FROM NEUTRONS TO DARK MATTER: DIRECTIONAL RECOIL DETECTION
AND UTILIZATION OF DEEP LEARNING FOR GASEOUS TIME PROJECTION
CHAMBERS

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAI'I AT MĀNOA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

PHYSICS

AUGUST 2022

By

Jeff Schueler

Thesis Committee:

Sven E. Vahsen, Chairperson
Thomas E. Browder
Jason Kumar
Gary Varner
Peter Sadowski, University Representative

Keywords: Directional recoil detection, Convolutional neural network, Head-tail effect,
Belle II, SuperKEKB, Fast neutron backgrounds

To my parents, Andy and Patty Schueler,

for the unwavering support and encouragement you've continued to provide well into my adult life.

ACKNOWLEDGMENTS

First off, I would like to thank my dissertation committee for taking the time to provide feedback on my thesis.

In addition to the collaborators and mentors whose guidance and support helped steer me in the direction toward graduation, I would like to acknowledge the people outside of my work-sphere I have been lucky enough to have in my life and who continually remind me that there's so much more to life than work!

To my friends old and new, thank you for always challenging me, inspiring me, and making the six years I've spent in Hawai'i and the one year in Japan memorable, meaningful, and fun enough to give me the strength to pick myself back up whenever I hit the wall and push through the marathon of grad school. I'd like to thank my former roommate Rob for his wisdom, the stimulating conversations we would have, and of course the many, many hikes up Koko Head. I'd also like to thank Alex Whys, Ben, Sam, and Travis for the frequent hang outs early on in graduate school. I'd especially like to thank Travis for all of the heart-to-heart conversations, long runs together, and frequent phone calls. I am also grateful to my current roommates Hosanna, Suzanne, and Rena for making our house feel like home. Thank you to the old friends who were able to pay me a visit while living in Hawai'i: Pat Leifer, Yan, Steve, Rafi, Alex Sanoja, Loc, Ryan, Allen, Cody, and Edith, it was awesome getting to see you all here! I'd additionally like to thank Alex Sanoja, Cody, Allen, Rafi, and Patrick Yagi for the two year stretch of weekly Dungeons and Dragons sessions. As difficult as it was to schedule those while I was flying back and forth between Hawai'i and Japan, those sessions gave me something to look forward to every week. And finally, one more thank you to Alex, Steve, and Rafi for the semi-regular phone calls.

The year I spent in Japan, though an experience I would never take back, was often times very lonely, so I am grateful that I was able to reconnect with older friends. Thank you Reid, Dezzie, Ryan, Patrick, and Kendra for meeting up with me either while living there or while I was visiting for an extended period of time. Thank you to the Hawai'i crew (Sam, Ben, Travis, and Alex Whys) who also paid me a visit there! I am particularly grateful to the friendships I made while working at KEK. Caleb, Hendrik, Zac, and Cate, thank you all for the bike rides to and from work, the hikes, the meals, the runs, the trips to Tokyo, and of course the weekly rounds Imperial Assault. I'd like to give an additional thank you to Cate for all of the memories we shared; you left this world too soon and are dearly missed.

Thank you to my parents Patty and Andy, my brother Michael, and sister in-law Kaitlin. I love you all very much and feel very lucky to have such a loving and supporting family. I am also extremely grateful for the love and support of my partner Casey. Thank you for all of the laughs we share, the weekly date nights, the fun weekend activities, and reminding me to take a step back and enjoy the time I'm spending here in Hawai'i with you. I love you!

putting up with my sometimes way-too-long presentations at group meeting!

Finally, I would like to thank my advisor, Prof. Sven Vahsen who I first met when teaching the recitation session for his quantum mechanics class in the Spring of 2016. I came into graduate school without much of an idea of what I wanted to do research wise but knew I wanted to work with somebody who was supportive and who I got along with. Sven believed in me from the start and asked if I would be interested in working for his group and for that I am extremely grateful. Since then, Sven has never ceased to be supportive of my goals, even if they deviate from our initial plans, and to this day, I wonder how he finds the time to engage in lengthy but useful discussions with me about minutia. I continue to learn a lot from Sven and am so grateful to have joined his group and had the opportunity to engage in work that I have come to love.

ABSTRACT

Modern gaseous time projection chambers (TPCs) with high readout segmentation are capable of reconstructing detailed 3D ionization distributions with voxel sizes of order $(100\,\mu m)^3$. This enables measurements of the 3D momentum vectors of short, mm-scale nuclear recoils, which is of interest for neutron measurements, as well as searches for dark matter, where directionality opens the possibility of identifying the galactic origin of weakly interacting massive particles (WIMPs), even below the so-called neutrino floor. We perform a variety of experiments and simulations with eight miniature TPCs filled with a 70:30 mixture of He:CO_2 gas at 1 atm pressure. Each so-called BEAST TPC is of identical design and contains two gas electron multiplier (GEM) amplification devices and a (2.00×1.68) cm^2 pixel-ASIC readout.

We first detail the measurement of neutron backgrounds at the SuperKEKB e^+e^- collider in Tsukuba Japan. We focus on measurements surrounding SuperKEKB's final focusing magnets (recorded in 2018) and in the accelerator tunnel surrounding the Belle II detector (recorded in 2020-2021). In our analyses we reject large X-ray backgrounds from the accelerator, resulting in >99% pure samples of nuclear recoils down to recoil energies as low as $8.0\,\mathrm{keV_{ee}}$. We find excellent agreement between measured and simulated nuclear recoil energy spectra indicating that our simulations model neutron production well. We additionally introduce a correction for charge integration bias in observed recoil tracks with high axial inclination. This correction leads to correct vector directional "head-tail" (sign of 3D vector) assignment for 91% of simulated He recoils ranging from $40\,\mathrm{keV_{ee}}$ to about $1\,\mathrm{MeV_{ee}}$, with a mean angular resolution of 8°; a significant improvement over the 72% head-tail efficiency achieved without these corrections. Applying this technique to measurement leads to an agreement between measured and simulated angular distributions that allows us to conclude the existence of a neutron production hotspot in the accelerator tunnel.

While the BEAST TPCs are highly sensitive to ionization, and can detect even single electrons, extending directionality to the keV-scale, as is desirable for dark matter searches, requires operating the detectors with lower-density gases, at higher gains, and developing improved analysis techniques. We here focus on the two latter aspects. We improve on existing head-tail classification methods through the introduction of deep-learning computer-vision algorithms called 3D convolutional neural networks (3DCNNs). We first perform a simulation benchmark study where we train a 3DCNN to assign directional head-tail to simulated neutron recoils with energies up to $515\,\mathrm{keV_r}$ and compare these results to three existing methods of head-tail assignment. We find a head-tail efficiency of 99.9% on this sample using the 3DCNN, compared to 97.8%, 93.7%, and 79.0% for existing methods.

Next, we measure neutrons from a ^{252}Cf source incident on separate sides of a TPC. We operate both at low gain and high gain. At low gain, the simulation-trained 3DCNN reliably identifies whether the observed recoil points toward or away from the ^{252}Cf source. On a small sample of

identified He recoils between $39\,\mathrm{keV_{ee}}$ and $49\,\mathrm{keV_{ee}}$, before correcting for residual background such as back-scattered events, we observe a head-tail efficiency of $(62.1 \pm 11.4)\%$. Using simulation, we show that the true head-tail efficiency after correcting for residual backgrounds should be greater than this, marking the first statistically significant observation of event-level head-tail sensitivity below $50\,\mathrm{keV_{ee}}$. At high gain, we attempt to improve our head-tail sensitivity to sub-10-$\mathrm{keV_r}$ recoils, and also introduce a 3DCNN for event identification. In simulation, we reject all X-ray backgrounds down to $5\,\mathrm{keV_{ee}}$ at 50% nuclear recoil selection efficiency and demonstrate head-tail efficiencies above 50% for He recoils down to $3\,\mathrm{keV_r}$. These results do not yet generalize to measurement, which is currently being investigated. If the 3DCNN robustness can be improved, this would be the first demonstration of directional recoil detection at energies relevant for the directional detection of $\mathcal{O}(\mathrm{GeV})$ dark matter particles.

Finally, we perform a study comparing the keV-scale electron background rejection performance of a 3DCNN to the traditional discriminant of track length, as well as discriminants obtained from state-of-the-art shallow learning methods in a simulated detector with an 80:10:10 mixture of $\mathrm{He:CF_4:CHF_3}$ at $60\,\mathrm{torr}$. We train the 3DCNN classifier using recoil charge distributions with ionization energies ranging from 0.5-10.5 $\mathrm{keV_{ee}}$ after $25\,\mathrm{cm}$ of drift. The charges are initially segmented into $(100\,\mathrm{\mu m})^3$ bins when determining track length and the shallow learning discriminants, but are rebinned with a reduced segmentation of $(850\,\mathrm{\mu m})^3$ for the 3DCNN. Despite the courser binning, compared to using track length, we find that classifying events with the 3DCNN reduces electron backgrounds by an additional factor of up to 1,000 and effectively reduces the energy threshold of our simulated TPC by 30% for fluorine recoils and 50% for helium recoils. We also find that the 3DCNN reduces electron backgrounds by up to a factor of 20 compared to the shallow machine learning approaches, corresponding to a $2\,\mathrm{keV_{ee}}$ reduction in the energy threshold.

Collectively, the results in this thesis highlight the unique measurements enabled by high-resolution ionization imaging, and how 3D convolutional neural networks appear ideally suited to maximally utilize the rich 3D data from detectors with this capability.

CONTENTS

LIST OF TABLES

xiii

xiv

LIST OF FIGURES

xxi

CHAPTER 1
BACKGROUND AND OVERVIEW

To provide context for the original work presented in later chapters, we briefly introduce here dark matter, directional detectors and their associated performance metrics, and give an overview of the content covered in this dissertation.

1.1 Dark matter

Astrophysical observations such as discrepancies in mass-to-luminosity ratios of galaxy clusters [1], and the observation of relatively flat galaxy rotation curves [2] imply the existence of non-luminous mass around galaxies and clusters that is now referred to as dark matter (DM). By the early 1980s, it was generally accepted that DM comprises the majority of mass in the universe. Around this time, so-called hot dark matter (HDM) cosmological models where DM is an ultrarelativistic low-mass particle, such as a neutrino, were mostly ruled out [3] in favor of cold dark matter (CDM) models, where dark matter is a heavier, sub-relativistic particle that interacts very weakly with ordinary matter. Small temperature anisotropies in the cosmic microwave background (CMB) that were predicted by CDM models, were confirmed by COBE in 1992 [4], further strengthening the CDM hypothesis. Precise measurements of the CMB temperature power spectrum were later released by WMAP in 2008 [5] and Planck in 2016 [6], and are in remarkable agreement with the predictions of the ΛCDM cosmological model (Λ is the cosmological constant that accounts for the accelerating expansion of the universe [7; 8]). While ΛCDM may not be in agreement with all cosmological observations [9], it is generally referred to as the standard model of Big Bang cosmology and it predicts that there are DM halo and subhalo structures that may contain galaxies [10]. Observations of orbits of halo stars in our galaxy have been used to constrain the mass of the Milky Way dark matter halo and determine a local DM density at the sun's distance from the galactic center of $\rho_0 = 0.35^{+0.08}_{-0.07}$ GeV/cm^3 [11].

While there is strong cosmological evidence for both the existence and abundance of DM, the true nature of DM remains to be determined and is one of the most fundamental open questions in particle physics. Over the past three decades, many experiments have investigated the hypothesis that DM consists of previously undiscovered elementary particles. While the setup and technologies used in these experiments are diverse, they can be broadly grouped into the following three categories (Figure 1.1):

(A) **Direct Detection**: $DM + SM \rightarrow DM + SM$

(B) **Indirect Detection**: $DM + DM \rightarrow SM + SM$

(C) **Collider Searches**: $SM + SM \rightarrow DM + DM$ or $SM + SM \rightarrow SM + SM + DM + DM$.
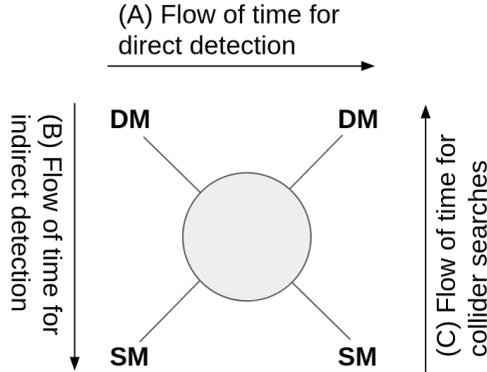
Figure 1.1: Dark matter (DM) and standard model (SM) particle interaction schematic highlighting three processes depending on the direction we consider time to flow. Case (A), where we consider time flowing to the right, is representative of direct detection of dark matter where a DM particle scatters off an SM particle, causing the SM particle to recoil. Case (B), where we consider time flowing downward, illustrates two DM particles annihilating into two SM particles. Measurements of the SM particles could lead to an indirect detection of DM. Finally, Case (C), where we consider time flowing upward, illustrates a particle collider DM search where two SM particles interact leading to the pair-production of two DM particles.

We're interested here in direct detection, where a DM particle scatters off a nucleus in the detector's target material, leading to a measurable nuclear recoil.

Direct DM detection experiments rely on the following assumptions [12]:

(1) There is a nonzero local DM density, ρ_0.

(2) The DM halo is relatively stationary with respect to sun's rotation around the galactic center.

The recent observations of Ref. [11] support (1), and while (2) is currently lacking observation, N-body simulations based on the predictions of ΛCDM predict figure rotation of DM halos associated with Milky Way-like galaxies to be very small [13], suggesting that the Milky Way is not co-rotating with the DM Halo.

1.2 Directional dark matter and neutron detection

The weakly interacting massive particle (WIMP) remains a compelling CDM candidate [14] that is expected to interact with SM particles via elastic scattering with nuclei [15]. If the DM flux incident on Earth is composed of WIMP-like particles, then we can expect that some of these DM particles will scatter off nuclei leading to nuclear recoils with energies ranging from $\mathcal{O}(1-100)$ keV.

Assuming the sun to travel along the galactic plane, then assumptions (1) and (2) imply that in the rest frame of our solar system, there is a DM velocity distribution incident on our solar system

coming roughly from the direction of the Cygnus constellation. The inherent time dependence of the DM velocity distribution observed on a terrestrial detector will modulate annually due to the Earth's orbit around the sun, and will also modulate over a sidereal day [16]. The diurnal modulation of event directions in detector coordinates (Figure 1.2) is equivalent to the presence of a galactic dipole (Figure 1.3) in galactic coordinates, with clearly distinct directional fluxes of DM from Cygnus and neutrinos from the sun; the most prominent anisotropic background expected in a direct DM detector.

Neutrinos can produce backgrounds in a direct DM detector via a process called coherent neutrino-nucleus elastic scattering (CEνNS) [17]. The energy spectrum of these nuclear recoils can mimic that of recoils from DM-nucleus scattering, leading to the so-called neutrino floor [18]. If a detector has sufficient ability to resolve nuclear recoil directions and event-times, however, then it should be possible to reconstruct the galactic dipole, which would lead to an unambiguous directional signal confirming the galactic origin of DM. Figure 1.2 illustrates this by showing the expected ϕ distribution of DM-induced nuclear recoils (blue) and CEνNS-induced nuclear recoils (red) in a detector at times t_0 and $t_0 + 12\,\text{h}$. A detector with event-level timing and the ability to resolve angle ϕ could then distinguish between DM and CEνNS-induced nuclear recoils over the course of a sidereal day and could also transform these events into galactic coordinates. If the detector additionally has *head-tail sensitivity*—the ability to distinguish between recoil direction $\hat{\mathbf{q}}$ and the opposite direction $-\hat{\mathbf{q}}$—then it should be possible to resolve the so-called galactic-dipole shown in Figure 1.3. In galactic coordinates, observed CEνNS backgrounds from solar neutrinos will vary along the Ecliptic over the course of the year, while DM particles will remain stationary near the Cygnus constellation, providing an unambiguous directional signal of the galactic origin of DM. Without head-tail sensitivity, the dipole shown in Figure 1.3 would be smeared across two hemispheres, reducing the contrast between the reconstructed neutrino and DM distributions, leading to an increase in the number of DM recoils required to reject a neutrino background hypothesis.

We're specifically interested here in *directional* DM detection, where directional gaseous time projection chambers (TPCs) [19] are leading detector candidates. Large scale efforts are currently underway for constructing large ($> 10\,\text{m}^3$) 3D directional gaseous TPCs [20]. Following the approach first introduced in Ref. [20] and later expanded upon in [12], we choose to characterize directional performance using two independent measures called *angular resolution* and *head-tail recognition efficiency*. Here we carefully define these directional performance metrics, as well as two additional metrics that are important for directional DM detection:

Angular resolution: The mean difference in 3D angle between the axis of the true recoil track and the axis of the reconstructed observed track. Angular resolution does not depend on sign of the recoil vector, so it can take on values ranging from 0 to $90°$, however, the expected angle between two random *axes* is defined on a hemisphere and given by $\langle \theta \rangle = \frac{\int_0^{\pi/2} \mathrm{d}\theta \, \theta \sin(\theta)}{\int_0^{\pi/2} \mathrm{d}\theta \sin(\theta)} = 1$, so we take
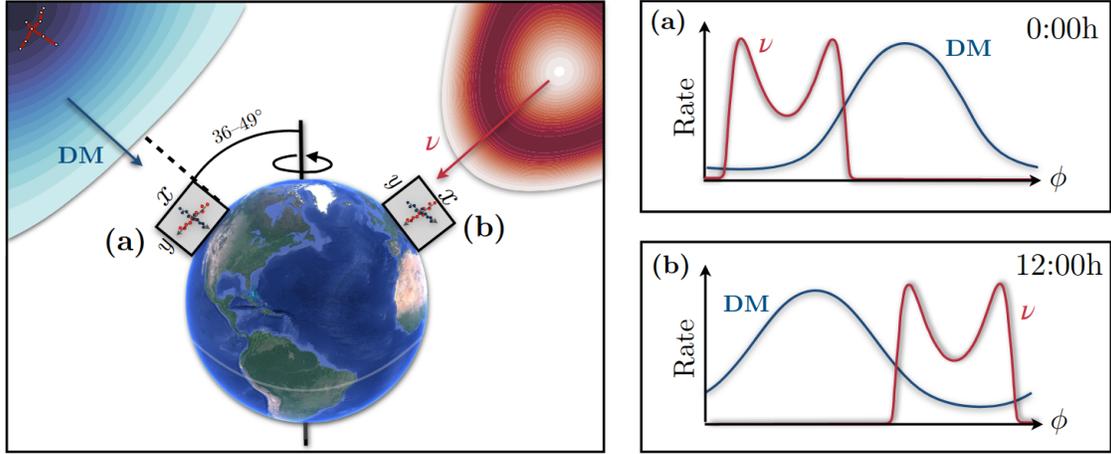
Figure 1.2: From Ref. [12]: Orientations of DM-induced and CEνNS-induced nuclear recoils observed in a terrestrial detector spaced 12 hours apart. The neutrino and DM-recoil signals will oscillate over the day as a function of ϕ but will remain distinct from one another.
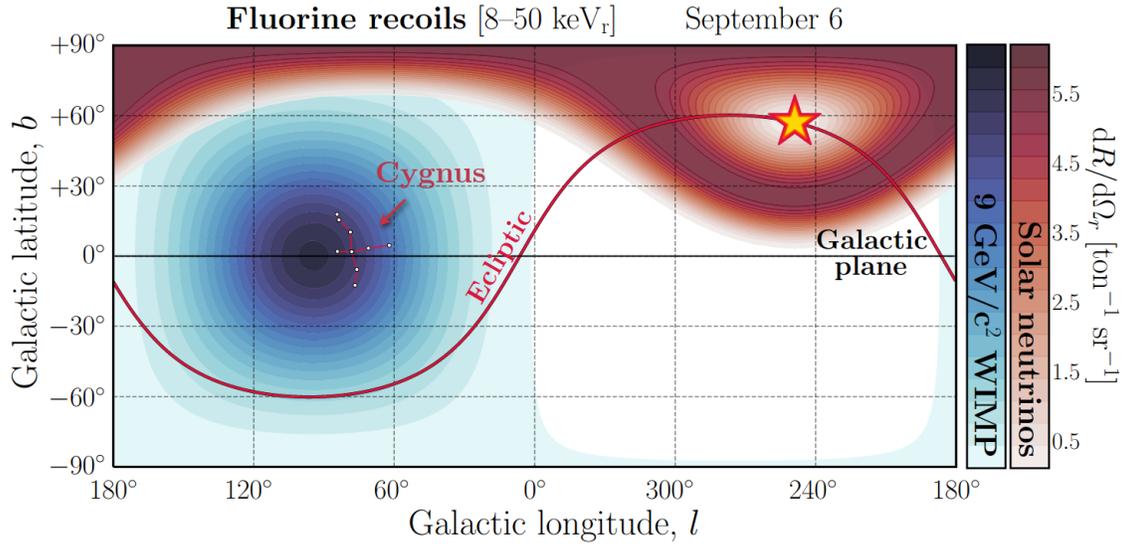
Figure 1.3: From Ref. [12]: Simulated distribution of nuclear recoils from 9 GeV DM particles and solar neutrino backgrounds on a given day in galactic coordinates. Our solar system moves toward $(l, b) = (90°, 0°)$, so the expected DM distribution incident on us from the Milky Way's DM Halo peaks in this direction. Observed solar neutrino backgrounds peak toward the direction of the sun of the sun while the position of the sun varies along the Ecliptic over the course of a year.

4

1 radian ($\sim 57°$) to be the limit of no 3D angular resolution and $0°$ to represent perfect angular resolution.

Head-tail recognition efficiency ($\varepsilon_{\mathbf{ht}}$): The fraction of events where the scalar (dot) product of the true recoil direction with the direction assigned to the observed event is positive. $\varepsilon_{\mathrm{ht}} = 0.5$ corresponds to no head-tail sensitivity and $\varepsilon_{\mathrm{ht}} = 1$ corresponds to perfect head-tail sensitivity.

Event-level time resolution: The ability to assign accurate timestamps to nuclear recoil events. Event-level timing is required to transform events into galactic coordinates to directly search for a galactic dipole signature. Ref. [12] suggests event-timing resolutions σ_t within 40 minutes are required to keep angular resolutions due to the spin of the Earth within $10°$. This performance requirement is easily met by modern gas TPCs where timing resolution is principally limited by the drift speed of ionization produced by nuclear recoils leading to $\mathcal{O}(10\,\mathrm{ms})$ timing resolution in the most pessimistic of cases [21].

Energy resolution: The fractional uncertainty of the energy of an observed event, σ_E/E. In general, energy resolution depends on primary ionization fluctuations as well as fluctuations in gain. Depending on the readout technology used, the dynamic range of the readout may also limit energy resolution. For our purposes, energy resolution is important for resolving $\mathrm{d}E/\mathrm{d}x$ well enough to discriminate between nuclear recoils and electron-recoil backgrounds.

Angular resolution, head-tail sensitivity, and energy resolution all deteriorate at low energy in gas TPCs, making it very challenging to achieve performance levels sufficient for resolving the keV-scale nuclear recoils expected from the scattering of $\mathcal{O}(\mathrm{GeV})$ DM particles. Furthermore, even in a deep underground detector environment, electron-recoil backgrounds from the Compton scattering of gamma rays is expected to overwhelmingly dominate over the expected rates of DM and CEνNS-induced nuclear recoils. These backgrounds get exponentially more difficult to reject with decreasing energy, so improving electron background rejection performance is necessary to elucidate any directionally anisotropic signal originating from DM or neutrinos, and improving angular resolution and head-tail recognition efficiency are important for distinguishing DM-recoils from solar neutrino-recoils.

These performance metrics can be readily measured and evaluated at the relevant keV-scale energies using neutron-induced nuclear recoils. Neutrons interact with gas nuclei in TPCs primarily via eleastic scattering. In particular, an incident neutron with energy E_n may elastically scatter off a gas nucleus with atomic mass A, causing the gas nucleus to recoil with energy E_r at an angle
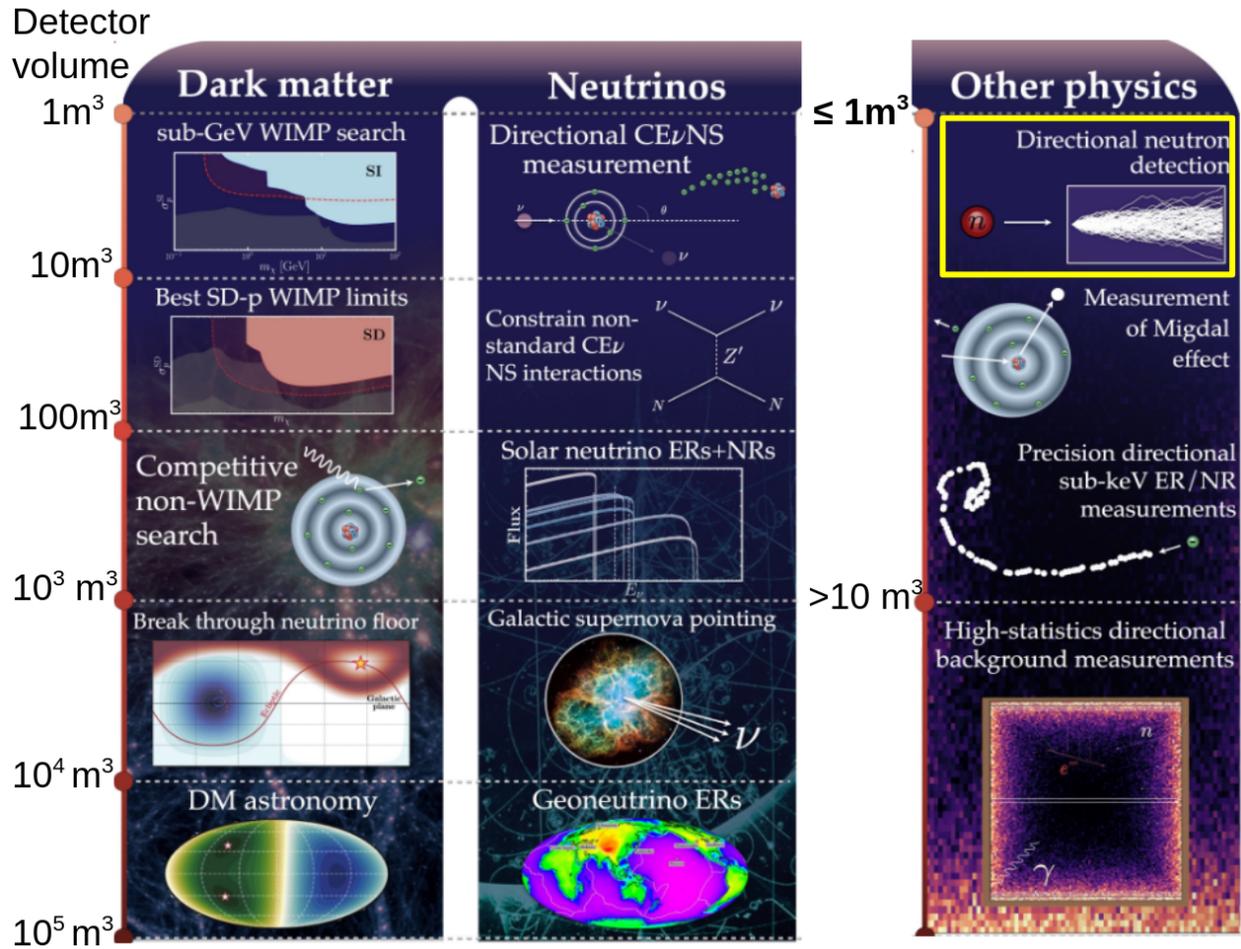
Figure 1.4: (Adapted from [12]) Selected physics applications possible with directional gas TPCs as a function of detector volume. We focus on the application of directional neutron detection (boxed in yellow) with $40\,\mathrm{cm}^3$ "micro" TPCs called the BEAST TPCs. The ingredients necessary for sensitive directional neutrino and DM detection can be demonstrated with directional neutron measurements in small gaseous TPCs, making these measurements interesting from a research and development standpoint.

of θ_r with respect to the direction of the incident neutron, where

$$E_r = \frac{4A}{(1+A)^2} E_n \cos(\theta_r). \tag{1.1}$$

The neutron-gas nucleus interaction probability is large enough (Figure 3.3) that high statistics samples of neutron-induced nuclear recoils can be recorded in a laboratory setting with sub-m^3 detectors (Figure 1.4).

1.3 Overview of dissertation content

As a whole, we split this work into two parts: In Part I, we demonstrate directional recoil detection for fast-neutron backgrounds at the SuperKEKB [22] e^+e^- collider, using $40\,\text{cm}^3$ gas-TPCs with 3D pixel readout known as the BEAST TPCs [23]. In Part II, we introduce the use of deep-learning-based computer vision techniques called 3D convolutional neural networks (3DCNNs) to improve both electron background rejection and head-tail recognition efficiency in the BEAST TPCs. While angular resolution is also important to improve DM sensitivity, Figure 5.1 demonstrates that head-tail recognition efficiency is the most important quantity to distinguish a DM-recoil event from a solar neutrino background, which is why we focus on improving head-tail here. We encourage readers principally interested in directional gas TPC research and development efforts to skim Chapter 2.4 for an overview of the BEAST TPCs and then skip directly to Part II (Chapter 5).

To break things down further, in Chapter 2, we set the stage for Part I by introducing SuperKEKB, Belle II, the BEAST TPCs, and beam-induced nuclear recoil measurements with these TPCs. We then present two detailed neutron background campaigns and the TPC systems associated with them: one using TPCs surrounding SuperKEKB's final focusing magnets (Chapter 3) and the other using TPCs in the accelerator tunnel surrounding the Belle II detector (Chapter 4). The measurements and analyses in Chapter 3 expand on results published in Ref. [24] and provided useful feedback of neutron backgrounds generated near the collision point of the beams. The results of this chapter highlight the BEAST TPCs' abilities of background discrimination and their ability to measure energy spectra.

The analyses of neutron backgrounds in the tunnel regions outside of Belle II that are presented in Chapter 4 are published in Ref. [25] and go into more depth and quantify background discrimination, evaluate angular resolution and head-tail recognition efficiencies of nuclear recoils, and present a full analysis of the angular distributions of nuclear recoils recorded in the TPCs. These analyses measure the key performance metrics of directional detection.

In Chapter 5, we utilize 3DCNNs to improve performance of particle identification and head-tail recognition. This chapter is devoted entirely to head-tail performance, and benchmarks four separate techniques for assigning head-tail, including a 3DCNN. Since traditional methods for head-tail identification work well for higher energy neutron-induced nuclear recoils, this is the regime

we focus on in this chapter so that we can benchmark performance in a familiar, well validated regime. We find that the 3DCNN considerably outperforms the other three methods of head-tail assignment in this simulation study.

Chapter 6 bridges the gap from simulation to measurement and details low gain nuclear recoil head-tail experiments performed in the Vahsen Group Lab at UH. Past and present directional DM detection experiments such as DMTPC [26], DRIFT [27], and most recently, NEWAGE [28], have demonstrated head-tail sensitivity for neutron-induced nuclear recoil tracks ranging from $\mathcal{O}(50\text{-}200\,\text{keV})$. Using 3DCNNs, we, for the first time, demonstrate event-level head-tail sensitivity for sub-50-keV$_\text{ee}$ measured nuclear recoils. These recoils were recorded in a BEAST TPC filled with an atmospheric-pressure mixture of He:CO$_2$ gas and operating at a gain of roughly 900. Typically higher gain and a lower density gas mixture are desirable for low-energy head-tail sensitivity, making these results all the more promising. Still, these results are a far-cry from the sub-10-keV$_\text{r}$ event-level head-tail sensitivity desired for $\mathcal{O}(\text{GeV})$ DM mass searches. Higher gains, where we are closer to being able to read out charge from single-electrons, should get us closer to the goal of demonstrating head-tail sensitivity in sub-10-keV$_\text{r}$ recoils.

In Chapter 7, we increase the gain until the charge from single electrons is detected, so that we can evaluate vector direction at the lowest possible energies. For sub-10-keV nuclear recoils, we find that particle identification becomes challenging so we introduce a 3DCNN to classify recoil species. Training and evaluating the 3DCNN on simulated nuclear recoils, we find excellent head-tail discrimination performance below $10\,\text{keV}_\text{r}$. All together, our results on simulation are the first demonstration of directionality at energies relevant for $\mathcal{O}(\text{GeV})$ DM searches, all with desirable performance. In the latter portion of this chapter, we attempt to bridge the gap between simulation and measurement and evaluate the performance of our particle identification and head-tail assignment 3DCNNs on high gain measurement. While doing this, we run into the so-called sim2real gap, where our performance in simulation does not translate into meaningful performance on measurement, so we still have work to do before we can claim any significant head-tail performance in the $\mathcal{O}(10\,\text{keV}_\text{r})$ regime.

Chapter 8 details a separate electron background rejection simulation study of a detector configured better for directional DM applications. This work is currently under peer review in Ref. [29]. We find that using a 3DCNN outperforms other state-of-the-art machine learning techniques that require feature engineering. The 3DCNN is also much more robust against added noise than traditional event discriminants like track length. We note that we are currently investigating possible bias in our event processing pipeline utilized for the 3DCNN in this study. The electron rejection performance of the 3DCNN may have been overestimated so the results in this chapter are preliminary and the reader should consult the forthcoming journal publication for final results. Finally, Chapter 9 summarizes our overall findings.

Part I: Neutron background measurements at SuperKEKB

CHAPTER 2
SUPERKEKB, BELLE II, AND THE BEAST TPCS

Parts of this chapter are taken from Ref. [25] of which I am first author.

2.1 Overview of SuperKEKB and Belle II

The SuperKEKB accelerator [30] is the upgrade of the KEKB [31] high-luminosity asymmetric-energy e^+e^- circular collider located at the KEK laboratory in Tsukuba, Japan. SuperKEKB has been in operation since 2016 [32] and has been providing collisions for the Belle II B-factory experiment [33] since March 2018 [24]. As of December 2021, SuperKEKB has reached an instantaneous luminosity of $3.8 \times 10^{34}\,\mathrm{cm^{-2}s^{-1}}$, and seeks to ultimately achieve a luminosity of $6.3 \times 10^{35}\,\mathrm{cm^{-2}s^{-1}}$ [34], a factor of 30 greater than the peak luminosity of KEKB. Two storage rings with a 3 km circumference called the high energy ring (HER) and low energy ring (LER) store 7 GeV electrons and 4 GeV positrons, respectively. The electrons and positrons collide at a center of mass energy corresponding to the mass of the $\Upsilon(4S)$ resonance ($\sqrt{s} = 10.58\,\mathrm{GeV}$), which decays into a $B\overline{B}$ pair with a branching ratio of 96% [35]. With a target integrated luminosity of $50\,\mathrm{ab^{-1}}$ over the course of its operating lifetime, SuperKEKB is the world's first so-called Super-B-factory.

To reach such ambitious luminosities, SuperKEKB employs a "Nano-beam scheme" proposed by P. Raimondi [37], where the horizontal and vertical beam sizes at the interaction point (IP) of the e^- and e^+ beams are squeezed to $\sigma_x^* \approx 10\,\mathrm{\mu m}$ and $\sigma_y^* \approx 50\,\mathrm{nm}$, respectively. To achieve vertical betatron functions at the IP (β_y^*) small enough to support such narrow beams, a precise final focusing system that includes a 1.5 T superconducting solenoid was developed. Achieving such a high luminosity through the squeezing of beams, however, comes at the cost of elevated beam-induced backgrounds. Continuous measurements of beam-induced backgrounds during accelerator operation are necessary to inform our background-modeling so that we can anticipate the necessary countermeasures to enable higher luminosity operation while maintaining background levels within the tolerances of Belle II. The primary countermeasures for mitigating beam backgrounds at SuperKEKB are a system of 31 beam collimators that are located at various points along the HER and LER storage rings (Figure 2.2), and strategically placed shielding around the beam pipes and on Belle II detectors. Adjustment of these collimators has a strong effect on single-beam background levels.

2.2 Overview of neutron backgrounds

Here we're concerned specifically with neutron backgrounds and refer the reader to the Refs. [32; 24; 39] for more general surveys of beam-induced backgrounds. Neutron backgrounds are particularly
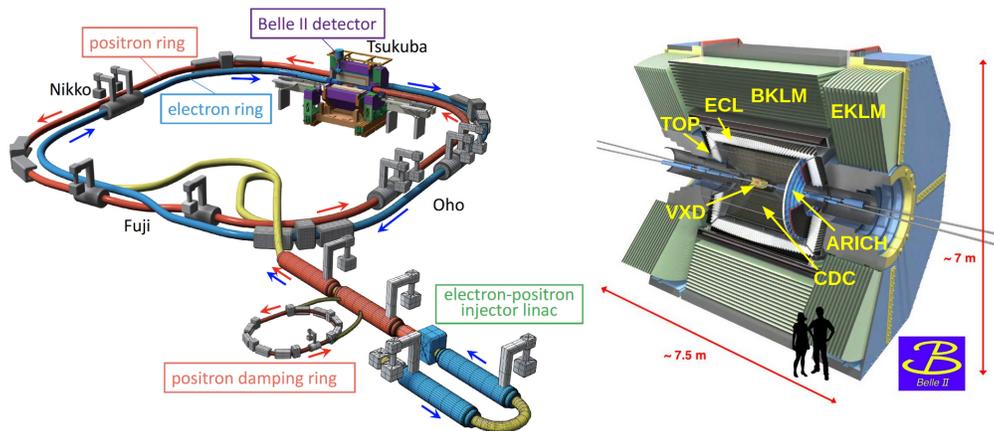
Figure 2.1: Adapted from [36]. Left: A schematic of the SuperKEKB collider. SuperKEKB consists of three main sections: (i) An injector LINAC, (ii) a positron damping ring, and (iii) the main ring which consists of the high energy e^- ring (HER) and low energy e^+ ring (LER). An RF gun generates a low emittance electron beam which passes through the LINAC directly into the HER. The positron beam is created from electron bunches colliding with a tungsten target. The positron beam is then routed into the damping ring to reduce its emittance before being injected into the LER. Right: The Belle II detector. The labeled subdetectors are the VerteX Detector (VXD), Central Drift Chamber (CDC), Electromagnetic CaLorimeter (ECL), Aerogel Ring-Imaging Cherenkov Detector (ARICH), Time Of Propagation (TOP) Cherenkov counters, and the K_L-μ end cap (EKLM) and barrel (BKLM) layers.
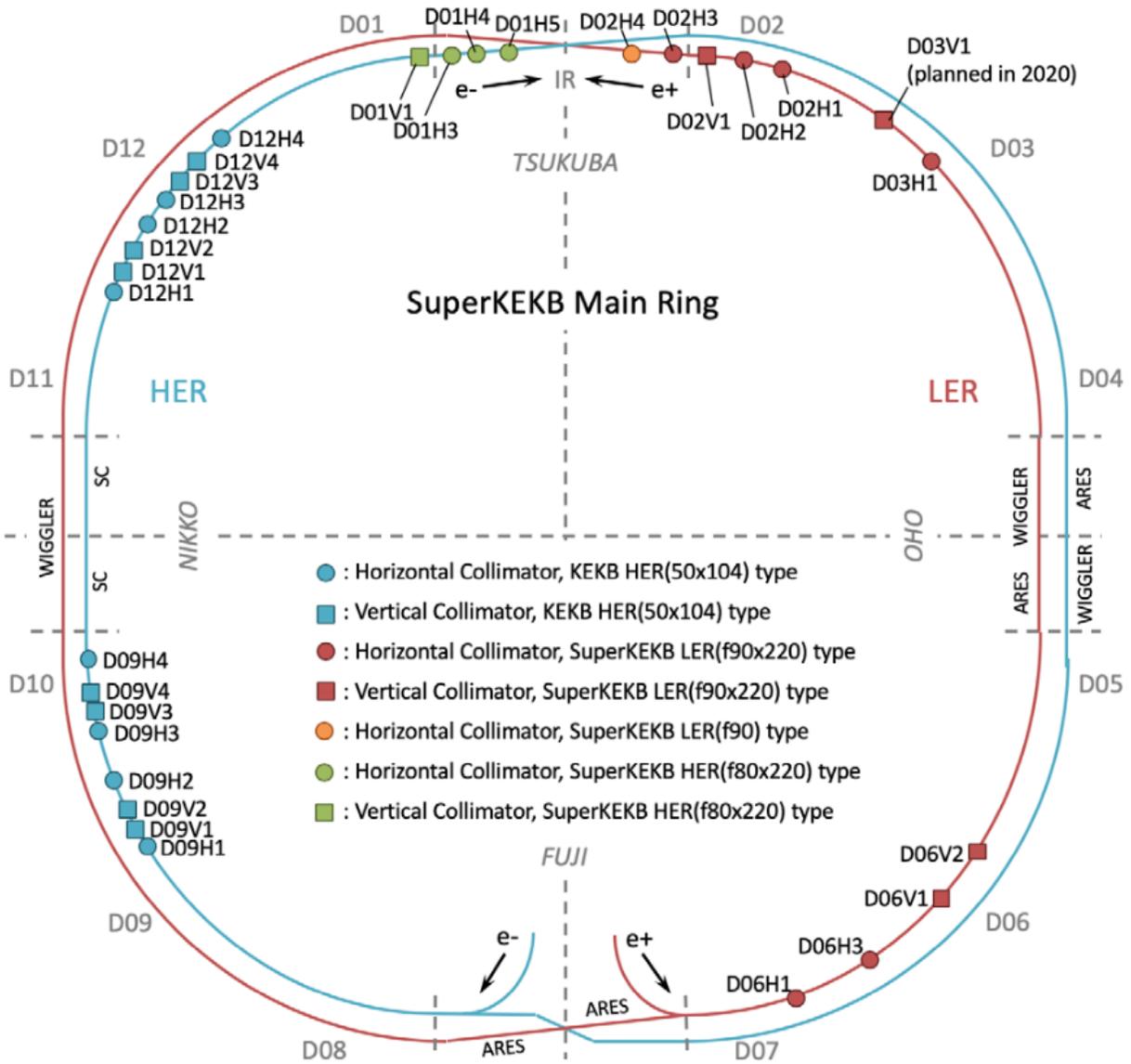
Figure 2.2: [38] The SuperKEKB collimation system is one of the primary means to mitigate beam backgrounds. The system contains 31 collimators at various points along the LER and HER. During the upgrade from KEKB to SuperKEKB, new collimators were installed both throughout the LER and along both rings near the interaction region (IR).

difficult to study and mitigate and were responsible for performance degradation in the KLM of Belle, the predecessor to Belle II [40; 41]. Since neutrons are electrically neutral, they lose energy primarily from elastic scattering off atomic nuclei, making them highly penetrating if not properly shielded by low-Z materials. To prevent a similar scenario in Belle II, polyethylene shielding was installed on the outer KLM end caps, and the RPC layers on both the outer KLM end caps and two innermost barrel layers were replaced with scintillator-based detectors tolerant of increased background hit rates [42]. Given the expectation of higher background rates at SuperKEKB over KEKB, and that these rate estimates have substantial uncertainties [34; 39], neutrons pose a risk not only to the KLM, but also to other Belle II detector systems where both fast and thermal neutrons can lead to single event upsets [43; 44]. It is thus important to understand neutron production at SuperKEKB in order to best assess neutron background remediation measures.

Neutron backgrounds originate from showers that result from off-orbit beam particles or photons interacting with the atoms in the walls of the beam pipe. Secondaries from these showers excite particular nuclei via the giant dipole resonance [45; 46], producing neutrons. Neutron production can arise both from circulating single beams and from collisions. Beam-gas scattering (Coulomb and Bremsstrahlung) and the Touschek effect are the principal single-beam-induced mechanisms leading to neutron-producing showers. When beams are colliding, radiative-Bhabha (RBB) scattering produces photons that travel along the straight section of beam pipes, ultimately colliding with the beam pipe walls in the region where the beam pipes start to curve. These RBB photon collisions are highly localized and produce copious amounts of neutrons leading to what we refer to as *radiative Bhabha hotspots*. Each mechanism leading to neutron production is difficult to simulate, making it important to directly measure neutron backgrounds and use these measurements to both improve our current understanding of neutron backgrounds, and also improve our ability to forecast neutron backgrounds in future scenarios.

With these goals in mind, we have deployed systems of BEAST TPCs throughout all three beam commissioning phases of SuperKEKB operation. These TPCs measure neutrons by producing 3D images of nuclear recoils from fast neutron scattering.

2.3 Overview of beam commissioning phases

Beam commissioning at SuperKEKB is performed in three distinct Phases, with each Phase consisting of both unique and overlapping goals for both the accelerator and Belle II detector groups.

Phase 1 Phase 1 was the first time beams were circulated in the main ring section of SuperKEKB. On the accelerator side, the goals of Phase 1 were to circulate beams in both the HER and LER, tune beam optics and collimators to improve beam lifetimes while reducing beam backgrounds, and vacuum scrub to bake out residual gas atoms present in the beam pipes and thereby reduce the gas pressure within the beam pipes. Vacuum scrubbing was especially important in the LER, since

Phase	Dates	L_{peak} [cm^{-2}s^{-1}]	# TPCs	TPC locations
1	Feb. 2016 - June 2016	No collisions	4	Surrounding near IP: $(r, z, \phi) = (45\,\text{cm}, 35\,\text{cm}, \{0°, 90°, 180°, 270°\})$
2	Feb. 2018 - July 2018	5.6×10^{33}	8	Dock ring surrounding the QCS: $(r, z, \phi) = (35\,\text{cm}, -1.3\,\text{m}, \{18°, 90°, 198°, 270°\})$ $(r, z, \phi) = (35\,\text{cm}, +1.9\,\text{m}, \{22°, 90°, 202°, 270°\})$
3	Mar. 2019 - 2030s	3.8×10^{34} (Current) 6.3×10^{35} (Target)	6	Inner concrete shield wall: $(r, z) \approx (2\,\text{m}, \{-14, -8.0, -5.6, +6.6, +14, +16\}\text{m})$

Table 2.1: Timeline of the three beam commissioning phases, the peak luminosities observed, and approximate locations of each of BEAST TPC during each Phase. Descriptions of the Phase 2 and Phase 3 TPC systems are found in Chapters 3 and 4, respectively.

the LER was completely replaced when KEKB was upgraded to SuperKEKB. Belle II was not yet "rolled in" nor were the final focusing Super Conducting Quadrupoles (QCS) installed. In the place of Belle II, a suite of dedicated beam background detectors known as BEAST II surrounded the interaction region (IR), with the principal goal of measuring beam backgrounds to assess whether backgrounds induced from circulating single beams were at a manageable level to move onto Phase 2 and install the QCS and Belle II.

Two of the four installed TPCs were ultimately used for measurement and analysis of nuclear recoil rates, energy spectra, and directional distributions during Phase 1. Angular resolution was found to be within $20°$ both for measured and simulated He recoils above $100\,\text{keV}_{\text{ee}}$, and a head-tail recognition efficiency of 78% was found using simulated He recoils ranging in ionization energies between $50\,\text{keV}_{\text{ee}}$ and $\mathcal{O}(1\,\text{MeV}_{\text{ee}})$. The Phase 1 BEAST TPC system predates any of the original work in this dissertation, so we refer the reader to Refs. [47] and [48] for further details about the system and results. Detailed descriptions of the Phase 1 systems and the results of Phase 1 commissioning from the SuperKEKB accelerator group are found in Ref. [22] and from the BEAST group in Ref. [32].

Phase 2 The Phase 2 system included the newly installed QCS with a partial installation of the Belle II detector surrounding the IR. This partial installation of Belle II included all of the outer layers of Belle II—the KLM, ECL, TOP, ARICH, and CDC—but only a small subset of the VXD which consists of highly sensitive silicon pixel and strip sensors. The remaining space within the partially filled VXD volume was populated with an updated suite of BEAST II background detectors to assess the safety of the planned installation of the full VXD in Phase 3. On the accelerator side, the main goals of Phase 2 were to produce first collisions, tune beam optics and collimators to increase luminosity as much as possible while keeping backgrounds safe for Belle II data collection, and to further vacuum scrub. On the detector side, the main goals of Phase 2 were to commission Belle II with beams, use the BEAST detectors to measure backgrounds induced from

hotspots in the region between the QCS final focusing magnets and the IP, and measure luminosity backgrounds produced by colliding beams. We will detail the Phase 2 TPC operation experience and an analysis of fast neutron backgrounds measured in the TPCs during this time in Chapter 3. Descriptions of the entire BEAST Phase 2 campaign and results are found in Ref. [24]. Detailed information about other BEAST Phase 2 detector systems can be found in Refs. [49; 50; 51; 52; 53].

Phase 3   Phase 3 began in 2019 and is the final stage of Beam Commissioning that will continue throughout the operating life cycle of Belle II. With Belle II fully installed, Phase 3 accelerator operation is now ultimately concerned with maximizing luminosity for collecting B-physics data, while maintaining suitable background levels to prevent detector performance degradation and enable long term detector operation. Background monitoring is now primarily performed by the Belle II subdetectors, with a smaller system of upgraded BEAST detectors performing dedicated background measurements surrounding the QCS and in the tunnel regions surrounding the Belle II detector. In Chapter 4, we detail analyses of neutron backgrounds measured with the Phase 3 TPC system. More general information regarding the current status of beam backgrounds and potential future SuperKEKB upgrades can be found in Refs. [34; 39; 54].

2.4   BEAST TPCs

The BEAST TPCs [23] are a second-generation [55] gas time projection chamber developed by the Vahsen lab at the University of Hawai'i. Two prototype models and eight identical production models were assembled and tested in-house and shipped to KEK. The Phase 1 TPC system utilized the two prototype models as well as two of the eight production models. The prototype models were shipped back to Hawai'i after the conclusion of Phase 1 and were not used in any of the original work in this dissertation. All eight of the production BEAST TPCs were brought to KEK for Phase 2 and six of those eight are currently operating as the Phase 3 TPC system at KEK. The remaining two production TPCs were shipped back to Hawai'i at the conclusion of Phase 2 and one of these TPCs is currently being used for the particle ID and directional head-tail studies detailed in Chapters 6 and 7.

Each TPC is a $10 \times 15 \times 31 \, \mathrm{cm}^3$ vessel with a $2.00 \times 1.68 \times 10.9 \, \mathrm{cm}^3$ fiducial volume. The vessels are filled with a 70:30 mixture of He:CO_2 gas, which serves as the target gas with which neutrons interact. We use the annotated diagram in the left-hand portion of Figure 2.3 to outline how nuclear recoil events are detected in a BEAST TPC:

(1) When a fast neutron enters a TPC, it may scatter off a ^{4}He, ^{12}C, or ^{16}O nucleus, causing the nucleus to recoil. The probability of a fast neutron scattering event to occur within the fiducial volume of a BEAST TPC is in general $\mathcal{O}(10^{-4} - 10^{-3})$, as shown in Figure 3.3.
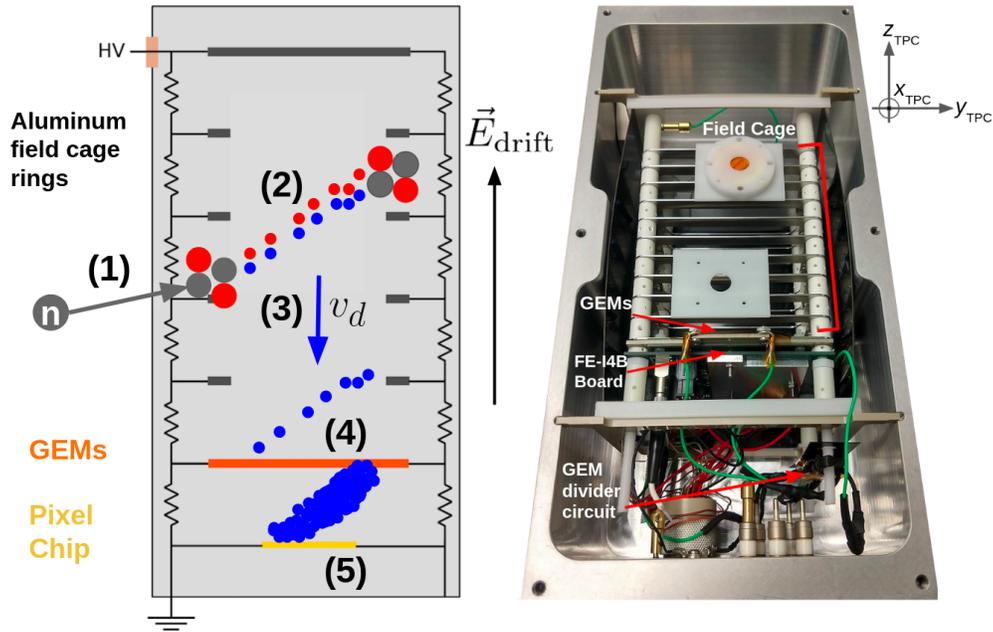
Figure 2.3: Left: A schematic illustrating the steps of detecting a nuclear recoil. These steps are described in the text. Right: The inside of a BEAST TPC. The key components in the detector are the field cage, the double gas electron multiplier (GEM) layer, and the ATLAS FE-I4B pixel chip. An internal GEM divider circuit was developed and installed in each TPC to reduce the TPC footprint in order to satisfy the spatial constraints of the Phase 2 TPC system. The white disk mounted on top of the field cage contains a ^{210}Po alpha emitting source that is used to calibrate the energy scale of the TPCs. We only use ^{210}Po sources to calibrate energy for the Phase 2 and Phase 3 neutron background analyses (Chapters 3 and 4). The coordinate axes shown define the TPC coordinate system. The origin is placed on the edge of the readout plane of the pixel chip, with positive z_{TPC} pointing upward toward the field cage.

16
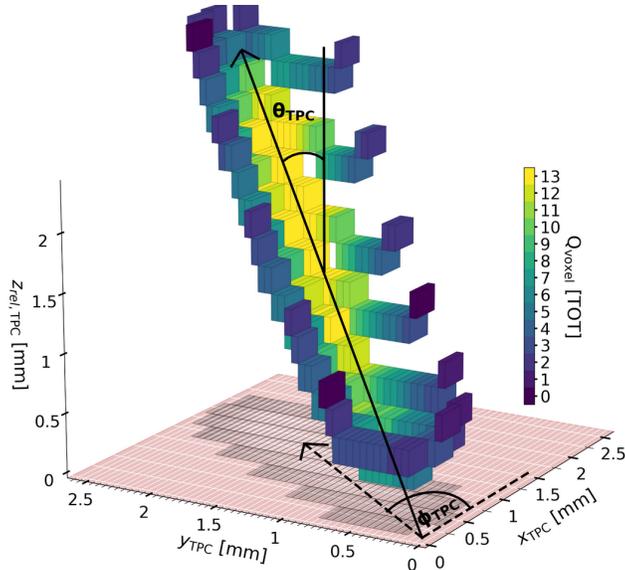
Figure 2.4: Visualization of a $210\,\mathrm{keV_{ee}}$ nuclear recoil track detected by a BEAST TPC. The color scale shows the detected charge in each voxel in units of time over threshold (TOT). θ_{TPC} is the zenith angle defined between z_{TPC} and the vector direction of the principal axis of the track, and ϕ_{TPC} is the azimuthal angle defined between x_{TPC} and the projection of the track's principal axis onto the readout pixel chip. The white grid marks illustrate the $250\,\mathrm{\mu m} \times 50\,\mathrm{\mu m}$ pixel dimensions.

(2) The recoiling nucleus will then collide with other gas atoms in the vessel until it eventually stops. Along the way, the recoiling nucleus strips electrons off gas atoms within the sensitive volume, forming an ionization trail. We call this ionization trail the **primary track**. We note here that the majority of nuclear recoil events recorded in the BEAST TPCs fall "beyond the Bragg Peak" (Figure 5.3), meaning the stopping power of a recoiling nucleus falls sharply along the length of the primary track. This leads to an asymmetry in the charge density along the length of a track's principal axis, with less ionization charge expected near the *head* direction of the track and more charge near the *tail* end of the track. This asymmetry can be exploited to assign a head-tail direction to events.

(3) A uniform electric field along z_{TPC} is provided by an aluminum field cage, causing the ionized nuclei to drift upward toward the cathode, while the electrons in the ionization trail drift against the field toward a double gas electron multiplier (GEM) layer [56]. Repeated collisions with gas atoms lead to a random-walk motion of the electrons drifting against $\vec{E}_{\mathrm{drift}}$ leading to, on average, a constant drift speed, v_d, but with statistical fluctuations in the transverse and longitudinal direction that is commonly referred to as diffusion.

(4) The charge is avalanche-multiplied as it passes through the double GEM layer. We mostly use modest double GEM gains of $\mathcal{O}(1,000)$ (Table 3.1 and Table 4.3 for Phases 2 and 3, respectively)

17

to avoid saturation due to the limited dynamic range of the pixel readout when detecting highly ionizing nuclear recoils. This choice also delays the onset of gas detector aging and lowers the risk of accidental sparking which might damage the sensitive pixel electronics. In Chapter 7, however, we increase the double GEM gain up to around 13,400—where the BEAST TPCs are sensitive to charge from single electrons—to directionally track low energy nuclear recoils

(5) The avalanche-multiplied charge is then detected by an ATLAS FE-I4B pixel chip with a custom metallization pattern [23; 57; 58; 59]. The ATLAS FE-I4B has a (2.00 cm × 1.68 cm) readout area consisting of an (80 × 336) grid of (250 µm × 50 µm) pixels. The constant drift speed (on average) of the ionization charge through the field cage volume allows for the construction of a relative z coordinate, providing a 3D reconstruction of the ionization distribution created by the recoiling gas nucleus as can be seen in Figure 2.4. Read-out charge is integrated on a 40 MHz clock leading to only a single hit per pixel, as opposed to a full 3D event shape (details in Chapter 3.3).

CHAPTER 3
FAST NEUTRON BACKGROUNDS NEAR THE SUPERKEKB FINAL FOCUSING MAGNETS

This chapter includes and expands upon the TPC results published in Ref. [24] of which I was a contributing author. It also includes some material from Ref. [25] of which I was first author.

In this chapter we detail an analysis of neutron backgrounds recorded by eight BEAST TPCs surrounding the QCS final focusing magnets where, prior to measuring them, beam-induced neutron backgrounds were expected to be high.

3.1 TPC system during Phase 2 of beam commissioning
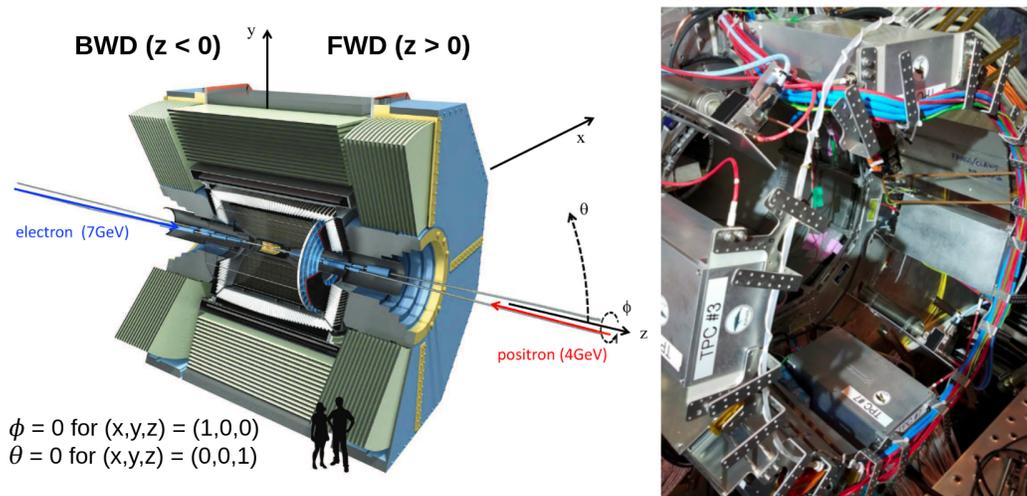
Figure 3.1: Left: Overview of the Belle II coordinate system. Near the interaction point of the beams (IP), $+z$ is defined to roughly coincide with the boost direction of colliding beams. Throughout this Chapter and Chapter 4, we will often use forward (FWD) and backward (BWD) as designations for $z > 0$ and $z < 0$ locations, respectively. Right: Four BEAST TPCs located on the BWD ($-z$) side of the IP before the QCS was rolled in. Moving clockwise from the TPC labeled "TPC #3", the TPCs are located at ϕ=18°, 90°, 198°, 270°, respectively. The locations of the FWD TPC system, essentially mirror the BWD system (Table 2.1).

The Phase 2 directional fast neutron detection system consists of eight BEAST TPCs, and their high voltage (HV), low voltage (LV), gas, and data acquisition (DAQ) systems. The TPCs during Phase 2 were located in the VXD dock ring surrounding the final focusing QCS on either side of the IR, with locations described in Table 2.1 and shown in Figure 3.1. In order to accommodate

the strict spatial requirements of the VXD dock ring, we designed and installed identical internal GEM voltage divider circuits (Figure 2.3) in each TPC [23] that reduced the TPC footprint and are still in use. Figure 3.1 also shows the Belle II coordinate system which will be used throughout this Chapter and Chapter 4. The Phase 2 TPC locations were critical for fast neutron monitoring, as Phase 2 marked the installation of the QCS system, so the measurements recorded by the BEAST TPCs would mark the first measurements of fast neutron backgrounds resulting from the both final focusing of the beams, and the decreasing in beam pipe radius between the QCS and the IP. Given that fast neutrons are highly penetrating, reasonable agreement between measurement and simulation is critical, because unexpectedly high neutron doses on Belle II electronics could lead to single event upsets (SEUs) or otherwise degrade detector performance.

3.1.1 Gas System

Maintaining adequate gas purity is essential for achieving relatively stable effective gains in each TPC. During Phase 1 operation [47], we flowed gas in parallel to four detectors. We auto-regulated the pressure in each detector separately, controlled only the total flow to all detectors, and mechanically adjusted flow impedance with valves to balance parallel flows. This approach was highly unstable, with large flow oscillations and even occasional reverse flow in some of the parallel branches. Despite this, gas purity and gain stability were better than expected: although there was a brief initial period of outgassing where gain would slowly rise, after several months of detector operation with gas flow, the avalanche gain would remain stable for weeks even at minimal or at no flow.

The Phase 2 and Phase 3 gas systems were re-designed based on this experience. Two parallel flow branches now have separately controlled flows to avoid oscillations, with the serial flow rate set between 15 and 20 sccm through the three TPCs present in each branch. A 70:30 mixture of He:CO_2 with a minimum purity of 99.999% has been used throughout all three Phases of beam commissioning. We now walk through the gas system and refer to Figure 3.2, which shows a schematic representation of all components present in this gas system, when describing the course of gas flow through the system.

Starting from the gas bottle regulator, 6.4-mm outer-diameter copper tubing with 0.8-mm wall-thickness is routed from the regulator of the gas bottle to a custom built rack-mounted Gas Box which serves as the central hub for the TPC gas system. Inside the Gas Box, a T-connector (not visible in the bottom image of Figure 3.2) splits the gas flow into two parallel paths, with the flow in each path controlled by a Brooks mass flow controller. Bellows valves are located directly upstream of each of these mass flow controllers allowing for flow to be manually turned on and off in each branch.

Immediately downstream of the mass flow controllers, the tubing exits the Gas Box and one of the two branches is routed to the forward (FWD, see Figure 3.1) TPCs and the other to the backward (BWD) TPCs. During Phase 2, about 15 m (20 m) of tubing was required to reach the
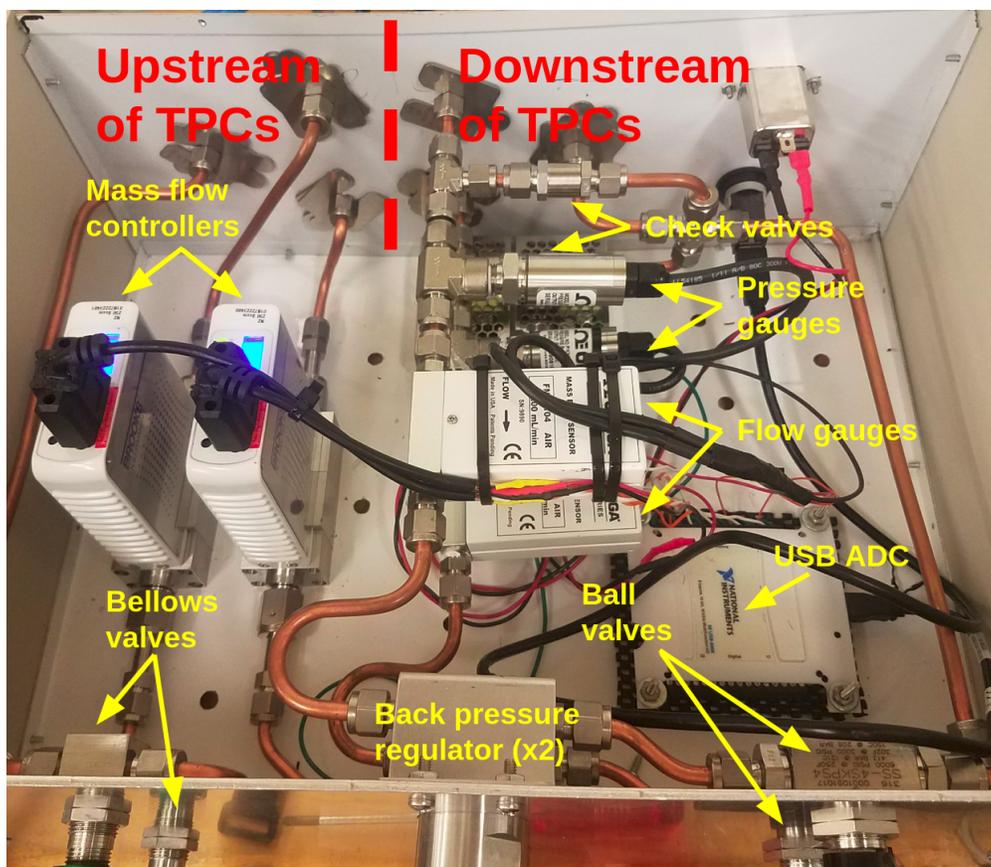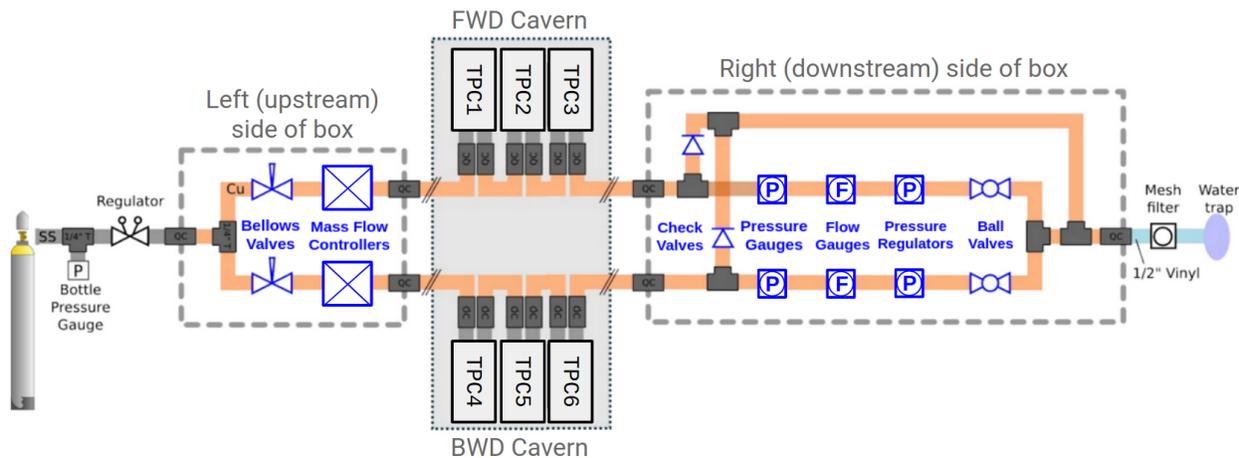
Figure 3.2: Top: Schematic of the Phase 3 gas system [not to scale]. The components shown within the bold gray dashed lines are contained within the custom-designed Gas Box. We used the same Gas Box in both Phases 2 and 3. Bottom: Inside of the Gas Box.

FWD (BWD) TPCs from the Gas Box, and in Phase 3, these numbers approximately double to reach the FWD and BWD tunnels, respectively. Two Swagelok Quick Connect (QC) stems are attached to the face-plate of each TPC vessel, allowing for easy connection of gas tubing to the vessels. The gas tubes between each of the four TPCs (three TPCs in Phase 3) in a given branch are connected in series. After exiting the furthest downstream TPC in each branch, the copper tubing is routed back to the downstream side of the Gas Box along the same path used to enter the FWD (BWD) tunnel.

Capacitance manometers (labeled "Pressure gauges" in Figure 3.2) and flow gauges are installed in each of the two branches on the downstream side of the box for the dual purpose of monitoring downstream pressure and flow and for identifying if leaks are present between the upstream and downstream sides of the box. Finally, back-pressure regulators are installed in each branch to ensure that the gauge pressure never exceeds 5 Psi, and ball valves are present at the downstream terminus. After exiting the Gas Box, copper tubing is routed an additional ~20 m away from the Gas Box (10 m in Phase 3 due to a different location of the Gas Box), where it is then connected to vinyl tubing that leads to a water trap and finally connects to an exhaust pipe where the gas is vented out of the building. Two check valves are also located in the downstream side of the Gas Box to prevent accidental over-pressuring inside the TPCs.

The Gas Box also employs a National Instruments USB-6001 Multifunction I/O device (labeled "'USB ADC'") in Figure 3.2) that digitizes the analog signals from the flow controllers, flow gauges, and pressure gauges for logging on the gas control computer. This device allows for remote control of the gas system via the mass flow controller, as well as remote monitoring of the pressure in each branch via the capacitance manometers; upstream flow in each branch via the mass flow controllers; and downstream flow in each branch via the flow gauges. Pressure and flow readout values are stored and displayed both locally on the TPC gas system PC and as Process Variables (PVs) via the Experimental Physics and Industrial Control System (EPICS) interface [60].

3.1.2 HV system

Each TPC contains two high voltage (HV) inputs: one SHV input used to bias the gas electron multipliers (GEMs) via an internal voltage divider circuit [23], and one UHV input to bias the field cage. These two HV inputs allow for independent control of the GEM gain and electric drift field. Two CAEN R1470ETD 8-kV rated 8-channel power supplies are used to deliver HV power to the TPCs. During Phase 2, both CAEN supplies were mounted on the same electronics rack directly on top of the Belle II detector on the BWD side of the IP. Two custom HV cables are constructed for each TPC: one cable, terminated with identical SHV plugs on each end, that provides 2.1 kV to the GEM voltage divider circuit, and another cable, terminated with an SHV plug on one end and a 10-kV rated UHV plug on the other end, to provide up to 8.0 kV to the field cage. 12-m length cables were made for the BWD TPCs and 15-m length cables were constructed for the FWD

TPCs. Spare 12 m and 15 m cables of both the HV cables biasing the GEMs and those biasing the field cage were constructed for contingency. Remote control of each HV channel is provided using CAEN's proprietary GECO2020 software.

3.1.3   LV and DAQ system

To communicate between the TPCs and the DAQ system (henceforth referred to as the DAQ box), we construct a custom LV cable for each TPC. Each LV cable includes two Cat-7 ethernet cables strung side by side and terminated on one end with 14 of the 16 inner wires attached to the socket of a 19 prong circular Mil-DTL connector, and the other end terminated with two RJ-45 ethernet plugs; one containing all 8 of its internal wires, and the other containing only 6 of its internal wires. The FE-I4B chips present inside each TPC are mounted on custom boards which route power and low voltage differential signal (LVDS) communications between the chip and the DAQ box through 14 distinct pins that are connected inside the vessel to a 19 prong circular Mil-DTL plug that is machined into the face-plate of the vessel. The Mil-DTL end of the cable assembly is thus connected directly to the face-plate of the TPC. The RJ-45 ends of the cable are then routed between the DAQ box and the TPC. The DAQ box contains RJ-45 sockets that are connected to LV power-supply channels for providing power to the FE-I4B chips; the RJ-45 end with 6 of its 8 wires is connected into one of these inputs. The remaining RJ-45 plug of the LV cable is connected to the ethernet input of a custom Multi-chip Module Card (MMC3) with an attached Xilinx Kintex-7 FPGA. These boards are both powered by a 5 VDC supply and serve as the conduit for LVDS communication between a PC and the TPC.

