

Optimal transport for e/π^0 particle classification in LArTPC neutrino experiments

David Caratelli^{1,*}, Nathaniel Craig^{1,2,†}, Chuyue Fang^{1,‡} and Jessica N. Howard^{1,2,§}

¹*University of California, Santa Barbara, USA*

²*Kavli Institute for Theoretical Physics, Santa Barbara, California, USA*



(Received 24 June 2025; accepted 18 February 2026; published 8 April 2026)

The efficient classification of electromagnetic activity from π^0 and electrons remains an open problem in the reconstruction of neutrino interactions in liquid argon time projection chamber (LArTPC) detectors. We address this problem using the mathematical framework of optimal transport (OT), which has been successfully employed for event classification in other high energy physics contexts and is ideally suited to the high-resolution calorimetry of LArTPCs. Using a publicly available simulated dataset from the MicroBooNE Collaboration, we show that OT methods achieve state-of-the-art reconstruction performance in e/π^0 classification. The success of this first application indicates the broader promise of OT methods for LArTPC-based neutrino experiments.

DOI: [10.1103/6zws-fnrf](https://doi.org/10.1103/6zws-fnrf)

I. INTRODUCTION

The observation of neutrino oscillations [1] is currently one of the most compelling motivations for the existence of physics beyond the Standard Model (BSM), making a better understanding of neutrino properties and interactions a crucial goal of particle physics in the coming years. Ongoing and future experiments such as those which are part of the Short Baseline Neutrino (SBN) program [2] or the Deep Underground Neutrino Experiment (DUNE) [3,4] rely on liquid argon time projection chamber (LArTPC) detectors for granular imaging and calorimetry of particle interactions over large volumes. It is therefore important to explore and understand potential data analysis strategies that can meet and expand on the physics goals of future LArTPC experiments.

A main feature of LArTPCs is their significantly enhanced resolution capabilities compared to other neutrino detector strategies [5–7]. This increased resolution is often not fully exploited by traditional analysis techniques. Many groups [8–13] have explored using machine learning (ML) to improve the data analysis pipeline.

While powerful, ML can also present difficulties related to interpretability and uncertainty quantification.

Optimal transport (OT) is a mathematical framework that furnishes a notion of the distance between two probability distributions in terms of the minimal “work” required to rearrange one into another. In recent years OT methods have proven powerful for image processing, which suggests their relevance to data analysis in high energy physics (HEP) experiments. Indeed, OT methods have been fruitfully applied to the analysis of Large Hadron Collider (LHC) data [14–44]. These applications have taken a variety of forms. For example, recently OT transport plans have been used to calibrate ATLAS flavor-tagging algorithms [15], to decorrelate observables [16], generate templates [17], and estimate backgrounds [18]. The majority of applications of OT to data analysis rely on calculating distances between distributions of interest. For example, OT distances have been used on Dalitz distributions to measure Charge-Parity (CP) violation [19] or on the loss functions of trained Variational Autoencoders (VAEs) to help identify anomalies [20]. However, the applications most relevant to the current work choose to represent a collider event or object (e.g., jet) as a discrete distribution over an underlying ground space [21–44]. Subsequently, OT distances are calculated between these event representations [21–40] or between the event representation and a constructed reference [39–44] to define a similarity metric. Note that for a certain choice of cost function the latter can implicitly approximate the pairwise distances between events [31,35].

In general, OT distances between events can be used for event classification and anomaly detection on their own or coupled to simple distance-based machine learning

*Contact author: dcaratelli@ucsb.edu

†Contact author: ncraig@ucsb.edu

‡Contact author: cfang@ucsb.edu

§Contact author: jnhoward@kitp.ucsb.edu

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

methods. The performance of these methods is often competitive with standard deep learning approaches, but more amenable to interpretation and uncertainty quantification. In this respect, OT provides a promising foundation for interpretable, physics-informed machine learning methods in HEP.

The success of OT methods applied to LHC datasets naturally suggests their relevance to other HEP calorimeters. In this work, we explore how OT can be used for data analysis in LArTPC experiments. As a test case, we explore the performance of OT on the task of e/π^0 classification. Neutral pion background rejection is particularly challenging for current algorithms, leading to significant backgrounds in analyses with electron [45] and single-photon [46,47] final states. We find that OT distances can be used to yield state-of-the-art classification performance on the task of e/π^0 separation. This is particularly true when OT distances are combined with support vector machines (SVMs), an interpretable ML method. More generally, we find that all OT-based methods explored in this work significantly outperform traditional reconstruction algorithms.

This paper is organized as follows. Section II describes the LArTPC technology and the physics that LArTPC detectors located in neutrino beams are exploring. Section III describes the simulation and low-level reconstruction that are used to produce Monte Carlo samples for analysis, and in particular discusses the “MicroBooNE Open Dataset” used by this work for accurate rendering of neutrino images in a LArTPC. Section IV discusses the task of e/π^0 classification, which is the focus of this work. Section V reviews OT distances and motivates our particular choice (balanced 2-Wasserstein distances) in the context of e/π^0 classification. Section VI overviews the data preprocessing, and Sec. VII discusses the postprocessing methods employed in this work. The results are presented in Sec. VIII, followed by a brief conclusion which discusses directions for future work in Sec. IX.

II. LIQUID ARGON TIME PROJECTION CHAMBER NEUTRINO DETECTORS

A. Physics landscape with LArTPC detectors

Because of their imaging capabilities, LArTPC detectors have become the technology of choice for HEP experiments operating in accelerator-based neutrino beams. LArTPCs are able to capture images of particle interactions with millimeter position accuracy and percent-level calorimetric resolution [48–50]. This, in turn, allows for excellent particle identification which is leveraged to perform precision measurement of neutrino oscillations [51] and searches for BSM particles [52–54]. These detectors complement other technologies employed for accelerator-based neutrino experiments, including water-Cherenkov detectors such as the next-generation Hyper-Kamiokande [55] experiment.

Common final states in LArTPCs exposed to neutrino beams consists of electromagnetic showers produced by $\mathcal{O}(0.01\text{--}1)$ GeV electrons and photons. Electron final states are the key signature of $\nu_\mu \rightarrow \nu_e$ oscillations being studied with both the DUNE and SBN programs. Another important component of this physics program are BSM signatures in the sub-GeV regime which are generally tied to “dark sector” models [56,57]. These models predict the production of new particles along the neutrino beamline. Because of the energy regime, these new particles often lead to final-state observables consisting of e^+e^- or photon electromagnetic (EM) showers. At the same time, photon EM showers are abundantly produced in standard model (SM) neutrino interactions, primarily through resonant pion production where a ν -Ar interaction excites a $\Delta(1232)$ or heavier baryon resonance [58,59] which can decay to a final-state π^0 producing two γ showers in the detector. Roughly 10% of neutrino interactions lead to a π^0 final state, far outnumbering the predicted signatures of neutrino oscillations or BSM signatures. Because of this, the physics program that these detectors hope to achieve relies critically on e - γ and e - π^0 separation.

B. Introduction to LArTPC detectors

Charged particles produced in neutrino interactions propagate through the LArTPC, ionizing and exciting the argon atoms that make up the active mass of the detector. An external electric field causes ionization electrons to drift uniformly from the bulk of the detector to the positive potential wall (anode) of the Time Projection Chamber (TPC). Wires [48,50,60], strips [61], or pixel-based [62,63] charge readout sensors measure the induced current caused by the propagating electron clouds. Charge reaches these sensors at different times, depending on the exact location where the energy was deposited. In the case of wire- or strip-based detectors, multiple overlapping planes allow for the formation of different two-dimensional (2D) views of the charge deposition, which are in turn used to reconstruct a three-dimensional (3D) image of the interaction [64,65]. Pixel-based detectors inherently provide 3D reconstructed charge deposits. In the GeV energy-scale range, different particle species give rise to different characteristic topologies. Particles such as protons, charged pions, or muons lose energy primarily through ionization, producing linear “tracklike” signatures in the detector. Electrons and photons, on the other hand, lose energy via ionization and radiation in roughly equal amounts, leading to the formation of EM showers which appear as “fuzzy” branches and e^+e^- pair cascades. Below 1 GeV, these EM showers are often highly segmented due to the stochastic nature of radiative energy loss. A more in-depth description of EM signatures in LArTPCs can be found in Refs. [66,67]. Figure 1 shows an example neutrino interaction image collected by the MicroBooNE LArTPC,

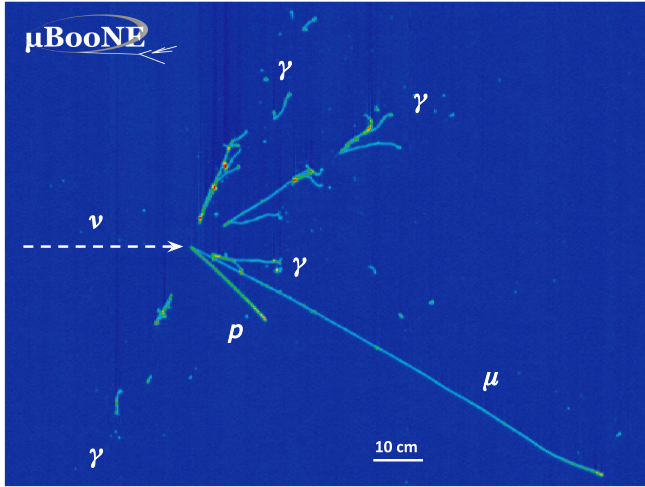


FIG. 1. Example event display of a neutrino interaction collected with one wire plane of the MicroBooNE LArTPC detector. Four photon EM showers, likely from the decay of two π^0 mesons, can be seen branching out from the interaction vertex. The x-axis represents the wire position while the y-axis indicates the time taken for the electrons to drift to the wire plane. The color scale in the event display shows the measured charge deposit.

where different final-state particles can be seen producing topologically distinct signatures.

III. DETECTOR SIMULATION AND LOW-LEVEL RECONSTRUCTION

In this section we describe general features of the simulation for accelerator-based LArTPC detectors as well as the details of the dataset used specifically in this work [68,69]. We also describe the 3D reconstruction applied to produce the inputs used in our analysis.

A. MicroBooNE Open Dataset

This work uses Monte Carlo simulations of electron and π^0 interactions obtained from the MicroBooNE Open Dataset [68,69]. These samples were released by the MicroBooNE Collaboration with the purpose of encouraging broader community participation and expanding the development of reconstruction and analysis methods for LArTPC detectors. The MicroBooNE Open Dataset consists of a sample of 10^6 simulated ν_μ and 10^5 ν_e interactions. In addition to providing a high-statistics and easily-accessible dataset, these samples are the same used by the MicroBooNE Collaboration in physics analyses, and incorporate noise and other detector effects that help provide a realistic modeling of the detector response. The results of this work therefore can be interpreted as a realistic assessment of the performance to be expected when analyzing real data in a LArTPC experiment, enhancing the robustness and value of the work presented.

B. Detector simulation and calibrations

The MicroBooNE simulation comprises three main components:

- (i) The Booster Neutrino Beamline flux [70], a ~ 1 GeV neutrino energy beamline comprised of 95% ν_μ , 4% $\bar{\nu}_\mu$, and $< 1\%$ $\nu_e/\bar{\nu}_e$.
- (ii) The simulation of neutrino interactions with the detector's argon target. This is carried out with the GENIE event generator [71] using a suite of neutrino interaction models and tailored tunes described in Ref. [72].
- (iii) The simulation of the detector response, which is carried out through custom code in the LArSoft framework [73]. This stage accounts for ionization charge propagation in the detector and signal formation on TPC wires.

The first two components of the simulation primarily influence the kinematics of particles produced in the detector from ν -Ar interactions. For the purpose of the particle classification task being performed in this work, we emphasize the characteristics of electron and π^0 events. Electron neutrinos, interacting through charged-current interactions, produce electrons with an energy spectrum which spans the 0.1–2 GeV range, peaking at approximately 0.5 GeV. Neutral pions, on the other hand, are generally produced with low momentum, peaking at 200 MeV. Figure 2 shows the true energy distribution of electrons and π^0 particles produced in neutrino interactions simulated in the MicroBooNE open datasets used for this work.

The last component of the simulation is responsible for incorporating all realistic effects that impact the detector. Effects that alter the amount of charge collected, or its propagation through the detector such as ion recombination, space charge, and diffusion are simulated based on data-driven measurements performed *in situ* or with dedicated experiments [74]. A set of calibration steps are finally applied to MicroBooNE data and simulation alike to

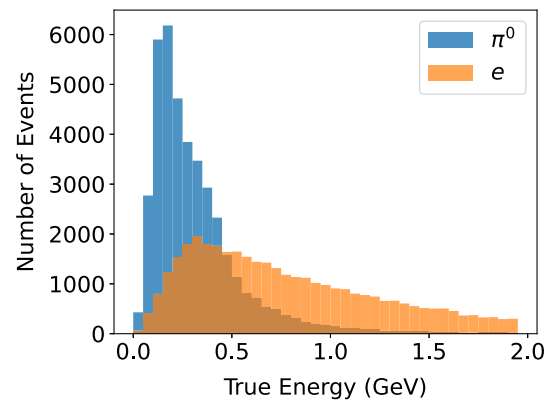


FIG. 2. True energy distribution of e and π^0 particles in the MicroBooNE open datasets used for this work.

account and correct for detector distortions. These are described in Ref. [49].

