Data validation for training a reconstruction ML model for IWCD

Gilberto Rodriguez Prado
Host Researcher: SAKASHITA Ken
Associated Researcher: Patrick de Perio
Associated Researcher: Ka Ming Tsui

Sokendai
High Energy Accelerator Research Organization
WatChMaL

July 23, 2025

Introduction

Introduction about me (part 1)

1. My name is Gilberto Rodriguez. I am a 22-year-old Physics Engineering Graduate (top of my class) from Tecnológico de Monterrey in Mexico.

Tecnológico de Monterrey

2. I am currently working with Machine Learning (ML) in Neutrino reconstruction for the IWCD detector from the Hyper-K experiment. I work alongside WatChMaL, a group dedicated to ML in neutrino reconstruction.

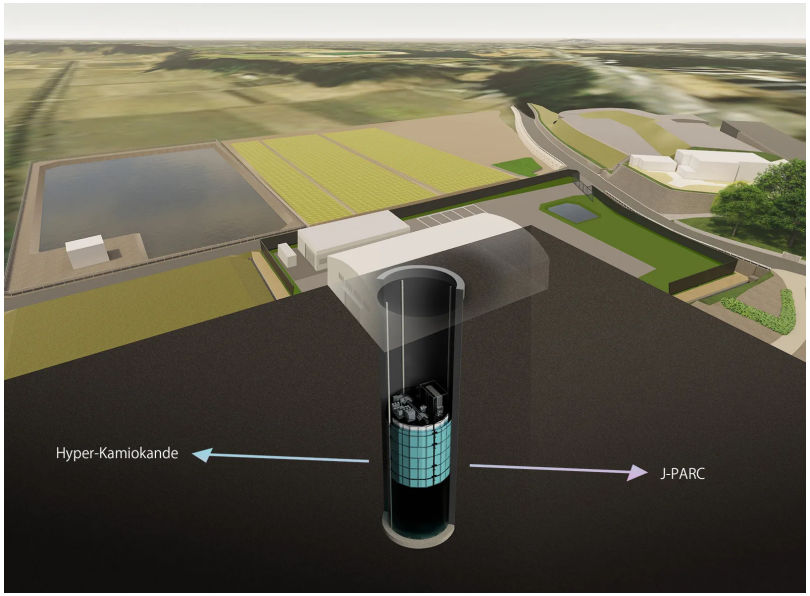
Introduction about me (part 2), Hyper-K CM

Introduction about me (part 3)

3. I am very into lifting weights, photography, reading, anime, and (seldomly) videogames. My favorite book series is **The Dark Tower** by Stephen King; my favorite anime is **Evangelion**, and my favorite video game is **Elden Ring**. My favorite album is **Kiss Land by The Weeknd**, I mostly listen to R&B.

Introduction

1. This presentation is about a brief validation of a sample of simulations of a particle gun inside the Intermediate Water Cherenkov Detector (IWCD).
2. The IWCD is a Cherenkov detector mounted on an elevator to measure different off-axis angles. The geometry that I used is 8 meters tall and has a diameter of 7 meters.



(a) IWCD's off axis elevator mechanism.

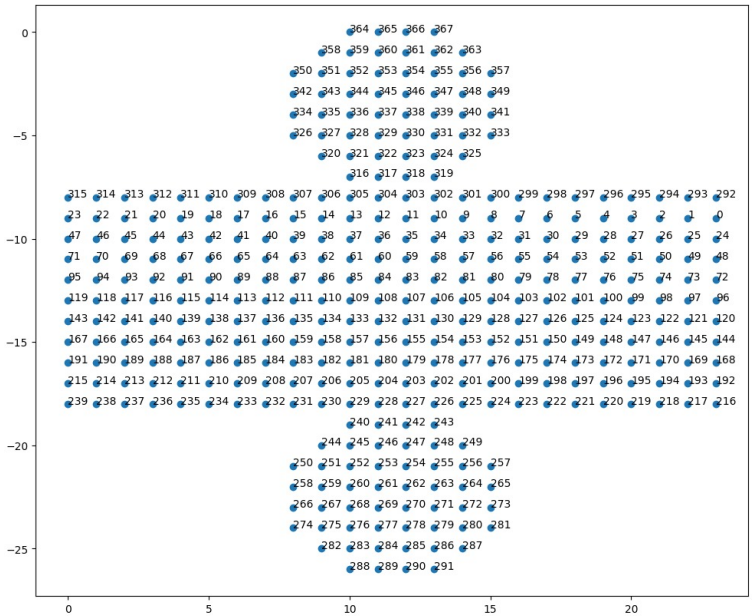
(b) IWCD's array of mPMT's.

