

# Fermilab

Demonstration and performance of an online data selection algorithm for liquid argon time projection chambers using MicroBooNE

FERMILAB-PUB-25-0954-PPD

arXiv:2602.11138

This manuscript has been authored by Fermi Forward Discovery Group, LLC under Contract No. 89243024CSC000002 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.

# Demonstration and performance of an online data selection algorithm for liquid argon time projection chambers using MicroBooNE

---

## MicroBooNE Collaboration

P. Abratenko<sup>mm</sup> D. Andrade Aldana<sup>n</sup> L. Arellano<sup>v</sup> J. Asadi<sup>ll</sup> A. Ashkenazi<sup>kk</sup>  
 S. Balasubramanian<sup>l</sup> B. Baller<sup>l</sup> A. Barnard<sup>cc</sup> G. Barr<sup>cc</sup> D. Barrow<sup>cc</sup> J. Barrow<sup>z</sup>  
 V. Basque<sup>l</sup> J. Bateman<sup>o,v</sup> B. Behera<sup>hh</sup> O. Benevides Rodrigues<sup>n</sup> S. Berkman<sup>y</sup> A. Bhat<sup>g</sup>  
 M. Bhattacharya<sup>l</sup> V. Bhelande<sup>t</sup> A. Binau<sup>p</sup> M. Bishai<sup>c</sup> A. Blake<sup>s</sup> B. Bogart<sup>x</sup> T. Bolton<sup>r</sup>  
 M. B. Brunetti<sup>q</sup> L. Camilleri<sup>j</sup> D. Caratelli<sup>d</sup> F. Cavanna<sup>l</sup> G. Cerati<sup>l</sup> A. Chappell<sup>oo</sup> Y. Chen<sup>gg</sup>  
 J. M. Conrad<sup>w</sup> M. Convery<sup>gg</sup> L. Cooper-Troendle<sup>dd</sup> J. I. Crespo-Anadón<sup>f</sup> R. Cross<sup>oo</sup>  
 M. Del Tutto<sup>l</sup> S. R. Dennis<sup>e</sup> P. Detje<sup>e</sup> R. Diurba<sup>b</sup> Z. Djurcic<sup>a</sup> K. Duffy<sup>cc</sup> S. Dytman<sup>dd</sup>  
 B. Eberly<sup>ii</sup> P. Englezos<sup>ff</sup> A. Ereditato<sup>g,l</sup> J. J. Evans<sup>v</sup> C. Fang<sup>d</sup> B. T. Fleming<sup>g</sup>  
 W. Foreman<sup>t</sup> D. Franco<sup>g</sup> A. P. Furmanski<sup>z</sup> F. Gao<sup>d</sup> D. Garcia-Gamez<sup>m</sup> S. Gardiner<sup>l</sup> G. Ge<sup>j</sup>  
 S. Gollapinni<sup>t</sup> E. Gramellini<sup>v</sup> P. Green<sup>cc</sup> H. Greenlee<sup>l</sup> L. Gu<sup>s</sup> W. Gu<sup>c</sup> R. Guenette<sup>v</sup>  
 P. Guzowski<sup>v</sup> L. Hagaman<sup>g,j</sup> M. D. Handley<sup>e</sup> M. Harrison<sup>t</sup> S. Hawkins<sup>y</sup> A. Hergenhan<sup>o</sup>  
 O. Hen<sup>w</sup> C. Hilgenberg<sup>z</sup> G. A. Horton-Smith<sup>r</sup> A. Hussain<sup>r</sup> B. Irwin<sup>z</sup> M. S. Ismail<sup>dd</sup>  
 C. James<sup>l</sup> X. Ji<sup>aa</sup> J. H. Jo<sup>c</sup> R. A. Johnson<sup>h</sup> A. Johnson<sup>p</sup> D. Kalra<sup>j</sup> G. Karagiorgi<sup>j</sup>  
 A. Kelly<sup>p</sup> W. Ketchum<sup>l</sup> M. Kirby<sup>c</sup> T. Kobilarcik<sup>l</sup> K. Kumar<sup>j</sup> N. Lane<sup>o,v</sup> J.-Y. Li<sup>k</sup> Y. Li<sup>c</sup>  
 K. Lin<sup>ff</sup> B. R. Littlejohn<sup>n</sup> L. Liu<sup>l</sup> S. Liu<sup>aa</sup> W. C. Louis<sup>t</sup> X. Luo<sup>d</sup> T. Mahmud<sup>s</sup> N. Majeed<sup>r</sup>  
 C. Mariani<sup>nn</sup> J. Marshall<sup>oo</sup> M. G. Manuel Alves<sup>n</sup> N. Martinez<sup>r</sup> D. A. Martinez Caicedo<sup>hh</sup>  
 F. Martinez Lopez<sup>p</sup> S. Martynenko<sup>c</sup> A. Mastbaum<sup>ff</sup> I. Mawby<sup>s</sup> N. McConkey<sup>ee</sup>  
 B. McConnell<sup>p</sup> L. Mellet<sup>y</sup> J. Mendez<sup>u</sup> J. Micallef<sup>w,mm</sup> A. Mogan<sup>i</sup> T. Mohayai<sup>p</sup> M. Mooney<sup>i</sup>  
 A. F. Moor<sup>e</sup> C. D. Moore<sup>l</sup> L. Mora Lepin<sup>v</sup> M. A. Hernandez Morquecho<sup>z</sup> M. M. Moudgalya<sup>v</sup>  
 S. Mulleriababu<sup>b</sup> D. Naples<sup>dd</sup> A. Navrer-Agasson<sup>o</sup> N. Nayak<sup>c</sup> M. Nebot-Guinot<sup>k</sup>  
 C. Nguyen<sup>ff</sup> L. Nguyen<sup>d</sup> J. Nowak<sup>s</sup> N. Oza<sup>j</sup> O. Palamara<sup>l</sup> N. Pallat<sup>z</sup> V. Paolone<sup>dd</sup>  
 A. Papadopoulou<sup>a,t</sup> V. Papavassiliou<sup>bb</sup> H. B. Parkinson<sup>k</sup> S. F. Pate<sup>bb</sup> N. Patel<sup>s</sup>  
 Z. Pavlovic<sup>l</sup> E. Piasetzky<sup>kk</sup> K. Pletcher<sup>y</sup> I. Pophale<sup>s</sup> X. Qian<sup>c</sup> J. L. Raaf<sup>l</sup> V. Radeka<sup>c</sup>  
 A. Rafique<sup>a</sup> M. Reggiani-Guzzo<sup>k</sup> J. Rodriguez Rondon<sup>hh</sup> M. Rosenberg<sup>mm</sup>  
 M. Ross-Lonergan<sup>t</sup> I. Safa<sup>j</sup> D. W. Schmitz<sup>g</sup> A. Schukraft<sup>l</sup> W. Seligman<sup>j</sup> M. H. Shaevitz<sup>j</sup>  
 R. Sharankova<sup>l</sup> J. Shi<sup>e</sup> L. Silva<sup>t</sup> E. L. Snider<sup>l</sup> S. Söldner-Rembold<sup>o</sup> J. Spitz<sup>x</sup> M. Stancari<sup>l</sup>  
 J. St. John<sup>l</sup> T. Strauss<sup>l</sup> A. M. Szelc<sup>k</sup> N. Taniuchi<sup>e</sup> K. Terao<sup>gg</sup> C. Thorpe<sup>v</sup> D. Torbunov<sup>c</sup>  
 D. Totani<sup>d</sup> M. Touns<sup>l</sup> A. Trettin<sup>v</sup> Y.-T. Tsai<sup>gg</sup> J. Tyler<sup>r</sup> M. A. Uchida<sup>e</sup> T. Usher<sup>gg</sup> B. Viren<sup>c</sup>  
 J. Wang<sup>aa</sup> L. Wang<sup>k</sup> M. Weber<sup>b</sup> H. Wei<sup>u</sup> A. J. White<sup>g</sup> S. Wolbers<sup>l</sup> T. Wongjirad<sup>mm</sup>  
 K. Wresilo<sup>e</sup> W. Wu<sup>dd</sup> E. Yandel<sup>t</sup> T. Yang<sup>l</sup> L. E. Yates<sup>l,rr</sup> H. W. Yu<sup>c</sup> G. P. Zeller<sup>l</sup>  
 J. Zennamo<sup>l</sup> C. Zhang<sup>c</sup> Y. Zhang<sup>c</sup>

<sup>a</sup>Argonne National Laboratory (ANL), Lemont, IL, 60439, USA  
<sup>b</sup>Universität Bern, Bern CH-3012, Switzerland  
<sup>c</sup>Brookhaven National Laboratory (BNL), Upton, NY, 11973, USA  
<sup>d</sup>University of California, Santa Barbara, CA, 93106, USA  
<sup>e</sup>University of Cambridge, Cambridge CB3 0HE, United Kingdom  
<sup>f</sup>Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid E-28040, Spain  
<sup>g</sup>University of Chicago, Chicago, IL, 60637, USA  
<sup>h</sup>University of Cincinnati, Cincinnati, OH, 45221, USA  
<sup>i</sup>Colorado State University, Fort Collins, CO, 80523, USA  
<sup>j</sup>Columbia University, New York, NY, 10027, USA  
<sup>k</sup>University of Edinburgh, Edinburgh EH9 3FD, United Kingdom  
<sup>l</sup>Fermi National Accelerator Laboratory (FNAL), Batavia, IL 60510, USA  
<sup>m</sup>Universidad de Granada, E-18071, Granada, Spain  
<sup>n</sup>Illinois Institute of Technology (IIT), Chicago, IL 60616, USA  
<sup>o</sup>Imperial College London, London SW7 2AZ, United Kingdom  
<sup>p</sup>Indiana University, Bloomington, IN 47405, USA  
<sup>q</sup>The University of Kansas, Lawrence, KS, 66045, USA  
<sup>r</sup>Kansas State University (KSU), Manhattan, KS, 66506, USA  
<sup>s</sup>Lancaster University, Lancaster LA1 4YW, United Kingdom  
<sup>t</sup>Los Alamos National Laboratory (LANL), Los Alamos, NM, 87545, USA  
<sup>u</sup>Louisiana State University, Baton Rouge, LA, 70803, USA  
<sup>v</sup>The University of Manchester, Manchester M13 9PL, United Kingdom  
<sup>w</sup>Massachusetts Institute of Technology (MIT), Cambridge, MA, 02139, USA  
<sup>x</sup>University of Michigan, Ann Arbor, MI, 48109, USA  
<sup>y</sup>Michigan State University, East Lansing, MI 48824, USA  
<sup>z</sup>University of Minnesota, Minneapolis, MN, 55455, USA  
<sup>aa</sup>Nankai University, Nankai District, Tianjin 300071, China  
<sup>bb</sup>New Mexico State University (NMSU), Las Cruces, NM, 88003, USA  
<sup>rr</sup>University of Notre Dame, Notre Dame, IN 46556, USA  
<sup>cc</sup>University of Oxford, Oxford OX1 3RH, United Kingdom  
<sup>dd</sup>University of Pittsburgh, Pittsburgh, PA, 15260, USA  
<sup>ee</sup>Queen Mary University of London, London E1 4NS, United Kingdom  
<sup>ff</sup>Rutgers University, Piscataway, NJ, 08854, USA  
<sup>gg</sup>SLAC National Accelerator Laboratory, Menlo Park, CA, 94025, USA  
<sup>hh</sup>South Dakota School of Mines and Technology (SDSMT), Rapid City, SD, 57701, USA  
<sup>ii</sup>University of Southern Maine, Portland, ME, 04104, USA  
<sup>kk</sup>Tel Aviv University, Tel Aviv, Israel, 69978  
<sup>ll</sup>University of Texas, Arlington, TX, 76019, USA  
<sup>mm</sup>Tufts University, Medford, MA, 02155, USA  
<sup>nn</sup>Center for Neutrino Physics, Virginia Tech, Blacksburg, VA, 24061, USA  
<sup>oo</sup>University of Warwick, Coventry CV4 7AL, United Kingdom