C. Low-level reconstruction

Signals recorded on TPC wires first undergo low-level reconstruction which is responsible for noise filtering [75] and signal processing [76,77]. These lead to the formation of reconstructed objects called “hits,” which represent individual pixels in images such as those of Fig. 1, and convey the position (wire number, on the horizontal axis), drift time to reach the wire (vertical axis), and charge measured on the wire (color scale). Hits are reconstructed independently on each plane and form the basis for most LArTPC reconstruction workflows.

D. Current status of LArTPC pattern recognition

After the identification of hits on each plane, different reconstruction workflows then diverge in their approach. Two commonly used reconstruction paradigms are the PANDORA multialgorithm workflow [64,78] and the WIRE-CELL toolkit [65,79]. PANDORA uses 2D hit inputs to perform tracking and 3D particle imaging. WIRE-CELL instead uses 2D inputs to reconstruct directly 3D ionization charge patterns through a tomographic approach, after which higher level pattern-recognition techniques are employed for particle track reconstruction. These reconstruction frameworks have been developed over a number of years with data from the MicroBooNE [60] and protoDUNE [48] detectors, and are now widely deployed across active LArTPC detectors including SBND and ICARUS. Finally, current ML approaches often directly use the 2D images made up by wire signals or reconstructed hits to perform high-level particle Identification (ID) and kinematics reconstruction. These ML tools often rely on convolutional neural networks (CNNs) and graphical neural networks to perform specific image-classification tasks that aim to classify features of LArTPC images at either particle-level [80–83] or pixel by pixel [84,85] or perform high-level reconstruction tasks such as energy reconstruction [86].

IV. e/π^0 PARTICLE CLASSIFICATION

As described in Sec. II, electrons and photons are key final-state observables for neutrino oscillations and BSM signatures in LArTPC experiments, but they are flooded by large backgrounds from π^0 decays to diphotons. One of the hallmarks of LArTPC detectors in neutrino experiments is their ability to distinguish electron from photon showers via two key features: (1) calorimetric dE/dx separation of e vs γ due to the rate of energy deposition at the start of the EM shower, and (2) the presence of a “gap” from the interaction vertex to the γ EM shower starting point due to the ~ 20 cm conversion distance in argon. These two features allow for powerful background rejection [87–89]. Yet, because the

vast majority of γ backgrounds originate from π^0 s, there is another, more fundamental distinguishing feature: π^0 events lead to two EM showers, one from each photon. In principle, π^0 backgrounds could be effectively mitigated by correctly identifying both photon showers. Successfully identifying both EM showers nonetheless poses several challenges. The $\pi^0 \rightarrow \gamma\gamma$ decay leads to two photons which may be collinear, and therefore hard to differentiate. In other cases, a very asymmetric decay leads to one of the photons having little energy. This is very common, in part because π^0 s produced in neutrino interactions typically have a low momentum, causing the subleading photon in the decay to often have less than 100 MeV of kinetic energy. Finally, sometimes photons from the decay escape the detector before depositing energy, which precludes identifying the second EM shower. Importantly, identifying *both* photons from a π^0 decay remains a significant challenge in LArTPC detectors, and is the primary limitation to further reducing π^0 backgrounds which currently limit the sensitivity of many LArTPC physics measurements [46,47]. In this section we will explore the current status of e/π^0 separation in LArTPC detectors, and motivate how OT can improve this.

A. Current status of LArTPC reconstruction for e/π^0

Many physics results produced with LArTPC neutrino experiments leverage the PANDORA, WIRE-CELL, or ML tools for electron-photon separation and π^0 rejection [45,47,87,89–91]. While these have demonstrated good performance, powerful π^0 rejection has come at the cost of a loss of efficiency that can negatively impact an analysis’ sensitivity. As an example, ν_e efficiencies for recent analyses in MicroBooNE hover around 10–40%, with room for improvement. Furthermore, specific topologies such as isolated single electrons with no hadronic activity [89], or more recent BSM searches for single photons [47,90,91] are still heavily dominated by π^0 backgrounds to the point that this is the primary limitation to the analysis. Novel ways of rejecting π^0 backgrounds can therefore help expand the physics reach of analyses targeting single-electron or single-photon final states. This is particularly true for methods that can overcome the reconstruction inefficiencies that often lead to not reconstructing one of the two photons from a π^0 decay.

Some of the ML tools developed aim to perform particle identification directly through a CNN, without the need to actually reconstruct the individual particles produced in an interaction (e.g., Ref. [81]). Such tools have been used to distinguish electrons from photons and other particles, but have not tackled the e/π^0 classification task. This paper explores OT as a new tool to tackle single-electron vs π^0 classification through an image-based interpretation of the data, without the need to identify and reconstruct the two photons individually. The approach therefore is novel in that it follows a philosophy similar to that of many ML

applications, which “bypass” several low-level reconstruction steps which may introduce significant inefficiencies, but does so introducing a new methodology based on OT distances, and tackling a classification problem that remains a key bottleneck in LArTPC physics analyses. Given the current lack of a ML baseline for the e/π^0 classification task, we instead compare our OT-based results to reconstruction-based strategies which are currently used in analyses [88,89,92].

V. OPTIMAL TRANSPORT DISTANCES FOR LARTPC DETECTORS

The granular, calorimeter nature of LArTPC detectors makes image analysis tools well suited for classification tasks. In turn, distinguishing different particle species by their visual topology in the detector is a task well suited for OT, which can quantify the “distance” between individual interaction images, leveraging features intrinsic to the physics of particle energy loss in a TPC. Importantly, OT can carry out this classification without the need to perform low-level reconstruction tasks which often suffer from significant inefficiencies.

A. Review of discrete OT distances

OT distances yield a mathematically rigorous notion of distance between probability distributions. Let $\mathcal{E} = \sum_{i=1}^n p_i \delta_{x_i}$ and $\mathcal{E}' = \sum_{j=1}^{n'} p'_j \delta_{x'_j}$ be two discrete probability distributions (which we will refer to as “events”) defined over the same underlying d -dimensional metric space.¹ In other words, the event, $\mathcal{E}$, is a collection of n points in d -dimensional space each weighted with probability p_i . OT distances elevate the pairwise distances between the points in $\mathcal{E}$ and $\mathcal{E}'$ ($|x_i - x'_j|$ for all possible combinations of i, j) to a distance between the events $\mathcal{E}, \mathcal{E}'$. Intuitively, this is done by calculating the “work” to transform $\mathcal{E} \rightarrow \mathcal{E}'$ by moving the individual probability point masses. Moving a single probability point mass with weight p from x to y costs an amount proportional to the distance traveled, $x-y$, and the amount of probability mass, p , moved. The total cost (work) of moving all points in $\mathcal{E} \rightarrow \mathcal{E}'$ is the sum of these individual costs. The optimal transport distance is the minimum cost incurred over all possible transport plans.

In this work we consider a particular kind of discrete, balanced OT distance, the 2-Wasserstein distance [35], which is defined as follows:

$$W_2(\mathcal{E}, \mathcal{E}') = \left[\min_{\gamma \in \Gamma(\mathcal{E}, \mathcal{E}')} \sum_{i,j} \gamma_{i,j} |x_i - x'_j|^2 \right]^{1/2} \quad (1)$$

¹Note that n need not be equal to n' . However, for balanced OT distances the total probability mass must be the same in each event, i.e., $\sum_{i=1}^n p_i = \sum_{j=1}^{n'} p'_j$.

where $\Gamma(\mathcal{E}, \mathcal{E}')$ is the set of all possible ways to transform $\mathcal{E} \rightarrow \mathcal{E}'$, so-called “transport plans,” and γ denotes a particular transport plan. In statistical terms, $\Gamma(\mathcal{E}, \mathcal{E}')$ is the set of all possible joint distributions with marginals $\mathcal{E}, \mathcal{E}'$. The γ^* which minimizes the above expression is unique and is known as the optimal transport plan.

In this work we focus on calculating discrete OT distances between events.² This is a natural choice since the energy deposits in an event can naturally be expressed as a discrete probability distribution over the detector volume. Furthermore, balanced 2-Wasserstein distances are one type of a broad array of discrete OT distances. In the next section we motivate this choice in the context of e/π^0 classification.

B. e/π^0 classification with OT

In order to apply OT to the task of e/π^0 classification, we first note that we can view the energy deposits as discrete probability distributions in 3D space. The amount of energy deposited (intensity) at a given spacepoint, up to a normalization factor, becomes the amount of probability mass. Since the shapes of these events are visually indicative of their identity, it is natural to wonder whether OT distances can help in this task.

There are many varieties of OT distances that one might think to apply. We have chosen the 2-Wasserstein distance for its nice mathematical properties, such as the uniqueness of the optimal transport plan and the ability to linearize the calculation [35]. 2-Wasserstein distance is also an example of a *balanced* OT strategy, meaning that the events must be normalized to have the same total intensity, which allows OT to focus on topological differences between events.

Electrons produced in ν_e interactions tend to have higher energies compared to the neutral pions (see Fig. 2), making the total amount of charge deposited by a shower a potentially powerful classification metric to separate signal ν_e from background π^0 interactions. While an unbalanced OT strategy [31] would be able to leverage this feature directly, for this work we aim to focus on the topological discrimination of e and π^0 events due to the spatial distribution of the energy loss, and not the total energy deposited. This could make the classification task more challenging, but allows us to focus on evaluating OT’s performance on tasks where existing methods are lacking. To this end, since balanced OT may still have indirect access to information on the total energy due to its correlation with the number of intensity points, we separate events into nine energy bins. Particle classification is carried out comparing events within each energy bin,

²We note that there has been interesting work on semidiscrete [41–44] alternatives in which events are represented as a discrete distribution but distances are calculated to a continuous manifold. The choice of manifold can help accentuate features in the events. However, in practice, this continuous manifold must be approximated by a discrete distribution.

avoiding classification dependencies on the energy of the particles.

We note that we also explored a different OT variant, Gromov-Wasserstein distances [93], as it is inherently insensitive to isometries (i.e., rotations and translations) of points in the ground space. Therefore, this strategy could serve as an alternative to 3D alignment (discussed in the next section). In a limited run on a single energy bin, Gromov-Wasserstein was shown to have similar performance to 3D alignment but did not see the same boost from postprocessing the OT distances with interpretable machine learning strategies. Due to this lower performance as well as the high computational cost of traditional Gromov-Wasserstein algorithms, we opted to use 3D alignment instead. Results of the limited run are shown and discussed further in Appendix A.

VI. DATA PREPROCESSING

Previous work [37] has shown that preprocessing data (e.g., through rotations) to minimize distributional variations corresponding to nonphysical differences is crucial for optimal transport. For example, a calculation of the optimal transport distance between a given event and that same event rotated by some angle will generally yield a nonzero result, even though they are physically the same event. Preprocessing the data to account for such physical symmetries (e.g., by aligning all tracks along the same direction) helps emphasize physically-meaningful distributional differences. This section describes the preprocessing steps implemented for this work with the aim of boosting the performance of OT-based methods. This includes how the 3D reconstructions of the events are obtained and aligned. Additionally, inspired by the success of jet images at the LHC [37], we also consider subsequent 2D projections of these 3D events along the principal axes. The removal of small, isolated energy deposits (see Appendix B) was also attempted, but was found to have no significant improvement.

A. Event filtering

Events with missing or incomplete showers will contaminate the sample and have a negative effect on separation using OT distances. For instance, a π^0 event with a missing second shower (for example when one of the photons escapes the detector) would appear to look like a single-shower event. Thus, having these π^0 events in the sample would be misleading for evaluating OT's performance. For this reason, we filter out events with showers (or large parts of the showers) which lie outside of the detector volume. This is done by removing events with more than 10 charge deposits falling within a 5 cm range from detector boundary.

B. 3D reconstruction

The three wire planes in MicroBooNE's LArTPC allow for 3D reconstructions of the events. Since this work

focuses on the supervised classification of e from π^0 shower signatures, only hits associated with true electron or photon energy deposits are used for 3D reconstruction via the SPACEPOINTSOLVER algorithm [94]. The output of such reconstruction are spacepoints with position vectors in 3D space, each associated with two or three hits from different wire planes. Examples of this are shown in Fig. 4. The intensities of these spacepoints are determined using the charge deposits of their corresponding hits in the collection plane. These 3D reconstructions are the starting point for our classification analysis with OT. While for this work we rely on truth labeling of EM showers to focus the evaluation of classification performance on the features that are most relevant to e vs π^0 classification, existing reconstruction methods would eventually be used to accomplish the task of isolating energy deposits associated with EM showers. ML methods which employ neural networks for hit-level semantic labeling can be integrated with OT for a holistic reconstruction. Several such tools showing great promise have become available, including NuGraph [9], as well as other examples developed within the DUNE [10] and MicroBooNE [84,85] Collaborations.