Figure: Two complementary views of the Intermediate Water Cherenkov Detector (IWCD).

Introduction

Introduction

1. The simulations were done by the Hyper-K production team with WCSim. The version used is v1.12.19. This is not the latest version.
2. My main objective is to train a regression model that can reconstruct accurately position, energy, and direction.
3. Before starting training, I had to study and process the data that I was working with. I had to make sure that it was an isotropic dataset so that the model I'm training is not biased.
4. The scope of this presentation is to show the distribution of input/output variables after processing for ML.



Progress

Progress Overview (e/mu events)

1. The dataset that I'm using has 3'097,499 events of both electron and muon events.
2. IWCD has an Outer Detector that helps determine if the event is fully contained (FC) or not. For this sample, FC is determined with true values.

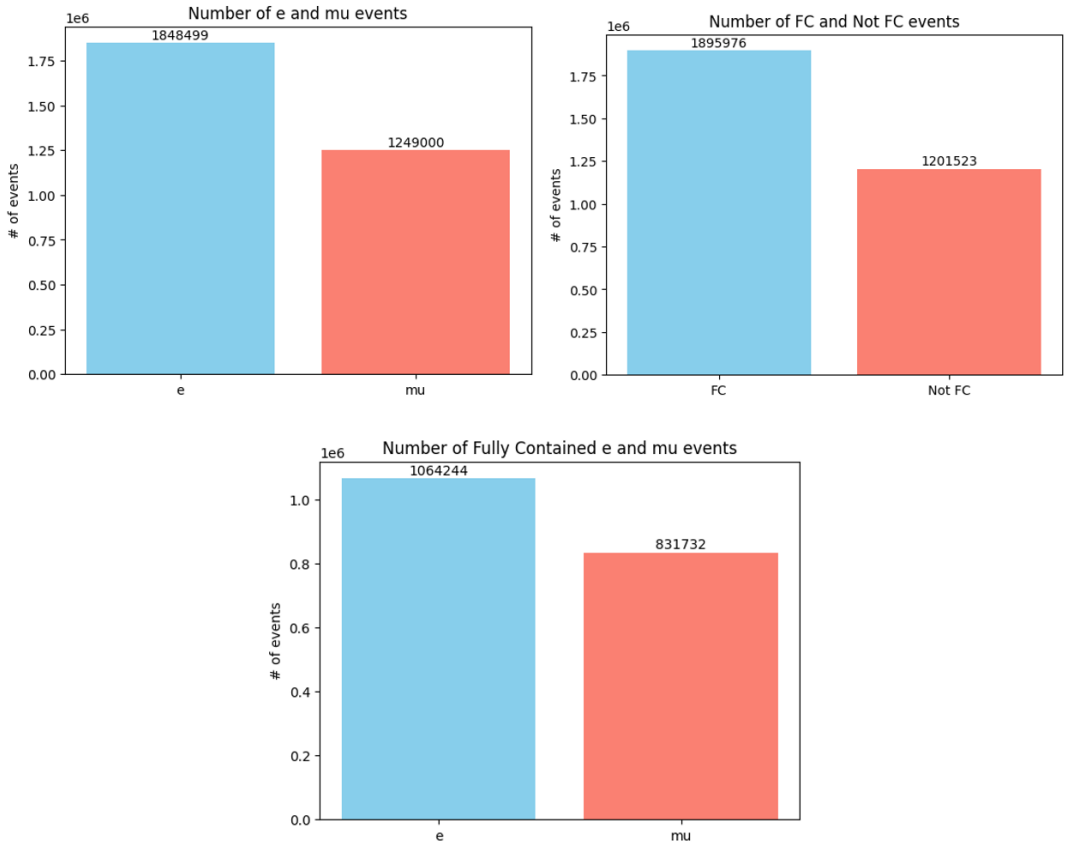


Figure: e/mu event distribution, FC/non FC event distribution, e/mu FC event distribution.