3.2   Simulation of beam-induced backgrounds and nuclear recoils

Throughout Phase 2, we utilized a "fast simulation" to model the rates and energy spectra of recoiling nuclei in each TPC. Here we describe the production steps of simulating neutron backgrounds, as well as their interactions in the sensitive volume of the TPCs:

1. **Generation of beam background events**: During Phase 2, we only simulate single-beam background sources, which include Bremsstrahlung, Coulomb, and Touschek backgrounds. Particle scattering and loss positions are simulated using the Strategic Accelerator Design (SAD) framework [61]. The kinematic information of these initial background particles is saved and passed into Geant4 [62; 63; 64].

2. **Simulation of background showers and neutron propagation**: The Belle II Analysis Software Framework (basf2) [65; 66] has geometry and material implementations of the entire interaction region of the SuperKEKB-Belle II system within $|z| < 4$ m from the IP, which includes the CDC dock space where the Phase 2 TPCs are located. Once initial background

particles are loaded, Geant4 simulates the physics of the interaction between these particles and the materials present within the geometry.

3. **Fast simulation of nuclear recoils in the TPCs**: The probability that an incident neutron interacts with a gas nucleus in the gas volume of a TPC depends both on the energy of incident neutron and the path length the neutron follows through the volume of the TPC as shown in Figure 3.3. Given the relatively low interaction probability of incident neutrons with target gas nuclei, obtaining sufficient statistics for the analysis of fully simulated nuclear recoils from these scattering events is computationally expensive. To save on computational resources, we instead opt to implement a fast simulation, where for each neutron passing through the sensitive volume a TPC we:

    (a) Assign weights corresponding to the interaction probability of each recoil constituent associated with the kinetic energy of the neutron incident on the sensitive volume of the detector. We assign separate weights for He, C, and O recoils.

    (b) Draw the nuclear recoil energy E_r associated with the neutron-nucleus scattering event weight from a uniform distribution between 0 and the maximum recoil energy predicted by neutron-nucleus elastic scattering. In particular we set
    $E_r = \text{RandUnif}\left[0, \frac{4A}{1+A^2}E_n\right]$, where E_n is the kinetic energy of the incident neutron, A is the atomic mass of the recoiling gas nucleus, and the notation $\text{RandUnif}[a, b]$ is to be understood as a random number selected from the uniform distribution between a and b inclusively.

Putting steps (a) and (b) together, for each neutron passing through the sensitive volume of the detector, our fast simulation provides a weight and a corresponding recoil energy for an He, a C, and an O recoil. The sum of the weights described in step (a) over the elapsed beam-time of a simulation campaign gives the simulated nuclear recoil rate, and step (b) gives recoil energy distributions.

Several improvements in the implementation of accelerator and detector components of SuperKEKB and Belle II in Geant4 were made. The details of these improvements, as well as the effect they had on reducing the discrepancies between measured and predicted rates in many of the BEAST detector systems, are outlined in Ref. [24].

3.3   Calibration and event selection

In this section we describe the steps taken to calibrate charge, gain, and ultimately determine particle identification (PID) criteria in each TPC. More detailed information about charge readout and general calibration procedures for the FE-I4B readout chips in these TPCs can be found in
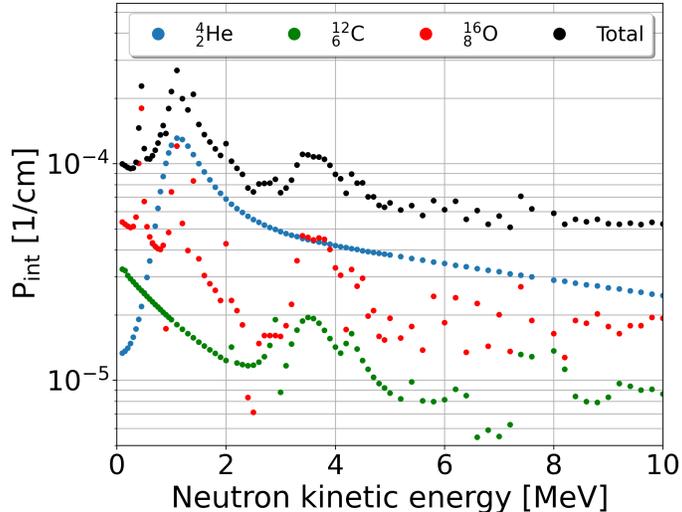
Figure 3.3: Neutron interaction probability per centimeter (P_{int}) of gas traversed versus neutron kinetic energy in a 70:30 mixture of He:CO_2 at 1 atm pressure. Elastic scattering cross sections to produce this figure were obtained from the `ENDF/B-VII` database [67] and may differ slightly from those listed in the `G4NDL4.6` library.

Ref. [23], and more general information about ATLAS FE-I4B performance and calibration can be found in Ref. [53].

3.3.1 Data processing

We utilize firmware packaged with the `pyBAR` readout software [68] that enables asynchronous triggering of each of the eight pixel chips in the Phase 2 TPC system. We collect data using the FE-I4 stop mode scan which is initialized by a HitOR signal that is processed by a Xilinx Kintex-7 FPGA connected to an MMC3 baseboard [23]. This stop mode setting allows for up to a 6.4 μs readout time per event. Throughout all of Phase 2 and Phase 3 TPC operation, this readout time window was reduced to 2.5 μs to allow for higher trigger rates. Timestamps are assigned to events with 0.05 s of precision which is sufficient for all beam background analyses. In Appendix A we detail how timestamps are assigned to events with our asynchronous triggering, as we initially mistakenly discarded a significant number of events in data processing and analysis campaigns between 2018 and early 2021. All data in this dissertation has been reprocessed so no events are discarded in this way.

3.3.2 Pixel-level calibration

Each pixel cell in a given FE-I4B contains a two-stage charge sensitive integrating amplifier, followed by a comparator with a set threshold voltage that samples on a 40 MHz clock. The comparator

outputs a logical high whenever the signal from the integrating amplifier is above the threshold voltage. The time over threshold (TOT) is the amount of time the comparator signal outputs logical high and thus depends on the amount of charge deposited on the pixel cell.

The first charge above threshold in an event triggers the readout. When an event is triggered, pixel hit data such as the column and row of the (80×336) pixel matrix, the TOT, and the readout time (recorded in multiples of $25\,\text{ns}$ due to the $40\,\text{MHz}$ comparator clock) are recorded. We use the comparator clock refresh rate to construct a relative z coordinate

$$z_{rel,\text{TPC}} = v_d n \times 25\,\text{ns}, \tag{3.1}$$

where $n \times 25\,\text{ns}$ is the readout time within the $2.5\,\mu\text{s}$ event readout window and v_d is the average drift speed of electrons for the gas and electric drift field strength used.

The charge read out in an FE-I4 pixel can be calibrated by first tuning the comparator threshold voltage and the TOT. A charge injection circuit located in each pixel injects charge in discrete voltage steps through two injection capacitors; these injection capacitors make the injected charge proportional to a variable injection voltage, thereby creating charge pulses of different, known, magnitude. The pyBAR readout software is packaged with a global tuning script that iteratively tunes both the threshold and TOT response of the integrating amplifier for all 26,880 pixels in the chip. The threshold is tuned with a target charge of $2700\,\text{e}$ in all TPCs and the TOT response is tuned to correspond to a saturation limit—the maximum TOT recorded in an event—above $45\,000\,\text{e}$. The results of the threshold and TOT scale tunings vary between FE-I4 chips (Table 3.1). The measured TOT in given pixel hit is represented by a 4 bit integer code ranging from 0 to 13, with $\text{TOT} = 0$ corresponding to a pixel near threshold and $\text{TOT} = 13$ corresponding to a saturated pixel.

After tuning the threshold and the TOT response in each TPC, we calibrate the remainder of the charge scale by sending two hundred injections at each of several distinct charge steps into each FE-I4B pixel and measure the TOT response in each pixel. The mean of the TOT over all pixels is recorded for each injection and the mean and standard error of each of these 200 pixel-averaged TOT values is plotted at each charge step. The TOT-to-charge mapping is not linear, so we use a bicubic spline interpolation of the injection charge versus mean TOT to determine the charge corresponding to integer TOT codes ranging between 0 and 13 in steps of 1. For reference, Figures 7.3 and 7.4 illustrate the process of mapping TOT to charge. These figures use different calibration settings than what is described here, but the general procedure still holds.

TPC location	$Q_{\mathrm{TOT}=0}$ [e$^-$]	$Q_{\mathrm{TOT}=13}$ [e$^-$]	G_{eff} [e$^-$]	E_{drift} $\left[\frac{\mathrm{V}}{\mathrm{cm}}\right]$	v_d $\left[\frac{\mu\mathrm{m}}{25\,\mathrm{ns}}\right]$
BWD 18°	2791 ± 33	47398	634 ± 75	358	173
BWD 90°	2819 ± 32	47492	542 ± 40	358	173
BWD 198°	2836 ± 42	47918	997 ± 58	358	173
BWD 270°	2810 ± 33	47009	729 ± 46	358	173
FWD 22°	2708 ± 31	47002	874 ± 82	358	173
FWD 90°	2713 ± 34	48416	668 ± 150	358	173
FWD 202°	2732 ± 31	46871	515 ± 82	267	130
FWD 270°	2727 ± 31	46001	978 ± 67	358	173

Table 3.1: Phase 2 TPC calibration results and settings. Columns from left to right describe: (i) TPC location in Belle II coordinate system, (ii) measured threshold charge, (iii) measured saturation charge, (iv) calibrated effective double GEM gains, (v) approximate drift field, and (vi) approximate drift speed. Threshold uncertainties (ii) are the standard deviation of the pixel threshold over every pixel in the chip. Gain uncertainties (iv) are given by the standard deviation of the ionization energy distributions of horizontal alphas in a given TPC. E_{drift} assumes 10.9 cm drift length and v_d is computed using Magboltz [69] for all TPCs except for FWD 202°, which experienced periodic high voltage trips, so its drift field was lowered to ensure stable operation. The drift speed for FWD 202° is estimated by assuming v_d scales proportionately as E_{drift}.

3.3.3 Gain calibration

We calibrate the effective gain of each TPC using alpha particles emitted from the ^{210}Po disk source installed in each detector (Figure 2.3). The ionization energy from an event in a TPC is related to the observed avalanche charge, Q, and effective gain of the TPC, G, via

$$E_{\mathrm{ionization}} = \frac{QW}{G}, \tag{3.2}$$

where $W = 34.45$ eV is the average energy per electron-ion pair of 70:30 He:CO$_2$. The location of the disk source is identical in each TPC and is chosen so that alpha tracks emitted from the ^{210}Po source span the entire 2 cm extent of x_{TPC}. This means the ionization energy deposited on the chip by an alpha with $(\phi_{\mathrm{TPC}}, \theta_{\mathrm{TPC}}) \sim (0°, 90°)$ will be constant within statistical fluctuations and can thus reliably be used as a reference value to determine G.

To ensure a pure sample of alphas, we calibrate using only alpha tracks that were recorded when beams were not circulating the main ring. The ^{210}Po calibration sources present in each TPC emit 5.4 MeV alphas, which create long enough ionization distributions to span the entire width of the chip. To select for alpha tracks, we only keep events that span the entire x length of the chip (events that hit both column $= 0$ and column $= 79$ on the pixel grid) and do *not* hit either of the y edges of the chip (neither row $= 0$ nor row $= 335$). Using these selections, in our TPC coordinate frame

(Figure 2.4), we expect the distribution of alpha events to peak at $\phi_{\mathrm{TPC}} = 0°$ and $\theta_{\mathrm{TPC}} = 90°$. Simulation shows the expected ionization energy for the portion of an alpha track that spans the x length of the FE-I4 at the ϕ_{TPC} and θ_{TPC} peak of $(\phi_{\mathrm{TPC}}, \theta_{\mathrm{TPC}}) = (0°, 90°)$ in 70:30 He:CO_2 is 1430 keV. We aim to select alphas near this peak, however during this calibration period where beams weren't circulating, the 1.5 T magnetic field from the QCS was still in operation, causing the alpha tracks in the TPCs to curve slightly from the presence of this magnetic field. As a result, the θ and ϕ distributions of alpha tracks were skewed slightly, so we opted to select alphas within $\pm 2°$ of both the mean θ_{TPC} and ϕ_{TPC} in each TPC. We expect the ionization energy distributions of these alphas be approximately Gaussian distributed. We thus plot these distributions and set our calibrated gain, G, to be the value that centers the ionization distribution of these selected alphas at $\langle E_{\mathrm{ionization}} \rangle = 1430$ keV in each TPC (Figure 3.4). The presence of curvature in these tracks leads to a slight increase in the expected ionization energy of each event, but we expect this effect to be at most a few percent, so we acknowledge our calibrated gains (labeled G_{eff} in Table 3.1) to be slight overestimates.
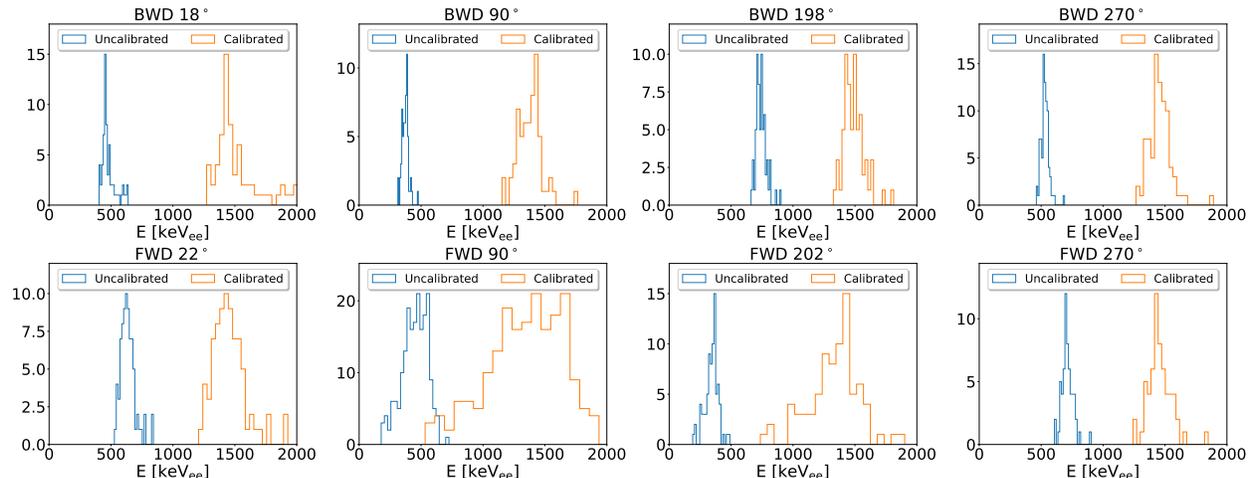
Figure 3.4: Ionization energy distributions of calibration alphas that are within $\pm 2°$ of both the mean θ_{TPC} and ϕ_{TPC} in each TPC before and after gain calibrations. We assume an uncalibrated gain of 2,000 leading to the distributions labeled "Uncalibrated" in each plot. The calibrated gains that lead to the distributions labeled "Calibrated" in each plot are listed in Table 3.1.

3.3.4 Event classification

For fast neutron analyses, the ionization energy of a track and its length in three dimensions provide sufficient information for selecting high purity nuclear recoil samples in a TPC. Once charge and gain have been calibrated, the ionization energy of a track is measured by summing over the energy deposited in each pixel hit in the event

$$E_{\text{ionization}} = \sum_{\text{hits}} E_{\text{hit}}. \tag{3.3}$$

Before determining nuclear recoil selections, we apply a "fiducialization preselection" where we reject all events that register pixel hits on any of the four edges of the FE-I4B chip. This removes calibration alpha tracks since they span the entire x_{TPC} extent of the chip and also helps ensure that the recorded ionization energy isn't biased by events with charge outside of the fiducial volume of the TPC. Using a Singular Value Decomposition (SVD) [70], we identify the principal axis of each track and take the difference between the highest and lowest values of position along this axis to be the 3D track length.
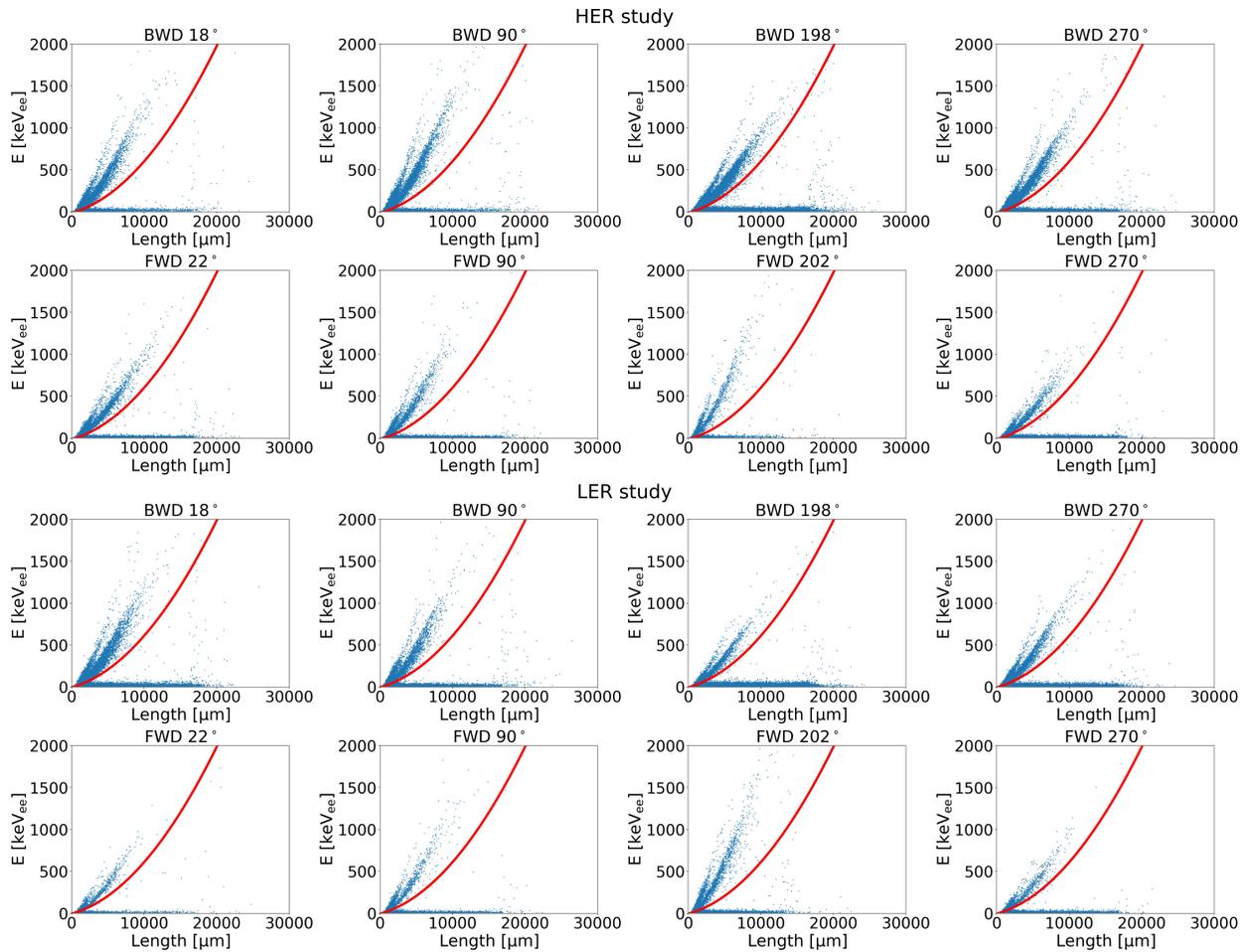
Figure 3.5: Energy versus length distributions of all tracks in each TPC during the HER (top two rows) and LER (bottom two rows) background studies. The red line in each plot shows our selection boundary; we keep all events with energies above this boundary.

29
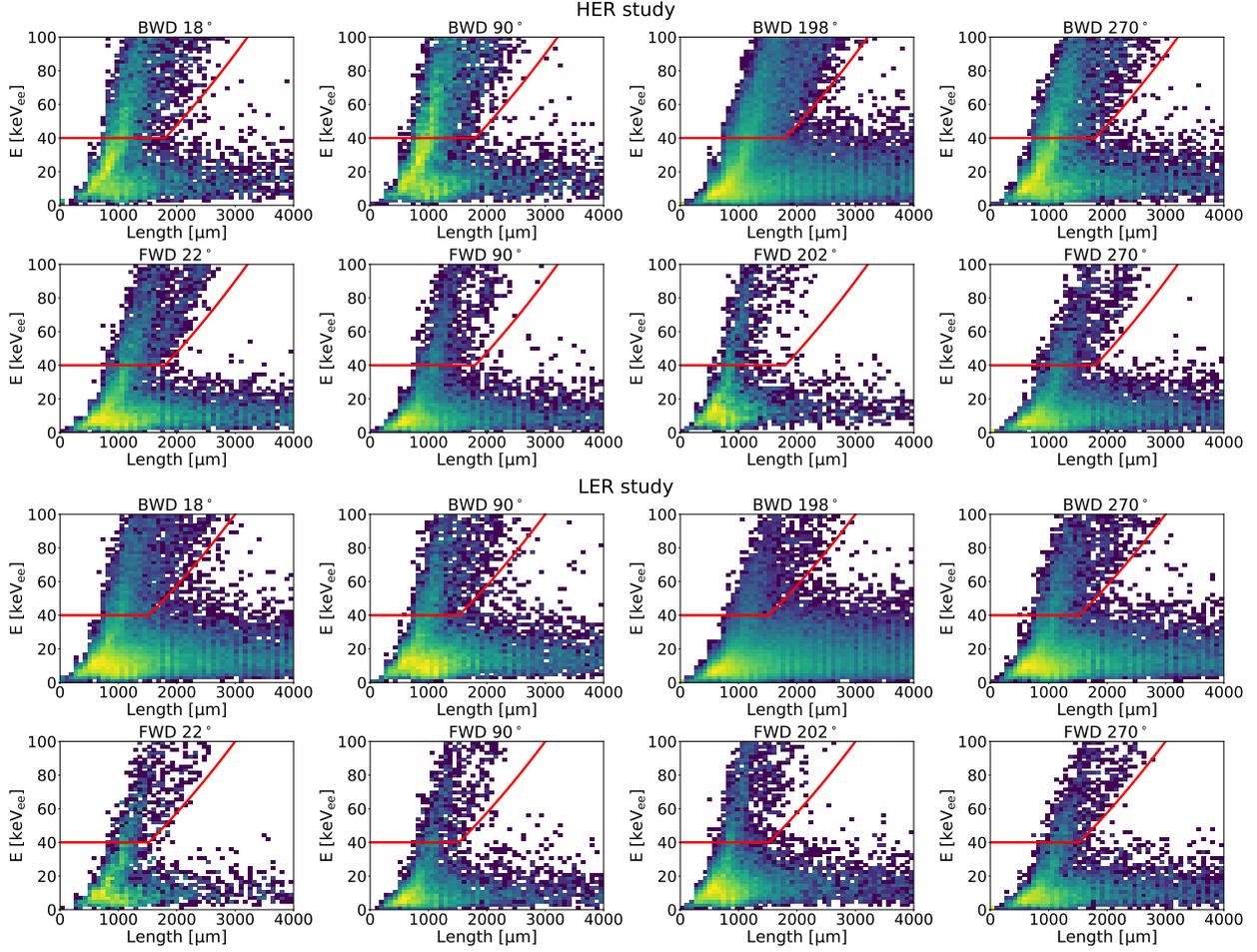
Figure 3.6: Same as Figure 3.5 except zoomed in and binned into $(80\,\mu\mathrm{m} \times 2\,\mathrm{keV_{ee}})$ bins. The color scale shows the relative number of events per bin on an increasing logarithmic scale with blue representing few events, and yellow representing the most events.

Figures 3.5 and 3.6 show distributions of track energy versus length in each of the eight TPCs after the fiducialization preselection. In each plot, there are three distinct $\mathrm{d}E/\mathrm{d}x$ bands. The relatively flat lowest energy band $(\mathrm{d}E/\mathrm{d}x \sim 0)$ corresponds to electron recoils from X-ray conversions, which is the predominant source of background in all TPCs. The remaining two curved $\mathrm{d}E/\mathrm{d}x$ bands correspond to, in order of increasing $\mathrm{d}E/\mathrm{d}x$: ^{4}He recoils and ^{12}C/^{16}O recoils. The $\mathrm{d}E/\mathrm{d}x$ distributions of recoiling ^{12}C and ^{16}O nuclei are similar enough that we do not distinguish between them in measurement. We use a single signal selection boundary (red line in Figures 3.5 and 3.6) for all TPCs and all Phase 2 background studies. We prioritize nuclear recoil purity when defining this selection boundary and start with a preselection where we only keep events with $E > 40\,\mathrm{keV_{ee}}$. Figure 3.6 shows a zoomed-in 2D histogram of energy versus length binned into $(80\,\mu\mathrm{m} \times 2\,\mathrm{keV_{ee}})$ bins with the number of entries in each bin denoted by color. A common logarithmic color scale is

shared among every plot in this figure, with blue representing few events and yellow representing lots of events. We can see that our $E > 40\,\mathrm{keV_{ee}}$ preselection rejects the vast majority of events as X-ray backgrounds. Due to our lack of fully simulating recoil events, we are unable to use simulation to inform the $\mathrm{d}E/\mathrm{d}x$ portion of our event selection boundary, so we generate this portion of the boundary "by eye" and keep events that are clearly within the $^4\mathrm{He}$ and $^{12}\mathrm{C}/^{16}\mathrm{O}$ recoil bands. In Chapter 4.3 we perform similar event selections in a much more systematic way for Phase 3 background studies where we do have fully digitized simulated TPC events.

3.3.5 Merging accelerator and TPC data

Summaries of key SuperKEKB parameters are stored as process variables using the Experimental Physics and Industrial Control System (EPICS PVs) [60] that update every second. These PVs are archived internally for Belle II and SuperKEKB collaborators using custom software built around the EPICS Archiver Appliance [71]. All accelerator parameters used in this analysis are extracted from this PV archiver and are merged with calibrated TPC data by matching integer timestamps of TPC data with integer timestamps of all accelerator PVs. For cases where there are multiple TPC events within a one second window, the accelerator data is duplicated for each TPC event. In this way, when we model nuclear recoils as a function of accelerator parameters, these models are *rate weighted*.

3.4 Modeling beam-induced backgrounds

Dedicated single-beam HER and LER studies were conducted on June 11th, and June 12th, 2018, which we will refer to as the HER study and LER study, respectively. Figure 3.7 shows how several accelerator parameters vary over the course of these studies. Each of the cyan regions in this figure show the beam storage periods where collected TPC data was analyzed. Within these storage fills, we see that beam currents $I_{\mathrm{HER}}/I_{\mathrm{LER}}$, and beam-pipe pressures, $\overline{P}_{\mathrm{HER}}/\overline{P}_{\mathrm{LER}}$ are varied, and between certain storage fills, vertical beam sizes (also called vertical bunch widths), $\sigma_{y,\mathrm{HER}}/\sigma_{y,\mathrm{LER}}$ and numbers of bunches, $n_{b,\mathrm{HER}}/n_{b,\mathrm{LER}}$, are varied. Since Touschek backgrounds vary with charge density, explicitly changing the total number of bunches or the vertical bunch width provides a probe for disentangling Touschek background contributions from beam-gas contributions. Our aim in the coming sections is to disentangle all contributions to beam-induced background rates, which allows for a direct comparison between the observed and predicted neutron background compositions, providing validation of our modeling of the mechanisms of neutron background production at SuperKEKB. Nuclear recoil rates from collision-induced luminosity backgrounds were found to be negligible in the TPCs during Phase 2 so we do not discuss Phase 2 luminosity backgrounds here.
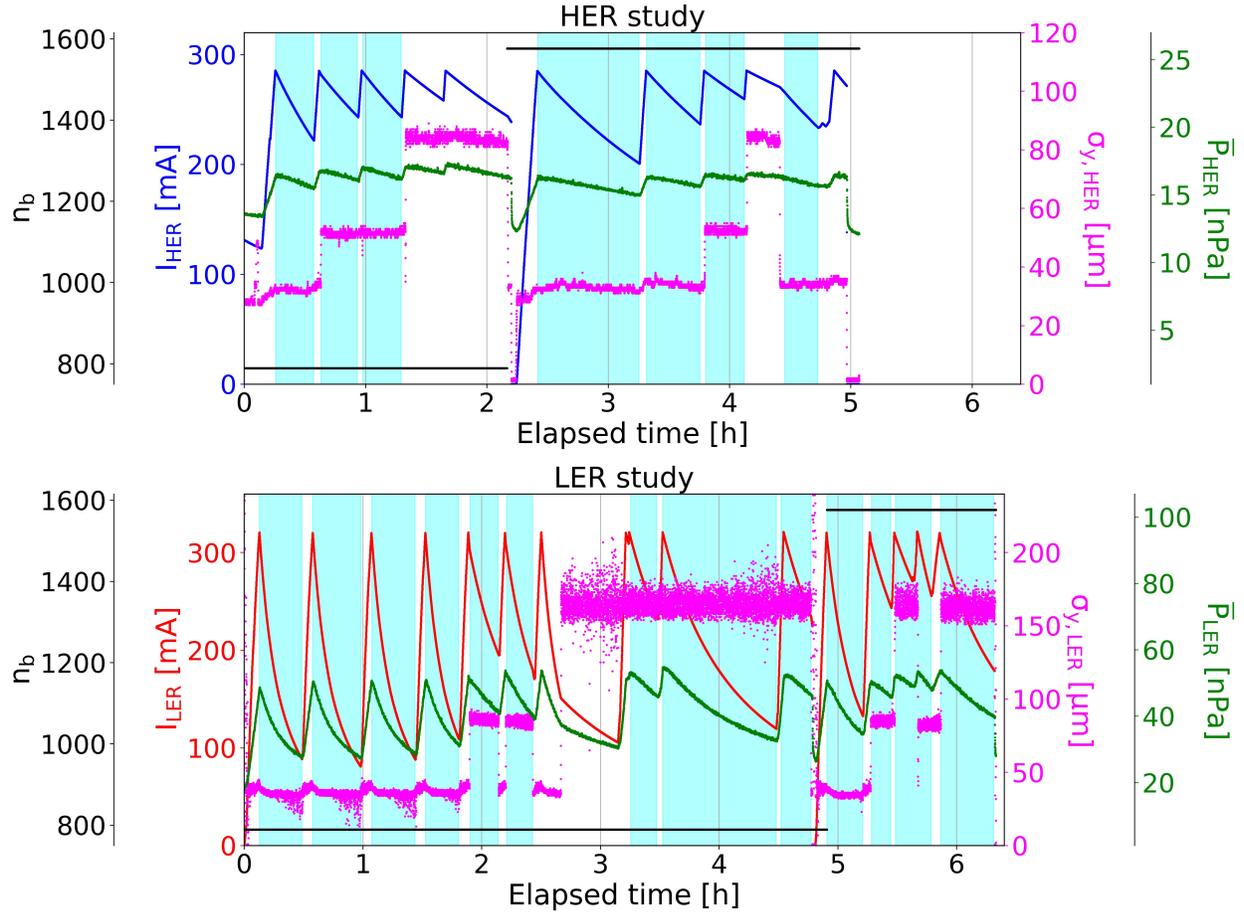
Figure 3.7: Beam currents, I, average beam-pipe pressures around the ring, $\overline{P}$, vertical beam sizes, σ_y, and numbers of bunches n_b plotted as a function of time over the course of the HER (top) and LER (bottom) studies. The cyan-filled regions represent beam storage periods where data was analyzed. We note that the largest HER beamsize periods are omitted from our analysis, due to an increase in background that wasn't understood at the time.

3.4.1 Beam-gas backgrounds

Beam-gas scattering occurs when e^+ and e^- beam particles interact with gas atoms present within the beam pipes. These interactions can either happen through Coulomb scattering between the beam particle and the gas atom, or through Bremsstrahlung. To good approximation, the rate of beam-gas scattering events in a given ring is proportional to IPZ^2, where I is the beam current, P is pressure at a given position along the ring, and Z^2 is the square of the atomic number of the gas constituent that the beam particles interact with.

We employ a physically motivated parametrization of nuclear recoil rates, R, that encapsulates the sensitivity of R to beam-gas and Touschek background components. Though beam-gas scattering rates scale as IPZ^2, there are only 3 residual gas analyzers present around the ring that can provide measurements of the gas composition inside the beam pipe. Since we are unable to directly measure this local gas composition variation throughout the majority of the accelerator rings, we parameterize the beam-gas contributions to nuclear recoil rates, R_{bg} as

$$R_{bg} = B \cdot IP, \tag{3.4}$$

where we absorb Z^2 into the *beam-gas sensitivity* parameter, B. Absorbing Z^2 into B means that B is not a constant and will in general vary with time and position along the beam pipe, so our treatment of B as a constant in our analyses is a source of uncertainty that we do not explicitly measure. In practice, however, we find that our constant-B model, when including Touschek backgrounds, describes our measurements well (Equation (3.8); Figures 3.8 and 3.9) so we expect that our absorption of Z^2 will simply bias the agreement between measured and simulated nuclear recoil rates away from unity.

We also don't account for local variations in beam pipe gas pressure around the ring during these Phase 2 analyses. We instead model pressure contributions to beam-gas scattering rates assuming a uniform pressure distribution given by the average beam-pipe pressures in the HER and LER beam pipes (during Phase 3, we do account for pressure variations as can be seen in Chapter 4.4.1). Equation 3.4 is thus a simplified parametrization that encapsulates the beam-gas contributions from a given beam pipe to the observed nuclear recoil backgrounds in a TPC. This model can be applied to both beam pipes so for a given beam pipe, we write

$$R_{bg,i} = B_i \cdot I_i \overline{P}_i; \quad i = \text{HER, LER}. \tag{3.5}$$

3.4.2 Touschek and total single-beam-induced neutron backgrounds

The Touschek effect describes Coulomb scattering between particles within an individual beam bunch, causing the momenta—and by extension, the orbits—of the Touschek scattered particles to deviate from those of the rest of the bunch [72]. For a single beam particle within a bunch, the Touschek scattering rate is proportional to the particle density of the bunch, $I_b/(\sigma_x \sigma_y \sigma_z)$, where I_b is the bunch current, σ_z is the longitudinal bunch width, and σ_x and σ_y are the horizontal and vertical transverse bunch widths. To model the Touschek background rates around the ring, we multiply the Touschek scattering rate of a single particle by the number beam particles around the ring, $I_b n_b$, where n_b is the number of bunches in the beam train, suggesting that Touschek scattering rates, R_T, can be parametrized as

$$R_T = T \cdot \frac{I_b^2 n_b}{\sigma_x \sigma_y \sigma_z}, \tag{3.6}$$

where T is the experimental *Touschek sensitivity* parameter that encodes all effects contributing to Touschek background rates that aren't explicitly adjusted in the experiment. In the analyses that follow, we measure T for nuclear recoils observed in the TPCs, so moving forward, R_T is understood to be the number of Touschek scattering-induced nuclear recoils measured by a TPC. During the data collection periods for these background studies, σ_x was not well measured and measurements for the longitudinal bunch length, σ_z, were not provided. We thus absorb the effects of σ_x and σ_z into T, leaving us with a final Touschek background rate parametrization in each ring of

$$R_{T,i} = T_i \cdot \frac{I_i^2}{\sigma_{yi} n_{bi}}; \quad i = \text{HER, LER}, \tag{3.7}$$

where we have used $I_b = I/n_b$ to express the bunch current in terms of beam current I and number of bunches n_b. Since we restrict our analyses to beam storage fills, beam-gas and Touschek contributions will overwhelmingly dominate the single-beam-induced nuclear recoil rates observed by the TPCs. Following the lead of Ref. [32], we thus combine Eqs. (3.5) and (3.7) and write

$$R_{\text{SB},i} = R_{bg,i} + R_{T,i} \tag{3.8}$$
$$= B_i \cdot I_i \overline{P}_i + T_i \cdot \frac{I_i^2}{\sigma_{yi} n_{bi}}; \quad i = \text{HER, LER},$$

as our *combined single-beam background parametrization*. Fitting measured nuclear recoil data with this parametrization provides empirical measurements of B_i, and T_i, which can then be compared with simulation and used to extrapolate expected single-beam-induced TPC event rates to different

accelerator conditions.

3.5   Analysis and results

Here we detail the composition, energy spectra, and spatial distributions of fast neutron background events measured in the TPCs. Together, this information paints a picture of how well we model the neutron production mechanisms, material description, and geometry of neutron production near the QCS final focusing, which is crucial for assessing the safety of Belle II electronics to fast neutron backgrounds produced near the IP that radiate outward toward the outer Belle II detectors.

3.5.1   Background composition and rates

Coefficients B_i and T_i from Equation 3.8 are used to estimate the fractional contributions of beam-gas and Touschek backgrounds in each ring for a given set of machine parameters $\{I_i, \overline{P}_i, \sigma_{yi}, n_{bi}\}$. We measure these coefficients by recasting Equation 3.8 as

$$\frac{R_{\mathrm{SB},i}}{I_i \overline{P}_i} = B_i + T_i \cdot \frac{I_i}{\sigma_{yi} n_{bi} \overline{P}_i}, \tag{3.9}$$

and performing a linear fit to $\frac{R_{\mathrm{SB},i}}{I_i \overline{P}_i}$ versus $\frac{I_i}{\sigma_{yi} n_{bi} \overline{P}_i}$. Figures 3.8 and 3.9 show the results of these fits for the HER and LER studies respectively, with the beam-gas contributions shaded in pale gray and Touschek background contributions shaded in dark gray. Data from the study periods with the beam-emittance control knob setting that corresponded to the largest HER vertical beam sizes were omitted from these fits (Figure 3.7 shows these study periods not highlighted in cyan indicating that they weren't included in these analyses), as this emittance control knob setting led to an unexpected increase in background rates.

Ring	I [mA]	P [nPa]	σ_y [µm]	n_b
HER	287	133	36	789
LER	341	133	38	789

Table 3.2: Key machine parameters used for SAD simulation of beam-induced backgrounds. We use B_i and T_i; $i =$ HER, LER to scale measured rates to these machine conditions.

Figure 3.10 summarizes the background composition versus TPC location for the HER study (top) and LER study (bottom). To generate this figure, we use coefficients B_i and T_i to scale measured LER and HER background rates to the conditions used to simulate beam-induced backgrounds in SAD that are shown in Table 3.2. In general, we find that simulation over-predicts LER neutron backgrounds and under-predicts HER neutron backgrounds. Looking at background
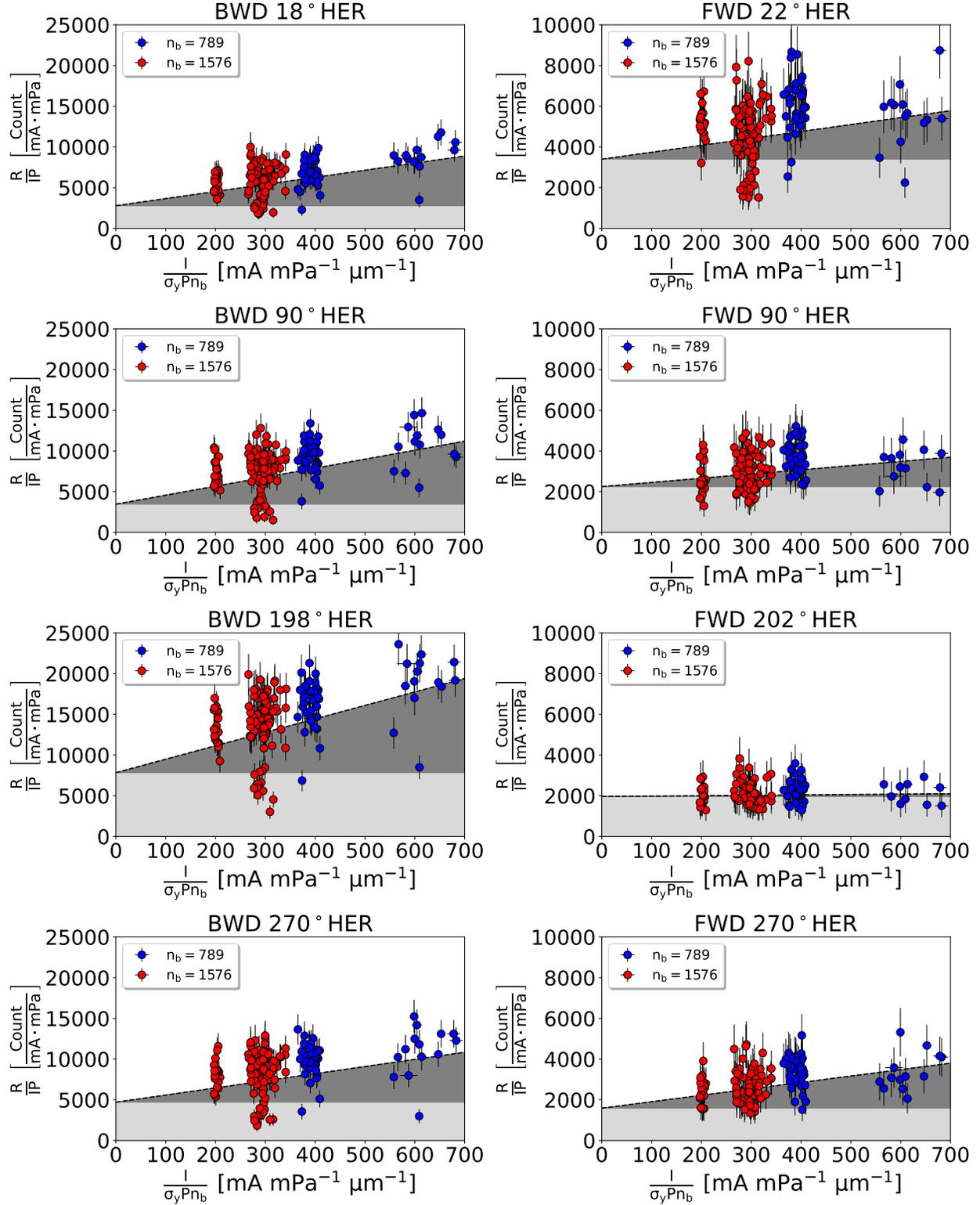
Figure 3.8: Fits from Equation (3.9) performed on data recorded during the HER study. The pale gray shaded region shows the beam-gas contribution to nuclear recoil background rates and the dark gray region shows the Touschek contribution to nuclear recoil background rates.
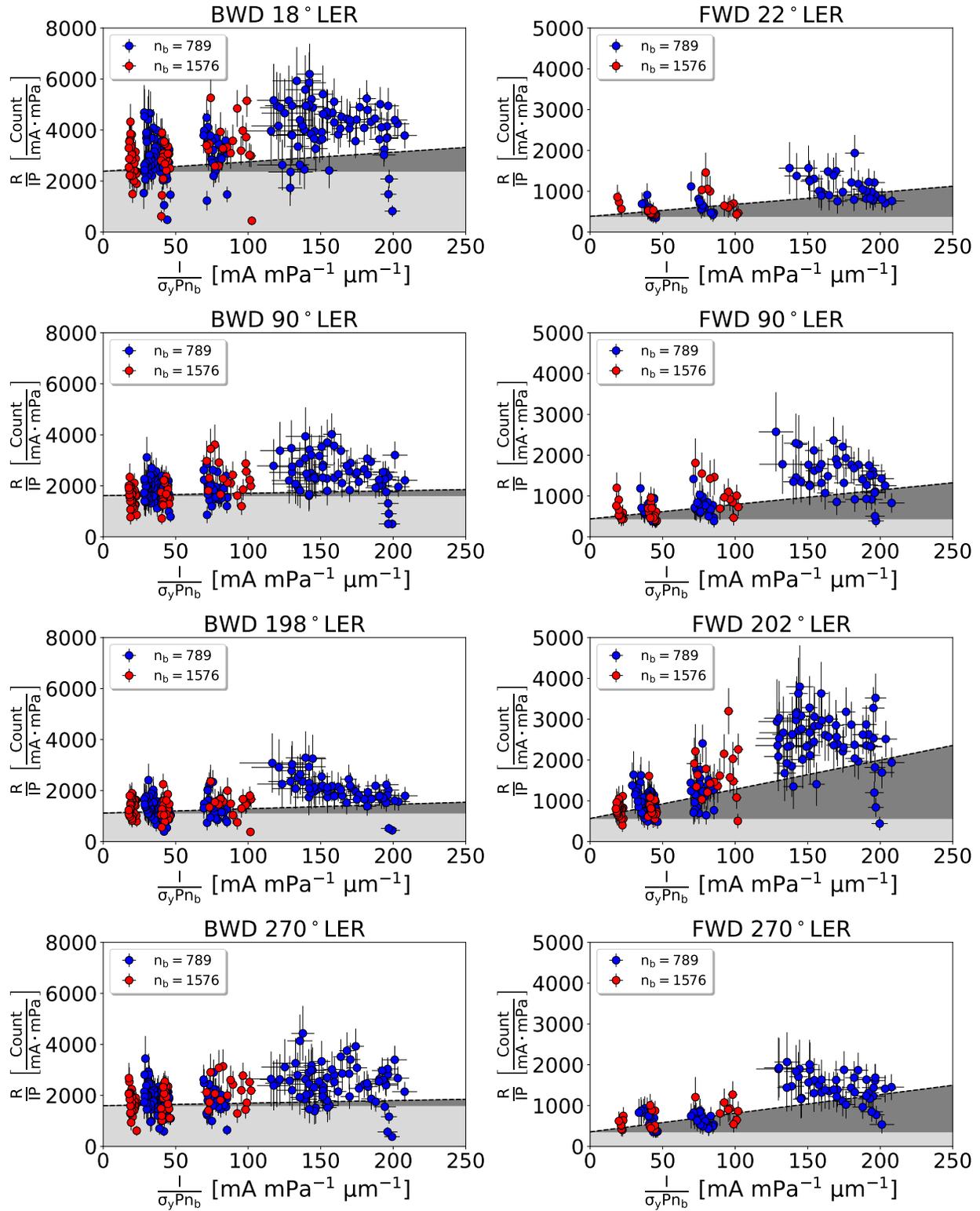
Figure 3.9: Same as Figure 3.8 but for the LER study.
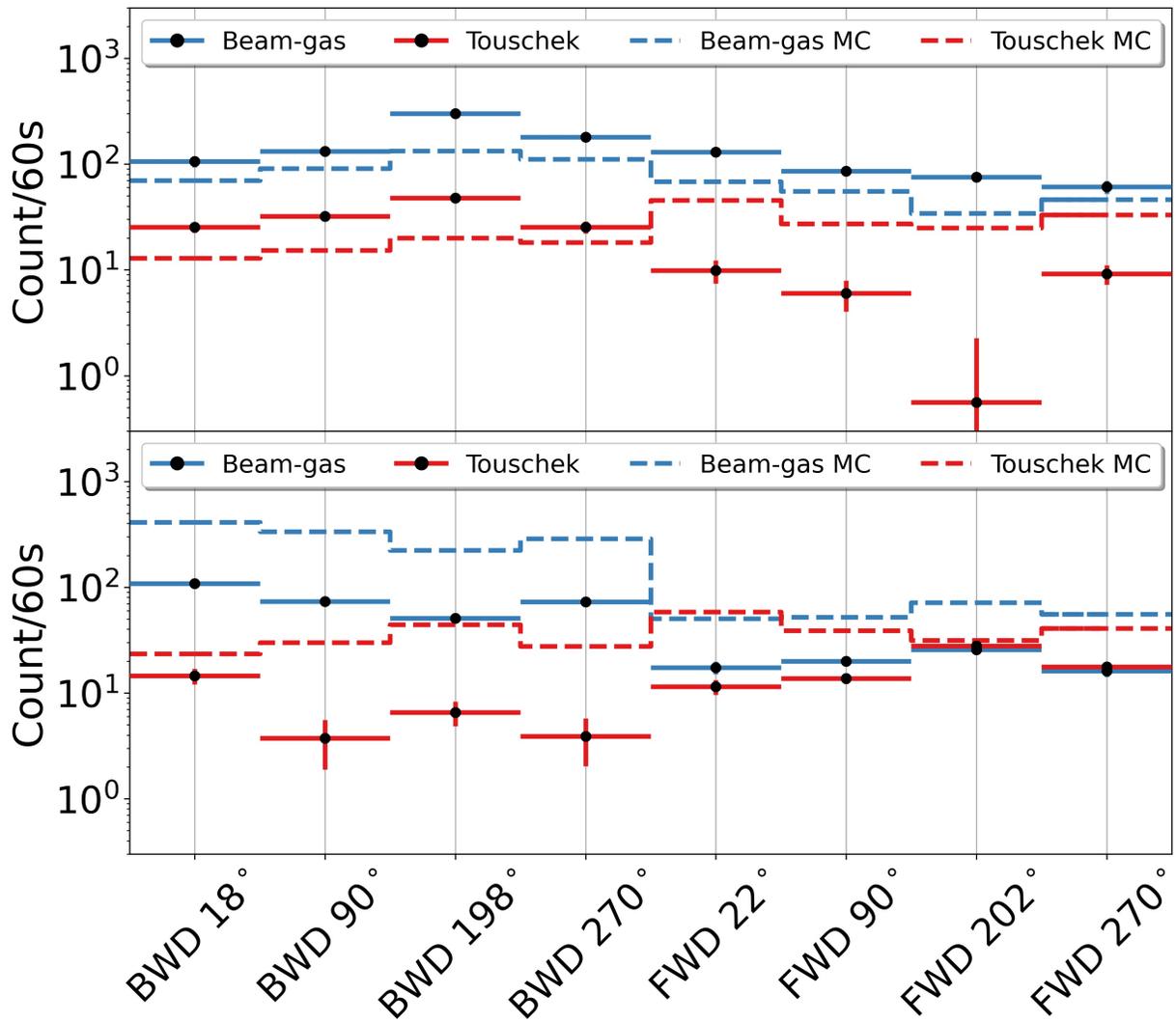
Figure 3.10: Comparison of modeled and measured beam-gas and Touschek backgrounds for the HER study (top) and LER study (bottom) in each TPC (TPC location listed as the horizontal axis label). The black markers with solid red and blue errorbars show measured rates and the dashed lines show simulated rates.

compositions, in both studies, we find that we correctly model that beam-gas backgrounds dominate over Touschek backgrounds, especially on the BWD side of Belle II. By comparing the relative measured and simulated rates between TPCs while ignoring normalization, we further find that overall neutron background localization is modeled well, particularly with the dominant beam-gas backgrounds and on the BWD side of Belle II. This relative agreement was not expected given the complexities of modeling neutron backgrounds in this region.

3.5.2 Energy spectra

Figures 3.11 and 3.12 show comparisons of measured and simulated nuclear recoil energy spectra above our threshold of $40\,\mathrm{keV_{ee}}$ for the HER and LER background studies, respectively. Despite using a fast simulation as described in Chapter 3.2, we find very good agreement between measured and simulated nuclear recoil energy spectra in most TPCs. This agreement suggests that the material description of components that neutrons interact with from their production to when they reach the TPCs is modeled reasonably.

3.6 Neutron flux estimates and extrapolations to future accelerator conditions

Ring	I [mA]	P [nPa]	σ_y [µm]	n_b
HER	1,820	17	0.164	1,576
LER	2,520	48	0.166	1,576

Table 3.3: Machine parameters of interest corresponding to SuperKEKB's target luminosity for 2027.

We can use the measured background sensitivity parameters B_{HER}, B_{LER}, T_{HER}, and T_{LER} to compute pessimistic estimates of neutron fluxes produced near the QCS final focusing both at the Phase 2 operating conditions shown in Table 3.2 and at machine conditions representative of the 2027 target luminosity of $2.8 \times 10^{35}\,\mathrm{cm^{-2}s^{-1}}$ shown in Table 3.3, which were obtained from Ref. [34]. We use the following procedure to determine the estimated neutron fluxes through each TPC:

1. Use Equations (3.5) and (3.7) to estimate the measured beam-gas and Touschek components contributing to the overall nuclear recoil rates from both the LER and HER at the machine conditions listed in Table 3.2 (for Phase 2 estimates) and Table 3.3 (for 2027 estimates).

2. Convert these nuclear recoil rates into an annual neutron fluxes. We use the following procedure to compute these:
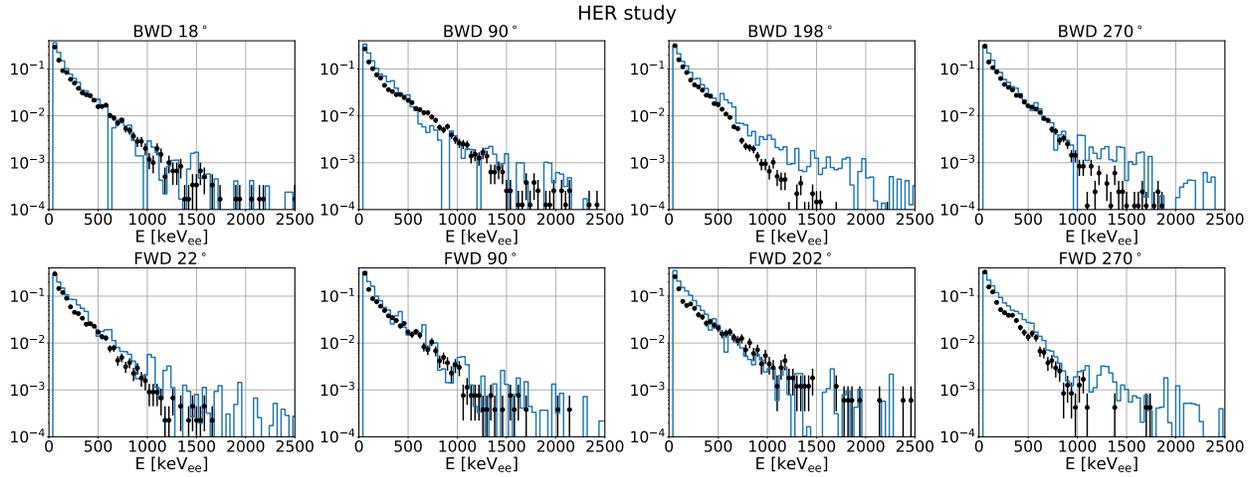
Figure 3.11: Comparison of measured (black points) and simulated (blue bars) nuclear recoil energy spectra above $40\,\text{keV}_{ee}$ for the HER background study. All distributions have been normalized to an integral of unity.
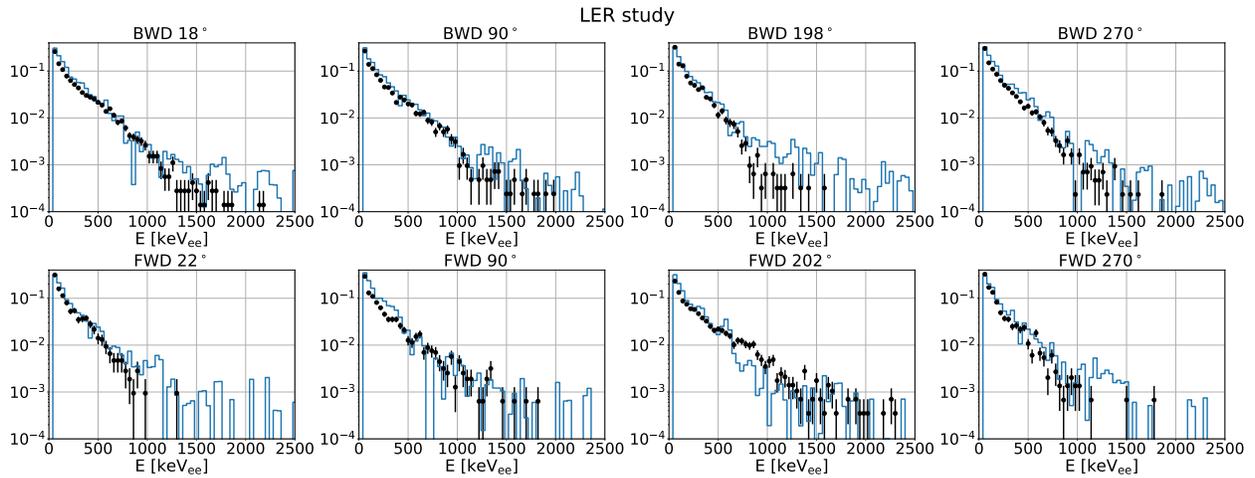
Figure 3.12: Comparison of measured (black points) and simulated (blue bars) nuclear recoil energy spectra above $40\,\text{keV}_{ee}$ for the LER background study. All distributions have been normalized to an integral of unity.

(a) Given the orientations of the TPCs surrounding the QCS, we would expect that most neutrons are incident on the radially innermost surface of a TPC, so they would travel approximately along the TPC y-direction. The neutron-gas nucleus interaction probability, P_{int}, is in general is $\mathcal{O}(10^{-4})$/cm so we would estimate the rate of neutrons passing through the TPC as the nuclear recoil rate divided by $P_{int}y$, where $y \sim 2$ cm is the fiducial y-length of the TPC.

(b) The x-z plane of the TPC is the area we would use to scale this neutron rate to an incident flux, so we divide the neutron rate from part (a) by $xz = (1.68\,\text{cm} \times 10\,\text{cm})$ to get the estimated neutron flux incident on a TPC.

Following this procedure, we obtain the fluxes shown in Tables 3.4 and 3.5.

TPC Location	$\Phi_{\text{BG, HER}}$ $[10^{11}/\text{cm}^2/\text{year}]$	$\Phi_{\text{T, HER}}$ $[10^{11}/\text{cm}^2/\text{year}]$	$\Phi_{\text{BG, LER}}$ $[10^{11}/\text{cm}^2/\text{year}]$	$\Phi_{\text{T, LER}}$ $[10^{11}/\text{cm}^2/\text{year}]$	Safety factor
BWD 18°	0.053±0.006	0.013±0.001	0.054±0.001	0.007±0.001	8
BWD 90°	0.066±0.007	0.016±0.002	0.036±0.001	0.002±0.001	8
BWD 198°	0.149±0.010	0.024±0.002	0.025±0.001	0.003±0.001	5
BWD 270°	0.089±0.007	0.013±0.002	0.036±0.001	0.002±0.001	7
FWD 22°	0.064±0.006	0.005±0.001	0.009±0.001	0.006±0.001	12
FWD 90°	0.043±0.004	0.003±0.001	0.010±0.001	0.007±0.001	16
FWD 202°	0.037±0.004	0.000±0.001	0.013±0.001	0.014±0.001	16
FWD 270°	0.030±0.004	0.005±0.001	0.008±0.001	0.009±0.001	19

Table 3.4: Estimated annual (Snowmass year $= 10^7$ seconds) neutron flux incident on each TPC for each background type and each storage ring scaled to the Phase 2 machine conditions shown in Table 3.2. Safety factors are computed using the sum of each of these fluxes assuming an annual Belle II flux tolerance of 10^{11} neutrons/cm^2.

TPC Location	$\Phi_{\text{BG, HER}}$ $[10^{11}/\text{cm}^2/\text{year}]$	$\Phi_{\text{T, HER}}$ $[10^{11}/\text{cm}^2/\text{year}]$	$\Phi_{\text{BG, LER}}$ $[10^{11}/\text{cm}^2/\text{year}]$	$\Phi_{\text{T, LER}}$ $[10^{11}/\text{cm}^2/\text{year}]$	Safety factor
BWD 18°	0.043±0.005	55.4±6.4	0.143±0.003	45.0±7.6	0.01
BWD 90°	0.053±0.006	70.3±7.1	0.097±0.003	11.6±5.7	0.01
BWD 198°	0.120±0.008	105±9.5	0.067±0.002	20.4±5.4	< 0.01
BWD 270°	0.072±0.006	55.7±7.1	0.096±0.003	12.1±5.8	0.01
FWD 22°	0.052±0.004	21.6±5.4	0.023±0.003	35.5±5.8	0.02
FWD 90°	0.034±0.004	13.2±4.3	0.026±0.002	42.5±5.0	0.02
FWD 202°	0.030±0.003	1.23±3.7	0.034±0.002	86.2±5.5	0.01
FWD 270°	0.024±0.003	20.1±4.2	0.021±0.003	55.0±5.5	0.01

Table 3.5: Estimated annual (Snowmass year = 10^7 seconds) neutron flux incident on each TPC for each background type and each storage ring scaled to the 2027 target luminosity conditions shown in Table 3.3. Safety factors are computed using the sum of each of these fluxes assuming an annual Belle II flux tolerance of 10^{11} neutrons/cm^2.

From these flux estimates, we ultimately concluded that fast neutron backgrounds near the QCS final focusing were safe enough to move onto Phase 3 operation, however additional tuning of horizontal collimators was needed to reduce Touschek backgrounds to sustainable levels in the long run.

CHAPTER 4
FAST NEUTRON BACKGROUNDS IN THE SUPERKEKB TUNNEL OUTSIDE OF BELLE II

This chapter is a modified version of an article published in the peer-reviewed journal Nuclear Instruments and Methods in Physics Research - section A (NIM-A) [25] of which I am first author. Some of the introductory material from that reference was moved to Chapters 2 and 3, and we also elaborate on certain portions of the analyses presented there, but otherwise much of the text remains unchanged.

4.1 Introduction

Of the principal neutron production regions at SuperKEKB, the tunnel regions surrounding either side of Belle II were not instrumented in Phases 1 or 2. In this chapter, we analyze fast neutron background measurements recorded by the Phase 3 TPC system which has been in operation since March 2019, and is comprised of six TPCs: three in the BWD tunnel, which corresponds to $z_{\mathrm{BELLE}} < -4\,\mathrm{m}$ and three in the FWD tunnel, which corresponds to $z_{\mathrm{BELLE}} > 4\,\mathrm{m}$ (Figure 4.1). Similar to Phase 2, we make comparisons between measurement and simulation to evaluate the accuracy of our fast neutron background modeling to assess appropriate remediation measures. For Phase 3, however, we produce a full simulation that includes digitized simulated recoil events in each TPC, allowing for a much more detailed analysis. In particular, we make direct comparisons between measured and simulated rates, energy spectra, and directional distributions of nuclear recoils in the TPCs. All reported measurements were recorded during two dedicated background study days conducted on May 9th, 2020 and June 16th, 2021, which we will refer to as Study A and Study B, respectively.

4.1.1 Radiative Bhabha Hotspots

In the previous chapter we discussed beam-gas and Touschek interactions as the principal single-beam induced neutron production mechanisms. Luminosity backgrounds, which are backgrounds produced from colliding beams, are expected to be a large source of neutron production in the tunnel regions surrounding Belle II. Through-going photons, emitted from radiative Bhabha (RBB) or two photon production processes from colliding beams at the IP, travel along the straight section of the beam pipes and ultimately collide with the walls of the beam pipes in the regions where the beampipes start to curve. When these photons have sufficiently high energy, they can cause excitations of atomic nuclei in the beam pipes, called giant resonances [45; 46], which emit copious amounts of neutrons in highly localized regions. The cyan traces in Figure 4.1 draw straight lines

between the simulated production points of such neutrons and where they pass through a given TPC. We call the green-shaded regions in this figure *Radiative Bhabha* (RBB) *hotspots*, as RBB-produced photons are expected to produce the majority of neutrons in these regions. The Phase 3 TPC system aims to confirm the existence of these hotspots in measurement, as they have never before been measured at SuperKEKB.

4.1.2   TPC system during Phase 3 of beam commissioning

The Phase 3 directional fast neutron detection system consists of six BEAST TPCs [23], and their high voltage (HV), low voltage (LV), gas, and data acquisition (DAQ) systems. Figure 4.1 shows the locations of each of the TPCs drawn to scale as yellow boxes. The location of each TPC was chosen with the goal that the TPC would detect background neutrons generated from the predicted RBB hotspots.
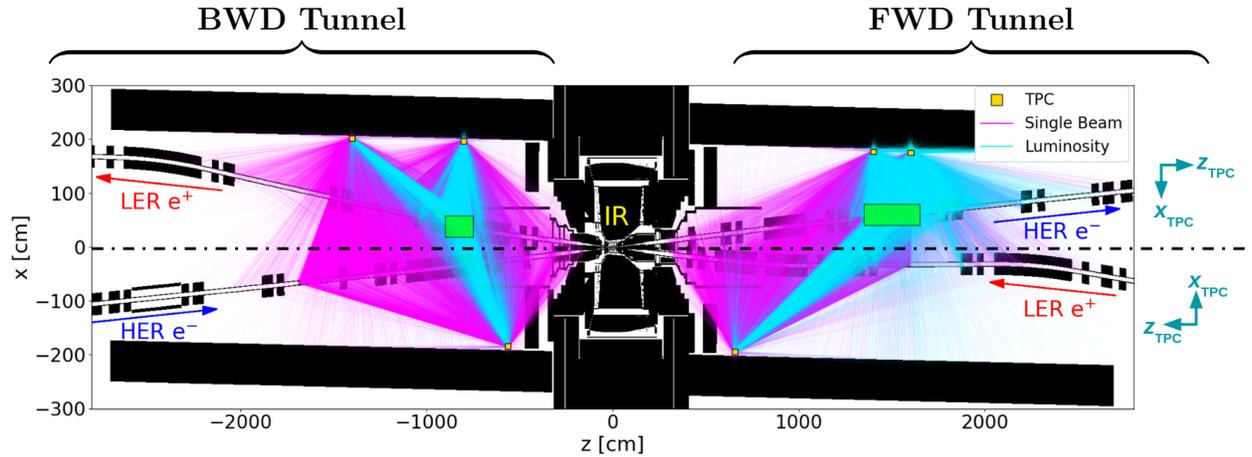
Figure 4.1: Geant4 material scan of the Belle II detector and the SuperKEKB accelerator rings with each of the six Phase 3 TPCs shown to scale as yellow boxes. The coordinate axes on the far right show local TPC coordinate systems in reference to Belle II coordinates. The magenta and cyan traces represent simulated background neutrons passing through the sensitive volume of a TPC, produced from single beams and collisions (luminosity), respectively. Each trace originates at the production point of the simulated neutron and terminates at the TPC vessel. The majority of luminosity-induced simulated neutrons come from two highly localized regions ("hotspots") shown in green.

Where possible, we repurposed the components of the Phase 2 TPC system for use with the Phase 3 TPC system. For gas delivery, we installed new tubing to route between the Gas Box and the new TPC locations, but otherwise use the remainder of the Phase 2 system, including the Gas Box (Chapter 3.1.1). Due to cable length restrictions for LVDS communications, we split our DAQ system into two parallel systems; one located near the BWD side of Belle II and one near the FWD

side. We repurposed the Phase 2 DAQ box to the "FWD DAQ room" and built an additional DAQ box with identical specifications and installed it in the "BWD DAQ room." All together these two DAQ rooms each housed a DAQ box with an associated DAQ computer, two LV power supplies, and one HV power supply. Using this parallel DAQ and HV system scheme, we were able to keep the length of our newly constructed LV and HV cables to between 35 m and 45 m.

4.2   Simulation of TPCs, beam-induced backgrounds, and nuclear recoils

During Phase 3, we employ a multi-step Monte-Carlo (MC) full simulation pipeline to produce files containing digitized recoils in each TPC, which can be compared directly with experimentally measured recoils. We describe the production steps here:

1. **Generation of beam background events**: Like with Phase 2, we use SAD to generate single-beam Bremsstrahlung, Coulomb, and Touschek backgrounds. In Phase 3 we also simulate luminosity backgrounds, as Luminosity backgrounds are expected to dominate over long term accelerator operation. We use `FORTRAN` based event generators [73] to simulate background particles originating from collision-based physics processes. The kinematic information of all initial background particles is saved and passed into Geant4 [62; 63; 64].

2. **Simulation of background showers and neutron propagation**: We use a custom "far beamline" geometry implementation in Geant4 that is integrated into the Belle II Analysis Software Framework (basf2) [65; 66]. This far beamline geometry extends the geometry and material description of accelerator and detector components from the $|z| < 4$ m region used in Phase 2, out to $|z| < 29$ m. Figure 4.1 shows a cross sectional material scan of this region which includes material descriptions of the beam pipes, magnets, collimators, Belle II detectors, and background detectors including BEAST TPCs. Once initial background particles are loaded, Geant4 simulates the physics of the interaction between these particles and the materials present within the geometry. Neutron interactions are simulated using the Geant4 Neutron Data Library (`G4NDL4.6`) [74]. The positions and momenta of all particles depositing energy within the sensitive volume of each TPC are saved. The initial production positions and momenta of all neutrons depositing energy within a TPC are also saved.

Since Phase 2, many improvements have been made to the implementation of tracking and recording SuperKEKB losses in SAD [54], as well as the implementation of accelerator and detector components for SuperKEKB and Belle II in Geant4 [24; 39].

4.2.1 Simulation of TPC detectors

Figure 3.3 shows the probability of an incident neutron interacting with a gas nucleus per centimeter of travel within the fiducial volume of the vessel. Given the low probability of interaction of a neutron passing through the fiducial volume of a TPC, we scale up the elastic scattering cross sections between neutrons and ^{4}He, ^{12}C, and ^{16}O nuclei in the G4NDL4.6 library, each by factors of 100. This factor of 100 scale-up reduces the computational resources necessary to produce a full simulation with adequate simulated nuclear recoil statistics for analysis and leads to a roughly 100-fold increase in the nuclear recoil detection efficiency in each TPC. When comparing measured and simulated nuclear recoil rates in the TPCs in Chapter 8.5, we compensate for this interaction cross section scale-up by dividing the rates predicted by simulation by 100.

The Geant4 simulation step provides energy deposits, positions, and momenta of simulated particles within the sensitive volume of a TPC. For each energy deposit, an ionization distribution is created with the number of electrons in the distribution determined from a random Gaussian distribution centered at $\mu = \frac{E_{\mathrm{dep}}}{W}$, with spread $\sigma = \sqrt{F\mu}$. Here W is the work function (average energy per electron-ion pair) of the gas mixture, F is the Fano factor, and E_{dep} is the ionization energy of the energy deposit. W is determined to be 34.45 eV using Garfield++ [75] and F is set to 0.19, as provided by Heed. Each ionization distribution is then read into a TPC simulation framework developed by our lab.

This custom TPC simulation models the ionization drifting through the simulated field cage volume toward a double GEM layer where the charge is amplified and then "pixelized" into a $2.00 \times 1.68\,\mathrm{cm}^2$ plane containing an 80×336 array of $250 \times 50\,\mathrm{\mu m}^2$ pixels, mimicking the FE-I4B sensitive area. All position resolution effects in the TPC are modeled using Gaussian effective resolutions, as described in [76]. Longitudinal and transverse diffusion constants for the drift volume, and for high-field regions internal to the readout plane were estimated using Magboltz [69]. Each primary electron is diffused individually in three dimensions, based on these diffusion constants and the drift distance to the readout plane. Avalanche gain is simulated at the single-electron level, using an exponential gain distribution. Secondary electrons after gain are again smeared individually, by a combined readout resolution. In the transverse direction, this readout resolution includes contributions from quantization into two GEM holes, diffusion in the transfer gap between the GEMs, and diffusion in the collection gap between the bottom GEM and the pixel chip. Longitudinal diffusion in the same two regions is also included. Finally, the charge is quantized into FE-I4B pixels and readout time bins, and converted into pixel chip-specific charge units called time over threshold (TOT; see Chapter 4.3), based on TPC-specific chip calibration configurations which mimic the experimentally determined charge calibration to convert TOT into charge. These simulated quantized charge distributions encode the same 3D information present in a measured event.