E-mail: [microboone\\_info@fnal.gov](mailto:microboone_info@fnal.gov)

**ABSTRACT:** The MicroBooNE detector is a liquid argon time projection chamber (LArTPC) that produces three-dimensional images of particle interactions using ionization charge collected by anode wire plane arrays and scintillation light collected by a light detection system. In addition to testing long-standing experimental neutrino anomalies and performing measurements of neutrino interactions with argon nuclei using the Fermilab Booster Neutrino Beam, MicroBooNE aims to develop methodologies for rare beyond the Standard Model and off-beam physics searches. Looking ahead to the upcoming Deep Underground Neutrino Experiment (DUNE), with MicroBooNE serving as a valuable testbed, achieving high sensitivity and livetime for off-beam physics while satisfying data processing and storage constraints will require data-driven, intelligent, and online or real-time data selection techniques. These techniques are essential for reducing data rates and preserving rare signals with high accuracy. In this paper, we describe a fast data selection algorithm suitable for online execution to identify electrons from stopping cosmic ray muons in the MicroBooNE detector utilizing ionization charge information, and present its performance. This represents the first demonstration of online data selection in a LArTPC using real data and charge information exclusively and provides an important proof-of-principle for applying such techniques to other LArTPC experiments such as the Short-Baseline Near Detector and DUNE.

**KEYWORDS:** Data selection; Image processing; Noble liquid detectors (scintillation, ionization, double-phase); Time Projection Chambers (TPC)

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The MicroBooNE Detector</b>	<b>2</b>
2.1	MicroBooNE TPC Readout System	3
2.2	“Emulated Online” Data Selection System	5
<b>3</b>	<b>Data and Simulation Samples for Supernova Data Stream</b>	<b>6</b>
3.1	Data Sample	6
3.2	Simulation Samples	8
3.3	Data to Simulation Comparison	10
<b>4</b>	<b>Workflow Overview</b>	<b>11</b>
4.1	Development and Validation of Data Selection Algorithm	11
4.2	Demonstration Using Real Data	15
4.2.1	Process 1: Trigger Primitive Generation	15
4.2.2	Process 2: Trigger Candidate Generation	16
4.2.3	Process 3: High-Level Trigger	18
<b>5</b>	<b>Future Applications and AI/ML Developments</b>	<b>20</b>
<b>6</b>	<b>Summary and Conclusions</b>	<b>21</b>

---

## 1 Introduction

The MicroBooNE experiment [1] employed a liquid argon time projection chamber (LArTPC) detector 470 m downstream in the Fermilab Booster Neutrino Beam (BNB), situated on-surface and operated without significant overburden shielding. The experiment collected beam data between 2015 and 2020 via two data streams—an externally-triggered neutrino (NU) data stream and a triggerless, continuously running supernova (SN) data stream. The SN data stream was commissioned in November 2017 [2]. The NU data stream was used to study beam-based physics including beyond-Standard Model (BSM) physics and neutrino interactions on argon using beam neutrinos. The SN data stream can be used to study a rich program of off-beam physics processes including rare ones, such as neutrinos from a nearby supernova burst (SNB), baryon number violating (BNV) processes such as neutron-antineutron ( $n \rightarrow \bar{n}$ ) transitions [3] and proton decay.

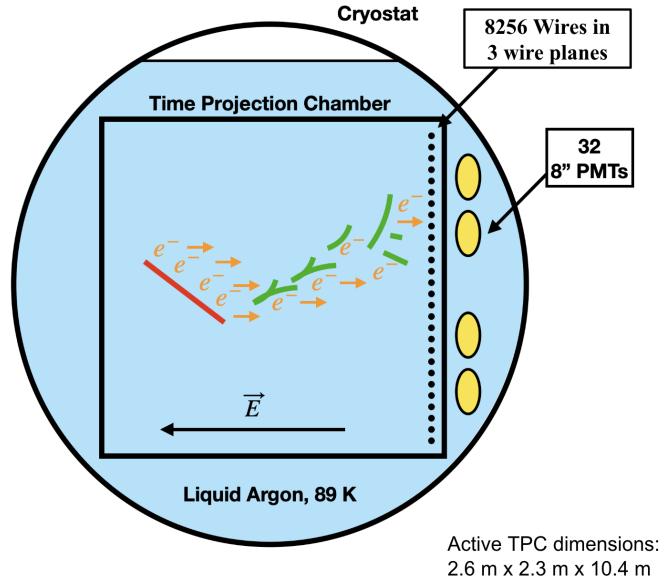
Rare off-beam physics processes occur randomly in time, and with no external timing information available to help promptly identify them in the data, making searches for these processes a challenging task. To perform this task successfully, data must be processed continuously with nearly 100% livetime, which requires implementing fast data selection techniques “online” (in software

when the experiment is running) or in “real-time” (in hardware when the experiment is running), capable of handling the detector-generated data rate. During MicroBooNE running, the total data rate generated by the detector’s charge readout system alone was approximately 33 GB/s [2]. Generated data rates will become even more challenging to process for future large LArTPC detectors such as the planned Deep Underground Neutrino Experiment (DUNE) [4–6], where raw generated data rates for the Far Detector, for example, are expected to reach several TB/s. The study presented in this paper uses data from the MicroBooNE detector to demonstrate and assess the performance of an online data processing system targeting the selection of a specific physics topology—that of decaying cosmic ray muons—based solely on TPC (ionization charge information) data. This work represents the first demonstration of fast data selection techniques using real TPC data from a LArTPC detector, and the first to use topological information available in the TPC data.

## 2 The MicroBooNE Detector

The MicroBooNE LArTPC detector [1] employs an active mass of 85 metric tons of liquid argon. The detector is a 10.4 m long, 2.6 m wide, 2.3 m high LArTPC, and is located on-surface and on-axis to the BNB. Due to its on-surface location, the MicroBooNE detector is exposed to a large flux ( $\sim 4\text{kHz}$  [7]) of cosmic rays, leading to a variety of cosmogenic activity within the detector. Each charged particle passing through or produced from any interaction within the LAr volume leaves behind a trail of ionization electrons that drift toward three anode wire planes under the effect of a uniform electric field with a maximum drift time of 2.3 ms. There are two induction wire planes ( $U$  and  $V$ ) with 2400 wires each, and one collection wire plane ( $Y$ ) with 3456 wires; all planes maintain a 3 mm wire pitch to detect the ionization electrons and produce three two-dimensional image projections of the interaction volume. Additionally, particle interactions within the LAr produce prompt scintillation light which is detected by 32 photomultiplier tubes (PMTs) located behind the anode wire planes as shown in figure 1. Light information complements the TPC charge information to enable 3D particle reconstruction. The combination of light-based triggering and ionization charge readout in the beam spill window forms the foundation of the NU data stream.

The SN data stream runs in parallel to and independently of the NU data stream. As such, it provides a useful platform for studying off-beam signals including rare and/or low-energy signals. The SN data stream can also be used for understanding the impact of data reduction and data selection schemes on associated low-energy [2] and other rare and exotic physics signals. Although the MicroBooNE detector, due to its surface location and size, is not highly sensitive to rare processes such as SNB neutrinos or BNV, it serves as an important R&D facility for developing and validating advanced online TPC-based triggering and data selection techniques. In this context, the presented study uses pre-collected data from the SN data stream (and, for simplicity, data from only the collection plane) to develop a fast data selection algorithm to identify Michel electrons from stopping cosmic ray muon decays. Online TPC data selection employing this algorithm is subsequently emulated and demonstrated to keep up with data throughput while meeting offline physics performance. Such developments are important for future large-scale LArTPC experiments like DUNE, whose massive detector will enable sensitive searches for SNBs [6], proton decay, and other BNV processes [5].



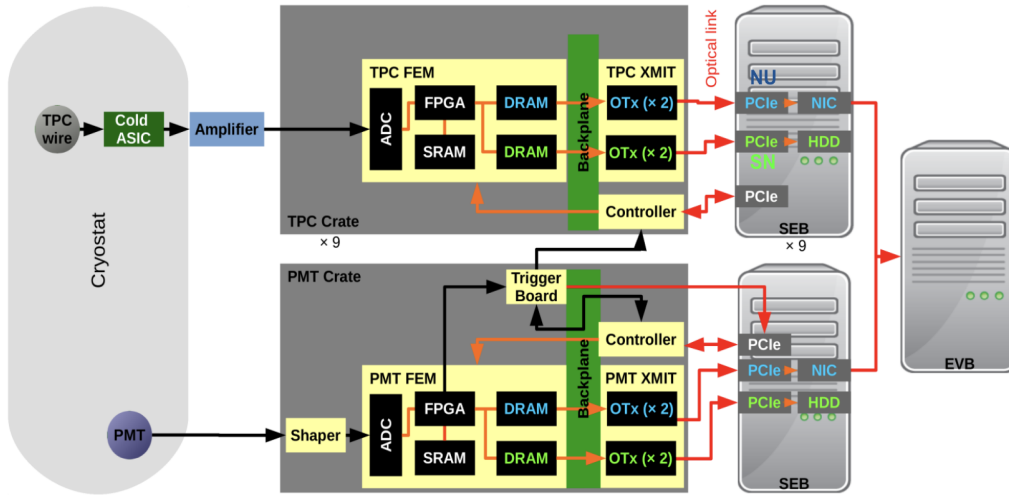
**Figure 1:** A schematic illustrating the MicroBooNE liquid argon time projection chamber. A neutrino interaction in the detector medium produces ionization electrons that drift toward the anode wire planes under a uniform electric field.

