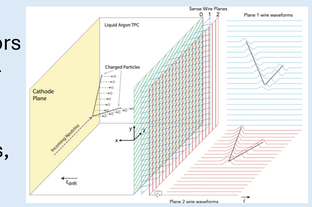
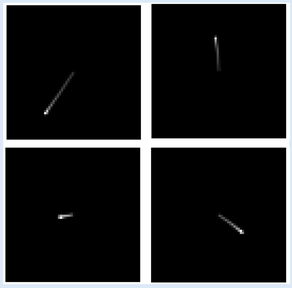
Data-Driven Generation & Inference of LArTPC Images Using Conditional Latent Diffusion Models

1. LArTPC Images

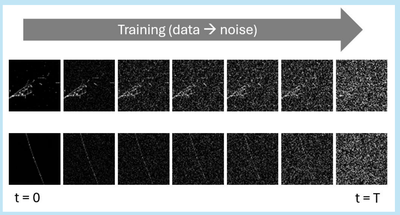
Liquid Argon Time Projection Chamber (LArTPC) detectors are used in many neutrino experiments to produce high-resolution images of interaction events within the medium. Charged particles ionize the liquid argon, and the resulting charges are drifted towards the wire planes, yielding 2D projection images of the 3D event.



For this work, we simulated protons in the 50–100 MeV kinetic energy range using Geant4. We used a simplified pixel-based readout model that projects 3D energy depositions onto two orthogonal 2D wire plane views without full detector effects (transverse diffusion only). Each event was cropped to a 64×64 pixel event image centered on the track vertex, with full containment enforced. The ground-truth 3D momentum vector $\mathbf{p} = (p_x, p_y, p_z)$ was recorded for each simulated event. Four representative examples are shown.

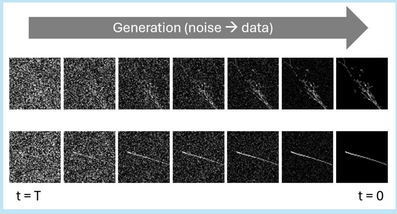


2. Conditional Latent Diffusion Model (LDM)

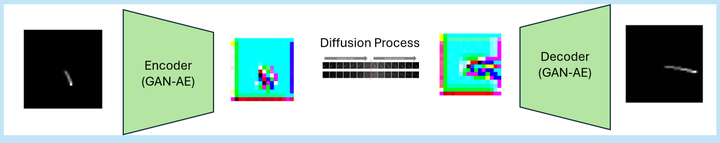


Diffusion is a process for transforming Gaussian noise into images. During training, the model learns to gradually add noise to an event image, associating a noise level with each timestep $t = 0$ to T. During training, true 3D momentum labels $\mathbf{p}$ are provided alongside each image to jointly train the conditioning network (see below).

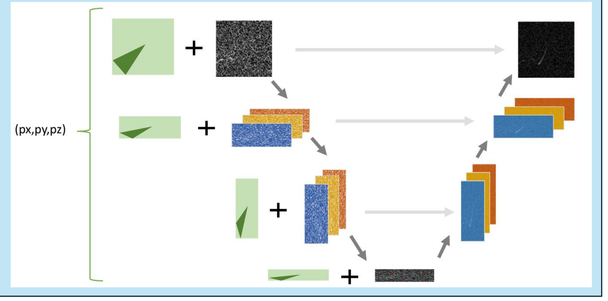
By learning the gradients associated with adding noise, we can reverse the process for generation. We sample from a Gaussian and then iteratively remove noise until we arrive at the denoised image at $t = 0$. The generated images are unique and align with the overall characteristics of our dataset.



Since diffusion computation scales with the number of pixels, we incorporate an autoencoder before and after the diffusion process to increase efficiency, running diffusion on the smaller latent space representations. The autoencoder was trained separately.