4.2.2   Description of simulated samples generated for analysis

Separate samples for Coulomb, Bremsstrahlung, and Touschek scattering were generated in each ring assuming beam optics and machine parameters representative of conditions recorded during Study A. The number of initial simulated particles passed into Geant4 from these single-beam sources is determined from the loss distributions computed by SAD. The effects of beam pipe pressure on beam-gas losses in SAD are weighted based on measurements from over 300 cold cathode gauges (CCGs) spread around each ring. This implementation [39] provides a more realistic modeling of the beam pipe gas pressure than assuming uniform pressure distributions as was done in Phase 2. Base and dynamic pressure contributions to the SAD simulated Coulomb and Bremsstrahlung losses are separated into individual components, with the base pressure sample assuming $P(I = 0 \, \mathrm{A})$ and the dynamic pressure sample assuming $P(I = 1.2 \, \mathrm{A}) - P(I = 0 \, \mathrm{A})$. Machine parameters used in determining losses are summarized in Table 4.1. Collimator tip scattering and a more accurate beam pipe shape were also implemented in SAD [54], giving better confidence in the modeling of losses from circulating beams around the ring.

	I[A]	σ_y[µm]	n_b	Luminosity [cm^{-2}s^{-1}]
LER	1.2	37	1576	2.5×10^{35}
HER	1.0	36	1576	

Table 4.1: Accelerator and luminosity conditions used to generate MC events.

Background Type	Simulated Beam Time [s] (LER,HER)
Coulomb	(4,40)
Bremsstrahlung	(40,400)
Touschek	(0.4,1.6)
Radiative Bhabha	0.0097
Two-Photon	0.01

Table 4.2: Total simulated beam time for each background process. Values within the parenthetical numerical pairs denote the beam time of single-beam simulation samples in each ring.

For luminosity background samples, we found that the overwhelming majority of simulated radiative Bhabha background events come from the BBBREM event generator [77] (events leading to neutron recoils in the TPCs from the BHWIDE generator [78] were negligible and are thus not included in this analysis). Two-photon background samples are generated using the AAFH event generator [79] with the final state set to $e^+e^-e^+e^-$. Both the radiative Bhabha and two-photon neutron background samples are generated assuming a luminosity of $2.5 \times 10^{35} \, \mathrm{cm}^{-2}\mathrm{s}^{-1}$. Given

the linear dependence of collision rates on luminosity, we can scale these rates to any luminosity as needed. Table 4.2 summarizes the amount of beam time used for each simulated background component.

4.3 Calibration and event selection

Here we discss the charge calibrations, gain calibrations, and event selections used leading up to the analyses for both Study A and Study B.

4.3.1 Pixel-level calibration

In both of our DAQ systems we use identical readout software and data processing procedures to those described in Chapter 3.1.3. We now tune the thresholds to a target value of $2100\,\mathrm{e}$ in all but one of the TPCs and the TOT response is tuned to correspond to a saturation limit—the maximum TOT recorded in an event—above $40\,000\,\mathrm{e}$. Due to operational oversight, the TPC at $z = -8.0\,\mathrm{m}$ was tuned to a target threshold of $2750\,\mathrm{e}$ instead of $2100\,\mathrm{e}$. The results of the threshold tunings and TOT scales vary between FE-I4 modules as can be seen in Table 4.3.

TPC Location	$Q_{\mathrm{TOT}=0}$ $[\mathrm{e}^-]$	$Q_{\mathrm{TOT}=13}$ $[\mathrm{e}^-]$	$G\ [\mathrm{e}^-]$	$E_{\mathrm{drift}}\left[\frac{\mathrm{V}}{\mathrm{cm}}\right]$	$v_d\left[\frac{\mu\mathrm{m}}{25\,\mathrm{ns}}\right]$	Correction Template	Threshold $[\mathrm{keV_{ee}}]$
$-14\,\mathrm{m}$	2120 ± 46	42827	783 ± 45	313	152	f_3	8.0
$-8.0\,\mathrm{m}$	2773 ± 40	47989	794 ± 37	452	216	f_3	6.0
$-5.6\,\mathrm{m}$	2110 ± 32	40821	1015 ± 42	452	216	f_1	9.5
$6.6\,\mathrm{m}$	2071 ± 50	46794	1476 ± 27	452	216	f_1	10
$14\,\mathrm{m}$	2084 ± 35	47304	883 ± 34	452	216	f_2	8.0
$16\,\mathrm{m}$	2083 ± 45	43625	863 ± 84	452	216	f_3	10

Table 4.3: TPC calibration results and settings. Columns from left to right describe: (i) TPC location in Belle II coordinate system, (ii) measured threshold charge, (iii) measured saturation charge, (iv) calibrated effective double GEM gains, (v) approximate drift field, (vi) approximate drift speed, (vii) best template model used for correcting energy due to gain drops during background studies, and (viii) the determined X-ray veto threshold using the procedure outlined in Chapter 3.3.4. Threshold uncertainties (ii) are the standard deviation of the pixel threshold over every pixel in the chip. Gain uncertainties (iv) are given by the standard error of the ionization energy distributions of horizontal alphas in a given TPC. Drift speeds (vi) are calculated using Magboltz [69]. We note that the TPC at $z = -14\,\mathrm{m}$ experienced periodic high voltage trips, so its drift field was lowered to ensure stable operation.

We then calibrate the remainder of the charge scale by sending two hundred injections at several distinct charge steps and measure the TOT response in each pixel. The mean of the TOT over all pixels is recorded for each injection and the mean and standard error of each of these 200 pixel-

averaged TOT values is plotted at each charge step. A bi-cubic spline interpolation of the injection charge versus mean TOT is used to determine the charge corresponding to TOT codes ranging between 0 and 13 in steps of 1.

4.3.2 Gain calibration

We use the same general procedure as described in Chapter 3.3.3 where we use the ^{210}Po alpha sources and calibrate the mean energy of roughly horizontal alpha tracks to a target of 1430 keV. We never replaced these sources between Phases 2 and 3, so given the ^{210}Po half like of 137 days, we had muc smaller alpha samples to calibrate gain during for the Phase 3 study. As a result of this, we only calibrated gain for Study A and use these same calibrations for Study B. The locations of the Phase 3 TPCs are far enough away from the QCS that we longer had to worry about the presence of the 1.5 T final focusing magnetic field during our calibration runs. Given the lower statistics and the fact that alpha ionization trajectories weren't distorted by the B-field, we select alphas satisfying $|\phi_{\text{TPC}}| < 5°$ and $85° < \theta_{\text{TPC}} < 95°$ and expect the ionization energy distributions of these to be approximately Gaussian distributed. As before, we set our calibrated gain, G, to be the value that centers the ionization distribution of these selected alphas at $\langle E_{\text{ionization}} \rangle = 1430$ keV in each TPC. The calibrated GEM gains are shown in Table 4.3. We note that effective gain variations between TPCs result both from inherent GEM gain differences and variation in gas purity, the latter being expected due to each set of three TPCs receiving gas in series.

During both Study A and Study B, we observe drops in $\mathrm{d}E/\mathrm{d}x$ at a fixed track length, ℓ, with increasing luminosity in several TPCs. This effect appears to result from drops in effective gain in these detectors during beam-collisions. A change in background particle composition alone would only change the distribution of events within an E versus ℓ band, or the relative normalization of the bands, but not the shape of individual bands comprised of a single particle species. Both the reduced $\mathrm{d}E/\mathrm{d}x$ for the observed He-recoils and the disappearance for the X-ray bands (which move below threshold at low gain) support the hypothesis that the effective gain is reduced at the highest luminosities.

Given that beam-induced background rates increase with luminosity, we speculate that the observed drop in effective gain is a consequence of reduced gas purity induced by stimulated X-ray desorption at higher luminosities. Since this is speculation and we do not know the exact cause of these drops in gain, we correct for them empirically to keep the $\mathrm{d}E/\mathrm{d}x$ distributions of nuclear recoils reasonably consistent in a given TPC during the entirety of the background studies. We thus introduce three simple template models to fit to the distribution of mean nuclear recoil energies for events with observed ionization energies above 12 keV with 3D lengths (Chapter 3.3.4) between 1.3 mm and 2.6 mm (1.7 mm and 3.0 mm for the TPC at $z = +16$ m due to its sharper drop in effective gain) binned by luminosity and choose the model among the three that gives a reduced χ^2 nearest to 1. The template models are
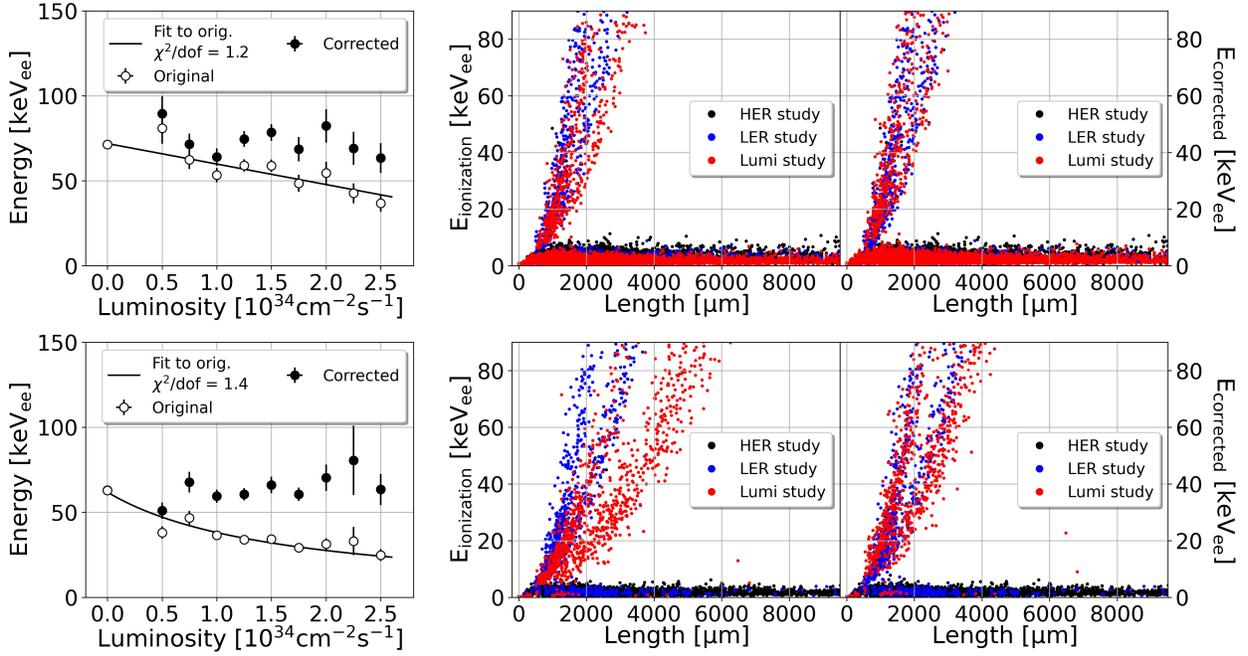
Figure 4.2: (color online) Left: average energy of identified nuclear recoils—both before and after the corrections described in Eqs. (4.1) and (4.2)—versus luminosity for recoils satisfying $1.3\,\mathrm{mm} < \ell < 2.6\,\mathrm{mm}$ (top) and $1.7\,\mathrm{mm} < \ell < 3.0\,\mathrm{mm}$ (bottom), where ℓ is the length along the principal axis of the recoil track. Middle and right: reconstructed ionization energy versus length distributions before and after gain corrections, respectively. The top and bottom rows of plots are associated with the TPCs at $z = +14\,\mathrm{m}$ and $z = +16\,\mathrm{m}$, respectively.

$$f_1(L) = a$$
$$f_2(L) = aL + b$$
$$f_3(L) = \frac{1}{aL + b} \tag{4.1}$$

where L is the measured luminosity and a and b are fit parameters. We then use $f_j(0)/f_j(L)$; $j = 1, 2,$ or 3, as the scale factor for which to modify the energy. Thus, for all events, we define our corrected energy due to gain corrections as

$$E_{\text{corrected}}(L) = \frac{f_j(0)}{f_j(L)} E_{\text{ionization}}, \tag{4.2}$$

where j corresponds to the template that minimizes $|\chi^2/\text{dof} - 1|$. Table 4.3 lists which fit template was used to correct for drops in gain in each TPC and Figure 4.2 shows the effect of these corrections on the TPCs at $z = +14\,\text{m}$ and $z = +16\,\text{m}$. Unless stated otherwise, when we refer to the energy of a measured recoil event in this chapter, we are referring to $E_{\text{corrected}}$.

4.3.3   Event classification

We use the distribution of gain-corrected energy versus track length to select for nuclear recoil events to include in our analyses. Like in Phase 2, we use an SVD to identify the principal axis of each track and take the difference between the highest and lowest values of position along this axis to be the 3D track length. We once again reject all events that register pixel hits on any of the four edges of the FE-I4B chip to remove calibration alpha tracks since they span the entire x_{TPC} extent of the chip.

Figure 4.3 shows the remaining distributions of track energy versus length. Like before, the relatively flat lowest energy band ($\mathrm{d}E/\mathrm{d}x \sim 0$) corresponds to electron recoils from X-ray conversions, which is the predominant source of background in all TPCs. The remaining two curved $\mathrm{d}E/\mathrm{d}x$ bands correspond to, in order of increasing $\mathrm{d}E/\mathrm{d}x$: ^{4}He recoils and ^{12}C/^{16}O recoils. The large region of parameter space between the X-ray and recoil bands contains a relatively small number of events compared to the three primary $\mathrm{d}E/\mathrm{d}x$ bands. Simulation suggests that this region contains a mixture of nuclear recoils and X-rays, but given that there are very few events in this region, we apply quadratic $\mathrm{d}E/\mathrm{d}x$ pre-selections that reject most of this region in order to prioritize nuclear recoil purity. After setting these $\mathrm{d}E/\mathrm{d}x$ pre-selections, we use a data driven approach to determine an "X-ray veto threshold" in each TPC for our analysis. The general procedure follows:

1. Start with an X-ray veto threshold of $5\,\text{keV}_{\text{ee}}$.
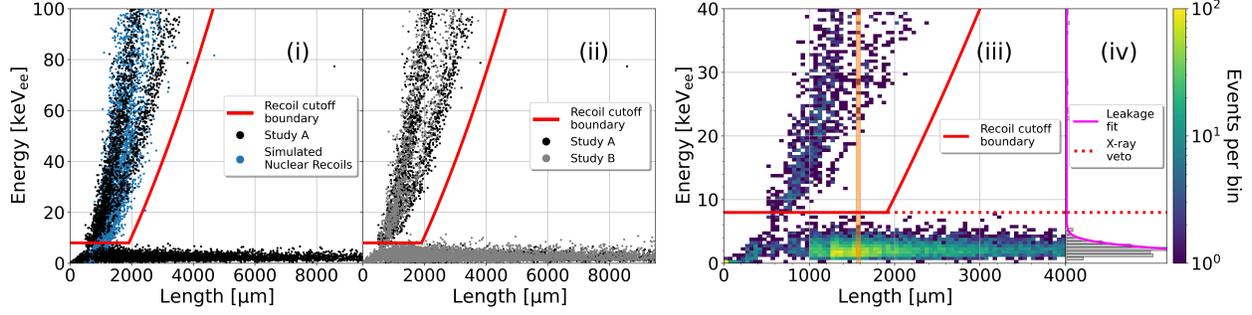
Figure 4.3: (color online) (i): Corrected energy versus 3D track length for simulated nuclear recoil events and all measured events passing our fiducialization preselection during Study A in the TPC at $z_{\mathrm{BELLE}} = +14\,\mathrm{m}$. Events above the red cutoff boundary are identified as nuclear recoils, and those below are rejected background. (ii): same as (i) but comparing measurements between Study A (black points) and Study B (gray points). (iii): zoomed in binned view of measured recoils in (i). (iv): binned energy distribution of all events within the orange shaded region shown in (iii). The energy distribution of events below the X-ray veto is fit with a half-Gaussian profile $\mathcal{F}_{\mathrm{bin}}(x)$ (described in text and shown as the magenta curve here) to estimate the leakage of X-ray events above this threshold.

2. Split the region between $\ell = 850\,\mu\mathrm{m}$ and $\ell = 4000\,\mu\mathrm{m}$ into bins of width $50\,\mu\mathrm{m}$ and in each of these length bins:

   (a) Plot a histogram of the energy distribution of all events. The gray bars in **??** (iv) show this energy distribution for all events within the length bin shaded in orange in **??** (iii). The distribution of events below the recoil cutoff boundary is asymmetric and has a long, approximately Gaussian, high-energy tail.

   (b) Estimate the number of x-ray events as a function of energy for a given length bin using a Gaussian profile of the form

   $$\mathcal{F}_{\mathrm{bin}}(E_{\mathrm{corrected}}) = \hat{A}_{\mathrm{bin}} \exp\!\left(-\hat{B}_{\mathrm{bin}} E^2_{\mathrm{corrected}}\right), \tag{4.3}$$

   fit to the higher-energy half of the observed energy distribution below the recoil cutoff boundary. Fit parameters $\hat{A}_{\mathrm{bin}}$ and $\hat{B}_{\mathrm{bin}}$ are determined by performing a χ^2 minimization of $\mathcal{F}_{\mathrm{bin}}$ evaluated between the energy bin with the most X-ray events and bin corresponding to the recoil cutoff boundary in the given length bin.

   (c) Estimate the leakage above the recoil cutoff boundary using

   $$\mathrm{Leakage}_{\mathrm{bin}} = \int_{\max_E\left(E_{\mathrm{veto}}, E_{\mathrm{d}E/\mathrm{d}x,\mathrm{bin}}\right)}^{\infty} \mathcal{F}_{\mathrm{bin}}(x)\,dx, \tag{4.4}$$

where E_{veto} and $E_{\text{d}E/\text{d}x,\text{bin}}$ are the energies of the proposed X-ray veto threshold and the $\text{d}E/\text{d}x$ pre-selection, respectively.

(d) Divide the estimated leakage above the recoil cutoff boundary by the total number of events above $\max_E\left(E_{\text{veto}}, E_{\text{d}E/\text{d}x,\text{bin}}\right)$ to get the estimated *leakage fraction* above $\max_E\left(E_{\text{veto}}, E_{\text{d}E/\text{d}x,\text{bin}}\right)$. Subtracting this leakage fraction from 1 gives the estimated recoil purity above the recoil cutoff boundary.

3. Adjust the energy of this flat X-ray veto threshold as needed until the estimated recoil purity in each length bin is greater than 99%.

We apply the above procedure to each TPC twice—once for Study A data and once for Study B data—and note that the determined thresholds for a given TPC are similar between these two studies, so we assign the larger of the two thresholds to each TPC with the final choice listed in column (viii) of Table 4.3. These thresholds are defined with respect to $E_{\text{corrected}}$[1]. Despite small differences in $\text{d}E/\text{d}x$ predictions between data and simulation, applying these same selections to simulated samples (Figure 4.4) suggests recoil purities of greater than 99% in all TPCs with nuclear recoil signal efficiencies greater than 75% in all TPCs. The predicted nuclear recoil signal purities
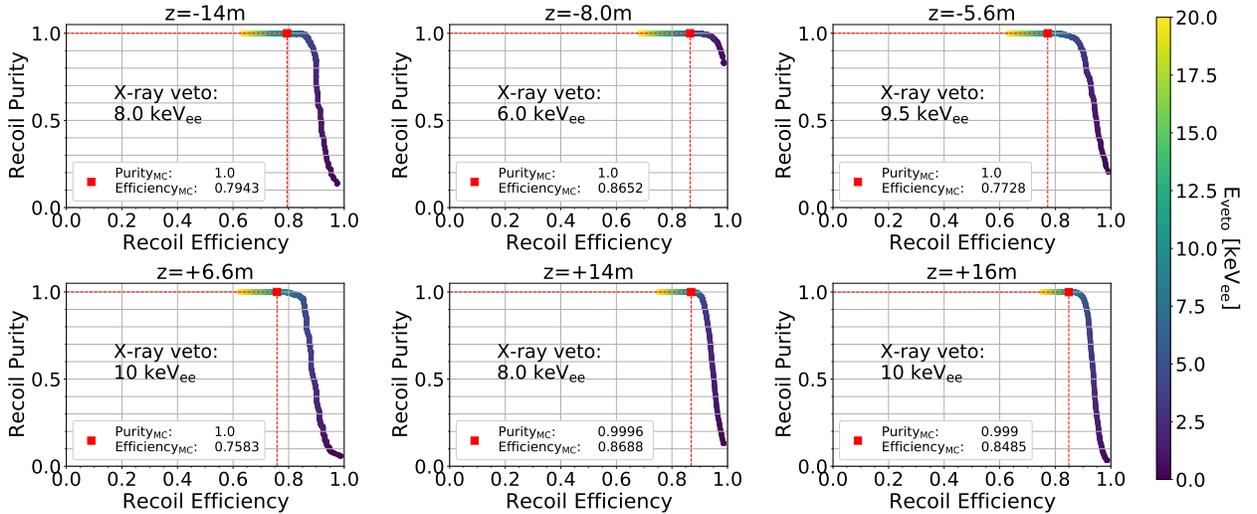
Figure 4.4: Simulated ROC curves showing nuclear recoil signal purity versus efficiency as a function of X-ray veto threshold energy for Study A. The boxes show the predicted nuclear recoil purity and efficiencies corresponding to the X-ray veto threshsold decided for each TPC. Our procedure prioritizes selecting a veto with higher nuclear recoil purities than the X-ray veto threshold point that maximizes the area under the ROC curve.

and efficiencies demonstrate that these TPCs are capable of measuring high purity samples of fast

---

[1]We increase the X-ray veto threshold in the TPC located at $z = +16$ m by an additional 60% to ensure that we are above the regime where nuclear recoil events drop below the FE-I4B detection threshold due to the drop in gain during collisions. The X-ray veto thresholds reported in Table 4.3 for this TPC are after the additional 60% increase.

neutrons down to $\mathcal{O}(10\,\text{keV})$ at effective double GEM gains of $\mathcal{O}(1000)$.

4.3.4 Merging accelerator and TPC data

As with our Phase 2 data processing, we extract one second summary EPICS PVs of key accelerator quantities from the PV archiver and merge them with calibrated TPC data by matching integer timestamps of TPC data with integer timestamps of all accelerator PVs. When there are multiple TPC events within a one second window, we duplicate the accelerator data for the given one second window, making our models of nuclear recoil rates versus accelerator parameters *rate weighted*.

4.4 Improved modeling of beam-induced backgrounds
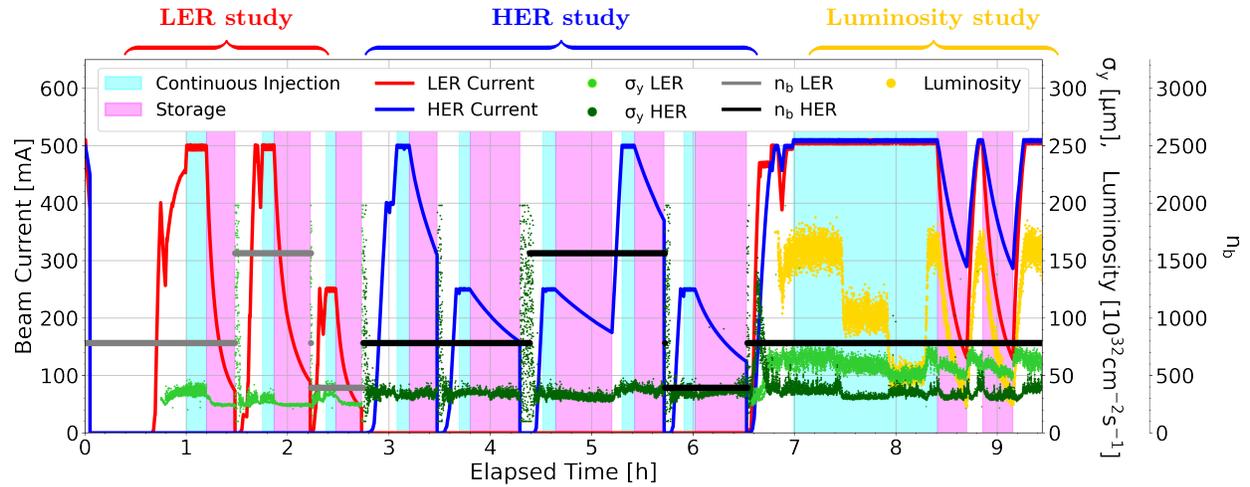
Figure 4.5: (color online) SuperKEKB machine parameters versus time. Beam currents, vertical beam sizes $\sigma_{y,\text{LER,HER}}$, numbers of bunches $n_{b,\text{LER,HER}}$, and luminosity, L, are shown over the course of all three study periods during Study A

In contrast to the Phase 2 background studies, Phase 3 Studies A and B each included an LER study portion, an HER study portion, and a luminosity study portion. Within each of these study periods background data was recorded during rapid continuous injection fills as well as storage ("decay") fills. Figure 4.5 shows how several accelerator parameters vary over the course of the three separate study periods conducted during Study A: we see that numbers of bunches (n_b) were varied in three steps during each of the LER and HER study periods. Rather than explicitly varying beamsize with an emittance control knob, during the Phase 3 background studies we instead use the change in number of bunches (fill pattern) as our probe for Touschek backgrounds. During the luminosity study period, the number of bunches were kept constant with 783 bunches in each ring, while luminosity was varied to control for the measurement of luminosity dependent

backgrounds. When analyzing data, we only consider beam storage fills in our analysis, as there are more systematic differences between the beam optics settings used during during continuous injection fills than those during decay fills.

In Chapter 4.2.2 we describe how we deal with ring sectional pressure weighting when simulating beam-gas backgrounds. As a result, we will treat pressure differently in our modeling of beam-gas backgrounds than we did in Phase 2. Additionally, despite us not explicitly varying σ_y with an emittance control knob during Phase 3, Figure 4.5 shows that during the Luminosity background study $\sigma_{y,\text{LER}}$ is considerably larger than during the single-beam LER study, due to a beam blowup effect during collisions. This beam blowup effect will also affect how we model Touschek backgrounds throughout these Studies. For these reasons we will spend some time motivating our single-beam background models, as these models are a prerequisite to disentangling collision-induced luminosity neutron background production.

4.4.1   Beam-gas backgrounds

As before, we employ a physically motivated parametrization of nuclear recoil rates, R, that encapsulates the sensitivity of R to beam-gas and Touschek background components. To good approximation, the rate of beam-gas scattering events in a given ring is proportional to IPZ^2, where I is the beam current, P is pressure at a given position along the ring, and Z^2 is the square of the atomic number of the gas constituent that the beam particles interact with. No new RGAs were installed between Phases 2 and 3, so we still do not explicitly measure Z. During phase 2, we ignored the fact that P is composed of a base component, P_0 due to the residual gases remaining in the beam pipe without the presence of beams, and a dynamic component P_1, which is proportional to I and $\mathrm{d}P/\mathrm{d}I$ [80]. Our Phase 3 simulations now separately simulate base and dynamic contributions to Coulomb and Bremsstrahlung beam-gas backgrounds, so we outline here how we model these base and dynamic pressure components in measurement.

In experiment, we treat pressure as follows: we first note that the pressure within the beam pipe, P, contains both a base component, P_0, and a current-dependent dynamic component, $P_1(I)$:

$$
\begin{aligned}
P &= P_0 + \frac{\mathrm{d}P}{\mathrm{d}I} \cdot I \\
&= P_0 + P_1(I).
\end{aligned} \tag{4.5}
$$

Previous machine simulation has shown that the dynamic pressure component measured by the CCGs, $P_{1,meas}(I)$ is about a factor of 3 lower than the dynamic pressure within the beam pipe, so

$$P_1(I) = 3P_{1,meas}(I),$$
$$P_0 = P_{0,meas},$$
$$\Rightarrow P = P_{0,meas} + 3P_{1,meas}(I).$$
(4.6)

The measured CCG pressure can be broken down into base and dynamic components analogously to the beam pipe pressure in Eq. (4.5) so we can rewrite Eq. (4.6) as

$$P = P_{0,meas} + 3(P_{meas} - P_{0,meas})$$
$$= 3P_{meas} - 2P_{0,meas}.$$
(4.7)

Substituting Eq. (4.7) into Eq. (3.4) gives

$$R_{bg} = B \cdot I(3P_{meas} - 2P_{0,meas}).$$

Noting that the base pressure is a constant, we define model parameters $B_0 \equiv 3B$ and $B_1 \equiv 2BP_{0,meas}$ leaving us with our beam gas background parametrization:

$$R_{bg,i} = B_{0i} \cdot I_i P_{meas,i} - B_{1i} \cdot I_i \quad i = \text{LER, HER}.$$
(4.8)

Cast in the form of Eq. (4.8), B_0 and B_1 are positive constants that are determined empirically and together encode the sensitivity of nuclear recoil rates to beam gas backgrounds.

4.4.2 Touschek backgrounds

The main differences between modeling Touschek backgrounds in Phase 2 versus Phase 3 is during Phase 3, we had reliable measurements of the longitudinal bunch length, σ_z, and there were also variations in σ_y in the LER that can be seen between the LER study period and the luminosity study period in Figure 4.5. This beam blowup effect that occurs during collisions can have a strong effect on observed rates, especially in the FWD tunnel where LER Touschek backgrounds from upstream collimators are particularly large. Due to this sensitivity to Touschek backgrounds from upstream collimators, we modify the vertical beam size dependence in our *LER* Touschek model from $\sigma_y \rightarrow \sigma_y^{\alpha_{\text{LER}}}$, where the exponent α_{LER} is determined empirically and corrects for discrepancies between Touschek predictions derived from single-beam LER studies and those during the luminosity study period. The assumption underlying this correction is that beam size differences

can cause larger variations in Touschek backgrounds for detectors that are sensitive to backgrounds from collimators than those that are not. With these changes, we update Equation 3.7 and write our Phase 3 Touschek parametrization as

$$R_{T,i} = T_i \cdot \frac{I_i^2}{\sigma_{yi}^{\alpha_i} \sigma_{zi} n_{bi}} \quad i = \text{LER, HER},$$
(4.9)

where we have used $I_b = I/n_b$ to express the bunch current in terms of beam current I and number of bunches n_b. Restricting our analyses to beam storage fills, we don't have to worry about additional injection background components, so beam-gas and Touschek contributions will be the dominant single-beam background sources. Following the lead of Chapter 3.4.2, we combine Eqs. (4.8) and (4.9) and write

$$
\begin{aligned}
R_{\text{SB},i} &= R_{bg,i} + R_{T,i} \\
&= B_{0i} \cdot I_i P_{meas,i} - B_{1i} \cdot I_i + T_i \cdot \frac{I_i^2}{\sigma_{yi}^{\alpha_i} \sigma_{zi} n_{bi}} \quad i = \text{LER, HER},
\end{aligned}
$$
(4.10)

as our Phase 3 *combined single-beam background parametrization*. Fitting measured nuclear recoil data with this parametrization provides empirical measurements of B_{0i}, B_{1i}, and T_i, which can then be compared with simulation and used to extrapolate expected single-beam-induced TPC event rates to different accelerator conditions.

4.4.3 Luminosity backgrounds

We model luminosity backgrounds as the backgrounds that remain during collisions after subtracting out LER and HER single-beam backgrounds. Total nuclear recoil rates, then, are given by

$$R = R_{\text{SB,LER}} + R_{\text{SB,HER}} + R_L,$$
(4.11)

where R is the total nuclear recoil rate and R_L is the luminosity component of nuclear recoil rates. Thus, during the luminosity study period, we measure the luminosity background rate as

$$R_L = R - \left(\sum_{i=\text{LER,HER}} B_{0i} I_i P_{meas,i} - B_{1i} I_i + T_i \frac{I_i^2}{\sigma_{yi}^{\alpha_i} \sigma_{zi} n_{bi}} \right) = m_L L,$$
(4.12)

where L is the instantaneous luminosity, and B_{0i}, B_{1i} and T_i represent the sensitivity coefficients measured in experiment from the single-beam LER and HER background studies, respectively. R_L is expected to be directly proportional to L—which we verify in Chapter 8.5—with *luminosity background sensitivity* parameter m_L.

4.5 Analysis and results

Here we analyze the rates, energy spectra, angular distributions, and directional distributions of recoils measured during Study A. We compare these measurements to simulation and use these comparisons to extrapolate simulated neutron fluxes to our peak target luminosity of $6.3 \times 10^{35}\,\mathrm{cm^{-2}s^{-1}}$. We note that since we do not have TPC simulation for Study B beam optics, we only use Study B measurement to set a systematic on luminosity backgrounds, as luminosity backgrounds are dependent only on collision rate.

4.5.1 Nuclear recoil background composition and rates

Eqs. (4.10) and (4.12) provide a framework for disentangling measured backgrounds into their individual contributions. Here we use these equations to determine the observed nuclear recoil background composition and compare with simulation.

Single-beam background composition

We start by applying Eq. (4.10) to single-beam LER data to measure α_{LER}. To do this, we fit LER data recorded during Study A in each of the three FWD TPCs using values of α_{LER} ranging from 0.5 to 4.5 in steps of 0.1 and record the mean reduced χ^2 of these fits between these three TPCs. We find a minimum reduced χ^2 of 1.44 at $\alpha_{\mathrm{LER}} = 2.4$ with $1.8 \leq \alpha_{\mathrm{LER}} \leq 3.0$ representing the 95% confidence interval for the value of α_{LER} that minimizes the reduced χ^2 of single-beam LER background fits in the FWD TPCs. In Study B, we find essentially no α_{LER} dependence on the reduced χ^2 fits to single-beam LER backgrounds in the FWD TPCs, but we still set $\alpha_{\mathrm{LER}} = 2.4$ since α_{LER} is intended to correct for the effects of LER beam size blow up during collisions. Given that there are no appreciable HER beam size changes during collisions in both Study A and Study B, we set $\alpha_{\mathrm{HER}} = 1$.

Using $\alpha_{\mathrm{LER}} = 2.4$ and $\alpha_{\mathrm{HER}} = 1$, we next apply Eq. (4.10) to 30 second averages of TPC nuclear recoil counts recorded during single-beam LER and HER decay fills respectively. Figure 4.6 shows a comparison between measured recoil counts binned into 30 s intervals during the LER study period of Study A, and the corresponding LER beam-gas and LER Touschek fit predictions from Eqs. (4.8) and (4.9), respectively in the TPC located at $z_{\mathrm{BELLE}} = +14\,\mathrm{m}$. The fit model appears to reasonably fit single-beam LER backgrounds, producing a reduced χ^2 of 1.3. Though not explicitly
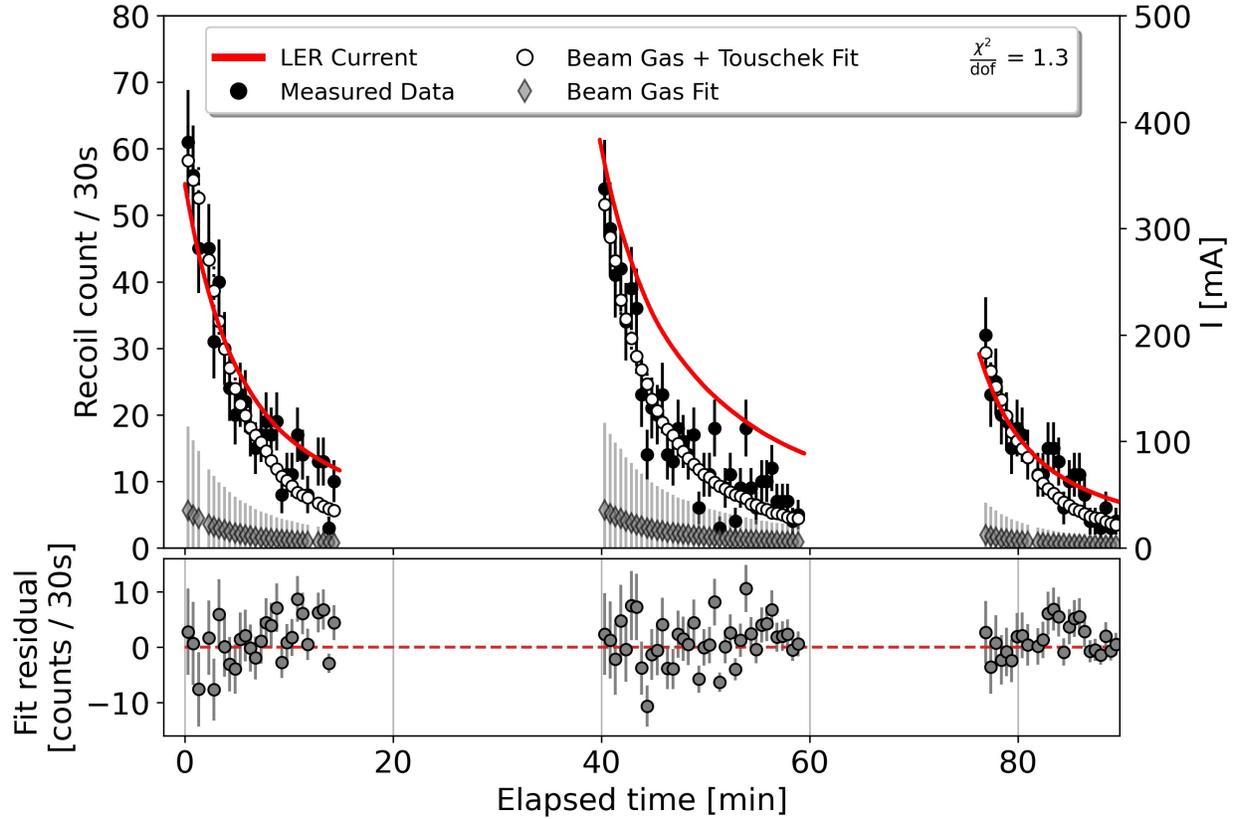
Figure 4.6: (color online) LER background composition fits versus time in the TPC located at $z_{\text{BELLE}} = +14\,\text{m}$ during Study A with $\alpha_{\text{LER}} = 2.4$. Top: black circles correspond to the measured recoil count, gray diamonds, and open circles represent the predicted recoil counts from LER beam gas and LER beam gas + LER Touschek backgrounds, respectively. Bottom: residual distribution defined as the difference between the measured recoil count and the fit-predicted recoil count using Eq. (4.10).

Study	Ring	(β_x^*, β_y^*) [mm,mm]	I [mA]	P [nPa]	(σ_y, σ_z) [μm, mm]	n_b	L [$10^{34}\text{cm}^{-2}\text{s}^{-1}$]
A	LER	(80,1)	510	30	(60, 5.9)	783	1.1
	HER	(60,1)	510	14	(35, 6.4)	783	
B	LER	(80,1)	730	35	(65, 5.8)	1174	2.5
	HER	(60,1)	650	14	(35, 6.2)	1174	

Table 4.4: Typical machine parameters during the luminosity background studies. All comparisons between data and simulation in this section assume these conditions.
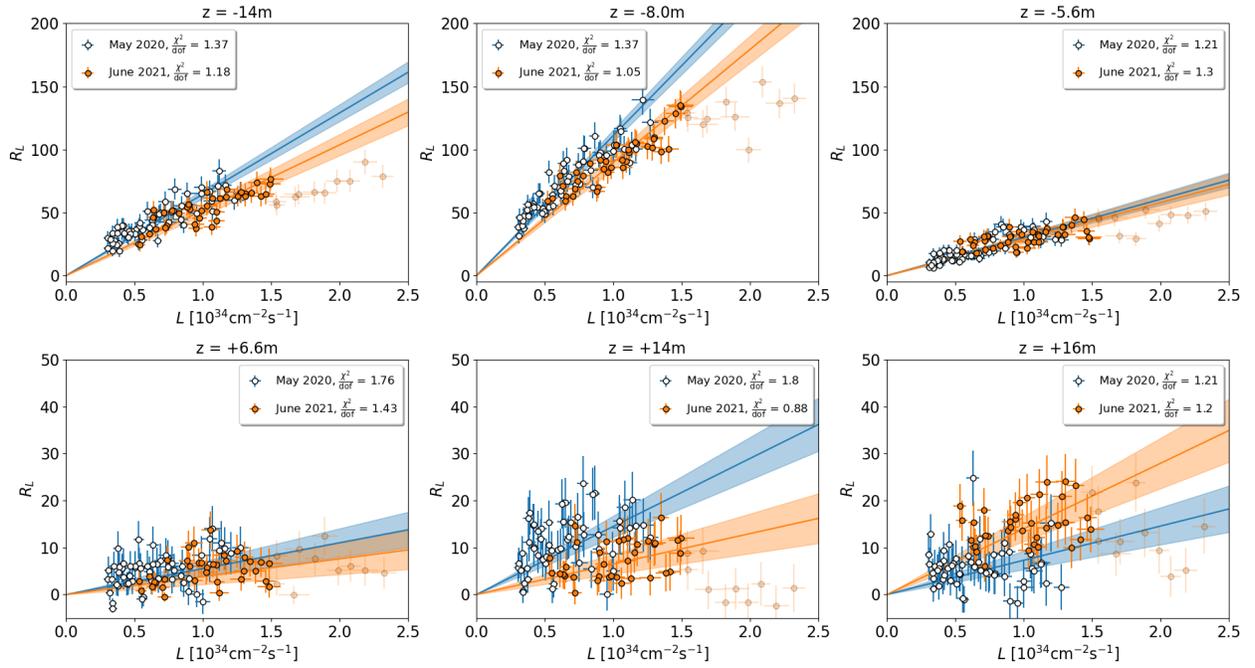
Figure 4.7: (color online) Fits to R_L versus luminosity during decay fills in each TPC. The solid blue and orange lines represent m_L for Studies A and B, respectively. The shaded regions represent statistical 95% confidence intervals on m_L for each study. Fits to m_L do not include contributions from the translucent points satisfying $L > 1.5 \times 10^{34}\,\mathrm{cm^{-2}s^{-1}}$ during Study B.

shown, we note that fits to LER and HER single-beam backgrounds yield reduced χ^2's ranging between 0.8 and 1.8 in all TPCs.

Luminosity background composition

After determining B_{0i}, B_{1i}, T_i, and α_i; $i =$ LER, HER, we use Eq. (4.12) to determine the luminosity sensitivity parameter m_L in each TPC. Figure 4.7 shows linear fits to R_L versus L in each of the six TPCs during the decay fills of the luminosity study for both Study A and Study B. Here R_L is the difference between the total number of recoils measured during each 30 second time bin during the luminosity study decay fills and the predicted number of recoils due to single-beam backgrounds during this same period. We obtain reasonable fits to R_L vs L in all TPCs during Study A. During Study B, we observe a drop in R_L at the highest luminosities (translucent points in Figure 4.7) in all TPCs. We speculate that this drop in R_L above $1.5 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ is related to the drop in effective gain at high luminosity described in Chapter 3.3.3. Indeed, a drop in effective gain would lower the average energy of events at high luminosity, resulting in fewer nuclear recoils above both the threshold of the chip and within our analysis selections. Furthermore, the two TPCs shown in Figure 4.7 with the least significant drops in R_L above $1.5 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ during Study B are $z = -5.6\,\mathrm{m}$ and $z = 6.6\,\mathrm{m}$, which, from Eqs. (4.1) and (4.2), and Table 4.3, are the two TPCs that did not require gain corrections. Since it is difficult to correct for all effects related to this drop in gain, we opt to exclude the translucent points when fitting for m_L.

Observed versus simulated background compositions

We next extrapolate the nuclear recoil counts for each background source observed in each TPC to machine conditions that are consistent with the decay fill periods during the luminosity study in Studies A and B. The top and bottom plots of Figure 4.8 show the extrapolated fractional background contributions at the conditions listed in Table 4.4 for the Study A and Study B luminosity study periods, respectively. Comparing the extrapolations at the conditions of Study B to the conditions of Study A, we observe that the luminosity background fraction increases in all TPCs except for the TPC at $z_{\mathrm{BELLE}} = 16\,\mathrm{m}$, where increases in LER Touschek backgrounds in Study B overshadow increases in luminosity-induced nuclear recoils at higher luminosity. Due to differences in beam optics settings that can have large effects on beam-induced background generation, directly comparing single-beam backgrounds between the two background study days is not always meaningful. For a given beam optics configuration, however, we can compare observed and predicted single-beam background rates. Comparing the unfilled data points in the top and bottom plots in Figure 4.9, we find that for the Study A beam optics configuration, observed single-beam recoil rates in the BWD tunnel are consistent with prediction, while simulation under-predicts single-beam recoil rates in the FWD tunnel.

Table 4.5 shows data/MC comparisons of nuclear recoil rates for all background sources of
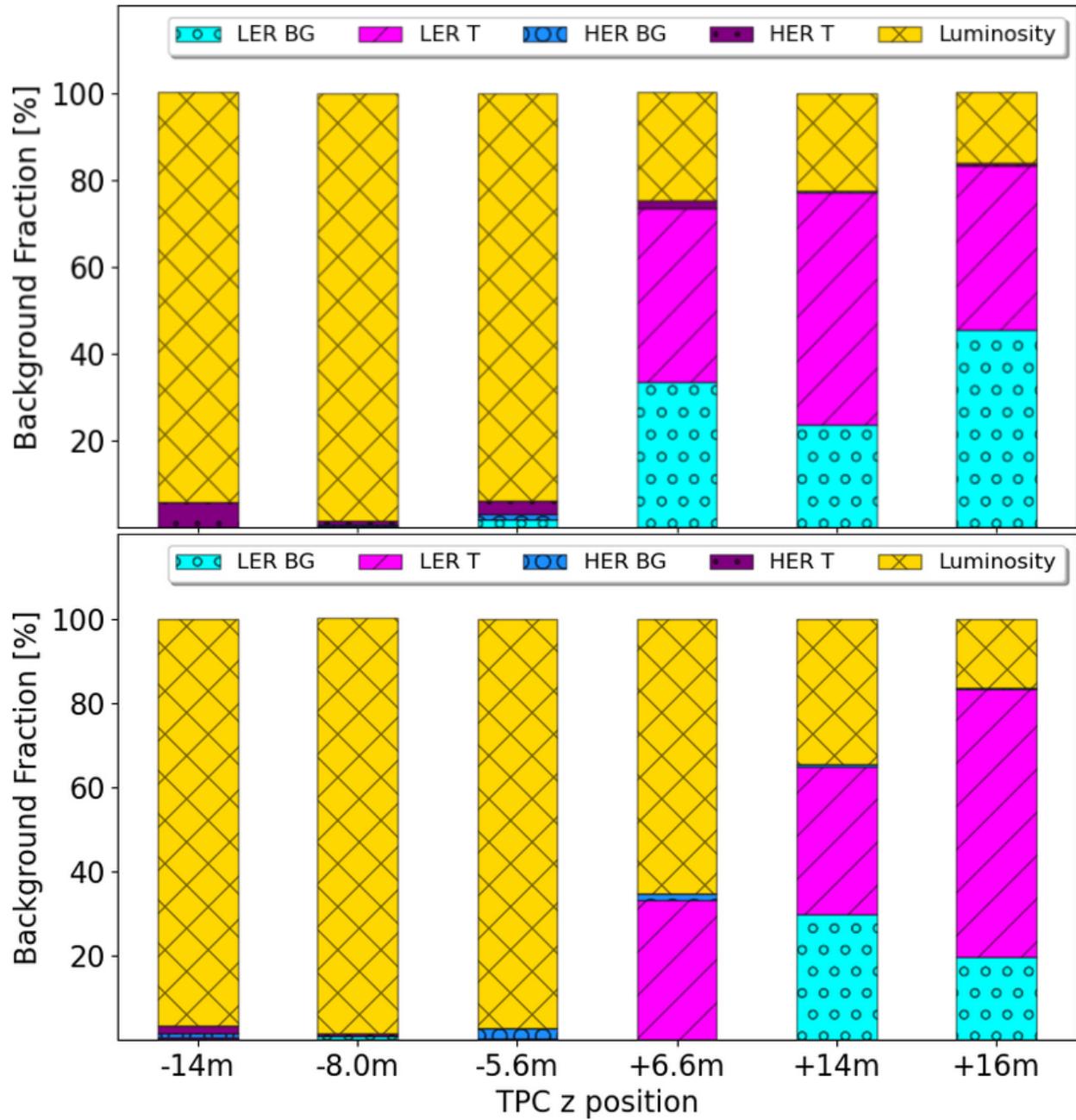
Figure 4.8: (color online) Background compositions determined from fits to storage fills extrapolated to the machine parameters shown in Table 4.4 for Study A (top) and Study B (bottom).
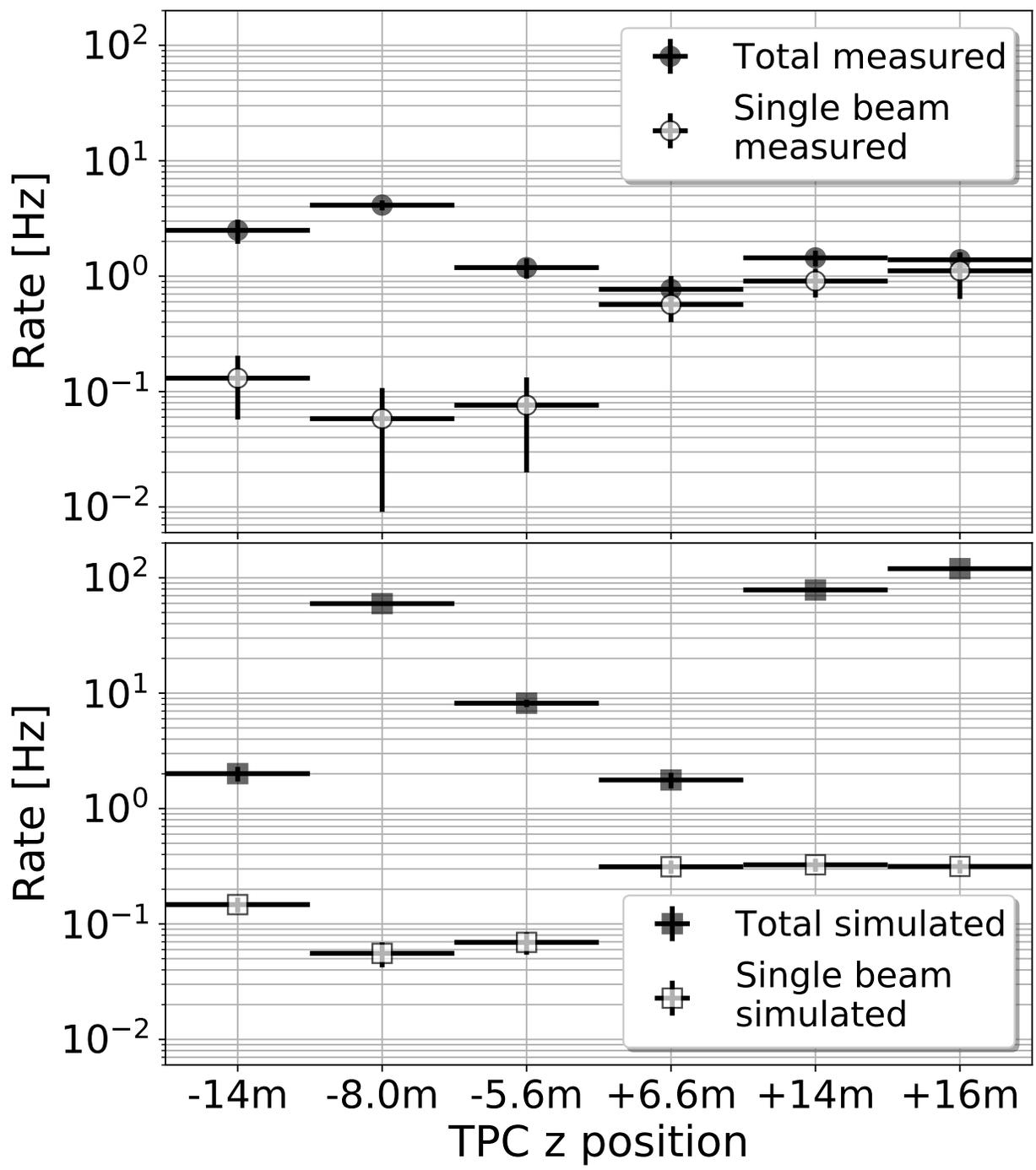
Figure 4.9: Comparison of single beam and total (single beam + luminosity) nuclear recoil rates extrapolated to the machine conditions listed in Table 4.4 for observed recoils during the luminosity decay fill periods of Study A (top) and simulated recoils (bottom).

TPC z[m]	LER Beam Gas	HER Beam Gas	LER Touschek	HER Touschek	$L_{\text{Study A}}$ $[^{+(\text{sys.})}_{-(\text{sys.})} \pm (\text{stat.})]$	$L_{\text{Study B}}$ $[^{+(\text{sys.})}_{-(\text{sys.})} \pm (\text{stat.})]$	Total Study A
-14	1.82 ± 5.2	0	—	0.90 ± 0.5	$1.27^{+0.03}_{-0.13} \pm 0.25$	$1.02^{+0.00}_{-0.43} \pm 0.18$	1.24 ± 0.22
-8.0	9.52 ± 17	$39.3 \pm 65.$	—	0.57 ± 0.6	$0.07^{+0.00}_{-0.00} \pm 0.01$	$0.06^{+0.00}_{-0.02} \pm 0.00$	0.07 ± 0.01
-5.6	22.9 ± 31	$16.0 \pm 24.$	—	0.56 ± 0.5	$0.14^{+0.00}_{-0.03} \pm 0.03$	$0.13^{+0.00}_{-0.05} \pm 0.02$	0.14 ± 0.03
+6.6	$186. \pm 80$	0	0.98 ± 0.4	4.61 ± 8.0	$0.14^{+0.13}_{-0.15} \pm 0.06$	$0.10^{+0.00}_{-0.10} \pm 0.04$	0.44 ± 0.11
+14	$471. \pm 240$	47.8 ± 140	1.92 ± 0.5	—	$(7^{+3}_{-5} \pm 2) \times 10^{-3}$	$(3^{+0.6}_{-3} \pm 1) \times 10^{-3}$	0.02 ± 0.00
+16	$480. \pm 260$	0	1.61 ± 0.5	—	$(2^{+3}_{-2} \pm 1) \times 10^{-3}$	$(4^{+0.0}_{-4} \pm 1) \times 10^{-3}$	0.01 ± 0.00

Table 4.5: Data/MC ratios scaled to the conditions shown in Table 4.4. Entries marked as 0 correspond to instances where Eq. (4.10) predicts no nuclear recoils. Entries marked as "—" indicate that no MC recoils were produced for the listed background component. Beam gas, Touschek, and total data/MC ratios were only computed for Study A. $L_{\text{Study A}}$ and $L_{\text{Study B}}$ include asymmetric systematic uncertainties accounting for potential misclassifications of single-beam backgrounds.

interest. Since our simulated samples are generated assuming beam optics settings consistent with Study A, we only include this study day for single-beam data/MC comparisons. We find very good agreement between data and MC Touschek backgrounds using $(\alpha_{\text{LER}}, \alpha_{\text{HER}}) = (2.4, 1)$ indicating that Touschek production mechanisms are modeled well in simulation. Furthermore, LER Touschek backgrounds appear to be the dominant single-beam background source in the FWD tunnel, which is consistent with the predictions of simulation. Beam gas background measurements, on the other hand, are underpredicted by simulation. We note that simulated statistics are in general low for beam gas backgrounds leading to large uncertainties in beam gas data/MC ratios.

The data/MC ratios for luminosity backgrounds include an additional systematic uncertainty meant to account for potential misclassification of single-beam backgrounds. These uncertainties on m_L are computed assuming 1σ uncertainty contributions to R_L from B_{0i}, B_{1i}, and T_i, that are truncated to ensure both single-beam and luminosity background rates do not drop below 0. In the absence of this additional systematic uncertainty, we find agreement of luminosity data/MC ratios between Study A and Study B within 1.4σ in all TPCs. Simulation, on the other hand, predicts significantly higher luminosity background rates than measurement in all TPCs except for the TPC at $z_{\text{BELLE}} = -14\,\text{m}$, as can be seen by comparing the filled data points in the top and bottom plots of Figure 4.9. The fact that the difference in observed luminosity recoil rates between Study A and Study B is much smaller than the difference between observed and simulated luminosity recoil rates may point to the known lack of materials such as magnet mounts, certain types of shielding, and other support structures, present in the simulated geometry description of the two accelerator tunnels, which the collaboration is currently working to improve.
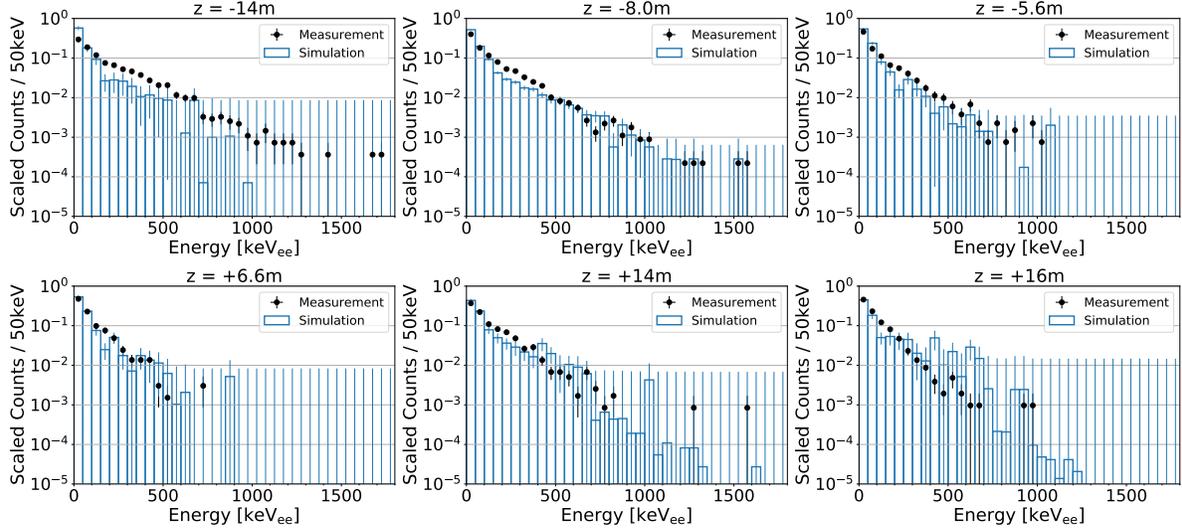
Figure 4.10: (color online) Comparison of observed (black points) and predicted (blue bars) recoil energy spectra in each TPC. The data shown are from the beam decay luminosity study periods from Study A. The lowest energy bin only includes events above the X-ray veto threshold listed in Table 4.3 in both measurement and simulation. Both distributions are normalized to an integral of unity. The title of each plot indicates the z location of its corresponding TPC.

4.5.2 Energy spectra

We next compare the energy spectra of measured nuclear recoils with their simulated counterparts. To perform this comparison, we first weight simulated recoil rates to provide a fair representation of the measured background composition during Study A. We apply the following steps:

1. Scale simulated recoils to their "one second" ($1\,\mathrm{s}$) equivalent rates. The simulated beam time used to generate MC samples varies considerably between background sources, so we apply a scaling factor of $1/t_{\mathrm{gen}}$ where t_{gen} is the simulated beam time, in seconds, shown in Table 4.2. Simulated luminosity recoils are further scaled down to the typical luminosity listed in Table 4.4.

2. Scale simulated rates of a given background type by their corresponding data/MC factor: due to differences in data/MC ratios with respect to background type, each simulated recoil is tagged with the associated background source from which it was generated (LER/HER beam gas, LER/HER Touschek, or luminosity). Data/MC ratios are computed for recoils satisfying the X-ray veto threshold listed in Table 4.3, and assuming machine parameters shown in Table 4.4.

3. Bin the samples by $E_{\mathrm{corrected}}$.

65

4. Normalize to a unit integral. Since we've already compared measured and simulated recoil rates (Table 4.5, Figure 4.9), we aim to compare the shapes of the measured and simulated nuclear recoil energy spectra.

Figure 4.10 shows the resulting histograms of measured and simulated recoil spectra. We use steps (1-3) above to assign weights to each simulated recoil.

We fill the jth bin of the simulated energy spectrum with the normalized sum of weights in the given bin,

$$S_j = \frac{\sum_i w_{ji}}{\mathcal{N}}, \tag{4.13}$$

where w_{ji} is the weight of the ith event in bin j, and $\mathcal{N} \equiv \sum_j \sum_i w_{ji}$ is the sum of all weights in the energy spectrum. We compute the uncertainty in the jth bin using the following procedure:

1. Compute the statistical uncertainty contribution, W_k, from background type k:

$$W_k = w_k \sqrt{N_k}, \tag{4.14}$$

where w_k is the weight associated with background type k and N_k is the number of events in the bin with background type k.

2. Compute the normalized uncertainty of the jth bin,

$$\sigma_j = \frac{\sqrt{\sum_k W_k^2}}{\mathcal{N}}. \tag{4.15}$$

The contents of the bins in the measured energy spectrum are unweighted. Due to measurements occurring over a substantially longer time frame than the equivalent beam time for all simulated recoil backgrounds, we include an additional "uncertainty floor" component where we assign the quadrature sum of the weight associated with each background type, $\sqrt{\sum_k w_k^2}$, as the uncertainty of each empty bin.

With the exception of the TPC furthest away from Belle II in the BWD tunnel ($z_{\text{BELLE}} = -14\,\text{m}$), we find that both the observed and simulated recoil spectra are approximately exponentially decaying with slopes in reasonable agreement, especially at energies below $200\,\text{keV}$, which constitute between roughly 80% and 90% of recoils measured during this period in these five TPCs. The agreement between measured and predicted energy spectra in these five TPCs suggests that the material interactions that lead to neutron production are modeled well in simulation out to $16\,\text{m}$ from Belle II in the FWD tunnel and $8.0\,\text{m}$ in the BWD tunnel.

4.5.3 Angular and directional analysis

Comparing the distributions of measured and simulated nuclear recoil angles θ_{TPC} and ϕ_{TPC} provides useful insight toward our understanding of cavern neutron production points in the absence of a full kinematic reconstruction of the distribution of *neutrons* incident upon a TPC. Though reconstructing incident neutron energies and angular distributions would be illuminating, doing so in a realistic detector with non-ideal performance is challenging and beyond the scope of this work. Agreement between measured and simulated recoil angles θ_{TPC} and ϕ_{TPC}, however, should only hold if there is agreement between the angular distributions of neutrons incident upon the TPCs, so we perform such comparisons as an indirect test of our modeling of angular distributions of neutrons in the two tunnels.

Axial directional performance: angular resolution

We use the principal axis of the 3D reconstructed ionization distribution determined using an SVD to determine the *axial* direction of a track. This "SVD fitter" performs well for long, higher energy tracks, where the principal axis is relatively unambiguous, but for sufficiently short tracks, the reconstructed 3D track is shaped like a round bowl, making the principal axis assignment of a track ambiguous. We quantify the axial angular resolution of the TPCs by computing the angular mismeasurement (difference in angle) between the axial direction of reconstructed track using the SVD fitter and the truth MC-simulated axial direction of the recoil.
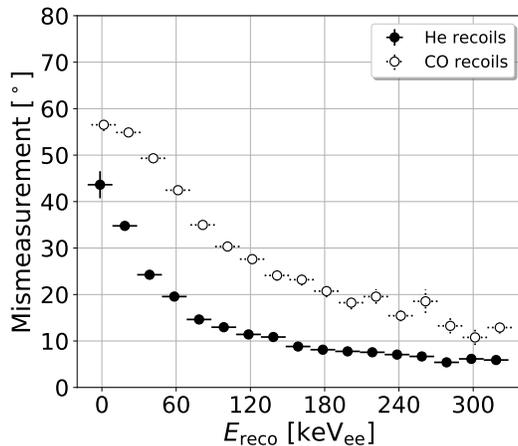
Figure 4.11: Mean (data points) and standard error (vertical error bars) of angular mismeasurement between SVD fits to the principal axis of the reconstructed simulated track and the truth MC-simulated direction of the recoil split up between ^{4}He recoils (filled points) and ^{12}C/^{16}O recoils (unfilled points). The angular resolution of ^{4}He recoils is considerably better than ^{12}C and ^{16}O recoils.

Figure 4.11 shows histograms of angular resolution versus E_{reco} for truth-matched ^{4}He and

^{12}C/^{16}O recoils. We find that the angular resolution of simulated ^{4}He recoils is considerably better than the angular resolution of ^{12}C and ^{16}O recoils. This finding is consistent with the fact that ^{12}C and ^{16}O recoils tend to produce lower energy and shorter tracks than He recoils, thus increasing the ambiguity of the identified principal axis in a measured track. During Phase 1, angular resolution was found to be within 20° for He recoils down to $100\,\mathrm{keV_{ee}}$ [48], whereas here we find angular resolution to be within 20° for He recoils down to $60\,\mathrm{keV_{ee}}$. We further find angular resolutions of $13.0° \pm 0.4°$ in the $100\,\mathrm{keV_{ee}}$ bin, indicating substantially improved angular resolution over the Phase 1 results. In the Phase 1 analyses, the principal axis of each recoil track was identified using a χ^2 minimization technique as opposed to the SVD used here. While we have not explicitly compared angular resolution performance of these two track-reconstruction techniques, we suspect that use of the SVD primarily contributes to the observed improvement in angular resolution.

Vector directionality

The principal axis of a recoil track assigns the axial direction of the track, however without a *vector* direction assigned to that principal axis, measurements of ϕ_{TPC} and θ_{TPC} are ambiguous. Here we outline a procedure for vector head-tail assignment to recoil tracks and assess the performance of this procedure in assigning vector directions to tracks.

Nuclear recoils measured in a TPC are said to be "beyond the Bragg Peak" [81], meaning that the stopping power of an event falls sharply at the stopping end of the track, leading to an expected asymmetric distribution of charge along the principal axis of the track. Previous work with these TPCs during Phase 1 [47; 48] has shown that this asymmetry is much more clear for ^{4}He recoils than for ^{12}C and ^{16}O recoils, so we only attempt to assign vector directions to ^{4}He recoil tracks. In particular, we restrict all remaining angular analyses to ^{4}He recoils with $E_{\mathrm{reco}} > 40\,\mathrm{keV_{ee}}$ in simulation and $E_{\mathrm{corrected}} > 40\,\mathrm{keV_{ee}}$ in measurement, as simulation predicts average axial angular resolutions to be within 8° for events satisfying this criteria.

Given that the He and C/O recoil bands differ in $\mathrm{d}E/\mathrm{d}x$ and that these $\mathrm{d}E/\mathrm{d}x$ bands appear to vary quadratically at energies below $100\,\mathrm{keV}$ (Figure 4.3), we make selections using E/ℓ^2 versus ℓ distributions for recoils above $40\,\mathrm{keV_{ee}}$, so that we can define a flat E/ℓ^2 versus ℓ selection boundary for He recoils. Figures 4.12 and 4.13 show the simulated and measured distribution of such recoils plotted in this space. From Figure 4.12, we see that there is clear separation between truth He and truth C/O recoils, and thus define the events within the red rectangular boundary to be He recoils. Looking at Figure 4.13, we find that the boundaries between C/O recoils and He recoils appear even more distinct, but since $\mathrm{d}E/\mathrm{d}x$ differ between measurement and simulation in some TPCs, we define the red boxes, within which we select He recoils, by eye. Figures 4.14 and 4.15 show the events identified as He and C/O recoils in E versus ℓ space.