## 2.1 MicroBooNE TPC Readout System

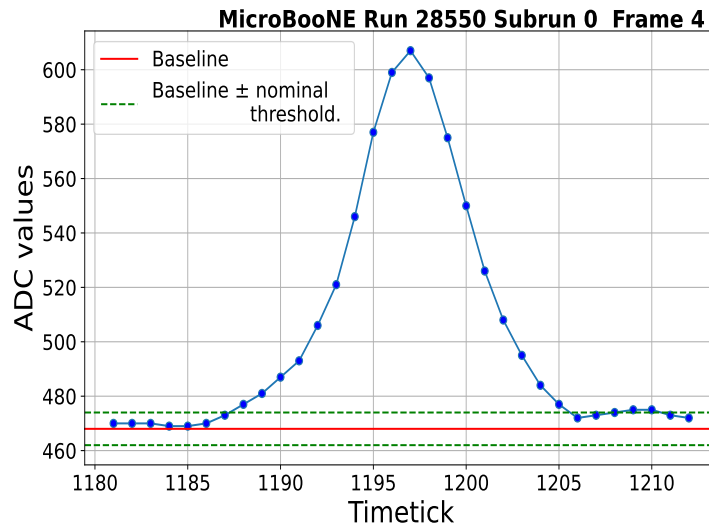
The MicroBooNE TPC readout electronics system is responsible for processing digitized waveforms from the anode wires, and sending this processed data to the downstream software-based data acquisition (DAQ) system; the schematic data flow is visualized in figure 2 [2]. In this section, only the TPC readout data flow (top dark gray block of figure 2) is discussed.

Continuous data (waveforms) from 8,256 TPC wires are distributed among nine TPC readout electronics crates [1]. Each crate houses 14-16 analog to digital converter (ADC) + front-end module (FEM) processing boards, where each board receives and handles data from a group of 64 wires. The ADC+FEM board is responsible for digitizing the data using a 16 MHz, 12-bit ADC. An Altera Stratix III field programmable gate array (FPGA) in each board then down-samples the data to 2 MHz and writes it in channel order to static RAM (SRAM). The data is then read back, in time order over the length of a predefined time “frame”, by the same FPGA. Furthermore, the FPGA duplicates the data into two readout streams (NU and SN) and, after compression, temporarily stores it in multi-buffer dynamic RAM (DRAM).

For the NU data stream, upon receiving an external trigger, 4.8 ms of data around the trigger time (1.6 ms before the trigger time and 3.2 ms after the trigger time) are read out from each ADC+FEM’s NU DRAM (shown in blue color in figure 2). The data are then sent via a transmitter (XMIT) module, which is equipped with four optical transceivers (OTxs), with two dedicated to the NU stream and two to the SN stream. The NU data are transmitted to a dedicated PCI Express (PCIe) card housed in a DAQ computer server, where they are handled by the DAQ system as described below. This NU data stream contains losslessly (Huffman) compressed data, and is used



**Figure 2:** A schematic illustrating MicroBooNE readout data flow. The blue-labeled path is part of the NU stream. The green-labeled path is part of the SN stream. See text for more details.



**Figure 3:** Example of a MicroBooNE ROI in the continuous SN data stream from one of the collection plane channels, showing samples saved as part of the ROI waveform (blue circles), once the threshold criterion is met. Baseline and threshold ADC values for this channel are shown in red and dashed green lines, respectively.

by all physics analyses targeting beam-related measurements.

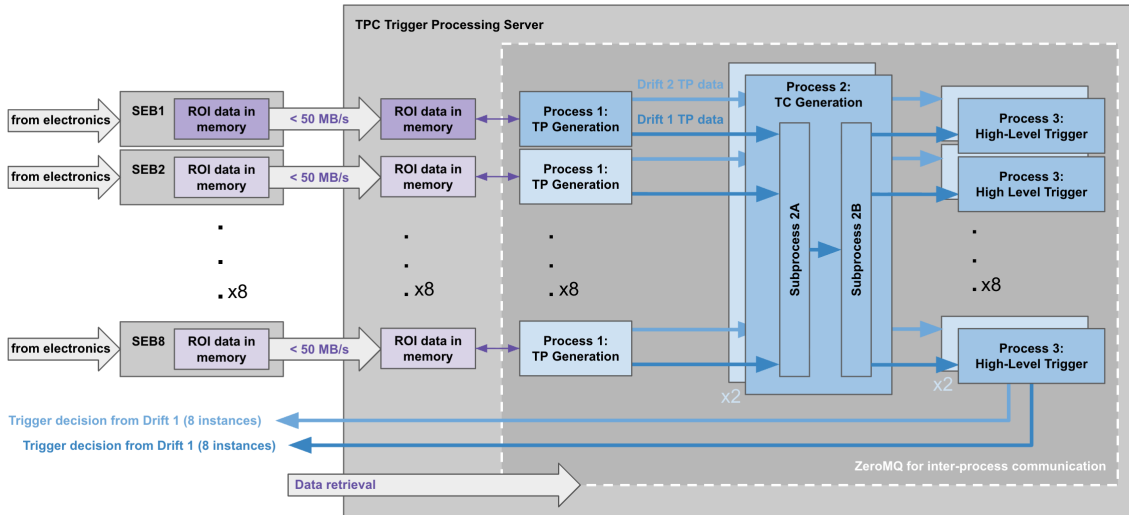
In parallel, SN data from each crate are streamed out continuously via the other two XMIT optical transceivers to a second dedicated PCI Express card in the same server. Unlike in the NU stream case, the continuous readout firmware implementation applies two data reduction algorithms sequentially [2]. The first one is a zero-suppression (lossy data reduction) scheme which checks whether an ADC sample passes a threshold after subtracting the channel baseline. The sign of the threshold can be chosen to be positive (passing samples are greater than the sum of the baseline and threshold), negative (passing samples are smaller than the baseline minus the threshold), or either. In addition, a configurable number of samples preceding the first one to pass the threshold (presamples), and following the last sample that passed the threshold (postsamples), are retained in order to better capture the waveform. The set of samples that pass the threshold, plus the presamples (set to 7) and postsamples (set to 7), is considered a region of interest (ROI). An example of a SN stream ROI is shown in figure 3. The second data reduction algorithm, applied to each ROI, is a lossless Huffman compression algorithm, in which consecutive ADC samples that differ by fewer than four ADC counts are encoded using a reduced number of bits. This lossy-compressed data stream benefits from a data reduction factor of approximately 80 times the total generated raw data stream of 33 GB/s, and is used to study off-beam events [2].

Both of these data streams are read out across nine Sub-Event Buffer (SEB) servers. The NU data from the nine SEB machines is transferred to the Event Builder (EVB) machine using a network interface card (NIC) where the events are built online and written to local disk on the EVB before being sent to permanent tape storage where they can be read offline for further processing. The SN data is stored in temporary buffers in each SEB’s hard-disk drive (HDD) for 48 hours at a time, and sent to permanent tape storage only upon receiving an external alert from the SuperNova Early Warning System (SNEWS) [2, 8].

## 2.2 “Emulated Online” Data Selection System

To enable CPU-based data selection on the SN stream, an “emulated online” system was developed, which read, buffered, and re-streamed pre-recorded SN stream data from the SEB disks at a rate comparable to the nominal SN streaming rate from the MicroBooNE readout electronics. The pre-recorded data was collected in 2021, after the MicroBooNE experiment completed beam operations. Then, a software-based, TPC-based data selection framework, comprising three main processes as illustrated in figure 4, was developed and used to process the SN data stream to identify Michel electrons from cosmic ray muon decays. The three main processes run sequentially and include: Process 1: Trigger Primitive (TP) Generation, Process 2: Trigger Candidate (TC) Generation, and Process 3: High-Level Trigger. This multi-stage data selection framework design aligns with DUNE’s proposed data selection strategy [9, 10], and therefore represents a critical test for TPC data-driven online data selection for DUNE.

The data selection framework begins with generating Trigger Primitives (TPs), which represent summaries of per-channel ROI data. Each TP object contains quantities such as integrated ROI charge, maximum amplitude of the ROI waveform, and width of the ROI waveform (i.e., the number of samples between the first and last sample). In true online implementation scenarios, such as those planned for SBND [11, 12], ProtoDUNE [13] and DUNE [9, 10], TPs will be produced directly by



**Figure 4:** A schematic showing the online TPC-based data selection framework developed and exercised on the MicroBooNE “emulated online” SN data stream. The online emulation refers to the ROI data input to Process 1 being read directly from memory, once already transferred over the network.

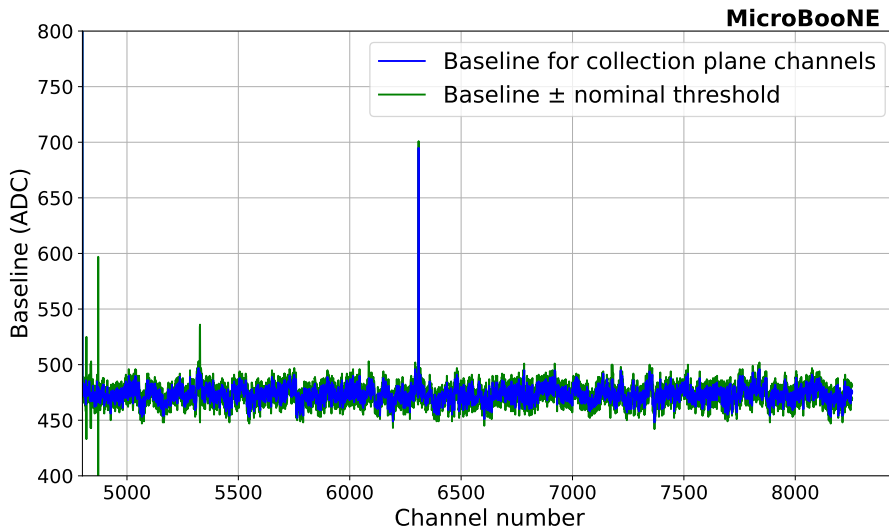
the electronics or DAQ system as an independent real-time data stream. The TPs are subsequently combined to form higher-level triggered objects referred to as Trigger Candidates (TCs), which are then further processed to generate a high-level trigger (HLT), enabling selective permanent storage of SN ROI data.

This framework was developed and demonstrated using only information from the collection plane channels in the MicroBooNE detector, which are distributed across eight TPC readout crates. In TPC crates 02–08, which each contain 15 or 16 ADC+FEM boards, collection-plane wire channels 4896–8255 were read across channels 32–63 of each ADC+FEM board. Crate 09 contains only 14 ADC+FEM boards, and in this crate only channels 32–63 of the last three ADC+FEM boards were used to read out collection-plane wire channels 4800–4895. The details of this framework and its performance are described in later sections.

### 3 Data and Simulation Samples for Supernova Data Stream

#### 3.1 Data Sample

The framework was developed and demonstrated using pre-collected MicroBooNE SN data obtained with a “(Nominal + 15) ADC” ROI threshold configuration. For this configuration, the nominal threshold values used in the firmware zero-suppression algorithms for ROI extraction, shown in figure 5, were all increased by a flat value of 15 ADC counts so as to further reduce electronics noise in the data. Since most channels (except the noisiest ones) are nominally configured with threshold values of  $\sim 4 - 5$  ADC, this corresponds to a net channel threshold of  $\sim 20$  ADC. This data sample then underwent pre-processing to remove several known and undesirable features in ROIs in real data, as described below.