1. Down sampling

Since the computational efficiency of OT depends on the number of spacepoints in an event, spacepoints within 1 cm distance of each other are merged into a single spacepoint with charge equal to the sum of the merged spacepoints. This down sampling is carried out to reduce the computation time. A 1 cm merging distance is small enough to still preserve the important features that distinguish EM showers. This reduces the number of spacepoints by a factor of ~ 3 , which results in an order of magnitude improvement in computational efficiency. See Sec. VIII B for more details on the computational cost scaling of OT distances.

2. Spatial alignment

Particles produced in neutrino interactions will propagate in different directions in the detector's coordinates. While a particle's orientation in the detector may carry valuable information on the kinematics of the interaction, it is not a feature that is relevant for particle classification, and may, in fact, complicate the particle classification task we wish to achieve. As previously described, preprocessing the data to account for these differences can enhance the ability for OT distances to pick up on the more physically relevant distributional differences. Therefore, we first align the 3D distributions consistently by defining a common principal axis for each 3D reconstructed event.

To calculate the principal axis we use weighted principal component analysis (WPCA) [95]. However, a subtlety arises when evaluating WPCA in the presence of the two showers in π^0 events: for some events, using WPCA directly results in first principal components which are not representative of any of the shower axes. To circumvent

this issue we first cluster spacepoints together if they are within 2.5 cm distance of each other and compute WPCA on the cluster with the largest number of spacepoints. The first principal component is then used to define the principal axis of the event.

Next, all events are rotated to align their principal axes, hence eliminating dependencies on the direction of the particle in the LArTPC. Similarly, a second rotation is carried out to align the second principal axes, in order to align the planar distribution of the events. Events are then moved translationally such that the centers of mass are at the origin. In Fig. 4, we show the electron and π^0 events from Fig. 3 after 3D reconstruction and alignment.

After these preprocessing steps, the OT distances can be computed. Two examples are shown in Fig. 5, where a π^0 and an electron event are each transported into the same electron event shown in blue. The black lines show the visualized transport plans [37], and the corresponding OT distances are computed. We can see that the π^0 in Fig. 5(a) takes a larger OT distance (279.7 cm) to transport all of its charge deposits into the electron event compared to the

comparison between the two electrons of Fig. 5(b) (distance of 36.7 cm).

C. Projections

Past applications of OT in collider physics [31,35] have largely treated calorimeter deposits as distributions on the 2D ground space comprising the longitudinal and azimuthal plane of the calorimeters in collider detectors. This 2D representation of calorimeter deposits at colliders is physically well motivated as much of the relevant information in the underlying event is captured by the projection of outgoing particles onto the celestial sphere. In contrast, the reconstruction of neutrino interactions enabled by the 3D calorimetry in LArTPC detectors motivates treating calorimeter deposits in neutrino experiments as distributions on the full 3D ground space, as shown in Fig. 5. This presumes that the full 3D distributions contain information relevant for event classification.

However, since showers tend to leave small residual charge deposits at a large distance from their shower start, [see Fig. 4(b)] the OT transport plan may be negatively impacted by the large cost to transport this residual charge deposited far along the shower axis. These large distances are not representative of differences between an e and π^0 topology, and may lead to suboptimal performance, particularly when relying on the full 3D calorimetric information available from the detector.

Hence, inspired by event classification for jets at the LHC, we also explore using 2D projections along the principal axis component, which is analogous to the so-called jet images in colliders. The primary similarity here is in the construction of a physically relevant axis along which to take the projection such that angular deviations carry physical meaning. Figure 6 shows a comparison of OT plans for an e - π^0 pair [Fig. 6(a)] and two electrons [Fig. 6(b)] performed both in 3D (insets in each figure) as well as within the 2D charge projection (main panel). The particularly long distance covered by a small component of the electron in cyan along the vertical axis [inset of Fig. 6(b)] causes a 3D OT distance which is much larger than the 3D OT distance for the corresponding e - π^0 pair [Fig. 6(a)]. The 2D projected charge mitigates this, leading (in this specific example) to a much smaller computed 2D OT distance for the e - e pair compared to the e - π^0 pair.

VII. METHODS

Optimal transport distance computations are carried out using the Python optimal transport (POT) library [96], which provides several generic OT solvers. In order to have a balanced sample, the same number of π^0 and e events are chosen for each energy bin so we have $N_{\pi^0} = N_e$. To allow for adequate sample statistics in each bin, the bin boundaries are chosen to enforce $N_{e,p}$ to each be greater than 1000. The bin edges are defined as [0.05, 0.1, 0.15, 0.2, 0.25, 0.3,

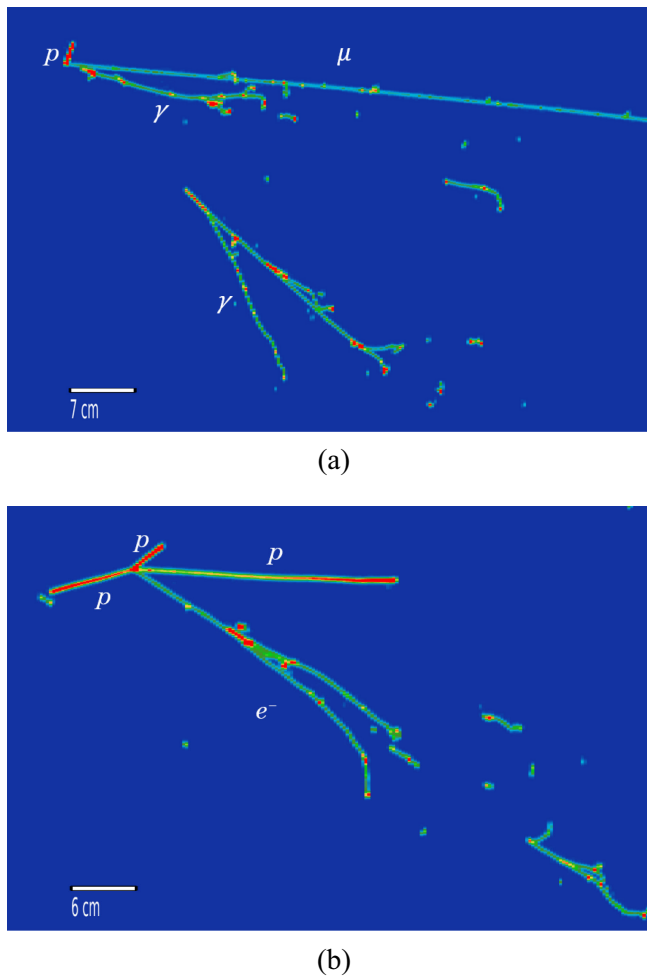


FIG. 3. Example event displays for π^0 (a) and e^- (b).

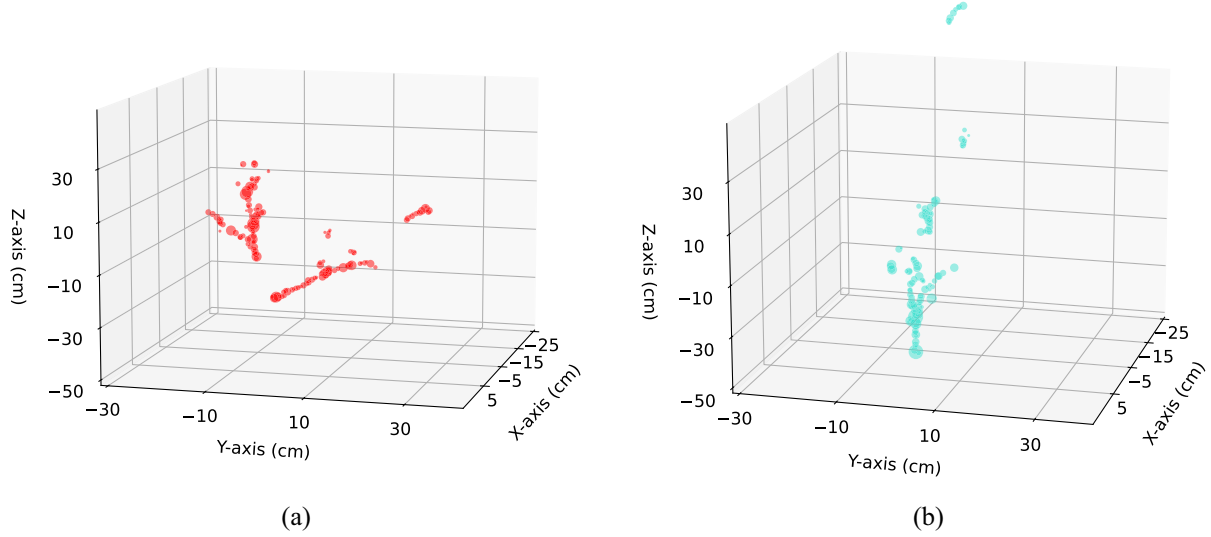


FIG. 4. Events from Fig. 3 after 3D reconstruction and alignment. The size of the dots represent the relative amount of energy deposited associated with each reconstructed 3D spacepoint. (a) Depiction of a π^0 event showing how the principal axis alignment based on the largest cluster in the event aligns the image with one of the two photon showers, keeping the second shower at an angle with respect to the $\hat{z}$ direction. (b) Depiction of an electron shower.

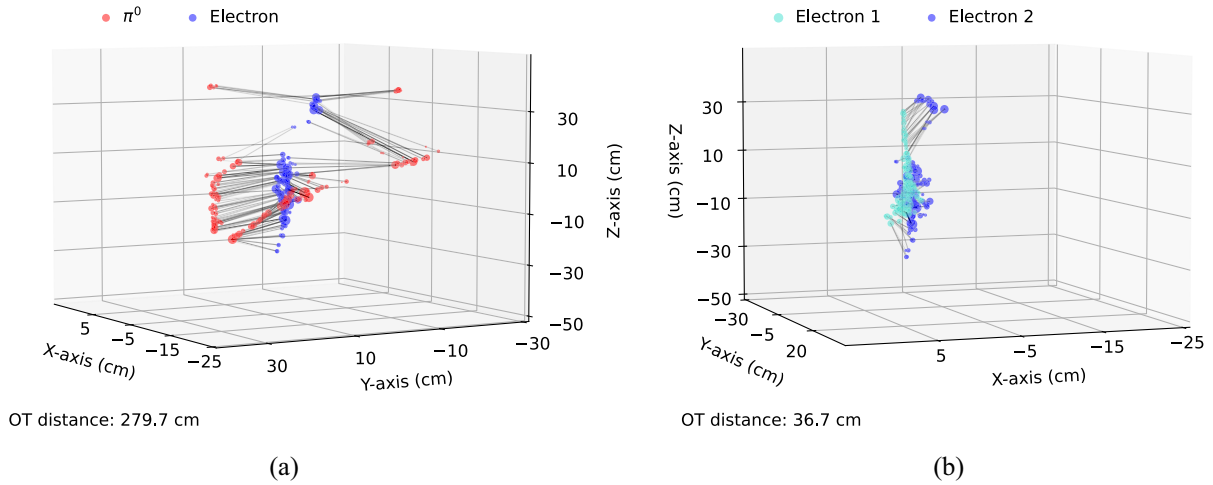


FIG. 5. Example transport plans for optimal transport distances computed between two 3D event reconstructions. (a) Transport plan between π^0 event (red) vs e^- event (blue). (b) Transport plan between the same e^- event (blue) and another e^- event (turquoise).

0.35, 0.4, 0.5, 0.9] GeV. For each bin, these $N_{\pi^0} + N_e$ events are divided into two equally sized training and testing sets, $N_{\pi^0, \text{train}} + N_{e, \text{train}}$ and $N_{\pi^0, \text{test}} + N_{e, \text{test}}$, respectively.

Given a test event, of unknown type, each OT method must assign a single number (score) which indicates the likely identity of that test event. The exact details of how the score is derived depend on the particular strategy employed. We explore “Minimum OT,” k-nearest neighbors, and support vector machine and describe each below.

A. Minimum OT: min(OT)

The first method we employ takes advantage of the fact that the OT distance between like distributions will be

smaller than the distance between unlike distributions. Since we expect physically similar events to have similar detector signatures, we can use this property to our advantage. In this case, the OT distance between a π^0 event and electron event will generally be larger than the distance between two electron events.