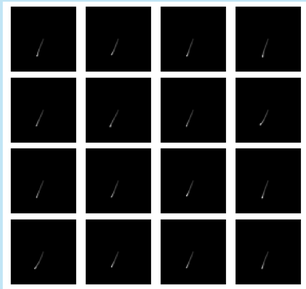


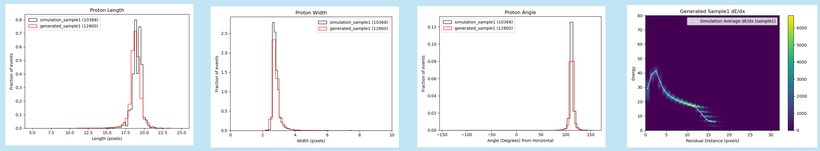
Our latent diffusion model uses a U-Net to predict the noise present in a latent image at each denoising timestep. Conditioning on momentum is achieved through cross-attention layers incorporated into the U-Net architecture. A transformer network encodes the momentum vector $\mathbf{p}$ into a learned embedding, which is injected via cross-attention into the intermediate layers of the U-Net. This transformer is trained jointly with the diffusion model, requiring momentum labels for each training image. This cartoon diagram illustrates a single denoising timestep; the green shapes represent cross-attention maps showing how the momentum condition modulates the spatial features of the denoising process.



3. Conditional Generation

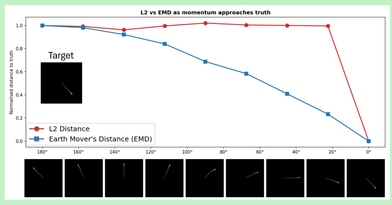
The conditional LDM can generate novel proton event images for any 3D momentum vector $\mathbf{p}$ within the 50–100 MeV kinetic energy range. Shown here are event images generated using a single momentum vector. The slight differences between images reflect the variations found in real-life protons with the same momentum. To quantify the quality of our generated images, we compare them to a sample of Geant4 simulated protons using the same momentum. Shown below are track length, width, angle, and dE/dx. These are simple homebrew reconstructions and are sensitive to individual pixel variations; in particular, the dE/dx curve does not adhere to typical normalization conventions. Nevertheless, all three metrics show good agreement between generated and simulated distributions, demonstrating that **our model produces physically accurate proton event images without any underlying physics simulation.**



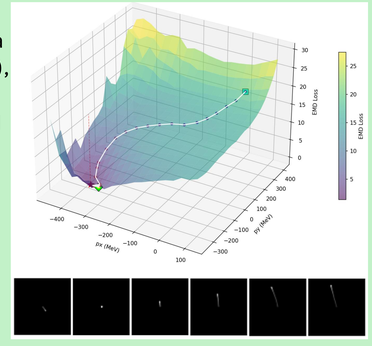


4. Track Inference

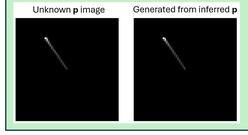
Now that we have a conditional generator for proton event images, we can reverse the process and infer the properties of an event image. To start, we need a metric for comparing images. An initial guess of using L_2 distance proves ineffective due to the sparsity of our images. Instead, we use Earth Mover's Distance (EMD), a.k.a. Wasserstein-1, which measures the minimum work needed to move mass (pixel intensity, deposited charge, etc.) from one region of the image to another. The name comes from the analogy of moving piles of dirt into holes with the least amount of effort. The plot shows that as a generated track approaches the target in orientation and position, the EMD decreases gradually, while the L_2 distance remains unchanged until near-exact overlap.

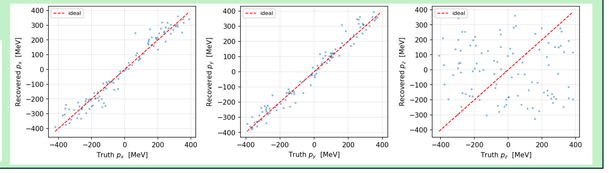


Using EMD to compare images, we construct a loss function $\mathcal{L} = \mathrm{EMD}(\textit{target image}\,,\mathrm{LDM}(\mathbf{p}))$, where $\mathrm{LDM}(\mathbf{p})$ denotes an event image generated by the conditional LDM conditioned on momentum $\mathbf{p}$. By backpropagating through the model to the momentum condition, we perform stochastic gradient descent to find a generated image that matches the target. This is shown below, with sample images from the gradient descent process illustrating how the generated track evolves toward the target.