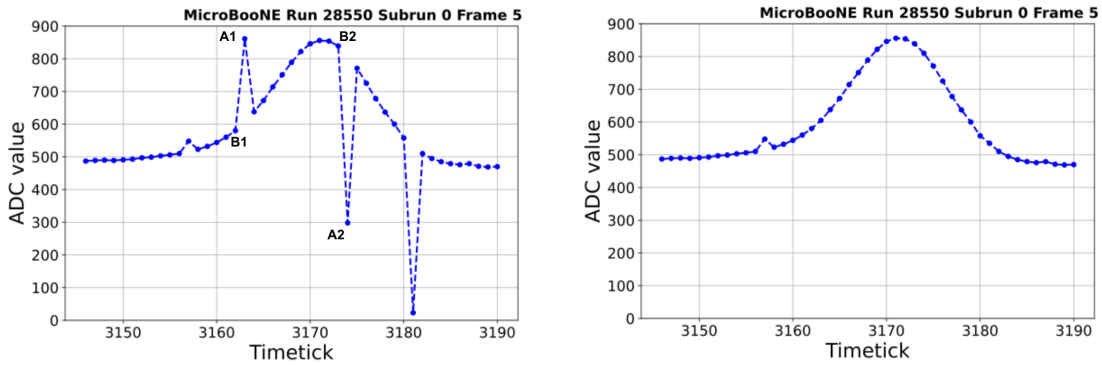


**Figure 5:** Baseline and nominal threshold values for collection plane channels. Some of the noisier channels have higher baseline settings.

- **Bit-flip correction:** In real data collected with the SN stream, the binary words containing ADC values exhibit “flipped bits”, where random bits in the 12-bit ADC word switch from 0 to 1, or vice-versa, relative to their true value [2]. This causes the recorded ADC values to shift by a sum of terms of the form  $2^n$ , where  $n$  is an integer, distorting the corresponding waveform. An example of an ROI exhibiting flipped bits is shown in figure 6 (left). Such bit-flip artifacts are observed in most channels and persist throughout the duration of all MicroBooNE SN runs, including the one from which the SN data sample was collected. The origin of this effect is unknown, but only affects the continuous SN data stream. The algorithm developed and applied to correct for these bit-flips is described next.

As shown in figure 6 (left), the ADC values at points A1 and A2 of the ROI waveform correspond to bit-flips occurring in opposite directions. To correct positive bit-flips as the one at A1, the ADC value at A1 ( $ADC_i$ ) is compared with the immediately preceding ADC value (at B1). If the difference between  $ADC_i - ADC_{i-1}$  is greater than 50 ADC counts, the nearest power of two to the difference is subtracted from the ADC value at A1. Similarly, to correct for a bit-flip in the opposite direction (as the one in A2), the ADC value at A2 is compared with the previous value (B2). If the difference  $ADC_i - ADC_{i-1}$  is less than  $-50$ , the nearest power of two to the difference is added to the ADC value at A2. This check is performed on all successive ADC values within an ROI. After this first iteration, the difference between the current and previous ADC values is checked again. If the difference between  $ADC_i - ADC_{i-1}$  is greater than 16 ADC counts or less than  $-16$  ADC counts, the same procedure described above is applied to further correct the ADC value. After applying these corrections, the resulting waveform is shown in figure 6 (right).

The threshold value of 50 was chosen empirically to balance sensitivity to true bit-flips against noise fluctuations; smaller thresholds tend to overcorrect noise, while higher thresholds



**Figure 6:** An example of a SN stream ROI exhibiting several flipped bits before (left) and after (right) correcting for this effect using the algorithm described in Section 3.1.

can miss smaller-amplitude flips. While the algorithm relies only on the preceding ADC value, incorporating both neighboring samples (before and after) could, in principle, improve robustness. However, the current approach was found sufficient for the SN data, as the remaining uncorrected cases, such as the small bit-flip near timetick  $\approx 3157$  in figure 6, have negligible impact on subsequent analysis.

- **Baseline subtraction:** The next step involved subtracting the baseline from the ROI waveforms. During MicroBooNE SN commissioning, the baseline was measured independently for each TPC channel and used as a configuration parameter in the firmware zero-suppression algorithm used for subsequent SN runs. These configured baseline values, shown in figure 5 as a function of collection channel number, remained stable over the course of MicroBooNE operations. As such, these same values were used to extract baseline-subtracted ROIs, before extracting TP information. Baseline-subtraction is desirable in order to improve signal to background discrimination using TP information such as maximum ADC amplitude and integrated charge.
- **Short waveform removal:** The SN stream data format occasionally exhibits a known issue where the end-of-waveform marker is missing, typically due to buffer overflows. This results in the presence of short, incomplete waveforms that do not contain valid data. To mitigate this, waveforms with fewer than 20 ADC words were excluded from the study. The threshold of 20 ADC words was selected based on the requirement that a valid waveform must include at least 14 samples (7 presamples and 7 postsamples) plus an additional minimum of 6 samples, consistent with the TPC signal  $\mu\text{s}$ -scale front-end electronics shaping time.

### 3.2 Simulation Samples

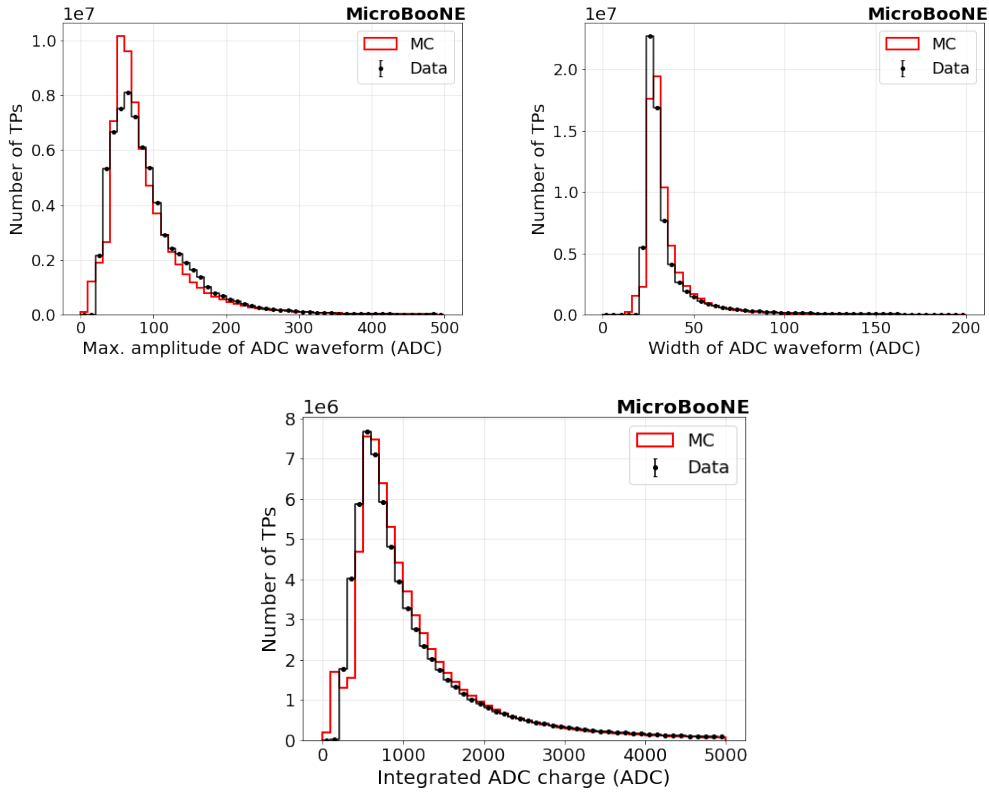
To optimize and validate the data selection algorithm, three distinct Monte Carlo (MC) simulated samples were produced. These samples represent typical physics scenarios expected in the detector and allow the study of the algorithm's performance under controlled conditions. Each sample underwent a generation stage, followed by detector simulation, zero suppression, and TP generation.

**Table 1:** Event fraction for muon capture and muon decay processes in the CORSIKA sample along with statistical uncertainties.

CORSIKA process	Fraction (%)
$\mu^-$ capture at rest	$8.5 \pm 1.1$
$\mu^-$ decay	$1.7 \pm 0.5$
$\mu^+$ decay	$10.4 \pm 1.2$

The simulated samples use the same thresholds as in the dedicated SN data sample. Effects such as electronics noise and unresponsive wires or channels were also simulated.

- Stopping muons:** A sample of muons with an average energy of 2 GeV (ranging between 1.5-2.5 GeV) was generated, with the muons required to stop inside the active TPC volume. The stopping muons were simulated with varying angles between the beam direction and the downward vertical direction. Stopping muons are a useful calibration and validation channel, as they deposit a well-defined amount of energy per unit length until they come to rest, producing a Bragg peak near the end of their track. This topology provides a clean test case for evaluating the algorithm’s efficiency for selecting Michel electrons, and for optimizing settings in the algorithm. Any Michel-like triggers generated from such a sample are regarded as true-positives.
- Crossing muons:** A sample of TPC crossing (through-going) muons was simulated with varying angles between the beam direction and the downward vertical direction. Crossing muons constitute the dominant cosmic background in on-surface operated detectors, creating a realistic and challenging trigger environment. This sample is used to evaluate the false-positive (background) trigger rate. Any Michel-like triggers generated from such a sample are regarded as false-positives.
- CORSIKA cosmic muons:** A sample of cosmic-ray muons was generated using the CORSIKA framework [14], which simulates the full air-shower development from primary cosmic rays in the atmosphere, and subsequently run through the MicroBooNE detector simulation. This sample provides a realistic distribution of muon energies and angular directions, matching the expected flux at the MicroBooNE detector site. It was used both to validate the data selection algorithm against more realistic conditions and to compare the performance of the algorithm when applied to data versus simulation. A small subset of the CORSIKA simulation sample was also analyzed to determine the fraction of true stopping muons that produce Michel electrons. The muon capture and muon decay processes were examined and are summarized in table 1, where stopping-muon decay processes correspond to the second and third rows. Approximately 12% of cosmic-ray muons were found to decay within the detector active volume, 8% underwent nuclear capture while the remaining 80% exited the active volume.



**Figure 7:** Data (black) and MC simulation (red) comparisons of trigger primitive distributions: maximum amplitude of ADC waveform (top left), width of ADC waveform (top right), and integrated ADC charge (bottom). Statistical uncertainties are shown for data.

### 3.3 Data to Simulation Comparison

Trigger primitive distributions from the SN data sample (87,936 drift regions, corresponding to 202 seconds of exposure, with each drift region spanning 2.3 ms) were compared with those from the CORSIKA cosmic muon MC sample (25,000 simulated readout windows, corresponding to an exposure of 57 seconds), as shown in figure 7. This was done as a validation check prior to running the data selection algorithm. The data distribution is normalized to the simulation by scaling the total number of trigger primitives in data so that it matches the total number of trigger primitives in the simulated sample. The broader spread in the maximum amplitude of the ROI waveform observed in data is likely due to uncorrected flipped bits introducing some smearing. In addition, the data distributions for the integrated ADC charge and waveform width appear shifted, which may be attributed to inefficiency in the short waveform removal cut described in section 3.1, since short waveforms only appear in data and are not simulated in the MC. Overall, there is good shape agreement between data and MC.