Charge integration effects in these TPCs are known to bias the measured charge asymmetry in a recoil away from its true ionization distribution when the recoil track is inclined with respect
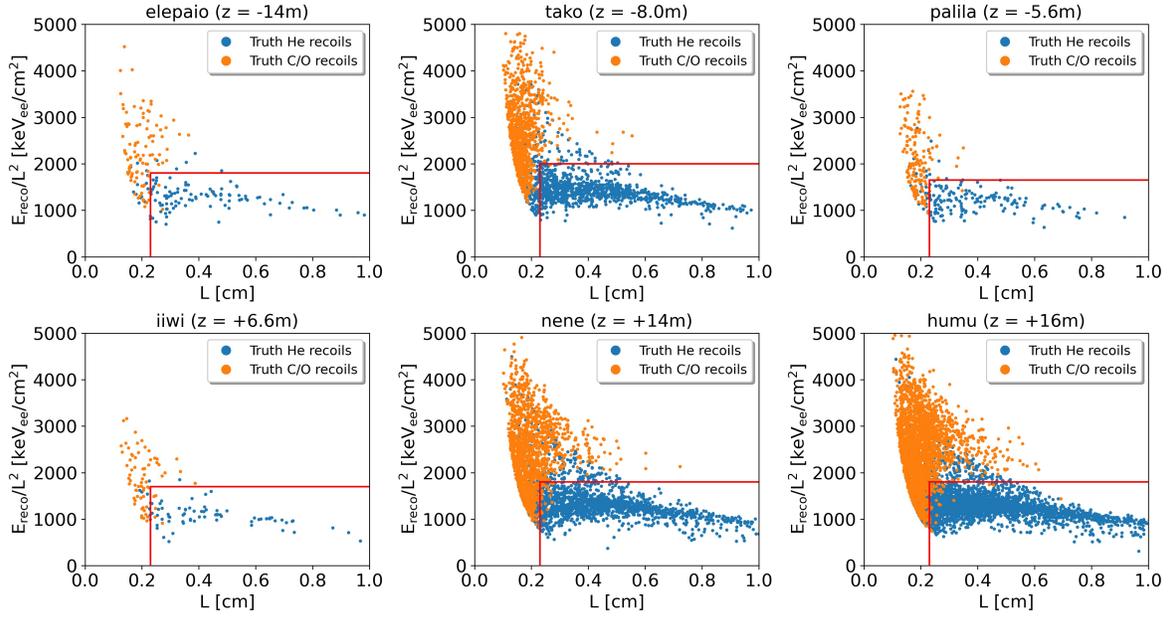
Figure 4.12: E/ℓ^2 versus ℓ with truth simulated He recoils in blue and truth simulated C/O recoils in orange. Events within the red rectangle in the lower right hand corner of each plot are selected as He recoils.
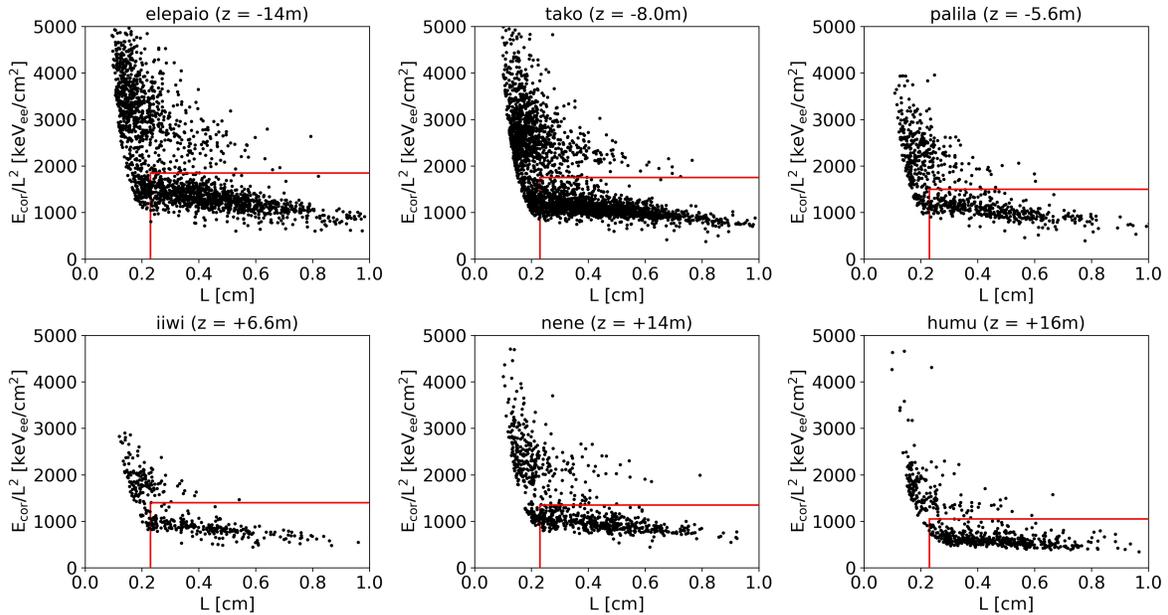
Figure 4.13: Measured E/ℓ^2 versus ℓ. Events within the red rectangle in the lower right hand corner of each plot are selected as He recoils.
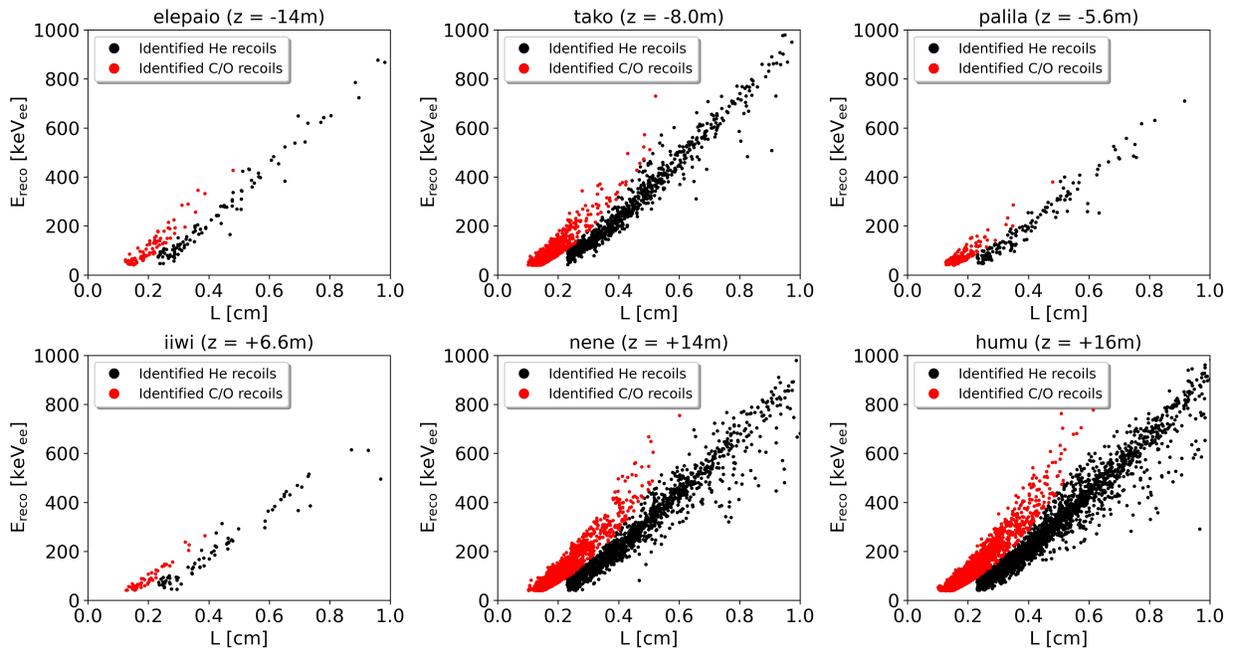
Figure 4.14: Selections from Figure 4.12 plotted as an E versus ℓ distribution.
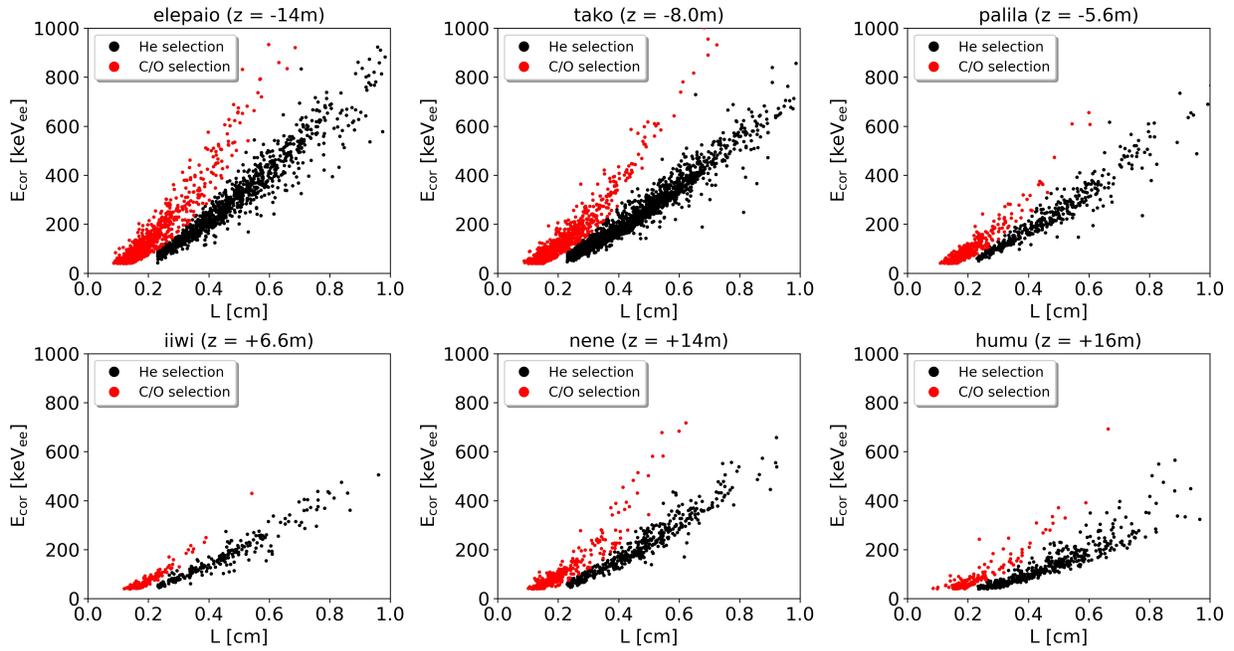
Figure 4.15: Selections from Figure 4.13 plotted as an E versus ℓ distribution.

70
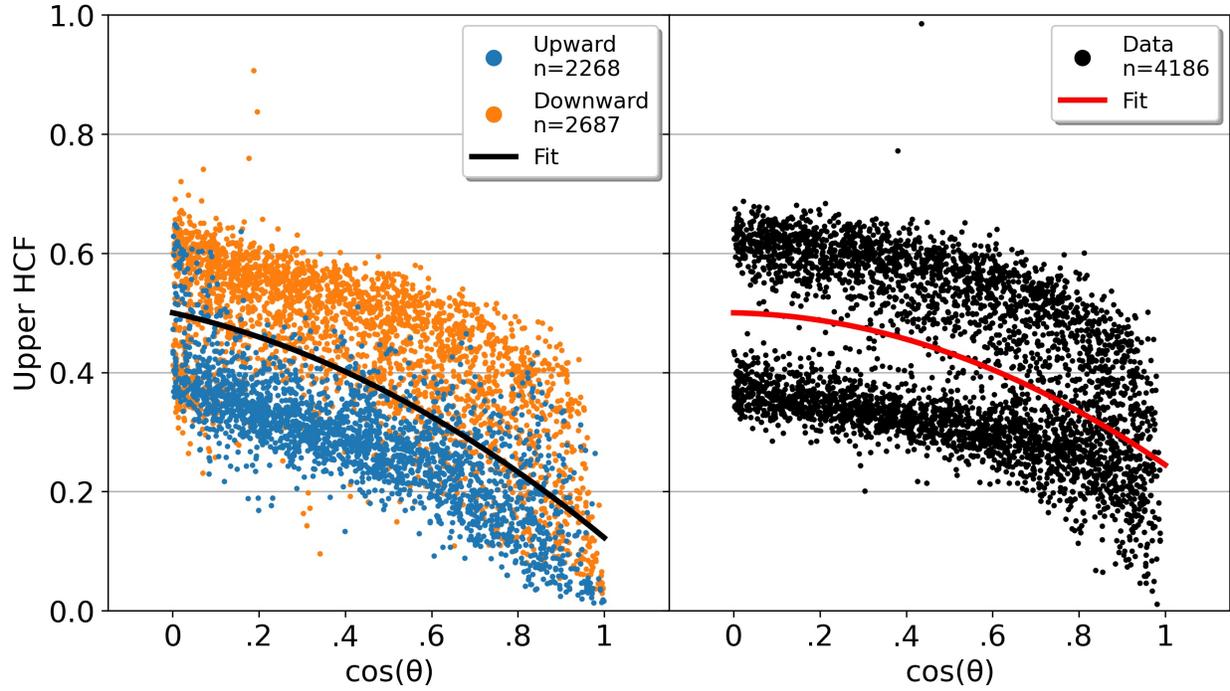
Figure 4.16: (color online) Fractional charge of ^{4}He recoils ($E_{\mathrm{reco}} > 40\,\mathrm{keV_{ee}}$) on the half of the track with larger z_{TPC} ("Upper HCF") in simulation (left) and measurement (right) as a function of the inclination of the track. In both plots we find that the upper HCF tends to decrease sharply as tracks become more inclined in z. In simulation we observe strong separation in Upper HCF between upward pointing tracks ($\cos(\theta_{\mathrm{TPC,\ truth}}) > 0$) and downward pointing tracks ($\cos(\theta_{\mathrm{TPC,\ truth}}) < 0$). The black and red fit lines in the left and right plots are used as boundaries to assign vector directions to the samples of measured and simulated nuclear recoils, respectively.

to the readout plane [82]. To investigate this bias, we split each measured and simulated track in half along the midpoint of its principal axis and count up the charge on either end. We make an *a priori* choice to initially define the head of the track to be the side of the track with larger average z_{TPC} (Figure 2.4). Figure 4.16 shows the fraction of charge on the half of each track containing the vector head (henceforth called Upper Head Charge Fraction, or Upper HCF) in all six TPCs as a function of axial track inclination[2]. We observe a steep drop in Upper HCF with increasing axial inclination in both measurement and simulation. Furthermore, two distinct Upper HCF bands arise in measured and simulated samples that are especially apparent in less inclined tracks. Simulation suggests that the band with larger (smaller) Upper HCF corresponds primarily to downward (upward) facing tracks, where downward and upward MC tracks satisfy $\cos(\theta_{\mathrm{TPC, \, truth}}) < 0$ and $\cos(\theta_{\mathrm{TPC, \, truth}}) > 0$, respectively. We use this separation between upward and downward facing tracks to implement our final vector head-tail assignments. The separation between these two bands appears to be stronger in measurement than in simulation, so we apply a data-driven fit boundary to implement our head-tail assignments in measurement.

The black and red fit boundaries shown for simulation and measurement, respectively in Figure 4.16 are determined by fitting quadratic polynomials, $p_{2, \, \mathrm{MC}}(\cos(\theta_{\mathrm{TPC}}))$ and $p_{2, \, \mathrm{data}}(\cos(\theta_{\mathrm{TPC}}))$ to the distribution of Upper HCF vs $\cos(\theta_{\mathrm{TPC}})$ of all ^{4}He recoils above $40\,\mathrm{keV_{ee}}$. In the limit of a perfectly flat track, we assume no charge integration bias in Upper HCF, so we force $p_{2, \, \mathrm{MC}}(0) = p_{2, \, \mathrm{data}}(0) = 0.5$. Using the separation between truth MC-simulated upward and downward facing tracks on either side of $p_{2, \, \mathrm{MC}}(\cos(\theta_{\mathrm{TPC}}))$, we form a *vector direction assignment hypothesis* that states that recoils with Upper HCF $> p_{2, \, \mathrm{MC, \, data}}(\cos(\theta_{\mathrm{TPC}}))$ are downward facing, or equivalently, recoils with correct vector assignment have head charge fractions (HCF) less than $p_{2, \, \mathrm{MC, \, data}}(\cos(\theta_{\mathrm{TPC}}))$. We can thus perform our final head-tail assignments by "flipping" the vector direction (switching the vector head and tail position) of all tracks with Upper HCF $> p_{2, \, \mathrm{MC, \, data}}(\cos(\theta_{\mathrm{TPC}}))$ to make them downward facing, thereby satisfying HCF $< p_{2, \, \mathrm{MC, \, data}}(\cos(\theta_{\mathrm{TPC}}))$. Figure 4.17 shows the percentage of correctly assigned vector directions of simulated tracks verses the assigned *axial* angle projections, $\phi_{\mathrm{Axial,TPC}}$ and $|\cos(\theta_{\mathrm{TPC}})|$, and reconstructed ionization energy E_{reco}. We find our assignment hypothesis of HCF $< p_{2, \, \mathrm{MC}}(\cos(\theta_{\mathrm{TPC}}))$ (filled points in Figure 4.17) leads to a significant performance improvement in vector direction assignment over the base assignment hypothesis of HCF < 0.5 (unfilled points in Figure 4.17) which would be expected to hold in the absence of the observed charge asymmetry bias with track inclination.

After assigning vector directions to all tracks, we then compute new *vector* angles θ'_{TPC} and ϕ'_{TPC}, which are the angles of the track vector after final head-tail assignments. We shift the ϕ'_{TPC} domain to range from 0 to 360 degrees so that $180°$ is the average direction expected for recoils caused by neutrons originating from the beam pipe. Moving forward we drop the $'$ designation of both of these recoil angles. Over our entire simulated sample of ^{4}He recoils in all six TPCs satisfying

---

[2]With the aforementioned initial head assignment, $0 \leq \cos(\theta_{\mathrm{TPC}}) \leq 1$ so we can call $\cos(\theta_{\mathrm{TPC}})$ the axial track inclination.
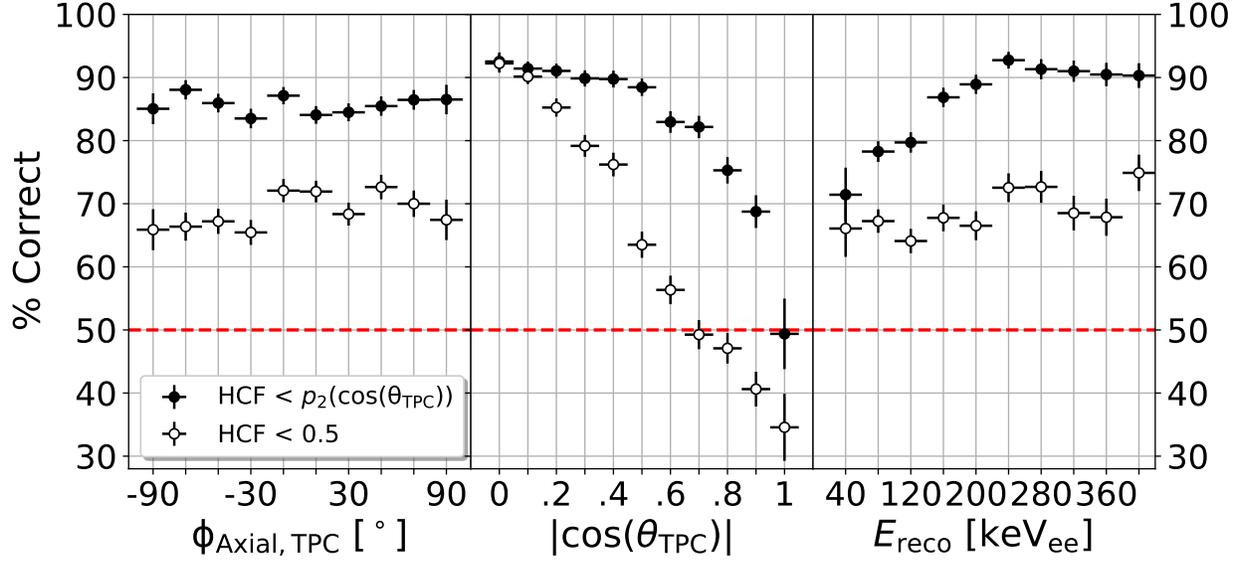
Figure 4.17: From left to right: percentage of simulated ^{4}He recoil tracks ($E_{\text{reco}} > 40\,\text{keV}_{\text{ee}}$) with assigned vector directions that match the truth MC-simulated direction of the recoil as functions of axial angle projections $\phi_{\text{Axial, TPC}}$ and $|\cos(\theta_{\text{TPC}})|$, and reconstructed ionization energy, E_{reco}, respectively. The filled [unfilled] points show the result using the assignment hypothesis of HCF $<$ $p_{2,\,\text{MC}}(\cos(\theta_{\text{TPC}}))$ [HCF $<$ 0.5].

$E_{\text{reco}} > 40\,\text{keV}_{\text{ee}}$ we find $\sim 86\%$ of recoils satisfying HCF $< p_{2,\,\text{MC}}(\cos(\theta_{\text{TPC}}))$ have vector head assignments consistent with the true simulated direction of the recoil. If we apply an additional $90° < \phi_{\text{TPC}} < 270°$ restriction to only include recoil events with origins tending toward the beam pipe, then our percentage of correctly assigned vector directions increases to $\sim 91\%$.

Though we estimated our angular reconstruction performance using simulation, we argue that these estimates are reliable for measurement as well. Refs. [47; 48] show consistency in axial angular resolution between measurement and simulation, finding that the angular mismeasurement between "half-tracks", that is, tracks split in half along their principal axis, agrees to within 6° between observed and simulated samples over all energies. Additionally, Ref. [48] shows consistency between the detected and simulated charge of recoil tracks as a function of the length along their principal axis ($\mathrm{d}Q/\mathrm{d}x$), indicating that the input quantities used to determine vector directional assignment are accurately simulated. Similarly, we find agreement in the shapes of $\mathrm{d}E/\mathrm{d}x$ distributions between measured and simulated recoils (Figure 4.3(i)) which supports that vector direction assignment is modeled accurately in simulation. Finally, as we'll see shortly, the agreement between observed and predicted angular distributions provides further evidence that our angular reconstruction performance is reliable for measured recoil tracks.
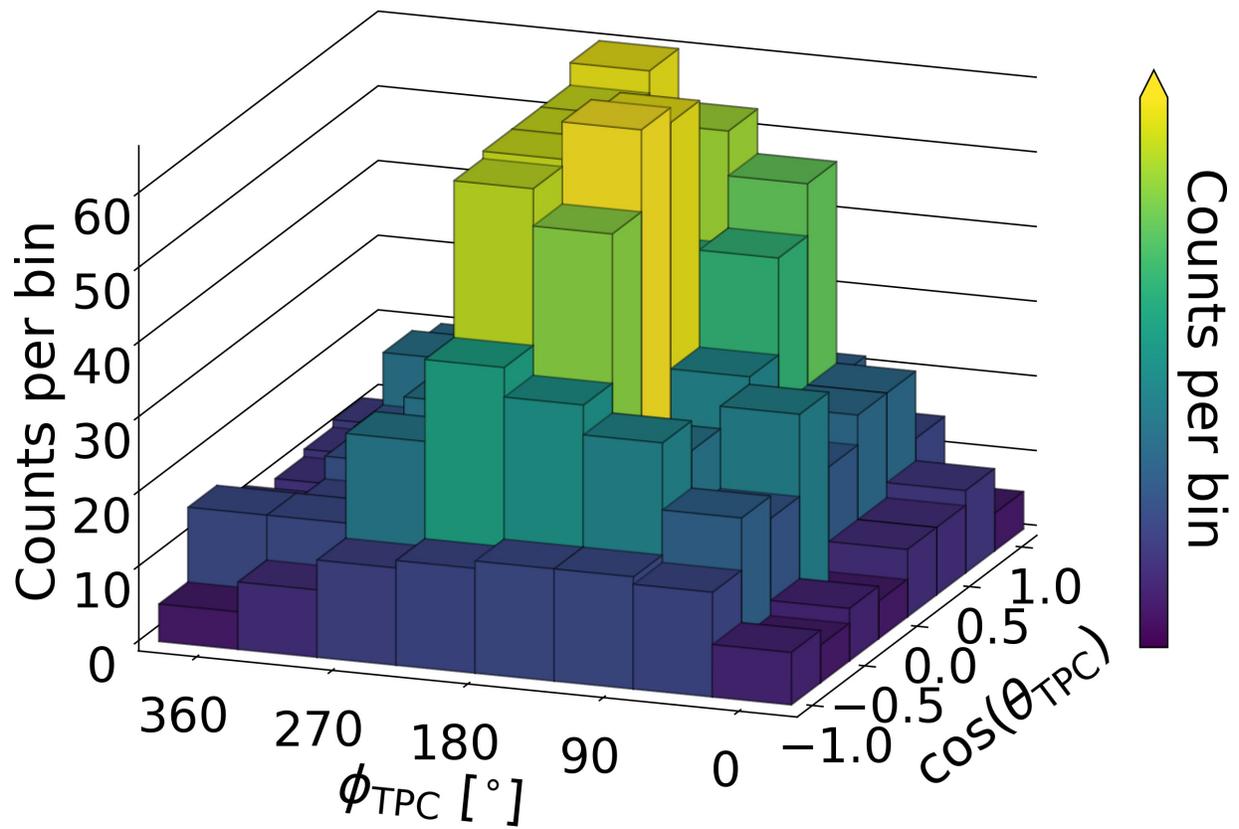
Figure 4.18: (color online) Binned ϕ_{TPC} versus $\cos(\theta_{\mathrm{TPC}})$ distribution of measured ^{4}He recoils ($E_{\mathrm{corrected}} > 40\,\mathrm{keV_{ee}}$) after final head-tail assignment in the TPC located at $z_{\mathrm{BELLE}} = -8.0\,\mathrm{m}$.
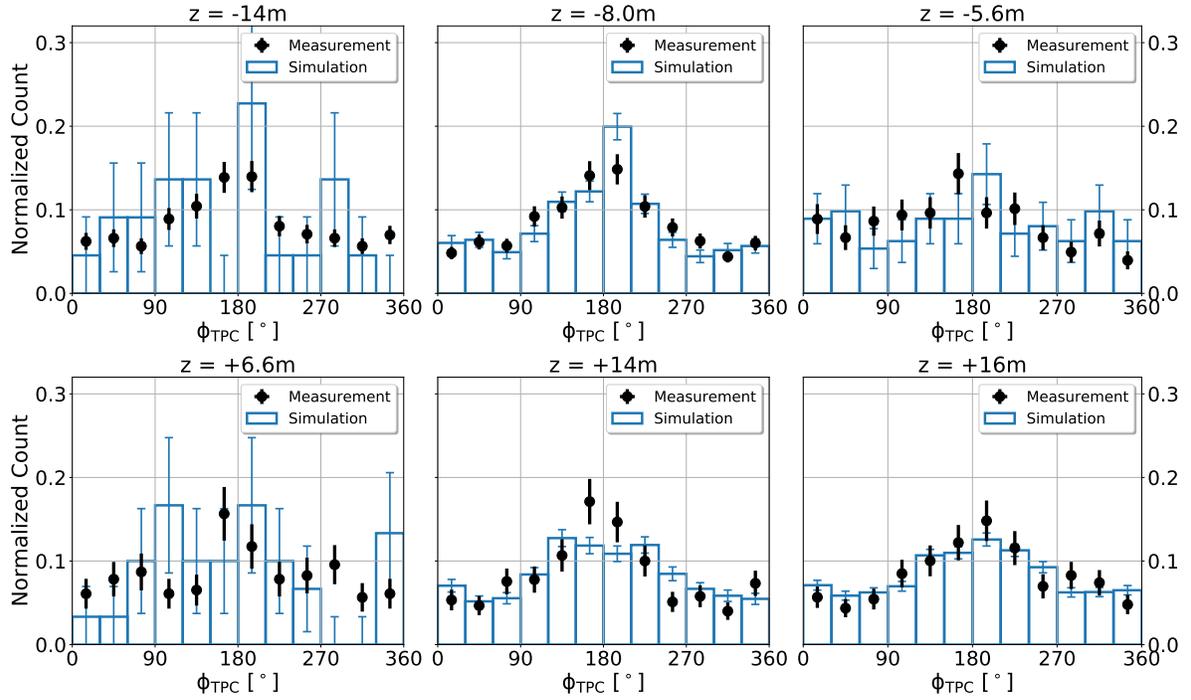
Figure 4.19: (color online) Comparison of ϕ_{TPC} between measurement (black points) and simulated luminosity background nuclear recoils (blue bars) for events satisfying $E_{\mathrm{ionization}} > 40\,\mathrm{keV_{ee}}$ during the luminosity decay fills of Study A. Histograms are constructed after applying final directional head-tail assignments.
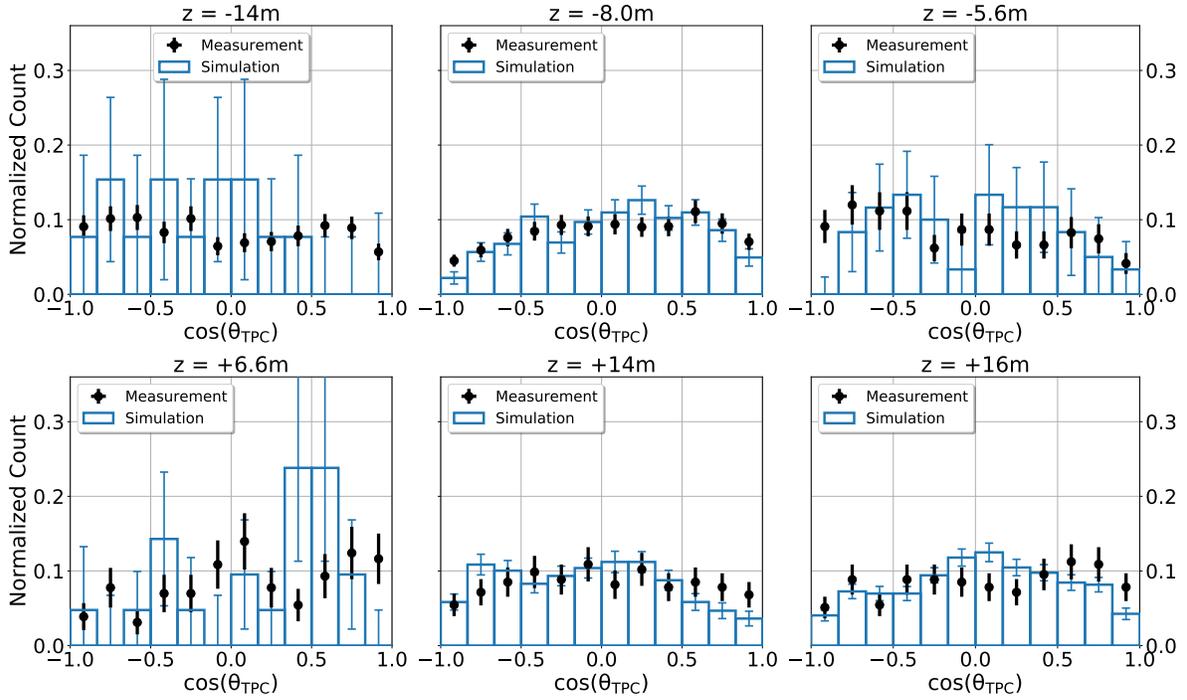
Figure 4.20: (color online) Comparison of $\cos(\theta_{\mathrm{TPC}})$ between measurement (black points) and simulated luminosity background nuclear recoils (blue bars) for events satisfying $E_{\mathrm{ionization}} > 40\,\mathrm{keV_{ee}}$ and $90° < \phi_{\mathrm{TPC}} < 270°$ during the luminosity decay fills of Study A. Histograms are constructed after applying final directional head-tail assignments.

Data versus MC comparisons of ϕ_{TPC} and θ_{TPC}

Figure 4.18 shows a 2D histogram of $\cos(\theta_{\mathrm{TPC}})$ and ϕ_{TPC} for measured ^{4}He recoils after vector head-tail assignment satisfying $E_{\mathrm{corrected}} > 40\,\mathrm{keV_{ee}}$ from the decay fills of the Study A luminosity study period in the TPC located at $z_{\mathrm{BELLE}} = -8.0\,\mathrm{m}$. We observe that this distribution has strong peaking along $90° < \phi_{\mathrm{TPC}} < 270°$, consistent with the expectation that the majority of recoil events point back to the beam pipe. The shape of the $\cos(\theta_{\mathrm{TPC}})$ distributions for these events coming from the beam pipe seems to also indicate that the majority of recoils point back to a cone along the beam pipe centered directly within the straight line of sight between this TPC and the beam pipe, which is qualitatively consistent with the radiative Bhabha neutron hotspot location predicted by simulation (Figure 4.1). With the vector directional assignment performance we've established, we can now indirectly test these claims by comparing ϕ_{TPC} and $\cos(\theta_{\mathrm{TPC}})$ between measurement and simulation.

Figures 4.19 and 4.20 show the resulting simulated ϕ_{TPC} and $\cos(\theta_{\mathrm{TPC}})$ distributions plotted with measurement collected during the decay fills of the Study A luminosity study period. The histograms are normalized to an integral of unity and we report uncertainties in the recoil counts for both measurement and simulation. We also assign an uncertainty of one event (before normalization) in each empty bin. The $\cos(\theta_{\mathrm{TPC}})$ distributions include an additional $90° < \phi_{\mathrm{TPC}} < 270°$ restriction to remove events less likely to originate near the beam pipe, as neutrons originating elsewhere are not expected to be modeled well in simulation. Restricting our angular analysis to ^{4}He recoils above $40\,\mathrm{keV_{ee}}$ reduces the sample size of nuclear recoils substantially compared to the samples used for the energy spectra in Chapter 4.5.2. To work around this limitation, we use only the luminosity background for the simulation data in Figures 4.19 and 4.20, as only the shape of the luminosity background angular distributions are well defined in MC after selecting directional recoils. We expect this to work well for the BWD TPCs, where luminosity background dominates, but not necessarily for the FWD TPCs.

We perform χ^2 hypothesis tests [83] comparing the normalized measured and simulated ϕ_{TPC} and $\cos(\theta_{\mathrm{TPC}})$ distributions shown in Figures 4.19 and 4.20. In particular, we test the null hypothesis that the simulated luminosity background angular distribution explains our observed angular distributions. We test this hypothesis separately for ϕ and $\cos(\theta)$ and we reject the null hypothesis if the p-value associated with the χ^2 test statistic is less than 0.05. In Table 4.6, p_ϕ, and $p_{\cos(\theta)}$, are the p-values associated with χ^2 tests for the ϕ_{TPC} and $\cos(\theta_{\mathrm{TPC}})$ distributions, respectively.

The results of these χ^2 tests suggest that at the current low level of statistics, the luminosity background alone can explain the observed distributions of recoils resulting from neutrons incident from the beam pipe. This is as expected for the BWD TPCs, where these backgrounds dominate. In the FWD TPCs, other background components are sizable, so we expect the inclusion of other backgrounds will be required to model angular distributions with larger statistics.

TPC	p_ϕ	$p_{\cos\theta}$
$z = -14\,\mathrm{m}$	0.25	0.95
$z = -8.0\,\mathrm{m}$	0.29	0.37
$z = -5.6\,\mathrm{m}$	0.64	0.37
$z = +6.6\,\mathrm{m}$	0.26	0.36
$z = +14\,\mathrm{m}$	0.05	0.38
$z = +16\,\mathrm{m}$	0.48	0.11

Table 4.6: Summary of p-values resulting from the χ^2 hypothesis tests.

TPC	Luminosity data/MC	Scaled TPC 1 MeV Equivalent Flux $[10^8/\mathrm{cm}^2/\mathrm{year}]$	Raw EKLM Flux $[10^8/\mathrm{cm}^2/\mathrm{year}]$	Scaled EKLM 1 MeV Equivalent Flux $[10^8/\mathrm{cm}^2/\mathrm{year}]$
$z = -14\,\mathrm{m}$	$1.27^{+0.03}_{-0.13} \pm 0.25^*$	1180 ± 230		
$z = -8.0\,\mathrm{m}$	$0.07^{+0.00}_{-0.00} \pm 0.01$	654 ± 62	3.8	4.8
$z = -5.6\,\mathrm{m}$	$0.14^{+0.00}_{-0.03} \pm 0.03$	367 ± 69		
$z = +6.6\,\mathrm{m}$	$0.14^{+0.13}_{-0.15} \pm 0.06^*$	101 ± 45		
$z = +14\,\mathrm{m}$	$(7^{+3}_{-5} \pm 2) \times 10^{-3}$	76 ± 19	148	20
$z = +16\,\mathrm{m}$	$(4^{+0}_{-4} \pm 1) \times 10^{-3}$	75 ± 13		

Table 4.7: Predicted luminosity neutron fluxes over one Snowmass year ($1 \times 10^7\,\mathrm{s}$) in each TPC and in the outermost KLM end cap layers in the FWD and BWD tunnels, scaled up to SuperKEKB's target luminosity of $6.3 \times 10^{35}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The Raw EKLM Flux column shows expected annual neutron flux in these outermost KLM end cap layers without any measurement-informed scalings. The rightmost column shows these Raw EKLM flux estimates scaled by the corresponding highest TPC data/MC ratio in each tunnel (ratios are starred in the table).

4.5.4 Neutron flux extrapolations at SuperKEKB's target luminosity

We close this section off with estimates of *neutron* fluxes, including those that don't produce nuclear recoils, in the tunnel regions surrounding Belle II at SuperKEKB's target luminosity of $6.3 \times 10^{35}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. We perform these extrapolations only for luminosity backgrounds under the assumption that luminosity backgrounds are independent of beam optics settings and scale proportionately to luminosity. In the interest of extrapolating neutron fluxes in the most pessimistic scenario, we choose to perform our extrapolations assuming the larger of the two data/MC ratios computed in each TPC (Table 4.5).

To estimate the measured neutron flux incident on each TPC from luminosity backgrounds, we simply take the total number of luminosity-induced simulated neutrons, multiply this number by the larger of the two measured data/MC ratios for that TPC, and scale it accordingly by the relevant TPC dimensions and the simulated equivalent beam-time. We then convert this into a

1 MeV equivalent Non-Ionizing Energy Loss (NIEL) damage-weighted flux.

We additionally scale the MC-predicted 1 MeV equivalent NIEL damage-weighted neutron flux in the outermost BWD and FWD KLM end cap layers by the largest data/MC ratio of the TPCs in the BWD and FWD tunnels, respectively, to provide an upper limit estimate of the neutron flux reaching Belle II. Table 4.7 shows both the raw predicted neutron fluxes in the outermost KLM end cap layers and the scaled prediction using the largest TPC data/MC fraction in each tunnel. We note that if shielding is added between the TPCs and the KLM end cap layers, the TPC data/MC fractions should remain constant, while the simulation-predicted neutron fluxes in the outermost KLM end caps will change, so Table 4.7 can still be used to predict upper limit neutron fluxes in the outermost KLM end cap layer after material description updates to simulation.

The annual upper limit neutron tolerance of the most neutron-sensitive Belle II subdetectors is $\mathcal{O}(10^{11})$ neutrons/cm^2/year [32], suggesting that all Belle II detectors are safe from luminosity-induced neutron backgrounds in the tunnel. Even so, simulation predicts that up to 95% of luminosity-induced neutrons are produced within the RBB hotspot regions shown in green in Figure 4.1, so given the localized nature of neutron production during collisions, additional shielding around these RBB hotspots could be useful as a further safeguard for detector longevity.

4.6 Summary and conclusions of beam-induced fast neutron backgrounds at SuperKEKB

The Phase 3 TPC system led to the first directional measurements of fast neutron backgrounds in the tunnel region surrounding Belle II at SuperKEKB. Using an expanded simulation suite that models the SuperKEKB-Belle II geometry out to $|z_{\text{BELLE}}| < 29\,\text{m}$, we provided direct comparisons between measured and simulated neutrons in these tunnels surrounding Belle II. Comparing observed and simulated Touschek rates, we find agreement within a factor of 2 between data and simulation in all TPCs, indicating that Touschek production is modeled well in simulation. Beam-gas backgrounds have much larger discrepancies, most notably LER beam-gas backgrounds in the FWD tunnel where measured rates exceeded predictions of simulation by factors of up to 500. When including contributions of neutrons generated from collisions, we find agreement between total observed and predicted nuclear recoil rates within a factor of roughly $\mathcal{O}(10)$ in all TPCs except for those at $z = +14\,\text{m}$ and $z = +16\,\text{m}$ where simulation greatly overestimates luminosity-dependent neutron production. We note that further improvements to both the geometry and material description of collimator heads, magnets, shielding, and other components in the $|z_{\text{BELLE}}| < 29\,\text{m}$ tunnel region that have recently been implemented in Geant4 are not included in this work and could lead to improvements in the agreement between measured and simulated background rates.

Comparing angular distributions of recoils, we conclude that the majority of observed and simulated neutrons in the SuperKEKB tunnel near Belle II originate near the beam pipe and further

find reasonable agreement between measured and simulated $\cos(\theta_{\mathrm{TPC}})$ distributions. Given that the simulated angular distributions only included luminosity backgrounds, and these backgrounds overwhelmingly dominate the measured recoil rates in the BWD tunnel, we suggest that these TPCs are sensitive to the predicted RBB hotspots.

When comparing the shapes of nuclear recoil energy spectra, in all except for the TPC furthest away from Belle II in the BWD tunnel, we find similar broadness between measured and simulated energy spectra. This indicates that the beam pipe and magnet material descriptions are modeled well in simulation out to at least $-8.0\,\mathrm{m}$ in the BWD tunnel and out to $+16\,\mathrm{m}$ in the FWD tunnel.

From the BEAST Phase 2 and Phase 3 TPC exercises, we have learned that fast neutron Touschek backgrounds are expected to dominate near the QCS final focus, while luminosity fast neutron backgrounds overwhelmingly dominate in the BWD tunnel and are expected to dominate in the FWD tunnel as luminosity increases. Since Touschek backgrounds are heavily dependent on beam optics and collimator tuning, continued careful tuning is needed to maintain reasonable Touschek background levels.

On the detector operation front, we demonstrated that the BEAST TPCs are capable of obtaining high purity nuclear recoil samples to reconstructed ionization energies as low as $6\,\mathrm{keV_{ee}}$ at effective gains of $\mathcal{O}(1000)$ (Table 4.3). We've also introduced a new method that reduced head-tail assignment bias for highly inclined tracks, leading to head-tail identification efficiencies of about 91% for simulated ^{4}He recoils with ionization energies above $40\,\mathrm{keV_{ee}}$ that also point back toward the beam pipe, compared to 72% on the same set of events without these corrections. This sample of ^{4}He events also has an average angular resolution within $8°$. Due to the finite dynamic range of the FE-I4 readout, low gain operation is ideal for fast neutron background studies, however if we want to push toward improving low energy directional sensitivity, angular resolution, and X-ray rejection performance, we need to push to higher gain. The BEAST TPCs are capable of operating at double GEM gains approaching 50,000—far beyond the threshold at which they are sensitive to single electrons—so in Part II of this dissertation, we study high gain performance near single electron efficiency and introduce new techniques for low energy electron rejection and directional sensitivity in the BEAST TPCs.

Part II: Deep learning for improved directional recoil detection

INTRODUCTION

In Part I we demonstrated excellent X-ray background rejection down to around $6\,\mathrm{keV_{ee}}$ and head-tail recognition efficiencies of about 91% at an average angular resolution within $8°$ using a sample of He recoils between $40\,\mathrm{keV_{ee}}$ and $1\,\mathrm{MeV_{ee}}$ in the BEAST TPCs operating at gains of $\mathcal{O}(1{,}000)$. With low gain amplification, the majority of charge present in keV-scale nuclear recoils falls below the FE-I4b threshold, meaning that in actuality, the true recoil energy is substantially higher than the observed reconstructed recoil energy. As a result, electron rejection and head-tail directional identification of a true $6\,\mathrm{keV_r}$ recoil will be more challenging than doing so for reconstructed $6\,\mathrm{keV_{ee}}$, especially at lower gains where a TPC isn't sensitive to charge from single electrons. The primary reason for this is because the expected length of the primary track of the $6\,\mathrm{keV_{ee}}$ recoil will be larger than that of the $6\,\mathrm{keV_r}$ recoil, leading to $6\,\mathrm{keV_{ee}}$ recoils on average, having better angular resolution and better defined principal axes compared to $6\,\mathrm{keV_r}$ events. Even at gains where the FE-I4 is sensitive to the charge of single electrons, reconstructed recoil energies for recoils below around $100\,\mathrm{keV_r}$ will be less than the true recoil energy due to ionization quenching. We have not measured the so-called ionization quenching factor (IQF) for He recoils in the BEAST TPCs, so we do not attempt to "undo" the energy loss due to quenching, but we do perform a high gain study where we are nearly sensitive to the charge of single electrons.

Directional searches for $\mathcal{O}(\mathrm{GeV})$ WIMPs will require directional sensitivity and background rejection for keV-scale nuclear recoils. When making the transition from neutron searches to WIMP searches, we ideally want to operate in the single-electron efficiency regime so that charge is not lost below threshold. Current readout technologies like the FE-I4 chips in the BEAST TPCs have a limited dynamic range, so operating at single electron-efficiency is ideal for keV-scale nuclear recoil searches, but not great for fast neutron measurements, as higher energy recoils are completely dominated by saturated pixel hits at these gains, leading to poor energy resolution.

Over the course of Part II, we develop a deep learning approach to both particle identification and directional head-tail assignment. In particular, we construct 3D convolutional neural networks (3DCNNs) that are directly fed in 3D images of recoil tracks and train them for these two applications. As we'll soon see, the native readout resolution of the FE-I4 chip lends itself nicely to the use of 3DCNNs with current-generation hardware. Our overall goal in Part II is to make the case for utilizing deep learning as an end-to-end approach for electron rejection and head-tail identification in recoil imaging TPCs.

CHAPTER 5
HEAD-TAIL IDENTIFICATION PERFORMANCE AT LOW GAIN: SIMULATION BENCHMARK

5.1 Introduction

Of the performance metrics of directional DM detectors listed in Chapter 1, head-tail recognition efficiency (coupled with event-level timing) is the most important quantity for improving DM sensitivity [12]. Figure 5.1 illustrates this by comparing the number of nuclear recoil events required to exclude a solar neutrino background hypothesis at 90% confidence limits versus head-tail recognition efficiency and angular resolution. Comparing the upper right-hand and lower left-hand corners of these plots, we see that a detector with no angular resolution but perfect head-tail identification requires significantly fewer events for rejecting solar neutrino backgrounds than a detector with perfect angular resolution but no head-tail recognition. Figure 5.1 is concerned with directional sensitivity beneath the neutrino floor, which corresponds to keV-scale nuclear recoils. While the ultimate goal is to demonstrate head-tail sensitivity for keV-scale nuclear recoils, any improvement in head-tail assignment performance is desirable. As a reminder, head-tail recognition efficiency, ε_{ht}, is the fraction of events where the scalar product between the true initial direction of the recoil and the reconstructed observed recoil event is positive.
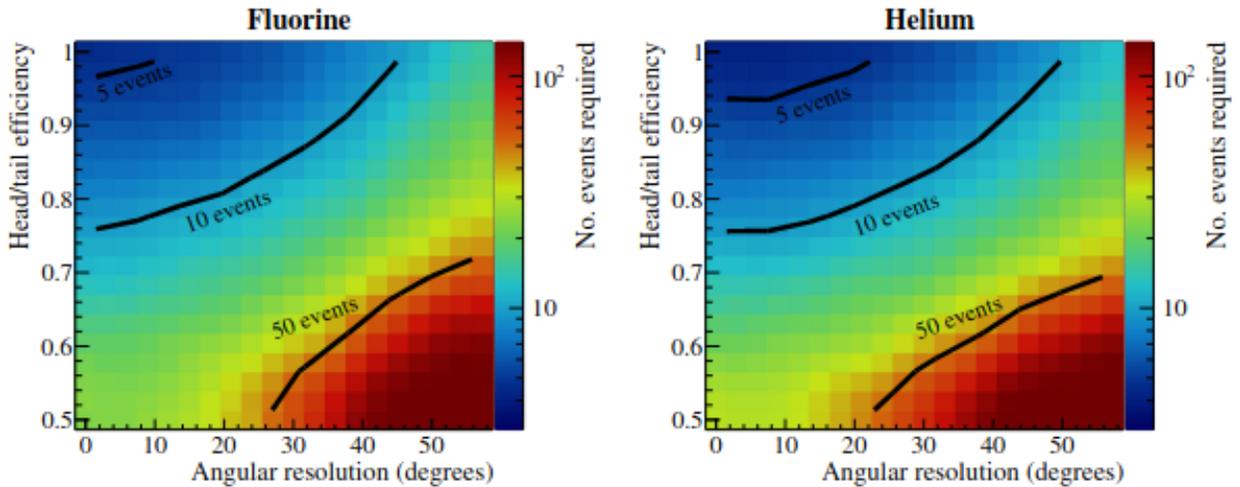
Figure 5.1: Taken from [12]: Estimated performance for the discrimination of solar neutrino background and DM signal events assuming a WIMP mass of $10\,\text{GeV}$, a negative ion drift He:SF$_6$ target gas and a threshold of $1\,\text{keV}_\text{r}$. The color scale represents the number of F recoils (left) and He recoils (right) needed to exclude a solar neutrino background hypothesis at a 90% confidence level.

Gas [70%:30%]	W [eV]	Gain	$\vec{v}_d$ [μm/BCID]	(σ_T, σ_L) [μm/$\sqrt{\text{cm}}$]	$(\sigma_{T,\text{GEM}}, \sigma_{L,\text{GEM}})$ [μm]
He:CO_2	35	909	216.25	(114, 114)	(180,180)

Table 5.1: Digitization parameters for our low gain head-tail simulation campaign. $\sigma_{T,\text{GEM}}$ and $\sigma_{L,\text{GEM}}$ represent the transverse and longitudinal point resolution of the readout plane, excluding the pixel chip.

Our aim in this chapter is to introduce and benchmark new analysis techniques that lead to improved head-tail assignment efficiencies in the BEAST TPCs. The techniques we benchmark are:

1. Baseline: Head charge fraction method (Chapter 5.3)

2. Log likelihood (LL) method (Chapter 5.4)

3. Primary track recovery (`ptr`) [82] (Chapter 5.5)

4. Deep learning (3DCNN) method (Chapter 5.6)

We focus on low gain in this chapter because our baseline method of head-tail assignment has been well validated on higher energy nuclear recoils both in simulation and in measurement (Chapter 4.5.3 and Refs. [47; 48]), making low gain settings a good starting point.

5.2 Overview of simulation

We use the Geant4 ParticleGun module present in basf2 [65] to shoot α particles inside the sensitive volume of a TPC. We generate 612,505 α tracks (He recoils) with a uniform distribution of energies between $0.035\,\text{keV}_\text{r}$ and $525\,\text{keV}_\text{r}$. The α tracks are fully fiducialized with initial vertices located at $(x_{\text{TPC}}, y_{\text{TPC}}, z_{\text{TPC}}) = (1.00\,\text{cm}, 0.84\,\text{cm}, 5.00\,\text{cm})$, and isotropic direction (uniform ϕ and uniform $\cos(\theta)$ distributions). After the primary ionization track is generated, the same Geant4 simulation processing described in Chapter 4.2 is used. We also use the digitizer described in Chapter 4.2 to drift, diffuse, amplify, and digitize each event with parameters shown in Table 5.1. Figure 5.2 shows six digitized simulated He recoil tracks with their corresponding reconstructed ionization energies (in keV_ee) and true recoil energies (in keV_r). Our simulated gain of 909 is well below single electron efficiency so there will always be charge near the perimeter of the tracks that falls under threshold. As track energy increases, the fraction of total event charge below threshold in general decreases. For the highest energy track in this event display, many pixels are saturated which limits the energy resolution, as the true amount of charge deposited on these pixels is larger than our saturation limit.
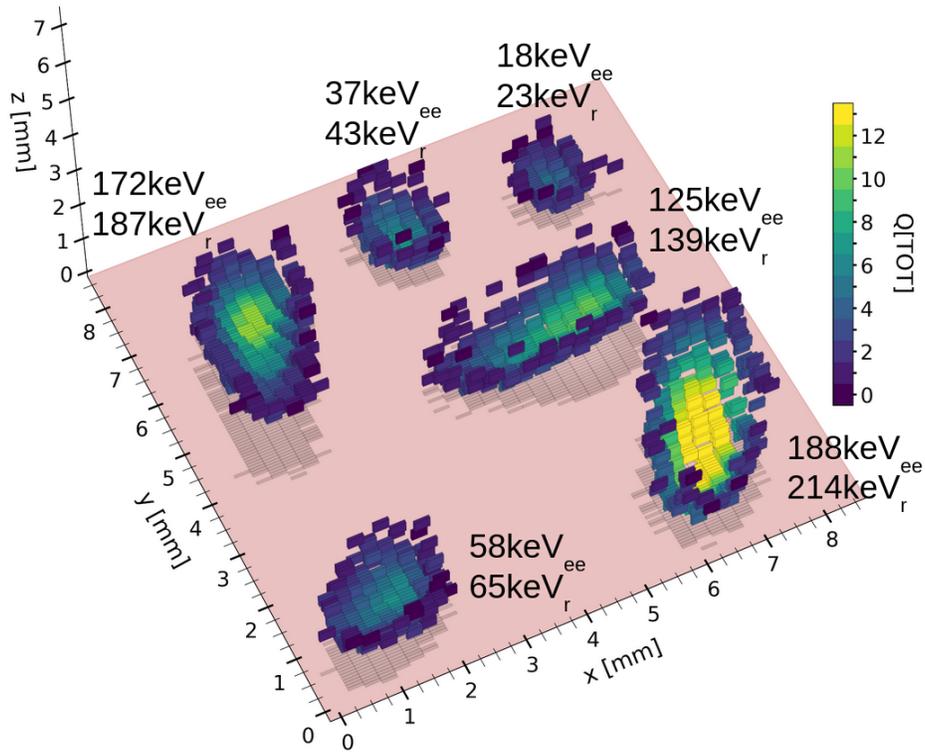
Figure 5.2: 3D charge distributions of six simulated He recoil tracks shown in the same event display. Each track is labeled with the reconstructed ionization energy (in keV_{ee}) and true ionization energy (in keV_r).

5.3 Head charge fraction (baseline) method

Recall from Chapter 4.5.3 that the stopping power of alpha recoils from fast neutrons fall beyond the Bragg peak (Figure 5.3), leading to a decrease in energy deposition of the recoiling nucleus the further it travels. This leads to a detectable charge asymmetry in nuclear recoil tracks that can be used to infer the vector direction of a track.
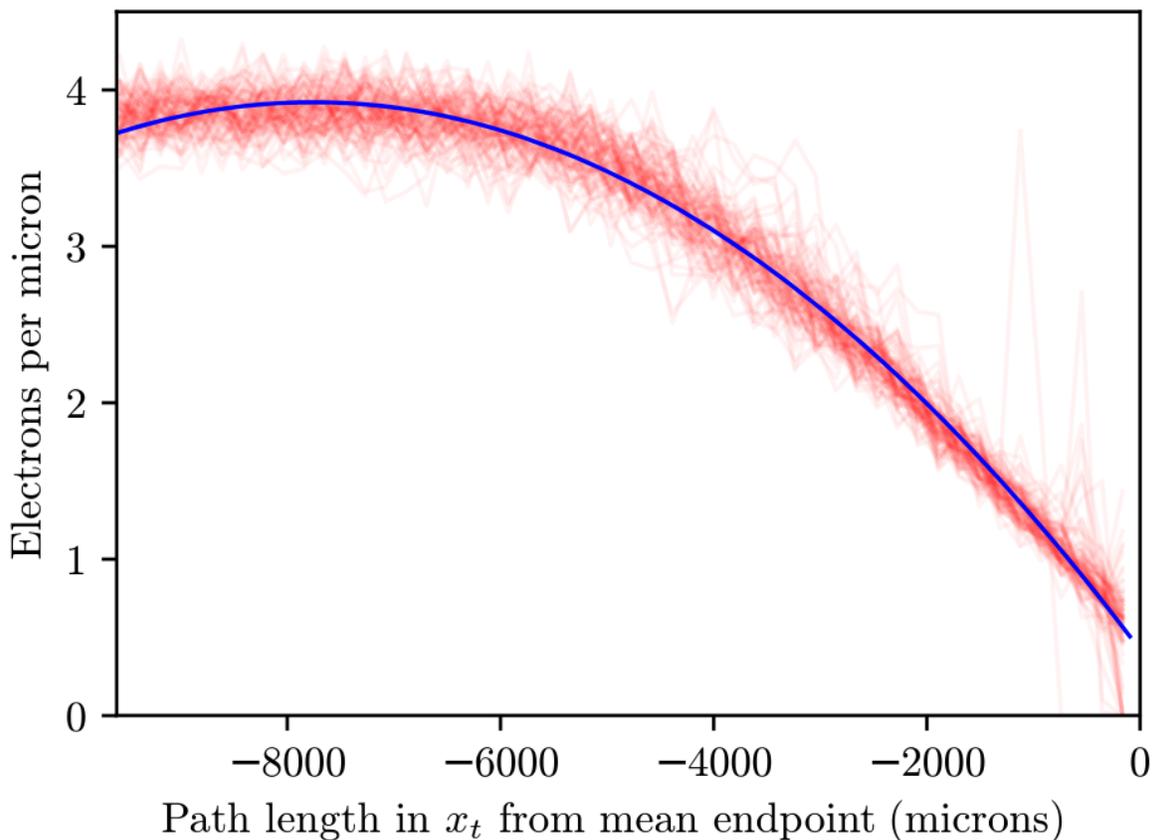
Figure 5.3: Taken from [82]: Simulated Bragg curves for one hundred $1\,\mathrm{MeV}$ alpha tracks (red lines) simulated in SRIM. Here positive x_t corresponds to the initial momentum direction of the recoil track with $x_t = 0$ being the track's stopping end (head). The blue curve is a fit representing the average behavior of these Bragg curves. The negatively sloped region of the blue curve is what we refer to as the region beyond the Bragg peak.

We use the following procedure to utilize this charge asymmetry to assign a vector direction to nuclear recoils:

1. Identify the principal axis of the track using an SVD.

2. Split the track in half along the midpoint of the track's principal axis.
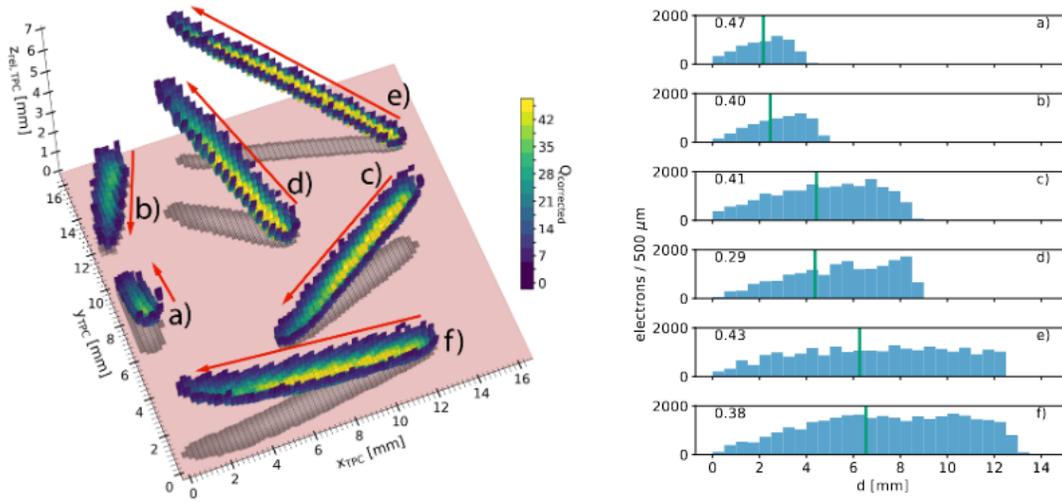
Figure 5.4: Taken from [48]: 3D charge distributions of nuclear recoil tracks with their assigned direction using the baseline HCF method (left) and charge distributions along the track's principal axis (right). In this figure, the principal axis was determined using a χ^2 minimization rather than the SVD we use throughout this work. The green line in the right hand plots shows the geometric center of these tracks. The left side of the green vertical line is the head-half of the track, with the head charge fraction labeled above the histogram.

3. Compute the fraction of charge on each half of the track.

4. Assign the vector head to the side of the track with less charge. We call the fraction of charge on the head-half of the track the head charge fraction (HCF) and with this procedure we have $0 < \text{HCF} < 0.5$.

Moving forward we will refer to this head-tail assignment procedure as the HCF method, or the baseline method. Figure 5.4 shows an example of the usage of the baseline HCF method to assign vector direction. The left plot in this figure shows six recoil tracks labeled (a)-(f) with associated arrows indicating their assigned vector direction. The right plot shows the charge distribution of each of the six tracks along the length of their principal axis. The green line indicates the geometric center of the track, so the head direction is assigned to the left side of the green vertical line. We may make selections on HCF to optimize head-tail assignment efficiency when assessing head-tail performance in our simulated sample.

5.4 Log likelihood (LL) method

In Chapter 4.5.3 we mentioned that charge integration effects can bias the detected charge asymmetry in highly inclined observed tracks, leading to frequent head-tail misassignment for highly

inclined downgoing ($\cos\theta < 0$) tracks using the baseline HCF method. Figure 5.5 illustrates this bias on a comparison between the x-z projection of a digitized recoil track and its truth primary ionization distribution. The detected track is split in half along its principal axis and the bin highlighted in red with a block border is filled with the charge highlighted in the pale red band of the primary track. We see that on the primary track, most of this charge comes from the upper half of the track, but because of the way charge is integrated in our detector, this detected charge is in the lower half of the measured track. As a result, the baseline HCF method would misassign this track as upgoing.
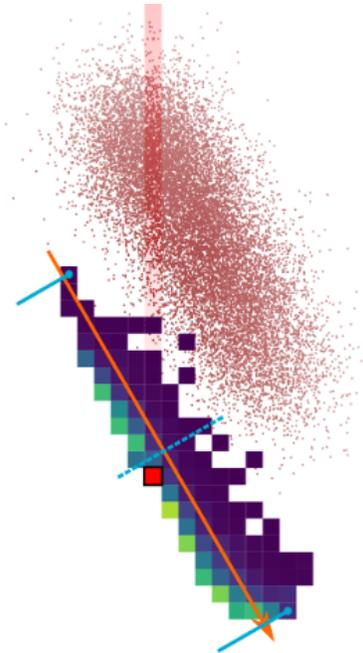
Figure 5.5: Taken from [82]: Illustration of head-tail mismeasurement resulting from the charge integration of a downgoing $400\,\mathrm{keV_r}$ recoil with $\theta_{\mathrm{TPC}} = 150°$.

The log-likelihood (LL) method is a recasting of the axial inclination correction method that was introduced in Chapter 4.5.3 to correct for this charge integration bias. The procedure for implementing the LL method is as follows:

1. Assign all tracks to point initially upward, so that $0 < \cos\theta_{\mathrm{assign}} < 1$. Let UCF represent the head charge fraction with this upward pointing assignment.

2. Use a Gaussian kernel density estimation (KDE) to generate unnormalized 2D likelihood functions, $\mathcal{L}_{\mathrm{up}}(\mathrm{UCF}, \cos\theta_{\mathrm{assign}})$ and $\mathcal{L}_{\mathrm{down}}(\mathrm{UCF}, \cos\theta_{\mathrm{assign}})$. $\mathcal{L}_{\mathrm{up}}(\mathrm{UCF}, \cos\theta_{\mathrm{assign}})$ is the KDE generated from the UCF vs. $\cos\theta_{\mathrm{assign}}$ for true upward pointing tracks and $\mathcal{L}_{\mathrm{down}}(\mathrm{UCF}, \cos\theta_{\mathrm{assign}})$ is the KDE generated from the UCF vs. $\cos\theta_{\mathrm{assign}}$ for true downward
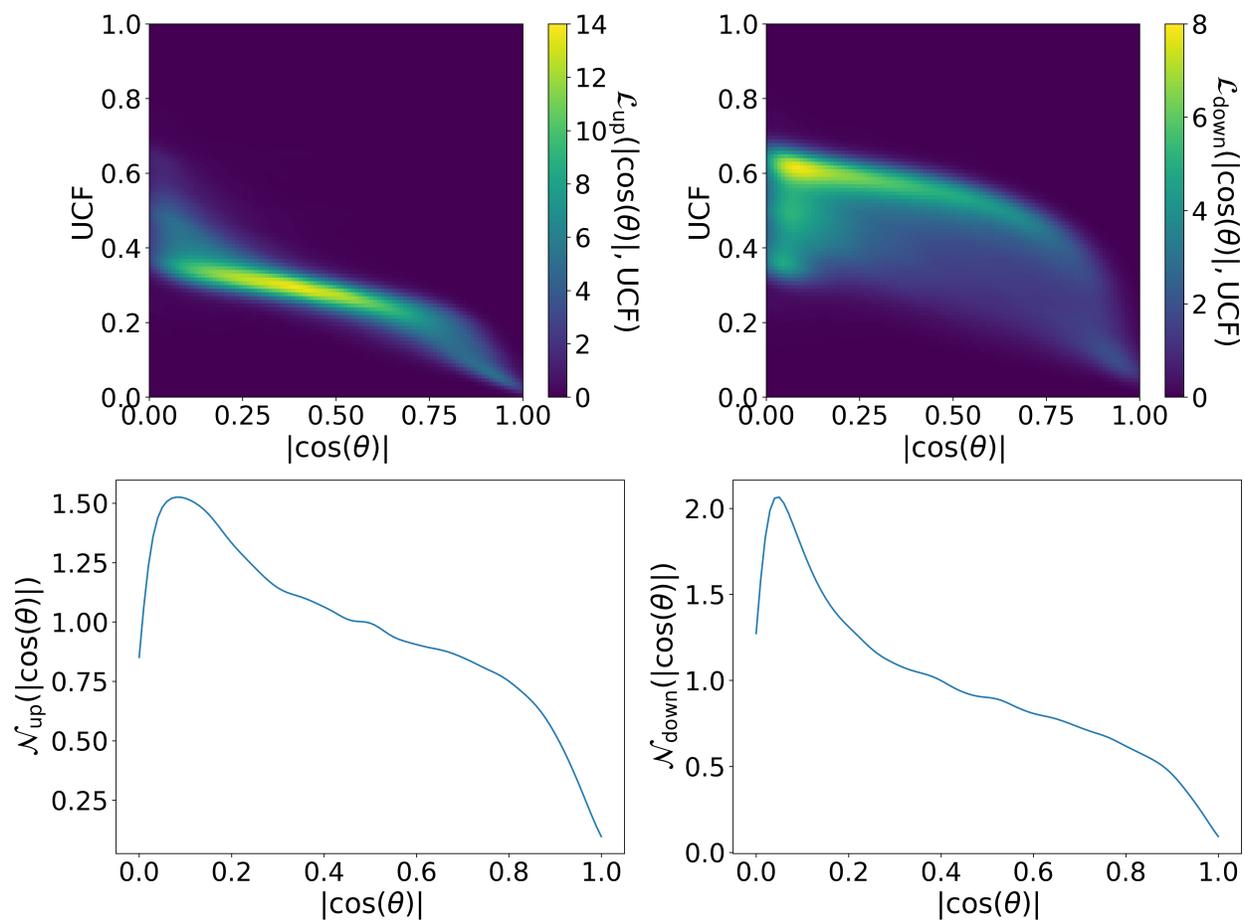
Figure 5.6: Unnormalized 2D likelihood distributions (top row) with corresponding normalization functions (bottom row) for upward (left column) and downward (right column) facing tracks.

pointing tracks (top row of Figure 5.6). We use Scott's rule [84] to determine the bandwidth of the KDE.

3. Use a Gaussian KDE to construct 1D likelihood functions $\mathcal{N}_{\text{up}}(\cos\theta_{\text{assign}})$ and $\mathcal{N}_{\text{down}}(\cos\theta_{\text{assign}})$, where $\mathcal{N}_{\text{up}}(\cos\theta_{\text{assign}})$ and $\mathcal{N}_{\text{down}}(\cos\theta_{\text{assign}})$ are generated using truth upward and truth downward facing tracks, respectively.

4. Normalize the 2D likelihood functions. To be specific, we compute normalized likelihood functions $\mathcal{L}'_{\text{up}}$ and $\mathcal{L}'_{\text{down}}$ where $\mathcal{L}'_{\text{up}} \equiv \mathcal{L}_{\text{up}}/\mathcal{N}_{\text{up}}$ and $\mathcal{L}'_{\text{down}} \equiv \mathcal{L}_{\text{down}}/\mathcal{N}_{\text{down}}$.

5. The confidence of our decision is given by $-2\log\left(\mathcal{L}'_{\text{up}}/\mathcal{L}'_{\text{down}}\right)$, which we use as a variable to assign head-tail. We assign tracks to be upward pointing if $-2\log\left(\mathcal{L}'_{\text{up}}/\mathcal{L}'_{\text{down}}\right) < c$ and tracks to be downward pointing if $-2\log\left(\mathcal{L}'_{\text{up}}/\mathcal{L}'_{\text{down}}\right) \geq c$. In this way, c can be called our directional decision boundary. We set c to be the value that maximizes the truth head-tail recognition efficiency in a given sample.

Figure 5.7 shows the distributions of log-likelihood confidences for truth upgoing and downgoing tracks. The black vertical line represents the value of c chosen where tracks to the right of this line are assigned to be downward and tracks to the left of this track are assigned to be upward. This value of c was determined to maximize the head-tail recognition efficiency, ε_{ht}. The normalized upgoing and downgoing likelihood templates, as well as the directional decision boundary, c, determined using simulation can also be applied to measurement.

5.5   Primary track recovery (ptr) method

Primary track recovery (ptr) [82] is a recently developed algorithm that was demonstrated on the BEAST TPCs as a technique to deconvolve detector effects and recover the primary ionization distribution of recoil tracks. Here we only briefly outline the steps that are needed to compute and optimize vector head-tail using ptr and refer the reader to Ref. [82] for a detailed description of the algorithm.

ptr algorithm for head-tail assignment

1. Define track unit vectors $(\hat{x}_{\text{track}}, \hat{y}_{\text{track}}, \hat{z}_{\text{track}})$. $\hat{x}_{\text{track}}$ is the detected track's principal axis, $\hat{y}_{\text{track}}$ is defined as

$$\hat{y}_{\text{track}} = \frac{\hat{z} \times \hat{x}_{\text{track}}}{\sin(\theta)}, \tag{5.1}$$

and $\hat{z}_{\text{track}}$ is defined as $\hat{z}_{\text{track}} = \hat{x}_{\text{track}} \times \hat{y}_{\text{track}}$. The detected track transformed into the track-coordinate system is shown in the upper left panel of Figure 5.8.
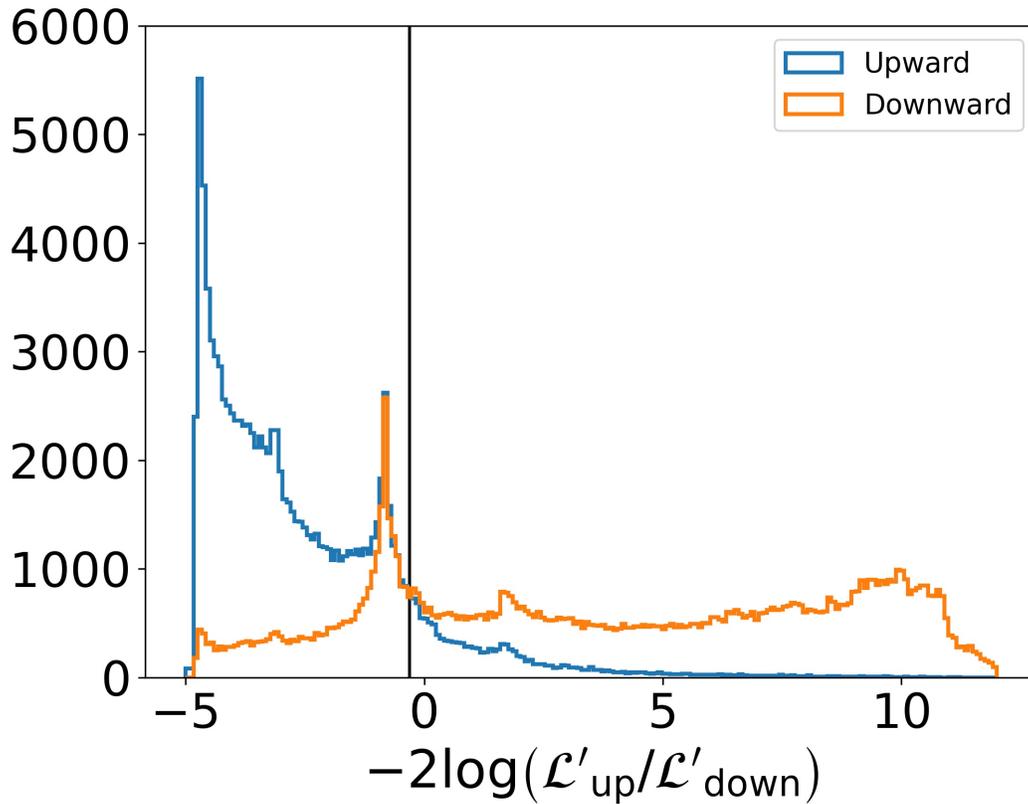
Figure 5.7: Log-likelihood confidence distribution of a random subset of about 230,000 of the simulated α tracks described in Chapter 5.2. True upward pointing tracks are shown in blue and true downward pointing tracks are shown in orange. The black vertical line represents the chosen value of c for which all events to the right (left) of this line are assigned to be downward (upward) facing by the LL method.

2. Slice the track along x_{track} and sample the charge at steps along y_{track} at each slice along x_{track}: We use $100\,\mu\text{m}$ spacing for the samples and slices along x_{track} and y_{track}, so we oversample in x and undersample in y. At each x_{track} slice (the vertical gray band in upper right panel of Figure 5.8 is an example of a slice), we collect the charge over all of the y_{track} samples within the slice. The y_{track} sample charges (indicated by color in Figure 5.8) are obtained by a bilinear interpolation from the charge in the pixel and its four surrounding pixels. If any one of the four surrounding pixels doesn't have charge or is saturated, the sample is discarded.

3. Fit a Gaussian profile of the form $A\hat{f}(y_{\text{track}};\mu,\sigma_T^s)$ to the charges sampled along each slice. Here σ_T^s is the estimated transverse diffusion of the slice. The bottom left panel of Figure 5.8 shows the charge profile fit that corresponds to the information within the vertical gray bands in the upper and lower right panels in this figure.

4. Derive the linear charge density at each x_{track} slice using

$$\lambda(x_{\text{track}}) = \frac{A\sigma_T^s\sqrt{2\pi}\sin(\theta)}{GL_xL_y}, \tag{5.2}$$

where A and σ_T^s are determined from the charge profile fit in step (3), G is the gain, and L_x and L_y are the length of the ionization distribution along the x and y axes, respectively.

5. The distribution of $\lambda(x_{\text{track}})$ versus x_{track} over each slice is the estimated "smeared"[1] Bragg distribution of the track (green bars in Figure 5.8). We fit this smeared Bragg distribution under two hypotheses: (i) that the track's vector direction matches the assigned direction from the SVD fit, and (ii) that the track's vector direction is flipped with respect to the orientation assigned by the SVD fit ($\hat{x}_{\text{track}} \to -\hat{x}_{\text{track}}$). We fit both of these hypotheses with a second order Chebyshev series (one hypothesis fit is shown as the solid magenta line in Figure 5.8) and assign the vector direction to the fit hypothesis that has the lowest $|\chi^2/\text{dof} - 1|$.

We can use the difference in the reduced χ^2 of the two Bragg fit hypotheses, $\Delta\chi^2/\text{dof}$, as a variable to optimize head-tail performance. These steps for determining vector direction may be applied to measured recoil tracks as was demonstrated in Ref. [82].