Progress Overview (nhits distribution)

1. Many events are empty events. This may be due to incorrect or corrupted triggers.
2. Deleting these events from the training sample is important to avoid bias.
3. In this test production, WCSim used a *finite trigger threshold*, so events bellow 22 hits are discarded.

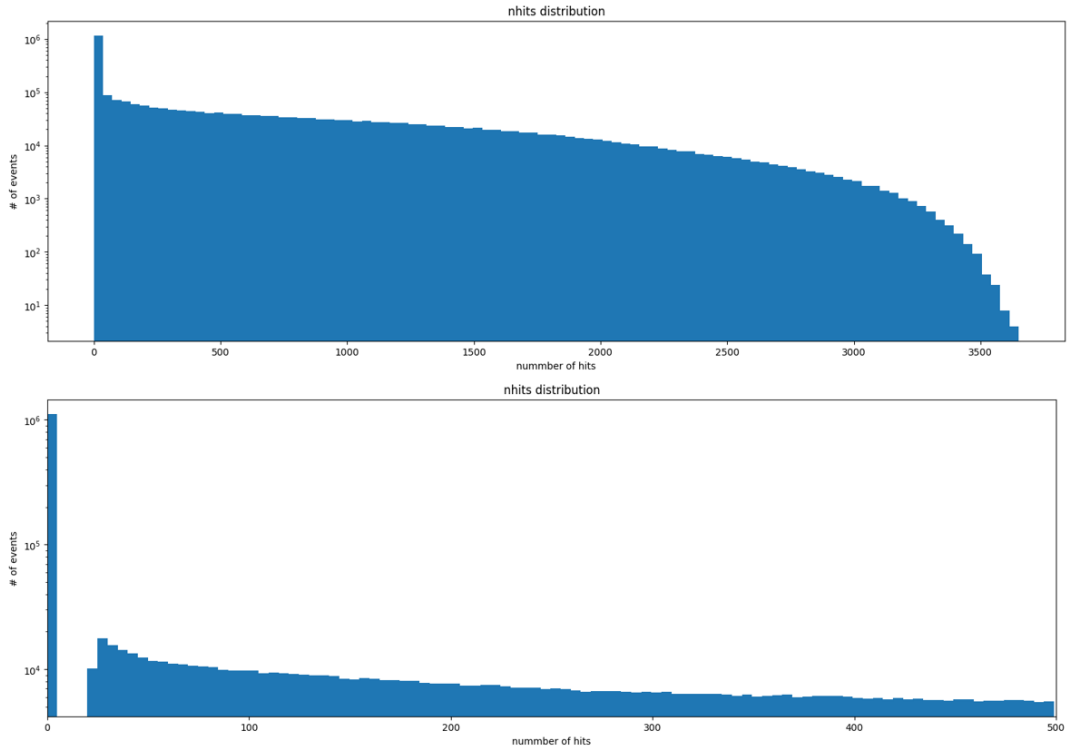


Figure: Distribution of number of hits per event

Progress Overview (Position distribution)

1. Since the IWCD geometry is a cylindrical tank. The position distribution in cylindrical coordinates proves that this MC is uniformly distributed for positions.

