fixed independently of this checking study. A native branch name does not establish keV units. This estimator differs from total gas deposition and from energy integrated inside a predefined readout circle.

The maximum-track mean coordinates are MaxTrack_XZ_MeanX and MaxTrack_YZ_MeanY, in millimeters. They define

$$r_{\max} = \sqrt{\bar{x}^2 + \bar{y}^2}, \qquad A_{\mathrm{fid}} = \pi (1\ \mathrm{cm})^2. \tag{F.2}$$

The origin is fixed at $(0, 0)$; no run-dependent recentering or additional z requirement is applied. The radius concerns the maximum-track center, not the all-hit centroid or containment of every hit.

Stage	Exact requirement
Valid, C_V	MaxTrack_XZ_OK, MaxTrack_YZ_OK and MaxTrack_XZ_YZ_OK are finite and each > 0.5; E_X, E_Y are finite and strictly positive; $E_\Sigma, E_{\mathrm{rec}}, \bar{x}, \bar{y}$ are finite.
Energy, C_E	$2 \le E_{\mathrm{rec}} \le 7$ keV.
Multiplicity, C_M	NTracksX= 1 and NTracksY= 1.
Fiducial, C_F	$r_{\max} < 10$ mm.

Table F.1: Frozen baseline $C_V C_E C_M C_F$. Energy boundaries are closed; the fiducial upper boundary is open. Branch suffixes share the prefix tckAna2D_.

Widths, skewness, hit balance, charge balance, pulse shape, saturation, maximum-track energy fraction, event identifier and classifier score supply no additional requirement. In particular, the earlier fifteen-feature finiteness gate is not part of this baseline. An unused undefined observable cannot remove an event from its denominator. Missing files or source records are reported as unavailable, rather than as selection failures.

F.2 Meaning of reconstruction validity and multiplicity

The OK flags in TRestTrack2DAnalysisProcess indicate that a listed maximum cluster has nonzero energy; the combined flag is their conjunction. They are not fit-quality probabilities and do not establish that the XZ and YZ maxima describe the same interaction. Explicit positive-energy and finite-observable tests define the usable numerical domain. An unavailable projection may be written as zero by the historical process; these placeholders

are not a measured central, zero-energy cluster. A source event returning no reconstructed row remains a reconstruction failure.

The native hits-to-track process clusters XZ and YZ hits separately with `clusterDistance`= 3.5 mm. Two hits are linked when

$$d_{ij}^2 < (3.5\ \text{mm})^2, \qquad d_{ij}^2 = \begin{cases} (x_i - x_j)^2 + (z_i - z_j)^2, & \text{XZ}, \\ (y_i - y_j)^2 + (z_i - z_j)^2, & \text{YZ}. \end{cases} \tag{F.3}$$

Connected components of these links define tracks. The relation is strict and transitive: a linked chain can span more than 3.5 mm, so this is not a maximum cluster diameter. An isolated hit also forms a cluster; the clustering routine supplies no minimum hit count or cluster-energy threshold.

`NTracksX` and `NTracksY` call the corresponding `GetNumberOfTracks` methods at the highest ancestry level present. The direct hits-to-track chain creates all clusters with parent identifier zero, so all belong to that level and are counted. A chain adding derived levels can change this meaning despite preserving the branch names. One cluster in each projection therefore does not mean one physical particle: separated energy deposits, noise or reconstruction fragmentation can create extra clusters, while nearby deposits can merge. Waveform thresholds, sampling, drift-velocity conversion and hit coordinates all affect the observed count. A disagreement at this stage remains a response diagnostic rather than a reason to change the criteria separately for data and simulation.

F.3 Selection populations and efficiency denominators

For an explicitly identified input population $\mathcal{S}_0$, the reporting order is

$$\mathcal{S}_0 \xrightarrow{C_V} \mathcal{S}_1 \xrightarrow{C_E} \mathcal{S}_2 \xrightarrow{C_M} \mathcal{S}_3 \xrightarrow{C_F} \mathcal{S}_4. \tag{F.4}$$

This order fixes the interpretation of intermediate losses without changing the final intersection. Writing $N_j = |\mathcal{S}_j|$ gives

$$\epsilon_{j|j-1} = N_j / N_{j-1}, \qquad \epsilon_{j|0} = N_j / N_0. \tag{F.5}$$

The conditional factors telescope to N_4/N_0 for the same population, without assuming independent criteria. A zero denominator gives an undefined efficiency. Efficiencies with different populations cannot be multiplied as if they were successive factors in this sequence.