In the min(OT) method, the strategy employed to identify π^0 events is to first designate a number of electron events from the training set as a reference, $\{\mathcal{E}_{e_i} | i = 1, \dots, N_{e, \text{ref}}\}$. We can then compare an event of an unknown type, $\mathcal{T}$, to this reference population. Namely, for each event $\mathcal{T}$ we can calculate $N_{e, \text{ref}}$ OT distances $\{\text{OT}(\mathcal{T}, \mathcal{E}_{e_i}) | i = 1, \dots, N_{e, \text{ref}}\}$. Electron events, $\mathcal{T}_e$, will tend to have a smaller OT distance than π^0 events, $\mathcal{T}_{\pi^0}$, i.e.,

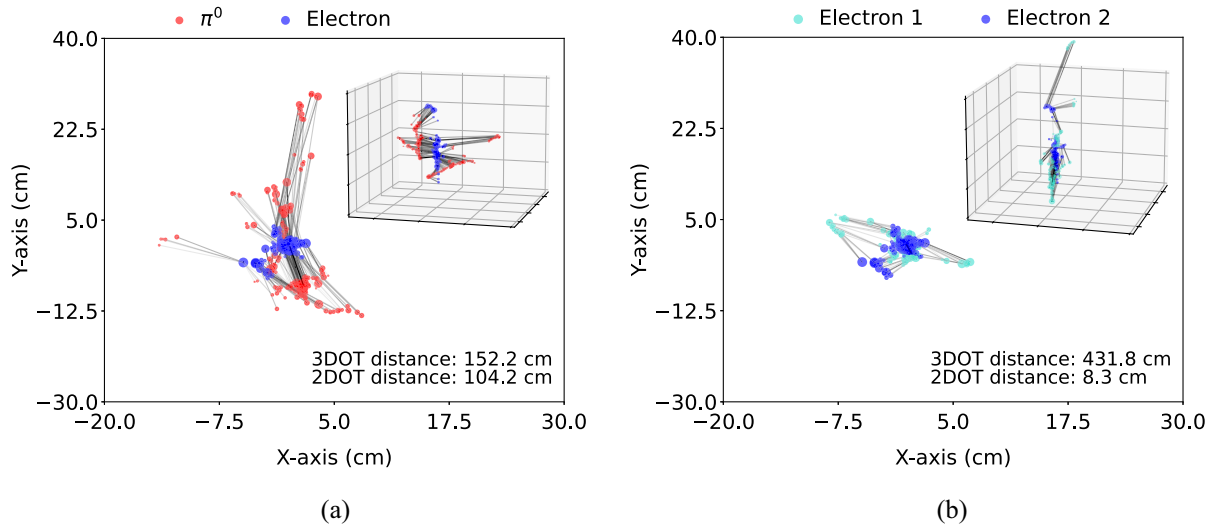


FIG. 6. Transport plans for optimal transport computed between 2D projections (main figure) and 3D reconstructions (inset) of the same events. (a) Transport plan between a π^0 event (red) and an e^- event (blue). (b) Transport plan between the same e^- event (blue) and another e^- event (turquoise).

the set $\{\text{OT}(\mathcal{T}_e, \mathcal{E}_{e_i})\}$ will generally be closer to 0 than $\{\text{OT}(\mathcal{T}_{\pi^0}, \mathcal{E}_{e_i})\}$. The task now is to turn $\{\text{OT}(\mathcal{T}, \mathcal{E}_{e_i}) | i = 1, \dots, N_{e,\text{ref}}\}$ into a single score for each test event $\mathcal{T}$.

We choose to take the minimum of these distances as our score for that test event. Namely, the score for an event $\mathcal{T}$ is the OT distance to the closest electron event in the reference sample. This approach follows from previous work on LHC data [24] which explored several methods of reduction. We also note that there have been several works in the context of particle colliders which have argued that the notion of smallest OT distance is a valid discriminating observable [24,39,40]. Reference [40] even found that in some cases traditional physics observables correspond to the closest OT distance to a manifold of events.

Figure 7 shows the distribution of $\min(\text{OT})$ scores calculated for the electron and π^0 test samples in the 0.2–0.25 GeV energy bin. As expected, the OT distance between like events (electron-electron) tends to be smaller than the OT distance between unlike events (π^0 -electron). Two different thresholds for $\min(\text{OT})$ values are shown representing cuts which give the highest accuracy (red) or a representative 80% π^0 rejection (gold).

We reserve half of the electron events to act as a reference sample and the other half are reserved for testing; the same number of π^0 events similarly are selected for testing. To get uncertainty estimates on our results, we further split our testing sample containing both π^0 s and electrons into 5 randomly shuffled sets each of size N_{test} . The $\min(\text{OT})$ distance between each test event and the $N_{e,\text{ref}}$ electrons is computed. We can then obtain receiver operating characteristic (ROC) curves and resulting metrics (e.g., accuracy) from various threshold cuts on $\min(\text{OT})$ scores. We repeat this process for each of the five test sets and report the average and standard deviations of the results.

B. Postprocessing pairwise OT distances with machine learning

Alternatively, interpretable ML methods can be used on pairwise OT distances to further boost the classification performance. The two methods we consider are k-nearest neighbors (kNN) and SVM. These methods overcome the fact that while two distributions may not be perfectly separated, OT distances for like events will tend to cluster together, providing additional information that can be used for event classification. These methods have also been shown to boost performance in certain contexts [24,35].

In contrast, the $\min(\text{OT})$ method relies on the distribution of $\min\{\text{OT}(\mathcal{T}_{\pi^0}, \mathcal{E}_{e_1}), \dots, \text{OT}(\mathcal{T}_{\pi^0}, \mathcal{E}_{e_{N_{e,\text{ref}}}})\}$

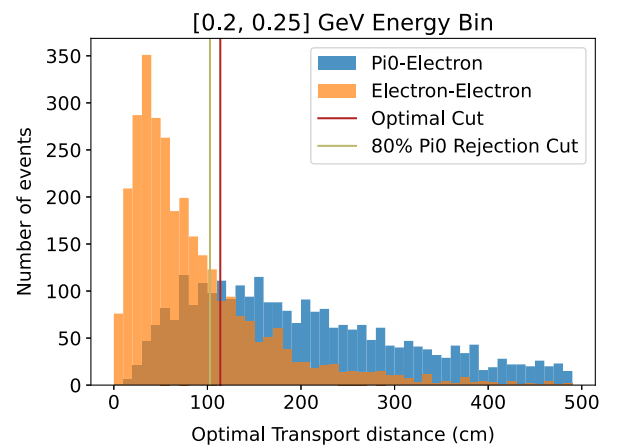


FIG. 7. Histogram of the minimum OT distance between π^0 and e events (blue) and between different electron events (orange). Two different cut thresholds for $\min(\text{OT})$ values are shown; one gives the highest accuracy (red) and the other achieves 80% π^0 rejection (yellow).

being sufficiently separated from the distribution of $\min\{\text{OT}(\mathcal{T}_e, \mathcal{E}_{e_1}), \dots, \text{OT}(\mathcal{T}_e, \mathcal{E}_{e_{N_{e,\text{ref}}}})\}$. While this allows you to operate in an unsupervised manner, you risk sacrificing additional separation power that supervised postprocessing (kNN and SVM) can yield. However, these methods now require both electron and π^0 events for training. Namely, the inputs to these models are the full set of pairwise OT distances, $\{\text{OT}(\mathcal{E}_{\pi^0}, \mathcal{E}_e)\}$, for all e, π^0 training events.

1. *K-nearest neighbors*

kNN [97] is a supervised algorithm which can be used for classification tasks, where the hyperparameter k is the number of nearest neighbors. An n -dimensional OT distance space is constructed, where n is the number of events in the training sample. kNN makes a prediction for a test event by finding the k events in the training sample which are closest to the test event we are trying to classify. Each neighbor casts a vote corresponding to its own type; the majority type is then assigned to the test event and a percentage given by the fraction of neighbors of that type is assigned as a measure of the probability that the test event is of that type. For example, if all k neighbors are electrons, the test event will be assigned the particle type “electron” with 100% probability.

To estimate the best hyperparameter values, we implement fivefold cross-validation on the randomly mixed training dataset $N = N_{\pi^0, \text{train}} + N_{e, \text{train}} = 500$. We first fix a hyperparameter value k . We split N into five separate sets, N^i for $i = 1, \dots, 5$. We select one of these to be our validation set and the remaining four to be our training, e.g., $N_{\text{val}} = N^0 = 100$ and $n = \sum_{j \neq i} N^j = 400$. We store the results of the kNN, then repeat the process until we have cycled through all possible choices of $N_{\text{val}} = N^i$. We then average the resulting five sets of stored kNN results. We repeat this process for a range of hyperparameter values k , and the best performing hyperparameter is selected. Initial scans of the hyperparameter k shows best performance around $k = 9$, which is why for all energy bins we scan over the hyperparameter space in the range $k \in [7, 11]$ with an increment size of 1. Similar to what was done with min(OT), the final results are obtained by splitting the $N_{\pi^0, \text{test}} + N_{e, \text{test}}$ into five sets of size 100. The model with the best hyperparameter k is retrained using the full $N_{\pi^0, \text{train}} + N_{e, \text{train}}$ set of events and is then evaluated on the five test sets; the average and standard deviation of the results are reported.

2. *Support vector machine*

Similarly, a SVM [98] can also be used as a supervised classifier. Events are again distributed in a hyperspace constructed from OT distances. However, instead of majority vote by the neighbors, a hyperplane (called the decision boundary) which most effectively separates the two classes is

found. A prediction for the class of an event is made based on which side of the hyperplane an event lies.

There are two hyperparameters in SVM:

- (i) C , a regularization parameter controlling the margin size around the decision boundary.
- (ii) γ , which determines the effect of a training event on the decision boundary.

In particular, we choose a Gaussian Radial Basis Function (RBF) kernel in which γ can be interpreted as the inverse variance of the kernel. A larger value of γ means training events which are far away from the decision boundary in OT-distance space are weighted less when constructing the decision boundary. In the case of using OT distances for e/π^0 classification, we found that the value of C had little effect on the results. As for γ , we first performed a coarse scan to determine the best γ range of [0.002, 0.014]. We then carried out a finer scan with an increment of 0.002 to find the optimal choice of γ . As in Sec. VII B 1, fivefold cross validation is performed to select the best hyperparameters, and the average and standard deviation of the results over five test sets are reported.

VIII. RESULTS

In this section, we show the performance of the methods [min(OT), OT + kNN, and OT + SVM] described in Sec. VII. Table I shows the performance of these three methods in each of the nine energy bins. The performance metrics we consider are the area under the curve (AUC), accuracy of overall classification, and e efficiency at 80% π^0 rejection. The ROC curves for the nine energy bins are shown in Fig. 8.

Overall, the best performing method has an accuracy ranging from 0.828 to 0.876 for events above 100 MeV. Performance is shown to be relatively stable across all energy bins, with a slight drop at higher energies. The first energy bin at 0.05–0.1 GeV is also shown to have lower relative performance. This is to be expected since this energy range is below the rest mass of a π^0 , meaning the π^0 events in this energy bin are missing a significant portion of their energy due to effects such as inefficiencies in the 3D reconstruction and a fraction of the charge escaping the detector boundary. Overall, all methods perform well, however postprocessing with both kNN and SVM boosts performance. SVM in particular makes the greatest improvement. In line with previous works [31,35], interpretable ML methods are shown to make good use of the features of OT distances and yield more competitive performance. Using 2D projections as input for OT (abbreviated as 2DOT) is shown to perform better than 3DOT when using min(OT) as the postprocessing method, which makes sense as min(OT) may be particularly sensitive to cases such as those discussed in Fig. 6. However, kNN and SVM are able to overcome this difference. In particular, 3DOT + SVM has the best performance across all energy bins. As we will see in

TABLE I. Optimal transport performance for min(OT), kNN and SVM combined with 2DOT and 3DOT in 9 energy bins. Performance metrics presented here include: area under the curve (AUC), accuracy of overall classification and e efficiency at 80% π^0 rejection.