5.6 Deep learning (3DCNN) method

Here we introduce a 3DCNN classifier for head-tail identification. We first introduce the core building blocks of 3DCNNs and then discuss the data processing and class labeling convention

---

[1] We call this smeared because it's smeared by the diffusion along x_{track}. The `ptr` algorithm attempts to undo this smearing and recover the primary track, but we don't discuss this here since we're only interested in using `ptr` to assign vector head-tail direction.
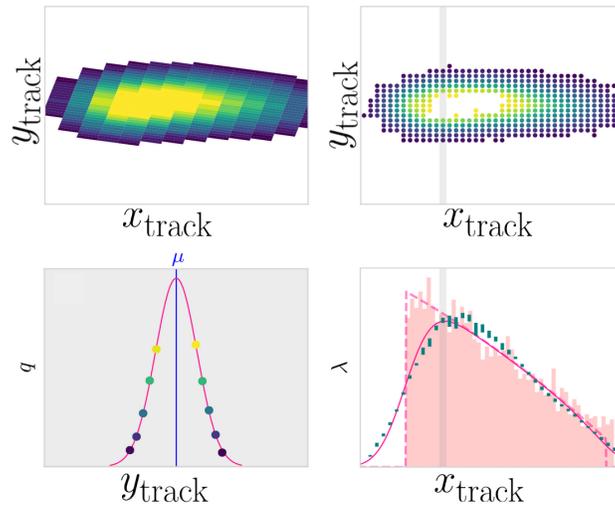
Figure 5.8: Adapted from [82]. Top left: Recorded ionization distribution rotated into the track coordinate frame. Colors represent charge with lowest charge in navy and highest charge in yellow. Top right: Slice and sample discretization as described in step (2). Bottom left: Profile of charge samples transverse to x_{track} within the gray shaded region in the top and bottom right panels. Bottom right: The green bars and solid magenta lines are relevant here and are described in steps (4) and (5). The magenta bars represent the true primary charge distribution of the track projected along its principal axis, and the dashed magenta lines represent the "unsmeared" Bragg fit which isn't relevant to this work.

used to train the 3DCNN to assign vector directions to tracks. We wrap this section up with a description of the network architecture we built, as well as our training procedure.

5.6.1 Building blocks of feature extraction in 3DCNNs
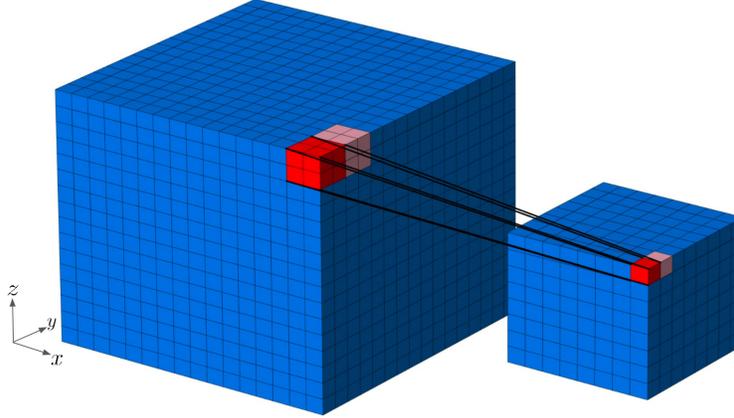
Figure 5.9: Schematic representation of the dimensionality reduction of a 3D convolution operation with a convolutional stride of $S = (2, 2, 2)$. Here $N \in \mathbb{R}^{16 \times 16 \times 16}$, and $K \in \mathbb{R}^{2 \times 2 \times 2}$. The black lines connecting the two red shaded regions illustrate the action of K on $N[14, 0, 14]$, a $2 \times 2 \times 2$ sub-block of N, leading to the $1 \times 1 \times 1$ element $C_{14,0,14}$ of $C \in \mathbb{R}^{8 \times 8 \times 8}$. Each index α, β, and γ of C runs along $\{0, 2, 4, 6, 8, 10, 12, 14\}$ due to the convolutional stride of $S = (2, 2, 2)$. The pink regions show the next element of the convolution operation after striding 2 units along the y direction on N.

3DCNNs are not yet widely used for natural image classification due to the large amounts of GPU memory required to store and perform operations on large 3D images with high voxel density. The lower energy recoil events that are of interest in directional DM searches, however, have their entire event topologies contained within a relatively small region, enabling the use of 3DCNNs with current-generation hardware, even in detectors with high spatial resolution. Here we briefly introduce 3D convolutions and 3D pooling; two core building blocks of our 3DCNN.

3D Convolutions

In 3DCNNs, the 3D convolution operation is a 3D cross correlation of an input grid of data, N, with a kernel (also called a filter), K, that slides along the input grid. To be more specific, let $N \in \mathbb{R}^{n_x \times n_y \times n_z}$ (often called an input feature map), and let $K \in \mathbb{R}^{k_x \times k_y \times k_z}$, where $k_j \leq n_j; \quad j = x, y, z$. We denote the shapes of N and K as $(n_x \times n_y \times n_z)$ and $(k_x \times k_y \times k_z)$, respectively. The action of K on $N[\alpha, \beta, \gamma]$, a $k_x \times k_y \times k_z$ sub-block of N, is an element, $C_{\alpha\beta\gamma}$, of the output "convolved" feature map, C, and is a mapping from $\mathbb{R}^{k_x \times k_y \times k_z} \rightarrow \mathbb{R}$ defined by

$$(K \star N[\alpha, \beta, \gamma])_{\alpha\beta\gamma} = \sum_{l=0}^{k_x-1} \sum_{m=0}^{k_y-1} \sum_{n=0}^{k_z-1} N_{\alpha+l,\beta+m,\gamma+n} K_{lmn} \equiv C_{\alpha\beta\gamma}, \tag{5.3}$$

where $\alpha + k_x \leq n_x$, $\beta + k_y \leq n_y$, $\gamma + k_z \leq n_z$, and $(K \star N[\alpha, \beta, \gamma])_{\alpha\beta\gamma}$ is a single element of the 3D cross correlation operation.

C is formed by sliding K in integer steps of S_x, S_y and S_z along the x, y, and z extents of N, respectively. In particular, C is an ordered grid composed of all $\lfloor \frac{n_x-k_x}{S_x}+1 \rfloor \times \lfloor \frac{n_y-k_y}{S_y}+1 \rfloor \times \lfloor \frac{n_z-k_z}{S_z}+1 \rfloor$ elements of $(R \star N[r_x, r_y, r_z])_{r_x r_y r_z}$, where $r_j \in \{0, S_j, 2S_j, \ldots, \lfloor \frac{n_j-k_j}{S_j}+1 \rfloor\}$; $\quad j = x, y, z$, and $\lfloor \rfloor$ represents the floor operator, telling us that we only compute elements of C when the shape of K is entirely contained within N. We call S_x, S_y, and S_z, the convolutional stride in x, y, and z, and will more compactly use $S \equiv (S_x, S_y, S_z)$ to denote the convolutional stride of a 3D convolution operation. Figure 5.9 shows an example of the shapes of input and output feature maps, where a $2 \times 2 \times 2$ filter acts on a $16 \times 16 \times 16$ input feature map with a convolutional stride of $S = (2, 2, 2)$, leading to an $8 \times 8 \times 8$ output grid. We can also use zero padding where we pad the outer perimeter of our input grid, N with $P = (P_x, P_y, P_z)$ layers of zeros. Here, P_j is a positive integer or zero and $j = x, y, z$. More generally, a $k_x \times k_y \times k_z$ kernel acting with a stride of (S_x, S_y, S_z) on an $n_x \times n_y \times n_z$ input feature map with additional zero padding of (P_x, P_y, P_z), will lead to an output feature map of shape $(C_x \times C_y \times C_z)$, where

$$C_j = \left\lfloor \frac{n_j - k_j + 2P_j}{S_j} + 1 \right\rfloor; \quad j = x, y, z. \tag{5.4}$$

In a 3DCNN, the elements of the convolutional filters, K_{lmn} are learnable parameters, so the network attempts to learn the features of interest to extract from a feature map.

3D Pooling

Pooling is an operation commonly used in convolutional neural networks to downsample a feature map while still retaining important information for classification. In our 3DCNN we include average pooling layers (called AvgPool; Chapter 5.6.3). Similar to a 3D convolution, the average pooling operation involves a 3D filter of a given size sliding over an input feature-map, N, but the average pooling operation simply computes the mean of all elements contained in the sub-block of N that the filter is sliding over. The average pooling operation will thus lead to an output feature map C of shape given by Equation (5.4). Since average pooling simply computes the mean of all elements in a sub-block of N, there are no learnable parameters associated with average pooling.

5.6.2 Data processing for the 3DCNN

The simulated recoil sample described in Chapter 5.2 includes 3D binned pixel hit information with associated charges (in TOT). Our aim is to feed 3D binned charge distributions of each event into the 3DCNN and have the 3DCNN output a class-assignment that corresponds to the head-tail direction of the recoil associated with its event image. In order to do this, we need to process the data into a grid of 3D voxels over which the operations contained in the 3DCNN (3D convolutions and average pooling) can be performed.

As mentioned in Chapter 5.2, the simulated events are fully fiducialized within the $2.00\,\text{cm} \times 1.68\,\text{cm} \times 2.16\,\text{cm}$ readout volume and the readout resolution is $250\,\mu\text{m} \times 50\,\mu\text{m} \times 216\,\mu\text{m}$. With this information, we could choose to process the entire readout into an $80 \times 336 \times 100$ grid of $250\,\mu\text{m} \times 50\,\mu\text{m} \times 216\,\mu\text{m}$ voxels but this would be computationally inefficient as none of our He recoils come close to spanning the entire readout in any dimension. To reduce computational expense, we shift each event so that the lowest binned charge in x, y, and z is at the origin. Shifting in this way ensures that the entire event topology of every event in our sample is fully contained within a $(34 \times 170 \times 34)$ grid of voxels, which we use for each recoil event.

We use the `PyTorch` [85] library for all neural network computations and store each event as a 2-tuple with the first entry being the $(34 \times 170 \times 34)$ voxelgrid filled with the binned charge of the centered-event and the second entry being the class-label of the event. We assign binary class-labels to events based on which ϕ hemisphere the event lies in. In particular, we assign a label of 1 if $\cos\phi_{\text{truth}} < 0$ and a label of 0 if $\cos\phi_{\text{truth}} > 0$. In this convention, forward-scattered nuclear recoils from neutrons incident on the lid (bottom) of a BEAST TPC are expected lie in the negative (positive) ϕ hemisphere, where the negative (positive) ϕ hemisphere is defined by $\cos\phi_{\text{truth}} < 0$ ($\cos\phi_{\text{truth}} > 0$). In Chapter 6 we run experiments with a neutron source incident both on the lid and bottom of a TPC, so this is why we choose this labeling convention here. Our reason for labeling tracks based on ϕ_{truth}, rather than ϕ_{measured} is so that we train the neural network to learn head-tail assignments on a recoil quantity that is independent of angular resolution and choice of principal axis.

Since raw *dense* voxelgrids are composed mostly of zeros, we save voxelgrids as *sparse* `PyTorch` tensors. Sparse tensors contain separate arrays of the x, y, and z bin numbers containing nonzero charge and the associated charge value at each (x, y, z) triplet. Our usage of sparse tensors saves a considerable amount of space, as storing each entry in a raw dense tensor format containing unsigned 8-bit integers would require around $200\,\text{kB}$ per event. This equates to about $120\,\text{GB}$ of space needed to store our entire sample of event images, compared to the $6.3\,\text{GB}$ when storing voxelgrids as sparse tensors. When we load images and feed them into the CNN, we first convert the sparse tensors back into dense tensors, as all convolution and pooling operations in the 3DCNN are performed over dense tensors.

5.6.3   Network architecture
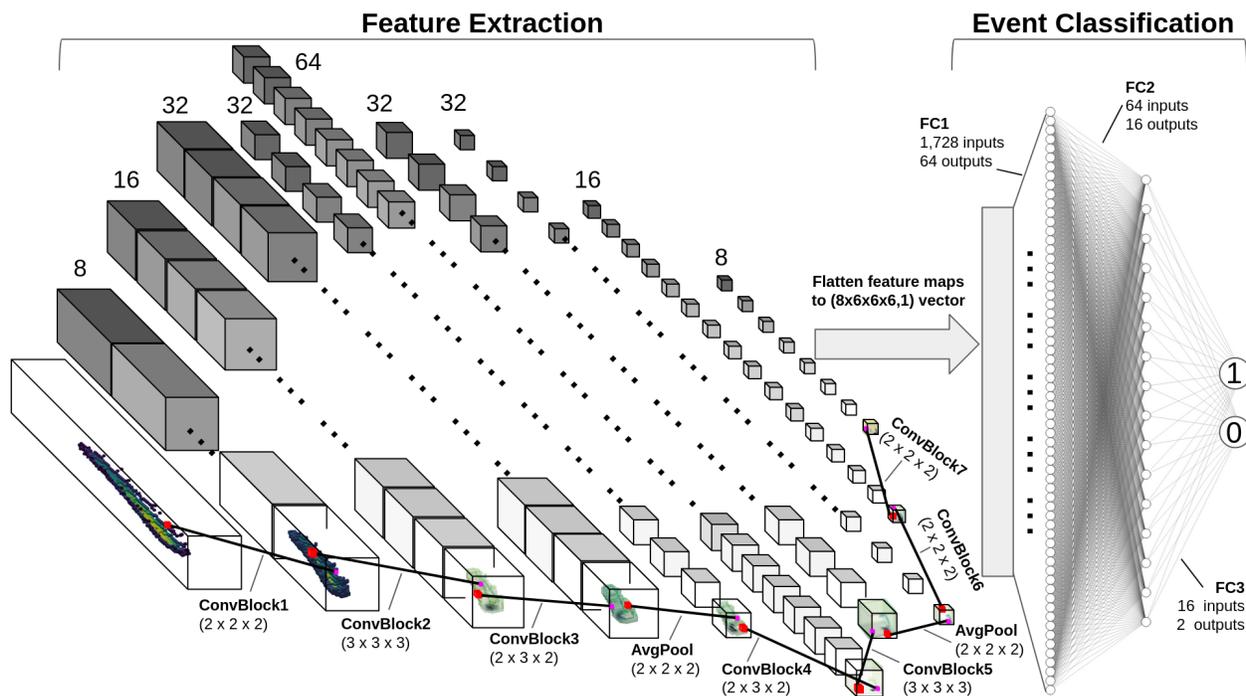
Figure 5.10: Architecture of our 3DCNN classifier. Within the feature extraction portion, each of the gray-scale prisms represents a unique feature map (indicated by the different shade of each map), with the size of the prisms shown roughly to scale in reference to the dimensions of the input image. The number of feature maps in each layer is listed above each layer and a feature map is illustrated at each layer through one possible traversal through the feature extraction layers. The illustrated feature maps downstream of ConvBlock2 show voxels of arbitrary and varying transparency to highlight distinct features. Solid black lines connect larger red prisms, which illustrate the convolutional or pooling filters acting on a portion of the feature map, to the smaller magenta cubes which are the outputs of the convolutional filter acting on the region shown in the input feature map. Each line is annotated with its associated computation name (description of each computation are in Table 5.2) and the size of the filter applied. The final eight feature maps are then flattened and passed into a fully connected neural network for classification, where in this diagram 1 represents tracks with $\cos\theta_{\text{truth}} < 0$ and 0 represents tracks with $\cos\theta_{\text{truth}} > 0$. The FCNN diagram was produced using [86].

Figure 5.10 and Table 5.2 together outline the network architecture of our 3DCNN. As is shown in Figure 5.10, our network consists of a series of convolutional and pooling layers for feature extraction, followed by a dense fully connected neural network (FCNN) for event classification. Within the feature extraction portion of the neural network, we employ seven convolutional blocks (ConvBlocks 1-7) and two pooling layers (AvgPool) to downsample the feature maps while still maintaining important features. Each of the seven convolutional blocks contain the following

Layer	# Filters	Filter size	S	P	# Learnable parameters	Dropout	Output shape
ConvBlock 1	8	$(2 \times 2 \times 2)$	(1,2,1)	0	88	0.05	$(8 \times 33 \times 85 \times 33)$
ConvBlock 2	16	$(3 \times 3 \times 3)$	(1,2,1)	0	3,504	0.05	$(16 \times 31 \times 42 \times 31)$
ConvBlock 3	32	$(2 \times 3 \times 2)$	1	0	6,240	0.05	$(32 \times 30 \times 40 \times 30)$
AvgPool	32	$(2 \times 2 \times 2)$	2	0	—	—	$(32 \times 15 \times 20 \times 15)$
ConvBlock 4	64	$(2 \times 3 \times 2)$	1	(1,0,1)	24,768	0.05	$(64 \times 16 \times 18 \times 16)$
ConvBlock 5	32	$(3 \times 3 \times 3)$	1	(1,0,1)	55,392	0.05	$(32 \times 16 \times 16 \times 16)$
AvgPool	16	$(2 \times 2 \times 2)$	2	0	—	—	$(32 \times 8 \times 8 \times 8)$
ConvBlock 6	16	$(2 \times 2 \times 2)$	1	0	4,144	0.05	$(16 \times 7 \times 7 \times 7)$
ConvBlock 7	8	$(2 \times 2 \times 2)$	1	0	1,048	0.05	$(8 \times 6 \times 6 \times 6)$
FC1	—	—	—	—	110,656	0.05	(64×1)
FC2	—	—	—	—	1,040	0.05	(16×1)
FC3	—	—	—	—	34	—	(2×1)

Table 5.2: More specific details of each layer that is shown in Figure 5.10. We assume a single $(1 \times 34 \times 170 \times 34)$ image is fed into the network. The output shape column gives the shape of the output after each layer. The output of ConvBlocks 1-7 and AvgPool is a tensor of dimension (#FeatureMaps $\times$ L $\times$ W $\times$ H), where L, W, and H are the length, width and height of each feature map, respectively. S is the convolutional stride of the layer and P is the zero padding. When S and P are given as integers, it is to be understood that the given amount is applied to all three spatial dimensions of the image, and when they are given as tuples, it is understood that the tuples are of the form (n_x, n_y, n_z), so $S = (1, 2, 1)$ means a convolutional stride of 1 in x and z and a stride of 2 in y. The output shapes of the FC layers are 1 dimensional vectors for each node in a given layer. In addition to the weights and biases associated with each node in the FCNN, the entries within each convolutional filter in the 3DCNN are also learnable parameters, so we list the total number of learnable parameters in each layer.

components: (i) A 3D convolution with a convolutional filter size listed in Table 5.2, (ii) a 3D batch normalization [87], (iii) a scaled exponential linear unit (SELU) activation layer [88], and (iv) a randomized dropout [89] of 0.05 to reduce overfitting.

Walking through our network architecture, we start with a $(34 \times 170 \times 34)$ input image. We implement a stride of $S_y = 2$ in both the ConvBlock1 and ConvBlock2 to downsample in y while picking out features of importance. Doing so reduces the computational overhead of training and evaluating our network over maintaining such a large y dimension compared to x and z. In ConvBlock3, we apply a prism shaped $(2 \times 3 \times 2)$ filter so that the output feature map of this layer has even number of voxels in all spatial dimensions. This makes it so each of the 32 AvgPool computations performed in this layer will sample each unique voxel in the image. ConvBlock4 and ConvBlock5 have their convolutional filter sizes and zero padding, P, carefully chosen so that the output feature map of ConvBlock5 will have an equal number of voxels along all three spatial dimensions. After ConvBlock5, we downsample with another AvgPool layer and finally, we perform two final convolutional blovks (ConvBlock6 and ConvBlock7) where we end up with eight $(6 \times 6 \times 6)$ feature maps. Since the values composing each convolutional filter are learnable parameters, we expect that when the network is trained, there will be useful features encoded in these most downstream feature maps. We finally flatten these downstream feature maps into an $((8 \times 6 \times 6 \times 6) \times 1) = (1728 \times 1)$ vector that contains each extracted feature at the end of the convolutional chain. This flattened (1728×1) feature vector is then fed into a fully connected dense neural network with two hidden layers and output class assignments corresponding to recoils assigned in the negative ϕ_{truth} hemisphere (labeled 1) and in the positive ϕ_{truth} hemisphere (labeled 0). Each of the three fully connected layers (FC1, FC2, and FC3) use a SELU activation function and FC1 and FC2 include a random dropout of 0.05. The raw model output of each event is a (2×1) vector, $\mathbf{z}$, with entries corresponding to each both class outputs $(\cos\phi_{\text{truth}} < 0$, and $\cos\phi_{\text{truth}} > 0)$. The softmax function

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{2} e^{z_j}}. \tag{5.5}$$

is applied to $\mathbf{z}$ to map the class outputs $z_i \in \mathbf{z}$ to class probabilities. We henceforth label $\sigma(\mathbf{z})_1$, and $\sigma(\mathbf{z})_2$, as p_-, and p_+, which represent the model-predicted class probabilities of $\cos\phi_{\text{truth}} < 0$, and $\cos\phi_{\text{truth}} > 0$, respectively.

5.6.4 Training the network

We first shuffle the order of all events and then split the data into distinct training, validation, and test sets with 306,252 events in our training sample, 76,563 events in our validation sample, and 229,690 events in our test sample. We set the test sample aside and train using stochastic gradient

descent with minibatch updates of size 128 and cross-entropy loss. In other words, we implement the following procedure to train our model:

1. Form a `PyTorch` tensor of dimension (128,1,34,170,34), which is composed of 128 randomly selected voxel images (each represented as a dense tensor) from the training sample. We call this tensor a minibatch.

2. Feed the minibatch and the corresponding truth label of each image of the minibatch into the 3DCNN.

3. Use `PyTorch`'s built in CrossEntropy loss function to compute the loss of the batch. We wish to minimize this loss. We use an Adam [90] optimizer with a learning rate of 0.0002.

4. Update model weights using backpropagation [91].

5. Repeat steps 1-4 until we've run through all events in the training set. This is called a training epoch.

6. At the end of each training epoch, repeat steps 1-3 for the validation sample. We do not implement step 4 as we don't wish to train the 3DCNN on the validation set. Compute the sum of the losses of each minibatch of the validation sample.

7. If the summed losses over the validation set minibatches are less than in the previous epoch, we treat this as the model learning and save all model weights.

8. Implement early stopping [92] where steps 1-7 are repeated until the total validation loss doesn't decrease at all over 10 successive epochs.

9. The model state corresponding to the epoch with the lowest validation loss is our trained model.

After our network is trained, we feed the test set through the network. By construction of the Softmax function, class probabilities p_- and p_+ satisfy $p_- + p_+ = 1$, so we only use p_- as our variable to optimize head-tail performance.

Figure 5.11 shows distributions of p_- and $\cos\theta_{\text{truth}}$ over various energy ranges subject to whether the 3DCNN correctly or incorrectly identified head-tail direction. Distributions of events with correctly (incorrectly) assigned head-tail direction are shown in blue (orange). Looking at the p_- distributions, we see that over all energies, this distribution is slightly asymmetric about 0.5, with more events incorrectly classified as being in the negative ϕ_{truth} hemisphere than being in the positive ϕ_{truth} hemisphere. This asymmetry is most distinctly seen at lower energies (middle-right panel of Figure 5.11) where there are more *misclassified* events than correctly classified events plotted in the $\cos\phi_{\text{truth}} > 0$ region, meaning these events have $\cos\phi_{\text{truth}} > 0$ but are falsely
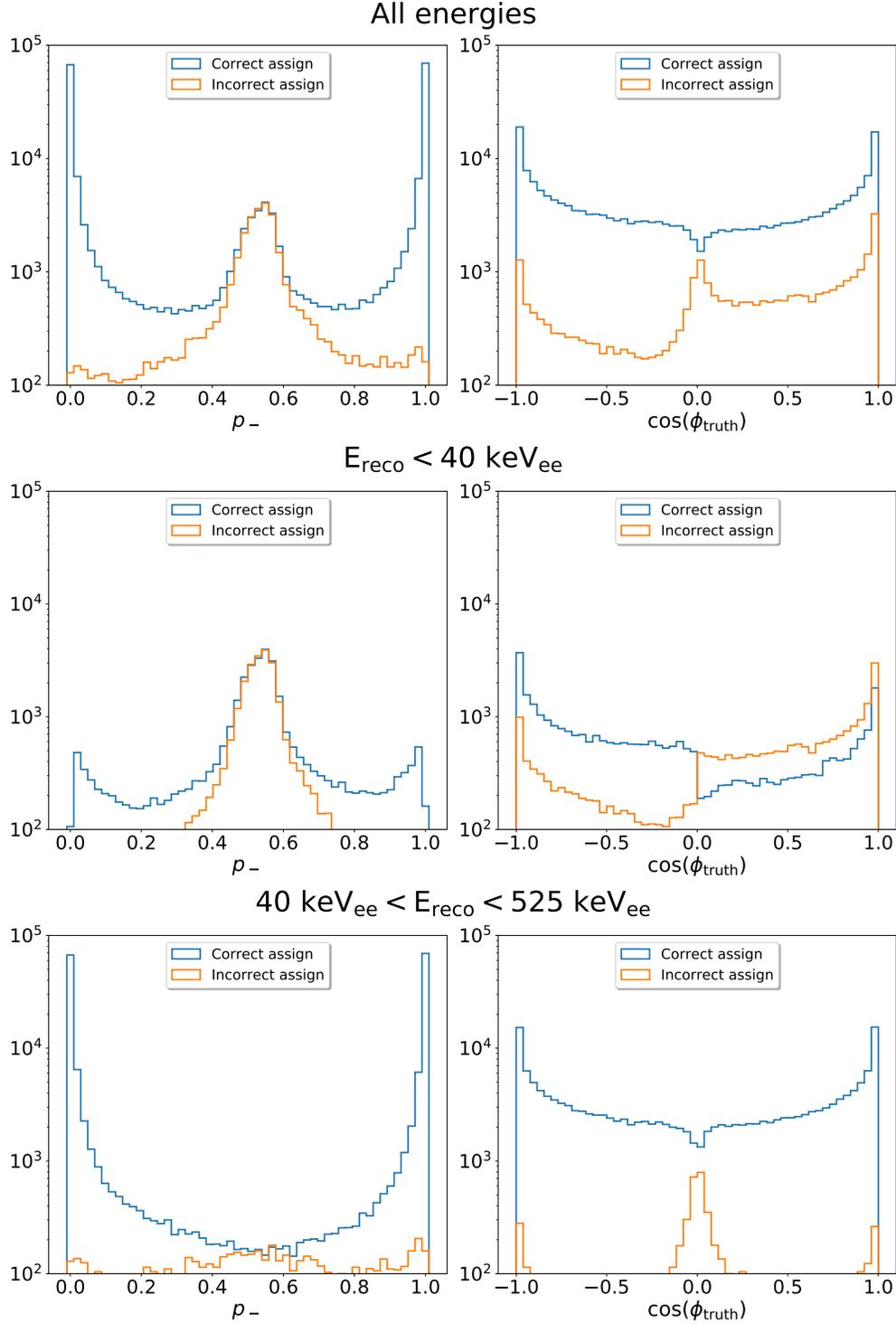
Figure 5.11: Distributions of 3DCNN output probability p_- (left), and the cosine of the true ϕ angle of the recoil (right). Distributions of events with correctly (incorrectly) assigned head-tail direction are shown in blue (orange). The top row of plots shows events over all energies, the middle row shows low energy events satisfying $E_{\text{reco}} < 40\,\text{keV}_{\text{ee}}$, and the bottom row shows higher energy events above $40\,\text{keV}_{\text{ee}}$.

predicted by the 3DCNN to have $\cos\phi_{\text{truth}} < 0$. Above $40\,\text{keV}_{\text{ee}}$ (bottom row of Figure 5.11) we observe approximately symmetric head-tail classification performance both in p_- and in $\cos\phi_{\text{truth}}$, which is what we would expect for a good head-tail classifier trained on isotropic simulated recoils.

Looking at the left panel in all three rows of Figure 5.11, we find that in the neighborhood of $p_- = 0.5$, the number of incorrectly assigned events essentially matches the number of correctly assigned events, so performance is essentially a coin-flip as expected. As p_- approaches 0 or 1, however, we find that the 3DCNN classifier performs extremely well. Given this behavior of the p_- distribution, we form a new confidence-based quantity

$$w \equiv 2\left|p_{\pm} - \frac{1}{2}\right|, \tag{5.6}$$

which by construction peaks near 0 when the classifier is maximally uncertain, and peaks near 1 when the classifier is certain. It can be shown that w will be the same regardless of whether we input p_+ or p_-, hence the $p_{\pm}$ designation in Equation (5.6). We use w as our quantity for optimizing ε_{ht} with the 3DCNN classifier.

5.7 Results

Here we compare the results of each head-tail assignment method on the test sample of 229,690 He recoil events. We use the test set so that we can fairly evaluate head-tail performance on an identical sample of recoils that the 3DCNN wasn't trained on. We start by generating ROC-like ε_{ht} versus ε_{He} distributions to assess performance of making selections based on the head-tail optimization variable we previously defined for each of the techniques. To allow for a direct comparison of performance, we compare the head-tail efficiencies of each method that correspond to 50% He recoil efficiency.

5.7.1 Baseline method

For the baseline method, our head-tail discriminant variable is the head charge fraction (HCF). The left panel of Figure 5.12 shows the HCF distribution of events after having their vector direction flipped so that $0 < \text{HCF} < 0.5$ as described in Chapter 5.3, and the right panel shows the ROC curve generated from this selection. The vertical black line on the left panel of this figure shows $\text{HCF} = 0.355$ which is the HCF selection value corresponding to $\varepsilon_{\text{He}} = 0.5$. The ROC curve on the right tells us that for the sample of recoils with $\text{HCF} < 0.355$, we keep 50% of He recoils and correctly assign head-tail for 79% of these recoils.
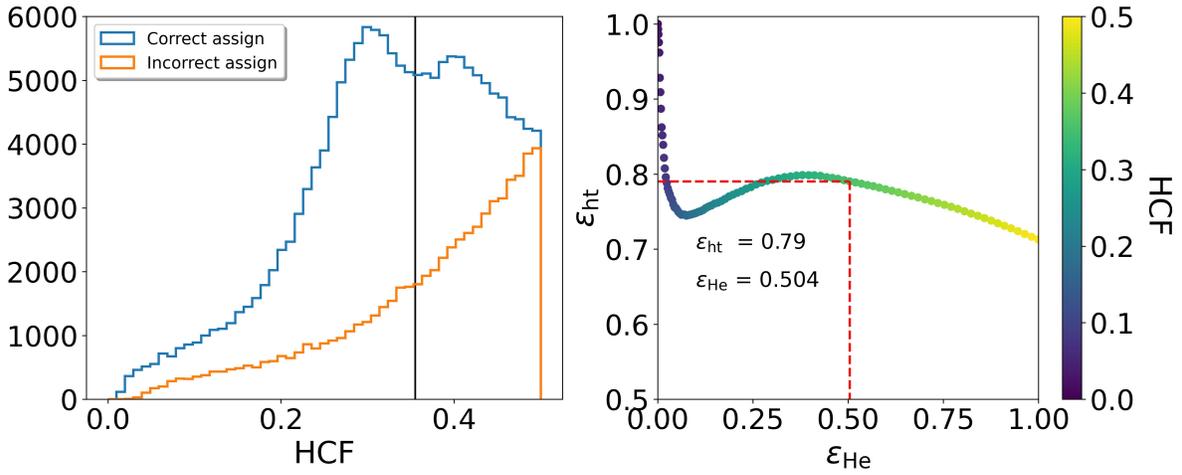
Figure 5.12: Left: HCF distributions with correctly assignments in blue and incorrect assignments in orange. The vertical black line shows the HCF value corresponding to $\varepsilon_{\text{He}} = 0.5$. Right: ROC curve showing ε_{ht} versus ε_{He} at various HCF selections. About 79% of events with HCF < 0.355 have correctly assigned vector direction.
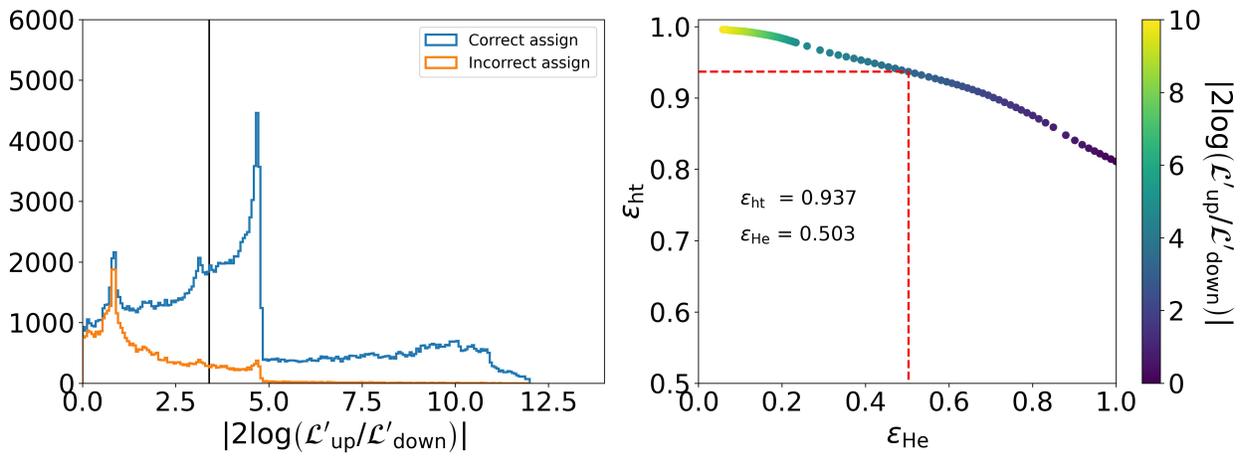
Figure 5.13: Left: Absolute value of the normalized log-likelihood ratios, with events with correctly assigned head-tail in blue and incorrectly assigned vector direction in orange. The vertical black line at $|2\log(\mathcal{L}'_{\text{up}}/\mathcal{L}'_{\text{down}})| = 3.4$ is the 50% He recoil efficiency line. We find that nearly 94% of events with $|2\log(\mathcal{L}'_{\text{up}}/\mathcal{L}'_{\text{down}})| > 3.4$ have correctly assigned vector direction.

5.7.2 LL method

Moving forward, we use the absolute value of the normalized log-likelihood ratio, $|-2\log\bigl(\mathcal{L}'_{\text{up}}/\mathcal{L}'_{\text{down}}\bigr)|$, as our selection variable for head-tail performance with the LL method.

Figure 5.13 shows the head-tail classification performance of choosing $|2\log\bigl(\mathcal{L}'_{\text{up}}/\mathcal{L}'_{\text{down}}\bigr)| > 3.4$ which corresponds to about 50% He recoil efficiency. From the right panel of this figure, we find that these selections correspond to a head-tail recognition efficiency of about 0.94. Furthermore we find a head-tail recognition efficiency of over 81% at 100% He recoil efficiency which is already higher than the best performing head-tail recognition efficiency using the baseline method.

5.7.3 `ptr` method

We evaluate head-tail performance using $\Delta(\chi^2/\text{dof})$, the difference in reduced χ^2 between the two Bragg fit hypotheses determined by the `ptr` algorithm. Figure 5.14 shows that a selection of $\Delta(\chi^2/\text{dof}) > 56$ corresponds to $\varepsilon_{\text{He}} = 0.5$ with ε_{ht} near 0.98.
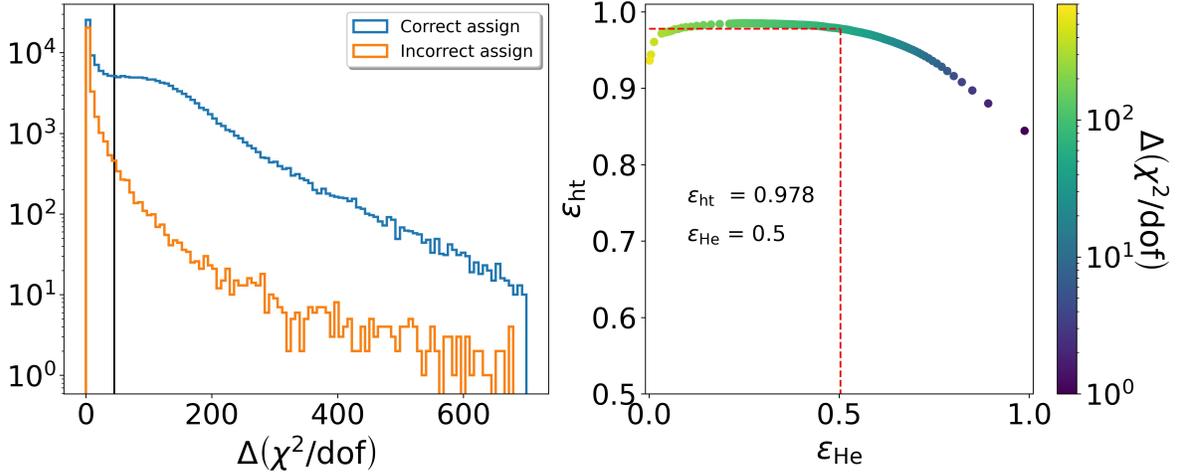
Figure 5.14: Left: Distribution of difference in reduced χ^2 of Bragg fit hypotheses with correctly assigned events in blue and incorrectly assigned in orange. The vertical black line indicates the value above which we have $\varepsilon_{\text{He}} = 0.5$. Right: ROC curve showing ε_{ht} versus ε_{He} at various $\Delta(\chi^2/\text{dof})$ selections. Nearly 98% of events with $\Delta(\chi^2/\text{dof}) > 56$ have correctly assigned vector direction.

5.7.4 3DCNN method

We use $w = 2|p_- - 1/2|$ as our head-tail selection variable. Figure 5.15 shows that selections on w lead to remarkable head-tail assignment performance even at high recoil efficiencies. In particular we find that a selection of $w > 0.9948$ corresponds to a 50% He recoil efficiency and a head-tail recognition efficiency of 99.9%. Moving further to the right along the ROC curve in Figure 5.15,

we find that at $\varepsilon_{\text{He}} = 0.75$, the 3DCNN correctly classifies the head-tail direction 97.9% of events ($w > 0.57$), meaning the 3DCNN classifier at 75% He recoil efficiency outperforms the head-tail performance of the other three methods at 50% efficiency.
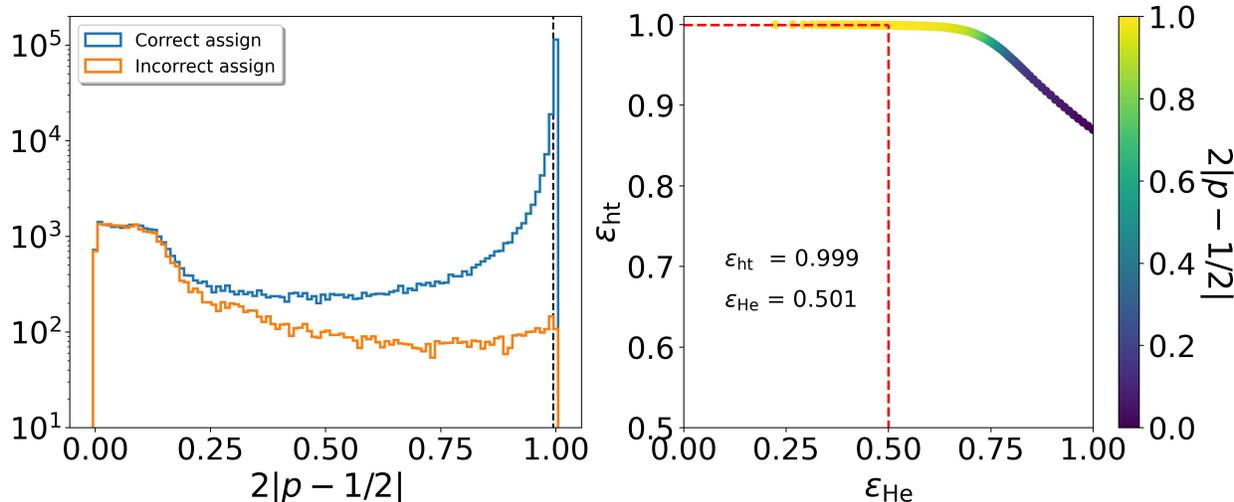
Figure 5.15: Left: Distribution of $w = 2|p_- - 1/2|$ with correctly assigned events in blue and incorrectly assigned events in orange. The vertical black dashed line indicates the value above which we have $\varepsilon_{\text{He}} = 0.5$. Right: ROC curve showing ε_{ht} versus ε_{He} at various w selections. The 3DCNN is a nearly perfect classifier for this sample of events at 50% He recoil efficiency.

5.8   Results summary

	Baseline	LL	ptr	3DCNN
ε_{ht} at $\varepsilon_{\text{He}} = 0.5$	0.790	0.937	0.978	0.999

Table 5.3: Head-tail efficiencies of each of the four methods at 50% He recoil efficiency.

Table 5.3 summarizes the head-tail efficiencies of each method at $\varepsilon_{\text{He}} = 0.5$. Over our entire sample of recoils, we find that our newly introduced 3DCNN performs significantly better than the other methods at this fixed He recoil efficiency.

We're ultimately interested in how head-tail assignment performance varies with energy. To study this, we group our recoil sample into energy bins of width $10\,\text{keV}$ with bin centers ranging from $5\,\text{keV}$ to $515\,\text{keV}$. We then use each of our head-tail selection variables to, within each bin, determine the selection that corresponds to 50% He recoil efficiency, and compute the corresponding head-tail recognition efficiency. Figure 5.16 shows these head-tail recognition efficiencies versus energy where we generated energy bins based on reconstructed ionization energy (keV_{ee}; top plot)

and truth recoil energy (keV$_r$; bottom plot). We call this the "grand summary" of our head-tail performance, as it illustrates optimized head-tail performance in each energy bin.
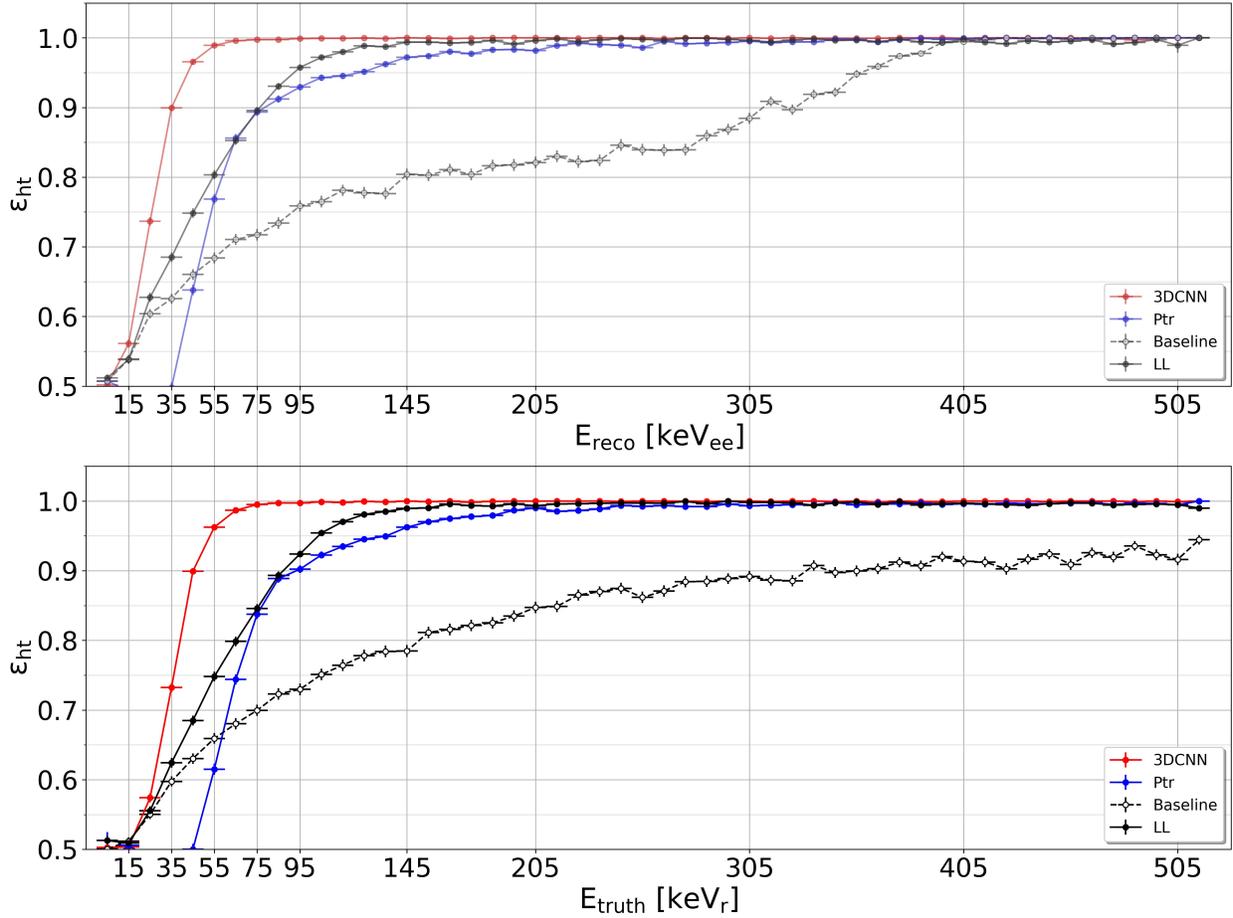
Figure 5.16: Head-tail recognition efficiency versus reconstructed recoil energy (top) and truth recoil energy (bottom) computed on isotropically-simulated He recoils at a gain of 910 for all four head-tail assignment methods. Each data point shows the head-tail recognition efficiency corresponding to 50% He recoil efficiency within the energy bin using the head-tail selection variable for the method shown. The 3DCNN gives both the best overall head-tail performance and the best head-tail performance at low energies.

Despite `ptr` outperforming the LL method when a single discriminant threshold is used at all recoil energies (Table 5.3), we find from this figure that the LL method performs much better than `ptr` at low energies, which are more relevant for dark matter detection. Indeed, the median energy of events remaining using LL selections over all energies corresponding to 50% He recoil efficiency is $210\,\mathrm{keV_{ee}}$ compared to $278\,\mathrm{keV_{ee}}$ for `ptr` selections, indicating that selections on $\Delta(\chi^2/\mathrm{dof})$ have a stronger preference for selecting higher energy nuclear recoils than selections on $|2\log(\mathcal{L}'_{\mathrm{up}}/\mathcal{L}'_{\mathrm{down}})|$. Our results show event-level head-tail recognition efficiencies of 90% between

$40\,\mathrm{keV_r}$ and $50\,\mathrm{keV_r}$ with the 3DCNN, and non-negligible head-tail recognition sensitivity down to around $25\,\mathrm{keV_r}$. While these results are a promising start and Figure 5.16 suggests that the 3DCNN significantly outperforms other head-tail determination strategies, these results must be validated with measurement before they are truly meaningful.

CHAPTER 6
HEAD-TAIL IDENTIFICATION PERFORMANCE AT LOW GAIN: EXPERIMENTAL VERIFICATION

In this chapter, we detail an experiment to measure the head-tail effect and evaluate the performance of our 3DCNN head-tail classifier evaluated on experimental data.

6.1 Overview and experimental setup

We use a $\sim$3-µCi ^{252}Cf fast-neutron source enclosed in a borated polyethylene shield to generate nuclear recoils in a single BEAST TPC. The shield has a small hole, providing a collimated neutron beam incident on the TPC. We run three measurement campaigns, labeled (i), (ii), and (iii) in Figure 6.1. Campaign (i) includes about 145 hours of continuous data collection with the ^{252}Cf source incident on the lid of the TPC; Campaign (ii) includes about 185 hours of continuous data collection with the ^{252}Cf source incident on the bottom of the TPC; and Campaign (iii) includes about 337 hours of continuous data collection with no neutron source present. Given the fixed source location during Campaigns (i) and (ii), we expect each campaign to give us a highly directional sample where the vast majority of detected nuclear recoils originate from neutrons coming directly from the source. Multiple scattering and back scattering can occur, which will lead to nuclear recoils whose true directions may not directly point back to the source, but we expect these to be minor effects. Table 6.1 summarizes the setup and measurements recorded during each of the three campaigns.
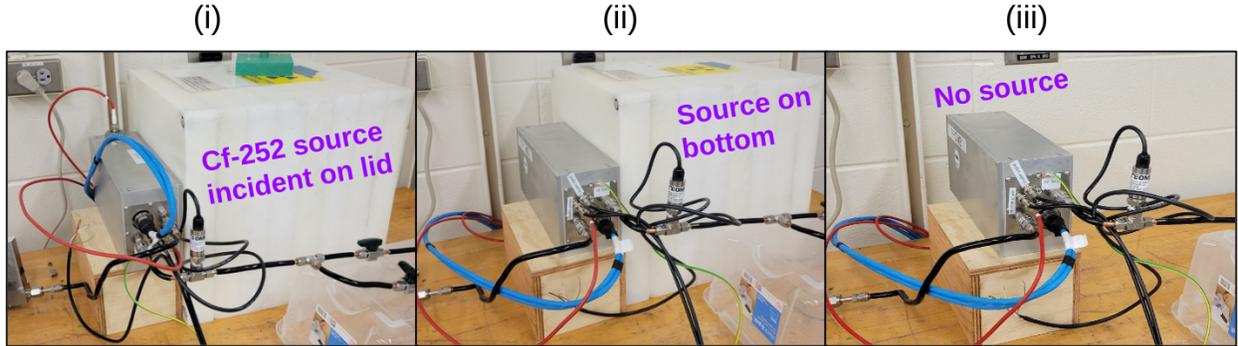
Figure 6.1: Lab set-up for low gain head-tail studies. We use a ^{252}Cf spontaneous fission source to generate neutron recoils for the TPCs. We record around one week of data both for the source-on-lid sample and source-on-bottom sample, and about two weeks of background "no source" data.

Before recording data, we calibrate the charge scale of the TPC using the same procedure detailed in Chapter 4.3 and obtain a threshold of around 2100 e and a saturation limit of around

Campaign	Meas. time [h]	Source location	Expected ϕ hemisphere	Fiducial events	Identified nuclear recoils	Identified He recoils	Identified C/O recoils
(i)	145	Lid	$-$	129,775	2,406	1,348	1,058
(ii)	185	Bottom	$+$	159,947	2,370	1,311	1,059
(iii)	337	None	Uniform ϕ	19,134	36	30	6

Table 6.1: Summary of data collection runs. In the absence of the neutron source during campaign (iii), we expect neutron directions to not have a preferred ϕ hemisphere. With the source incident on the lid (bottom) of the TPC, we expect measured recoils to, on average, point in the $-x_{\mathrm{TPC}}$ ($+x_{\mathrm{TPC}}$) direction, which corresponds to the negative (positive) ϕ hemisphere. The particle ID procedure to determine the number of each recoil species is described in Chapter 6.2.

$47\,000\,\mathrm{e}$. An ^{55}Fe X-ray source is present inside the TPC which would nominally be used for an absolute energy-scale calibration, however the operating gain in our TPC is low enough that the majority of hits from ^{55}Fe X-rays fall below threshold, making our energy resolution too poor for events with this low of energy. We instead perform a comparison between simulated and measured He recoils to determine our gain and thus our energy scale. While we describe the process of determining the gain in Chapter 6.3, we note here that we found our gain to be 910 and use that for all event selections and analyses in this chapter.

6.2 Data processing and event selection

After calibrating the TPCs, we record data and process our data using the same procedure described in Chapters 3 and 4. We start by applying fiducialization cuts where we reject all events that contain any pixel hits along the outer perimeter of the FE-I4 and then make further nuclear recoil selections based on event energy and length. Figure 6.2 shows the energy versus length distribution of recorded events during Campaign (i) (source-on-lid), with the nuclear recoil selection boundary used in all three Campaigns drawn in red. All events in the gray shaded region below the red boundary are rejected as background. The right panel in this figure is a zoomed in version of the left panel, where we observe that below $12\,\mathrm{keV_{ee}}$ it is difficult to distinguish between X-ray backgrounds and nuclear recoils. We thus set a minimum signal threshold of $12\,\mathrm{keV_{ee}}$ and define additional quadratic selections to ensure a relatively pure sample of nuclear recoils with minimal stray noise hits.

Next we wish to exclude the less-directional C and O recoils. Given the approximate quadratic dependence of track energy versus length in the energy regime shown in Figures 6.2 and 6.3, we draw the constant selection boundary shown in Figure 6.3 of $E/\ell^2 = 19.5\,\mathrm{keV_{ee}/mm^2}$ and classify all events below this boundary as He recoils and all events above this boundary as C or O recoils. The right panel of this figure shows our final He recoil sample for Campaign (i) in blue and the rejected C and O sample in orange. We apply this same selection procedure in Campaigns (ii) and
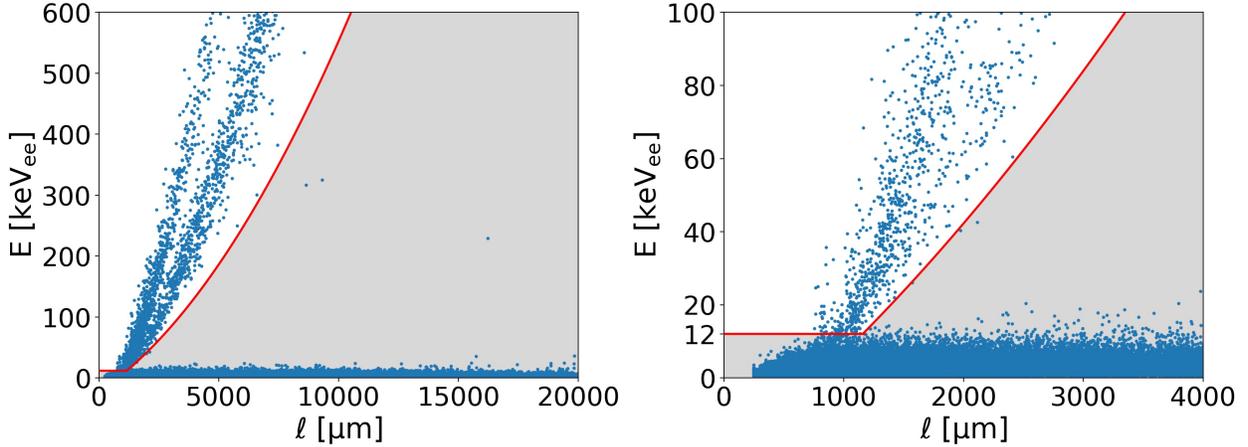
Figure 6.2: Event energy versus length along the principal axis during Campaign (i).

(iii).

6.3 Neutron beam simulation and gain determination

We use the ParticleGun module in Geant4 to generate two samples of simulated neutrons mimicking the geometry of Campaigns (i) and (ii). In both simulation campaigns, we simulate a neutron beam where neutrons are shot in a cone of width $3°$ in both θ and ϕ to approximate the collimation of neutrons emitted from the ^{252}Cf source. We did not measure the source location or diameter of the opening on the actual source, so we do not expect our estimation to be accurate. In both campaigns, the beam vertex is located at $z_{\text{TPC}} = 8.0\,\text{cm}$, $y_{\text{TPC}} = 0.84\,\text{cm}$, and in Campaign (i), the x_{TPC} position corresponds to just outside of lid of the TPC and in Campaign (ii), the x_{TPC} position corresponds to just outside of the bottom plate of the TPC. The energy spectrum of the neutron beam does not match the energy spectrum of the shielded ^{252}Cf source, so we expect there to be some substantial differences in overall head-tail performance due to differences in the recoil energy spectra, but assuming we select a relatively pure sample of He recoils in measurement, when we compare the head-tail performance versus energy of measurement and simulation, we expect simulation to be fairly representative of measurement, as we only keep truth He recoils in simulation.

To calibrate gain, we generated several samples of simulated He recoils at various gains and selected the gain where simulated dE/dx distributions approximately match. Figure 6.4 shows comparisons of the energy versus length distributions of simulated and measured nuclear recoils with the Campaign (i) setup assuming a gain of 910 both in measurement and simulation. We find that these distributions approximately agree, so we freeze our gain at 910 in both measurement and simulation. We apply the same recoil selection boundary used on measurement to our simulation
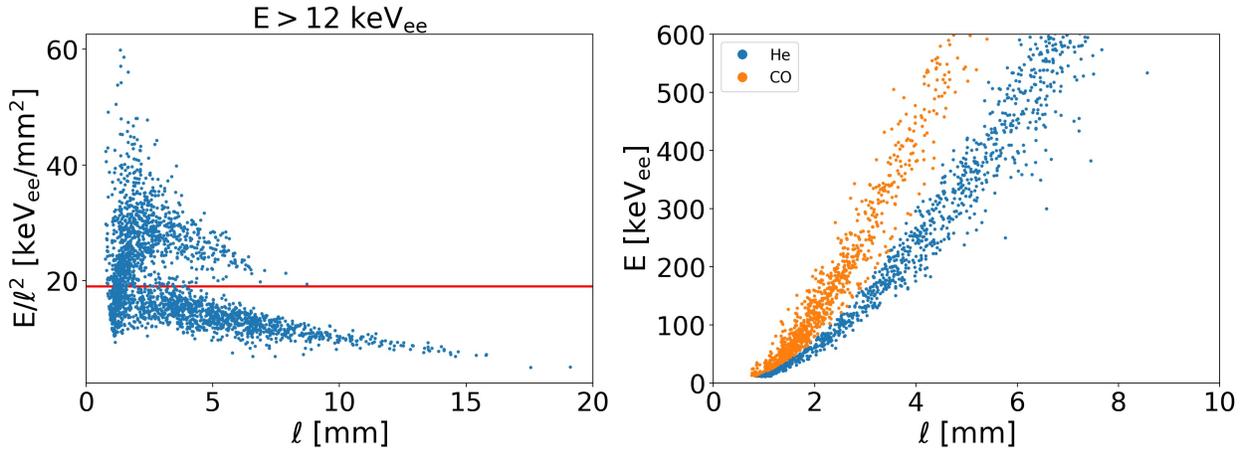
Figure 6.3: Discriminating between He recoils and C/O recoils. Left: E/ℓ^2 versus ℓ for all recoils satisfying our initial X-ray rejection selections. Recoils below the red boundary line are the He recoils we keep in this analysis. Right: Same events as the left panel plotted as an E versus ℓ distribution. Events below the red boundary in the left panel are identified He recoils (blue) and events above the red boundary in the left panel are identified C and O recoils (orange).
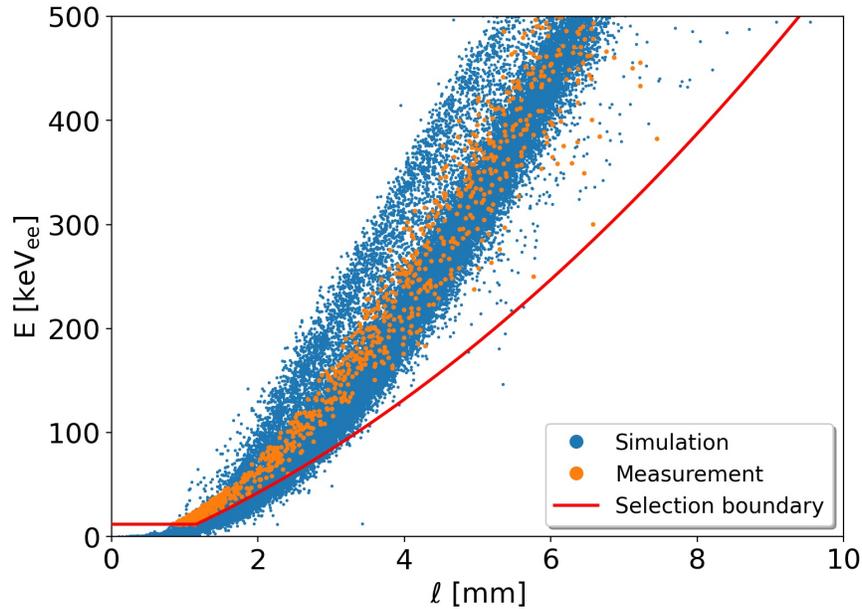
Figure 6.4: Energy versus length along principal axis distributions of measured and simulated tracks. Both measurement and simulation assume a gain of 910. Simulated tracks include only truth He recoils.

sample to make as direct of a comparison between measurement and simulation as we reasonably can.

6.4 Processing 3DCNN inputs

We process both measured and simulated events as 2-tuples containing (i) a $(34 \times 170 \times 34)$ sparse Pytorch tensor, and (ii) the associated class-label of the event. We choose the same voxelgrid dimensions as what was used in Chapter 5.6 so that we can evaluate head-tail for both the measured and neutron-beam simulated samples using the same 3DCNN that was already trained on isotropic simulated α recoils. We do not perform any additional training of the 3DCNN. Like before, we assign a label of 1 (0) to simulated events with true vector directions pointing in the negative (positive) ϕ hemisphere. Since we do not know the true direction of measured recoils, we assign a null class-label at this stage. After forming these 2-tuples for all events, we pass each of these 2-tuples into the pre-trained 3DCNN for head-tail assignment.

6.5 Results

Figure 6.5 shows the 3DCNN output class-assignment probability distributions of our direction-specific measured and simulation samples. Looking at the bottom row of this figure, we observe the expected behavior in both measurement and simulation, where the output probability distribution strongly peaks at 1 for the Campaign (i) sample and peaks at 0 for Campaign (ii). Since we recorded so few background events over the 2 week Campaign (iii), we plot these distributions on a logarithmic scale in the top row of this figure so we can compare the background distribution with the source distributions. In the top left panel of this figure, we observe that the background sample is roughly symmetric as we would expect from an isotropic background and unlike the asymmetric source-on-lid/source-on-bottom distributions, indicating that the 3DCNN is correctly identifying the source when it is present. Comparing measurement and simulation, we see that there are proportionally more events with uncertain classifications (p_- near 0.5) in measurement than in simulation. This is most likely due to differing nuclear recoil energy spectra between measurement and simulation, where there is a larger proportion of higher energy events in simulation than in measurement. Despite this difference between simulation and measurement, the 3DCNN still does its job well and identifies the source location in the output probability distributions with strong peaking.

We've identified that the 3DCNN statistically points back to the correct source location (some examples of tracks are shown in Figure 6.6), so next we attempt to quantify the head-tail performance of the 3DCNN evaluated on measurement. While we don't know the true recoil direction in measurement, we expect the majority of recoils observed in the TPC when the source is incident on the lid (bottom) to be in the negative (positive) ϕ hemisphere. With this in mind, we come up
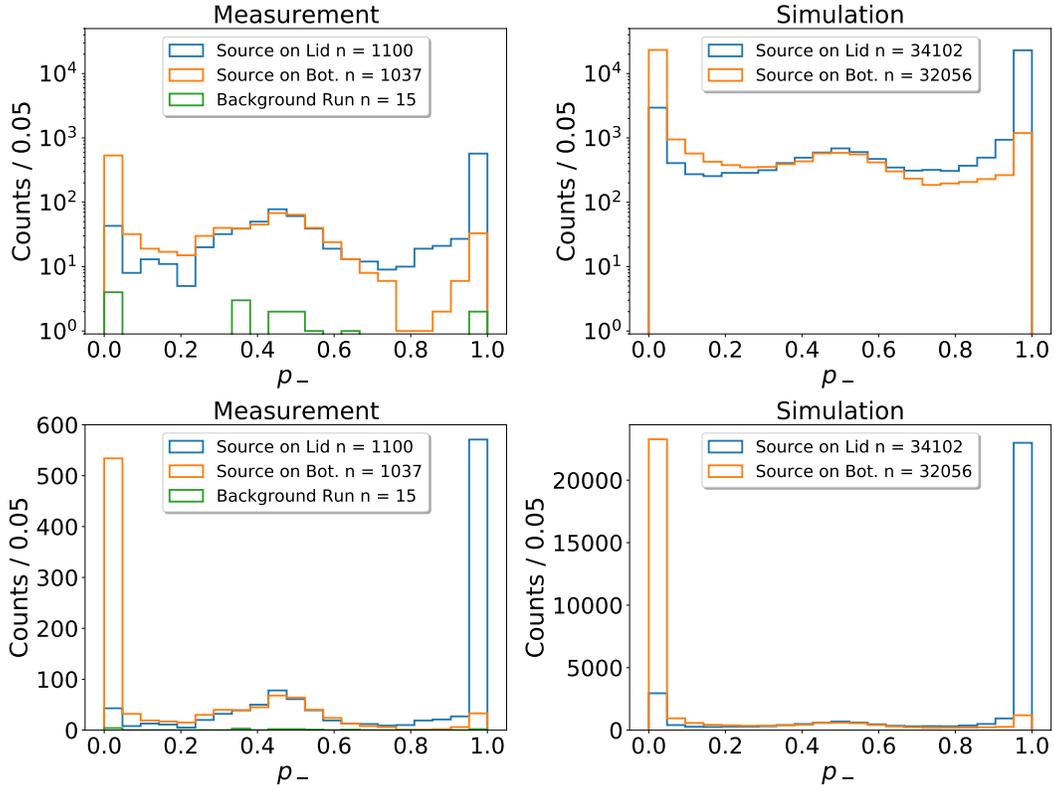
Figure 6.5: 3DCNN head-tail assignment for measured (left) and simulated (right) recoils. Individually the source-on-lid and source-on-bottom samples have strongly asymmetric p_- distributions in both measurement and simulation with the probabilities peaking in the expected direction of recoils coming from the source.

Source location	Lower limit correctly assigned measured events [%]	Lower limit correctly assigned simulated events [%]	True correctly identified simulated events [%]	Correctly assigned simulated events satisfying lower-limit criteria [%]
Lid	69.1 ± 2.5	80.6 ± 0.5	87.6 ± 0.5	98.6 ± 0.6
Bottom	83.8 ± 2.8	87.4 ± 0.5	90.2 ± 0.5	99.4 ± 0.6

Table 6.2: Columns from left to right: (i) Side of the TPC that the ^{252}Cf source was incident on. (ii) Percentage of measured events that are labeled as correctly assigned using our lower-limit assignment convention of $p_- > 0.5$ ($p_- < 0.5$) for the source-on-lid (bottom) sample. (iii) Same as (ii) but for our simulated samples. (iv) True percentage of correctly assigned simulated events. (v) Percentage of simulated events satisfying our lower-limit criteria $p_- > 0.5$ ($p_- < 0.5$) for the source-on-lid (source-on-bottom) sample with assigned directions that match truth. See text for more complete description of this column.
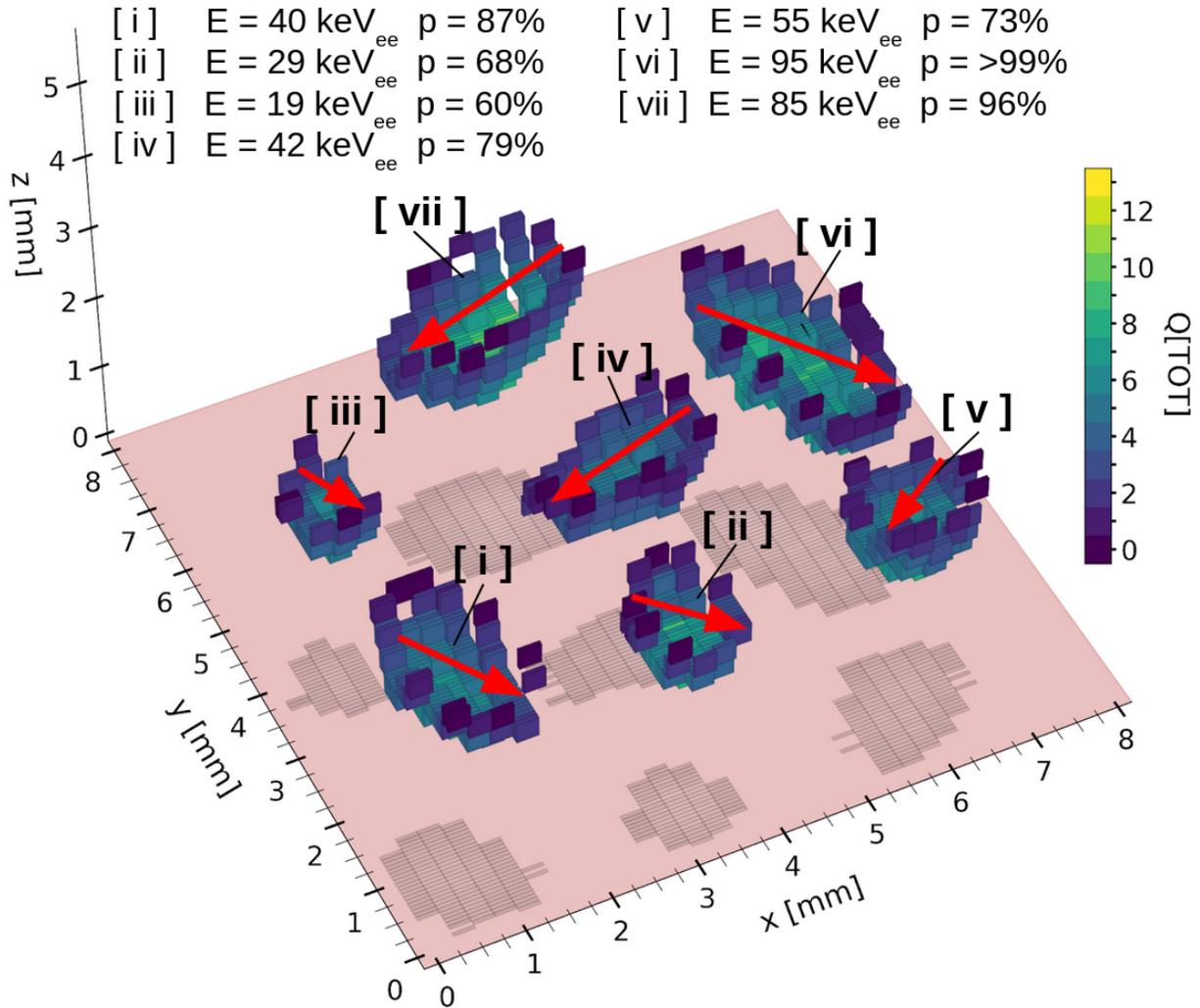
Figure 6.6: 3D event display of measured tracks with head-tail assigned by the 3DCNN [93]. Tracks [i]-[iii] and [vi] come from Campaign (i) (source-on-lid sample) and the remaining three tracks come from Campaign (ii). The vectors overlaid on top of each track have directions corresponding to ϕ_{SVD}. We note that angular resolution is poor below $\sim$40 keV$_{\mathrm{ee}}$ so the most important feature is whether these vectors point toward $+x$ or $-x$. The track energy and 3DCNN output probabilities are labeled at the top of the figure. For tracks from Campaign (i), the listed probability is p_- and for tracks from Campaign (ii), the listed probability is p_+. The track locations have been translated so that they all fit in a single display, but have otherwise not been altered.

with a lower-limit head-tail performance estimate by defining any event in the source-on-lid sample to be correctly assigned if $p_- > 0.5$ and any event in the source-on-bottom sample to be correctly assigned if $p_- < 0.5$. Table 6.2 describes the breakdown of these samples which we walk through in detail. Labeling the columns of this table from left to right as (i)-(v), comparing columns (iii) and (iv) tells us that our lower-limit head-tail selection criteria are indeed lower limits, as they produce lower head-tail assignment efficiencies than the true head-tail efficiencies. Furthermore, as is evident from Figure 6.5, there are proportionally more measured events that the CNN identifies with p_- in the neighborhood 0.5, so we observe a smaller fraction of events labeled with correct head-tail assignment in column (ii) than in column (iii), meaning these lower-limit assignments in measurement will in general show worse head-tail performance than they do in simulation. Column (v) gives the percentage of simulated recoils from the source-on-lid (bottom) sample with correctly assigned head-tail direction when the lower-limit selection criteria is applied. For instance, of the simulated source-on-lid (source-on-bottom) recoils satisfying $p_- > 0.5$ ($p_- < 0.5$), 98.6% (99.4%) of them have head-tail directions that match truth, suggesting this lower-limit selection criteria to be an excellent head-tail assignment criteria for evaluating performance in the absence of knowledge of the true recoil direction. We also notice that across the board, we get a lower correct assignment rate for the source-on-lid sample than the source-on-bottom sample. Similar to what we observed in Chapter 5.6, this asymmetry in performance goes away at higher energies, but to avoid biasing our low energy results within individual Campaigns, we combine the source-on-lid and source-on-bottom samples when evaluating overall performance versus energy.