For saved experimental analysis entries, N_0 counts those entries and the acceptance is conditional on acquisition and upstream processing. A complete waveform ledger can separately measure reconstruction losses; a surviving analysis tree cannot identify absent source records. Calibration acceptance also depends on source illumination and background

contamination and is not automatically the efficiency for all incident monoenergetic photons.

Absolute simulation efficiency requires every generated primary, including non-interactions and primaries producing no saved TPC event, with the source phase space stated. A transport-filtered or saved-waveform denominator gives conditional acceptance. Response replay preserves original identities and records reconstruction failures and unsupported domains separately. Delayed deposits in different triggerable windows remain distinct electronics events even if they share a primary. Absolute rates additionally require source activity or exposure, the 5 keV window width and the normalization area; neither area nor a uniform-area ratio determines efficiency for a nonuniform calibration source.

F.4 Individual criteria and correlations

An $N-1$ comparison omits one criterion while keeping the other baseline requirements. For criterion C_k on an evaluable domain $\mathcal{D}_k$,

$$\epsilon_k^{N-1} = \frac{N(\mathcal{D}_k \cap \bigcap_j C_j)}{N(\mathcal{D}_k \cap \bigcap_{j\neq k} C_j)}. \tag{F.6}$$

Energy, multiplicity and fiducial diagnostics retain C_V. Removing validity while replacing missing positions by zero does not define a valid diagnostic. Before C_M, the independently chosen maximum XZ and YZ clusters can come from different deposits, limiting the physical interpretation of their combined center in a multiplicity $N-1$ study.

The archived event record also retains separate energy and radius scans of one-track survival, joint (N_X, N_Y) counts, all eight energy/multiplicity/fiducial pass–fail combinations, and Spearman correlations before and after selection. Finite-pair counts accompany correlation estimates without imposing an undeclared all-feature finiteness gate. No two-dimensional energy–radius acceptance map or complete multivariate agreement is inferred from these one-dimensional checks. The compact figures below retain the quantities that determine the response comparison; the full distribution grids remain in the separate validation record.

F.5 Samples, calibrations and reconstruction checks

The sample assignments, source identities and original calibration records are preserved in the analysis manifest. Measured checking samples use odd blocks among twelve entry-rank blocks: canonical processed file 1 for R02756, and the available processed records for R02704. These contain 32450 and 11120 rows before validity, respectively. A 60-waveform control per run checked the archived native scalars and exact readout treatment; it did not reconstruct every waveform in either full run.

Garfield v8 checking samples contain 3843 rows per run, selected by odd collected-electron-bank index rather than numeric event-identifier parity. The same primaries receive the two frozen run-specific response settings. A 399-event control found unchanged scalar columns under the exact-readout check, validating that interface on the control sample rather than establishing detector-physics agreement. The REST detector-response comparator uses 1000 odd rows of 2000 sampled saved waveform events tuned for R02756. The R02704 comparator is exactly the same sample transferred without retuning; it is not a new simulation adjusted to R02704. Neither repeated comparator nor the shared Garfield bank supplies independent Monte Carlo evidence across runs.

Accepted factors c_r, in keV per native energy unit, are $5.775\,911\,24 \times 10^{-4}$ for R02756 data and Garfield, $5.854\,401\,08 \times 10^{-4}$ for R02704 data and Garfield, and $5.637\,236\,54 \times 10^{-4}$ for the REST comparator in both comparisons. No factor was fitted in this study; the manifests preserve full precision and original calibration records. All three populations use native REST reconstruction. The response labels identify the simulated detector model, while measured data supply laboratory waveforms to the same observable calculation.

F.6 Checking-sample results

Table F.2 preserves the cumulative event counts. Table F.3 gives baseline/valid retention and isolates multiplicity after energy and fiducial requirements. Pointwise 95% Wilson intervals are retained from this historical validation checkpoint; they measure conditional counting uncertainty rather than simultaneous coverage or detector-response systematics.

Sample	Input	Valid	Energy	One/view	Baseline
R02756					
Data	32 450	32 439	25 715	15 003	13 298
REST response	1 000	1 000	904	677	635
Garfield response	3 843	3 842	2 821	1 557	1 374
R02704					
Data	11 120	10 953	9 066	5 816	5 168
REST transfer	1 000	1 000	904	677	635
Garfield response	3 843	3 835	3 003	1 818	1 601

Table F.2: Cumulative selection counts for the frozen checking partitions. All samples use REST reconstruction. The R02704 REST row repeats the R02756-tuned sample without retuning; the Garfield samples share primaries.

