Generative Graph Neural Networks for Reconstructing Parton-Level Jets after Hadronization

Umar Sohail Qureshi (uqureshi@cern.ch)
Michael Taleb, Raghav (Rithya) Kunnawalkam Elayavalli

Department of Physics and Astronomy, Vanderbilt University, Nashville, TN, USA

Hot Jets: Advancing the Understanding of High Temperature QCD with Jets
Loomis Lab, Urbana Champaign, Illinois, United States

January 8, 2025

U.S. DEPARTMENT OF ENERGY | Office of Science

VANDERBILT UNIVERSITY

The Anatomy of a Hadron-Hadron Collision

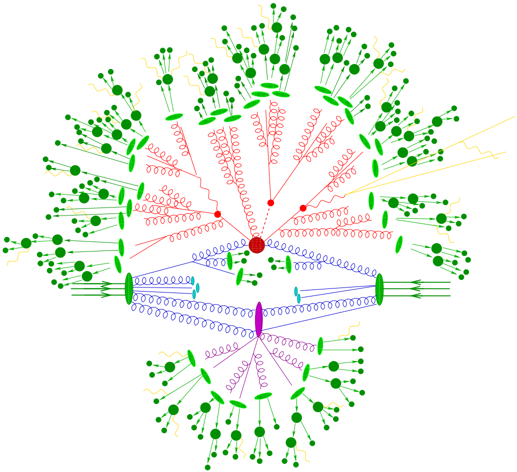


Figure 1: Schematic of a hadron-hadron collision.
Image Credit: Stefan Hoche

Perturbative

- High Q^2 scattering.
- Parton showering.

Possible because of factoring theorems.

Non-Perturbative

- Hadronization.
- Multi-parton interactions.
- Underlying events.

Hadronization Processes

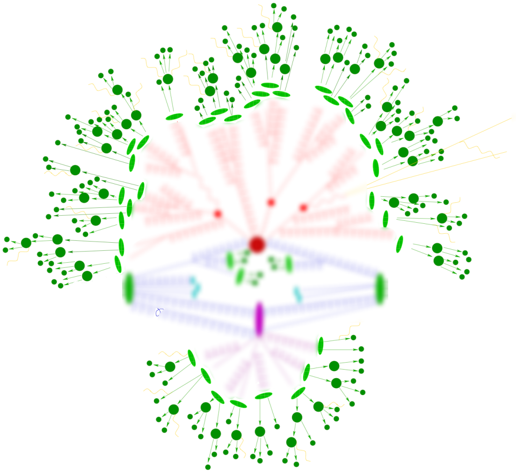


Figure 2: Schematic of a hadron-hadron collision.
Image Credit: Stefan Hoche

Hadronization

- Formation of hadrons from quarks and gluons.
- Incalculable using pQCD!

Phenomenological Models

- Parameterized fits to data.
- Intractable to recover partonic event analytically.

Probing the Intrinsic Parton Shower

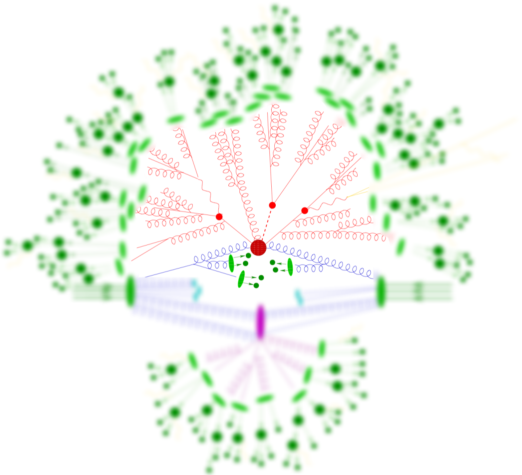


Figure 3: Schematic of a hadron-hadron collision. Image Credit: Stefan Hoche

The Goal

- Reconstruct the intrinsic (and immeasurable) parton shower from experimentally accessible quantities.

Predicting Parton-Level Jets

- We find the expected parton-level jet for a given hadron-level jet.
- Learn a mapping $f : \mathscr{H} \to \mathscr{P}$.