## 4 Workflow Overview

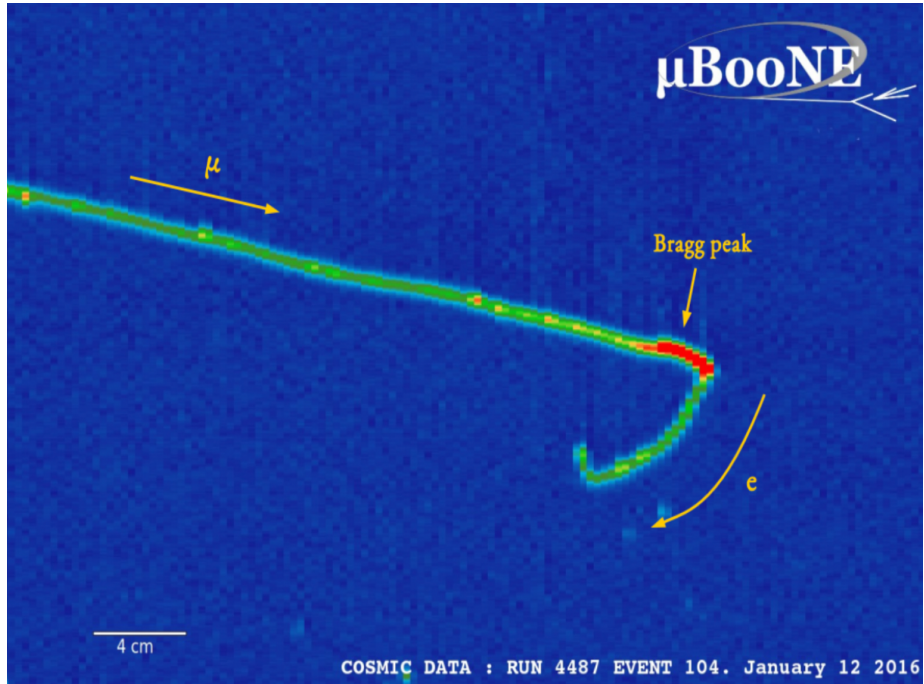
The goal of this development and demonstration effort is a fast TPC-based data selection algorithm that could be deployed in an online environment for other LArTPC experiments such as SBND and/or DUNE in the near future, with sensitivity to an exclusive physics signal topology. The physics signal topology being targeted by the algorithm is Michel electrons produced from stopping cosmic-ray muons. Michel electrons are particularly valuable to study because their energy spectrum closely resembles that of electrons generated in charged-current or elastic scattering interactions of SNB neutrinos, making them an effective proxy for one of the key low-energy physics targets of current and next-generation LArTPC neutrino experiments. Furthermore, Michel electrons provide an excellent low-energy calibration source ( $\lesssim 50$  MeV), reinforcing the motivation to study this topology through targeted high-statistics samples that can be collected in off-beam data.

The effort proceeds by first developing and validating the physics performance of the data selection algorithm using simulation samples, as described in subsection 4.1, and then implementing and applying the algorithm to the emulated stream of pre-recorded SN data, as part of an online workflow described in subsection 4.2.

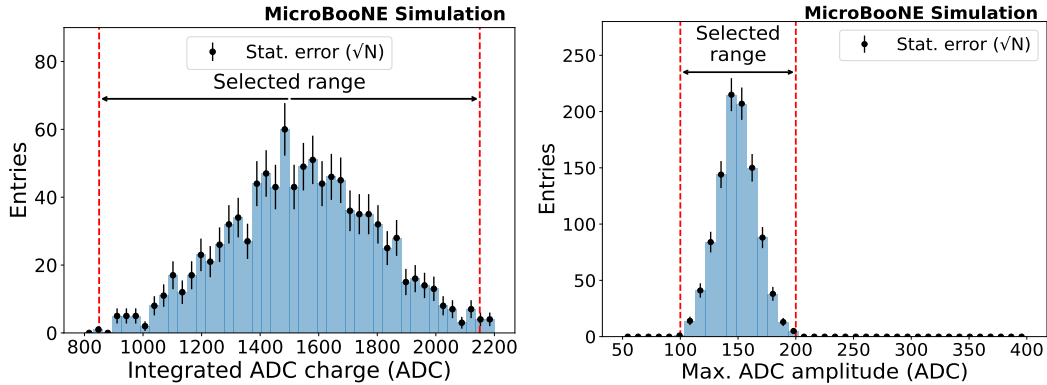
### 4.1 Development and Validation of Data Selection Algorithm

When a muon slows inside the detector medium, it deposits its remaining energy via ionization, producing a characteristic Bragg peak, before coming to rest and decaying into a Michel electron (as illustrated in figure 8). This decay typically introduces a noticeable kink between the muon and electron track directions. The data selection algorithm exploits two key features of this topology: (1) the localized increase in energy deposition near the stopping point (Bragg peak) and (2) the change in ionization track linearity at the decay kink. The algorithm has been developed and tested using the simulation samples described in subsection 3.2, and validated with simulation and data comparisons.

1. **Step 1: Aggregating TPs over a drift region:** Since this study relies only on collection-plane wire waveforms, the first step involves gathering and re-ordering TPs from the 3456 collection-plane channels over a full 2.3 ms drift window (drift region). Because the online TPC-based data selection operates on TPs organized per drift region, the algorithm is correspondingly developed and validated using the same drift region structure.
2. **Step 2: Identifying potential Bragg peak locations within a drift region:** In this step, the algorithm identifies candidate Bragg peaks—points of maximum energy deposition—as seeds for locating the characteristic kinked topology of the decay. First, the drift region is subdivided into “slices,” each covering 36 consecutive (in physical space) channels and 575 consecutive time-ticks. Slicing reduces the amount of information that must be processed to find the Bragg peak associated with stopping muons and enables further parallelization, while also ensuring reliable Bragg peak identification, particularly as more than one Bragg peak candidate can be identified in a given drift region. Within each slice, the TP with the highest integrated ADC charge and peak amplitude is selected. A TP is classified as a Bragg peak candidate if its integrated ADC charge lies within 850–2150 ADC counts and if its maximum



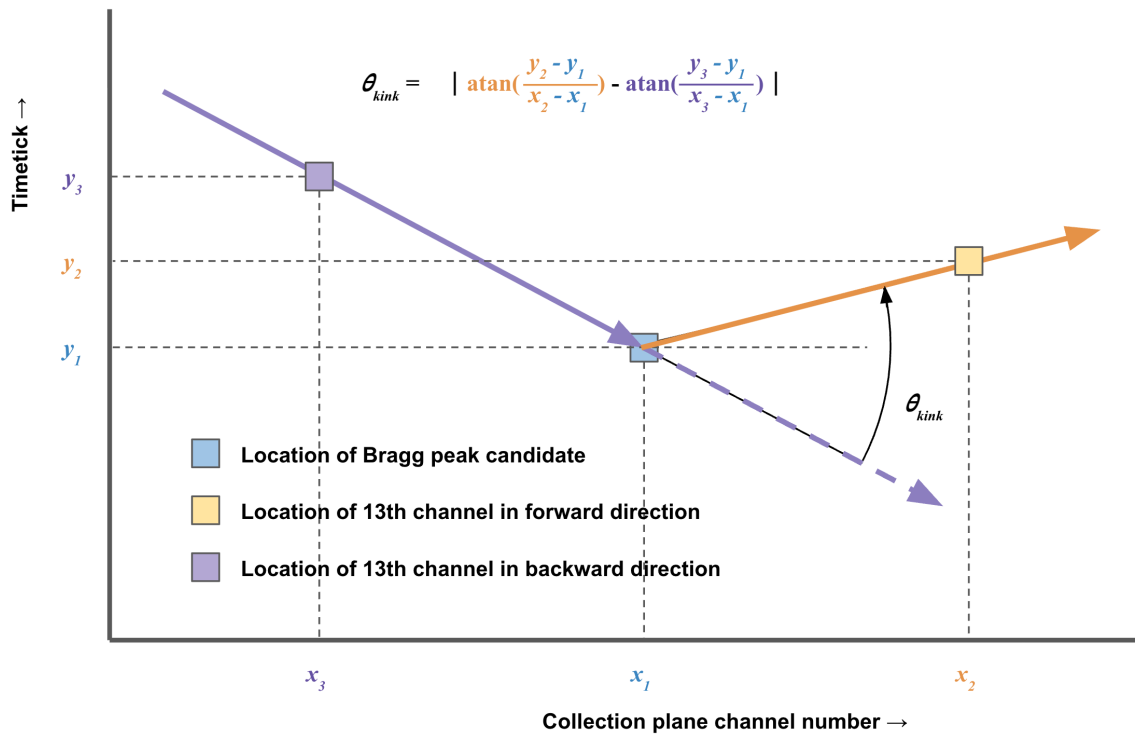
**Figure 8:** A MicroBooNE event display zoomed in to show a stopping and decaying muon topology. The  $x$ -axis represents the collection plane channels and the  $y$ -axis represents the time-tick space. The pitch between the wires is 3 mm. Progressively yellow-to-red colors represent an increasing amount of deposited ionization charge.



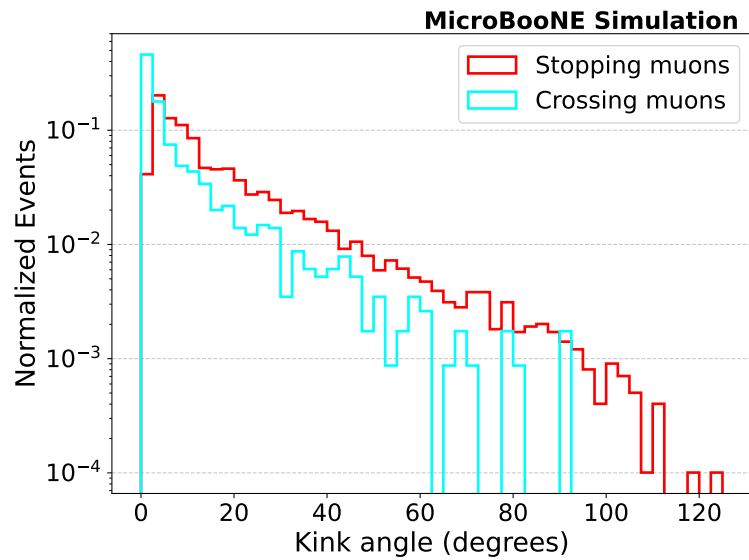
**Figure 9:** Distribution of integrated ADC charge (left) and maximum amplitude (right) at the Bragg peak for simulated 2.3 ms drift time periods with true stopping muons decaying to Michel electrons. The red lines indicate the selection requirements.

amplitude falls within 100 – 200 ADC counts. These values were optimized by manually scanning 300 event displays from the stopping muon MC sample; relevant distributions are shown in figure 9.