Sample	Baseline / valid [%]	One/view given E, r [%]
R02756		
Data	41.0 [40.5, 41.5]	59.4 [58.8, 60.1]
REST response	63.5 [60.5, 66.4]	75.0 [71.9, 77.8]
Garfield response	35.8 [34.3, 37.3]	56.1 [54.2, 58.1]
R02704		
Data	47.2 [46.2, 48.1]	65.1 [64.0, 66.1]
REST transfer	63.5 [60.5, 66.4]	75.0 [71.9, 77.8]
Garfield response	41.7 [40.2, 43.3]	61.6 [59.7, 63.4]

Table F.3: Conditional retained fractions and 95% Wilson intervals. The final column uses the multiplicity $N - 1$ denominator, which differs from the ordered multiplicity stage. Shared simulations cannot be pooled as independent run samples.

The measured minus Garfield baseline-retention differences are +5.23 percentage points for R02756, with a 95% interval [+3.61, +6.82], and +5.44 for R02704, with [+3.61, +7.25]. The corresponding data-minus-REST differences are −22.51 [−25.48, −19.43] and −16.32 [−19.39, −13.15] percentage points. These within-run Newcombe intervals combine the two Wilson intervals for independent proportions; they do not make the shared simulations independent between run comparisons. Garfield is closer in total conditional retention, while still differing from the data.

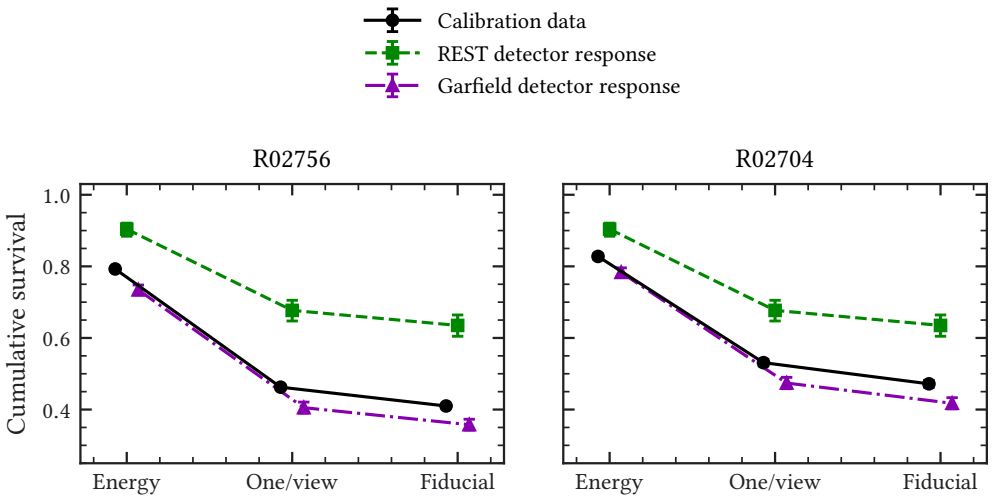


Figure F.1: Cumulative survival relative to each sample's valid reconstructed input, retaining the historical pointwise 95% Wilson intervals. Data are black markers; REST and Garfield detector responses use distinct colors, markers and line styles. R02704 REST is the unretuned R02756 transfer, and the Garfield panels share primaries.

The $N - 1$ checks identify the remaining discrepancy. After validity, multiplicity and

radius, the energy-window survival is 84.2% in R02756 data, 73.4% for Garfield and 89.6% for REST; R02704 gives 87.0%, 78.8% and the same 89.6% transfer. Thus Garfield loses more events at the energy boundary even though it better reproduces total baseline retention. After validity, energy and multiplicity, the fiducial survival is 88.6% in R02756 data and 88.2% for Garfield, and 88.9% versus 88.1% in R02704; REST gives 93.8% in both comparisons. These conditional fractions do not separately identify illumination, diffusion or readout-edge effects.