Samples and Simulation

Event Generation (PYTHIA 8.312)

- pp beams with $\sqrt{s} = 14$ TeV.
- Photon-tagged events $qg \to q\gamma$.
- $\widehat{p}_T > 1000$ GeV.
- Anti-k_t $R = 0.8$ parton-level and hadron-level jets.
- Visible final-state particles.
- $1000 <$ Jet $p_\perp < 2000$ GeV.
- 100K events to ensure sufficient statistics.

pythia.org/latest-manual/welcome.html

Graph Representation of Pythia Quark Jets

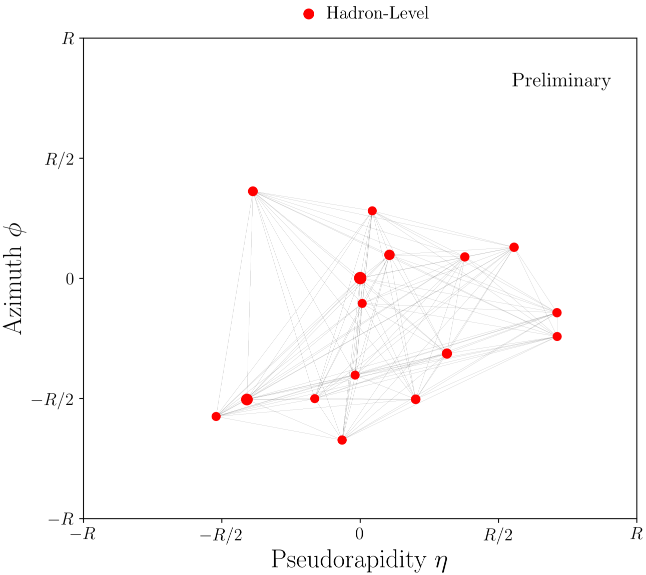
Jets represented as graphs, connected by ΔR:

$$\text{Vertices} : \mathcal{J} = \left\{ \left(p_\perp^i, \eta^i, \phi^i \right)_{i=1}^n \right\}$$

$$\text{Edges} : E = \left\{ \Delta R(i,j)_{i,j=1}^n, i \neq j \right\}$$

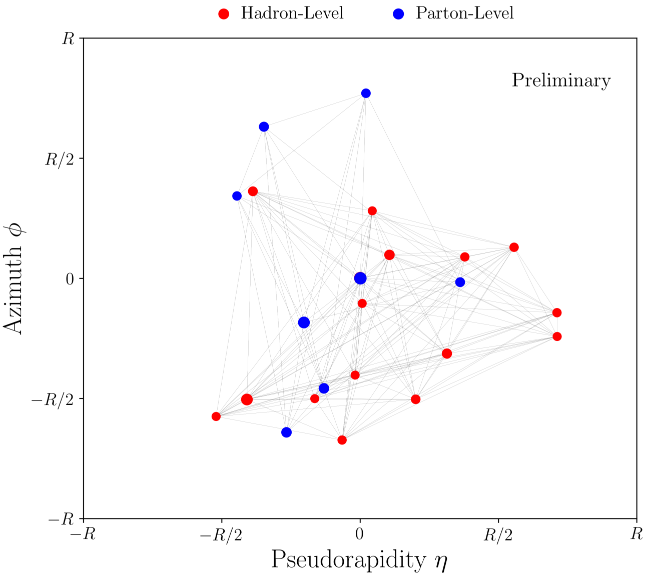
Fully connected graphs, no self-loops.

Graph Representations of Quark Jets



- Centered using the E-scheme axis.
- Pythia 8.312, pp $\sqrt{s} = 14$ TeV.
- Anti-k_T, $R = 0.8$
- $1000 < p_T < 2000$ GeV.

Graph Representations of Quark Jets



- Pythia 8.312, pp $\sqrt{s} = 14$ TeV.
- Anti-k_T, $R = 0.8$.
- $1000 < p_T < 2000$ GeV.

Machine Learning Model

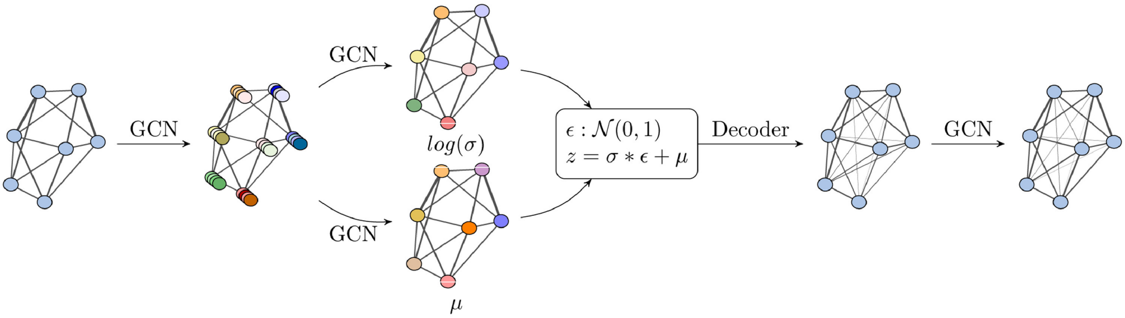


Image Credit: Tina Behrouzi et. al.