E Bin (GeV)	Performance Metric	2DOT			3DOT		
		min(OT)	kNN	SVM	min(OT)	kNN	SVM
0.05–0.1	AUC	0.786 ± 0.017	0.761 ± 0.013	0.783 ± 0.016	0.770 ± 0.022	0.785 ± 0.007	0.822 ± 0.007
	Accuracy	0.738 ± 0.022	0.699 ± 0.016	0.717 ± 0.019	0.717 ± 0.020	0.694 ± 0.008	0.742 ± 0.012
	e^- efficiency	0.649 ± 0.052	0.574 ± 0.044	0.612 ± 0.026	0.611 ± 0.058	0.624 ± 0.023	0.689 ± 0.022
0.1–0.15	AUC	0.826 ± 0.021	0.852 ± 0.015	0.852 ± 0.015	0.816 ± 0.024	0.867 ± 0.007	0.913 ± 0.009
	Accuracy	0.765 ± 0.022	0.780 ± 0.023	0.791 ± 0.018	0.755 ± 0.015	0.784 ± 0.009	0.828 ± 0.024
	e^- efficiency	0.712 ± 0.047	0.754 ± 0.044	0.780 ± 0.043	0.693 ± 0.034	0.788 ± 0.032	0.848 ± 0.036
0.15–0.2	AUC	0.836 ± 0.007	0.850 ± 0.012	0.856 ± 0.008	0.820 ± 0.008	0.887 ± 0.016	0.930 ± 0.015
	Accuracy	0.765 ± 0.009	0.774 ± 0.019	0.792 ± 0.007	0.754 ± 0.009	0.803 ± 0.019	0.858 ± 0.012
	e^- efficiency	0.718 ± 0.018	0.752 ± 0.025	0.789 ± 0.013	0.688 ± 0.006	0.837 ± 0.023	0.910 ± 0.025
0.2–0.25	AUC	0.839 ± 0.007	0.856 ± 0.018	0.860 ± 0.018	0.824 ± 0.009	0.874 ± 0.009	0.938 ± 0.010
	Accuracy	0.767 ± 0.006	0.782 ± 0.017	0.788 ± 0.014	0.751 ± 0.007	0.788 ± 0.021	0.876 ± 0.008
	e^- efficiency	0.720 ± 0.015	0.755 ± 0.050	0.777 ± 0.034	0.683 ± 0.014	0.811 ± 0.027	0.917 ± 0.015
0.25–0.3	AUC	0.847 ± 0.013	0.850 ± 0.014	0.852 ± 0.015	0.829 ± 0.019	0.865 ± 0.014	0.926 ± 0.010
	Accuracy	0.774 ± 0.013	0.776 ± 0.014	0.784 ± 0.012	0.765 ± 0.013	0.768 ± 0.012	0.853 ± 0.010
	e^- efficiency	0.734 ± 0.026	0.768 ± 0.024	0.760 ± 0.018	0.714 ± 0.033	0.787 ± 0.042	0.900 ± 0.020
0.3–0.35	AUC	0.853 ± 0.022	0.862 ± 0.020	0.879 ± 0.030	0.833 ± 0.019	0.860 ± 0.013	0.940 ± 0.012
	Accuracy	0.786 ± 0.023	0.797 ± 0.028	0.815 ± 0.013	0.762 ± 0.010	0.778 ± 0.016	0.873 ± 0.018
	e^- efficiency	0.749 ± 0.063	0.783 ± 0.053	0.828 ± 0.024	0.703 ± 0.025	0.747 ± 0.049	0.925 ± 0.029
0.35–0.4	AUC	0.818 ± 0.023	0.834 ± 0.035	0.848 ± 0.040	0.788 ± 0.034	0.831 ± 0.035	0.916 ± 0.023
	Accuracy	0.758 ± 0.024	0.759 ± 0.027	0.765 ± 0.035	0.732 ± 0.027	0.710 ± 0.022	0.837 ± 0.032
	e^- efficiency	0.686 ± 0.047	0.705 ± 0.066	0.751 ± 0.080	0.628 ± 0.063	0.685 ± 0.079	0.881 ± 0.047
0.4–0.5	AUC	0.822 ± 0.016	0.851 ± 0.015	0.857 ± 0.014	0.761 ± 0.035	0.845 ± 0.009	0.927 ± 0.006
	Accuracy	0.756 ± 0.014	0.777 ± 0.022	0.787 ± 0.011	0.711 ± 0.028	0.741 ± 0.013	0.857 ± 0.012
	e^- efficiency	0.688 ± 0.026	0.742 ± 0.025	0.760 ± 0.027	0.57 ± 0.068	0.708 ± 0.036	0.907 ± 0.020
0.5–0.9	AUC	0.783 ± 0.033	0.822 ± 0.015	0.833 ± 0.022	0.737 ± 0.034	0.778 ± 0.015	0.891 ± 0.014
	Accuracy	0.733 ± 0.037	0.751 ± 0.021	0.763 ± 0.026	0.689 ± 0.031	0.674 ± 0.029	0.828 ± 0.016
	e^- efficiency	0.646 ± 0.073	0.694 ± 0.044	0.738 ± 0.044	0.553 ± 0.044	0.571 ± 0.045	0.854 ± 0.033

Sec. VIII A, all three methods with both 2D and 3D OT outperform traditional reconstruction in e/π^0 classification.

To contextualize these results, it is important to note that in a full analysis where only one photon can be clearly identified, other handles for e/γ separation would be subsequently employed to remove remaining backgrounds. Handles such as calorimetric dE/dx , or the presence of a gap allow for additional discriminating power. Importantly, however, the more e/π^0 classification can be improved by correctly identifying the characteristic two-shower vs one-shower topology of the event early in the classification, the less one needs to rely on subsequent cuts such as dE/dx , the presence of a gap, or other kinematic-based discriminating variables. These subsequent cuts may introduce several drawbacks including larger inefficiencies and potential kinematic and topological biases. Boosting e/π^0 separation through the topological discrimination of the full energy deposit as OT is able to do can therefore lead to meaningful performance gains when applied to physics measurements in LArTPCs.

A. Performance compared with PANDORA

In order to benchmark results with current state-of-the-art tools, performance of OT is compared to the PANDORA reconstruction framework [64,78] as implemented in the reconstruction available in MicroBooNE's Open Datasets. PANDORA is chosen because it is currently the most widely used reconstruction framework across neutrino LArTPC experiments, and because its reconstruction output is automatically stored in the MicroBooNE open dataset files used for this work. In comparing the performance of OT to a more traditional reconstruction approach, we use the number of reconstructed EM showers in an event as a way to classify electrons (one shower) vs π^0 (two showers). We find that overall OT outperforms PANDORA in all phase space for certain kinematic variables. In this comparison it is important to keep in mind that the preprocessing employed for this work (Sec. VI) relies on truth labeling of EM shower activity which is not available from the already stored PANDORA output. The impact this has in the performance comparisons presented was tested by studying

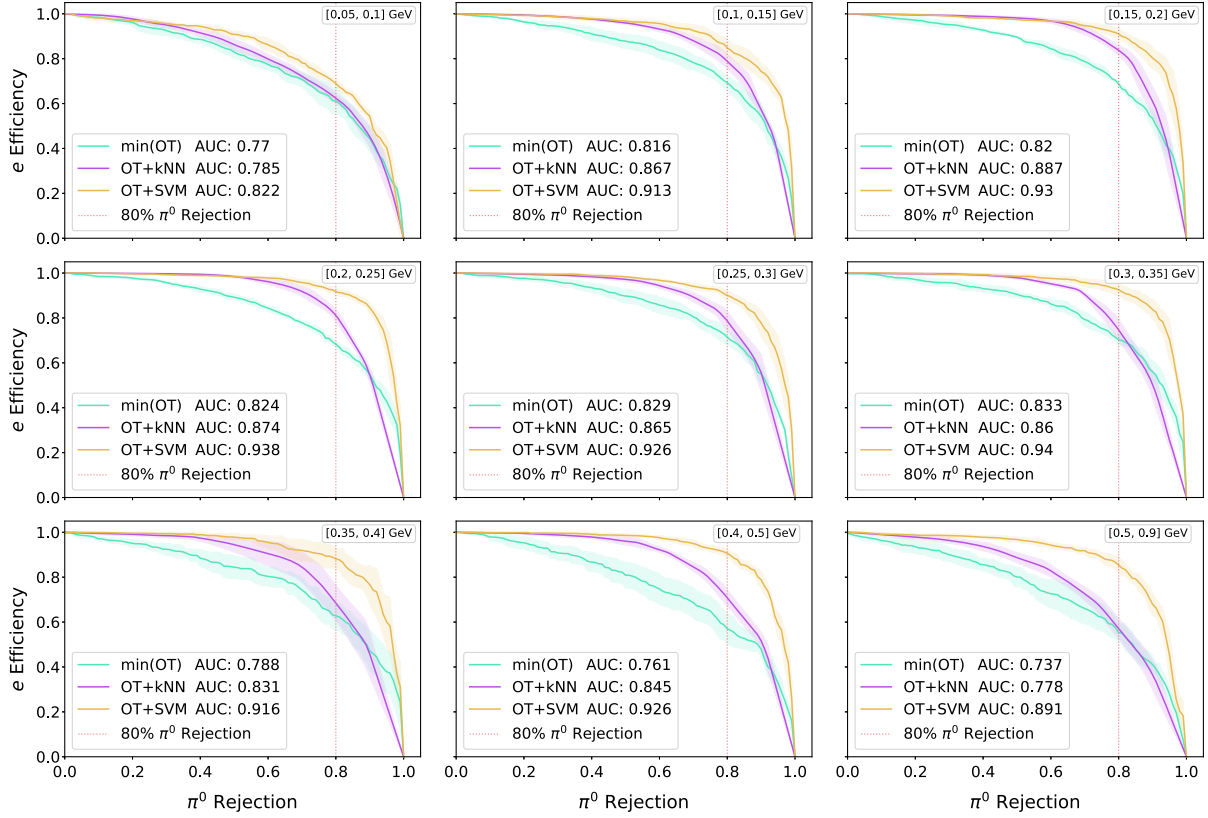


FIG. 8. ROC curves for 3D OT methods between events in each energy bin (GeV).

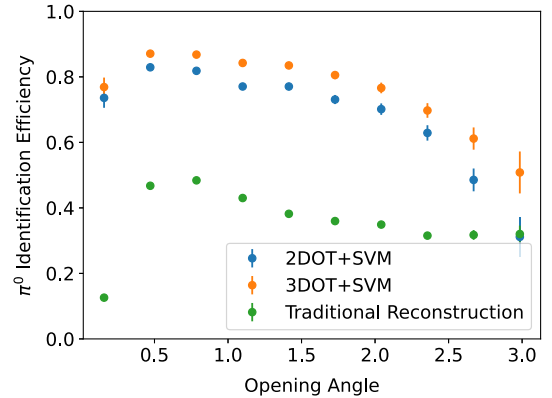
based on truth information on how classification results change as the quality of the reconstruction increases. We find that overall this does not meaningfully alter the conclusions of this comparison. As mentioned previously, future integration of OT in event classification would be paired with existing complementary reconstruction packages that can perform hit-level track-shower classification with high accuracy, such as NuGRAPH2 [9]. This would remove the reliance on truth-based information leveraged for this study.

1. π^0 classification

To successfully classify an event as a π^0 in PANDORA, we require that two or more reconstructed showers are present in the event. For OT, we rely on the methods described earlier. The reason we are requiring two or more showers instead of exactly two showers is to aim for fairly evaluating PANDORA's performance at distinguishing electrons from π^0 s without demanding stricter reconstruction accuracy for the individual photon showers in the event.

The results are presented as a function of the relevant kinematic variables for $\pi^0 \rightarrow \gamma\gamma$ decays. The first variable studied is the momentum ratio of the two photons, defined as $R = E_{\gamma 2}/E_{\gamma 1}$ where $E_{\gamma 1}$ and $E_{\gamma 2}$ are the energy of the leading and subleading photons, respectively. Events with low momentum ratio indicate a very low-energy, subleading photon, which is typically challenging to identify.

The second variable used is the opening angle between the photons produced by the π^0 decay. Photon showers which are closely aligned with a small opening angle are another typically challenging topology to identify and reconstruct. The results are shown in Figs. 9 and 10, and compare PANDORA's performance (green) to that of OT using 3D (orange) and 2D (blue) projections with SVM postprocessing. Both 2DOT and 3DOT have more competitive performance in π^0 classification. As compared to

FIG. 9. Performance compared with opening angle. For PANDORA, π^0 efficiency is defined as its efficiency in finding ≥ 2 showers. Error bars represent statistical uncertainties of the samples in each bin.

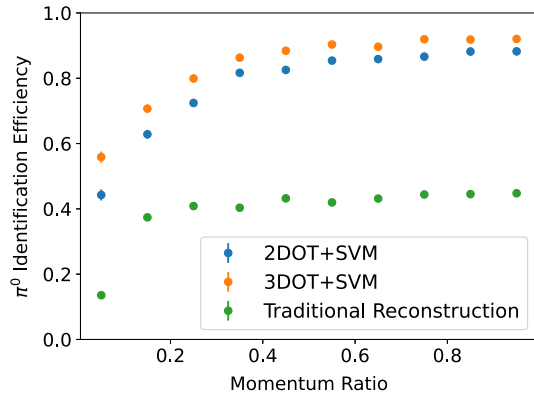


FIG. 10. Performance compared with subleading shower momentum ratio. For PANDORA, π^0 efficiency is defined as its efficiency in finding ≥ 2 showers. Error bars represent statistical uncertainties of the samples in each bin.

PANDORA which has an overall efficiency of 40%, OT is able to correctly classify 80% of the π^0 s in general. The improvement is particularly pronounced for events with small opening angles, where PANDORA is only able to correctly classify $\sim 10\%$ of π^0 s (see Fig. 9). The three different examples of π^0 events with small opening angles in Fig. 11 show why this could be the case. In Figs. 11(a) and 11(b) the electron showers are spatially located between the two γ showers. In Fig. 11(c), the electron shower overlaps with one photon shower while being at a distance away from the other one. In all three cases, we see that much work was done to overcome the planar distributional differences between one-shower and two-shower structures, despite the small opening angles for the π^0 events, leading to good event classification even in this challenging region of phase space. When evaluating the performance as a function of the momentum ratio (Fig. 10), we similarly notice a particularly large performance improvement at low values of momentum ratio, improving the classification efficiency from around 15% to almost 60% for the lowest bin.