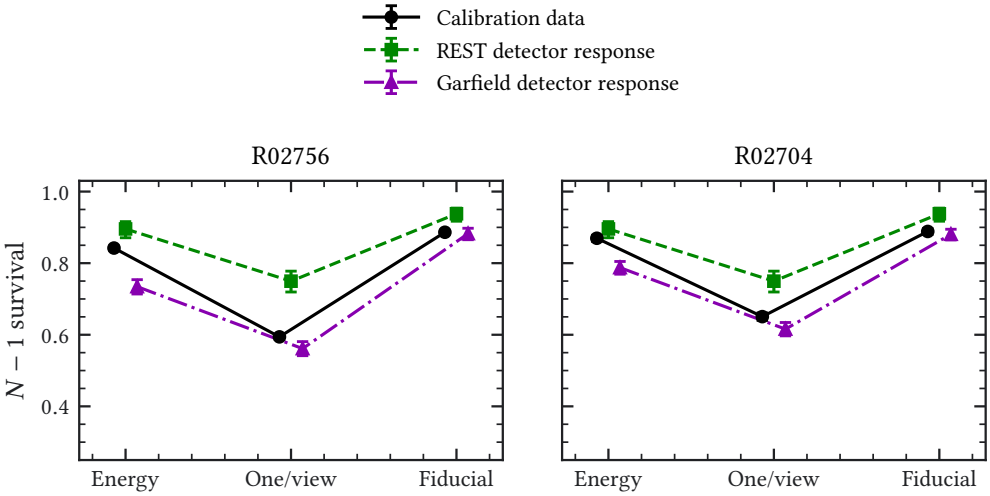


Figure F.2: Survival of each omitted criterion after all other baseline requirements, including validity, have been applied. Error bars retain the checkpoint's pointwise 95% Wilson intervals. Energy and fiducial survival use different populations from the multiplicity check. R02704 REST and both Garfield comparisons retain the shared-sample limitations of Figure F.1.

After the complete baseline, the largest data–Garfield Kolmogorov–Smirnov distances among the six checked observables are 0.085 for R02756 and 0.064 for R02704; these are distances, not p-values. The largest absolute Spearman-correlation differences are 0.162 for energy/radius in R02756 and 0.232 for radius/mean transverse width in R02704. Small marginal distances do not establish multivariate equivalence after different fractions of the original populations have been removed. The archived distributions retain all finite selected values in normalization, including tails outside plotted ranges.

F.7 Interpretation and preserved analysis record

This evaluation tests the frozen baseline, including its losses. Neither the energy window, radius, multiplicity criterion nor clustering distance is adjusted to suppress the discrepancies. A response change after inspection starts a new development iteration; renaming or repartitioning previously inspected runs does not make them blind. The reserved evaluation

samples remain outside this study.

The configuration, native-source audit, accepted calibrations, source hashes, event identities, criterion decisions and missing-event status remain with `code/baseline-cuts` and the response archive. Historical 15 mm readout-energy or hit-centroid selections and the earlier fifteen-feature finiteness gate are distinct methods. Repeated electronics realizations or simulations shared between runs retain parent/event correlations. Counting intervals, shape distances and correlation differences constrain different aspects of the response; none alone validates source composition, hardware triggering or absolute photon efficiency. Muon-rich laboratory background likewise does not become a pure tagged-muon sample through use of the common reconstruction chain.

Appendix G

Software Implementation and Reproducibility

G.1 Contributions and retained outputs 151
G.2 Reproducible production inputs 152
G.3 Readout metadata 153
G.4 Acquisition compression 154
G.5 Process-isolated `Geant4` prototyping 154
G.6 Exploratory learned waveform residuals 155

This appendix records the contributions and implementation choices that support Chapter 3. Detailed source examples, exploratory waveform plates and execution records are preserved in the separate software supplement. The physics response and its validation are documented in Appendix E.

G.1 Contributions and retained outputs

Tables G.1 and G.2 distinguish framework integration, original developments and operational tools. Quantified gains refer to their stated tests rather than every production.

Component	Contribution and evidence
REST analysis	Configured common RML chains and checked event/analysis-tree consistency against data.
`restG4`/`Geant4Lib`	Multithreading, pruning, subevent splitting, interrupt handling and hashed volume resolution; checked in cosmic productions.
Readout	Generated IAXO-D0/D1 strip and veto metadata; audited 60 simulation aliases, 59 prototype positions and 57 operating prototype channels.
TPC response	Developed `Garfield` response with direct amplification and individual collection; numerical and reconstructed checks in Section 3.6.
Geometry	Integrated the `Kotlin` GDML generator into `restG4`; recorded geometry revisions and checked channel correspondence.
Track pruning	Retained ancestry leading to relevant deposits; tested output reductions of 1–10 times while preserving the analyzed observables.
PDP generator	Developed the published enclosing-sphere trajectory sampler, with tested yield improvements of 3–37 times [38].
CRY source histograms	Generated reusable particle-dependent energy/angle distributions; compared with EXPACS, HENSA and cosmic productions.
Batch production	Developed job splitting, dependency management, staging and merging through `HTCondor`; production accounting accompanies Chapter 5.

Table G.1: Production and analysis contributions. Use of the REST framework and geometry integration is distinguished from development of the response and generators.