- Step 3: Identifying tracks and kink angle:** In this step, the algorithm searches for track-like patterns originating from each of the potential Bragg peak candidates in both the forward and



**Figure 10:** A schematic illustrating the method used to calculate the kink angle. The kink angle,  $\theta_{kink}$ , is the angle between the track segments on either side of the Bragg peak. This is defined as an angle in the space of timetick against collection-plane channel number, rather than as a physical angle.



**Figure 11:** Distribution of kink angle for stopping muons (red) and crossing muons (cyan) with a Bragg peak candidate.

**Table 2:** Number of drift regions selected at Steps 2 and 3, and for different kink angle requirements, shown separately for the simulated stopping muon, crossing muon, and CORSIKA muon samples. The selection efficiency (in parentheses) is defined as the ratio of drift regions remaining after each step to the total number of drift regions in the given sample. The kink angle is defined as in figure 10.

	<b>Stopping muons</b>	<b>Crossing muons</b>	<b>CORSIKA muons</b>
Total drift regions	10000	2000	25000
After Bragg peak	9939 (99.4%)	1149 (57.4%)	24410 (97.6%)
After kink angle > 10°	5757 (57.6%)	316 (15.8%)	14056 (56.2%)
After kink angle > 20°	3300 (33.0%)	162 (8.1%)	7741 (30.9%)
After kink angle > 30°	2036 (20.4%)	91 (4.5%)	4964 (19.8%)
After kink angle > 40°	1265 (12.6%)	63 (3.1%)	3440 (13.7%)

backward channel directions, and requires TPs in 13 contiguous channels in each direction. During the TP search, the algorithm accounts for detector dead regions by skipping over dead channels. Each forward or backward track segment is required to contain only one TP per channel with start times within  $\pm 5$  time-ticks of the candidate Bragg peak TP start time. This helps suppress false positives from delta rays, which are often identified as Bragg peak candidates. The algorithm then calculates the slopes and angles of both the forward and backward tracks (in channel space) as shown in figure 10. Their relative angle (defined as the kink angle in figure 10) is computed.

The kink angle distributions for true stopping muons and true crossing muons are shown in figure 11. The distributions show, as expected, that the majority of crossing muons generally lead to smaller (false) kink angles. The distribution of (false) kink angles at  $> 10^\circ$  for true crossing muons could be due to the presence of delta rays, where the algorithm mistakenly identifies the delta ray energy deposition as the Bragg peak.

Table 2 summarizes the number of drift regions selected at Steps 2 and 3, and for different kink angle requirements, for all three MC samples: stopping muons (true positive), crossing muons (false positive) and CORSIKA muons. The simulated selection efficiency for each case, quoted in parentheses, is defined as the ratio of the number of simulated drift regions passing a given selection step to the total number of drift regions in that sample. At smaller kink angles of  $10^\circ$  and  $20^\circ$ , the misidentification rate due to crossing muons is relatively high. Further restricting the angle selection requirement from  $> 20^\circ$  to  $> 30^\circ$  reduces the misidentification rate from 8.1% to 4.5%—a reduction of nearly a factor of 2—while the Michel electron efficiency drops from 33.0% to 20.4%. Tightening the requirement further to  $40^\circ$  reduces the efficiency from 20.3% to 12.6%, but only marginally improves the misidentification rate (from 4.5% to 3.1%). Based on this trade-off, a kink angle requirement of  $> 30^\circ$  was chosen for optimal Michel electron selection. The data selection algorithm’s corresponding performance on the CORSIKA muon sample is shown in the last column of table 2.

## 4.2 Demonstration Using Real Data

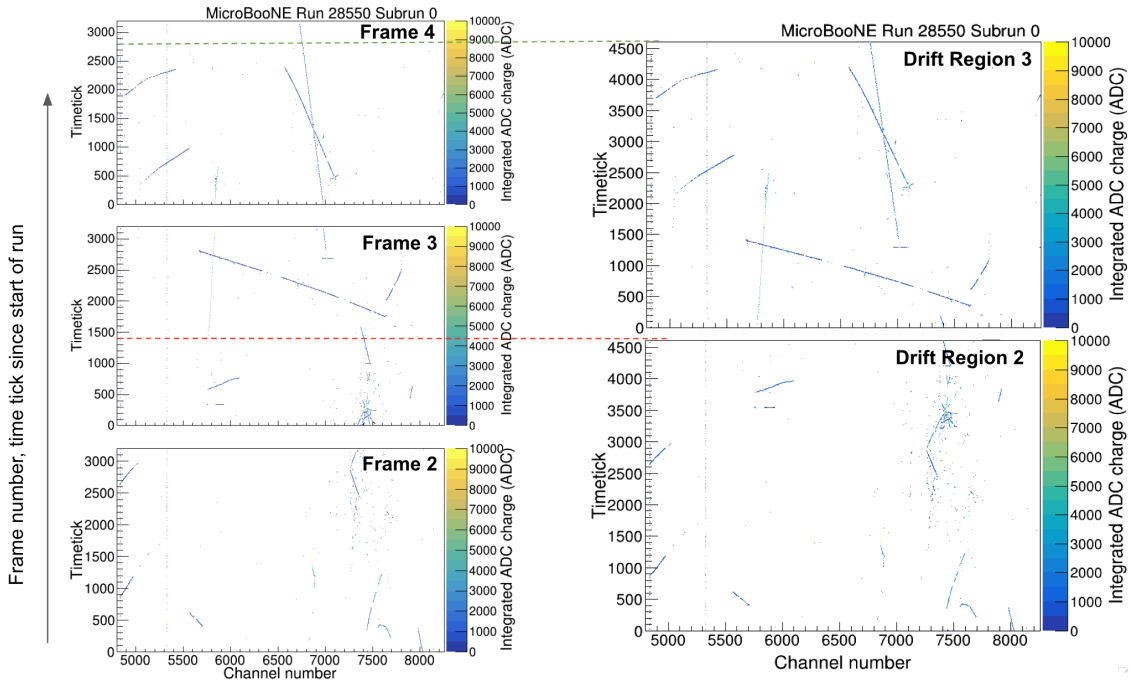
The MicroBooNE detector completed beam data-taking and operations in 2020. To facilitate these online data selection demonstration studies beyond MicroBooNE’s end of operations, an “emulated online” TPC-based data selection framework, as introduced in subsection 2.2, was developed and exercised on pre-recorded MicroBooNE SN ROI data. As part of this framework, the SN ROI data was processed through three sequential software processes, described below, which were executed online on CPUs.

### 4.2.1 Process 1: Trigger Primitive Generation

Process 1 begins by pre-processing the SN data sample (collected after the experiment stopped collecting beam data), as described in subsection 3.1. After pre-processing, TPs, which are summaries of per-channel waveform ROI information (including maximum amplitude, ADC integral, and width of the ROI, accompanied by ROI channel number and ROI start timetick), are extracted from the SN ROIs. Process 1 then maps the channel numbers of each ADC+FEM board (in the range of 0 – 63) to collection plane channel numbers (4800 – 8256) using a dedicated channel map. Subsequently, the TP data is rearranged into “drift regions”. The SN ROI data in MicroBooNE is streamed in packets of 1.6 ms time intervals, referred to as “frames”, whereas a drift region in MicroBooNE corresponds to 2.3 ms. To increase the probability of capturing all causally-connected charge deposited within a single interaction topology, it is therefore beneficial to reorganize the SN stream data from successive frames into successive drift regions, such that signals over an entire drift length can be considered. This rearrangement is referred to as “stitching”.

As such, Process 1 begins by pre-processing SN data, reading three consecutive frames worth of SN data, evaluating TPs, stitching TPs from consecutive frames to generate a contiguous drift region’s worth of TP data and then masking the noisier channels which appear as vertical lines in the event displays. In practice, three frames generate TPs corresponding to two full drift regions (each equivalent to 4,600 time ticks), and the remaining TPs are held in memory until the next set of three frames are received by the process. Figure 12 (left) shows the distributions of TPs per frame (3,200 ticks), before stitching, for three consecutive frames (frame 2, frame 3, and frame 4), with the color  $z$  scale representing the integrated ADC charge, and  $x$  and  $y$  coordinates representing the (physical) ROI channel number and start time. The panel on the right shows the rearrangement of the TPs into drift regions (4,600 ticks), after stitching. The dashed red line marks the end of the first drift region, built from all TPs in frame 2 plus 1,400 ticks from frame 3 to complete a 4,600-tick drift region. The dashed green line marks the end of the second drift region, formed by combining 1,800 ticks from frame 3 with 2,800 ticks from frame 4. The remaining 400 ticks from frame 4 are carried over to the next set of three consecutive frames. Some channels exhibiting noise as vertical bands around channel number 4,800 and 5,300, as seen in figure 12, are masked.

Since the streaming of SN data is distributed across eight TPC readout crates, eight separate instances of Process 1 execute concurrently to generate TPs for each drift region, as shown in figure 4. The distribution of processing times per two drift regions for each of these instances is shown in figure 13 (left), confirming that Process 1 can be executed with sufficiently low latency (on average, well within  $2 \times$  drift time = 4.6 ms) so as to keep up with the detector-generated data rates.



**Figure 12:** (Left) The distribution of TPs in three consecutive frames, each corresponding to 3,200 time ticks in length. (Right) The distributions of trigger primitives in two successive drift regions obtained following the stitching step of Process 1. The dashed red and green lines represent the ending of the first and second drift region respectively.

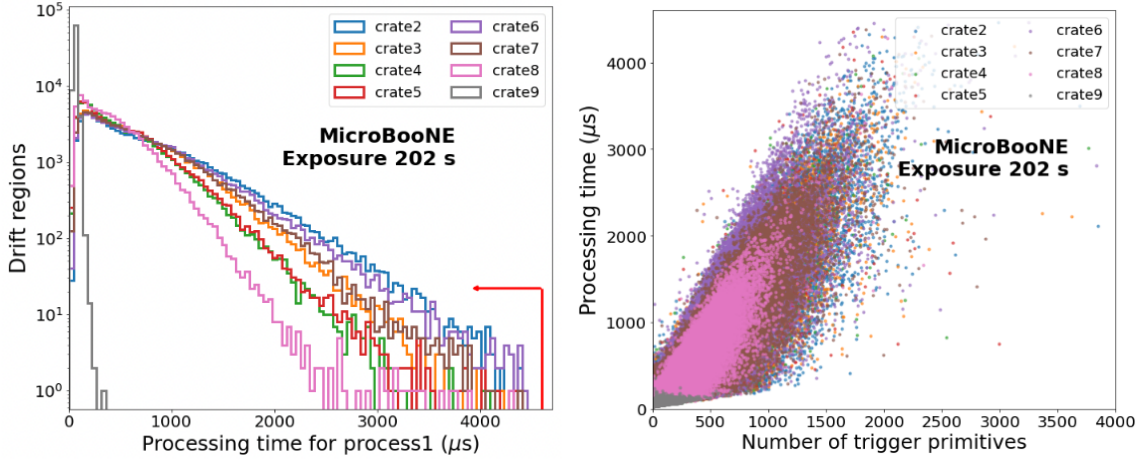
The processing time is correlated with the number of TPs generated by each instance as shown in figure 13 (right).