We start by evaluating our overall performance with ROC curves. As before, we generate these ROC curves by determining $\varepsilon_{\mathrm{ht}}$ at $\varepsilon_{\mathrm{He}}$'s resulting from selections on w given in Equation (5.6). Here $\varepsilon_{\mathrm{ht}}$ for the measured samples is defined by the fraction of events with $p_- > 0.5$ in the source-on-lid sample and $p_- < 0.5$ in the source-on-bottom sample (Figure 6.7). We produce the same curve for simulation in the right panel of Figure 6.8, as well as the ROC curve that corresponds to the true $\varepsilon_{\mathrm{ht}}$. Figure 6.9 directly compares the ROC curve generated by measurement and the two ROC curves generated using simulation. Consistent with Table 6.2, we find that the lower-limit measured head-tail efficiency is lower than in simulation at $\varepsilon_{\mathrm{He}} = 1.0$, but as we increase the requirement on w, its performance eventually catches up to and surpasses the lower-limit simulation estimate.

Figure 6.10 compares the head-tail recognition efficiency of these samples as a function of energy. As before, the points within each energy bin show the head-tail efficiencies at 50% He-recoil efficiency of the sample within the given bin. Despite modest measurement statistics in most energy bins, we find acceptable agreement between $\varepsilon_{\mathrm{ht}}$ lower-limit performance in our combined (source-on-lid with source-on-bottom) measured and combined simulated samples. The trend is as expected in both the measurement and simulation lower-limit curves where head-tail performance in general increases with energy.

Comparing to other experiments, in 2008, the DRIFT-IIc detector using $40\,\mathrm{torr}$ of CS_2 reported
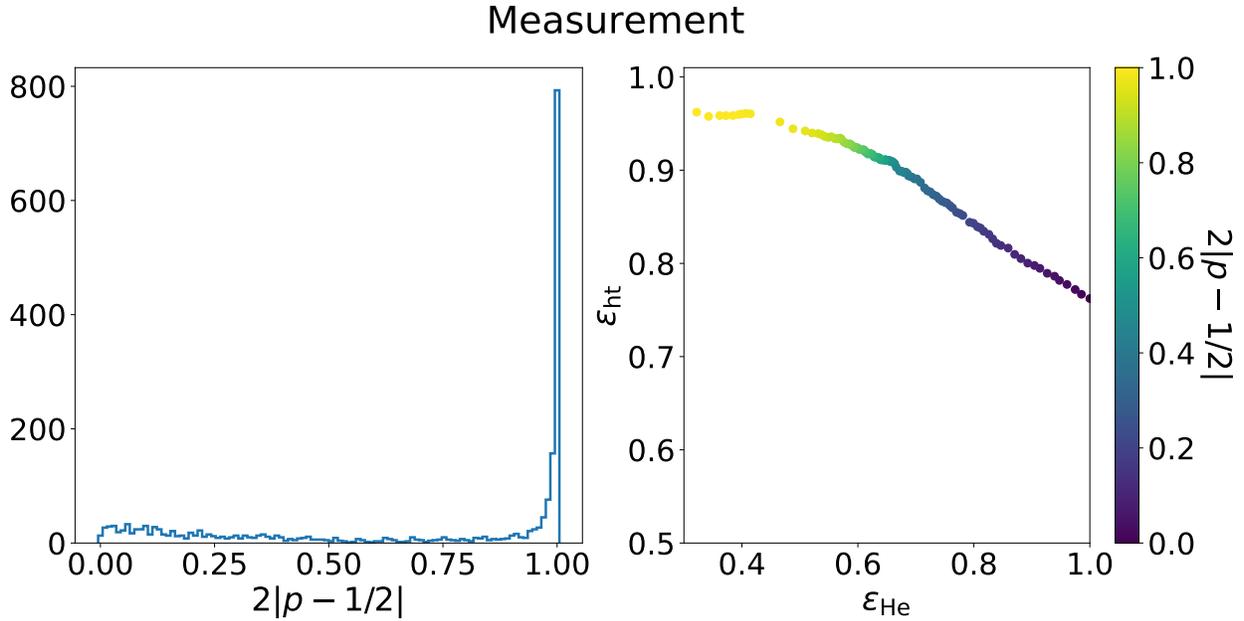
Figure 6.7: Left: w distribution of measured events. Right: ROC curve with $\varepsilon_{\mathrm{ht}}$ defined using the lower-limit correct assignment criteria described in the text.
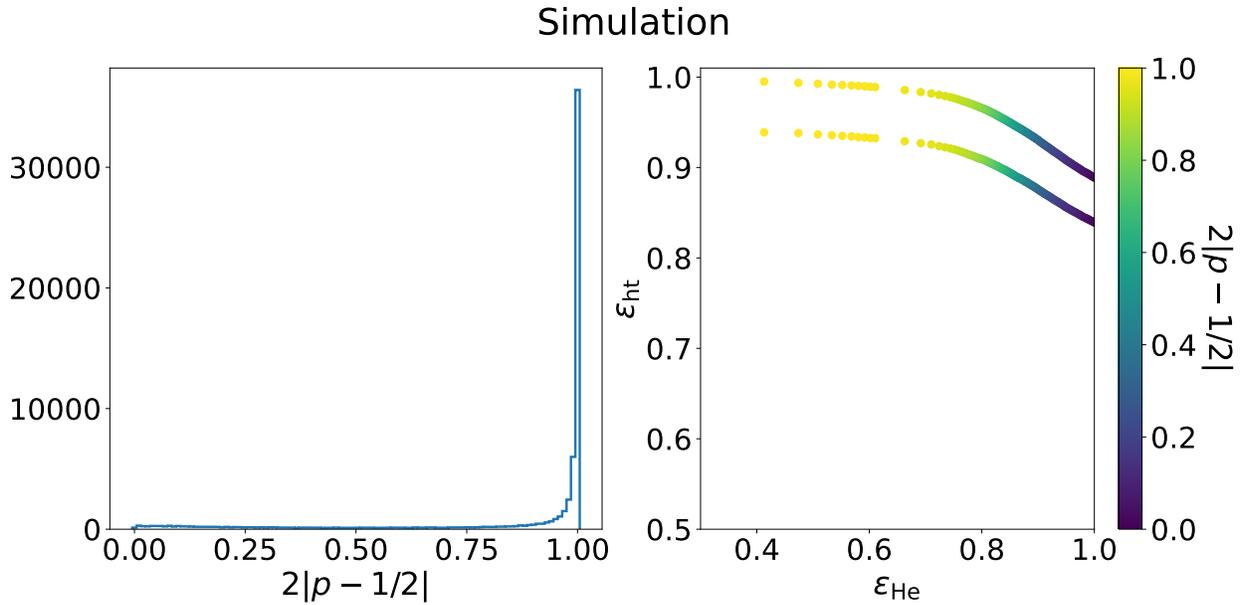
Figure 6.8: Same as Figure 6.7 but for simulation. The ROC curve with lower $\varepsilon_{\mathrm{ht}}$ uses the lower-limit correct assignment criteria, and the ROC curve with larger $\varepsilon_{\mathrm{ht}}$ shows the true head-tail recognition efficiency.
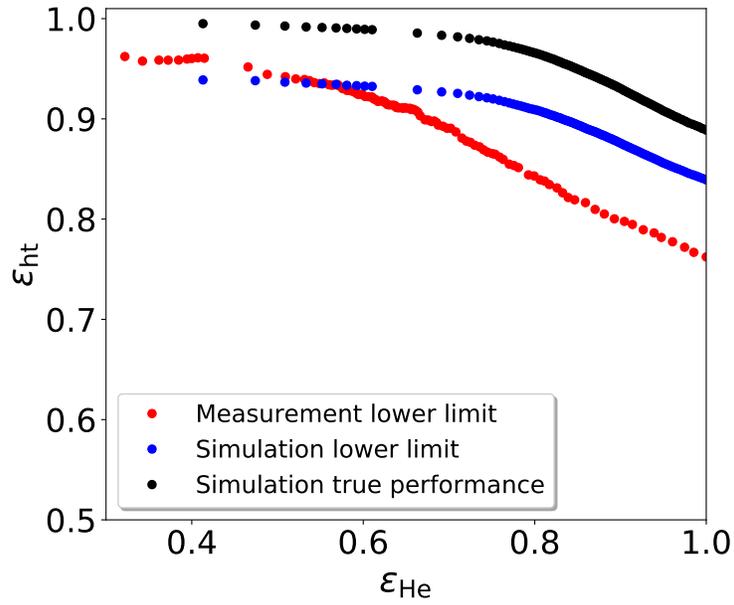
Figure 6.9: Direct comparison of the three ROC curves generated in Figures 6.7 and 6.8.
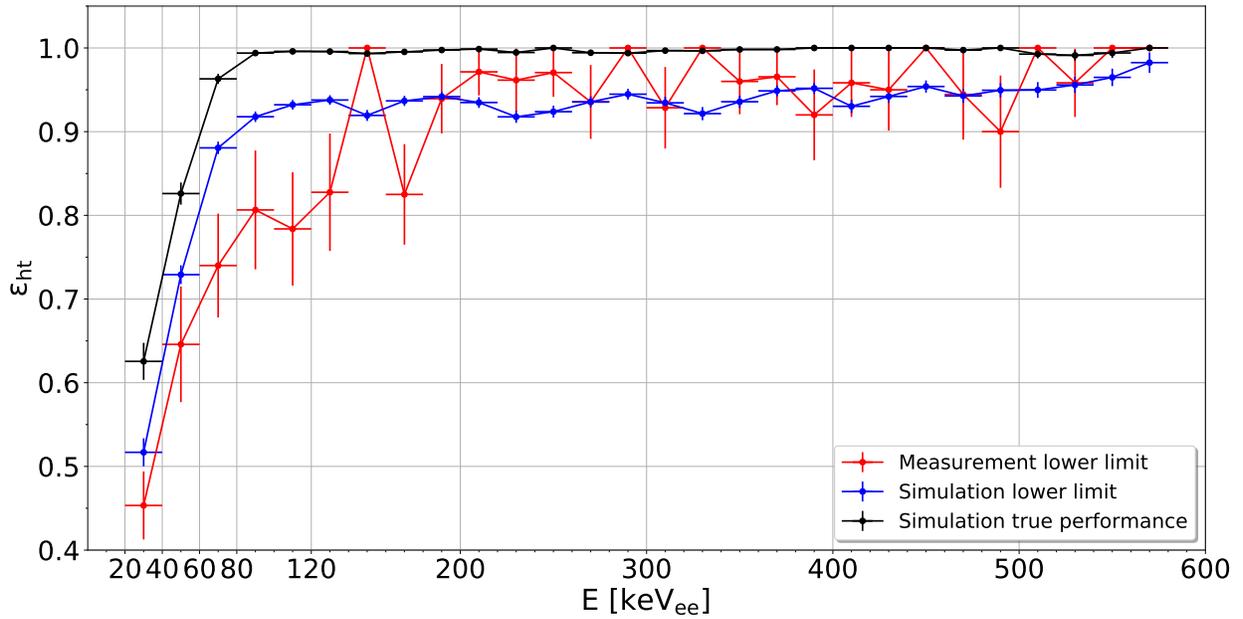
Figure 6.10: "Grand summary" comparison of head-tail assignment performance versus energy. Energy bins have width $20\,\text{keV}_{\text{ee}}$ and for the events within each energy bin, we make selections on w corresponding to 50% He recoil efficiency and compute the head-tail recognition efficiency.

head-tail sensitivity for sulfur recoils down to about 50 keV [27], and in 2020, the NEWAGE-03b detector filled with 76 torr of CF_4 gas, reported head-tail efficiencies of $(53.4 \pm 0.5)\%$ for 50-100 keV recoils, $(57.7 \pm 0.4)\%$ for 100-200 keV recoils, and $(65.1 \pm 0.5)\%$ for 200-400 keV recoils [28]. Both of these experiments used low pressure and low diffusion gas mixtures which are more favorable than our 1 atm $He:CO_2$ gas mixture for head-tail sensitivity. Nevertheless, our 3DCNN results are a considerable improvement over the DRIFT-IIC and NEWAGE-03b head-tail performance with lower-limit estimates of $\varepsilon_{ht} = (63.3 \pm 4.3)\%$ at 100% He recoil efficiency and $\varepsilon_{ht} = (73.7 \pm 5.2)\%$ at 50% He recoil efficiency for 50-100 keV_{ee} recoils. Furthermore, at 50% He recoil efficiency, we find 18 He recoils between 39 keV_{ee} and 49 keV_{ee} and determine a lower-limit head-tail recognition efficiency of $(62.1 \pm 11.4)\%$, indicating the first (albeit barely) statistically significant evidence of event-level head-tail recognition for sub-50-keV_{ee} nuclear recoils.

CHAPTER 7
HEAD-TAIL IDENTIFICATION PERFORMANCE AT HIGH GAIN: SIMULATION AND EXPERIMENT

7.1 Introduction

There isn't enough statistical structure in keV-scale recoil tracks for head-tail recognition at low avalanche gain, as a significant portion of charge in the recoil track will fall below the readout threshold. With the atmospheric pressure He:CO_2 mixture used by the BEAST TPCs, we need to be operating at near single electron efficiency to have any chance of observing the head-tail effect in keV-scale recoils. Even at single electron efficiency, there are a number of challenges to overcome in order to observe the head-tail effect. For instance, sub-10-keV He recoils at 1 atm gas pressure are short enough in length that even over relatively short drift lengths, the tracks diffuse enough that we have little to no angular resolution by the time the track is detected. This is further complicated by the effect of the pixel electronics shaping time, which turn such lowest-energy recoils into bowl-like shapes (Figure 6.6). Traditional principal axis-based methods for assigning axial direction and for determining head-tail assignment will break down in that regime. Nevertheless, at high gain, detected recoils in this energy regime register enough pixel hits that it may be possible for a 3DCNN to reliably identify head-tail. In principle, a 3DCNN can learn to account for biases due to effects such as diffusion and shaping electronics. If the 3DCNN approach is successful, the benefits would not only be improved performance, but also much easier development for the end-user compared to traditional deconvolution approaches that must be custom-tailored to gas conditions and electronics settings.

7.2 Experimental setup, calibrations, and simulation

To test for head-tail sensitivity at high gain, we use the same experimental set up as with low gain and again run three campaigns: Campaign (i) with the ^{252}Cf fast neutron source incident on the lid of the TPC, Campaign (ii) with the source incident on the bottom of the TPC, and Campaign (iii) with no source present. We take careful consideration of the location of the hole where the neutron beam from the source exists the plastic shield and set it at $z_{\text{TPC}} = 8.0$ cm. In practice there is a trade-off between saturation and diffusion, where with more diffusion (longer drift lengths) we expect relatively fewer saturated pixels, so while angular resolution may improve with the source opening closer to the charge readout-plane of the TPC, our energy resolution would suffer due to larger fractions of saturated pixels in events. We found that a drift length of $z_{\text{TPC}} = 8.0$ cm led to little or no saturation for $6\,\text{keV}_{\text{ee}}$ recoils at our operating gain (determined in Chapter 7.2.2).

7.2.1 Charge calibration

The finite dynamic range of the FE-I4 readout is a serious consideration when operating at high gain. Lowering the FE-I4 threshold by a factor of two, for instance, should lead to a proportional lowering of the lowest gain that corresponds to single electron efficiency. A factor of two drop in gain, then, leads to a substantial decrease in fractions of saturated pixels we would observe for events of a given true recoil energy. Lowering the pixel threshold, however, comes at the cost of an increase in the fractional uncertainty of the threshold in each pixel due to noise.
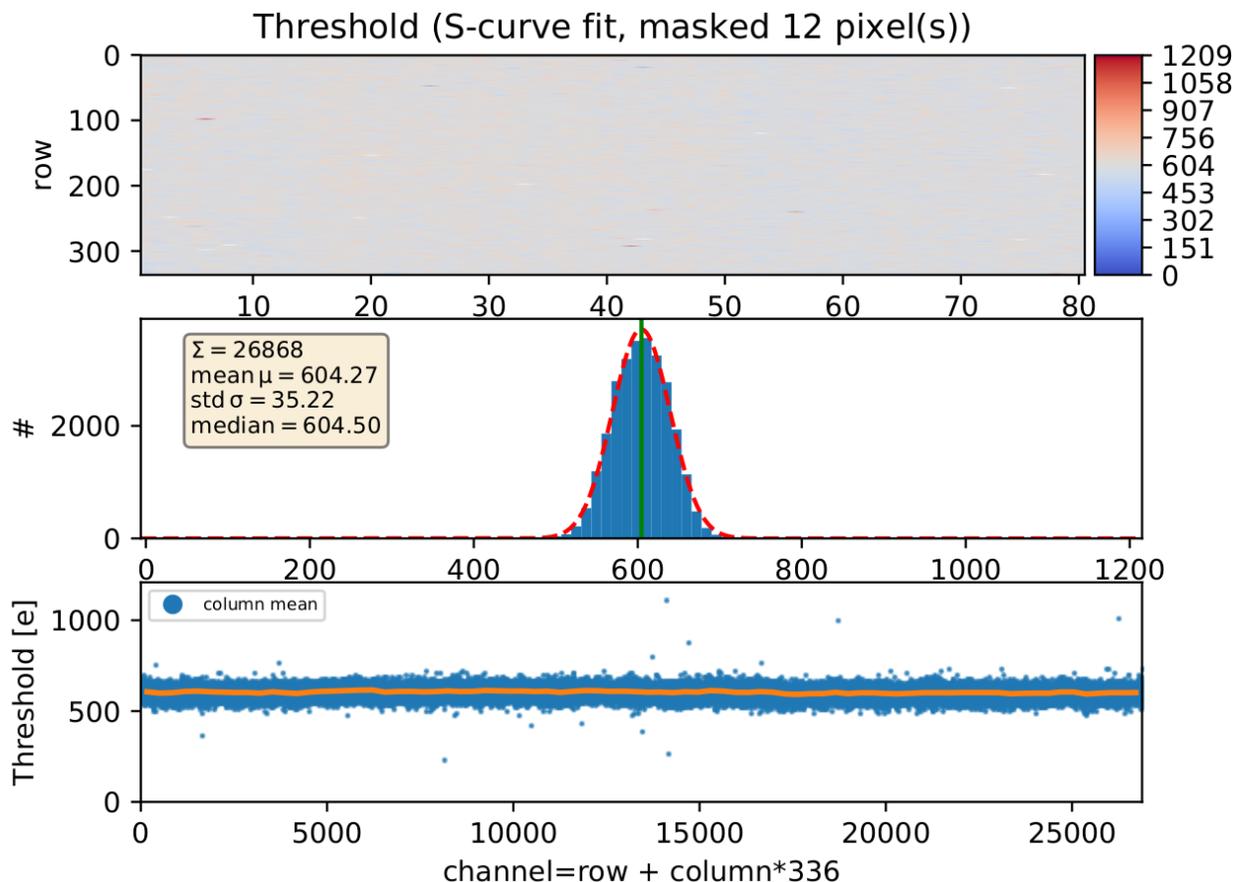
Figure 7.1: Top: Row versus column of the FE-I4 pixel grid with the color scale showing the threshold of each pixel. Middle: Threshold charge distribution with a fitted Gaussian profile. The variance in threshold is in general low. Bottom: Threshold distribution plotted versus 1D pixel number.

Figures 7.1 and 7.2 show the threshold and noise distributions for all pixels about two months after tuning the chip to a target threshold of 625 e per pixel. We see that the average threshold over all pixels is about 3% lower than this target. Comparing the threshold charge distribution and noise charge distribution histograms in these figures, we find that the mean noise is about 30%

of the mean threshold. Ideally we would want to have a higher threshold to noise ratio (we had > 10 for the Phase 2, Phase 3, and low gain head-tail studies), but we hope this trade-off is worth lowering the gain required for single electron efficiency.
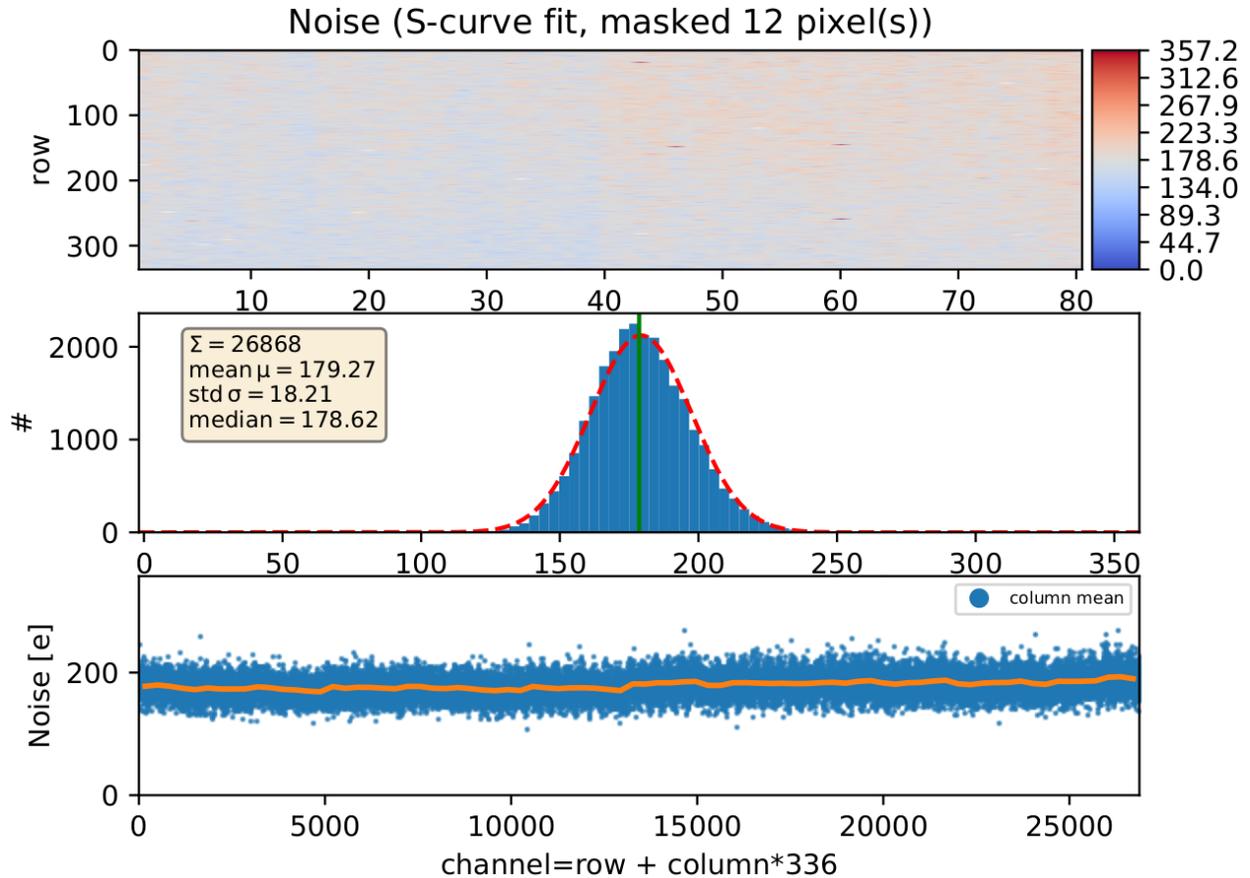
Figure 7.2: Same as Figure 7.1 but for noise distribution. The mean noise is about 30% of the mean threshold leading to a large uncertainty in threshold due to noise.

When tuning the FE-I4, we attempt to tune to the highest possible saturation limit to maximize the dynamic range of our readout. When adjusting the target dynamic range, we found that lowering the threshold also caused the saturation limit to lower, despite attempting to tune to a saturation limit of $50\,000\,\mathrm{e}$. Figure 7.3 shows the result of measuring the mean and standard error of TOT codes over 100 injections at each of 70 distinct charge steps. We see that injected charge corresponding to $\mathrm{TOT} = 13$ (saturation limit) is below our target of $50\,000\,\mathrm{e}$, but conclude this is acceptable since the lower gain resulting from the lower threshold improves our energy resolution at single electron efficiency over further increasing the raw saturation limit.

To calibrate our charge (TOT) scale, we perform a bicubic spline interpolation over all data points in Figure 7.3, and then record the interpolated electron charge at each integer TOT step.
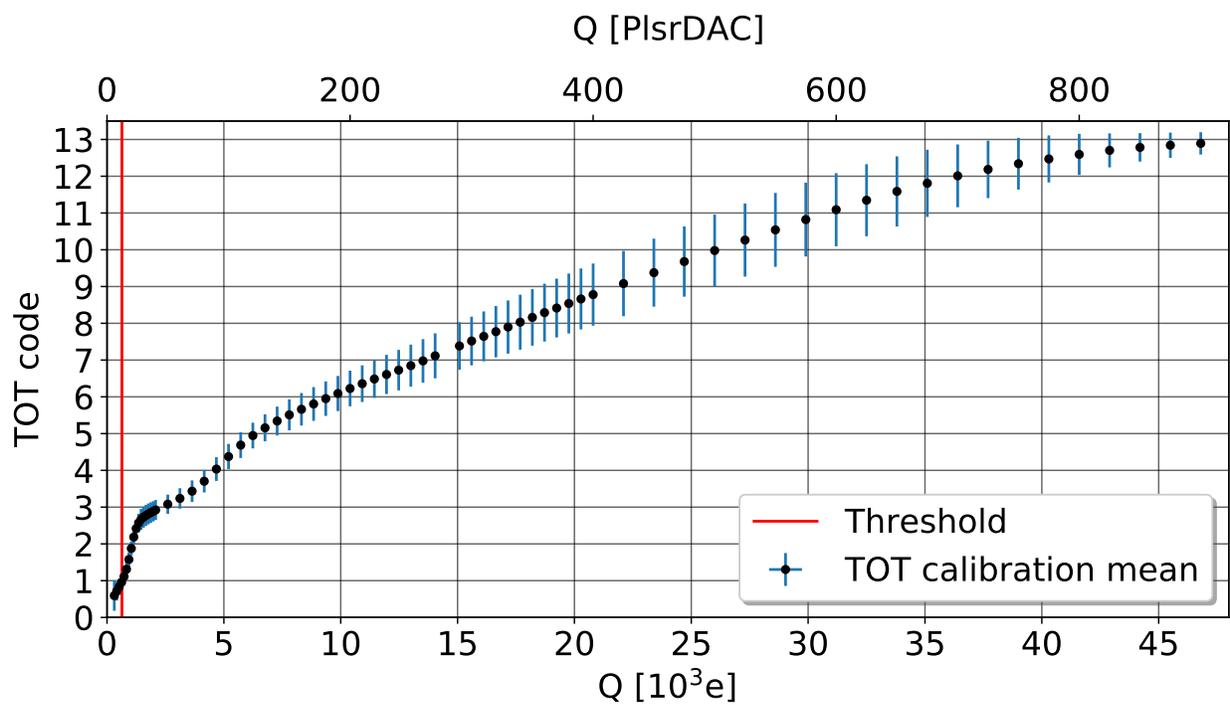
Figure 7.3: Average TOT response to 100 charge injections in all pixels over 70 distinct charge steps (charge steps are represented by points). The vertical red line shows our threshold which is what we set TOT = 0 to be.

The interpolated curve along with the charge threshold is shown in Figure 7.4. Notice that the charge difference corresponding to pixel hits registering TOT $= 0$ and TOT $= 1$ is very small. As TOT increases, the difference in charge between subsequent TOT values also increases, so our TOT to charge response is not linear. This was also the case in all previous studies.
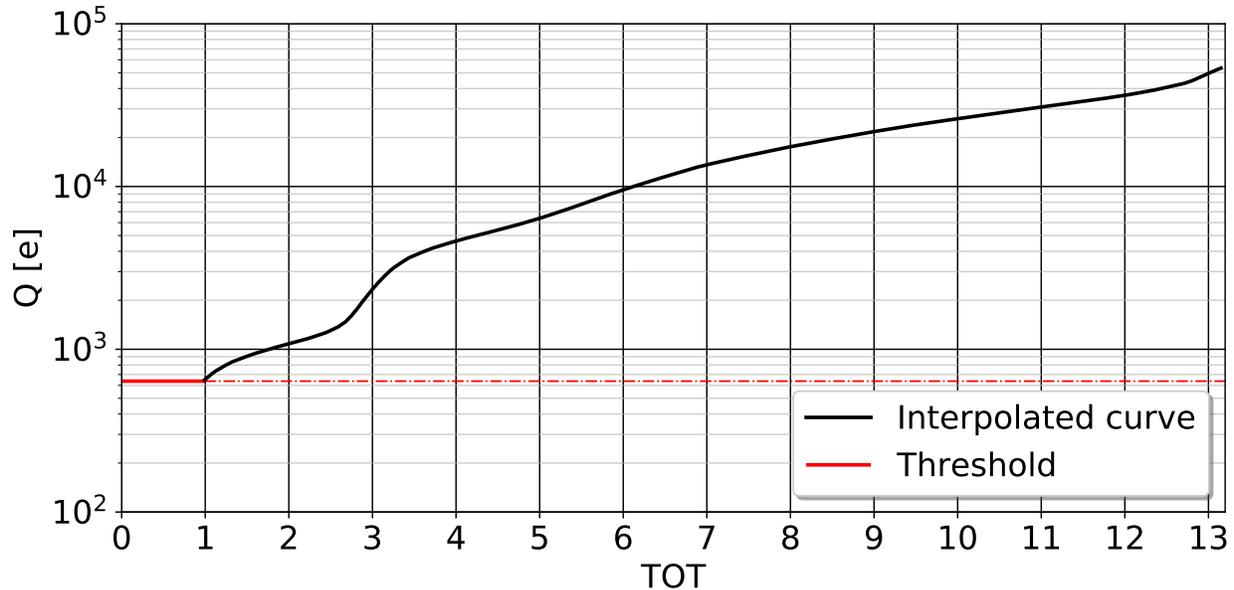
Figure 7.4: Bicubic spline interpolation over the points shown in Figure 7.3 shown as a black line. The charges where the interpolated curve matches an integer TOT code are what we set as the calibrated charge for a given integer TOT code.

7.2.2 Gain calibration

Our goal is operate near single electron efficiency with a calibrated energy scale. Before deciding our target gain point, we first need to ensure that the gain is stable. We evaluate gain stability by performing 15 min runs at constant GEM high voltage settings, where we alternate between exposing the active volume of the TPC to X-rays from an ^{55}Fe source and not. To control the ^{55}Fe source exposure, we engineered a remote controllable Delrin shutter that we close every other run, allowing us to assess the purity of the ^{55}Fe X-ray sample when the shutter is open so that we can use its energy spectrum to assess gain stability. The goal is to achieve gas quality where the effective gain (which includes reduction due to gas impurities) versus time has plateaued after sealing the TPC vessel. Performing 115 hours of such runs with the TPC sealed and our 70:30 mixture of He:CO$_2$ gas flowing at 12.5 sccm, we achieve $> 99\%$ of the predicted maximum gain after about 96 hours of operation. The details of this gain stability study are in Appendix B.

After we have flowed gas long enough to reach a sufficiently stable gain, we then use the rate of

cosmic rays observed over fixed time exposure runs at various GEM high voltage settings (which translate to different uncalibrated gains). Since cosmic rays are minimally ionizing, we plot the average rate of cosmic rays as a function of effective gain, and define the minimum single electron efficiency gain point to be the lowest gain where this distribution is sufficiently close to the nominal maximum cosmic ray rate.
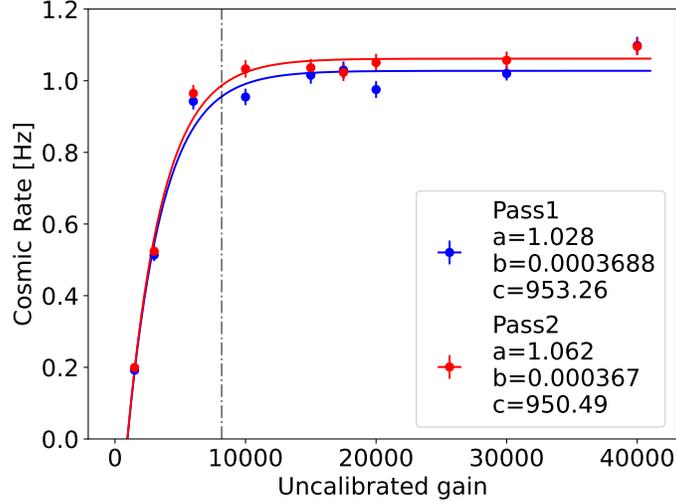
Figure 7.5: Average cosmic rate over 30 data collection periods versus uncalibrated gain. The FE-I4b threshold on average is aroun SI600e. Pass 1 starts with runs at low gain and dials high voltage settings gradually upward to high gain. Pass 2 starts at high gain and ramps high voltage settings down to lower gain. The continuous curves are fits to these data using Equation (7.1). The vertical dot-dashed line shows the gain point of $G_{\mathrm{uncal,set}} = 8170$ that we determined to correspond to 95% of the cosmic ray rate predicted at single electron efficiency.

Figure 7.5 shows the results of this study. To control for gain variations due to any residual time dependence in the gain due to variations in gas-purity, we perform two passes of runs: One pass where we start with low gain and gradually increase the gain, and a second pass where we start at high gain, and gradually decrease the gain. Each data point shown in this figure is the mean cosmic ray rate recorded over 30 minutes with the ^{55}Fe source shutter closed. The uncalibrated gain, G_{uncal} shown on the horizontal axis of this figure is determined using the figure of merit (B.1) described in Appendix B. To determine our single electron efficiency gain point, we fit an exponential of the form

$$R(G_{\mathrm{uncal}}) = a \left(1 - e^{-b(G_{\mathrm{uncal}}-c)} \right), \tag{7.1}$$

to our distribution of cosmic ray rates, R versus G_{uncal}. Fit parameter a represents the nominal

cosmic ray rate at single electron efficiency and fit parameter b represents the rise time of cosmic ray rates versus gain. We set our near-single electron efficiency gain point to be

$$G_{\mathrm{uncal,set}} = \frac{3}{b}, \tag{7.2}$$

as $3/b$ corresponds to $(1 - 1/e^3) \approx 95\%$ of the nominal maximum cosmic ray rate. We choose to use the fit to Pass 2 to set our uncalibrated gain to $G_{\mathrm{uncal,set}} = 8170$ (shown as the vertical dot-dashed line in Figure 7.5), as the difference in gain between Passes 1 and 2 is effectively negligible, but Pass 2 predicts a slightly higher gain.

Now we determine our calibrated gain using a sample of ^{55}Fe X-rays collected over 8 hours at $G_{\mathrm{uncal,set}} = 8170$. We follow a variant of the procedure described in Appendix B to determine the GEM divider and field cage input voltages needed to give us this target gain. The energy spectrum of ^{55}Fe X-rays peaks at $5.9\,\mathrm{keV}$ in He:CO_2 [76], so lining up the peak of the uncalibrated energy spectrum at $G_{\mathrm{uncal,set}} = 8170$ with $5.9\,\mathrm{keV}$ provides an absolute energy calibration. To ensure that this energy calibration is accurate, we first need to test that there isn't a significant portion of saturated pixels in these X-ray tracks.
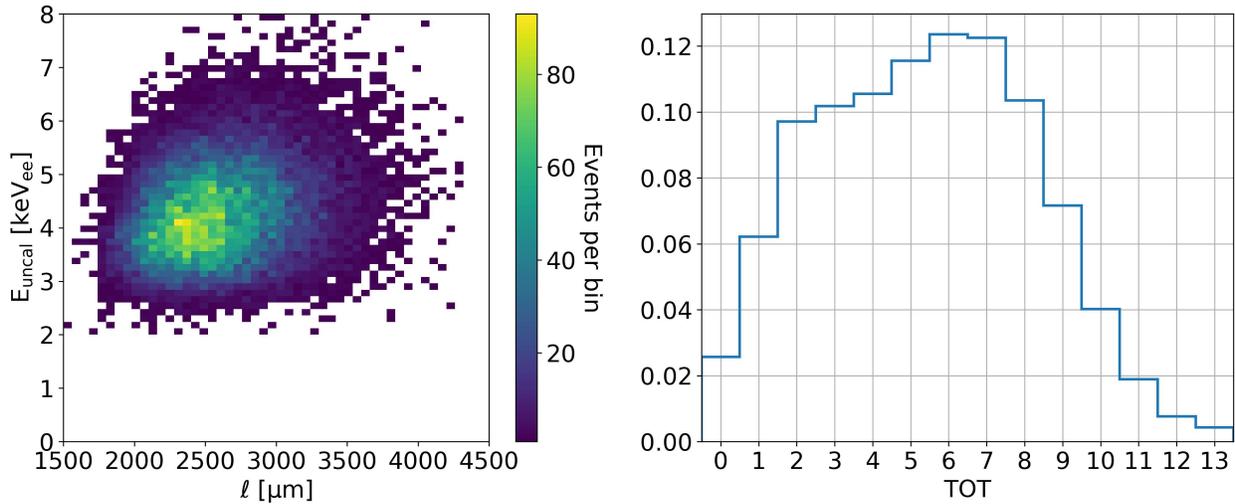
Figure 7.6: Left: Uncalibrated energy versus track length distribution of ^{55}Fe X-rays recorded over this 8 hour period. Right: Normalized TOT distribution of all pixel hits from this sample. The low frequency of TOT = 13 hits suggests that events on average do not have significant fractions of saturated pixel hits.

From Figure 7.6 we observe that the normalized TOT distribution (right panel) over all events in the sample shown in the left panel of the figure doesn't have an excessive peak at TOT = 13, suggesting that we are not limited by saturation in this X-ray sample. Additionally, since we are

operating close to single electron efficiency, we expect that very few hits after amplification will fall below our energy threshold, suggesting that we're operating at near optimal energy resolution for this gain calibration. We finally determine our calibrated gain, G, by identifying the peak of the uncalibrated energy spectra (red vertical line penetrating through the red histogram in Figure 7.7), and shifting that peak to line up with $5.9 \, \mathrm{keV_{ee}}$ (blue vertical line penetrating through the blue histogram in Figure 7.7). Doing this, we find a calibrated gain of $G = 13400 \pm 400$.
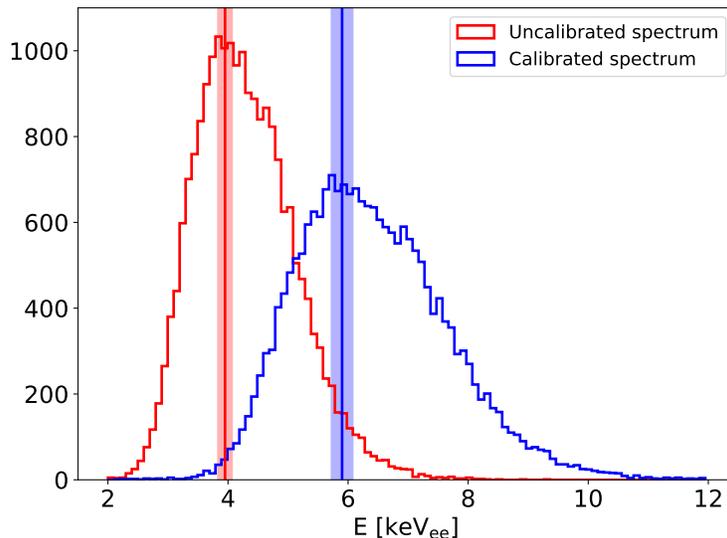
Figure 7.7: Uncalibrated (red) and calibrated (blue) energy spectrum of the ^{55}Fe X-rays. The pale red and blue vertical bands show the uncertainty in the peak location.

7.3   Simulation

We use the same procedure described in Chapter 5.2 to generate samples of neutrons mimicking the geometry of Campaigns (i) and (ii). Particle identification for low energy recoils is considerably more challenging than for the higher energy recoil sample we considered in our low gain study. As a result, we also aim to train a 3DCNN to identify recoils and reject X-ray backgrounds, so we simulate a sample of X-ray backgrounds. To generate this background, we use the basf2 implementation of the Geant4 ParticleGun module to fire electrons with a uniform momentum distribution isotropically from the center of the fiducial volume of the TPC. We additionally generate an isotropic sample of α tracks to augment our neutron beam simulation sample, so that we don't bias our head-tail training with just directional neutron-recoil samples. All recoil samples are drifted and digitized using the digitizer introduced in Chapter 4.2. We digitize using the parameters shown in Table 7.1. We use a gain of 15,000 instead of 13,400 and also add a constant offset of $2600 \, \mathrm{e}$ to the TOT to charge mappings used for measured recoils to get better agreement between measured and simulated

Gas [70%:30%]	W [eV]	Gain	$\vec{v}_d$ [µm/BCID]	(σ_T, σ_L) [µm/$\sqrt{\text{cm}}$]	$(\sigma_{T,\text{GEM}}, \sigma_{L,\text{GEM}})$ [µm]
He:CO$_2$	35	15,000	216.25	(134.8, 128.2)	(143,97)

Table 7.1: Digitization parameters for the high gain head-tail simulation campaign. $\sigma_{T,\text{GEM}}$ and $\sigma_{L,\text{GEM}}$ represent the transverse and longitudinal point resolution of the readout plane, excluding the pixel chip. The diffusion parameters were computed using `Magboltz` and should be more representative of measurement than the diffusion coefficients used in Table 5.1.

dE/dx distributions (details in Chapter 7.4 and Figure 7.18).

Ideally we would train our head-tail 3DCNN classifier using isotropic He, C, and O recoils, and then evaluate the trained 3DCNN on a directional neutron-recoil sample like what we did in Chapter 5, but the ParticleGun module we used to generate events does not generate heavy ions, so the only way we were able to simulate C and O recoils out of the box was by simulating neutron beams. Reliably rejecting C and O recoils for the event samples in this study is very difficult, so we cannot simply ignore C and O recoils in our 3DCNN head-tail training and reject those events as backgrounds at the analysis stage like we did in Chapter 5.

7.3.1 Electron rejection on simulation

Data samples and processing for 3DCNN

We combine all recoils from the samples listed in Table 7.2 (and plotted in Figure 7.8) into a single data set and shuffle the events so that they are randomized when fed into the 3DCNN. The true nuclear recoil signal to X-ray background ratio in our measured sample is considerably lower than what we generated for MC, but as long as the 3DCNN has adequate statistics of both to learn features that discriminate recoil constituents well, then differences in signal to noise ratios between simulation and measurement shouldn't matter. Figure 7.9 shows the reconstructed energy versus length distribution of these simulated events. The highest dE/dx band is a mixture of all three nuclear recoil species that are difficult to distinguish (we're interested in distinguishing He recoils from C or O recoils) using selections on E versus ℓ. The lower dE/dx band is primarily composed of electron recoils from X-rays. At energies below $4\,\text{keV}_{\text{ee}}$, the nuclear recoil and X-ray bands bleed together and are indistinguishable using selections on energy and length.

Like before, we store the events as 2-tuples where the first entry is a $(34 \times 170 \times 34)$ sparse `PyTorch` tensor, and the second entry is the class-label of the event. For event classification, we assign labels of 0, 1, and 2 for truth electrons, truth He recoils and truth C or O recoils, respectively.

3DCNN architecture and training

We use a nearly identical architecture for our event classification 3DCNN (Table 7.3) as our previous 3DCNNs used for head-tail assignment, with the main difference being that we now have three

Recoil species	Simulation sample	# Events
e	Isotropic	692,139
He	Isotropic	101,739
He	Source-on-lid	16,099
He	Source-on-bottom	15,663
C/O	Source-on-lid	120,138
C/O	Source-on-bottom	117,845

Table 7.2: Number of simulated recoils generated for each recoil constituent and each generation procedure. Reconstructed ionization energy spectra for these samples are shown in Figure 7.8.
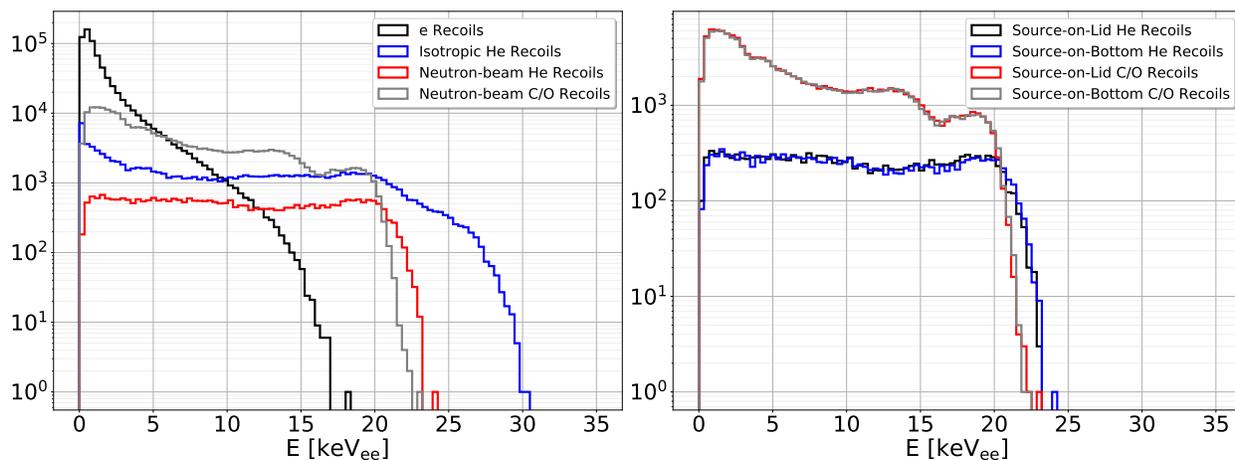
Figure 7.8: Left: Reconstructed ionization energy spectra of simulated e, He, and C/O recoils. The neutron-beam samples show a combination of source-on-lid and source-on-bottom samples. Right: Neutron-beam reconstructed ionization energy spectra with the source-on-lid and source-on-bottom samples distinctly shown. The energy spectra of the source-on-lid and source-on-bottom samples appear consistent. Truth ionization energies were not logged for the electron recoil sample, so we opt to plot reconstructed ionization energies for all samples.
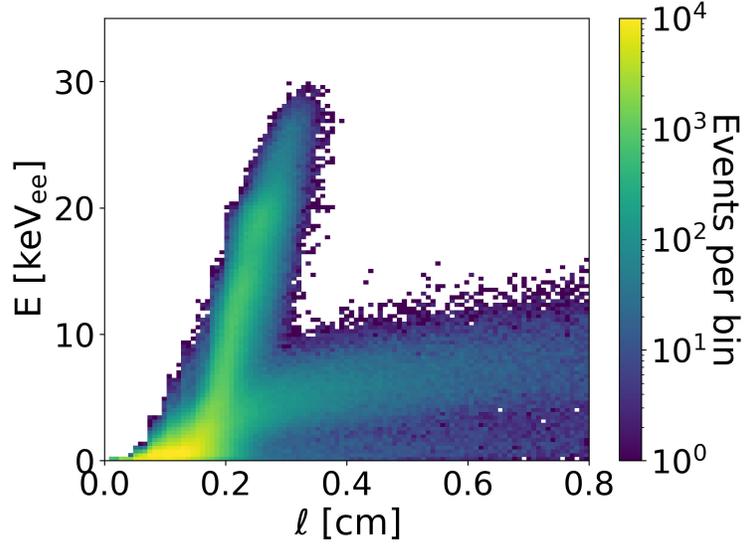
Figure 7.9: Reconstructed energy versus length of all simulated recoils in the high-gain campaign.

Layer	# Filters	Filter size	S	P	# Learnable parameters	Dropout	Output shape
ConvBlock 1	8	$(2 \times 2 \times 2)$	(1,2,1)	0	88	0.03	$(8 \times 33 \times 85 \times 33)$
ConvBlock 2	16	$(3 \times 3 \times 3)$	(1,2,1)	0	3,504	0.03	$(16 \times 31 \times 42 \times 31)$
ConvBlock 3	32	$(2 \times 3 \times 2)$	1	0	6,240	0.03	$(32 \times 30 \times 40 \times 30)$
AvgPool	32	$(2 \times 2 \times 2)$	2	0	—	—	$(32 \times 15 \times 20 \times 15)$
ConvBlock 4	64	$(2 \times 3 \times 2)$	1	(1,0,1)	24,768	0.03	$(64 \times 16 \times 18 \times 16)$
ConvBlock 5	32	$(3 \times 3 \times 3)$	1	(1,0,1)	55,392	0.03	$(32 \times 16 \times 16 \times 16)$
AvgPool	16	$(2 \times 2 \times 2)$	2	0	—	—	$(32 \times 8 \times 8 \times 8)$
ConvBlock 6	16	$(2 \times 2 \times 2)$	1	0	4,144	0.03	$(16 \times 7 \times 7 \times 7)$
ConvBlock 7	8	$(2 \times 2 \times 2)$	1	0	1,048	0.03	$(8 \times 6 \times 6 \times 6)$
FC1	—	—	—	—	110,656	0.05	(64×1)
FC2	—	—	—	—	1,040	0.05	(16×1)
FC3	—	—	—	—	51	—	(3×1)

Table 7.3: Architecture of the 3DCNN used for particle ID at high gain. Note the similarity to the architecture of the 3DCNN used for head-tail assignment (Table 5.2) except FC3 now has three output classes instead of two, and the dropout used in the ConvBlocks was reduced to 0.03.

output classes instead of two. Applying a SoftMax function to the model output now gives us output probabilities of p_e, p_{He}, and p_{CO}, which are the model classification probabilities of electron recoils, He recoils, and C or O recoils, and satisfy $p_e + p_{He} + p_{CO} = 1$. We split our combined sample that includes all recoil constituents into a training set with 654,256 events, a validation sample with 150,952 events, and a test sample with 214,572 events and train using stochastic gradient descent (see Chapter 5.6.4 for a description of how this works) with minibatch updates of size 128 and cross-entropy loss and use the same early stopping criteria as before.

Results

Figure 7.10 shows comparisons of distributions of length, and the three output probability hypotheses, p_e, p_{CO}, and p_{He} for the test set. We immediately see that the 3DCNN classifier gives significantly stronger separation with much higher signal yields than our traditional method of event selection on track length.

To further quantify event selection performance, we introduce a rejection factor, R, which we define as

$$R = \frac{N_{bg}}{N'_{bg}}, \tag{7.3}$$

where N_{bg} is the total number of background candidates in a sample and N'_{bg} is the number of background candidates that remain after a particular selection. The top panel of Figure 7.11 shows the results of *electron rejection factor*, R_e versus energy, where in each energy bin, we make selections that lead to 50% nuclear recoil (mixture of He and C and O) efficiency. The blue points use selections on p_e and the gray points use selections on length. We find that electron rejection performance is orders of magnitude better using the 3DCNN over length.

Ultimately, for head-tail studies, we're interested in selecting maximally pure samples of He recoils, so we next evaluate rejection factors R_e and R_{CO} for selections on p_{He} corresponding to 50% He recoil efficiency, which are shown in Figure 7.12. With these selections, we find that electron rejection remains above 100 at all energies (red points) and find that CO rejection (blue points) increases exponentially with energy. From these results, we anticipate that the fractional compositions of CO and He in measurement may further limit directional sensitivity at low energies. For instance, if we record a factor of 10 more C/O recoils than He recoils at $8\,\mathrm{keV_{ee}}$, then even after selections with the performance shown in Figure 7.12, N'_{He}/N'_{CO} would be less than 0.5. Nevertheless, if the 3DCNN event classifier performs remotely similar to our measured samples as it does on simulation, then at higher energies, we should expect very high purity He-recoil samples.
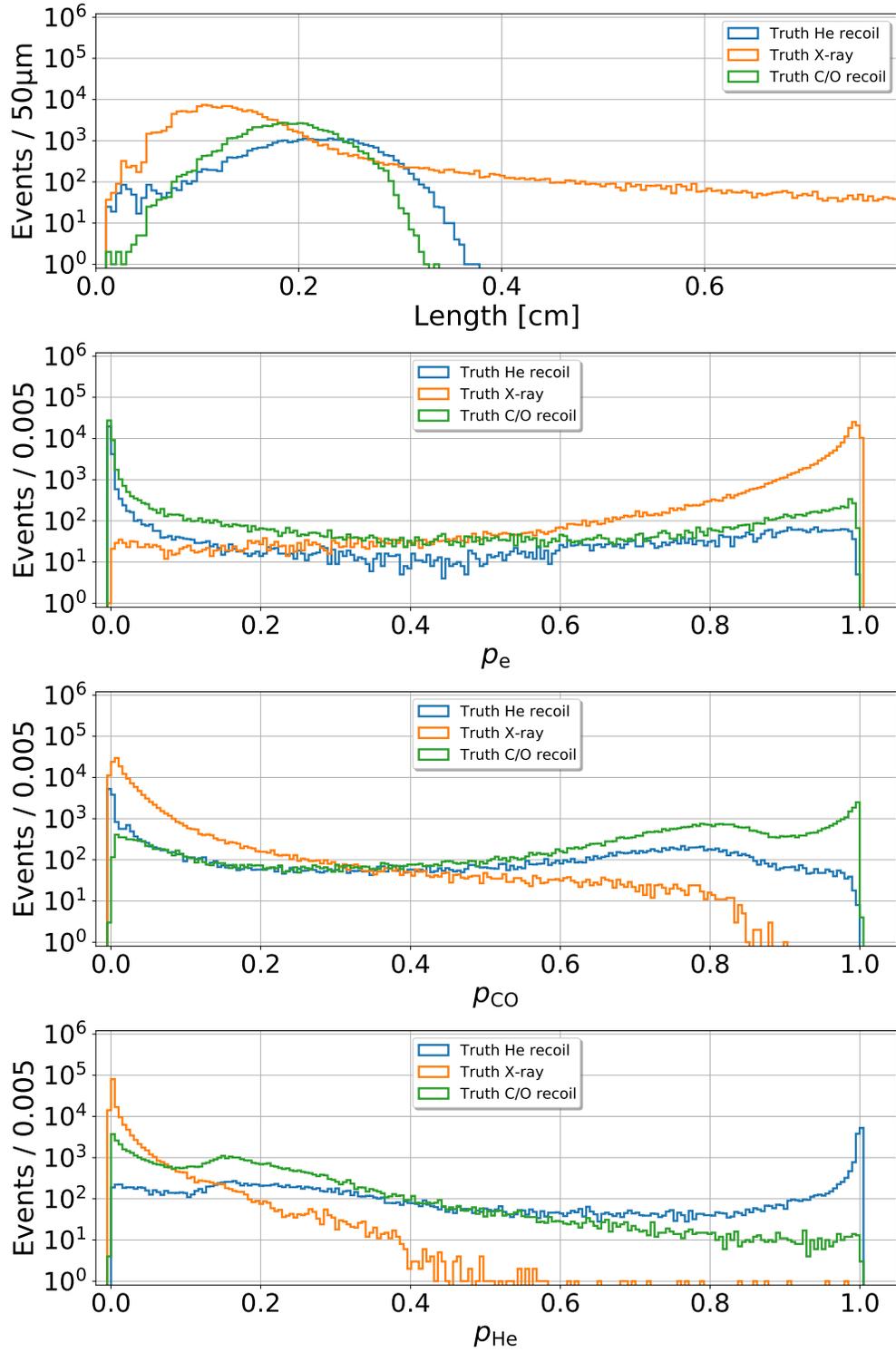
Figure 7.10: Truth recoil distributions of length and 3DCNN output probability hypotheses.
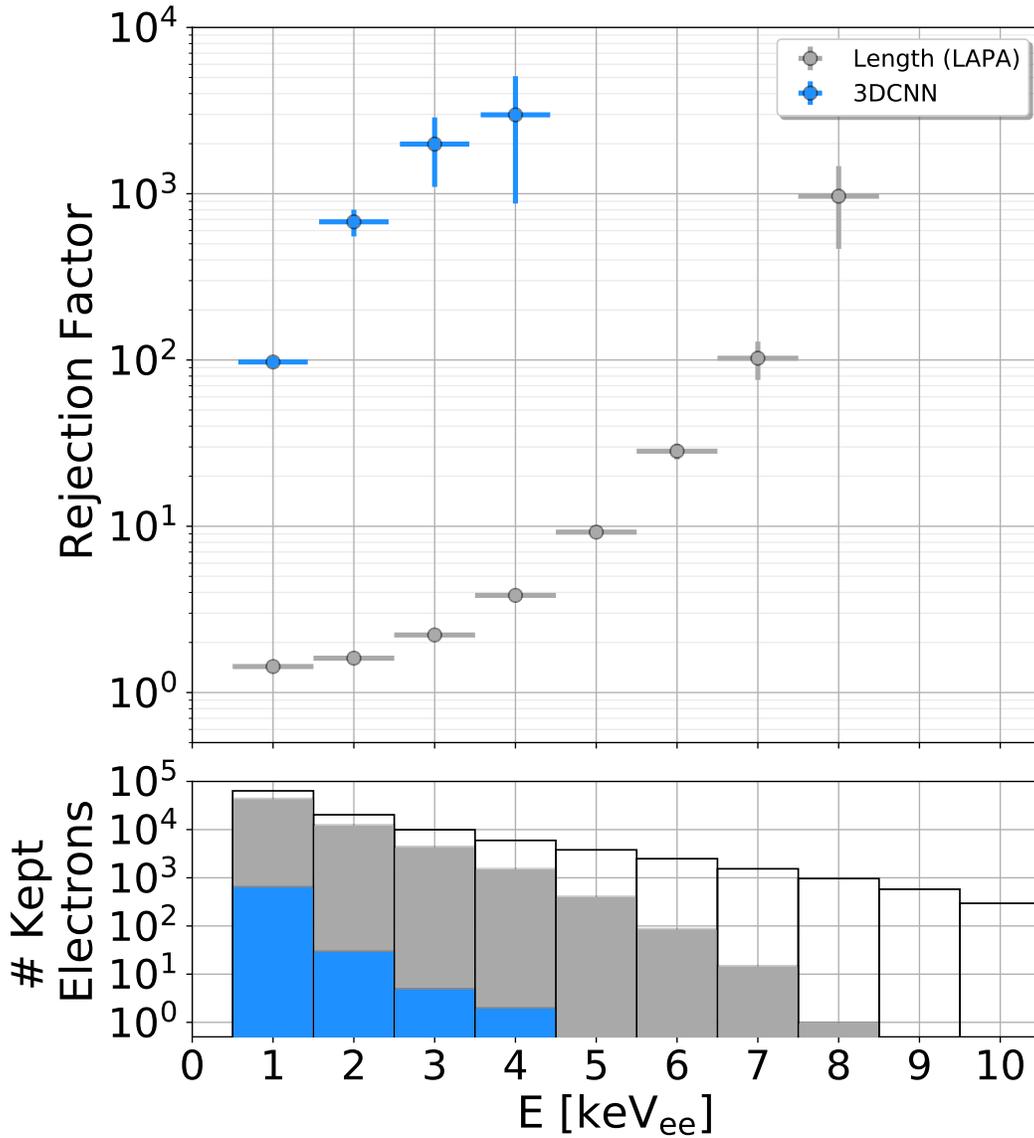
Figure 7.11: Top: Electron rejection factor versus energy using selections that correspond to 50% nuclear recoil efficiency. If a point is not shown on the plot, it means all electrons were rejected. For instance, there are no blue points above $4\,\mathrm{keV_{ee}}$, indicating that all electrons were rejected using selections on p_e corresponding to $\varepsilon_{\mathrm{He}} = 0.5$. Note: in the legend, LAPA stands for "Length along principal axis" and is our metric for track length. Bottom: Number kept electrons, N'_e after selections on length (gray) and p_{e} (blue). The clear bars show the total number of electrons in the energy bin.
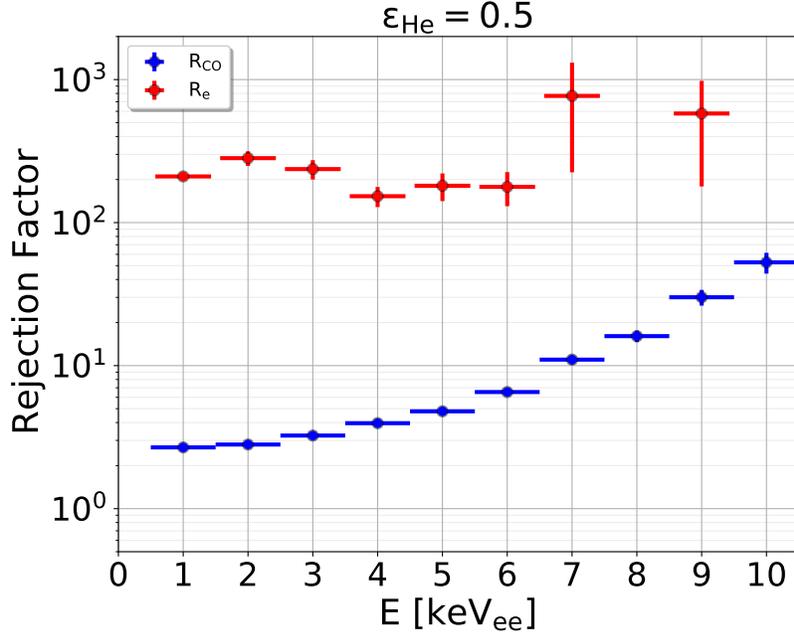
Figure 7.12: Rejection factors R_e and R_{CO} versus energy using selections of p_{He} that correspond to $\varepsilon_{He} = 0.5$ in each energy bin. The 8 and $10\,keV_{ee}$ bins have all electrons rejected.

7.3.2 Head-tail on simulation

3DCNN architecture and training

We use the exact same architecture as our low gain head-tail classifier described in Table 5.2, except we increase the random dropout from 0.05 to 0.15 in each of the convolutional blocks. Using this higher dropout value led to better performance on the test set once the network was trained. We speculate that the reason for this is because the features used for head-tail extraction are subtle enough that with too low of a dropout, the network may pick up on misleading features that are sub-optimal for head-tail classification. Increasing the randomized dropout increases the amount of training iterations required before early stopping is triggered but it also reduces the chance of overfitting.

When we evaluate high gain performance on measurement, we first feed our data through the 3DCNN trained for particle identification to obtain a pure sample of nuclear recoils and then assign head-tail to the nuclear recoil candidate sample using a separate 3DCNN. Since we aim to reject all backgrounds before evaluating head-tail, we train our high gain 3DCNN head-tail classifier using only the He and C/O recoil samples listed in Table 7.2. We combine each recoil type into a single set of data, shuffle the order of events, and then split the sample up into training, validation, and test sets containing 232,177, 53,579, and 85,728 events, respectively. As with the low gain exercise,

we process each event as a 2-tuple containing a $(34 \times 170 \times 34)$ tensor and an associated class-label corresponding to the ϕ_{truth} hemisphere of the recoil track. We train the neural network using stochastic gradient descent with minibatch updates of size 128 using cross-entropy loss, and the same early stopping criteria as in Chapter 5.6. Applying a SoftMax function to the output gives p_- and p_+ with the same definitions as before.

Results

Since our training sample was a mixture of an isotropic α-recoil sample with samples of neutron-source simulation that include He, C, and O recoils, we evaluate our performance in three separate samples labeled, "Isotropic", "Source-on-lid", and "Source-on-bottom". Figure 7.13 shows the ROC performance of He recoils. As a whole, we find that performance is biased favorably toward the source-on-lid sample, which is evident both in the ROC curves and in Figure 7.14.
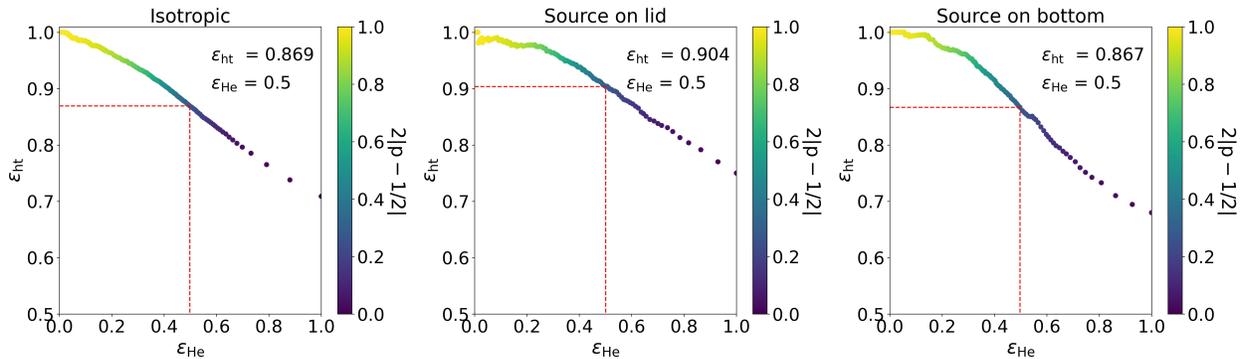
Figure 7.13: ROC curves for He recoil samples. Despite consistent recoil energy spectra between the source-on-lid and source-on-bottom samples (Figure 7.8), we find that performance is asymmetric and favors the source-on-lid sample.

We check for the same bias in our C/O recoil sample and find that the bias is even stronger in favor of identifying source-on-lid for these recoils than for He recoils (Figure 7.15). In all cases, the misclassification rates for events with high $w = 2|p_- - 1/2|$ is still very low, so this bias shouldn't pose a problem for events with high classification confidence.

We next evaluate performance as a function of energy using the same "Grand Summary" criteria defined earlier. Namely, we partition our samples into energy bins of width $2\,\text{keV}$ with bin centers ranging in steps of $2\,\text{keV}$ between $1\,\text{keV}$ and $35\,\text{keV}$. We do this both for our reconstructed energies (in keV_{ee}) and truth recoil energies (in keV_{r}). Then, within each energy bin, we compute ε_{ht} at selections on w that correspond to a nuclear recoil efficiency of 50%. We do this separately for the set of only He recoils, the set of only C/O recoils, and the set of all nuclear recoils combined. The results are shown in Figure 7.16, where for He recoils, we find head-tail efficiencies in excess of 80% down to $9\,\text{keV}_{\text{r}}$ and non-negligible event-level head-tail sensitivity down to $3\,\text{keV}_{\text{r}}$. The
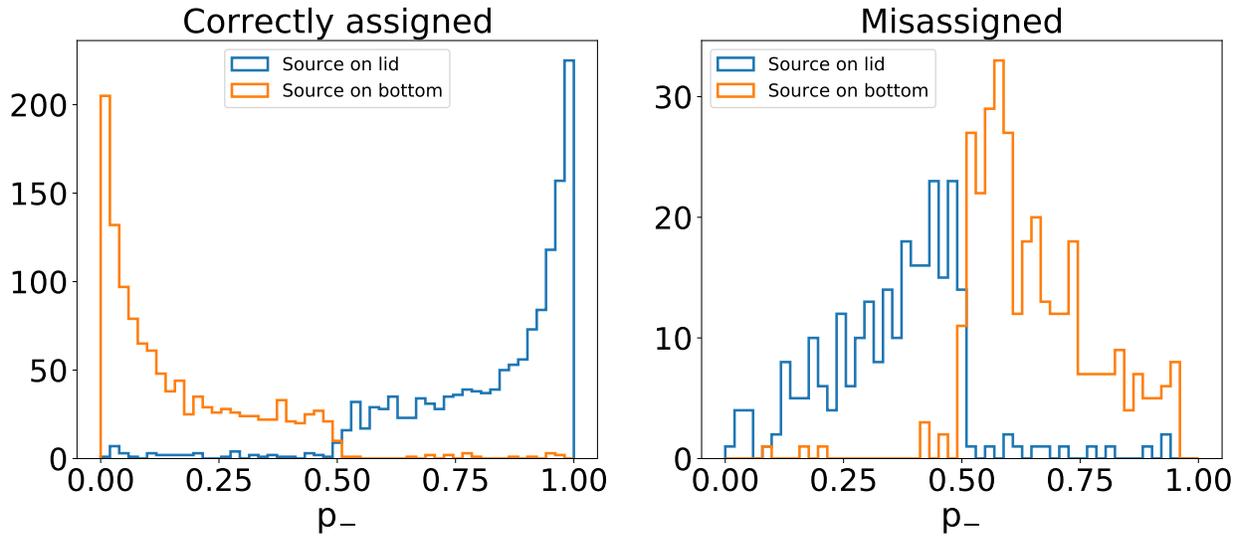
Figure 7.14: Illustration of the head-tail assignment asymmetry on simulated He recoils with reconstructed ionization energies between $10\,\mathrm{keV_{ee}}$ and $20\,\mathrm{keV_{ee}}$. The training sample is roughly symmetric in terms of number of events with truth direction in each ϕ_{truth} hemisphere, but the trained 3DCNN misassigns more events for the source-on-bottom sample than the source-on-lid sample.
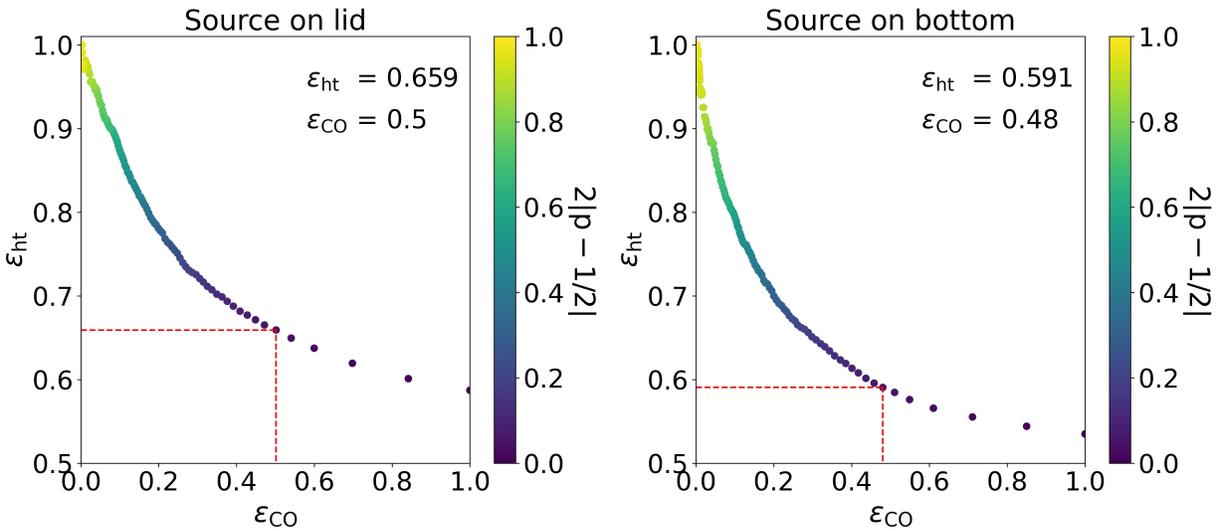
Figure 7.15: ROC curves for C/O recoil samples. Like with the He sample, despite nearly identical recoil energy spectra between the two samples (Figure 7.8), we find that performance is asymmetric and favors the source-on-lid sample.