Tool	Contribution and use
`Uproot`/`fsspec`	Delegated file access to `fsspec`; upstreamed in `Uproot` 5.2.0 and used for remote waveform viewing.
Browser event viewer	Developed a `three.js` viewer and REST-to-JSON export for geometry, topology and veto-map inspection.
DAQ/viewer	Refactored `feminos-daq` for ROOT output, `Prometheus` monitoring [39], `CMake` and live waveform inspection during IAXO-D0/D1 operation.
`hvps`	Developed CAEN/iseg control, command-line and `Node-RED` interfaces; published packages and tested command generation.
`Python Geant4` wrapper	Developed process-isolated notebook execution and binary wheels; used for attenuation studies, outside production simulations.

Table G.2: Auxiliary, operational and prototyping software.

G.2 Reproducible production inputs

REST stores analysis configuration and framework/dependency metadata in ROOT files, enabling reanalysis from the saved event representation. An earlier discarded hit or waveform stage cannot be reconstructed from scalar observables alone. The open source, releases, containers and continuous-integration tests support reproducibility; we migrated the CI

infrastructure to `GitHub Actions`. A nominal release is insufficient when local patches, staged inputs or gas tables differ. Table G.3 specifies the required records; completeness remains a per-campaign question.

Record	Purpose and storage
Framework and local patches	Revision, build record and patch hashes identify algorithms and event conventions.
External software/data	ROOT, `Geant4`, `Garfield` and physics-data versions identify the transport and I/O environment.
Geometry and RML	Repository revision, staged hashes and serialized settings identify source, geometry, physics thresholds, response and reconstruction.
Source and gas inputs	File hashes and commands record source units, bin/angular measures, support and gas-transport tables.
Seeds and accepted jobs	Per-job identifiers, counters and accepted file list distinguish independent histories, failures and restarted products.
Exposure accounting	Generated-parent counters and source integrals connect selected events to physical exposure.
Final outputs	Merged metadata, analysis trees and merge logs identify saved observables and accepted inputs.

Table G.3: Records required to identify a production and its analysis. Their storage mechanisms do not establish that every historical file contains all inputs.

Stage	Configuration and check
Split/submit	`restG4ToCondor.py`: independent seeds and event/time budgets, optionally with DAGMan dependencies.
Transport	`restG4`: frozen RML, geometry, source and environment.
Process	`restManager`: fixed analysis RML applied to each transport output.
Merge	Accepted outputs combined with counters, identities and failed jobs retained separately.
Analyze	Saved observables determine selections; generated particles and source normalization determine rates.

Table G.4: HTCondor production stages. The generated-parent population is preserved through processing and merging rather than replaced by the saved-event count.

G.3 Readout metadata

`TRestDetectorReadout` maps DAQ identifiers to detector positions and types. Its planes specify world position, orientation, drift-side normal, active depth and semantic type; modules contain channels and elementary ownership polygons. Multiple pixels can belong to one strip or pad, including microbulk border pieces. Dedicated generation builds one plane per veto panel with its location, outward normal, depth and experimental alias. This permits physical layer and panel selections without external channel spreadsheets.

Forward hits-to-signal conversion and experimental waveform reconstruction use the same mapping. `TRestRawReadoutMetadataProcess` serializes channel information so later stages distinguish TPC and veto signals while combining waveform, topology and coincidence observables. Correct mapping must still be checked against each acquisition and geometry: a shared metadata class does not establish identical historical wiring.

The veto comparison uses separate maps for the 60 simulated panel aliases and the 59 positions in the archived IAXO-D0 paper readout. Two broken-channel placeholders remain defined in the latter, leaving 57 operating channels. The common 59-coordinate neural input is formed by intersecting aliases after transport and reconstruction; this does not remove detector material or define an operating-channel mask. Section A.2 gives the face-by-face counts, inactive identifiers and source records, and distinguishes the preserved 59-panel design and 60-panel September GDML revision from each other and from historical productions whose exact inputs must be checked separately.

G.4 Acquisition compression

The ROOT 6.32.02 benchmark in Table G.5 uses 52 194 calibration events, about 272 signals per event and 512 samples per signal. The default LZMA output is smaller than default ZLIB but slower to write; maximum LZMA compression is too costly for routine acquisition. Packing into bytes gave no practical compressed-size benefit over unsigned shorts. Downstream conversion to REST showed no measurable difference between settings. The selected compression therefore balances safe acquisition throughput and storage.

Samples	Codec	Write [s]	Read [s]	Size [GB]	Factor
`unsigned short`	ZLIB/default	142.86	51.88	5.61	2.41
`unsigned short`	LZMA/default	620.96	182.72	4.53	2.99
`unsigned short`	LZMA/9	4 621.88	144.56	3.76	3.61
`unsigned char`	ZLIB/default	196.59	75.03	5.74	1.77
`unsigned char`	LZMA/default	721.47	187.08	4.79	2.12
`float`	ZLIB/default	302.71	86.95	8.04	3.37
`double`	ZLIB/default	399.62	134.91	10.95	4.95

Table G.5: Acquisition compression benchmark. Factor refers to compression of the waveform-sample branch, not the complete file.

