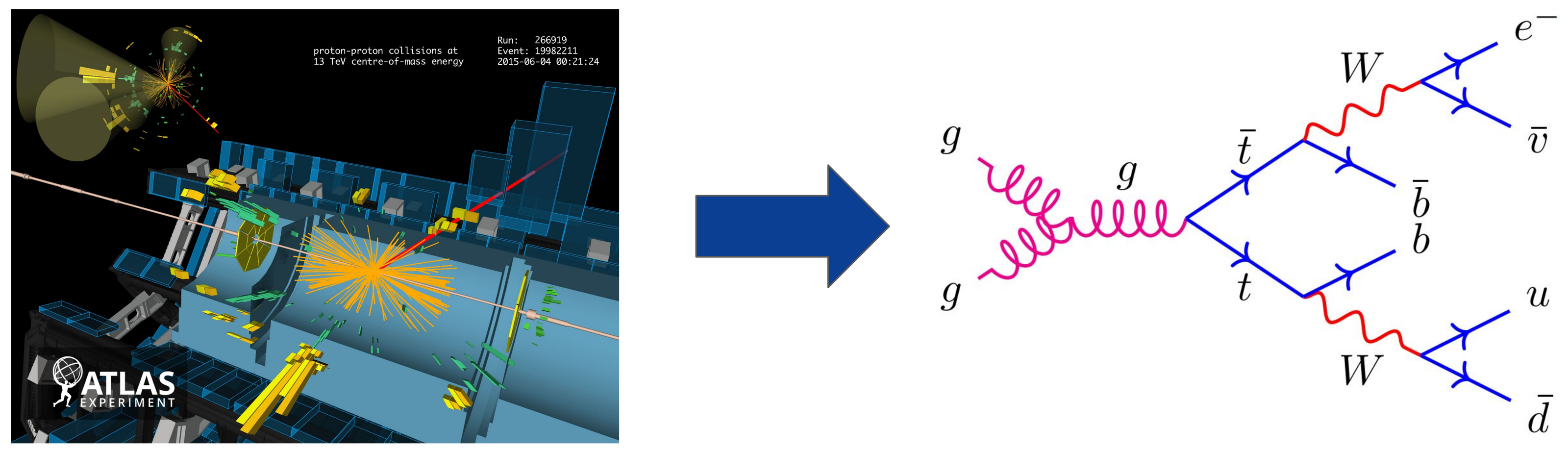
End-To-End Latent Variational Diffusion Models for Inverse Problems in High Energy Physics

Alexander Shmakov, Kevin Greif, Michael Fenton, Aishik Ghosh, Pierre Baldi, Daniel Whiteson