135
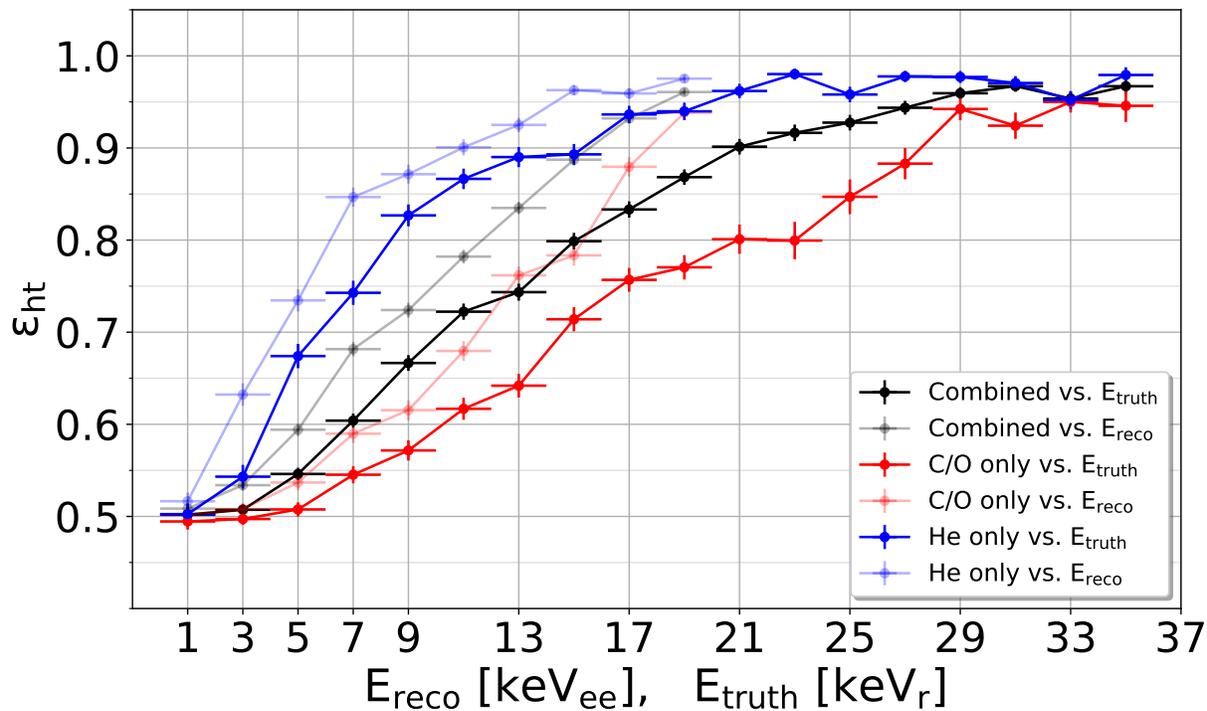
Figure 7.16: 3DCNN-determined head-tail recognition efficiencies on simulated nuclear recoils versus reconstructed nuclear recoil energy (pale colors; units in keV$_{\text{ee}}$) and true ionization energy (bold colors; units in keV$_{\text{r}}$). In each sample, we compute $w = 2|p_- - 1/2|$ that corresponds to a nuclear recoil efficiency of 50% and compute ε_{ht} after these selections.

3DCNN event selection performance for this high gain sample suggests that we should be able to obtain a reasonably pure sample of He recoils in this energy regime. At minimum, these results suggest that in the limit where we cannot distinguish He recoils from C/O recoils, our head-tail recognition efficiency will be about 66% at $9\,\mathrm{keV_r}$ (black line in Figure 7.16). Assuming our ability to distinguish He recoils from C/O recoils is better than chance like we have found, then we should expect considerably higher head-tail efficiencies that approach the He recoil-only performance. All together, if the results on measurement are at all similar to our 3DCNN event selection and head-tail classification performance holds on measurement, then this will be the first demonstration of directionality for keV-scale recoils.
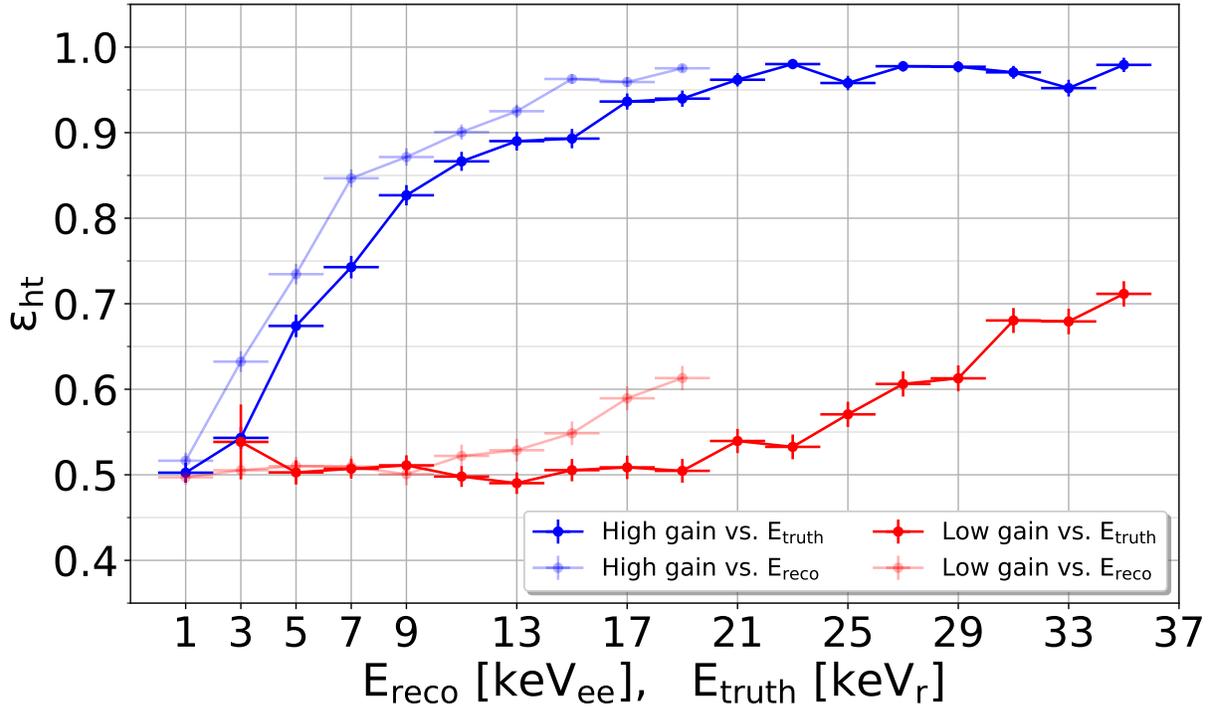
Figure 7.17: Comparison of 3DCNN-determined head-tail recognition efficiencies on simulated He recoils at low gain and high gain. $\varepsilon_{\mathrm{ht}}$ is computed using selections on w that correspond to 50% He recoil efficiency in each bin. The high gain results use the simulated isotropic He recoil sample described in this chapter and the low gain results use the simulated isotropic He recoil sample with digitization settings described in Chapter 5.

Figure 7.17 shows a direct comparison of head-tail assignment performance for the He recoil sample considered here versus the low gain He-recoil sample evaluated in Chapter 6. Both samples are evaluated with the appropriate 3DCNN. We find that at low gain, head-tail performance only starts exceeding a coin flip at $21\,\mathrm{keV_r}$ and doesn't pass the 70% efficiency mark until $35\,\mathrm{keV_r}$. On the other hand, at high gain, head-tail performance exceeds a coin flip at as low as $3\,\mathrm{keV_r}$ and

exceeds 70% by $7\,keV_r$, verifying that high gain operation is necessary for observing the head-tail effect for sub-10-keV_r nuclear recoils. We remind the reader that low gain head-tail recognition on measurement was not quite as performant as on simulation, so while these high gain results on simulation show promise, we must verify that they generalize to measurement.

7.4 Experimental results

Here we evaluate the results of both the high-gain particle ID 3DCNN and the head-tail assignment 3DCNN on measurements from all three campaigns. We recorded 166 hours of high-gain data during Campaign (i) (source-on-lid), 186 hours during Campaign (ii) (source-on-bottom), and 357 hours during Campaign (iii) (background data with no source present). In all of these campaigns the internal ^{55}Fe source shutter was closed, so we expect very few X-rays leaking from this source.

7.4.1 Electron rejection
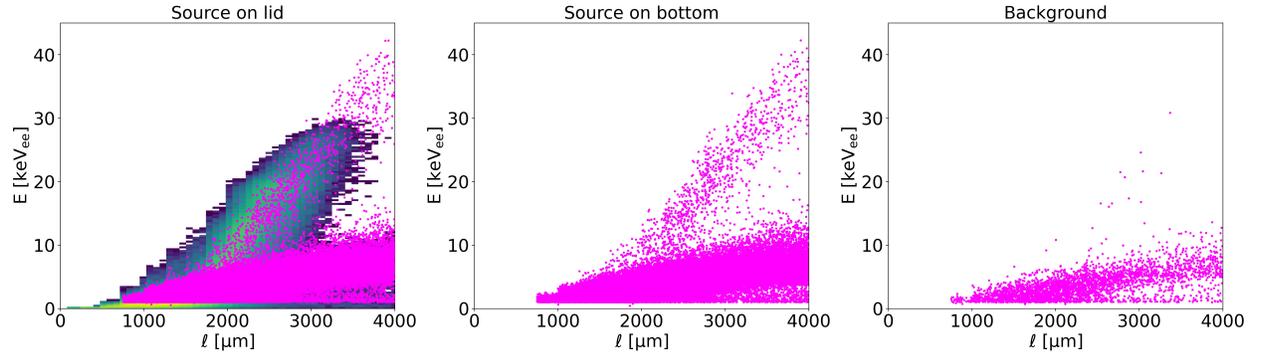
Figure 7.18: Observed recoil energy versus length distributions in each of the three measurement campaigns (magenta). The leftmost panel also shows a 2D histogram of simulated recoils with a logarithmic colorscale to highlight the agreement in dE/dx between measurement and simulation.

Figure 7.18 shows the distribution of measured recoils for each of the three campaigns. The leftmost panel shows measured points (in magenta) overlaid on top of a 2D histogram of simulated events (histogram uses a logarithmic colorscale), where we observe dE/dx agreement in both the nuclear recoil and X-ray bands. Moving forward, we restrict our event range to $E < 30\,keV_{ee}$, as we did not train our 3DCNN on higher energy events than this. Identifying events in the region where nuclear recoils and electron recoils are indistinguishable in terms of dE/dx is challenging.

We opt to select signal candidates using selections on p_e corresponding to 50% nuclear recoil efficiency on our simulated sample at $8\,keV_{ee}$. Figure 7.19 shows events identified as nuclear recoils with these selections in orange overlaid on top of the rest of the measurements in blue. At higher energies, we maintain high nuclear recoil efficiencies, however around $8\,keV_{ee}$, our efficiency appears

138

to be much lower than 50%, indicating that performance on simulation doesn't generalize well to measurement. Nevertheless, we continue with these selections, as dE/dx profile of the selected recoil sample appears to minimally overlap with the X-ray background band. A few events labeled as nuclear recoils with these selections appear suspect due to their relatively longer length and lower energy, however it's possible that these events are correctly identified nuclear recoils that contain noise hits, so we do not rule them out at this stage.
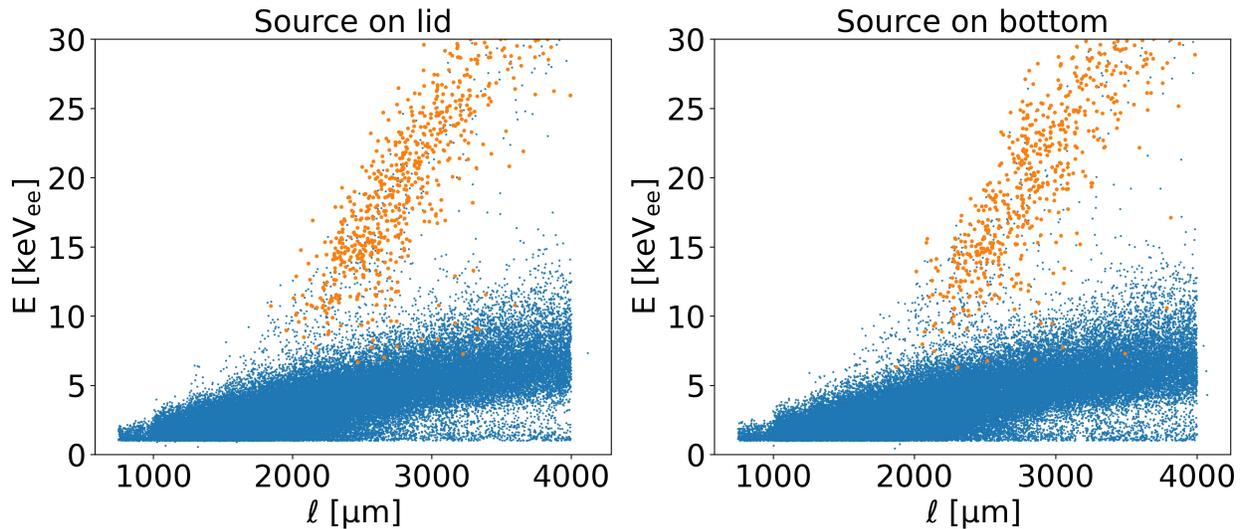
Figure 7.19: Selected signal sample of nuclear recoil candidates (orange) overlaid on top of all measured events (blue).

7.4.2 Head-tail assignment

We now use our 3DCNN classifier to assign head-tail to recoils satisfying these p_e selections. Figure 7.20 shows the p_- distributions for all three measurement campaigns at three different energy ranges. In the highest energy case $E_{reco} < 30\,\text{keV}_{ee}$ where we would expect the most peaking toward the source locations, we find little to no head-tail performance. Indeed, we would expect the source-on-bottom sample to peak at $p_- = 0$, but we instead find that both source orientations peak at the $p_- = 1$, indicating that the model is not well enough trained to pick out the true source directions in measurement. This appears to be an example of the sim2real gap in artificial intelligence where model performance does not translate effectively between simulation and measurement.

7.5 Discussion and conclusions

The 3DCNN classification performance for both event selection and head-tail identification on simulation beats out any technique in the current literature, but unfortunately this performance
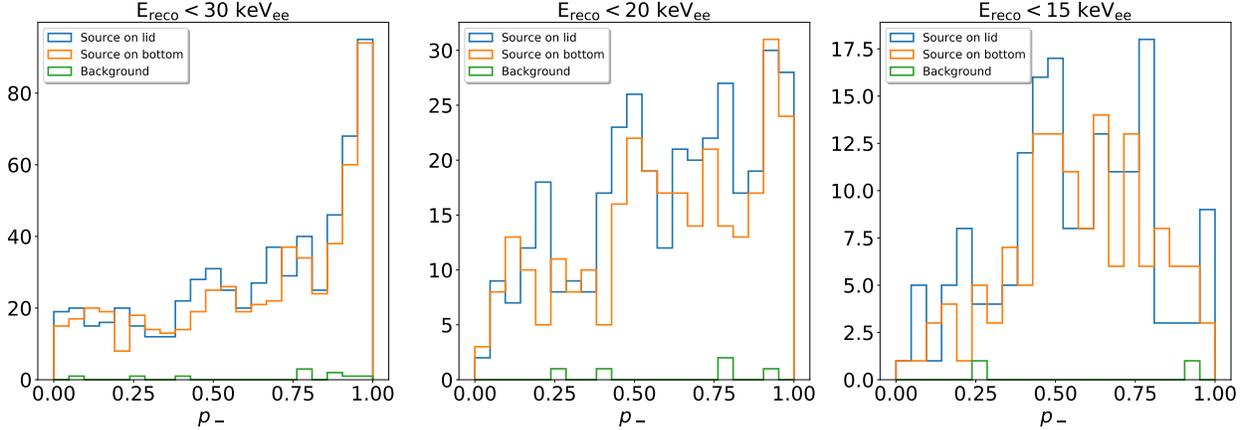
Figure 7.20: Distribution of output probabilities that the events lie in the negative ϕ_{truth} hemisphere (p_-). We find little difference in the source-on-lid and source-on-bottom samples indicating that our trained 3DCNN does not provide meaningful head-tail assignment performance on measurement.

does not yet generalize to measurement. We believe that our performance on simulation shows promise for 3DCNNs becoming a standard tool for directional DM analyses in 3D recoil-imaging TPCs, however more work is needed to ensure that performance generalizes to measurement. There are a number of features that simulation doesn't model accurately. One such example is the net charge distribution over all events. Though we observe dE/dx agreement between measurement and simulation, the charge distributions of events that the 3DCNN uses to make classifications do not agree over a fixed event length range ($2.0\,\text{cm} < \ell < 2.2\,\text{cm}$ shown in Figure 7.21). By extension, the number of pixel hits per event at a given energy and length range do not agree, suggesting that the statistical information the 3DCNN uses for classification may differ considerably between measurement and simulation.

The model complexity of 3DCNNs is a double-edged sword. On one hand the model complexity allows the network to find subtle correlations and use these to classify events that would not be possible to reliably classify with a simpler model, but on the other hand, this model complexity makes it difficult to assess the correlations that lead to poor generalization of performance to measured data. Improving simulation is one approach to try to narrow the sim2real gap. For instance, if the FE-I4's noise, threshold, and saturation limits were more accurately modeled, then the charge distributions would better agree which might improve generalization. But charge distributions aren't the only discrepancies. The 3D spatial distribution also matters, so differences between measured and simulated gains and diffusion through the GEMs can have effects that are difficult to tune.

We could also work to "strengthen" the features that the network learns from. The most effective way to do this would be to build a detector that's more optimized for low energy directionality. In
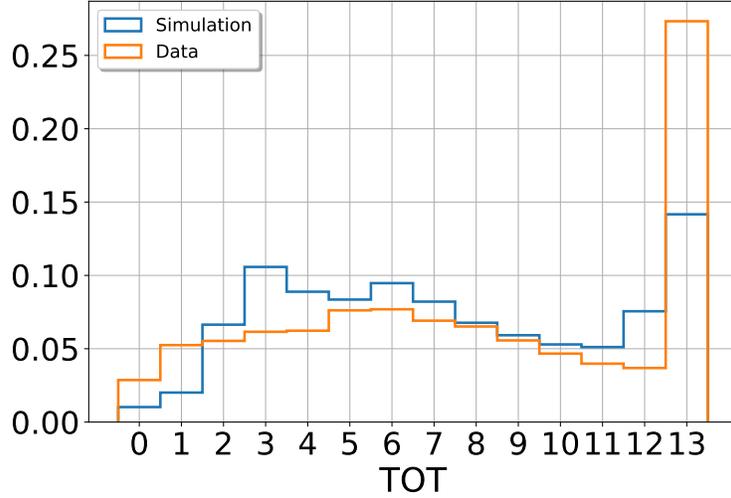
Figure 7.21: Comparison of charge (TOT) distributions of measured and simulated events satisfying $2.0\,\mathrm{cm} < \ell < 2.2\,\mathrm{cm}$. Both histograms are normalized to an integral of unity.

the absence of a more ideal detector, with our current setup there are ways that we may be able to improve the generalization potential of our 3DCNNs. Shortening the drift length by moving the source cube closer to the TPC readout is something we could immediately try that would reduce the diffusion due to drift of recoil events. This would improve angular resolution but come at a cost of energy resolution unless we decreased the gain slightly. Alternatively we could lower the gas pressure of our He:CO_2 mixture which would lead to longer tracks at a fixed energy, and thus better angular resolution. Adjusting either of these in simulation and measurement could lead to the 3DCNN learning stronger correlations for classification that may generalize better to measurement.

At the time of writing, we are investigating detailed differences between experimental data and simulation, and working to make the 3DCNN head-tail classifier generalize better to experimental data, but that work is beyond the scope of this dissertation.

CHAPTER 8
DEEP LEARNING FOR IMPROVED KEV-SCALE
SIMULATED RECOIL IDENTIFICATION

In Chapters 5-7 we developed and trained 3DCNNs for both event classification and directional head-tail in the BEAST TPCs. What follows here is a self-contained simulation study comparing electron rejection performance for He and F recoils in a 60-Torr mixture of He:CF$_4$:CHF$_3$, which is a gas mixture that is better suited for low energy directional sensitivity. This chapter has been submitted as a journal article to JCAP [29] and is currently being revised. The sections introducing 3DCNNs that will appear in the journal article have been removed, as they have already been previously introduced. We're currently investigating bias in our event processing that may affect the electron rejection performance of the 3DCNN. The other electron rejection techniques investigated here are unaffected by this potential bias.

Abstract

Recoil-imaging gaseous time projection chambers (TPCs) with directional sensitivity are attractive for DM searches. Detectors capable of reconstructing 3D nuclear recoil directions would be uniquely sensitive to the predicted dipole angular distribution of DM recoils. Observation of this directional distribution would unambiguously establish the galactic origin of a claimed DM signal. Recoil directionality also would provide powerful discrimination against background recoils from solar neutrino scattering. These advantages of directionality can only be exploited however, if electron recoil backgrounds from gamma rays can be sufficiently suppressed. We introduce a deep learning-based recoil event classifier that uses a 3D convolutional neural network (3DCNN) to identify event species based on their recoil images. We compare electron background rejection performance of discriminants determined by the 3DCNN both to the traditional discriminant of track length, as well as discriminants obtained from state-of-the-art shallow learning methods. We train the 3DCNN classifier using recoil charge distributions with ionization energies ranging from 0.5-10.5 keV$_{ee}$, for 25 cm of drift in an 80:10:10 mixture of He:CF$_4$:CHF$_3$. The charges are initially segmented into $(100 \times 100 \times 100)$µm^3 bins when determining track length and the shallow learning discriminants, but are rebinned with a reduced segmentation of about $(850 \times 850 \times 850)$µm^3 for the 3DCNN. Despite the courser binning, compared to using track length, we find that classifying events with the 3DCNN reduces electron backgrounds by a factor of up to 1,000 and effectively reduces the energy threshold of our simulated TPC by 30% for fluorine recoils and 50% for helium recoils. We also find that the 3DCNN reduces electron backgrounds by up to a factor of 20 compared to the shallow machine learning approaches, corresponding to a 2 keV$_{ee}$ reduction in the energy threshold.

8.1 Introduction

Directly detecting the constituents of dark matter (DM) remains one of the key goals of contemporary physics. Directional detection would provide unique, robust, and unambiguous confirmation of the galactic origin of a signal in the form of a dipole distribution in galactic coordinates [16; 94].

As more regions of DM parameter space are being ruled out, and cross-section limits approach the solar neutrino fog, directional detection is also increasingly of interest for maximizing DM sensitivity within the fog, and for studying the solar neutrinos themselves. Detectors capable of reconstructing the full 3D vector direction, energy, and time of individual recoil events are preferable in this context. Such "recoil-imaging" detectors are widely applicable and hence garnering increasing interest [12; 93].

Gaseous time projection chambers (TPCs) are the most mature recoil imaging technology with event time measurement. Unlike liquid noble gas detector, which benefit from strong self-shielding, one key issue in gas detectors is the rejection of electron recoil backgrounds from gamma rays. TPCs with high-definition charge readout (HD TPCs) are, however, capable of reconstructing the topology of both nuclear recoils and electron recoils in great detail, which we expect should maximize particle ID capabilities and again help to reject such backgrounds. A key question then, is whether the electron rejection in gas TPCs is sufficient to achieve a background free DM search for a given experiment and exposure.

Studies of the proposed $1000\,\mathrm{m}^3$ gas TPC experiment CYGNUS [20], running for six years, suggest the electron background will be flat in energy, and electron backgrounds must be rejected offline by factors exceeding 6×10^4 per $\mathrm{keV_{ee}}$. Due to the diffusion in gas-based detectors, electron identification performance falls exponentially with recoil energy. As a result, the effective analysis energy threshold above which a large gas TPC will remain background free is determined by the particular energy at which the electron rejection meets the performance requirement. A relatively small improvement in electron rejection, and consequent reduction of the analysis energy threshold can have large benefits on DM reach, because the expected DM recoil spectrum is generally steeply falling with energy.

Because recoil-imaging HD TPCs offer extremely rich three-dimensional charge density measurements, one obvious question is how to best exploit this information to maximize recoil identification. This is a problem ideally suited for machine learning techniques.

There have been many efforts to improve electron background rejection for keV-scale recoil events, which often use multivariate combinations of discriminant observables for event classification. The MIMAC group, for instance, used boosted decision trees (BDTs) [95] to improve electron rejection factors by a factor of 20 over more traditional methods [96]. More recently, we introduced a set of nine observables that are based on the shape of the recoil charge cloud measured in a 3D recoil-imaging TPC [97]. A joint combination of these nine event-shape variables using a hard-cuts-based approach led to up to a two order of magnitude improvement in electron rejection over using

143

the length along a track's principal axis, which is a common observable for event classification.

Given the electron background rejection improvement observed both by multivariate combinations of discriminant variables and the usage of event-shape variables, we attempt to combine the best aspects of both of these approaches by introducing a deep learning-based classifier for event identification. In particular, we introduce a 3D convolutional neural network (3DCNN, Chapter 8.3) that is directly fed the 3D ionization density distribution of events, binned into a $32 \times 32 \times 32$ voxel grid, and outputs class probabilities of the recoil species of the event. We expect this end-to-end approach to enable us to capture more information than using predefined observables [98; 99; 100; 101], thus leading to better background rejection performance. In Chapter 8.4, we separately combine the nine event-shape observables with two shallow learning techniques, (1) a BDT and (2) a fully connected neural network (FCNN), both of which also output recoil-species class probabilities that can be used as multivariate classification discriminants. Comparing the electron background rejection performance between these shallow learning-produced discriminants and the hard-cuts-based combined observable from Ref. [97] allows us to assess the relative effectiveness of different techniques of combining the nine event-shape observables. Comparing the electron rejection performance of the 3DCNN with the best-performing combination of these observables provides insight on if any crucial event-shape information is missing from the set of nine observables.

8.2   Overview of simulation

Before introducing our new classifiers, we first describe the simulated detector and recoil characteristics used in our study. We build off of previous work from our group and generate a large simulation sample using identical parameters to those in Ref. [97]. Doing this allows us to directly compare the electron rejection performance of our new classifiers with already established electron rejection improvements.

We use `SRIM` [102] and `retrim` [103] to simulate recoiling He and F nuclei and `DEGRAD` [104] to simulate electron recoils in an 80:10:10 mixture of He:CF_4:CHF_3 at a total pressure of 60 Torr and temperature of 25°C. The `SRIM` computation requires a compound correction for every nucleus in the gas mixture (He, H, C, and F). The calculation of these compound corrections and their associated values are detailed in Ref. [97]. The `retrim` step requires the average energy per ion pair, W, and Fano factor, $\mathcal{F}$, which were calculated using `Garfield++/Heed` [75] as $W = 35$ eV and $\mathcal{F} = 0.19$, respectively. `DEGRAD` can simulate recoils isotropically; however, for `SRIM` and `retrim` we isotropize the recoil simulations after generating the initial ionization distributions. We assume a drift field of 40.6 V/cm parallel to the drift direction and use `Magboltz` [69] to determine the corresponding transverse and longitudinal diffusion coefficients as $(\sigma_T, \sigma_L) = (398, 425)$µm/$\sqrt{\text{cm}}$. The diffusion is applied to our simulations assuming a uniform drift length of 25 cm.

After applying diffusion, we assume each individual electron is detected. We do not simulated any charge amplification or digitization, but instead simulate a highly efficient, pixelated readout

by binning each primary track into $(100\,\mu m \times 100\,\mu m \times 100\,\mu m)$ voxels. We note that our choice of a $40.6\,V/cm$ drift field corresponds to $100\,\mu m$ per clock cycle on a $40\,MHz$ clock which is equivalent to the readout refresh rate used in existing TPCs with pixel readout [23].

8.3 Convolutional neural network classifier

Recoil-imaging TPCs with high readout segmentation reconstruct detailed 3D images of ionization charge distributions of recoil tracks. These tracks have many identifiable characteristics that can be used for particle identification. For sufficiently high energy recoil tracks, the length of the track along its principal axis is often enough to reliably distinguish between electron recoils and nuclear recoils, as electron recoils tend to create longer tracks following a more meandering path than nuclear recoils of equivalent energy. At lower energies, however, diffusion during drift has a proportionately larger effect on the overall event topology, leading to more spherical ionization distributions. Figure 8.1 shows $\sim 6\,keV_{ee}$ examples of each of the three recoil species investigated here. Comparing the background electron recoil (left) with the signal fluorine and helium recoils (middle and right, respectively), we see that both of the nuclear recoil species have a dense cluster of charge near the center of the event, whereas the electron recoil appears to be a more diffuse charge cloud with larger gaps. It's difficult to unambiguously identify the principal axis direction of low energy tracks, leading to poor angular resolution. Furthermore, a few stray charges can significantly bias the length along the principal axis, which may lead to event misclassification with this traditional approach. Since there are so many features present in these tracks, using a deep learning classifier where the classifier learns the best patterns for event selection on its own, independent of an identified principal axis, is attractive. To this end, we construct and train a 3DCNN classifier that is directly fed 3D voxel grids of the ionization distribution of individual recoil events with the recoil species assigned as the class-label of the event.

8.3.1 Data processing for the 3DCNN

We simulate 12,380,422 electron recoil events ranging in ionization energies between $0.5\,keV_{ee}$ and $10.5\,keV_{ee}$, 367,984 F recoil events, and 338,909 He recoils. The nuclear recoils are simulated with a uniform energy spectrum between $5\,keV_r$ and $50\,keV_r$. While our simulated detector is sensitive to individual electrons, nuclear recoils still lose energy due to quenching, so in the absence of reconstructing the true recoil energy, we compute the ionization energy of all nuclear recoil species in terms of electron equivalent energy to allow for a direct comparison of electron rejection at a given ionization energy in our detector. Expressed in terms of electron equivalent energy, the ionization energy of a nuclear recoil $E_{ionization}$, is computed as
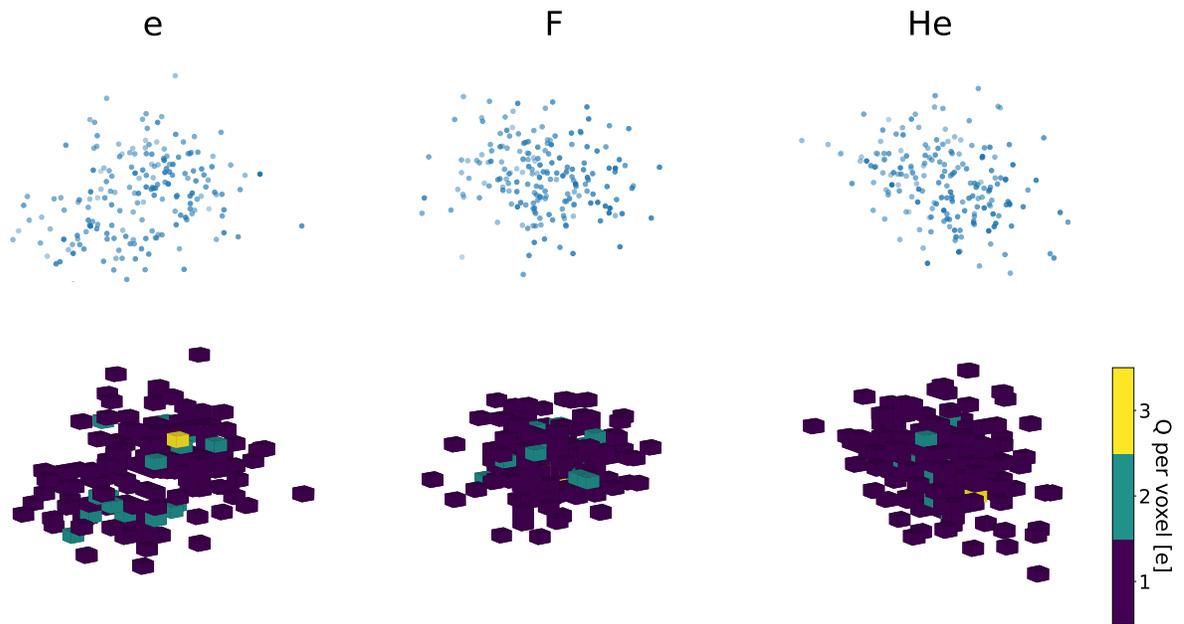
Figure 8.1: 3D charge distributions of $6\,\mathrm{keV_{ee}}$ electron, fluorine, and helium recoils after diffusion. The top row shows the original $(100\,\mu\mathrm{m} \times 100\,\mu\mathrm{m} \times 100\,\mu\mathrm{m})$ binning of these events, which is fine enough that there is only a single charge in each filled bin. The bottom row shows the events re-binned into the $(32 \times 32 \times 32)$ voxel grid that are input into the 3DCNN.

Recoil species	Original energy range [keV$_r$]	Energy range [keV$_{ee}$]	# Events training set	# Events validation set	# Events test set
e	—	0.5-10.5	2,947,702	491,429	8,941,291
F	5-50	0.5-10.5	53,760	8,812	161,727
He	5-50	0.5-10.5	22,443	3,743	67,588

Table 8.1: Event samples used with the 3DCNN classifier. Events in each sample are binned in $1\,\mathrm{keV_{ee}}$ wide bins centered at integer ionization energy steps between $1\,\mathrm{keV_{ee}}$ and $10\,\mathrm{keV_{ee}}$. We train and evaluate the 3DCNN classifier separately for each energy bin.

$$E_{\mathrm{ionization}} = N_e W_e, \tag{8.1}$$

where N_e is the number of electrons produced in the recoil event, and W_e is the average energy per electron-ion pair. Following the approach in Ref. [97] we set W_e to $32.4\,\mathrm{eV}$. After computing $E_{\mathrm{ionization}}$, we restrict our nuclear recoil sample to $0.5\,\mathrm{keV_{ee}} \leq E_{\mathrm{ionization}} \leq 10.5\,\mathrm{keV_{ee}}$, leaving us with 224,299 F recoils and 93,774 He recoils.

The charges in our simulated events are already binned with $(100 \times 100 \times 100)\mathrm{\mu m}^3$ segmentation, however because diffusion is larger than our bin size, we use a low density gas, and we do not simulate any charge amplification, we find that the charge clouds are relatively sparse at this resolution. While it is feasible to train a 3DCNN on current-generation hardware using the native $(100 \times 100 \times 100)\mathrm{\mu m}^3$ resolution of our simulation, we opt to reduce the spatial segmentation of our events to a $32 \times 32 \times 32$ grid of bins evenly spaced within a cube of width $2.72\,\mathrm{cm}$, leading to bin sizes of about $(850 \times 850 \times 850)\mathrm{\mu m}^3$. We do this for two reasons: First, all fluorine recoil events are entirely contained within this cube, so no recoil information is lost due to cropping, and second, the reduced spatial segmentation leads to more than a factor of 600 reduction in grid volume compared to the original $(100 \times 100 \times 100)\mathrm{\mu m}^3$ segmentation, leading to a substantial reduction in computational cost. Comparing the top and bottom rows of Figure 8.1 it appears that the structure of recoil tracks isn't significantly altered with this reduced resolution, suggesting that event classification performance for the simulated detector configuration will not be significantly hampered by our reduction of spatial segmentation. We note that uncertainties in bin placement from re-binning may slightly alter performance compared to binning the unbinned initial coordinates of 3D charge into the $32 \times 32 \times 32$ grid.

We use the `PyTorch` [85] software library for all neural network computations and store each event as a tuple containing two entries: (1) a $(32 \times 32 \times 32)$ voxel grid filled with binned charge that is stored as a `PyTorch` tensor data structure, and (2) an integer representing the class-label of the event. We label electron recoils as 0, F recoils as 1, and He recoils as 2. Charge in each bin of the voxel grid is stored as unsigned 8 bit integers leading to an effective dynamic range of 0 to 255

electrons per voxel. No bins are saturated in any event with this dynamic range.

8.3.2 Network architecture
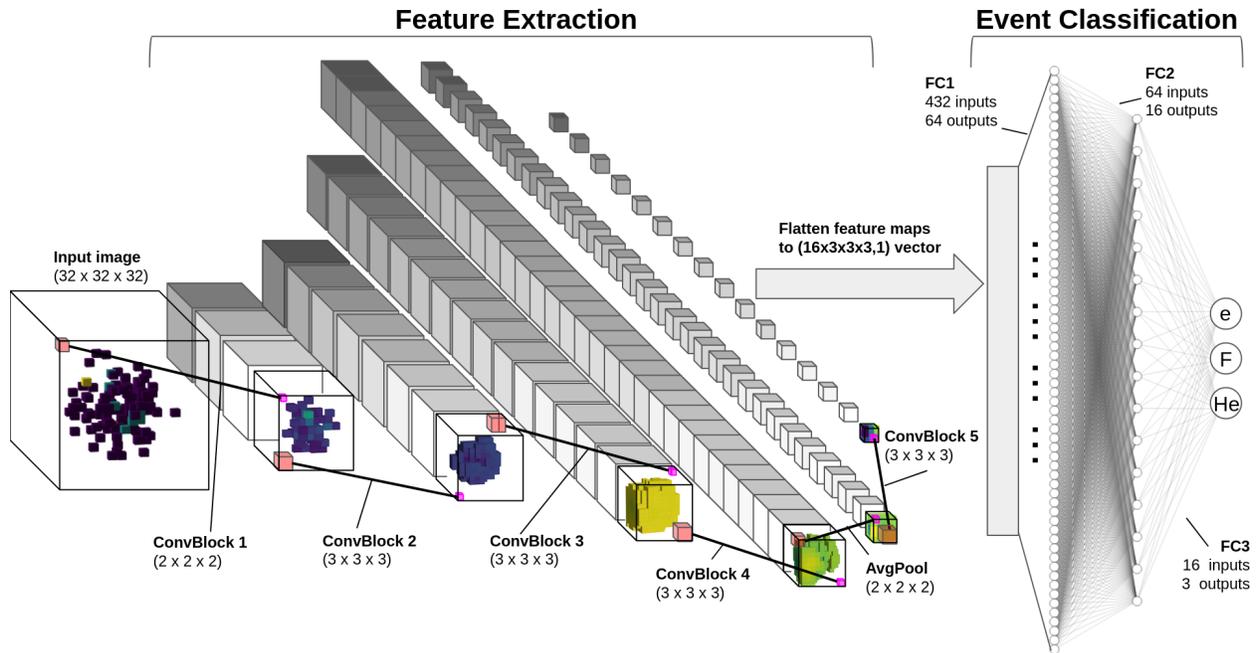
Figure 8.2: Architecture of our 3DCNN classifier. Within the feature extraction portion, each cube represents a unique feature map (represented by the different shade of each map), with the size of the cubes shown approximately to scale in reference to the dimensions of the input image. We show the bin content of one such-feature map in each layer to represent the steps of a convolutional chain. Solid black lines connect larger pale red cubes, which illustrate the convolutional or pooling filters acting on a portion of the input feature map, to the smaller magenta cubes which are the outputs of the convolutional filter. The final 16 feature maps are then flattened and passed into a fully connected neural network for classification. The FCNN diagram was produced using [86].

Figure 8.2 and Table 8.2 together outline the network architecture of our 3DCNN. As is shown in Figure 8.2, our network consists of a series of convolutional and pooling layers for feature extraction, followed by a dense fully connected neural network (FCNN) for event classification. Within the feature extraction portion of the neural network, we employ five convolutional blocks (ConvBlocks 1-5) and a pooling layer (AvgPool) to downsample the feature maps while still maintaining important features. Each of the five convolutional blocks contain the following components: (i) A 3D convolution with a convolutional filter size listed in Table 8.2, (ii) a 3D batch normalization [87], (iii) a scaled exponential linear unit (SELU) activation layer [88], and (iv) a randomized dropout [89] of 0.03 to reduce overfitting.

Walking through the network architecture, we start with a $(32 \times 32 \times 32)$ input image. The vast majority of voxels in all input images are filled with 0, so we implement a stride of 2 in

Layer	# Filters	Filter size	S	# Learnable parameters	Dropout	Output shape
ConvBlock 1	4	$(2 \times 2 \times 2)$	(2,2,2)	44	0.03	$(4 \times 16 \times 16 \times 16)$
ConvBlock 2	8	$(3 \times 3 \times 3)$	(1,1,1)	888	0.03	$(8 \times 14 \times 14 \times 14)$
ConvBlock 3	16	$(3 \times 3 \times 3)$	(1,1,1)	3,504	0.03	$(16 \times 12 \times 12 \times 12)$
ConvBlock 4	32	$(3 \times 3 \times 3)$	(1,1,1)	13,920	0.03	$(32 \times 10 \times 10 \times 10)$
AvgPool	32	$(2 \times 2 \times 2)$	(2,2,2)	—	—	$(32 \times 5 \times 5 \times 5)$
ConvBlock 5	16	$(3 \times 3 \times 3)$	(1,1,1)	13,872	0.03	$(16 \times 3 \times 3 \times 3)$
FC1	—	—	—	27,712	0.05	(64×1)
FC2	—	—	—	1,040	0.05	(16×1)
FC3	—	—	—	51	—	(3×1)

Table 8.2: More specific details of each layer shown in Figure 8.2. We assume a single $(1 \times 32 \times 32 \times 32)$ image is fed into the network. The output shape column gives the shape of the output after each layer. The output of ConvBlocks 1-5 and AvgPool is a tensor of shape $(D \times L \times W \times H)$, where D is the layer depth (number of feature maps), and L, W, and H are the length, width and height of each feature map, respectively. S is the convolutional stride of the layer. The output shapes of the FC layers are 1 dimensional vectors for each node in a given layer. In addition to the weights and biases associated with each node in the FCNN, the entries within each convolutional filter are also learnable parameters, so we list the total number of learnable parameters associated with each layer in the network.

ConvBlock 1 to immediately downsample this block's four output feature maps to $(16 \times 16 \times 16)$ to reduce the computational overhead of training and evaluating our network. Following the chain in Figure 8.2, we next perform three successive convolutional blocks (ConvBlocks 2-4) where we gradually increase the number of convolutional filters employed in each layer to produce more feature maps for classification. ConvBlocks 2-4 each use $(3 \times 3 \times 3)$ convolutional filters with a stride of 1 and no padding, so each of these convolutions decreases the feature map dimension by two, leaving us with 32 feature maps of size $(10 \times 10 \times 10)$ at the end of ConvBlock4. After this, we perform average pooling. We use a $(2 \times 2 \times 2)$ pooling filter with a stride length of 2, leaving us with 32 $(5 \times 5 \times 5)$ feature maps that we feed into our final convolutional block (ConvBlock5) which leaves us with 16 $(3 \times 3 \times 3)$ feature maps. Since the values composing each convolutional filter are learnable parameters, we expect that when the network is trained, there will be useful features encoded in some of these feature maps. We finally flatten these feature maps into a $((16 \times 3 \times 3 \times 3) \times 1) = (432 \times 1)$ vector, that contains each extracted feature at the end of the convolutional chain. This flattened (432×1) feature vector is then fed into a fully connected dense neural network with two hidden layers and output class assignments corresponding to e, F, and He recoils. Each of the three fully connected layers (FC1, FC2, and FC3) use a SELU activation function and FC1 and FC2 include a random dropout of 0.05. The raw model output of each event is a (3×1) vector, $\mathbf{z}$, with entries corresponding to each of the three class outputs (e recoil, F recoil, and He recoil). The softmax function

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{3} e^{z_j}}. \tag{8.2}$$

is applied to $\mathbf{z}$ to map the class outputs $z_i \in \mathbf{z}$ to class probabilities. We henceforth label $\sigma(\mathbf{z})_1$, $\sigma(\mathbf{z})_2$, and $\sigma(\mathbf{z})_3$, as p_e, p_F, and p_{He}, which represent the model-predicted class probabilities of e, F, and He recoils, respectively.

8.3.3 Training the network

We first shuffle the order of all events and then split the data into distinct training, validation, and test sets with 3,023,905 events in our training sample, 503,984 events in our validation sample, and 9,172,513 events in our test sample. We set the test sample aside and implement the following procedure to train our model:

1. Form a `PyTorch` tensor of shape $(256 \times 32 \times 32 \times 32)$, which is a *minibatch* consisting of 256 randomly selected voxel images from the training sample.

2. Feed the minibatch and the corresponding truth label of each image of the minibatch into the 3DCNN.

3. Use `PyTorch`'s built in CrossEntropy loss function to compute the loss of the batch. We wish to minimize this loss. We use an Adam [90] optimizer with a learning rate of 0.0002.

4. Update model weights using backpropagation [91].

5. Repeat steps 1-4 until we've run through all events in the training set. This is called a training epoch.

6. At the end of each training epoch, repeat steps 1-3 for the validation sample. We do not implement step 4 as we don't wish to train the 3DCNN on the validation set. Compute the sum of the losses of each minibatch of the validation sample.

7. If the summed losses over the validation set minibatches are less than in the previous epoch, we treat this as the model learning and save all model weights.

8. Implement early stopping [92] where steps 1-7 are repeated until the total validation loss doesn't decrease at all over 10 successive epochs.

9. The model state corresponding to the epoch with the lowest validation loss is our trained model.

We were able to train the 3DCNN with an Nvidia GeForce RTX 2070 consumer-grade laptop GPU in less than 6 hours.
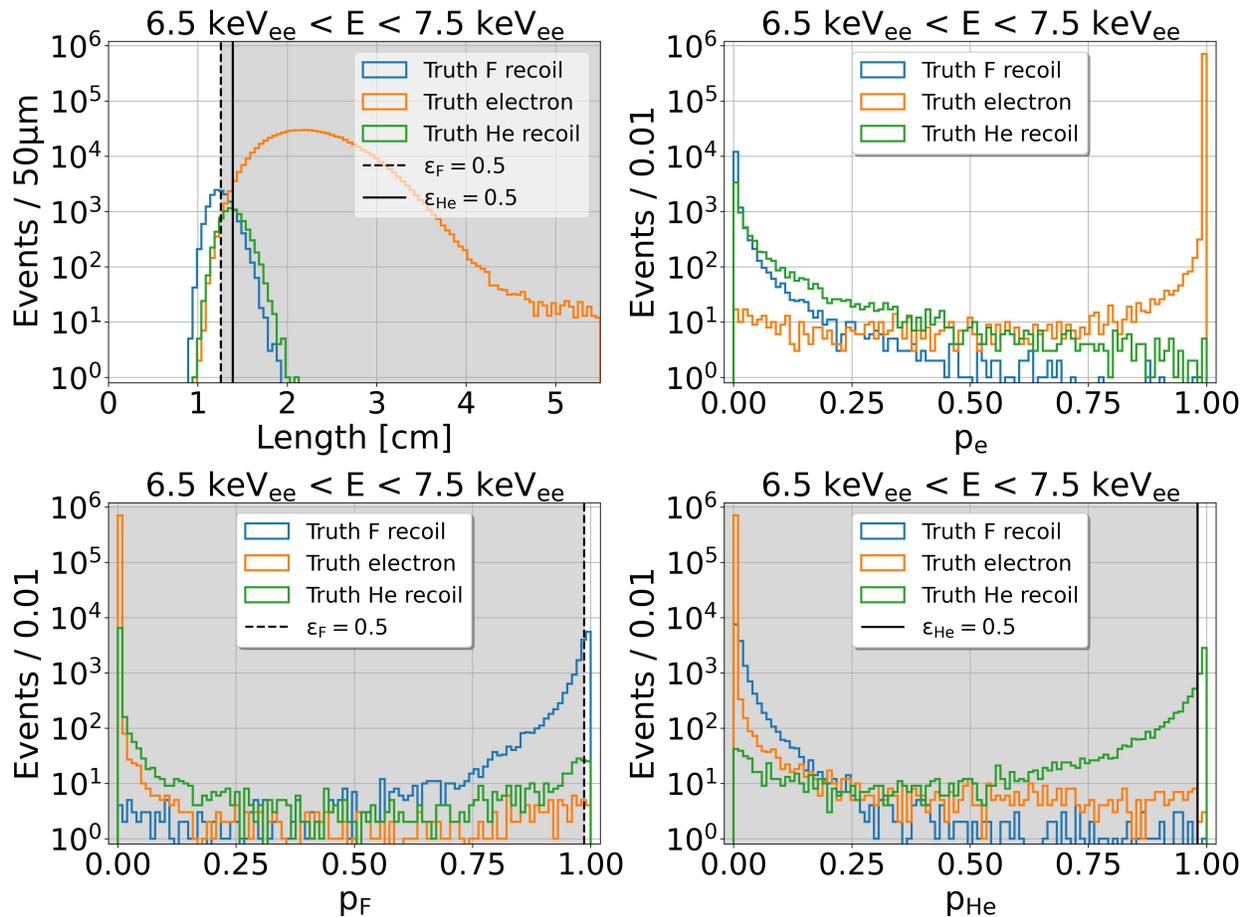
Figure 8.3: Comparison of traditional (track length) and 3DCNN-based (p_e, p_F and p_{He}) discriminants for events in the test sample satisfying $6.5\,keV_{ee} < E_{ionization} < 7.5\,keV_{ee}$. The shaded regions in the panels indicate events that are rejected after selections corresponding to 50% F recoil efficiency (dashed black vertical line) and/or 50% He recoil efficiency (solid black vertical line).

8.3.4   Interpreting model output

We use the 3DCNN output variables p_e, p_F, and p_{He} as event classification discriminants. Figure 8.3 shows a comparison of the distributions of track length, p_e, p_F, and p_{He} for all true F recoils (blue), all true He recoils (green), and all true electrons (orange) in the test sample of data. Since p_e, p_F, and p_{He} represent class probabilities, for each event, $p_e + p_F + p_{He} = 1$, meaning if our 3DCNN is a good classifier, we should then expect that each class probability peaks strongly toward 1 for its corresponding class, and strongly toward 0 for the other two classes. This expected behavior is observed in Figure 8.3, suggesting that p_e, p_F, and p_{He} are strong event classification discriminants.

In addition to rejecting electron backgrounds, these discriminants can also be used to distinguish nuclear recoil signal candidates. To quantify classification performance, we define the signal efficiency as

$$\varepsilon_s = \frac{N_s'}{N_s},\tag{8.3}$$

where N_s is the total number of signal events in the sample and N_s' is the number of remaining signal events after making a selection. Now, suppose we want to quantify the signal purity of F recoils at an F recoil efficiency of $\varepsilon_F = 0.5$. We can do this by determining p_F', the median p_F of true F recoils. Then, we can determine the number of F, He, and e recoils where $p_F > p_F'$, and compute the F recoil purity as $N_F'/(N_e' + N_F' + N_{He}')$. The bottom left panel of Figure 8.3 shows an example of these selections for the set of recoil events between $6.5\,\mathrm{keV_{ee}}$ and $7.5\,\mathrm{keV_{ee}}$, where the black vertical dashed line corresponds to p_F'. We can perform an analogous procedure with p_{He} to compute the He recoil at a fixed He recoil efficiency.

In certain instances it may be advantageous to treat any nuclear recoil species as signal and electrons as background. In these cases we could treat p_e as a binary classification variable between electrons versus not electrons (i.e. signal nuclear recoils for this ternary classification model). For example, if we wanted to calculate the nuclear recoil purity at a nuclear recoil efficiency of $\varepsilon_{\mathcal{R}} = 0.5$, we could determine, p_e', the median value of p_e for all He or F recoils. We would then compute $(N_F' + N_{He}')/(N_e' + N_F' + N_{He}')$ where N_F', N_{He}', and N_e' represent the number of events of each of these three classes after applying a selection of $p_e < p_e'$.

Since signal purity depends on the relative composition of recoil constituents, we opt to mostly use the notion of a rejection factor, R, to quantify background rejection performance. We define the sample size-independent *electron* rejection factor R_e as the ratio of the total number of electrons in a sample N_e to the number of electrons remaining after a selection, N_e'

$$R_e = \frac{N_e}{N_e'}.\tag{8.4}$$

For the $7\,\mathrm{keV_{ee}}$ sample shown in Figure 8.3, when using length as the event classification variable we find rejection factors of 135 ± 2 and 870 ± 30 for $\varepsilon_{\mathrm{He}} = 0.5$ and $\varepsilon_{\mathrm{F}} = 0.5$, respectively. Using 3DCNN output probability as our event classification variable, we find that only five electrons remain both for He recoils at $\varepsilon_{\mathrm{He}} = 0.5$ and F recoils at $\varepsilon_{\mathrm{F}} = 0.5$, leading to an electron rejection factor of $141,400 \pm 63,000$ for both of these cases, a more than 3 order of magnitude improvement over the traditional track length discriminant for He recoils.

8.4 Shallow learning classifiers

Here we introduce the nine event-shape discriminants used in Ref. [97] and train both a BDT (Chapter 8.4.2) and an FCNN (Chapter 8.4.3) to combine these observables into multivariate classification discriminants.

8.4.1 Defining electron rejection discriminants

Each of the following variables are computed using our simulated TPC's native $(100 \times 100 \times 100)\mathrm{\mu m}^3$ resolution, as opposed to the reduced $(850 \times 850 \times 850)\mathrm{\mu m}^3$ segmentation used for the 3DCNN classifier.

1. Length along the principal axis (LAPA): We use a singular value decomposition (SVD) to identify the principal axis of the 3D track, and then take the difference between the maximum and minimum of the track's ionization distribution coordinates projected onto this principal axis.

2. Standard deviation of charge distribution (SDCD): The standard deviation of the 3D position vectors of all charges in the event.

3. Maximum charge density of the event (ρ_{Max}): This discriminant is optimized by varying the bin width of the charge in cubic voxels and recording the maximum amount of charge in a voxel. ρ_{Max} was separately optimized to maximize electron rejection for weakly-directional $7\,\mathrm{keV_{ee}}$ F recoils and for directional $12\,\mathrm{keV_{ee}}$ He recoils as described in Ref. [97].

4. Charge uniformity (QUnif): The standard deviation of the distribution of mean distances between each charge and all other charges in a recoil event.

5. Cylindrical thickness (CylThick): Sum of the squared transverse distances from each charge to the principal axis of the track.

6. Number of clusters (NClust): Number of clusters determined by the DBSCAN [105] clustering algorithm. Two of the input parameters to this algorithm were adjusted to find the pair that optimizes electron rejection both for weakly-directional $7\,\mathrm{keV_{ee}}$ F recoils and for directional $12\,\mathrm{keV_{ee}}$ He recoils as described in Ref. [97].

$\rho_{\mathrm{Max,F}}$	SDCD	$\mathrm{CThres_F}$	$\rho_{\mathrm{Max,He}}$	$\mathrm{CThres_{He}}$	CylThick	QUnif	LAPA	NClust
0.437	0.231	0.169	0.111	0.032	0.009	0.007	0.003	0.002

Table 8.3: Normalized feature importance of the nine discriminants in the trained BDT.

7. Clustering threshold (CThres): The threshold that the fraction of total event charge in the largest cluster of an event must be above. Clustering is performed `DBSCAN` and the threshold value is optimized separately to maximize electron rejection for weakly-directional $7\,\mathrm{keV_{ee}}$ F recoils and for directional $12\,\mathrm{keV_{ee}}$ He recoils.

Given the separate directional and weakly-directional optimizations for the ρ_{Max} and ClustThres observables, we have a total of nine discriminants to work with.

8.4.2 Boosted decision tree

We use the `XGBClassifier` model in the `XGBoost` [106] software library as our BDT classifier for recoil classification. For this BDT analysis, we first randomize the order of our data and then partition the data into a 23.8% / 4.0% / 72.2% training/validation/testing sample split. The validation data sample is not used for the BDT analysis; however we still opt to create a validation set so that we have an identical testing set to what's later used in the FCNN analysis. For each event, we compute the nine discriminants and store the results as a (9×1) vector to feed into the BDT classifier. We assign class-labels of 0 for e recoils, 1 for F recoils, and 2 for He recoils.

Given the relatively quick training and evaluation time of `XGBoost` compared to our 3DCNN classifier, we perform a course, two-step partial grid search to optimize some of the hyperparameters in the `XGBClassifier` model. We make the *a priori* decision to choose the set of hyperparameters $(N_{\mathrm{trees}}, D_{\mathrm{tree}}, \ell)$ that maximizes the electron rejection factor R_{e} for $7\,\mathrm{keV_{ee}}$ F recoils at $\varepsilon_{\mathrm{F}} = 0.5$, where N_{trees}, D_{tree}, and ℓ represent the number of trees, the maximum depth of a tree, and the learning rate of the model, respectively. The first step of our grid search is to fix ℓ at 0.2, and then train and evaluate our BDT using all 16 combinations of $N_{\mathrm{trees}} \in \{50, 100, 200, 500\}$ and $D_{\mathrm{tree}} \in \{2, 3, 5, 7\}$. We find that $N_{\mathrm{trees}} = 200$ with $D_{\mathrm{tree}} = 3$ to be the best combination at $\ell = 0.2$, leading to $R_{\mathrm{e}} = 101,000 \pm 38,000$. Since this combination of N_{trees} and D_{tree} does not sit at the boundary of our grid, we keep this combination and move on to our next step of optimizing ℓ. In our second step of the grid search, we train our classifier using $\ell \in \{0.1, 0.15, 0.2, 0.25, 0.3\}$ and find our initial choice of $\ell = 0.2$ to be the best, so we freeze our optimal set of hyperparameters for our BDT classifier at $(N_{\mathrm{trees}} = 200, D_{\mathrm{tree}} = 3, \ell = 0.2)$.

We trained our BDT model using 53,959 F recoils, 22,265 He recoils and 2,992,534 e recoils in our training sample[1]. Similar to the 3DCNN classifier, the BDT outputs ternary class probabilities,

---

[1]Data for the shallow learning classifiers uses the same simulation as the 3DCNN, but was processed independently, leading to small differences in the training sample statistics compared to Table 8.1. We ensured the number of each

Layer	Input shape	Output shape	Dropout
Input	(9×1)	(32×1)	0.05
Hidden 1	(32×1)	(64×1)	0.05
Hidden 2	(64×1)	(128×1)	0.05
Hidden 3	(128×1)	(32×1)	0.05
Output	(32×1)	(3×1)	—

Table 8.4: FCNN classifier architecture. As mentioned in the text, the input vectors are formed with normalized z-scores of each of the nine electron rejection observables. We use a ReLU activation function and a randomized dropout of 0.05 at the input and each of the hidden layers. When test sample events are passed through the trained network, a softmax function is applied to the model output to give class probabilities.

$p_{e,BDT}$, $p_{F,BDT}$, and $p_{He,BDT}$, allowing for the same model output interpretation as the 3DCNN (Chapter 8.3.4).

A benefit of using a BDT classifier is the ability to track the decisions made throughout all branches in the trained classifier model. Table 8.3 shows the normalized feature importance of each of the nine discriminants. Here we define the feature importance for a particular observable to be the sum of the information gain [107] of all splits where the observable is used, normalized so that the sum of the feature importance of all nine observables is 1. The two observables with the highest feature importance, $\rho_{Max,F}$ and SDCD, are the same two discriminants that perform best in terms of electron rejection factor versus $E_{ionization}$ at 50% F recoil efficiency in Ref. [97].

8.4.3 Fully connected neural network

We next train and evaluate the performance of an FCNN using identical training, validation, and test sets to those generated for the BDT classifier to derive a fair comparison between the two classifiers. Our FCNN consists of an input layer with 9 nodes representing each of the electron rejection observables we've defined, 3 hidden layers, and an output layer with 3 nodes that represent our three output classes (Table 8.4). We use a ReLU activation with batch normalization after the input and hidden layers, and apply a randomized dropout of 0.05 in all layers except the output layer. We train this FCNN classifier by forming minibatches of 512 events, where each event consists of a (9×1) `PyTorch` tensor of normalized z-scores of each of the nine observables and the associated class-label. We then follow the same training procedure (steps 3-9 shown in Chapter 8.3.2) with the same early stopping trigger that we used for the 3DCNN classifier training, except we use a learning rate of 0.001.

We performed a moderate amount of hand-tuning of the hyperparameters for the FCNN, rather than a full hyperparameter search, and find that in general the FCNN gives better electron rejec-

---

recoil species in the *test* sample here matches exactly with Table 8.1. Furthermore, when binning in integer steps of electron equivalent energy, the number of each recoil species in the test sample here is identical to the number of each recoil species in the 3DCNN test sample within each energy bin.

tion performance than the BDT classifier, especially for He recoils. Similar to the 3DCNN, the output of each event passed through the FCNN classifier is a (3×1) vector containing outputs associated with each of the three recoil species, so we apply a softmax to this vector to obtain class probabilities $p_{e,FCNN}$, $p_{F,FCNN}$, and $p_{He,FCNN}$ that may be interpreted analogously to the output class probabilities given by the 3DCNN and BDT.

8.5   Event identification performance results

We now have a set of class probabilities from the 3DCNN, BDT, and FCNN classifiers that can be used to quantify and compare event selection performance. We also include the multivariate observable from Ref. [97] in our comparisons where available.

For a future directional DM detector, it is useful to consider electron background rejection both as a function of ionization energy at fixed nuclear recoil selection efficiency and as a function of nuclear recoil efficiency at fixed energies. Comparing electron rejection performance versus energy provides insight toward the energy threshold required for near background-free operation at a given exposure, while comparing versus nuclear recoil efficiency allows us to compare detector exposure required for a fixed signal yield. Figures 8.4 and 8.5 show the results for both of these approaches.

Figure 8.4 shows R_e performance versus energy at a fixed 50% nuclear recoil efficiency for F recoils (left) and He recoils (right). In cases where all electrons are rejected, we plot R_e as 1×10^6 with zero uncertainty. The black dashed line labeled "No ML" is the combined observable from Ref. [97]. For the three machine learning discriminants and the traditional track length discriminant, we apply the following procedure to determine R_e in each energy bin:

1. Determine the values, $p'_{\mathcal{R}}$, $p'_{\mathcal{R},BDT}$, $p'_{\mathcal{R},FCNN}$, and track length, $L'_{\mathcal{R}}$, that correspond to a 50% efficiency of recoil $\mathcal{R}$. In the left panel of Figure 8.4, $\mathcal{R} = F$, and in the right panel, $\mathcal{R} = He$.

2. N_e is the total number of electrons within the energy bin and N'_e for the 3DCNN, BDT, FCNN, and track length classifiers, is the number of electrons in the energy bin satisfying selections of $p_{\mathcal{R}} > p'_{\mathcal{R}}$, $p_{\mathcal{R},BDT} > p'_{\mathcal{R},BDT}$, $p_{\mathcal{R},FCNN} > p'_{\mathcal{R},FCNN}$, and $L_{\mathcal{R}} < L'_{\mathcal{R}}$, respectively.

3. R_e is computed for the 3DCNN, BDT, FCNN, and track length classifiers with Eq. (8.4) using the N'_e associated with the chosen classifier.

In general, we find that the 3DCNN outperforms all other methods of electron background rejection, especially for He recoils. The CYGNUS collaboration mentions in their feasibility study that electron rejection will effectively determine their energy threshold [20], and argue that with a 6 year exposure and flat electron background energy spectrum of $1 \times 10^4 \, \mathrm{keV_{ee}^{-1} year^{-1}}$, they will be essentially background-free at energies corresponding to an electron rejection factor in excess of 6×10^4. With $\varepsilon_F = 0.5$, we find $R_e > 6 \times 10^4$ starting around $10 \, \mathrm{keV_{ee}}$ using length as a discriminant, compared to somewhere between $6 \, \mathrm{keV_{ee}}$ and $7 \, \mathrm{keV_{ee}}$ with the 3DCNN, meaning the
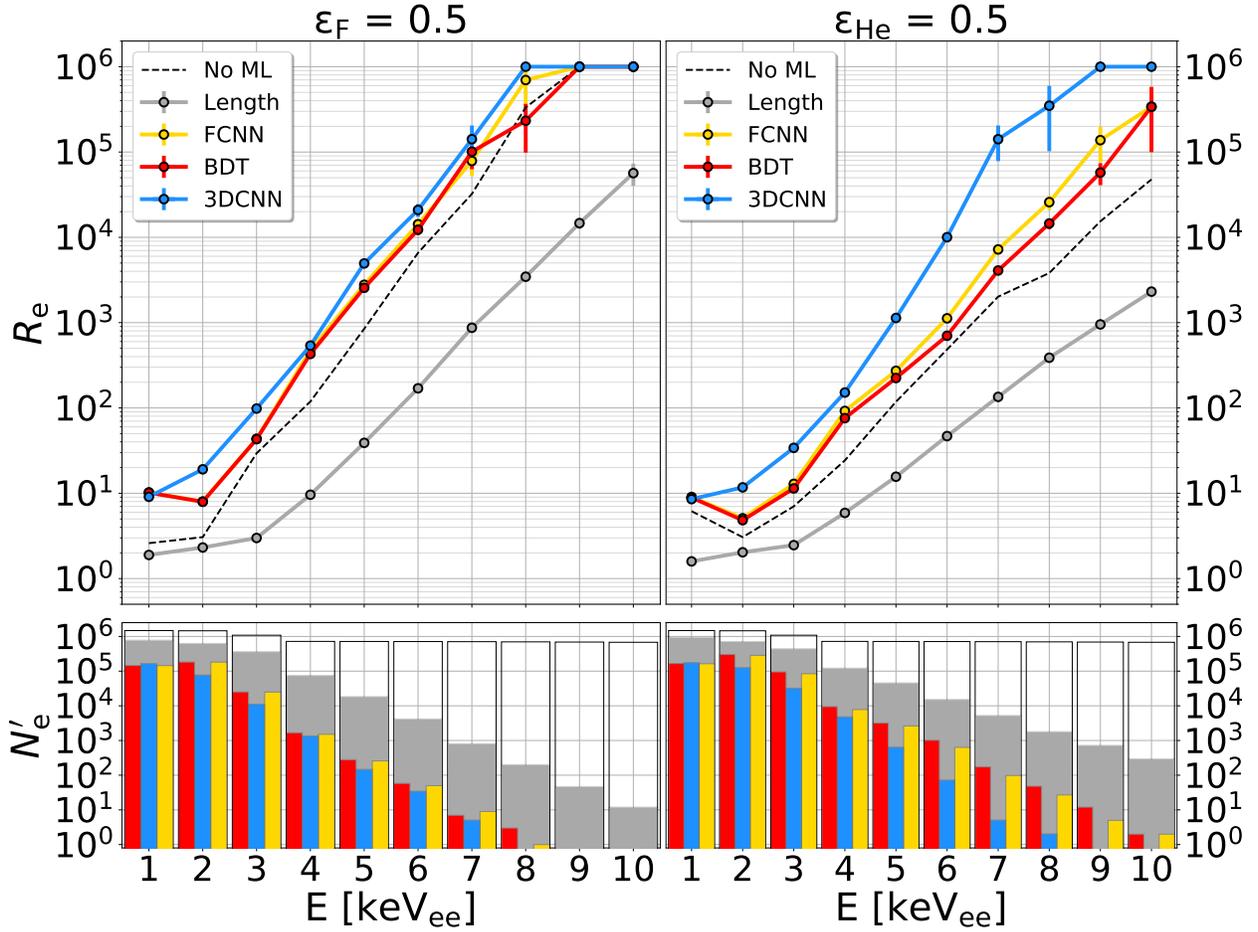
Figure 8.4: Comparison of electron rejection performance of all discriminants versus electron-equivalent energy at 50% fluorine recoil efficiency (left) and 50% helium recoil efficiency (right). The top and bottom row plots share a common horizontal axis with energy-bin centers at integral energy steps from 1 to $10\,\mathrm{keV_{ee}}$. The black dashed lines show the combined observable introduced in Ref. [97]. The bottom bar plots show the number of remaining electrons after a given recoil selection in each integer energy bin. The transparent bars indicate the total number of electron recoils in each energy bin.

use of the 3DCNN for electron rejection lowers the "background-free" energy threshold by more than $3\,\mathrm{keV_{ee}}$ over using length. At $\varepsilon_{\mathrm{He}} = 0.5$ Ref. [97] found that R_e doesn't exceed 6×10^4 until around $14\,\mathrm{keV_{ee}}$ using length as a discriminant, while with the 3DCNN classifier it exceeds 6×10^4 at $7\,\mathrm{keV_{ee}}$. We note here that given the logarithmic scale shown in Figure 8.4, electron rejection performance drops off exponentially with energy, so a factor of two reduction of the "background-free" energy threshold for He recoils when using the 3DCNN classifier over using length is a significant improvement. The otherwise state-of-the-art FCNN and BDT combinations of observables have "background-free" thresholds of around $9\,\mathrm{keV_{ee}}$ for He recoils so the 3DCNN reduces this threshold by around $2\,\mathrm{keV_{ee}}$. Even a seemingly small reduction in energy threshold has a significant impact on DM reach, given the that expected DM recoil energy spectrum falls steeply with energy.
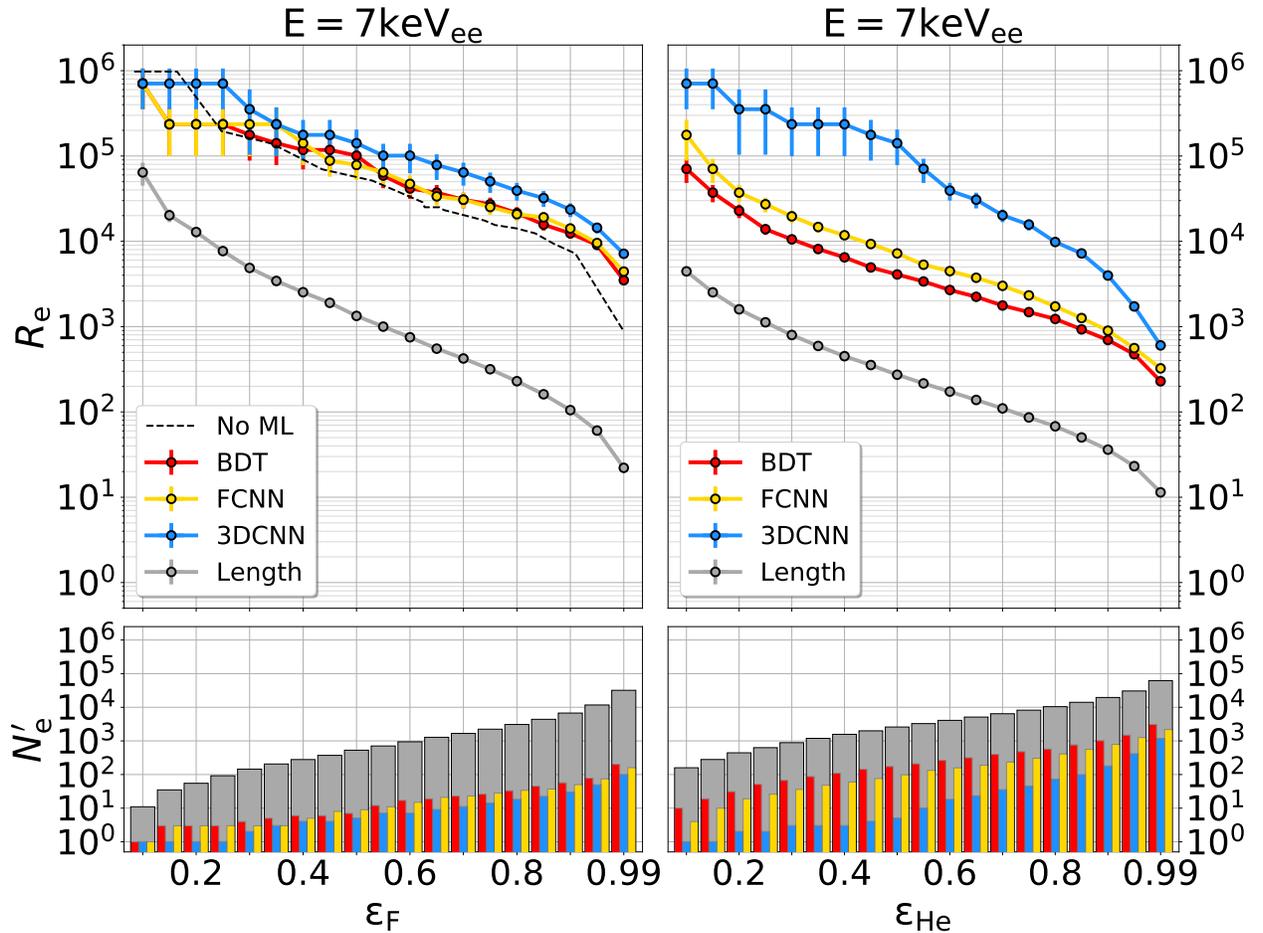
Figure 8.5: Comparison of electron rejection performance of all discriminants versus nuclear recoil efficiency for $7\,\mathrm{keV_{ee}}$ fluorine recoils (left) and $7\,\mathrm{keV_{ee}}$ helium recoils (right). The top and bottom row plots share a common horizontal axis. The black dashed line in the upper left hand plot shows the combined observable from Ref. [97]. The bottom bar plots show the number of remaining electrons after a given recoil selection in each efficiency bin.