G.5 Process-isolated Geant4 prototyping

We developed `geant4-python-application` for small notebook and teaching studies. A `Python` proxy exchanges commands and event buffers with an independent

worker through a duplex pipe; `Pybind11` connects the worker to `Geant4` and `Application.run()` assembles nested `Awkward Arrays` [40]. Process isolation addresses repeated initialization in this particular wrapper, rather than a general toolkit limitation. Binary wheels provide pip installation on Linux, Windows and macOS; parallel notebook tasks launch separate workers.

The attenuation example transports 477.6 keV photons through lead in batches of 5000, up to 100 000, stopping when $\sqrt{N_{\text{valid}}}/N_{\text{valid}} < 1\%$. For fixed N independent photons the transmitted count is binomial, with standard-error estimate $\sqrt{\hat{p}(1-\hat{p})/N}$ and exact zero-count one-sided 68% bound $1 - 0.32^{1/N}$. The notebook's rare-transmission Poisson approximation, $1.14/N$, and adaptive stopping demonstrate execution but do not establish sequential confidence coverage [23]. The example does not normalize a production background result. REST's separate `PyROOT` interface uses C++ dictionaries; array operations or compiled backends avoid `Python` per-hit loop overhead in analysis.

G.6 Exploratory learned waveform residuals

A convolutional generative adversarial network adds a channel–time residual $N_{\text{GAN}}(c, t)$ to simulated calibration waveforms before baseline correction and reconstruction. Training uses unpaired simulated and measured batches, with constraints on pulse changes, correlations and spectra. The correction can absorb gain, shaping, diffusion or calibration mismatch; it is not uniquely identified electronics noise. A visually matched waveform illustration does not establish distribution-level agreement or unchanged selection efficiency. No validated predictive efficiency is assigned to this method. The current `Garfield` calculation uses a separate measured residual-noise bank; learned residuals are not part of its reported validation.

Bibliography

[1] K. Altenmüller, J. F. Castel, S. Cebrián, et al. "Background discrimination with a Micromegas detector prototype and veto system for BabyIAXO". In: *Frontiers in Physics* 12 (2024). DOI: 10.3389/fphy.2024.1384415 (cited on pp. 4, 28, 45, 88, 129).

[2] D. A. Brown et al. "ENDF/B-VIII.0: The 8th Major Release of the Nuclear Reaction Data Library with CIELO-project Cross Sections, New Standards and Thermal Scattering Data". In: *Nuclear Data Sheets* 148 (2018), pp. 1–142. DOI: 10.1016/j.nds.2018.02.001. URL: https://www.nndc.bnl.gov/endf-b8.0/ (visited on 08/30/2026) (cited on p. 6).

[3] Geant4 Collaboration. *Geant4 Neutron Data Library 4.6*. Neutron data library distributed with Geant4 11.0. 2021. URL: https://geant4.web.cern.ch/download/11.0.0.html (visited on 05/13/2026) (cited on p. 6).

[4] Geant4 Collaboration. *Geant4 Physics Reference Manual*. Release 11.0 documentation. 2021. URL: https://geant4.web.cern.ch/docs/ (visited on 05/13/2026) (cited on pp. 6, 69).

[5] International Atomic Energy Agency Nuclear Data Section. *EXFOR: Experimental Nuclear Reaction Data*. Dataset distributed through the EXFOR retrieval system. 2026. URL: https://nds.iaea.org/exfor/ (visited on 05/14/2026) (cited on p. 7).

[6] A. Takahashi, E. Ichimura, Y. Sasaki, and H. Sugimoto. *Measurement of Double Differential Neutron Emission Cross Sections for Incident Neutrons of 14 MeV*. Tech. rep. OKTAVIAN Report No. 87-03. Natural lead data distributed as EXFOR dataset 220750382. Osaka University, 1987 (cited on p. 7).

[7] R. B. Firestone and Z. Revay, eds. *Database of Prompt Gamma Rays from Slow Neutron Capture for Elemental Analysis*. IAEA STI/PUB/1263. Vienna: International Atomic Energy Agency, 2007. URL: https://www-nds.iaea.org/pgaa/ (visited on 08/30/2026) (cited on p. 9).

[8] L. McInnes, J. Healy, and J. Melville. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. 2018. DOI: 10.48550/arxiv.1802.03426. arXiv: 1802.03426 [stat.ML] (cited on p. 43).