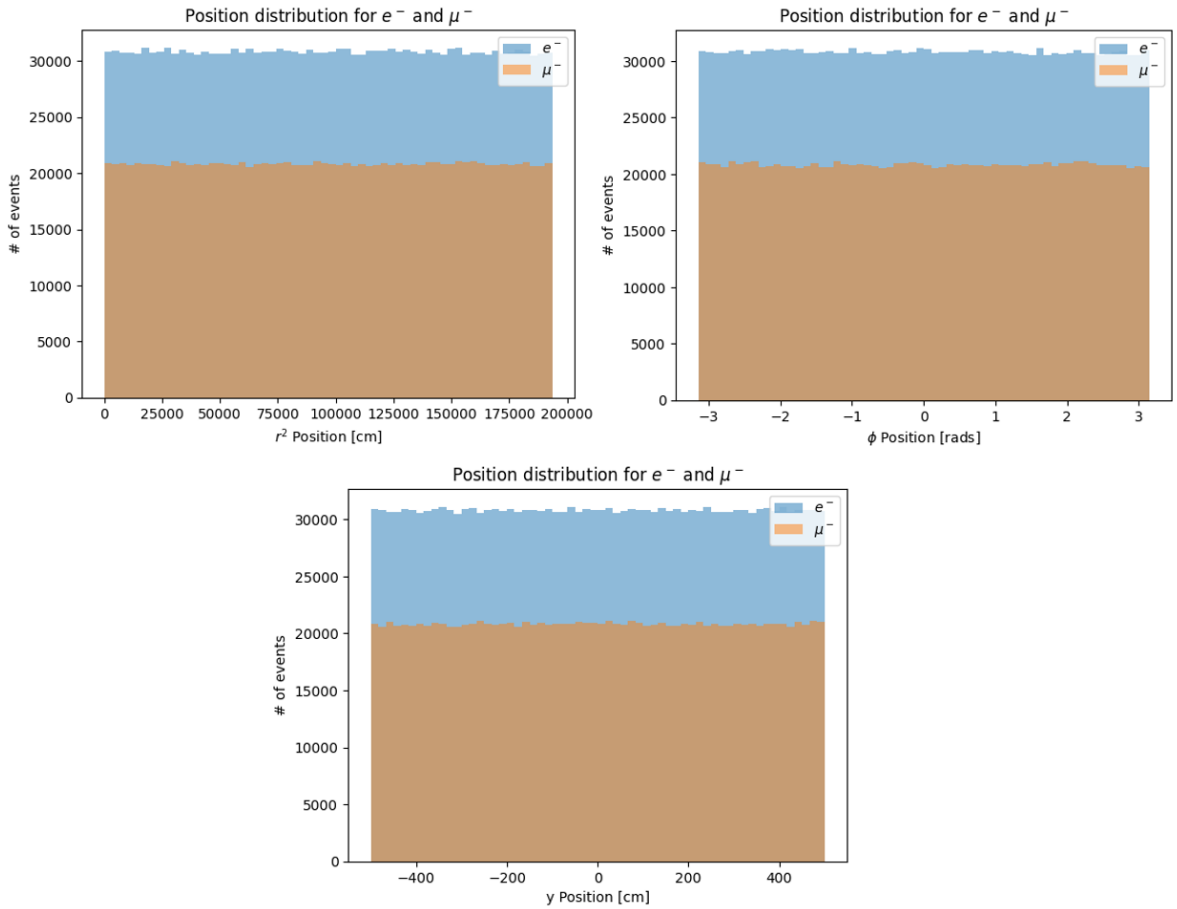


Figure: Position uniformly distributed in cylindrical coordinates

Progress Overview (Direction distribution)

1. Since both position and direction are uniformly distributed, this is an isotropic MC, meaning that there is no *preferred* direction in space, and every solid-angle patch on the sky is equally likely.