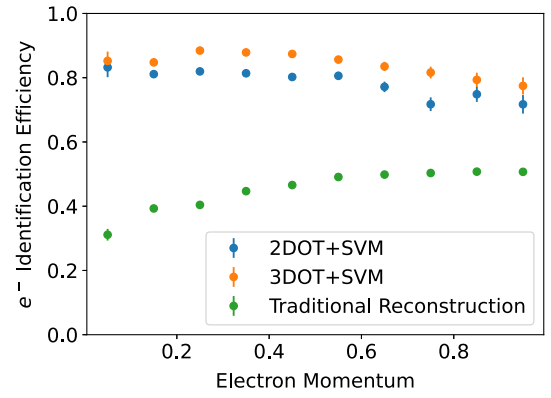


FIG. 12. Performance as a function of electron momentum. For PANDORA, e^- efficiency is defined as its efficiency in reconstructing one shower in the event.

2. Electron classification

We additionally evaluate the performance as a function of the electron momentum, as shown in Fig. 12. Here too, we see that OT outperforms traditional reconstruction across all values of electron momentum, with a relatively stable performance around 80%, as opposed to PANDORA which has an overall efficiency of about 40%.

B. Computational scaling of OT calculation

In this work, we have used a standard strategy to calculate OT distances [99] which is implemented in the POT library [96]. The average CPU time for a single OT distance computation for each energy bin is shown in Table II. By single OT distance computation we mean calculating the OT distance between two events, $OT(\mathcal{E}, \mathcal{E}')$. However, more efficient OT distance strategies exist. In this section, we briefly review the computational time complexity of OT distance calculations in general and in the context of this work. Our hope in doing so is to motivate the wide array of OT approximation methods that could be utilized to deploy these methods at scale in LArTPC experiments.

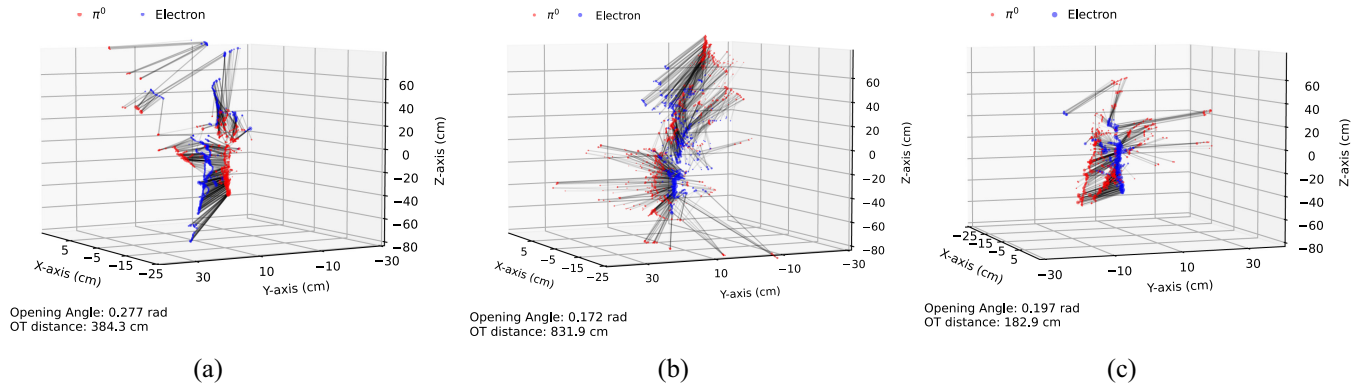


FIG. 11. Transport plans between π^0 with small opening angle and electrons in the same energy bin. (a), (b), and (c) show examples of transport plans between π^0 with small opening angles and electrons in the same energy bin.

TABLE II. CPU time for a single OT distance computation in ms. These times were averaged over 1,000 computations on a standard laptop. The average and standard deviation of the time is reported.

E Bins (GeV)	2DOT(ms)	3DOT(ms)
0.05–0.1	0.364 ± 0.015	0.382 ± 0.024
0.1–0.15	0.629 ± 0.012	0.624 ± 0.012
0.15–0.2	0.986 ± 0.033	0.963 ± 0.028
0.2–0.25	1.46 ± 0.02	1.62 ± 0.48
0.25–0.3	2.18 ± 0.37	2.22 ± 0.69
0.3–0.35	2.32 ± 0.51	2.58 ± 0.052
0.35–0.4	3.26 ± 0.75	3.68 ± 0.35
0.4–0.5	8.24 ± 0.56	8.78 ± 0.88
0.5–0.9	29.7 ± 1.9	27.8 ± 3.3

The time complexity of an OT distance calculation is dominated by the dependence on the number of discrete probability masses, n . In the most pessimistic case, the computational complexity of OT distance calculations grows as $\mathcal{O}(n^3 \log n)$ [100]. However, the algorithm implemented in Ref. [99] lowers this to $\mathcal{O}(n^3)$. Additionally, many approximate algorithms exist which are even more efficient. For example, Sinkhorn distances [100] scale as $\mathcal{O}(n^2)^3$ with a sample complexity of $\mathcal{O}(n^{-1/2})$ [102] and sliced Wasserstein distances scale as $\mathcal{O}(n \log n)$ also with a sample complexity of $\mathcal{O}(n^{-1/2})$ [103]. Interestingly, neither of these strategies suffer from the curse of dimensionality and thus still work efficiently when the dimension d of the underlying metric space is large.

In addition to the time complexity due to n , we must also consider how many OT distance calculations are required. For a single test event and N_{ref} reference events, we must calculate N_{ref} OT distances. Thus, the overall complexity per test event is at worst $\mathcal{O}(N_{\text{ref}} n^3 \log n)$ and at best near linear in n (i.e. $\approx \mathcal{O}(N_{\text{ref}} n)$). In this work, our complexity is $\mathcal{O}(N_{\text{ref}} n^3)$ per test event [99].

This computational time cost can be converted into a memory cost by using linearized OT (LOT) distances [35]. LOT takes advantage of the pseudo-Riemannian structure of 2-Wasserstein distances to simplify the calculation of OT distances. This is done by first defining the 2-Wasserstein tangent plane at a reference event, $\mathcal{R}$, living on the space of possible events. An event $\mathcal{E}$ can then be projected onto this plane by solving the OT problem between $\mathcal{R}$ and $\mathcal{E}$ to obtain the transport plan, $\gamma_{\mathcal{R}\mathcal{E}}$. After the events are embedded, simpler L_2 distances can be calculated between events on this tangent plane which faithfully approximate the 2-Wasserstein distance between events. Practically, $\mathcal{R}$ is chosen to be an unphysical reference event which adequately spans the ground space (i.e., a grid of n point masses in the d -dimensional space). The OT distances between $\mathcal{R}$ and

each of the N_{ref} physical reference events are precomputed and the $n \times n$ transport plans are stored, $\gamma_{\mathcal{R}\mathcal{E}_1}, \dots, \gamma_{\mathcal{R}\mathcal{E}_{N_{\text{ref}}}}$. For a given test event, $\mathcal{T}$, you then calculate one OT distance between $\mathcal{T}$ and $\mathcal{R}$ and store this transport plan, $\gamma_{\mathcal{R}\mathcal{T}}$. The last step is to calculate the Euclidean distance between $\gamma_{\mathcal{R}\mathcal{T}}$ and each of the N_{ref} transport plans, $\gamma_{\mathcal{R}\mathcal{E}_1}, \dots, \gamma_{\mathcal{R}\mathcal{E}_{N_{\text{ref}}}}$. Thus, the time to evaluate each test event decreases from $\mathcal{O}(N_{\text{ref}} n^3)$ to $\mathcal{O}(n^3 + N_{\text{ref}} n)$. However the memory cost is increased since the transport plans must be stored [$\mathcal{O}(N_{\text{ref}} n^2)$] rather than the point-mass coordinates themselves [$\mathcal{O}(nd)$]. Note that since LOT requires the transport plans themselves, distance-only approximation strategies like sliced Wasserstein distances cannot be used.

Additionally, under the min(OT) strategy, we must also find the smallest of the N_{ref} OT distances for each test event; with standard algorithms this has linear time complexity. Whereas, under the OT + ML strategies, this extra time cost is replaced by the time cost of the model (kNN or SVM) making its prediction.

IX. CONCLUSION

OT has been shown to perform well on event classification tasks for collider physics data [24,31,35], which motivated this study in which we performed the first implementation of OT on neutrino LArTPC data. In this work, we have applied OT to e/π^0 separation, a challenging pattern recognition task. We developed preprocessing techniques crucial for OT to be effective and tested 2D and 3D OT, both on their own and as inputs for interpretable ML methods, including kNN and SVM. MicroBooNE Open Datasets [68,69] were used, which incorporate a realistic simulation of the MicroBooNE detector providing a robust framework to test performance. All methods significantly outperform the traditional reconstruction algorithm PANDORA, and in particular combining 3DOT with SVM postprocessing is shown to have the best performance out of all methods considered, with accuracy ranging from 74.2–87.6% and AUC ranging from 82.2–94.0% depending on the energy bin. The fact that OT significantly outperforms traditional reconstruction methods in this task can be explained by its ability to quantify features that distinguish single-shower (e) vs two-shower (π^0) events by utilizing the calorimetric charge distribution in the event. This bypasses high-level EM shower reconstruction algorithms and hence avoids many bottlenecks and inefficiencies which have, to date, limited the performance of traditional methods.

There are a variety of future directions motivated by this work. The success of OT methods applied to e/π^0 particle classification suggests applying OT to other reconstruction problems in LArTPC experiments such as e/γ separation. The computational complexity of standard OT distance calculations warrants exploring strategies such as LOT [35] for better computational efficiency. Finally, it would be worthwhile to explore the prospects for integrating OT into the reconstruction pipeline of neutrino LArTPC experiments

³Although more recent variants claim near linear performance [101].

comprising the SBN program and DUNE using a variety of charge readout methods. While this work has shown promising performance using wire-based TPC images, comparable performance is expected for strip-based detectors and OT may be even more suited to pixel-based readout thanks to its intrinsic 3D readout.

ACKNOWLEDGMENTS

We acknowledge the MicroBooNE Collaboration for making publicly available the data sets [68,69] employed in this work. These data sets consist of simulated neutrino interactions from the Booster Neutrino Beamline overlaid on top of cosmic data collected with the MicroBooNE detector [2017 JINST 12 P02017]. We would like to thank Giuseppe Cerati for helping interpret the MicroBooNE public datasets and Pedro Machado for suggesting 2D projections. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics under Awards No. DE-SC-0023730 and No. DE-SC-0011702. Parts of this work were performed at the Kavli Institute for Theoretical Physics, supported by the National Science Foundation under Grants No. NSF PHY-1748958 and NSF PHY-2309135. J. N. H. was supported by the National Science Foundation under Grant No. NSF PHY-1748958, the Gordon and Betty Moore Foundation through Grant No. GBMF7392, and by a UCSB Chancellor's postdoctoral fellowship.

DATA AVAILABILITY

The data that support the findings of this article are openly available [68,69].

APPENDIX A: GROMOV-WASSERSTEIN LIMITED RUN

Gromov-Wasserstein distances [93] were initially developed to calculate OT distances between distributions over different underlying metric spaces; for example, the OT distance between a 2D and a 3D distribution. To achieve this it relies only on relative distance information within each metric space. Concretely, let $\mu = \sum_i a_i \delta_{x_i}$ and $\nu = \sum_j b_j \delta_{y_j}$ be two discrete probability distributions over the ground spaces X and Y , respectively. Let $D_{i,k} := d_X(x_i, x_k)$ where $d_X(\cdot, \cdot)$ is the Euclidean distance between points x_i, x_k in the ground space X . Similarly, we have $D_{j,l} := d_Y(y_j, y_l)$. Then the p -Gromov-Wasserstein distance is given by

$$GW_p(\mu, \nu) = \left(\min_{\gamma \in \Pi(\mu, \nu)} \sum_{i,j,k,l} |D_{i,k} - D_{j,l}|^p \gamma_{i,j} \gamma_{k,l} \right)^{\frac{1}{p}}, \quad (\text{A1})$$

where $\Pi(\mu, \nu)$ denotes the set of all distributions over the product space $X \times Y$ with marginals μ, ν (i.e., the set of all transport plans). Since Gromov-Wasserstein distances only

rely on relative information, $D_{i,k}, D_{j,l}$, they are manifestly invariant to isometric transformations (i.e., rotations and translations) of points in the ground spaces.

This useful feature comes at a price. Unfortunately, direct computation of Gromov-Wasserstein distances is a nonconvex quadratic optimization problem, making it Nondeterministic Polynomial time (NP) hard [104]. While more efficient approximation strategies exist [105,106], Gromov-Wasserstein distances are still far more computationally expensive than classic Wasserstein distances. This can be prohibitive for discrete distributions with many point masses.