Following Process 1, the processed TP data from each drift region are handed to a second, TC generation process (as shown in figure 4) via zeroMQ [15], a high-performance messaging library, without any loss in data. The TC generation step (Process 2) is described in the following subsection. The information sent out to Process 2, organized by drift region, includes run number, subrun number, crate number, frame number, drift region number, ADC+FEM board number, TP channel numbers within each ADC+FEM board, TP times, and corresponding TP information (maximum amplitude, ADC integral and width of the ROI).

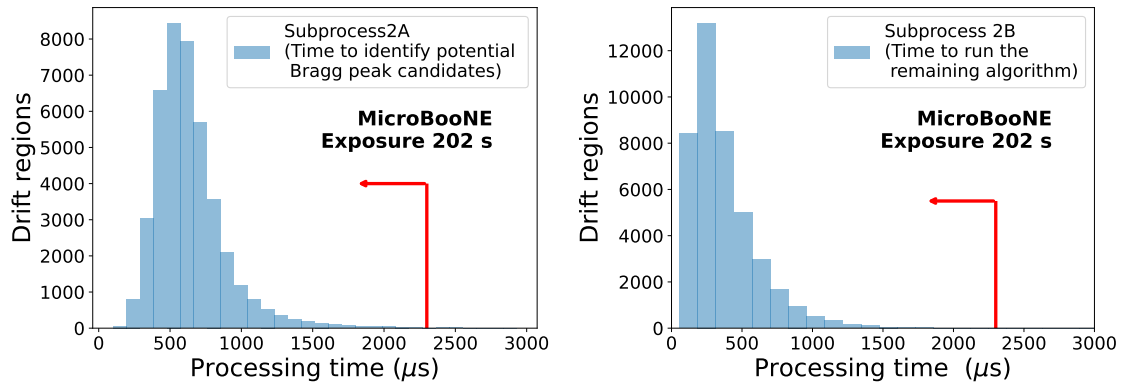
#### 4.2.2 Process 2: Trigger Candidate Generation

Process 2 receives a full packet of information including TPs per drift region from Process 1 via zeroMQ. Because Process 1 generates two drift regions every time it is executed, two instances of Process 2 run simultaneously to process the two drift regions in parallel, as shown in figure 4. Each Process 2 instance is further subdivided into two subprocesses, 2A and 2B, which execute steps 2 and 3 of the data selection algorithm, respectively, as described in section 4.1. Drift regions that meet the selection criteria of the algorithm are down-selected as TCs.

The distribution of execution times per drift region for each of the subprocesses 2A and 2B is shown in figure 14. The processing latency per drift region for each of the two subprocesses is

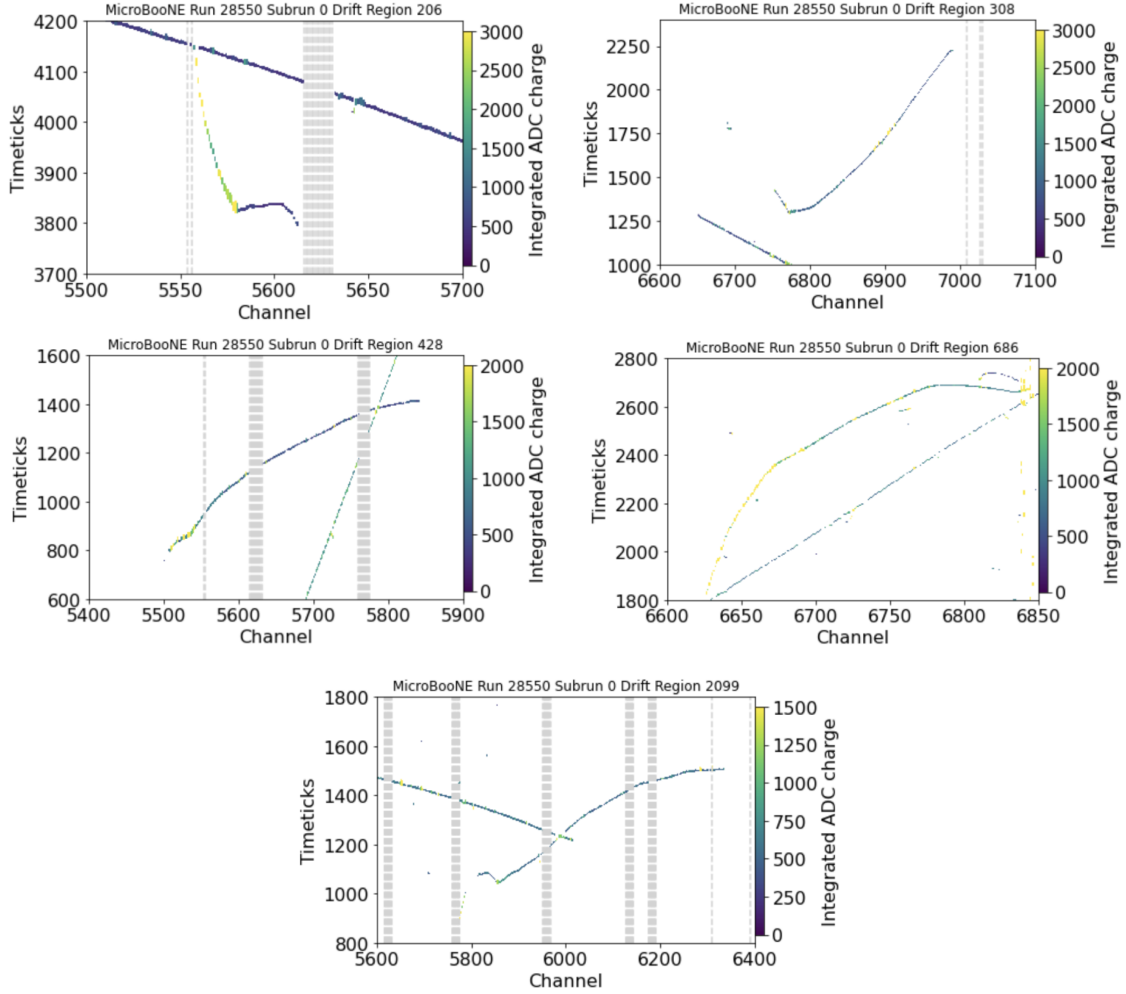


**Figure 13:** (Left) Distribution of execution time of Process 1 for each of the eight instances running in parallel. The red arrow shows the maximum allowed latency of 4.6 ms to keep up with data throughput. (Right) Distribution of number of TPs per drift region generated by each of the eight instances of Process 1.



**Figure 14:** Distribution of execution times for Process 2 separated into each of the subprocesses; subprocess2A (left) and subprocess2B (right). The red arrow shows a conservatively set maximum allowed latency of 2.3 ms, as information from Process 1 is handed to each instance of Process 2 in a time-serial manner.

within acceptable limits. In terms of physics performance validation, out of 87,936 consecutive drift regions (corresponding to 202 seconds of SN stream exposure), 18,465 regions (21%) are selected, consistent with the CORSIKA performance shown in table 2. Figure 15 shows example zoomed-in event displays of Process 2-selected drift regions, containing Michel candidates that satisfy the Bragg-peak and kink-angle selection criteria of Process 2. The gray bands in each display correspond to dead channels.

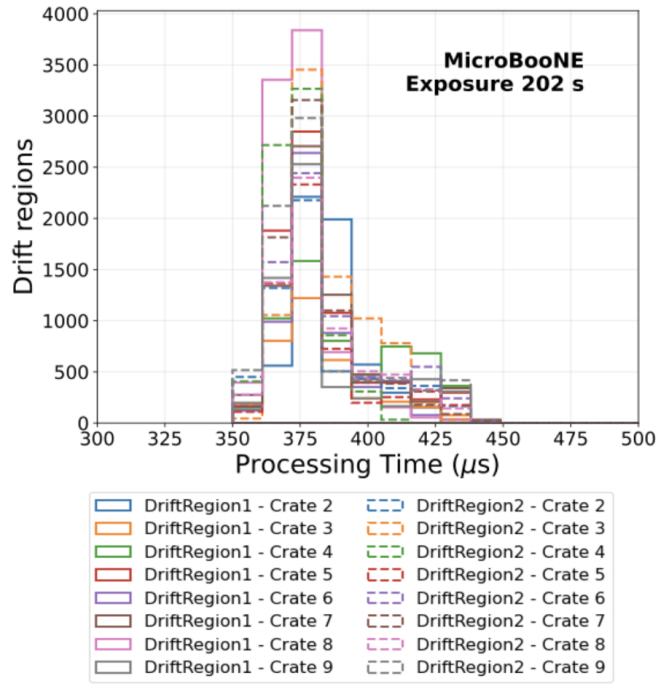


**Figure 15:** Examples of drift regions with Michel candidates (satisfying the Bragg peak and kink angle selection criteria), as selected by Process 2.

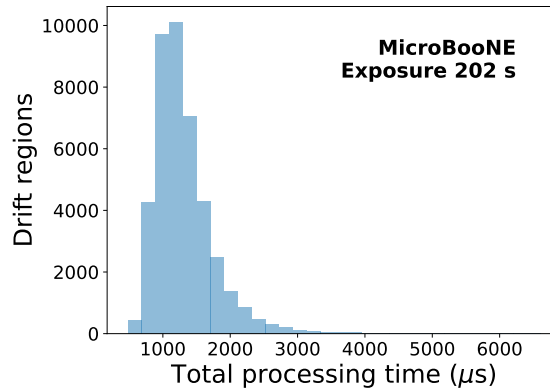
### 4.2.3 Process 3: High-Level Trigger

Process 3 receives TC information which contains the frame number, drift region number, TPC crate number, run number and subrun number from Process 2-selected drift regions, and for each Process 2 instance. Using the TC information, Process 3 parses through the disk-stored data from the SN readout stream, in order to retrieve the entire SN ROI data for the particular frame in which the TC was identified, as well as the frame before and the frame after the TC-identified frame, and stores this data back to disk. As shown in figure 4, in total, sixteen instances of Process 3 run simultaneously to process candidate information from two drift regions (from Process 2) so as to save data from eight readout crates, which are distributed across separate SEB servers; the distribution of execution times per TC for each of the sixteen instances is shown in figure 16.