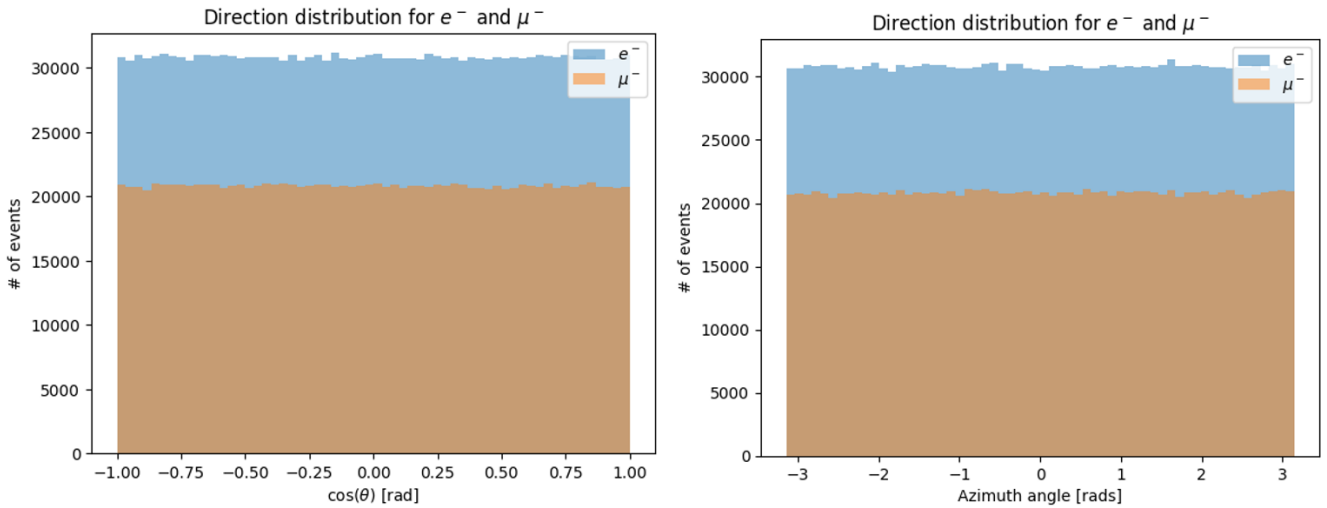


Figure: Direction uniformly distributed

Progress Overview (Energy distribution)

1. Energy is distributed uniformly.
2. Electron Cherenkov threshold in water is 0.8 MeV, while muon is 160 MeV. That explains why the electron sample covers a bigger energy range.
3. I use mostly sub-GeV events. At IWCD this roughly corresponds to the highest lepton energy, due to flux and cross sections of neutrino interactions.