Our hypothesis was that Gromov-Wasserstein distances might be particularly useful in this setting because they would obfuscate the need to align our 3D events in preprocessing. This possibility, and the fact that these could be run in off-line analyses, motivated us to explore this as an alternative. We performed a limited run to see if the performance of Gromov-Wasserstein distances would out compete our 3D alignment of events. We used the POT [96] implementation of the entropically regularized algorithm described in Ref. [105].

We found that the min(OT) methods for both Gromov-Wasserstein and 3D balanced OT after alignment performed

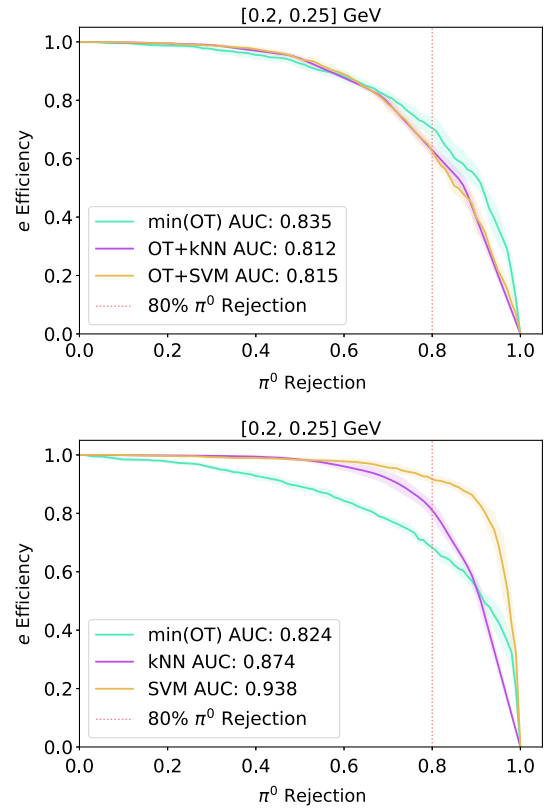


FIG. 13. (Top) Results of a limited run to test the performance of Gromov-Wasserstein distances. (Bottom) The balanced 3D OT results on the same energy bin, [0.2, 0.25], repeated from Fig. 8 for comparison.

similarly, with Gromov-Wasserstein having a slightly higher AUC as shown in Fig. 13. This similarity gives us confidence that our preprocessing is accurately aligning these 3D events. Interestingly, when postprocessing the Gromov-Wasserstein distances with interpretable machine learning methods (kNN and SVM) we do not see a significant performance jump. Whereas, in the balanced 3D OT case, we see a significant increase in performance. This could be explained by the fact that Gromov-Wasserstein solvers, like the ones used here, are approximate which may lead to inexact representations of relative distances.

APPENDIX B: ISOLATED ENERGY DEPOSIT REMOVAL

Certain events contain isolated small energy deposits which are far away from the main energy loss of the EM

shower. These correspond to small branches of the shower produced by photons which travel a large distance before depositing their energy. Similar to the last section, the inclusion of these energy deposits may result in OT spending significantly more cost in the transport plan, leading to a high OT distance even between events of the same type. To mitigate this effect, small isolated energy deposits are removed. A number of configurations are tested: clusters of isolated energy deposits are constructed by collecting together spacepoints within a range of distances (5, 8, 10, 12, 15, or 30 cm), and after this step those clusters with less than 3, 4, or 5 spacepoints are removed. No such configuration however leads to meaningful improvements in performance, leading us to avoid the isolated energy deposit removal for the main results presented here.

-
- [1] P. Huber *et al.*, Snowmass neutrino frontier report, in *Snowmass 2021* (2022); [arXiv:2211.08641](#).
 - [2] R. Acciarri *et al.* (MicroBooNE, LAr1-ND, and ICARUS-WA104 Collaborations), A proposal for a three detector short-baseline neutrino oscillation program in the Fermilab booster neutrino beam, [arXiv:1503.01520](#).
 - [3] B. Abi *et al.* (DUNE Collaboration), Deep underground neutrino experiment (DUNE), far detector technical design report, volume I introduction to DUNE, *J. Instrum.* **15**, T08008 (2020).
 - [4] B. Abi *et al.* (DUNE Collaboration), Long-baseline neutrino oscillation physics potential of the DUNE experiment, *Eur. Phys. J. C* **80**, 978 (2020).
 - [5] A. A. Aguilar-Arevalo *et al.* (MiniBooNE Collaboration), The MiniBooNE detector, *Nucl. Instrum. Methods Phys. Res., Sect. A* **599**, 28 (2009).
 - [6] P. Adamson *et al.* (NOvA Collaboration), First measurement of muon-neutrino disappearance in NOvA, *Phys. Rev. D* **93**, 051104 (2016).
 - [7] D. G. Michael *et al.* (MINOS Collaboration), The Magnetized steel and scintillator calorimeters of the MINOS experiment, *Nucl. Instrum. Methods Phys. Res., Sect. A* **596**, 190 (2008).
 - [8] F. Drielsma, K. Terao, L. Dominé, and D. H. Koh, Scalable, end-to-end, deep-learning-based data reconstruction chain for particle imaging detectors, in *Proceedings of the 34th Conference on Neural Information Processing Systems* (2021), [arXiv:2102.01033](#).
 - [9] A. Aurisano, V. Hewes, G. Cerati, J. Kowalkowski, C. S. Lee, W. Liao, D. Grzenda, K. Gumpula, and X. Zhang, Graph neural network for neutrino physics event reconstruction, *Phys. Rev. D* **110**, 032008 (2024).
 - [10] A. Abed Abud *et al.* (DUNE Collaboration), Separation of track- and shower-like energy deposits in ProtoDUNE-SP using a convolutional neural network, *Eur. Phys. J. C* **82**, 903 (2022).
 - [11] P. Lutkus, T. Wongjirad, and S. Aeron, Towards designing and exploiting generative networks for neutrino physics experiments using liquid argon time projection chambers, in *Proceedings of the 9th International Conference on Learning Representations* (2022), [arXiv:2204.02496](#).
 - [12] F. Drielsma, Q. Lin, P. C. de Soux, L. Dominé, R. Itay, D. H. Koh, B. J. Nelson, K. Terao, K. V. Tsang, and T. L. Usher (DeepLearnPhysics Collaboration), Clustering of electromagnetic showers and particle interactions with graph neural networks in liquid argon time projection chambers, *Phys. Rev. D* **104**, 072004 (2021).
 - [13] L. Dominé, P. C. de Soux, F. Drielsma, D. H. Koh, R. Itay, Q. Lin, K. Terao, K. V. Tsang, and T. L. Usher (DeepLearnPhysics Collaboration), Point proposal network for reconstructing 3D particle endpoints with subpixel precision in liquid argon time projection chambers, *Phys. Rev. D* **104**, 032004 (2021).
 - [14] HEP ML Community, A Living Review of Machine Learning for Particle Physics, [arXiv:2102.02770](#).
 - [15] G. Aad *et al.* (ATLAS Collaboration), A continuous calibration of the ATLAS flavour-tagging classifiers via optimal transportation maps, *Eur. Phys. J. C* **85**, 1272 (2025).
 - [16] M. Algren, J. A. Raine, and T. Golling, Decorrelation using optimal transport, *Eur. Phys. J. C* **84**, 579 (2024).
 - [17] M. Leigh, D. Sengupta, B. Nachman, and T. Golling, Accelerating template generation in resonant anomaly detection searches with optimal transport, *J. High Energy Phys.* **12** (2025) 105.
 - [18] T. Manole, P. Bryant, J. Alison, M. Kuusela, and L. Wasserman, Background modeling for double Higgs boson production: Density ratios and optimal transport, [arXiv:2208.02807](#).
 - [19] A. Davis, T. Menzo, A. Youssef, and J. Zupan, Earth mover's distance as a measure of CP violation, *J. High Energy Phys.* **06** (2023) 098.

- [20] T. Cheng, J.-F. Arguin, J. Leissner-Martin, J. Pilette, and T. Golling, Variational autoencoders for anomalous jet tagging, *Phys. Rev. D* **107**, 016002 (2023).
- [21] A. J. Larkoski, Factorization for collider dataspace correlators, *Phys. Rev. D* **112**, 054012 (2025).
- [22] A. J. Larkoski, Non-gaussianities in collider metric binning, [arXiv:2503.03809](#).
- [23] T. Cai, N. Craig, K. Craig, and X. Lin, Multi-scale optimal transport for complete collider events, *Phys. Rev. D* **112**, 036021 (2025).
- [24] N. Craig, J. N. Howard, and H. Li, Exploring optimal transport for event-level anomaly detection at the large hadron collider, [arXiv:2401.15542](#).
- [25] R. Gambhir, A. J. Larkoski, and J. Thaler, SPECTER: Efficient evaluation of the spectral EMD, *J. High Energy Phys.* **12** (2024) 219.
- [26] T. Gaertner and J. Reiten, Unsupervised learning in the metric space of jets, [arXiv:2312.06948](#).
- [27] A. J. Larkoski and J. Thaler, A spectral metric for collider geometry, *J. High Energy Phys.* **08** (2023) 107.
- [28] D. Ba, A. S. Dogra, R. Gambhir, A. Tasissa, and J. Thaler, SHAPER: Can you hear the shape of a jet?, *J. High Energy Phys.* **06** (2023) 195.
- [29] L. Gouskos, F. Iemmi, S. Liechti, B. Maier, V. Mikuni, and H. Qu, Optimal transport for a novel event description at hadron colliders, *Phys. Rev. D* **108**, 096003 (2023).
- [30] S. E. Park, P. Harris, and B. Ostdiek, Neural embedding: Learning the embedding of the manifold of physics data, *J. High Energy Phys.* **07** (2023) 108.
- [31] T. Cai, J. Cheng, K. Craig, and N. Craig, Which metric on the space of collider events?, *Phys. Rev. D* **105**, 076003 (2022).
- [32] S. Tsan, R. Kansal, A. Aportela, D. Diaz, J. Duarte, S. Krishna, F. Mokhtar, J.-R. Vlimant, and M. Pierini, Particle graph autoencoders and differentiable, learned energy Mover's distance, in *Proceedings of the 35th Conference on Neural Information Processing Systems* (2021), [arXiv:2111.12849](#).
- [33] G. Di Guglielmo *et al.*, A reconfigurable neural network ASIC for detector front-end data compression at the HL-LHC, *IEEE Trans. Nucl. Sci.* **68**, 2179 (2021).
- [34] J. H. Collins, An exploration of learnt representations of W jets, [arXiv:2109.10919](#).
- [35] T. Cai, J. Cheng, N. Craig, and K. Craig, Linearized optimal transport for collider events, *Phys. Rev. D* **102**, 116019 (2020).
- [36] M. Crispim Romão, N. F. Castro, J. G. Milhano, R. Pedro, and T. Vale, Use of a generalized energy Mover's distance in the search for rare phenomena at colliders, *Eur. Phys. J. C* **81**, 192 (2021).
- [37] P. T. Komiske, E. M. Metodiev, and J. Thaler, Metric space of collider events, *Phys. Rev. Lett.* **123**, 041801 (2019).
- [38] P. T. Komiske, R. Mastandrea, E. M. Metodiev, P. Naik, and J. Thaler, Exploring the space of jets with CMS open data, *Phys. Rev. D* **101**, 034009 (2020).
- [39] K. Fraser, S. Homiller, R. K. Mishra, B. Ostdiek, and M. D. Schwartz, Challenges for unsupervised anomaly detection in particle physics, *J. High Energy Phys.* **03** (2022) 066.
- [40] P. T. Komiske, E. M. Metodiev, and J. Thaler, The hidden geometry of particle collisions, *J. High Energy Phys.* **07** (2020) 006.
- [41] C. Cesarotti and M. LeBlanc, A field guide to event-shape observables using optimal transport, *J. High Energy Phys.* **12** (2025) 014.
- [42] G. Aad *et al.* (ATLAS Collaboration), Measurements of multijet event isotropies using optimal transport with the ATLAS detector, *J. High Energy Phys.* **10** (2023) 060.
- [43] C. Cesarotti and J. Thaler, A robust measure of event isotropy at colliders, *J. High Energy Phys.* **08** (2020) 084.
- [44] C. Cesarotti, M. Reece, and M. J. Strassler, The efficacy of event isotropy as an event shape observable, *J. High Energy Phys.* **07** (2021) 215.
- [45] P. Abratenko *et al.* (MicroBooNE Collaboration), Search for an anomalous production of charged-current ν_e interactions without visible pions across multiple kinematic observables in MicroBooNE, *Phys. Rev. Lett.* **135**, 081802 (2025).
- [46] A. M. Abdullahi *et al.* (MicroBooNE Collaboration), First search for dark sector e^+e^- explanations of the Mini-BooNE anomaly at MicroBooNE, [arXiv:2502.10900](#).
- [47] P. Abratenko *et al.* (MicroBooNE Collaboration), Inclusive search for anomalous single-photon production in MicroBooNE, [arXiv:2502.06064](#).
- [48] B. Abi *et al.* (DUNE Collaboration), First results on ProtoDUNE-SP liquid argon time projection chamber performance from a beam test at the CERN Neutrino Platform, *J. Instrum.* **15**, P12004 (2020).
- [49] C. Adams *et al.* (MicroBooNE Collaboration), Calibration of the charge and energy loss per unit length of the MicroBooNE liquid argon time projection chamber using muons and protons, *J. Instrum.* **15**, P03022 (2020).
- [50] P. Abratenko *et al.* (ICARUS Collaboration), ICARUS at the Fermilab short-baseline neutrino program: Initial operation, *Eur. Phys. J. C* **83**, 467 (2023).
- [51] P. Abratenko *et al.* (MicroBooNE Collaboration), First constraints on light sterile neutrino oscillations from combined appearance and disappearance searches with the MicroBooNE detector, *Phys. Rev. Lett.* **130**, 011801 (2023).
- [52] P. Abratenko *et al.* (MicroBooNE Collaboration), Search for heavy neutral leptons decaying into muon-pion pairs in the MicroBooNE detector, *Phys. Rev. D* **101**, 052001 (2020).
- [53] P. Abratenko *et al.* (MicroBooNE Collaboration), Search for long-lived heavy neutral leptons and Higgs portal scalars decaying in the MicroBooNE detector, *Phys. Rev. D* **106**, 092006 (2022).
- [54] P. Abratenko *et al.* (MicroBooNE Collaboration), First search for dark-trident processes using the MicroBooNE detector, *Phys. Rev. Lett.* **132**, 241801 (2024).
- [55] K. Abe *et al.* (Hyper-Kamiokande Collaboration), Hyper-Kamiokande design report, [arXiv:1805.04163](#).
- [56] A. Abdullahi *et al.*, From oversimplified to overlooked: The case for exploring Rich Dark Sectors, *Nucl. Phys. B* **1020**, 117148 (2025).
- [57] B. Batell *et al.*, Dark sector studies with neutrino beams, in *Snowmass 2021* (2022); [arXiv:2207.06898](#).
- [58] D. Rein and L. M. Sehgal, Neutrino-excitation of baryon resonances and single pion production, *Ann. Phys. (N.Y.)* **133**, 79 (1981).