Figure 17 shows the total processing time per drift region for running all three processes for the selected drift regions. The total processing time distribution peaks at around  $1195 \mu\text{s}$ , while the mean value is slightly higher ( $1315 \mu\text{s}$ ), reflecting a small population of events with longer



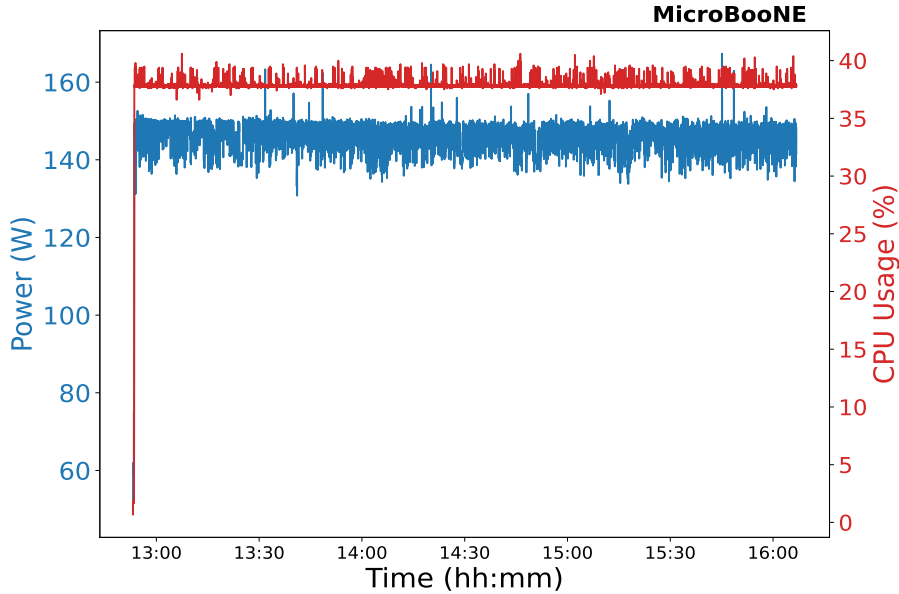
**Figure 16:** Processing time per three frames to run Process3.



**Figure 17:** Total processing time to run all three processes for the selected drift regions.

than average processing durations, but remains lower than the maximum allowed latency of 2.3 ms. Since the average processing time is below 2.3 ms, the system processes events faster than they arrive and therefore keeps up without backing up. This demonstrates that the algorithm is fast enough to keep up with the detector rates in an online system.

The framework was executed on a server equipped with dual-socket Intel<sup>®</sup> Xeon<sup>®</sup> Gold 6130 CPUs (32 physical cores, 64 threads total) and 93 GB of RAM, running CentOS 7 with Linux kernel 3.10.0-1160.102.1.el7.x86\_64. While running this framework, CPU power consumption and utilization were monitored using accumulated energy values from Intel RAPL (Running Average



**Figure 18:** CPU power consumption (blue, left axis) and utilization (red, right axis) over time during framework execution.

Power Limit) and the top command, respectively. Figure 18 shows the variation of CPU power consumption and utilization over time, with the  $x$ -axis representing elapsed time in hour:minutes from the start of the framework run. The left  $y$ -axis shows instantaneous CPU power consumption (in Watts, shown as the blue color), while the right  $y$ -axis indicates total CPU utilization (in percentage, shown as the red color). When the framework is started, the CPU operates at very low utilization levels (below 2%), with corresponding power consumption around 60 W. When all the processes are started, CPU activity increases sharply, reaching approximately 40% utilization, which results in a corresponding rise in power consumption to nearly 150 W. After this increase, both power and CPU usage remain relatively stable, indicating a sustained workload and consistent power draw during the remainder of the running period.

## 5 Future Applications and AI/ML Developments

Building upon the developments and results presented in this work, several future applications and extensions are envisioned, particularly in the context of “online” or “real-time” data processing and selection incorporating artificial intelligence and machine learning (AI/ML) techniques, for improved physics performance. Such approaches can be developed and tested or applied to ongoing and future experiments such as SBND, ICARUS [16] or DUNE. In particular, SBND shares a functionally identical TPC readout electronics design with MicroBooNE, ensuring that the framework developed and validated within MicroBooNE can be seamlessly transferred and deployed for “online” or “real-time” demonstrations in SBND.

Looking ahead to DUNE, the demonstrated feasibility of topologically-sensitive TPC-based data selection in MicroBooNE provides a strong foundation for integrating similar traditional or AI/ML techniques to further enhance trigger and data selection performance. Beyond Michel

electrons, this framework could be applied to efficiently identify neutrino interactions online across a broad energy range — from a few tens of MeV up to several GeV — where backgrounds from low-energy sources are significantly reduced, enabling clearer identification of high-energy signals such as beam and atmospheric neutrinos (exploiting, for example track multiplicity [17]).

To extend these capabilities toward the few-MeV to sub-MeV regime, further improvements in TP generation can be pursued, including the integration of advanced deep-learning-based noise filtering algorithms [18, 19] and the development of more robust, intelligent data selection methods. Deep learning architectures, such as convolutional and graph neural networks, can exploit the full spatial and temporal correlations present in TPC signals to identify rare physics signatures with high efficiency and low latency. One such algorithm has already been developed and demonstrated by MicroBooNE to search for the rare neutron–antineutron transition process [3]. Integrating similar algorithms into the online data acquisition framework could enable adaptive, self-optimizing triggers that dynamically respond to changing detector and beam conditions.

Consequently, the work presented here not only provides an essential foundation but also serves as a key enabler for establishing the advanced triggering capabilities that will be required for future LArTPC experiments to fully exploit their physics potential.

## 6 Summary and Conclusions

We present the first demonstration of an “emulated online”, physics-topology-sensitive, TPC-based data selection, using MicroBooNE data. The developed framework is validated through the successful identification of targeted low-energy Michel electrons from stopping cosmic-ray muons, using both topological and calorimetric information. We have demonstrated a data selection based on LArTPC features that is fast enough to operate online—even within a challenging on-surface operating detector environment. While future implementations will need to be adapted to specific detector designs and use cases, this work shows that the challenging problem of “online” or “real-time” TPC-based data selection is indeed solvable, and that such data selection can effectively incorporate both topological and calorimetric information from the TPC. This study thus provides an important proof of concept for currently operating LArTPCs and a strong foundation for future large-scale LArTPC applications such as DUNE.

## Acknowledgments

This document was prepared by the MicroBooNE collaboration using the resources of the Fermi National Accelerator Laboratory (Fermilab), a U.S. Department of Energy, Office of Science, Office of High Energy Physics HEP User Facility. Fermilab is managed by Fermi Forward Discovery Group, LLC, acting under Contract No. 89243024CSC000002. MicroBooNE is supported by the following: the U.S. Department of Energy, Office of Science, Offices of High Energy Physics and Nuclear Physics; the U.S. National Science Foundation; the Swiss National Science Foundation; the Science and Technology Facilities Council (STFC), part of United Kingdom Research and Innovation (UKRI); the Royal Society (United Kingdom); the UKRI Future Leaders Fellowship; the NSF AI Institute for Artificial Intelligence and Fundamental Interactions; and the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie

grant agreement No. 101003460 (PROBES). Additional support for the laser calibration system and cosmic ray tagger was provided by the Albert Einstein Center for Fundamental Physics, Bern, Switzerland. We also acknowledge the contributions of technical and scientific staff to the design, construction, and operation of the MicroBooNE detector as well as the contributions of past collaborators to the development of MicroBooNE analyses, without whom this work would not have been possible. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) public copyright license to 9434636208338275448065any Author Accepted Manuscript version arising from this submission.

## References

- [1] MICROBooNE collaboration, *Design and Construction of the MicroBooNE Detector*, *J. Instrum.* **12** (2017) P02017 [[1612.05824](#)].
- [2] MICROBooNE collaboration, *The continuous readout stream of the MicroBooNE liquid argon time projection chamber for detection of supernova burst neutrinos*, *J. Instrum.* **16** (2021) P02008 [[2008.13761](#)].
- [3] MICROBooNE collaboration, *First application of a liquid argon time projection chamber for the search for intranuclear neutron-antineutron transitions and annihilation in  $^{40}\text{Ar}$  using the MicroBooNE detector*, *J. Instrum.* **19** (2024) P07032 [[2308.03924](#)].
- [4] DUNE collaboration, *Long-baseline neutrino oscillation physics potential of the DUNE experiment*, *Eur. Phys. J. C* **80** (2020) 978 [[2006.16043](#)].
- [5] DUNE collaboration, *Prospects for beyond the Standard Model physics searches at the Deep Underground Neutrino Experiment*, *Eur. Phys. J. C* **81** (2021) 322 [[2008.12769](#)].
- [6] DUNE collaboration, *Supernova neutrino burst detection with the Deep Underground Neutrino Experiment*, *Eur. Phys. J. C* **81** (2021) 423 [[2008.06647](#)].
- [7] MICROBooNE collaboration, *Measurement of the atmospheric muon rate with the MicroBooNE Liquid Argon TPC*, *J. Instrum.* **16** (2021) P04004 [[2012.14324](#)].
- [8] SNEWS collaboration, *SNEWS 2.0: a next-generation supernova early warning system for multi-messenger astronomy*, *New J. Phys.* **23** (2021) 031201 [[2011.00035](#)].
- [9] DUNE collaboration, *Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume IV: Far Detector Single-phase Technology*, *J. Instrum.* **15** (2020) T08010 [[2002.03010](#)].
- [10] DUNE collaboration, *Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume II: DUNE Physics*, [[2002.03005](#)].
- [11] SBND collaboration, *Liquid Argon TPC Trigger Development with SBND*, in *Meeting of the Division of Particles and Fields of the American Physical Society*, 10, 2019 [[1910.08218](#)].
- [12] MICROBooNE collaboration, *Liquid Argon Time Projection Chamber Trigger Development with MicroBooNE*, *J. Phys. Conf. Ser.* **2374** (2022) 012163.
- [13] DUNE collaboration, *Design, construction and operation of the ProtoDUNE-SP Liquid Argon TPC*, *J. Instrum.* **17** (2022) P01005 [[2108.01902](#)].
- [14] D. Heck et al., *CORSIKA: A Monte Carlo code to simulate extensive air showers*, Tech. Rep. FZKA-6019, Forschungszentrum Karlsruhe (1998).

- [15] P. Hintjens, “ZeroMQ – The Guide.” <http://zguide.zeromq.org/page:all>, 2010.
- [16] ICARUS collaboration, *ICARUS at the Fermilab Short-Baseline Neutrino program: initial operation*, *Eur. Phys. J. C* **83** (2023) 467 [2301.08634].
- [17] S. Chung et al., *Real-time Anomaly Detection for Liquid Argon Time Projection Chambers*, [2509.21817].
- [18] ARGONEUT collaboration, *A deep-learning based raw waveform region-of-interest finder for the liquid argon time projection chamber*, *J. Instrum.* **17** (2022) P01018 [2103.06391].
- [19] Y.-J. Jwa, G. Di Guglielmo, L. Arnold, L. Carloni and G. Karagiorgi, *Real-Time Inference With 2D Convolutional Neural Networks on Field Programmable Gate Arrays for High-Rate Particle Imaging Detectors*, *Front. Artif. Intell.* **5** (2022) 855184 [2201.05638].