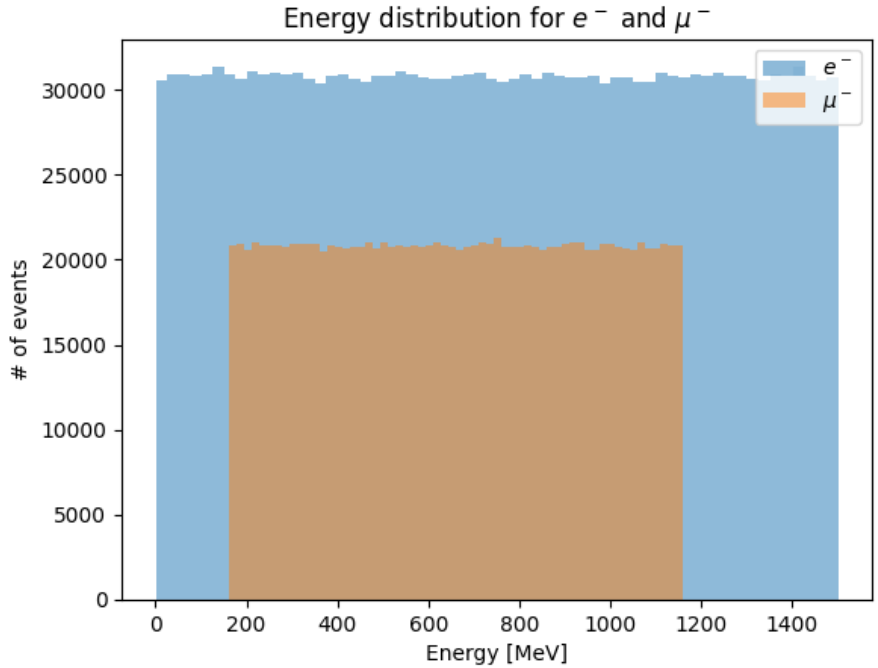
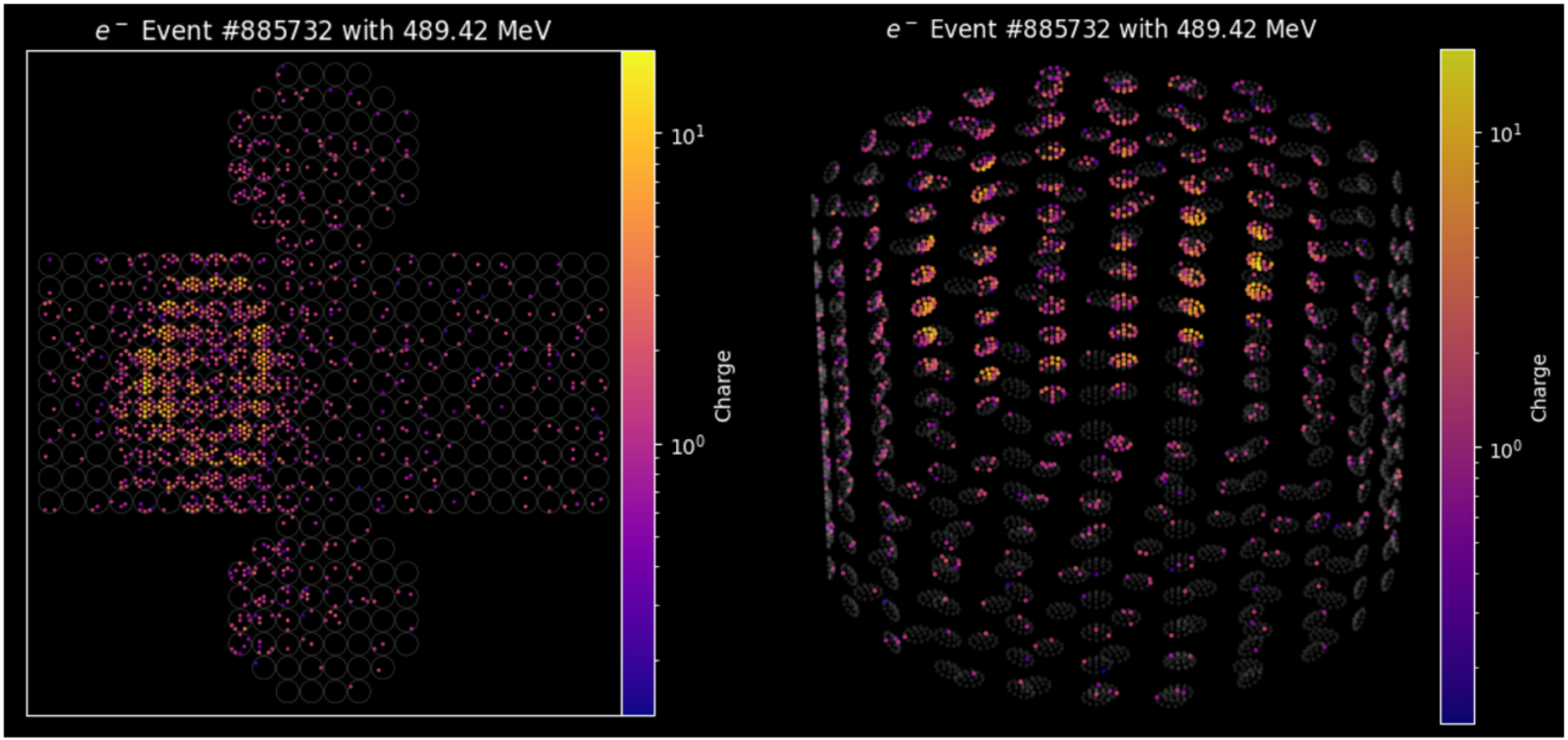