- [59] T. Leitner, O. Buss, L. Alvarez-Ruso, and U. Mosel, Electron- and neutrino-nucleus scattering from the quasielastic to the resonance region, *Phys. Rev. C* **79**, 034601 (2009).
- [60] R. A. *et al.* (MicroBooNE Collaboration), Design and construction of the microboone detector, *J. Instrum.* **12**, P02017 (2017).
- [61] A. Abed Abud *et al.* (DUNE Collaboration), The DUNE far detector vertical drift technology. Technical design report, *J. Instrum.* **19**, T08004 (2024).
- [62] S. Abbaslu *et al.* (DUNE Collaboration), Operation of a modular 3D-pixelated liquid argon time-projection chamber in a neutrino beam, [arXiv:2509.07012](https://arxiv.org/abs/2509.07012).
- [63] A. Abed Abud *et al.* (DUNE Collaboration), Performance of a modular ton-scale pixel-readout liquid argon time projection chamber, *Instruments* **8**, 41 (2024).
- [64] R. Acciarri *et al.* (MicroBooNE Collaboration), The PANDORA multi-algorithm approach to automated pattern recognition of cosmic-ray muon and neutrino events in the MicroBooNE detector, *Eur. Phys. J. C* **78**, 82 (2018).
- [65] P. Abratenko *et al.* (MicroBooNE Collaboration), WIRE-CELL 3D pattern recognition techniques for neutrino event reconstruction in large LArTPCs: Algorithm description and quantitative evaluation with MicroBooNE simulation, *J. Instrum.* **17**, P01037 (2022).
- [66] R. Acciarri *et al.* (MicroBooNE Collaboration), Michel electron reconstruction using cosmic-ray data from the MicroBooNE LArTPC, *J. Instrum.* **12**, P09014 (2017).
- [67] C. Adams *et al.* (MicroBooNE Collaboration), Reconstruction and measurement of $\mathcal{O}(100)$ MeV energy electromagnetic activity from $\pi^0 \rightarrow \gamma\gamma$ decays in the MicroBooNE LArTPC, *J. Instrum.* **15**, P02007 (2020).
- [68] MicroBooNE, Microboone BNB inclusive overlay sample (no wire info), [10.5281/zenodo.8370883](https://zenodo.org/record/8370883) (2023).
- [69] MicroBooNE, Microboone BNB electron neutrino overlay sample (no wire info), [10.5281/zenodo.7261921](https://zenodo.org/record/7261921) (2022).
- [70] I. Stancu, Technical design report for the 8 GeV beam, Report No. FERMILAB-DESIGN-2001-03, 2001, [10.2172/1212167](https://arxiv.org/abs/10.2172/1212167).
- [71] C. Andreopoulos *et al.*, The GENIE neutrino Monte Carlo generator, *Nucl. Instrum. Methods Phys. Res., Sect. A* **614**, 87 (2010).
- [72] P. Abratenko *et al.* (MicroBooNE Collaboration), New $CC0\pi$ GENIE model tune for MicroBooNE, *Phys. Rev. D* **105**, 072001 (2022).
- [73] E. L. Snider and G. Petrillo, LArSoft: Toolkit for simulation, reconstruction and analysis of liquid argon TPC neutrino detectors, *J. Phys. Conf. Ser.* **898**, 042057 (2017).
- [74] C. Adams *et al.* (MicroBooNE Collaboration), A method to determine the electric field of liquid argon time projection chambers using a UV laser system and its application in MicroBooNE, *J. Instrum.* **15**, P07010 (2020).
- [75] R. Acciarri *et al.* (MicroBooNE Collaboration), Noise characterization and filtering in the MicroBooNE liquid argon TPC, *J. Instrum.* **12**, P08003 (2017).
- [76] C. Adams *et al.* (MicroBooNE Collaboration), Ionization electron signal processing in single phase LArTPCs. Part I. Algorithm Description and quantitative evaluation with MicroBooNE simulation, *J. Instrum.* **13**, P07006 (2018).
- [77] C. Adams *et al.* (MicroBooNE Collaboration), Ionization electron signal processing in single phase LArTPCs. Part II. Data/simulation comparison and performance in MicroBooNE, *J. Instrum.* **13**, P07007 (2018).
- [78] A. Abed Abud *et al.* (DUNE Collaboration), Reconstruction of interactions in the ProtoDUNE-SP detector with PANDORA, *Eur. Phys. J. C* **83**, 618 (2023).
- [79] P. Abratenko *et al.* (MicroBooNE Collaboration), Neutrino event selection in the MicroBooNE liquid argon time projection chamber using WIRE-CELL 3-D imaging, clustering and charge-light matching, *J. Instrum.* **16**, P06043 (2021).
- [80] R. Acciarri *et al.* (MicroBooNE Collaboration), Convolutional neural networks applied to neutrino events in a liquid argon time projection chamber, *J. Instrum.* **12**, P03011 (2017).
- [81] P. Abratenko *et al.* (MicroBooNE Collaboration), A convolutional neural network for multiple particle identification in the MicroBooNE liquid argon time projection chamber, *Phys. Rev. D* **103**, 092003 (2021).
- [82] B. Abi *et al.* (DUNE Collaboration), Neutrino interaction classification with a convolutional neural network in the DUNE far detector, *Phys. Rev. D* **102**, 092003 (2020).
- [83] A. Abed Abud *et al.* (DUNE Collaboration), Separation of track- and shower-like energy deposits in ProtoDUNE-SP using a convolutional neural network, *Eur. Phys. J. C* **82**, 903 (2022).
- [84] C. Adams *et al.* (MicroBooNE Collaboration), Deep neural network for pixel-level electromagnetic particle identification in the MicroBooNE liquid argon time projection chamber, *Phys. Rev. D* **99**, 092001 (2019).
- [85] P. Abratenko *et al.* (MicroBooNE Collaboration), Semantic segmentation with sparse convolutional neural network for event reconstruction in MicroBooNE, *Phys. Rev. D* **103**, 052012 (2021).
- [86] P. Abratenko *et al.* (MicroBooNE Collaboration), Improving neutrino energy estimation of charged-current interaction events with recurrent neural networks in MicroBooNE, *Phys. Rev. D* **110**, 092010 (2024).
- [87] R. Acciarri *et al.* (ArgoNeuT Collaboration), First observation of low energy electron neutrinos in a liquid argon time projection chamber, *Phys. Rev. D* **95**, 072005 (2017).
- [88] P. Abratenko *et al.* (MicroBooNE Collaboration), Search for an anomalous excess of charged-current ν_e interactions without pions in the final state with the MicroBooNE experiment, *Phys. Rev. D* **105**, 112004 (2022).
- [89] P. Abratenko *et al.* (MicroBooNE Collaboration), Search for an anomalous production of charged-current ν_e interactions without visible pions across multiple kinematic observables in MicroBooNE, *Phys. Rev. Lett.* **135**, 081802 (2025).
- [90] P. Abratenko *et al.* (MicroBooNE Collaboration), First search for neutral current coherent single-photon production in MicroBooNE, [arXiv:2502.06091](https://arxiv.org/abs/2502.06091).
- [91] P. Abratenko *et al.* (MicroBooNE Collaboration), Enhanced search for neutral current Δ radiative single-photon

- production in MicroBooNE, *Phys. Rev. D* **112**, L091101 (2025).
- [92] P. Abratenko *et al.* (MicroBooNE Collaboration), First measurement of ν_e and $\bar{\nu}_e$ charged-current single charged-pion production differential cross sections on argon using the MicroBooNE detector, *Phys. Rev. Lett.* **135**, 061802 (2025).
- [93] F. Mémoli, Gromov–wasserstein distances and the metric approach to object matching, *Found. Comput. Math.* **11**, 417 (2011).
- [94] B. A. *et al.* (DUNE Collaboration), Deep underground neutrino experiment (dune), far detector technical design report, volume ii: Dune physics, [arXiv:2002.03005](https://arxiv.org/abs/2002.03005).
- [95] L. Delchambre, Weighted principal component analysis: A weighted covariance eigendecomposition approach, *Mon. Not. R. Astron. Soc.* **446**, 3545 (2014).
- [96] R. Flamary, N. Courty, A. Gramfort, M. Z. Alaya, A. Boissunon, S. Chambon, L. Chapel, A. Corenflos, K. Fatras, N. Fournier, L. Gautheron, N. T. Gayraud, H. Janati, A. Rakotomamonjy, I. Redko, A. Rolet, A. Schutz, V. Seguy, D. J. Sutherland, R. Tavenard, A. Tong, and T. Vayer, Pot: PYTHON optimal transport, *J. Mach. Learn. Res.* **22**, 1 (2021).
- [97] T. Cover and P. Hart, Nearest neighbor pattern classification, *IEEE Trans. Inf. Theory* **13**, 21 (1967).
- [98] B. Schölkopf, R. C. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, Support vector method for novelty detection, in *Advances in Neural Information Processing Systems*, edited by S. Solla, T. Leen, and K. Müller (MIT Press, Cambridge, MA, 1999), Vol. 12, https://papers.nips.cc/paper_files/paper/1999/hash/8725fb777f25776ffa9076-e44fcfd776-Abstract.html.
- [99] N. Bonneel, M. van de Panne, S. Paris, and W. Heidrich, Displacement interpolation using Lagrangian mass transport, *ACM Trans. Graph.* **30**, 1 (2011).
- [100] M. Cuturi, Sinkhorn distances: Lightspeed computation of optimal transportation distances, [arXiv:1306.0895](https://arxiv.org/abs/1306.0895).
- [101] J. Altschuler, J. Weed, and P. Rigollet, Near-linear time approximation algorithms for optimal transport via sinkhorn iteration, [arXiv:1705.09634](https://arxiv.org/abs/1705.09634).
- [102] A. Genevay, L. Chizat, F. Bach, M. Cuturi, and G. Peyré, Sample complexity of sinkhorn divergences, [arXiv:1810.02733](https://arxiv.org/abs/1810.02733).
- [103] K. Nguyen and N. Ho, Energy-based sliced wasserstein distance, [arXiv:2304.13586](https://arxiv.org/abs/2304.13586).
- [104] N. Kravtsova, Note on computational complexity of the Gromov-Wasserstein distance, [arXiv:2408.06525](https://arxiv.org/abs/2408.06525).
- [105] G. Peyré, M. Cuturi, and J. M. Solomon, Gromov-wasserstein averaging of kernel and distance matrices, in *International Conference on Machine Learning* (2016), <https://proceedings.mlr.press/v48/peyre16.html>.
- [106] T. Vayer, R. Flamary, R. Tavenard, L. Chapel, and N. Courty, Sliced Gromov-Wasserstein, [arXiv:1905.10124](https://arxiv.org/abs/1905.10124).