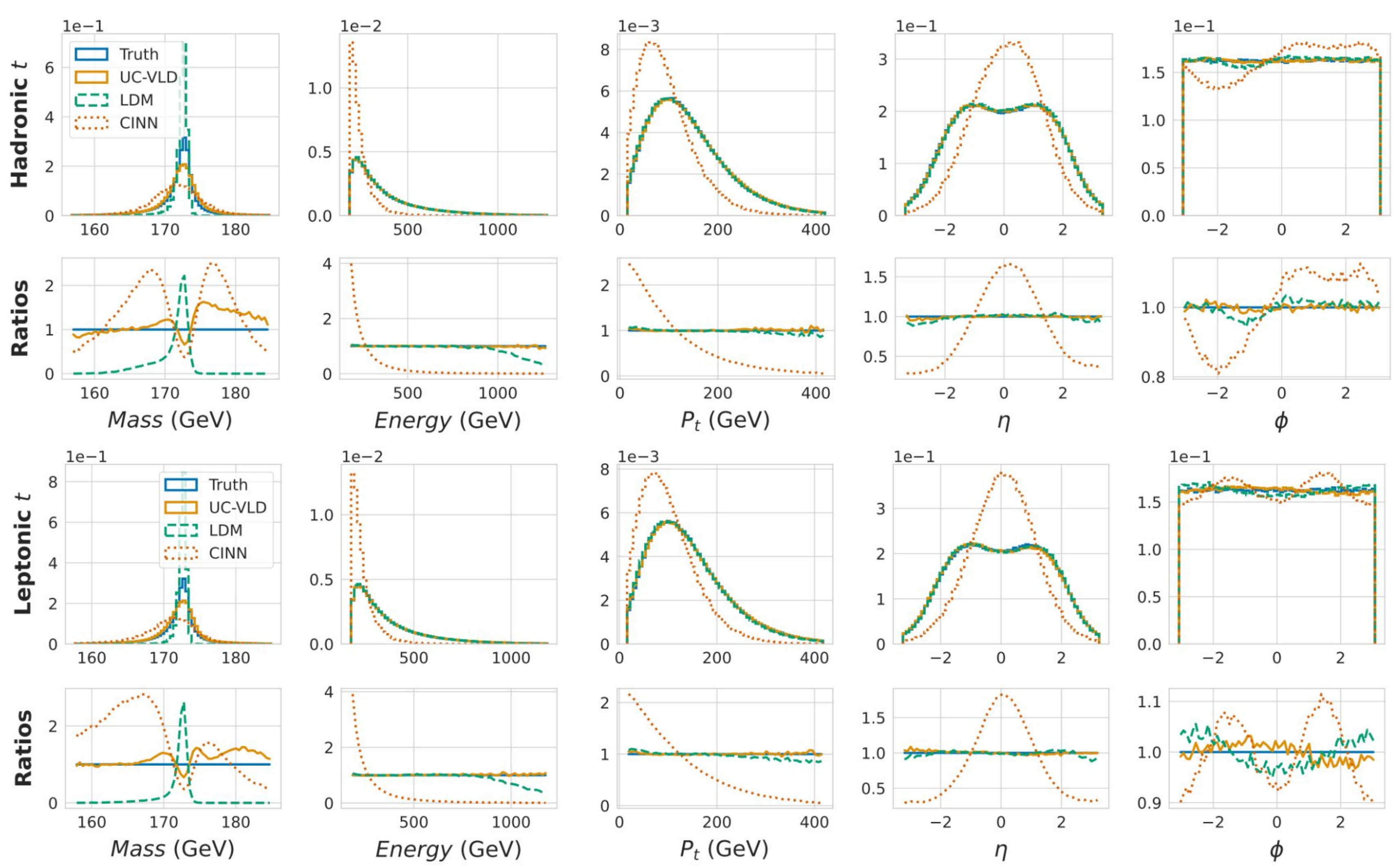
Background

- Large Hadron Collider (LHC) measures particle collisions which are key for answering open questions in particle physics.
- Detector effects introduce bias into results and must be corrected for theoretical measurements.
- **Unfolding** is an inverse problem of converting *detector observations* into more fundamental theoretical quantities.
- We employ state-of-the-art **latent diffusion models** to tackle this generative inverse problem.
- Focus on the semileptonic ttbar decay channel. Map from ATLAS **detector** measurements to fundamental **parton** momenta.



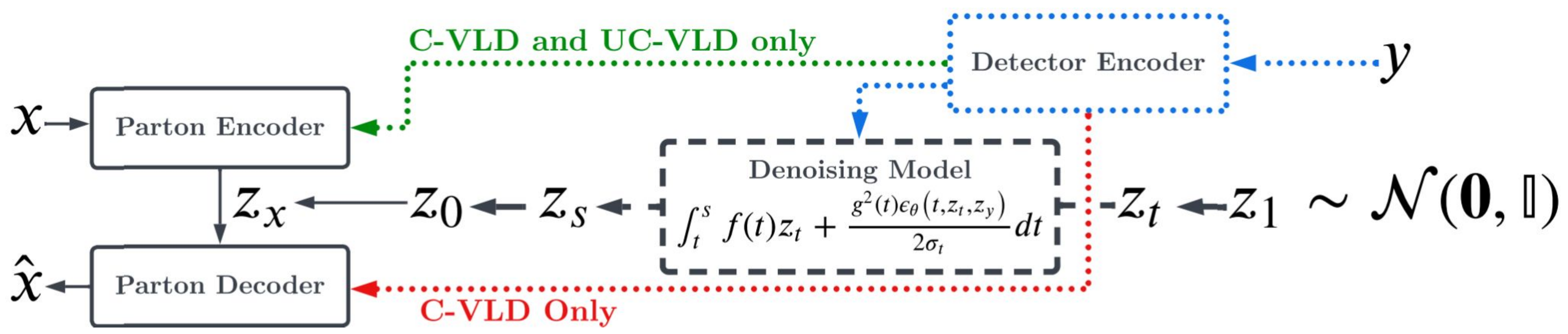
Results

Distribution and ratios to truth for top quark kinematics across the entire testing dataset. Each event was sampled once for each model.