Figure 8.5 shows R_{e} and N'_{e} after selections at various F recoil efficiencies (left) and He recoil efficiencies (right) for these same discriminants. We choose to evaluate R_{e} and N'_{e} at a fixed electron-equivalent energy range of $6.5\,\mathrm{keV_{ee}} < E < 7.5\,\mathrm{keV_{ee}}$ because this is the lowest energy range exceeding the CYGNUS background-free criteria of $R_{\mathrm{e}} > 6 \times 10^4$ at both $\varepsilon_{\mathrm{F}} = 0.5$ and $\varepsilon_{\mathrm{He}} = 0.5$. We use the same procedure to compute R_{e} and N'_{e} as in Figure 8.4, except now the values of $\mathrm{p}'_{\mathcal{R}}$, $\mathrm{p}'_{\mathcal{R},\mathrm{BDT}}$, $\mathrm{p}'_{\mathcal{R},\mathrm{FCNN}}$, and L' are separately computed for each $\varepsilon_{\mathcal{R}}$. Since the $\rho_{\mathrm{Max,F}}$, $\mathrm{CThres_F}$, and NClust variables were optimized specifically for $7\,\mathrm{keV_{ee}}$ F recoils, we expect this $7\,\mathrm{keV_{ee}}$ bin will produce the best relative R_{e} of the shallow learning classifiers compared to the 3DCNN. We note here that the combined observable from Ref. [97] was computed at $12\,\mathrm{keV_{ee}}$ for He recoils versus $\varepsilon_{\mathrm{He}}$. Given that the 3DCNN rejects all but two electrons down to $8\,\mathrm{keV_{ee}}$ at $\varepsilon_{\mathrm{He}} = 0.5$ (Figure 8.4), we expect the 3DCNN to reject essentially all $12\,\mathrm{keV_{ee}}$ electron recoils, so we do not investigate the $12\,\mathrm{keV_{ee}}$ scenarios discussed in [97].
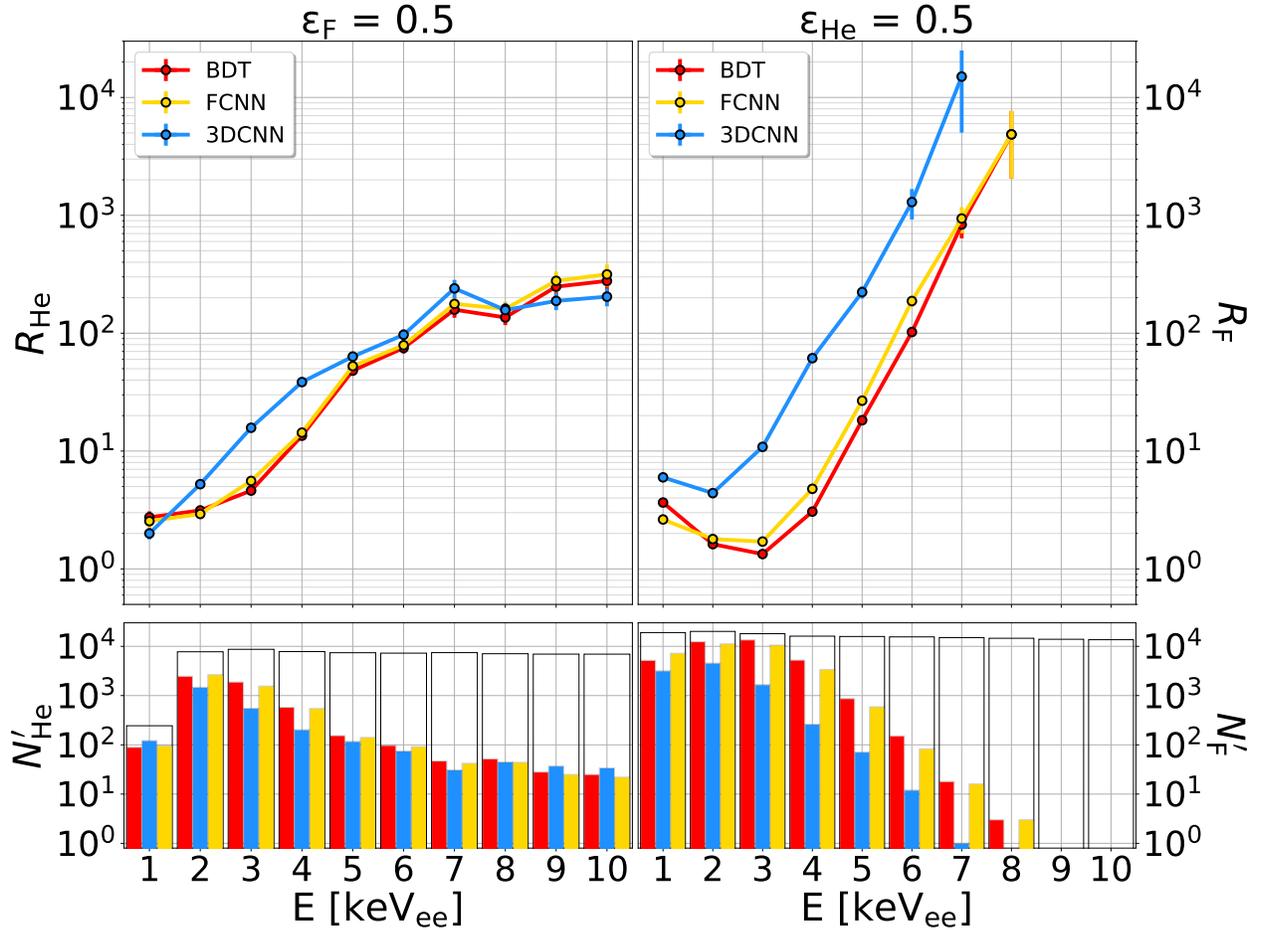
Figure 8.6: Comparison of He and F recoil rejection at fixed F and He signal efficiencies. R_{He} and R_{F} are rejection factors for He and F recoils, respectively and are both defined analogously to R_{e}. The bins in the upper right plot with no data points plotted have all F recoils rejected.

Comparing each multivariate classifier constructed from combining the nine event-shape variables, we find that introducing machine learning for multivariate analysis leads to an improvement in electron rejection above $\varepsilon_F \approx 0.2$. We find further improvement with the use of the 3DCNN, where it improves R_e by more than 2 orders of magnitude over selections on length at higher F recoil efficiencies. For He recoils, we find that the 3DCNN classifier significantly outperforms all other models, often leading to an order of magnitude increase in R_e compared to the BDT and FCNN combinations of the nine electron rejection observables. Furthermore, we find that selecting on 3DCNN output probability at 85% He recoil efficiency leads to a higher R_e than selecting on length at 10% He recoil efficiency, meaning that usage of the 3DCNN classifier could allow an experiment to run with a factor of 8.5 smaller exposure and obtain the same He recoil signal, at an improved background level over using track length to classify events.

An additional benefit of training the machine learning models as ternary classifiers is the ability to use the three output class probabilities to not only reject electron backgrounds, but also classify nuclear recoil species. Figure 8.6 summarizes the nuclear recoil classification ability of the three machine learning classifiers. Above $6\,\text{keV}_{ee}$ all three classifiers maintain 50% F recoil efficiency while rejecting greater than 99% of He recoils. Above $5\,\text{keV}_{ee}$, each classifier starts to do a better job selecting pure samples of He recoils than F recoils, especially the 3DCNN classifier. Selecting pure samples of He recoils is desirable, as low energy He recoils tend to have better angular resolution than F recoils of the same energy, so the ability to select a background-free, almost entirely pure sample of He recoils down to $7\,\text{keV}_{ee}$ at $\varepsilon_{He} = 0.5$ is significant. To quantify this, there are 3,721 He recoils, 1 F recoil, and 5 e recoils in our sample that satisfy $p_{He} > p'_{He}$, where p'_{He} is the 3DCNN He class probability corresponding to $\varepsilon_{He} = 0.5$ in the $7\,\text{keV}_{ee}$ energy bin. This means we have a He recoil purity in excess of 99.8% for $7\,\text{keV}_{ee}$ He recoils at $\varepsilon_{He} = 0.5$. Though the fraction of each event type present in our simulation doesn't model a physically expected event composition, these numbers suggest that use of the 3DCNN will lead to a high He recoil purity for $7\,\text{keV}_{ee}$ He recoils at $\varepsilon_{He} = 0.5$ in a directional DM experiment.

8.6 Robustness of 3DCNN performance

The event identification performance demonstrated with the 3DCNN is promising, however no simulation is perfect, so it is important to test the robustness of classification performance to perturbations in our simulation. We test this by adding $\mathcal{N}$ noise hits to our events and then compare the electron rejection performance of the 3DCNN evaluated on our noisy events ($\mathcal{N} = \mathcal{N}'$) with the electron rejection performance evaluated on the raw events ($\mathcal{N} = 0$). In particular, we implement the following procedure using $\mathcal{N}' = 1, 5, 10$, and 50 noise hits:

1. For each raw event binned for the 3DCNN classifier, generate an empty $2.72 \times 2.72 \times 2.72\text{cm}^3$ voxel grid with the same $850\,\mu\text{m}$ segmentation used previously for the 3DCNN analysis and
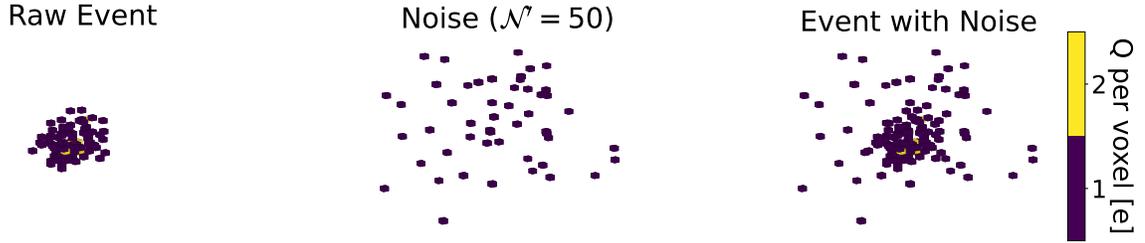
Figure 8.7: Comparison of raw event, a random noise distribution, and the "noisy event" when $\mathcal{N}' = 50$. Left: Raw 3D event display using the $(850 \times 850 \times 850)\mu m^3$ segmentation. Middle: Random 3D noise distribution computed on a $32 \times 32 \times 32$ grid with the same segmentation. Right: The combined "noisy event" containing all points from the raw charge distribution as well as the noise distribution.

      fill it with a random distribution of $\mathcal{N}'$ distinct noise hits. The random noise distribution is in general different for each event.

2. Add the noise voxel grid to the raw event voxel grid to produce "noisy events". Figure 8.7 shows an example of this when $\mathcal{N}' = 50$, where the event in the rightmost panel is the noisy event.

3. Feed the noisy event analogs of each raw event from the test set described in Table 8.1 into the 3DCNN to classify the noisy events. We do not train the 3DCNN on any of the noisy events.

As a baseline reference, we also follow an analogous procedure to test the robustness of electron rejection performance using track length as our classification discriminant. Since the length of an event's principal axis is heavily dependent on geometry, we crop all charge of the centered $100\,\mu m$ segmented events used with our track length classifier to fit within the $2.72 \times 2.72 \times 2.72$ cm^3 voxel grid to match the readout boundaries used for the 3DCNN classifier. We then add $\mathcal{N} = \mathcal{N}'$ noise points within this grid. Figure 8.8 summarizes the electron rejection versus energy performance for both of these classifiers using identical samples of raw events with $\mathcal{N} = \mathcal{N}'$ noise points added, where $\mathcal{N}' = 0$, 1, 5, 10, and 50.

We first note that with 50 noise points added to each event, the electron background rejection performance of the 3DCNN significantly exceeds using track length on events without any noise added. More generally, we find that electron rejection performance of the 3DCNN falls off gracefully with the addition of noise, whereas performance using track length degrades catastrophically. While the addition of up to 10 noise points doesn't have a significant effect on the 3DCNN's classification performance, the addition of just a single noise point significantly reduces the performance of track

161
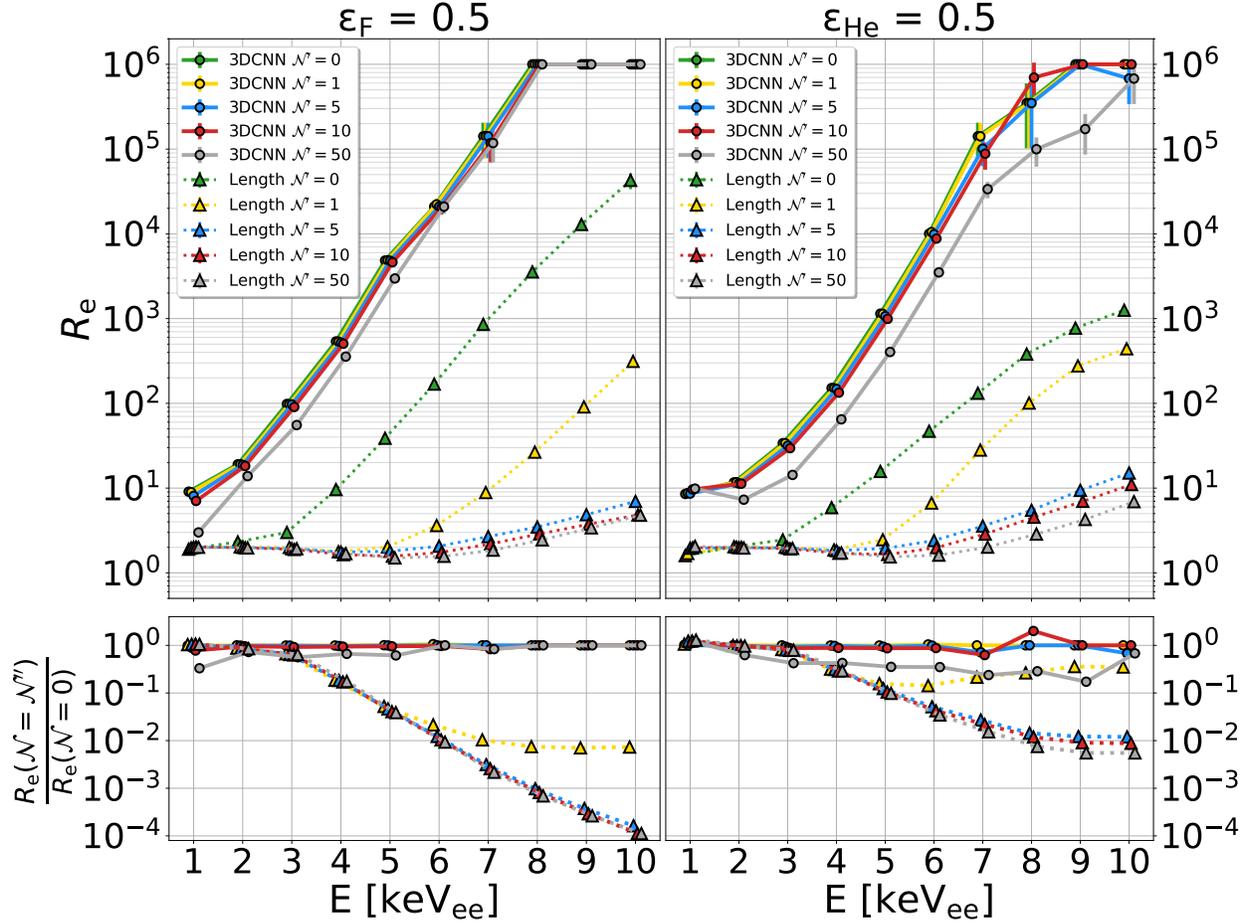
Figure 8.8: Electron rejection versus electron-equivalent energy performance generalization on noisy events using the 3DCNN (circles with solid lines) and track length (triangles with dashed lines) discriminants at 50% fluorine recoil efficiency (left) and 50% helium recoil efficiency (right). The top and bottom row plots share both common horizontal axes and common legends. Points within each energy bin are shifted horizontally for visual clarity. The bottom row plots show the ratio of electron rejection factors with $\mathcal{N}'$ noise points to the electron rejection factors computed using raw events without noise. Note: When $\mathcal{N} = 0$, R_e using track length is slightly lower here than in Figure 8.4 due to us cropping charge outside of the $2.72 \times 2.72 \times 2.72 \text{cm}^3$ grid.

length as a background discriminant. Indeed, when $\mathcal{N} = 1$, below $5\,\mathrm{keV_{ee}}$ track length doesn't have any significant discriminatory power, and above $5\,\mathrm{keV_{ee}}$ electron rejection factors drop by around 2 orders of magnitude for F recoils and by nearly an order of magnitude for He recoils compared to $\mathcal{N} = 0$. Only when $\mathcal{N} = 50$ do we start to see a significant drop in electron rejection performance for the 3DCNN, however the proportional performance drop is still lower than the drop between $\mathcal{N} = 0$ and $\mathcal{N} = 1$ for track length, demonstrating that the 3DCNN is a robust classifier against the addition of random noise.

8.7 Conclusion

We have introduced a deep learning approach to event selection in 3D recoil-imaging TPCs and compared its performance to other state-of-the-art machine learning-based multivariate classifiers. Training a 3D convolutional neural network on the 3D charge distributions of recoil tracks, we found a significant improvement in electron rejection performance between $0.5\,\mathrm{keV_{ee}}$ and $10.5\,\mathrm{keV_{ee}}$ compared to shallow learning methods that form multivariate classification variables from predefined event-shape observables. Notably, the improved performance of the 3DCNN holds when using input data with a reduced spatial segmentation of $(850 \times 850 \times 850)\mathrm{\mu m}^3$ compared to $(100 \times 100 \times 100)\mathrm{\mu m}^3$ segmentation used to compute the observables fed into the shallow learning classifiers. We find that the 3DCNN classifier outperforms other methods by a larger margin for He recoils than for F recoils indicating that the nine event-shape observables used in the BDT and FCNN classifiers are better suited for electron rejection in F recoil samples than in He recoil samples. Since the 3DCNN classifier decides which features for event classification are most useful on its own, using a classifier like this is advantageous, as it seems to have found more useful sets of features to classify nuclear recoils (especially He recoils) than our predefined observables. Defining the background-free energy threshold of our simulated detector to be the lowest energy corresponding to $R_\mathrm{e} \geq 6 \times 10^4$, we find our threshold to be between $6\,\mathrm{keV_{ee}}$ and $7\,\mathrm{keV_{ee}}$ for both F and He recoils at 50% nuclear recoil efficiency when using the 3DCNN output probability to select for events. Using this 3DCNN to classify events effectively reduces the energy threshold of our simulated detector by over 30% for F recoils and around 50% for He recoils compared to classification using track length; a significant improvement over traditional keV-scale recoil identification techniques considering the steeply falling energy spectrum expected for DM recoils. Furthermore the 3DCNN is robust against the addition of random noise and maintains excellent overall electron rejection performance even when classifying events with 50 randomly assigned additional noise hits. The 3DCNN classifier is also able to assign classification probabilities to multiple recoil species simultaneously, making it a flexible approach for event classification that should be of general interest for directional DM searches.

CHAPTER 9
IMPACT AND CONCLUDING REMARKS

We have demonstrated the versatility and robustness of directional recoil detection experimentally and substantially extended its performance through improved analysis techniques and use of machine learning. Operating the 40-cm^3 BEAST TPCs at the SuperKEKB collider—an extremely high background setting—we demonstrated electron background-free operation down to $6\,\mathrm{keV_{ee}}$. Using a singular value decomposition to assign principal axes to recoil tracks, we reduced the energy threshold corresponding to angular resolutions within $20°$ for simulated He recoils, from $100\,\mathrm{keV_{ee}}$ to $60\,\mathrm{keV_{ee}}$. We additionally introduced corrections to charge integration bias for highly inclined nuclear recoil tracks that improved head-tail recognition efficiencies from 72% to 91% for a sample of simulated He recoils ranging from $40\,\mathrm{keV_{ee}}$ to about $1\,\mathrm{MeV_{ee}}$. Applying this technique to assign vector direction on a similar sample of measured nuclear recoils, we found agreement between measured and simulated angular distributions of nuclear recoil tracks, leading to the experimental verification of a radiative Bhabha neutron-production-hotspot in the SuperKEKB tunnel.

On the operational front, we reduced the footprint of each TPC by developing and installing internal GEM divider circuits and improved gas-flow control and stability by adopting serial flow between TPCs in two parallel branches. Operating at modest avalanche-gains of $\mathcal{O}(1{,}000)$, the BEAST TPC system at SuperKEKB has far outlasted its originally planned expiration date of 2018, with six TPCs continuing to collect neutron-background data through the first half of 2022 without the need to replace any internal detector components. Put together, we've demonstrated that the BEAST TPCs, as a technology, are a portable and flexible diagnostic tool that can be easily deployed for directional recoil detection applications. We plan to continue operating the BEAST system at SuperKEKB after the Summer 2022-2024 long shutdown period, which is a testament to the robustness of the TPCs and the effectiveness of the low-gain operation strategy we've employed for fast neutron measurements with these TPCs.

We also presented the first attempt to bridge the performance gap between directional neutron measurements and the expected requirements for directional DM detection using 3DCNNs. We showed that 3DCNNs lead to a significant performance improvement over other approaches in directional head-tail identification in the "neutron detection" regime that includes higher energy nuclear recoils at double-GEM gains of $\mathcal{O}(1{,}000)$. Our ^{252}Cf neutron-source studies demonstrated that the head-tail identification performance of the trained 3DCNN generalizes well to measurement, where we observed strong peaking in the head-tail assignment probability distributions that match the expected direction of recoils from neutrons originating at the source. This peaking translates to the first statistically significant demonstration of event-level head-tail sensitivity below $50\,\mathrm{keV_{ee}}$ for measured nuclear recoils in an atmospheric pressure gas mixture.

Increasing the gain to around 13,000 where we are sensitive to pixel hits from single primary

electrons, we find that the 3DCNN is able to correctly assign head-tail to over 80% of $9\,\text{keV}_\text{ee}$ simulated He recoils. The head-tail effect has yet to be observed for sub-10-keV recoils, so this level of performance is unprecedented, especially in a detector that isn't optimized for low energy directionality. Unfortunately, this performance does not yet generalize to measurement, so more work is needed to improve the robustness of the 3DCNN to low energy recoils.

Finally, we compared the electron background rejection performance of a 3DCNN trained on $0.5\,\text{keV}_\text{ee}$-$10.5\,\text{keV}_\text{ee}$ recoil charge distributions with $(850\,\mu\text{m})^3$ segmentation to using track length and to shallow learning classifiers that combine multiple predefined discriminants that are computed on the same recoil events but using $(100\,\mu\text{m})^3$ segmentation. The recoil events were generated in an 80:10:10 mixture of He:CF$_4$:CFH$_3$ at 60 torr, and despite the courser segmentation of the events the 3DCNN was trained and evaluated on, we found the electron rejection perofrmance of the 3DCNN to significantly outperform the other methods. In particular, if we define a background-free energy threshold as the lowest energy corresponding to $R_\text{e} \geq 6 \times 10^4$, then at 50% F and He recoil efficiency, respectively, this threshold is between $6\,\text{keV}_\text{ee}$ and $7\,\text{keV}_\text{ee}$ using the 3DCNN, which is about a 30% (50%) reduction in threshold compared to using track length for F (He) recoils. Furthermore the 3DCNN is robust against the addition of random noise and maintains excellent overall electron rejection performance even when classifying events with 50 randomly assigned additional noise hits. Given the steeply falling energy spectrum expected for WIMP recoils and that electron rejection becomes exponentially more difficult with decreasing energy, the electron background rejection performance improvements demonstrated with the 3DCNN are significant.

The work presented in this dissertation will appear in four peer-reviewed publications. Two have already been accepted [24; 25], one is currently under peer review [29], and the final one is in preparation.

Appendices

APPENDIX A
ASSIGNING TIMESTAMPS TO EVENTS

In earlier data processing and analysis campaigns between 2018 and early 2021, we were unintentionally discarding good events recorded in the TPCs. Here we describe how the asynchronous triggering in the TPC system works in order to illustrate how we unintentionally discarded events and how that problem has since been resolved.

A common first in first out (FIFO) trigger counter is shared between each enabled FE-I4B chip during a run. This trigger counter starts at 0 at the beginning of a run and increments by 1 each time an event is triggered in a given TPC. The trigger counter logs information into a "Metadata" table on a roughly 20 Hz clock, meaning every ∼0.05 s, the trigger count is logged as the current event number, and timestamp of both the beginning and end of this 20 Hz cycle is logged into this table. Whenever an event is triggered in a given TPC, pixel hit data such as row and column of the (80×336) pixel matrix, the charge deposited in each pixel cell, and the readout time (read out in multiples of 25 ns and not to be confused with the event timestamp) is stored into a "Hits" table. The trigger count from the common FIFO trigger counter is also stored in this table and used as an event counter.

Timestamps are assigned to individual events by merging the data from the Hits table in each TPC with the Metadata table that is shared among all TPCs. Since data recorded in the Hits tables update each time a TPC receives a trigger, it is possible that multiple events will occur within the 20 Hz update window for the Metadata table. In cases like this, subsequent entries in the Metadata table will show trigger count values that differ by more than 1. We illustrate this using Tables A.1 and A.2.

Referring to these two tables, our problem of unintentionally discarding events arose when merging the information between the Metadata table and the Hits table. Originally, when merging these tables, we would simply add the timestamp info from the Metadata table to events in the Hits table with Event Numbers exactly matching. For example Tables A.1 and A.2, this means we would have discarded all events except for event numbers 1, 5, 12, and 14. Given our fixed 20 Hz update period for the Metadata table, one can see that on average, higher proportions of events would be discarded during high beam-background periods using our original merging scheme. We have since fixed our procedure for merging the Hits and Metadata tables by keeping all events in the Hits table and assigning empty timestamps to events where the Event Number in the Hits and Metadata tables don't match up. We then perform a linear interpolation between each subsequent event with filled timestamps to fill the empty timestamp entries. Whenever none of the event numbers listed in the Metadata table match a given event number in a Hits table, we simply assign the timestamp of nearest neighbor event number from the Metadata table as the timestamp for that event in the Hits table. Since the Metadata table updates on a 20 Hz clock, filling timestamps

Event Number	Timestamp [s]
1	ts_0
5	$ts_0 + 0.05$
12	$ts_0 + 0.10$
14	$ts_0 + 0.15$

Table A.1: An example Metadata table showing the trigger counter values at each 20 Hz update starting at our initial timestamp ts_0. The fact that the event numbers increment by more than one is attributed to multiple events being triggered in any number of TPCs within a single 0.05 s window as can be seen in the hits table below.

TPC ID	1	1	1	2	3	1	1	2	1	2	1	2	1	3
Event Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14

Table A.2: An example Hits table omitting all of the pixel hit information and just showing the TPC ID (integer identifier for a given TPC module) and the Event Number to illustrate the process of assigning timestamps to events.

using this procedure gives us a timing uncertainty of 0.05 s.

APPENDIX B
GAIN STABILITY MEASUREMENT FOR HIGH GAIN RUNS

B.1 Run conditions

Before determining the single electron efficiency point detailed in Chapter 7, we performed a gain stability study where we recorded 115 hours of data using an ^{55}Fe electron-capture X-ray source immediately after sealing the inside of the TPC off from outside air and beginning He:CO_2 gas flow. We had not yet calibrated the gain at this point, but wanted to be operating in the single electron efficiency regime, so we made the *a priori* choice to set our input gain point to 20,000. Using the figure of merit from Ref. [76], we can estimate the voltage across both GEMs that corresponds to this target gain:

$$G = 10^{\frac{V_{\mathrm{GEM}} - V_1}{V_2}}, \tag{B.1}$$

where V_{GEM} is the voltage across the two GEMs, and V_1 and V_2 were determined to be $356\,\mathrm{V}$ and $136\,\mathrm{V}$ in this reference for $5.9\,\mathrm{keV}$ X-rays in $1\,\mathrm{atm}$ of He:CO_2. Substituting 20,000 in for G in Equation (B.1), we find $V_{\mathrm{GEM}} = 941\,\mathrm{V}$. To get the GEM divider circuit input voltage, V_{in}, that corresponds to V_{GEM}, we refer to previous gain calibrations with ^{210}Po-α emission sources. These calibrations suggest gains of $\mathcal{O}(1000)$ with $V_{\mathrm{in}} = 2100\,\mathrm{V}$ and a field-cage voltage $V_{\mathrm{FC}} = 7000\,\mathrm{V}$. Substituting G into Equation (B.1), and using the previously mentioned V_1 and V_2 values, we find $V_{\mathrm{GEM}} = 760\,\mathrm{V}$ when $V_{\mathrm{in}} = 2100\,\mathrm{V}$. Using this, we determine the V_{in} that corresponds to our target gain of 20,000

$$V_{\mathrm{in}}|_{\mathrm{G}=20000} = \frac{V_{\mathrm{GEM}}|_{\mathrm{G}=20000}}{V_{\mathrm{GEM}}|_{\mathrm{G}=1000}} V_{\mathrm{in}}|_{\mathrm{G}=1000}, \tag{B.2}$$

and find $V_{\mathrm{in}} = 2600\,\mathrm{V}$. To maintain a constant drift field, whenever we make adjustments to V_{in}, we make an equal adjustment to the input field-cage voltage V_{FC}. To maintain the $450\,\mathrm{V/cm}$ drift field used in other studies (Table 4.3), we set $V_{\mathrm{FC}} = 7500\,\mathrm{V}$.

During our 115 hours of data collection, we primarily record runs that are 15 minutes in length. The ^{55}Fe X-ray source inside the TPC is housed in Delrin shielding with a movable shutter. We test the purity of the ^{55}Fe X-ray sample by closing the shutter every other 15-minute run. Figure B.1 shows a 2D histogram of fiducialized ^{55}Fe X-rays with the shutter open. The magenta points overlaid on top show the background sample of events recorded with the source shutter closed over the same amount of time. Comparing the counts in the samples, we find nearly 3 orders of
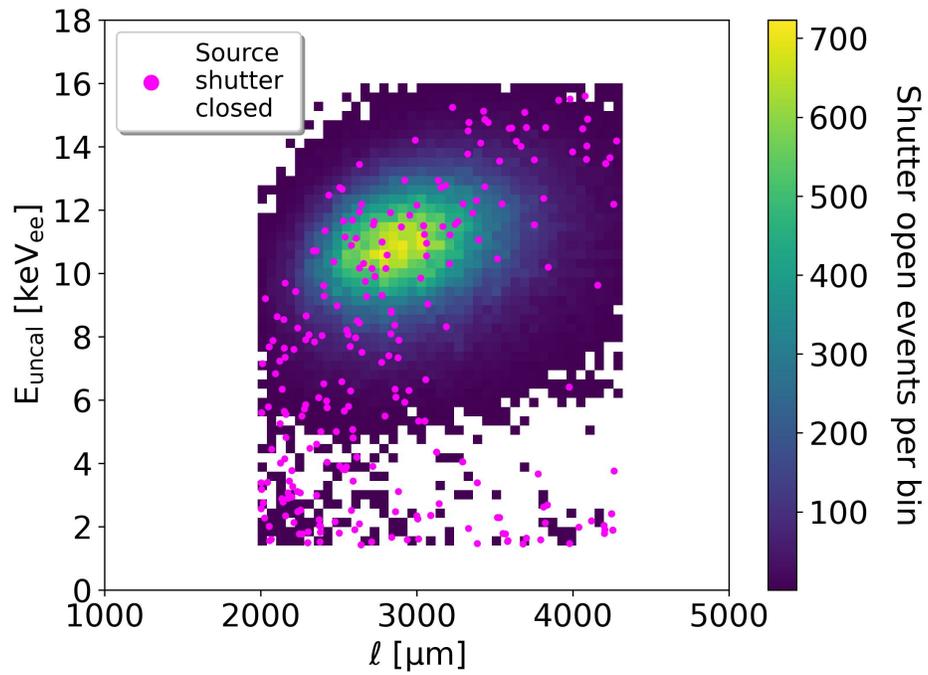
Figure B.1: Uncalibrated energy versus track length distribution of fiducialized events near the expected X-ray energy spectrum peak from the ^{55}Fe source. The 2D histogram shows events recorded when the source shutter was open and the magenta points show events recorded over the same amount of elapsed time with the source shutter closed.

magnitude more events in the sample with the source shutter open, suggesting a very pure sample of ^{55}Fe X-rays when the shutter is open that we can use for gain measurements.

B.2   Gain measurements

We split the shutter-open sample of events by timestamp and partition the sample into sub-samples that each contain 5 hours of successive measurements. We then plot the energy spectra of the events in each of the sub-samples and fit a Gaussian profile with linear background of the form

$$\hat{P}(E_{\text{uncal}}; A, B, C, \hat{\mu}, \hat{\sigma}) = A + BE_{\text{uncal}} + Ce^{-\left(\frac{E_{\text{uncal}}-\hat{\mu}}{\sqrt{2}\hat{\sigma}}\right)^2}, \tag{B.3}$$

to the histogram of each sub-sample, where parameters A, B, C, $\hat{\mu}$, and $\hat{\sigma}$ are fit parameters to the given sub-sample energy spectra. We compute the gain of each sub-sample as

$$G = \frac{\hat{\mu}}{5.9\,\text{keV}} \times 20000 \pm 1.96\frac{\sigma_G}{\sqrt{N}} \tag{B.4}$$

where we have scaled the X-ray energy spectra to a mean of $5.9\,\text{keV}$, $\sigma_G = \hat{\sigma} \times \frac{20000}{5.9\,\text{keV}}$, and N is the number of events in the given sub-sample. Fits to 21 such sub-samples are shown in Figure B.2, where we observe the peak of the spectra to increase in energy with time and eventually level off (we omitted two samples due to lower measurement statistics). To quantify gain stability, we fit the gain recorded in the time bin of each sub-sample using an exponential of the form

$$\hat{G}(t) = \hat{\alpha}\left(1 - e^{-\hat{\beta}(t-\hat{\gamma})}\right). \tag{B.5}$$

Here $\hat{\alpha}$ corresponds to the nominal gain, $\hat{\beta}$ encodes the gain rise time, and $\hat{\gamma}$ is a fit offset parameter. From this fit, we find $5/\hat{\beta} = 96$ hours and corresponds to $1 - 1/e^5 \approx 99.3\%$ of the nominal max gain, so we conclude we have achieved sufficiently stable gain at this point. Figure B.3 shows the gain versus time distribution with its corresponding fit from Equation (B.5), as well as the gain resolution (σ_G/G) versus time. The gain resolution drops off quickly and stabilizes at around 11% after about 20 hours of flow.
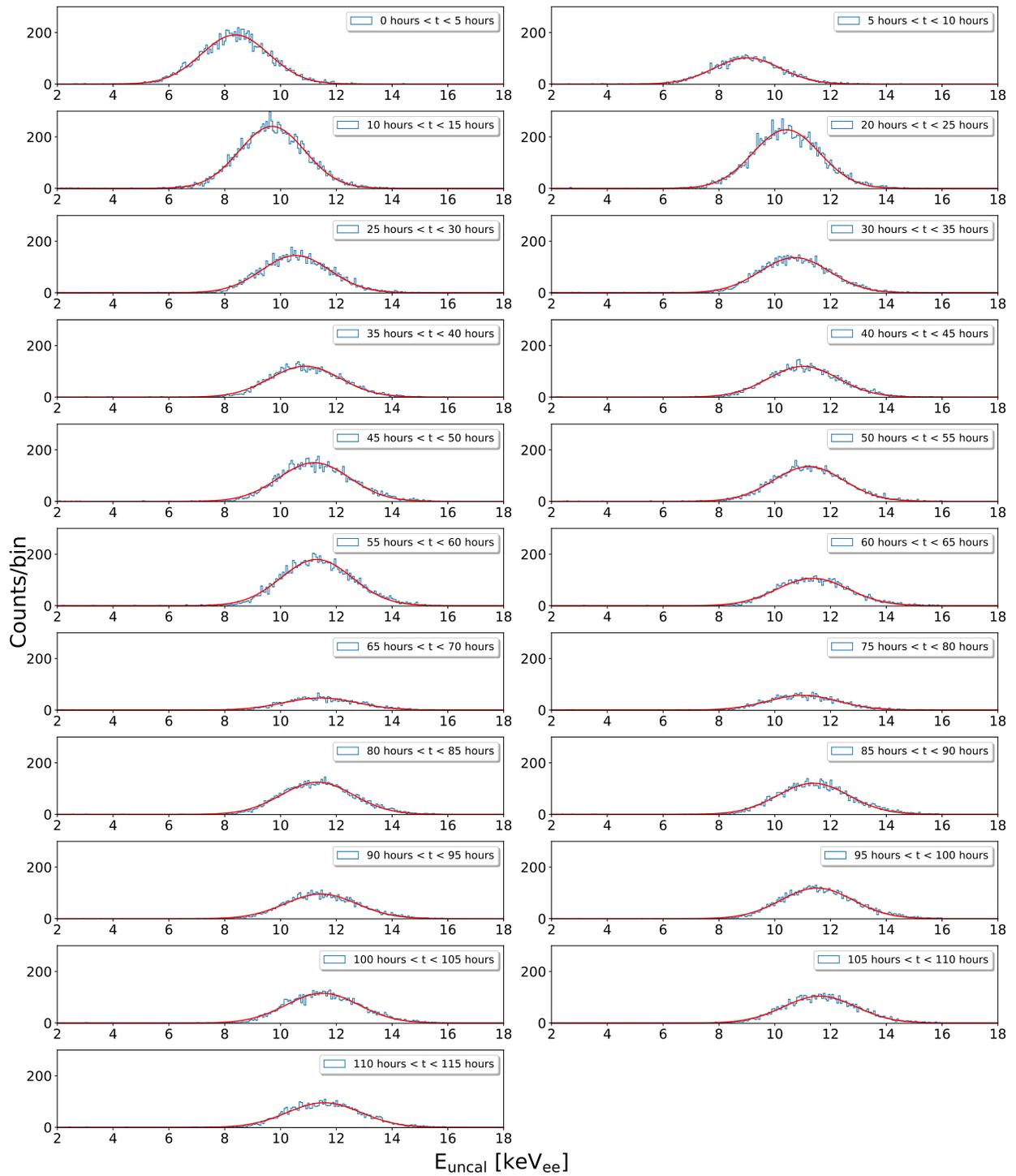
Figure B.2: Uncalibrated X-ray energy spectra at various time slices. The red curve shows a Gaussian profile fit with linear background to each distribution. The spectra shifting to the right over time is an indication of the gain increasing due to increasing He:CO_2 gas purity.
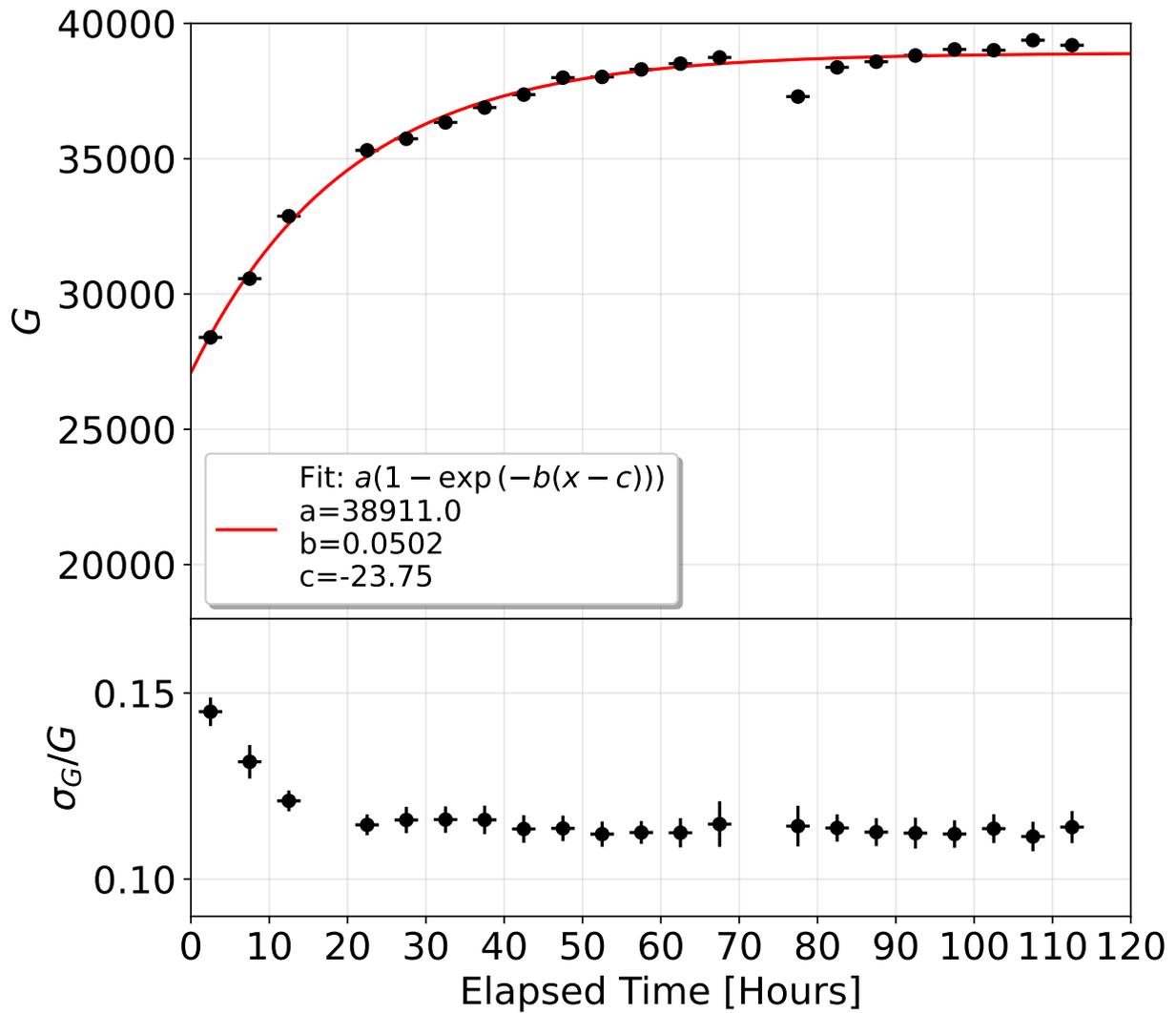
Figure B.3: Top: Gain determined by scaling the profile fits from Equation (B.3) to $5.9\,\mathrm{keV}$ versus time. Bottom: Gain resolution determined from the profile fits of Equation (B.3) versus time.

APPENDIX C
FURTHER DETAILS OF SIMULATION PACKAGES AND PARAMETERS

Here we outline the tools and parameters used for select simulations in this thesis. We include all simulations except for the fast simulation used in Phase 2 (Chapter 3), as these simulations were not thoroughly logged.

C.1 Glossary of tools

1. Event generators: Geant4, SAD, BBBREM, AAFH, SRIM, retrim, and DEGRAD

2. TPC simulators: basf2: BEAST geometry

3. Code for digitizer used throughout this thesis

C.2 Chapter 3: Phase 3 neutron simulation

For these studies we use SAD, BBREM, and AAFH to generate beam backgrounds that are passed into Geant4. The primary ionization distributions of recoil tracks in the TPC are generated in Geant4 and then passed into the digitizer.

Toolkits	G4PhysicsList	Neutron cross sections and final states	He, C, and O cross section scale factor	W [eV]	F
Geant4 v10.6.1 SAD (Single beam) BBBREM (RBB) AAFH (2γ)	FTFP_BERT_HP	G4NDL4.6 JEFF3.3	100	35	0.19

Table C.1: Summary of simulation of neutron interactions and primary tracks in the TPCs for Phase 3 Study A. The interaction cross section between neutrons and He, C, and O nuclei are dialed up by a factor of 100 in Geant4 to increase nuclear recoil statistics. Simulated rates are then scaled down by a factor of 100 when comparing observed and simulated rates. W and F are the average energy per electron-ion pair and Fano factor, respoectively and are used to generate the primary recoil ionization distribution in the TPCs.

	I[A]	σ_y[μm]	n_b	Luminosity $[\mathrm{cm}^{-2}\mathrm{s}^{-1}]$
LER	1.2	37	1576	2.5×10^{35}
HER	1.0	36	1576	

Table C.2: Copy of Table 4.1: Machine conditions and luminosity used for SAD simulation and BBBREM / AAFH event generators, respectively.

Background Type	Coulomb	Brems	Touschek	RBB	2γ
Simulated Beam Time [s] (LER,HER)	(4,40)	(40,400)	(0.4,1.6)	0.0097	0.01

Table C.3: Copy of Table 4.2: Total simulated beam time for each background process. Values within the parenthetical numerical pairs denote the beam time of single-beam simulation samples in each ring.

TPC z Location	Gas [70%:30%]	Gain	$\vec{v}_d$ [μm/BCID]	(σ_T, σ_L) [μm/$\sqrt{\mathrm{cm}}$]	$(\sigma_{T,\mathrm{GEM}}, \sigma_{L,\mathrm{GEM}})$ [μm]
$-14\,\mathrm{m}$	$He:CO_2$	797	151.68	(127, 127)	(180,180)
$-8.0\,\mathrm{m}$	$He:CO_2$	807	216.25	(127, 127)	(180,180)
$-5.6\,\mathrm{m}$	$He:CO_2$	1033	216.25	(127, 127)	(180,180)
$6.6\,\mathrm{m}$	$He:CO_2$	1476	216.25	(127, 127)	(180,180)
$14\,\mathrm{m}$	$He:CO_2$	899	216.25	(127, 127)	(180,180)
$16\,\mathrm{m}$	$He:CO_2$	878	216.25	(127, 127)	(180,180)

Table C.4: Digitization parameters for each TPC. Primary ionization distributions were generated in Geant4 using the appropriate generators (SAD for single beams, BBBREM for RBB, and AAFH for 2γ).

C.3 Chapter 4: Isotropic α-particle low gain simulation

α particles are produced at the center of the fiducial (x,y) volume of the TPC in Geant4 forming the primary ionization distributions that are passed into the digitizer. The vertices of all α particles are at $z_{\mathrm{TPC}} = 5\,\mathrm{cm}$. We use $W = 35\,\mathrm{eV}$ and $F = 0.19$.

Toolkit	G4PhysicsList	Momentum range [GeV]	θ range [°]	ϕ range [°]	E_{recoil} range [keV$_{\mathrm{r}}$]
Geant4 v10.6.1 (ParticleGun)	FTFP_BERT_HP	$[-0.051, 0.051]$	$[0, 180]$	$[0, 360]$	$[0.035, 525]$

Table C.5: Summary of the ParticleGun module parameters used to simulate the 612,505 He recoils described in Chapter 5.2. The momentum distribution and ϕ distributions are uniform and the θ distribution is isotropic (uniform $\cos(\theta)$).

Gas [70%:30%]	Gain	$\vec{v}_d$ [µm/BCID]	(σ_T, σ_L) [µm/$\sqrt{\mathrm{cm}}$]	$(\sigma_{T,\mathrm{GEM}}, \sigma_{L,\mathrm{GEM}})$ [µm]
He:CO$_2$	909	216.25	(114, 114)	(180,180)

Table C.6: Copy of Table 5.1: Digitization parameters for the sample of 612,505 He recoils described in Chapter 5.2. $\sigma_{T,\mathrm{GEM}}$ and $\sigma_{L,\mathrm{GEM}}$ represent the transverse and longitudinal point resolution of the readout plane, excluding the pixel chip. The W is the low energy cutoff we use for recording ionization.

C.4 Chapter 5: Neutrons generated with particle gun

We only show the ParticleGun neutron simulation parameters used to mimic the configuration of the ^{252}Cf source setup in our lab. The neutron production vertices are $(x_{\text{TPC}}, y_{\text{TPC}}, z_{\text{TPC}}) = (\pm 13, 0.84, 5.0)$ cm with the '+' and '−' in x_{TPC} corresponding to the source incident on the lid and bottom of the TPC, respectively. The digitization parameters and all other simulation parameters are identical to what was used in Chapter 5.

Toolkit	Neutrons Simulated [$\times 10^7$] (lid, bottom)	Momentum range [GeV]	θ range [°]	ϕ range [°]	Recoils after selections (lid, bottom)
Geant4 v10.6.1 (ParticleGun)	(6,6)	$[0.0015, 0.081]$	$[88.5, 91.5]$	$[178.5, 181.5]$	(34102,32056)

Table C.7: Summary of the ParticleGun module parameters used to simulate the neutrons producing the simulated neutron-induced nuclear recoil sample described in Chapter 6. We use the FTFP_BERT_HP physics list. The momentum distribution and ϕ distributions are uniform and the θ distribution is isotropic (uniform $\cos(\theta)$). Figure 6.4 shows the nuclear recoil selection boundary used. We only keep truth He recoils in our simulated samples.

C.5 Chapter 6: High gain simulation

We generate all primary ionization distributions passed into the digitizer using Geant4's ParticleGun module. We run campaigns shooting the following from the ParticleGun:

1. Electrons (isotropic distribution)

2. α particles (isotropic distribution)

3. Low energy neutron beam with the same θ and ϕ envelope as in Table C.7

The electrons and α particles have vertices at the center of the fiducial (x,y) volume of the TPC and $z_{\text{TPC}} = 8$ cm. We use $W = 35$ eV and $F = 0.19$.

Toolkit	Momentum range [GeV]	θ range [°]	ϕ range [°]	E_{recoil} range [keV$_{\text{ee}}$]
Geant4 v10.6.1 (ParticleGun)	$[-0.02, 0.02]$	$[0, 180]$	$[0, 360]$	$[0.035, 15.9]$

Table C.8: Summary of the ParticleGun module parameters used to simulate the 675,326 electrons used for the event selection 3DCNN.

Toolkit	θ range [°]	ϕ range [°]	E_{recoil} range [keV$_{\text{r}}$]
Geant4 v10.6.1 (ParticleGun)	$[0, 180]$	$[0, 360]$	$[0.035, 50]$

Table C.9: Summary of the ParticleGun module parameters used to simulate the 101,739 α particles used for the event selection and head-tail 3DCNN. The momentum range was not logged for this.

Toolkit	θ range [°]	ϕ range [°]	E_{recoil} range [keV$_{\text{r}}$]	n production vertex (TPC coordinates [cm])
Geant4 v10.6.1 (ParticleGun)	$[88.5, 91.5]$	$[178.5, 181.5]$	$[1, 35]$	$(\pm 13, 0.84, 8.0)$

Table C.10: Summary of the ParticleGun module parameters used to simulate the 269,745 He, C, and O nuclear recoil samples generated from the neutron beam generated from the ParticleGun module. The neutron production vertices with $x_{\text{TPC}} > 0$ correspond to the source incident on lid of the TPC ($x_{\text{TPC}} < 0$ vertex is source-on-bottom). The number of neutrons generated to produce these recoils and the momentum range of the neutrons were not logged.

Gas [70%:30%]	Gain	$\vec{v}_d$ [µm/BCID]	(σ_T, σ_L) [µm]	$(\sigma_{T,\text{GEM}}, \sigma_{L,\text{GEM}})$ [µm]
He:CO$_2$	15,000	216.25	(134.8, 128.2)	(143,97)

Table C.11: Copy of Table 7.1: Digitization parameters for the high gain head-tail simulation campaign. $\sigma_{T,\text{GEM}}$ and $\sigma_{L,\text{GEM}}$ represent the transverse and longitudinal point resolution of the readout plane, excluding the pixel chip. The diffusion parameters were computed using `Magboltz` and should be more representative of measurement than the diffusion coefficients used in Table C.6.

C.6 Chapter 7: Simulation for electron rejection study in He:CF$_4$:CHF$_3$

We use `SRIM` and `retrim` to simulate recoiling He and F nuclei and `DEGRAD` to simulate electron recoils in an 80:10:10 mixture of He:CF$_4$:CHF$_3$ at a total pressure of 60 Torr and temperature of 25°C

Gas [80%:10%:10%]	W [eV]	W_e [eV]	F	$\vec{E}_d$ [V/cm]	(σ_T, σ_L) [µm/$\sqrt{\text{cm}}$]	C_{ij}			
He:CF$_4$:CHF$_3$	35.0	32.4	0.19	40.6	(398, 425)	1.00 0.957 0.969	1.00 0.959 0.965	1.00 0.974 0.991	1.00 0.974 0.991

Table C.12: Key simulation parameters used for this study. W_e is the average energy per electron-ion pair used to compute electron equivalent energies, and F is the Fano factor. C_{ij} shows the compound correction factors used in `SRIM` and `retrim`. The columns of C represent $j = \{$He, H, C, and F$\}$, respectively and the rows of C represent $i = \{$He, CH$_3$, and CHF$_3\}$, respectively. The second row and fourth column of C would thus be read as the compound correction of F recoils in CF$_4$.

APPENDIX D
CODE AVAILABILITY

Select simulation and data processing code is available here upon request.

BIBLIOGRAPHY

[1] F. Zwicky, *On the Masses of Nebulae and of Clusters of Nebulae*, *The Astrophysical Journal* **86** (1937) 217.

[2] V.C. Rubin and J. Ford, W. Kent, *Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions*, *The Astrophysical Journal* **159** (1970) 379.

[3] S.D.M. White, C.S. Frenk and M. Davis, *Clustering in a neutrino-dominated universe*, *The Astrophysical Journal Letters* **274** (1983) L1.

[4] G.F. Smoot, C.L. Bennett, A. Kogut, E.L. Wright, J. Aymon, N.W. Boggess et al., *Structure in the COBE Differential Microwave Radiometer First-Year Maps*, *The Astrophysical Journal Letters* **396** (1992) L1.

[5] WMAP collaboration, *Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Data Processing, Sky Maps, and Basic Results*, *Astrophys. J. Suppl.* **180** (2009) 225 [`0803.0732`].

[6] Planck collaboration, *Planck 2015 results. XIII. Cosmological parameters*, *Astron. Astrophys.* **594** (2016) A13 [`1502.01589`].

[7] Supernova Cosmology Project collaboration, *Measurements of Ω and Λ from 42 high redshift supernovae*, *Astrophys. J.* **517** (1999) 565 [`astro-ph/9812133`].

[8] Supernova Search Team collaboration, *Observational evidence from supernovae for an accelerating universe and a cosmological constant*, *Astron. J.* **116** (1998) 1009 [`astro-ph/9805201`].

[9] L. Perivolaropoulos and F. Skara, *Challenges for ΛCDM: An update*, *New Astron. Rev.* **95** (2022) [`2105.05208`].

[10] R.H. Wechsler and J.L. Tinker, *The Connection between Galaxies and their Dark Matter Halos*, *Ann. Rev. Astron. Astrophys.* **56** (2018) 435 [`1804.03097`].

[11] P.R. Kafle, S. Sharma, G.F. Lewis and J. Bland-Hawthorn, *On the Shoulders of Giants: Properties of the Stellar Halo and the Milky Way Mass Distribution*, *Astrophys. J.* **794** (2014) 59 [`1408.1787`].

[12] S.E. Vahsen, C.A.J. O'Hare and D. Loomba, *Directional Recoil Detection*, *Ann. Rev. Nucl. Part. Sci.* **71** (2021) 189 [`2102.04596`].

[13] J. Bailin and M. Steinmetz, *Figure rotation of cosmological dark matter halos*, *The Astrophysical Journal* **616** (2004) 27.

[14] G. Arcadi, M. Dutra, P. Ghosh, M. Lindner, Y. Mambrini, M. Pierre et al., *The waning of the WIMP? A review of models, searches, and constraints*, *Eur. Phys. J. C* **78** (2018) 203 [1703.07364].

[15] M.W. Goodman and E. Witten, *Detectability of certain dark-matter candidates*, *Phys. Rev. D* **31** (1985) 3059.

[16] D.N. Spergel, *Motion of the Earth and the detection of weakly interacting massive particles*, *Phys. Rev. D* **37** (1988) 1353.

[17] J. Billard, E. Figueroa-Feliciano and L. Strigari, *Implication of neutrino backgrounds on the reach of next generation dark matter direct detection experiments*, *Phys. Rev. D* **89** (2014) 023524.

[18] C.A.J. O'Hare, *New definition of the neutrino floor for direct dark matter searches*, *Phys. Rev. Lett.* **127** (2021) 251802.

[19] D.R. Nygren and J.N. Marx, *The time projection chamber*, *Physics Today* **31** (1978) 46 [https://doi.org/10.1063/1.2994775].

[20] S.E. Vahsen et al., *CYGNUS: Feasibility of a nuclear recoil observatory with directional sensitivity to dark matter and neutrinos*, 2008.12587.

[21] P.M. Lewis, S.E. Vahsen, I.S. Seong, M.T. Hedges, I. Jaegle and T.N. Thorpe, *Absolute Position Measurement in a Gas Time Projection Chamber via Transverse Diffusion of Drift Charge*, *Nucl. Instrum. Meth. A* **789** (2015) 81 [1410.1131].

[22] Y. Funakoshi and Y. Ohnishi, *Highlights from SuperKEKB Commissioning Phase 1 and Plan for Phase 2*, *ICFA Beam Dyn. Newslett.* **72** (2017) 11.

[23] I. Jaegle et al., *Compact, directional neutron detectors capable of high-resolution nuclear recoil imaging*, *Nucl. Instrum. Methods Phys. Res., A* **945** (2019) 162296.

[24] Z. Liptak et al., *Measurements of beam backgrounds in SuperKEKB Phase 2*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **1040** (2022) .

[25] J. Schueler, S. Vahsen, P. Lewis, M. Hedges, D. Liventsev, F. Meier et al., *Application of recoil-imaging time projection chambers to directional neutron background measurements in the superkekb accelerator tunnel*, *Nuclear Instruments and Methods in Physics Research*

Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **1040** (2022) 167291.

[26] D. Dujmic, S. Ahlen, P. Fisher, S. Henderson, A. Kaboth, G. Kohse et al., *Improved measurement of the 'head-tail' effect in nuclear recoils, Journal of Physics: Conference Series* **120** (2008) 042030.

[27] S. Burgos, E. Daw, J. Forbes, C. Ghag, M. Gold, C. Hagemann et al., *First measurement of the head–tail directional nuclear recoil signature at energies relevant to WIMP dark matter searches, Astroparticle Physics* **31** (2009) 261.

[28] R. Yakabe et al., *First limits from a 3D-vector directional dark matter search with the NEWAGE-0.3b' detector, PTEP* **2020** (2020) 113F01 [`2005.05157`].

[29] J. Schueler, M. Ghrear, S.E. Vahsen, P. Sadowski and C. Deaconu, *Deep learning for improved keV-scale recoil identification in high resolution gas time projection chambers,* `2206.10822`.

[30] Y. Ohnishi, et al., *Accelerator design at SuperKEKB, Progress of Theoretical and Experimental Physics* **2013** (2013) [`https://academic.oup.com/ptep/article-pdf/2013/3/03A011/4439973/pts083.pdf`].

[31] N. Toge, *KEK B-factory Design Report*, Tech. Rep. KEK-Report-95-7, KEK, Tsukuba (1995).

[32] P.M. Lewis et al., *First Measurements of Beam Backgrounds at SuperKEKB, Nucl. Instrum. Meth.* **A914** (2019) 69 [`1802.01366`].

[33] The Belle II Collaboration, *Belle II Technical Design Report*, Tech. Rep. arXiv:1011.0352. KEK REPORT 2010-1 (Nov, 2010).

[34] A. Natochii et al., *Beam background expectations for Belle II at SuperKEKB*, in *2022 Snowmass Summer Study*, 3, 2022 [`2203.05731`].

[35] PARTICLE DATA GROUP collaboration, *Review of Particle Physics, PTEP* **2020** (2020) 083C01.

[36] SUPERKEKB collaboration, *SuperKEKB Collider, Nucl. Instrum. Meth. A* **907** (2018) 188 [`1809.01958`].

[37] SUPERB collaboration, *SuperB: A High-Luminosity Asymmetric e+ e- Super Flavor Factory. Conceptual Design Report,* `0709.0451`.

[38] T. Ishibashi, S. Terui, Y. Suetsugu, K. Watanabe and M. Shirai, *Movable collimator system for SuperKEKB*, *Phys. Rev. Accel. Beams* **23** (2020) 053501.

[39] A. Natochii et al., "Measured and projected beam backgrounds in the Belle II experiment at the SuperKEKB collider." (In Progress).

[40] Y. Hoshi, N. Kikuchi, T. Nagamine, K. Neichi and A. Yamaguchi, *Performance of the endcap RPC in the Belle detector under high luminosity operation of the KEKB accelerator*, *Nuclear Physics B - Proceedings Supplements* **158** (2006) 190 .

[41] K. Abe, Y. Hoshi, T. Nagamine, K. Neichi, K. Onodera, T. Takahashi et al., *Neutron sensitivity of the endcap RPC modules in Belle detector*, IEEE Transactions on Nuclear Science **50** (2003) 831.

[42] T. Aushev, D. Besson, K. Chilikin, R. Chistov, M. Danilov, P. Katrenko et al., *A scintillator based endcap KL and muon detector for the Belle II experiment*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **789** (2015) 134?142.

[43] D. Kotchetkov et al., *Front-end electronic readout system for the Belle II imaging Time-Of-Propagation detector*, *Nucl. Instrum. Meth. A* **941** (2019) 162342 [1804.10782].

[44] M. Cecchetto et al., *Thermal Neutron Induced SEUs in the LHC Accelerator Environment*, *IEEE Transactions on Nuclear Science* **PP** (2020) 1.

[45] B.R. Bortignon P.F., Bracco A., *Giant Resonances, in: Contemporary Concepts in Physics*, vol. 10, CRC Press (1998).

[46] *Giant dipole resonance neutron yields produced by electrons as a function of target material and thickness*, *Health Phys* **70** (1996) 207 .

[47] M.T. Hedges, *Performance and First Deployment of Novel 3d Nuclear Recoil Detectors*, Ph.D. thesis, Hawaii U., 2018.

[48] M.T. Hedges, S.E. Vahsen, I. Jaegle, P.M. Lewis, H. Nakayama, J. Schueler et al., *First 3D vector tracking of helium recoils for fast neutron measurements at SuperKEKB*, *Nucl. Instrum. Meth. A* **1026** (2022) 166066 [2106.13079].

[49] S. de Jong, *Study of Thermal Neutron Flux from SuperKEKB in the Belle II Commissioning Detector*, Ph.D. thesis, University of Victoria, https://dspace.library.uvic.ca//handle/1828/8203, 2017.

[50] D. Cuesta et al., *Operation of a double-sided CMOS pixelated detector at a high intensity e^+e^- particle collider*, *Nucl. Instrum. Meth. A* **967** (2020) 163862 [2002.06941].

[51] H. Windel, *Claws - An Injection Background Monitoring System for the Second and Third Phase of the SuperKEKB Commissioning*, Ph.D. thesis, Max Planck Institute for Physics, 2021.

[52] M. Gabriel et al., *A time resolved study of injection backgrounds during the first commissioning phase of SuperKEKB*, *Eur. Phys. J. C* **81** (2021) 972 [2012.10948].

[53] P. Ahlburg, *Development of a FE-I4-based module for radiation monitoring with BEAST II during the commissioning phase of the Belle II detector*, Master's thesis, Universität Bonn, 2016.

[54] A. Natochii, S.E. Vahsen, H. Nakayama, T. Ishibashi and S. Terui, *Improved simulation of beam backgrounds and collimation at SuperKEKB*, *Phys. Rev. Accel. Beams* **24** (2021) 081001 [2104.02645].

[55] S. Vahsen et al., *The Directional Dark Matter Detector (D3)*, *EAS Publications Series* **53** (2012) 43.

[56] F. Sauli, *The gas electron multiplier (GEM): Operating principles and applications*, *Nucl. Instrum. Meth. A* **805** (2016) 2.

[57] M. Garcia-Sciveres, D. Arutinov, M. Barbero, R. Beccherle, S. Dube, D. Elledge et al., *The FE-I4 pixel readout integrated circuit*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **636** (2011) S155.

[58] The ATLAS IBL collaboration, *Prototype ATLAS IBL modules using the FE-i4a front-end readout chip*, *Journal of Instrumentation* **7** (2012) P11010.

[59] The ATLAS IBL collaboration, *Production and integration of the ATLAS insertable b-layer*, *Journal of Instrumentation* **13** (2018) T05008.

[60] "Experimental physics and industrial control system." Available at https://epics.anl.gov.

[61] "Strategic accelerator design." Available at http://acc-physics.kek.jp/SAD/.

[62] GEANT4 collaboration, *GEANT4–a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250.

[63] J. Allison et al., *Geant4 developments and applications*, *IEEE Trans. Nucl. Sci.* **53** (2006) 270.

[64] J. Allison et al., *Recent developments in Geant4*, *Nucl. Instrum. Meth. A* **835** (2016) 186.

[65] Belle-II Framework Software Group collaboration, *The Belle II Core Software*, *Comput. Softw. Big Sci.* **3** (2019) 1 [`1809.04299`].

[66] Belle-II Framework Software Group collaboration, *Belle II Analysis Software Framework (basf2): release-06-00-00*, Oct., 2021. 10.5281/zenodo.5574116.

[67] M. Chadwick et al., *ENDF/B-VII.1 Nuclear Data for Science and Technology: Cross Sections, Covariances, Fission Product Yields and Decay Data*, *Nuclear Data Sheets* **112** (2011) 2887 .

[68] J. Janssen and D. Pohl, "pyBAR - Bonn ATLAS Readout in Python."

[69] S. Biagi, "Magboltz-transport of electrons in gas mixtures." Available at https://magboltz.web.cern.ch/magboltz/.

[70] G.H. Golub and C. Reinsch, *Singular value decomposition and least squares solutions*, *Numer. Math.* **14** (1970) 403–420.

[71] S. Park, Y. Kwon, M. Nakao, U. Sadaharu and T. Konno, *Environmental Monitoring for Belle II*, in *21st IEEE Real Time Conference*, 6, 2018 [`1806.08933`].

[72] A. Piwinski, *The Touschek effect in strong focusing storage rings*, `physics/9903034`.

[73] T. Ferber and P. Urquijo, *Overview of the Belle II Physics Generators*, .

[74] E. Mendoza and D. Cano-Ott, *Update of the Evaluated Neutron Cross Section Libraries for the Geant4 Code*, June, 2018.

[75] R. Veenhof and H. Schindler, *Garfield++*, 2018.

[76] S. Vahsen, M. Hedges, I. Jaegle, S. Ross, I. Seong, T. Thorpe et al., *3-D Tracking in a Miniature Time Projection Chamber*, *Nuclear Instruments and Methods in Physics Research Section A Accelerators Spectrometers Detectors and Associated Equipment* **788** (2015) .

[77] R. Kleiss et al., `BBBREM`—*Monte Carlo simulation of radiative Bhabha scattering in the very forward direction*, *Computer Physics Communications* **81** (1994) 372.

[78] S. Jadach et al., `BHWIDE` *1.00:$\mathcal{O}(\alpha)$ YFS exponentiated Monte Carlo for Bhabha scattering at wide angles for LEP1/SLC and LEP2*, *Physics Letters B* **390** (1997) 298.

[79] F. Berends et al., *Complete lowest-order calculations for four-lepton final states in electron-positron collisions*, *Nuclear Physics B* **253** (1985) 441.

[80] Y. Suetsugu, K. Shibata, T. Ishibashi, K. Kanazawa, M. Shirai, S. Terui et al., *First commissioning of the SuperKEKB vacuum system*, *Phys. Rev. Accel. Beams* **19** (2016) 121001.

[81] J.B.R. Battat et al., *Readout technologies for directional WIMP Dark Matter detection*, *Phys. Rept.* **662** (2016) 1 [1610.02396].

[82] P.M. Lewis, M.T. Hedges, I. Jaegle, J. Schueler, T.N. Thorpe and S.E. Vahsen, *Primary track recovery in high-definition gas time projection chambers*, *Eur. Phys. J. C* **82** (2022) 324 [2106.15829].

[83] *Pearson's Chi-Square Test Modifications for Comparison of Unweighted and Weighted Histograms and Two Weighted Histograms*, PoS(ACAT)060, Apr., 2007.

[84] D. Scott, *Multivariate density estimation: Theory, practice, and visualization: Second edition* (03, 2015), 10.1002/9781118575574.

[85] A. Paszke et al., *Pytorch: An imperative style, high-performance deep learning library*, in *Advances in Neural Information Processing Systems 32*, H. Wallach et al., eds., pp. 8024–8035, Curran Associates, Inc. (2019), http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[86] A. LeNail, *NN-SVG: Publication-Ready Neural Network Architecture Schematics*, *Journal of Open Source Software* **4** (2019) 747.

[87] S. Ioffe and C. Szegedy, *Batch normalization: Accelerating deep network training by reducing internal covariate shift*, 2015.

[88] G. Klambauer, T. Unterthiner, A. Mayr and S. Hochreiter, *Self-Normalizing Neural Networks*, *arXiv e-prints* (2017) arXiv:1706.02515 [1706.02515].

[89] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, *Dropout: A simple way to prevent neural networks from overfitting*, Journal of Machine Learning Research **15** (2014) 1929.

[90] D.P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2017.

[91] D.E. Rumelhart, G.E. Hinton and R.J. Williams, *Learning Representations by Back-propagating Errors*, *Nature* **323** (1986) 533.

[92] L. Prechelt, *Early stopping — but when?*, in *Neural Networks: Tricks of the Trade: Second Edition*, G. Montavon, G.B. Orr and K.-R. Müller, eds., (Berlin, Heidelberg), pp. 53–67, Springer Berlin Heidelberg (2012), DOI.

[93] C.A.J. O'Hare et al., *Recoil imaging for dark matter, neutrinos, and physics beyond the Standard Model*, in *2022 Snowmass Summer Study*, 3, 2022 [2203.05914].

[94] F. Mayet et al., *A review of the discovery reach of directional Dark Matter detection*, *Phys. Rept.* **627** (2016) 1 [1602.03781].

[95] F. Breiman, L., O. J.H., R.A. and C. Stone, *Classification And Regression Trees*, Routledge, 1st ed. (1984), https://doi.org/10.1201/9781315139470.

[96] J. Billard, F. Mayet and D. Santos, *Low energy electron/recoil discrimination for directional dark matter detection*, *Journal of Cosmology and Astroparticle Physics* **2012** (2012) 020–020.

[97] M. Ghrear, S.E. Vahsen and C. Deaconu, *Observables for recoil identification in high-definition Gas Time Projection Chambers*, *JCAP* **10** (2021) 005 [2012.13649].

[98] P. Baldi, P. Sadowski and D. Whiteson, *Searching for exotic particles in high-energy physics with deep learning*, *Nature Communications* **5** (2014) .

[99] P. Baldi, K. Bauer, C. Eng, P. Sadowski and D. Whiteson, *Jet substructure classification in high-energy physics with deep neural networks*, *Phys. Rev. D* **93** (2016) 094034.

[100] P. Sadowski, B. Radics, Ananya, Y. Yamazaki and P. Baldi, *Efficient antihydrogen detection in antimatter physics by deep learning*, *Journal of Physics Communications* **1** (2017) 025001.

[101] P. Sadowski and P. Baldi, *Deep learning in the natural sciences: applications to physics*, in *Braverman Readings in Machine Learning. Key Ideas from Inception to Current State*, pp. 269–297, Springer (2018).

[102] J.F. Ziegler, M. Ziegler and J. Biersack, *Srim–the stopping and range of ions in matter (2010)*, *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **268** (2010) 1818 .

[103] C.Ş. Deaconu, *A model of the directional sensitivity of low-pressure CF_4 dark matter detectors*, Ph.D. thesis, Massachussetts Institute of Technology, United States, Jan., 2015.

[104] Stephen Biagi, "Degrad - transport of electrons in gas mixtures."

[105] M. Hahsler, M. Piekenbrock and D. Doran, *dbscan: Fast density-based clustering with R*, *Journal of Statistical Software* **91** (2019) 1.

[106] T. Chen and C. Guestrin, *XGBoost: A scalable tree boosting system*, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, (New York, NY, USA), pp. 785–794, ACM, 2016, DOI.

[107] J.R. Quinlan, *Induction of decision trees*, *Machine Learning* **1** (1986) 81.