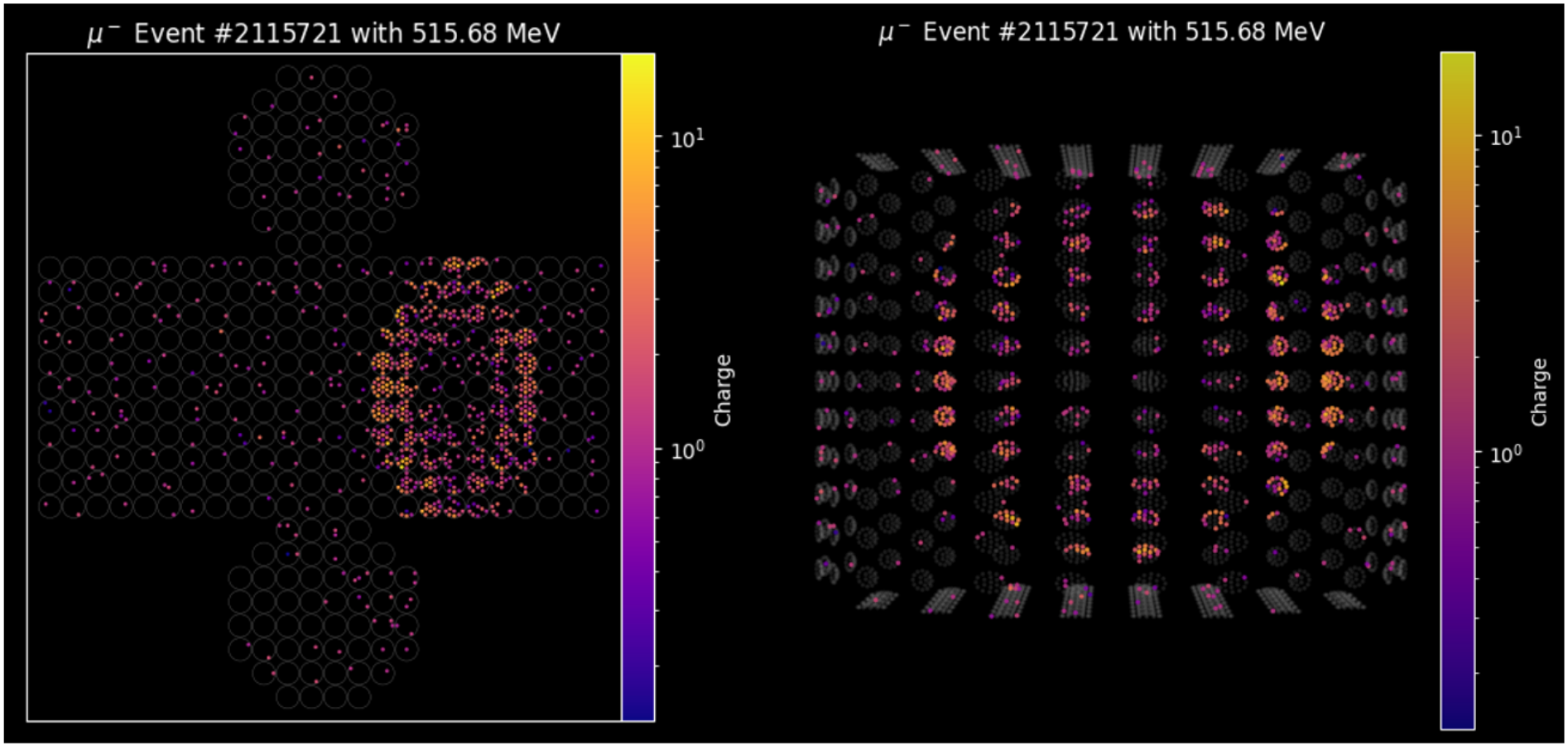
Figure: Energy uniformly distributed

Results

Electron event display



Muon event display



Next Steps

Next Steps

1. Train a ResNet-50 ML model to reconstruct the output variables (position, energy, and direction).
2. Start using the latest production sample made with WCSim version 1.12.22.

Thank You!