[9] S. Navas et al. “Review of Particle Physics”. In: *Physical Review D* 110 (2024), p. 030001. DOI: 10.1103/physrevd.110.030001 (cited on pp. 50, 91).

[10] M. Jimenez Puyuelo et al. “IAXO-D1 Micromegas Setups at LSC and Zaragoza: Status of Background Model”. Presentation at the 23rd IAXO Collaboration Meeting, DESY. Mar. 2026 (cited on p. 55).

[11] G. Bohm and G. Zech. “Statistics of weighted Poisson events and its applications”. In: *Nuclear Instruments and Methods in Physics Research Section A* 748 (2014), pp. 1–6. DOI: 10.1016/j.nima.2014.02.021. arXiv: 1309.1287 [physics.data-an] (cited on p. 57).

[12] L. Pattavina, J. W. Beeman, M. Clemenza, et al. “Radiopurity of an archeological Roman lead cryogenic detector”. In: *The European Physical Journal A* 55, 127 (2019). DOI: 10.1140/epja/i2019-12809-0. arXiv: 1904.04040 (cited on p. 62).

[13] J. H. Hubbell and S. M. Seltzer. *X-Ray Mass Attenuation Coefficients: Lead, Copper and Material Constants*. Elemental lead and copper coefficients in Table 3; densities in Table 1. National Institute of Standards and Technology. July 2004. DOI: 10.18434/t4d01f. URL: https://www.nist.gov/pml/x-ray-mass-attenuation-coefficients (visited on 09/23/2026) (cited on p. 65).

[14] S. Westerdale and P. D. Meyers. “Radiogenic Neutron Yield Calculations for Low-Background Experiments”. In: *Nuclear Instruments and Methods in Physics Research A* (2017). DOI: 10.1016/j.nima.2017.09.007. arXiv: 1702.02465 (cited on p. 66).

[15] V. A. Kudryavtsev, P. Zakhary, and B. Easeman. “Neutron Production in α, n Reactions”. In: *Nuclear Instruments and Methods in Physics Research A* (2020). DOI: 10.1016/j.nima.2020.164095. arXiv: 2005.02499 (cited on p. 66).

[16] M. Guan, M.-C. Chu, J. Cao, K.-B. Luk, and C. Yang. “A parametrization of the cosmic-ray muon flux at sea-level”. Sept. 2015. arXiv: 1509.06176 [hep-ex] (cited on pp. 81, 127).

[17] C. Hagmann, D. Lange, and D. Wright. “Cosmic-ray shower generator (CRY) for Monte Carlo transport codes”. In: *2007 IEEE Nuclear Science Symposium Conference Record*. Vol. 2. 2007, pp. 1143–1146. DOI: 10.1109/nssmic.2007.4437209 (cited on p. 81).

[18] Á. Quero-Ballesteros et al. “Characterization of the Neutron Background Spectrum at DESY with HENSA++: Status and Future Steps”. Presentation at the 22nd IAXO Collaboration Meeting; preliminary results and systematic qualifications, pp. 12–17. Sept. 17, 2025 (cited on p. 82).

[19] J. B. Birks. *The Theory and Practice of Scintillation Counting*. Oxford: Pergamon Press, 1964 (cited on p. 87).

[20] Geant4 Collaboration. *Geant4 Book for Application Developers: Birks Quenching*. 2026. URL: https://geant4.web.cern.ch/documentation/dev/bfad_html/ForApplicationDevelopers/Detector/birks.html (visited on 05/09/2026) (cited on pp. 87, 129).

[21] G. Ros, G. Sáez-Cano, G. A. Medina-Tanco, and A. D. Supanitsky. "On the Design of Experiments Based on Plastic Scintillators Using Geant4 Simulations". In: *Radiation Physics and Chemistry* 153 (2018), pp. 140–151. DOI: 10.1016/j.radphyschem.2018.09.021. arXiv: 1804.08975 [astro-ph.IM] (cited on p. 87).

[22] T. Pöschl, D. Greenwald, M. J. Losekamm, and S. Paul. "Measurement of Ionization Quenching in Plastic Scintillators". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 988 (2021), p. 164865. DOI: 10.1016/j.nima.2020.164865. arXiv: 2007.08366 [physics.ins-det] (cited on pp. 87–88).

[23] G. Cowan and L. Lista. *Statistics*. Review of Particle Physics, 2025 update; review revised June 2025. Particle Data Group. 2025. URL: https://pdg.lbl.gov/2025/reviews/rpp2025-rev-statistics.pdf (visited on 09/05/2026) (cited on pp. 90–91, 155).