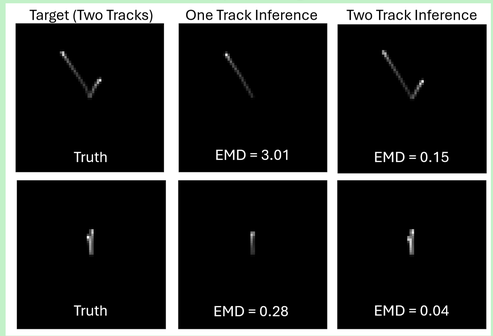
From this, we infer that the final momentum $\mathbf{p}$ producing the best match is the momentum of the target image. As shown, this accurately reconstructs the p_x and p_y momentum components. Since our images are 2D wire plane projections, the out-of-plane component p_z has an inherent sign degeneracy and an overall lesser effect on the event image. Recovering the magnitude of p_z is theoretically possible with further development of the inference procedure, but the sign ambiguity cannot be resolved from a single projection alone. Instead, we present an alternate solution.





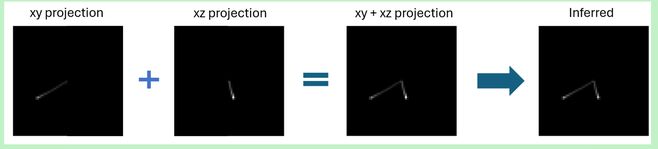
5. Multi-Track Inference

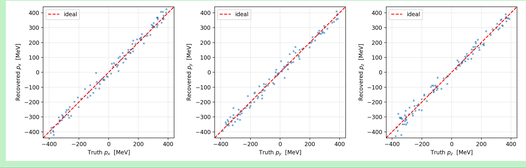
Single-track inference applied to a multi-particle target fits only the dominant track. We extend the inference process by generating two single-track event images and adding them before computing the EMD, thereby allowing us to backpropagate to all six momentum components. As shown, the final EMD is significantly lower in this two-track inference case, and this holds regardless of the separation angle between tracks. This inference approach generalizes to N tracks, providing a natural method for identifying the number of tracks in an event image. Given a target image with N tracks, a generated image with N-1 tracks will yield significantly worse EMD than one with N tracks, while adding superfluous tracks beyond N has minimal effect as the extra tracks are pushed to zero momentum.



6. Projection Trick

Standard LArTPC detectors produce two 2D wire plane projections of each event, in our setup, the XY and XZ views of the same 3D interaction. We exploit this geometry to improve inference of the full 3D momentum vector $\mathbf{p}$, particularly the p_z component. By treating the two projections as a synthetic two-track event and applying the two-track inference framework from Section 5, we recover two momentum estimates: $(p_x, p_y, p_z)$and (p_x, p_z, p_y). We enforce a shared gradient between the two tracks to constrain them to a single consistent momentum solution. As shown in the recovered vs. true momentum plots, **this approach accurately infers all three momentum components without any traditional reconstruction methods.**





7. Conclusions

I. **Conditional Latent Diffusion Models enable event generation without underlying physics simulation (i.e. data-driven approach).**

II. **Inference of event properties without traditional reconstruction methods using EMD loss and stochastic gradient descent.**

III. **Demonstrated with LArTPC protons. Goal of applying to regimes that traditional event generators struggle (neutrino-nucleus interactions).**

8. Sources & Acknowledgments

[1] Acciarri, R., et al. "Design and Construction of the MicroBooNE Detector." Journal of Instrumentation, vol. 12, no. 02, Feb. 2017, p. P02017, doi:10.1088/1748-0221/12/02/P02017.
[2] Agostinelli, S., et al. "GEANT4 - A Simulation Toolkit." Nucl. Instrum. Meth. A, vol. 506, 2003, pp. 250–303, doi:10.1016/S0168-9002(03)01368-8.
[3] Feydy, Jean, et al. "Fast and Scalable Optimal Transport for Brain Tractograms." MICCAI 2019, 2019, doi:10.1007/978-3-030-32248-9_71.
[4] Rombach, Robin, et al. " High-Resolution Image Synthesis with Latent Diffusion Models ." 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) , IEEE Computer Society, 2022, pp. 10674–85, doi:10.1109/CVPR52688.2022.01042.
[5] Imani, Zeviel, et al. "Score-Based Diffusion Models for Generating Liquid Argon Time Projection Chamber Images." Phys. Rev. D, vol. 109, no. 7, American Physical Society, Apr. 2024, p. 072011, doi:10.1103/PhysRevD.109.072011.

Thanks to Prof. Taritree Wongjirad (Tufts Physics) and Prof. Shuchin Aeron (Tufts ECE/CS)
Support from Tufts University & The NSF Institute for Artificial Intelligence and Fundamental Interactions (IAIFI)

Contact info: Zeviel.Imani@tufts.edu