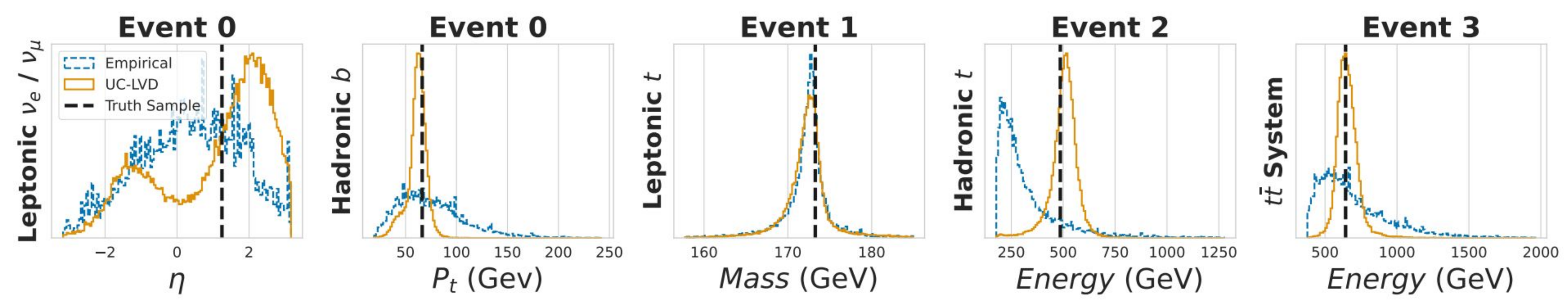
Latent Variational Diffusion

- **Unified Variational Approach** We combine latent diffusion models [1] with the variational diffusion framework [2] to create a unified end-to-end variational latent diffusion model.
- **Detector Encoder** Use permutation invariant transformer architecture [3] to embed the variable length detector measurements into informative, latent fixed-length representation.
- **Parton VAE** Use a Gaussian VAE to encode and decode fixed-length parton representations of event into abstract latent space.
  - Unlike most VAEs, we embed into a *higher dimension* than our data!
  - Train VAE simultaneously with other network components to learn a fine-tuned latent space for diffusion.
  - Explore several variations for conditioning the VAE.
- **Denoising Network** Diffusion models learn a target distribution by learning to reverse a Gaussian noise diffusion process. Feed in the detector embedding and encoded particle to learn a conditional diffusion process. We use a continuous variance-preserving diffusion process.



Produce posterior distributions using generative models to examine the possible partons configurations for a single event.

1. Lets us measure the *confidence* of the network in its estimation of the parton kinematics.
2. Allows us to test if the network captured symmetries and known physics constraints.
   a. We notice a bimodal symmetric distribution of the light quarks in the event, indicating that the network is learning charge symmetry.
   b. Notice that the neutrino eta posterior is also (non-symmetric) bimodal! The neutrino is not detected and must be inferred from the missing energy. This induces a quadratic constraint with two possible solutions, and the network picked up on this constraint!



Experiments

- Generate a training dataset with Madgraph, Pythia, and Delphes simulators. We generate 10 million training examples to train all models to completion.
  - **Detector Observations** 4-Momenta of up to 20 hadronic jets and leptons. Also measure the missing energy to estimate neutrino.
  - **Parton Variables** 4-Momenta of intermediate and decay particles for a total of 55 dimensions. Generate 5-component representation (M, E, Px, Py, Pz) to handle challenging mass component.
- Train several generative learning approaches, building up to the unified latent variational diffusion architecture.
  - **CVAE** Basic conditional VAE with a Gaussian latent prior [4].
  - **VDM** Variational diffusion model in data-space [2].
  - **LDM** Latent diffusion model with pre-trained CLIP VAE.
  - **CINN** Conditional normalizing flow and current *SOTA* for unfolding [5].
- Evaluate methods using distribution-free metrics between the true parton distributions and the generated samples across a 1 million example testing dataset.
  - Unified approach performs best; effectiveness of end-to-end training.
  - Latent methods outperform data-space methods.

	Wasserstein	Energy	K-S	KL_{64}	KL_{128}	KL_{256}
VLD	108.76	7.59	4.08	**3.47**	**3.74**	**4.53**
UC-VLD	**73.56**	**6.35**	**3.41**	5.77	7.10	8.48
C-VLD	389.62	25.39	4.65	9.54	10.09	10.79
LDM	402.32	24.09	5.91	14.71	16.34	17.92
VDM	2478.35	181.35	17.14	29.28	32.29	35.60
CVAE	484.56	32.29	6.37	7.79	9.17	10.60
CINN	3009.08	185.13	15.74	28.55	30.19	32.37

References

[1] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models, 2021.
[2] Diederik P Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. On density estimation with diffusion models. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, Advances in Neural Information Processing Systems, 2021.
[3] Alexander Shmakov, Michael James Fenton, Ta-Wei Ho, Shih-Chieh Hsu, Daniel Whiteson, and Pierre Baldi. SPANet: Generalized Permutationless Set Assignment for Particle Physics using Symmetry Preserving Attention. SciPost Phys., 12:178, 2022.
[4] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning structured output representation using deep conditional generative models. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, Advances in Neural Information Processing Systems, volume 28, 2015.
[5] Marco Bellagente, Anja Butter, Gregor Kasieczka, Tilman Plehn, Armand Rousselot, Ramon Winterhalder, Lynton Ardizzone, and Ullrich Köthe. Invertible networks or partons to detector and back again. SciPost Phys., 9:074, 2020.