[24] F. Garwood. "Fiducial Limits for the Poisson Distribution". In: *Biometrika* 28.3–4 (1936), pp. 437–442. DOI: 10.1093/biomet/28.3-4.437 (cited on p. 91).

[25] F. Pedregosa, G. Varoquaux, A. Gramfort, et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830 (cited on p. 95).

[26] I. Waudby-Smith and A. Ramdas. "Estimating means of bounded random variables by betting". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 86.1 (2024), pp. 1–27. DOI: 10.1093/jrsssb/qkad009 (cited on p. 107).

[27] B. Singh and J. Chen. *ENSDF: 85Kr Beta Decay to 85Rb (10.739 y)*. Evaluated decay dataset; Nuclear Data Sheets 116, 1 (2014). National Nuclear Data Center, Brookhaven National Laboratory. 2014. URL: https://www.nndc.bnl.gov/nudat3/getdecaydataset.jsp?dsid=85kr+bM+decay+(10.739+y)&nucleus=85RB (visited on 09/21/2026) (cited on p. 109).

[28] P. Benetti, F. Calaprice, E. Calligarich, et al. "Measurement of the specific activity of 39Ar in natural argon". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 574.1 (2007), pp. 83–88. DOI: 10.1016/j.nima.2007.01.106 (cited on p. 109).

[29] T. Zerguerras, B. Genolini, V. Lepeltier, J. Peyré, J. Pouthas, and P. Rosier. “Single electron response and energy resolution of a Micromegas detector”. In: *Nuclear Instruments and Methods in Physics Research Section A* 608 (2009), pp. 397–402. DOI: 10.1016/j.nima.2009.07.015. arXiv: 0901.2882 (cited on p. 113).

[30] C. Margalejo Blasco. “Searching for solar axions with ultra-low background Micromegas detectors in the IAXO pathfinder system at CAST”. PhD thesis. Universidad de Zaragoza, Mar. 2026 (cited on p. 114).

[31] OpenJS Foundation and contributors. *Node-RED*. 2024. URL: https://nodered.org (visited on 05/20/2024) (cited on p. 115).

[32] P. Baron, S. Anvar, B. Blank, et al. “AGET, the Get Front-End ASIC, for the readout of the Time Projection Chambers used in Nuclear Physic Experiments”. In: *2011 IEEE Nuclear Science Symposium Conference Record*. 2011 (cited on p. 115).

[33] D. Calvet. “A Versatile Readout System for Small to Medium Scale Gaseous and Silicon Detectors”. In: *IEEE Transactions on Nuclear Science* 61.1 (2014), pp. 675–682. DOI: 10.1109/tns.2014.2299312 (cited on p. 115).

[34] L. A. Obis Aparicio. “Measuring electron drift velocities using a Micromegas detector”. Internal presentation of the CEA Saclay internship study. Dec. 20, 2022. URL: https://drive.google.com/file/d/1TECE-NTw-lrwP4QrvV4JHLu9SmnKRvNG/view (visited on 10/15/2026) (cited on p. 116).

[35] H. Schindler and R. Veenhof. *Garfield++ User Guide*. 2024. URL: https://garfieldpp.web.cern.ch/garfieldpp/ (visited on 05/20/2024) (cited on p. 118).

[36] D. Pfeiffer et al. “Interfacing Geant4, Garfield++ and Degrad for the Simulation of Gaseous Detectors”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 935 (2019), pp. 121–134. DOI: 10.1016/j.nima.2019.04.110. arXiv: 1806.05880 [physics.ins-det] (cited on p. 118).

[37] Particle Data Group. *Atomic and Nuclear Properties of Polyvinyltoluene*. Includes linked muon stopping-power and range tables. 2022. URL: https://pdg.lbl.gov/2022/AtomicNuclearProperties/HTML/polyvinyltoluene.html (visited on 09/12/2026) (cited on p. 131).

[38] D. Díez Ibáñez and L. A. Obis Aparicio. “Efficient cosmic ray generator for particle detector simulations”. In: *Computer Physics Communications* 317 (2025), p. 109805. DOI: 10.1016/j.cpc.2025.109805. arXiv: 2501.07452 [hep-ex] (cited on p. 152).

[39] B. Rabenstein and J. Volz. “Prometheus: A Next-Generation Monitoring System (Talk)”. Conference presentation. Large Installation System Administration Conference (Dublin). May 2015 (cited on p. 152).

[40] Pivarski, Jim, Elmer, Peter, and Lange, David. “Awkward Arrays in Python, C++, and Numba”. In: *EPJ Web of Conferences* 245 (2020), p. 05023. DOI: 10.1051/epjconf/202024505023 (cited on p. 155).