Variational Graph Autoencoder (VGAE)

- Input hadron-level jets $\mathcal{H}$.
- Output parton-level jets $\mathcal{P}$.

- Encoder: learns an embedding (z, μ) for $\mathcal{H}$ in latent space.
- Decoder: learns reconstructing parton-level jets $\mathcal{P}$ from embedding.

ML Model Training

Model Training

- Implemented using `PyTorch-geometric`.
- Trained using on an Nvidia A100.
- Train-validation split of 90%-10%.
- Adam optimizer, 5000 epochs.

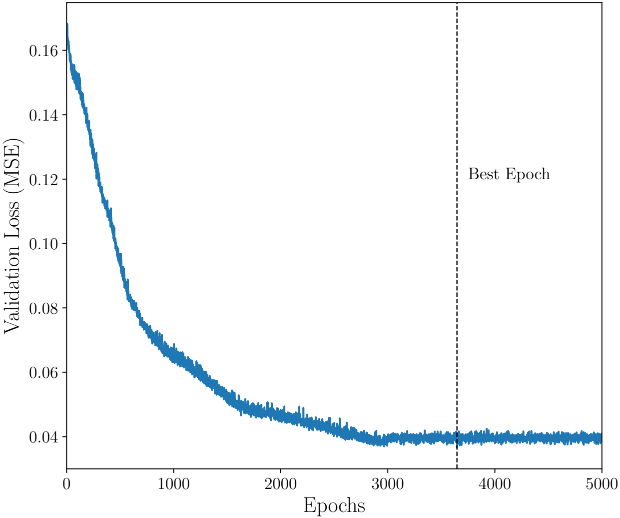
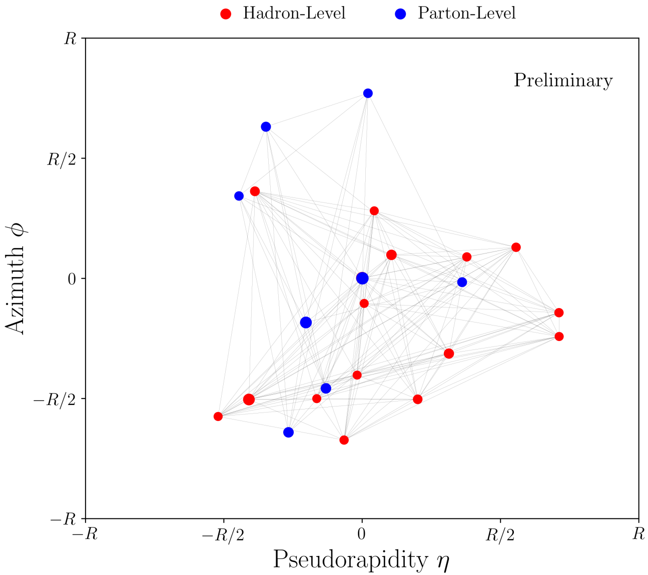


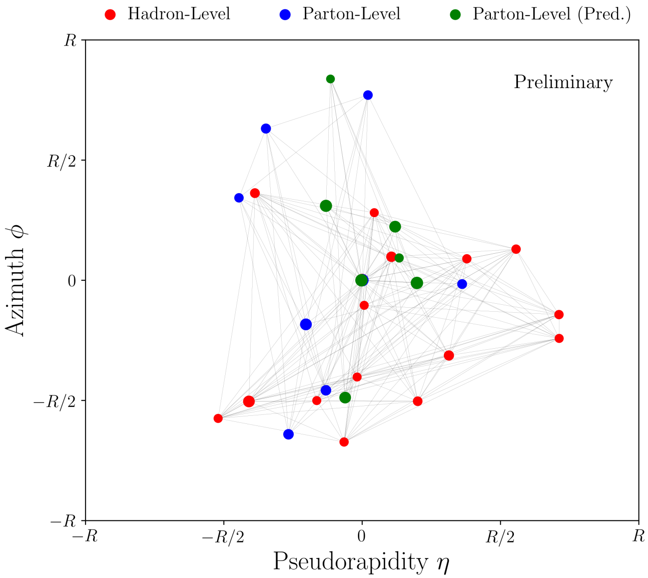
Figure 4: Validation loss over time.

Results: Predictions on Unseen Data



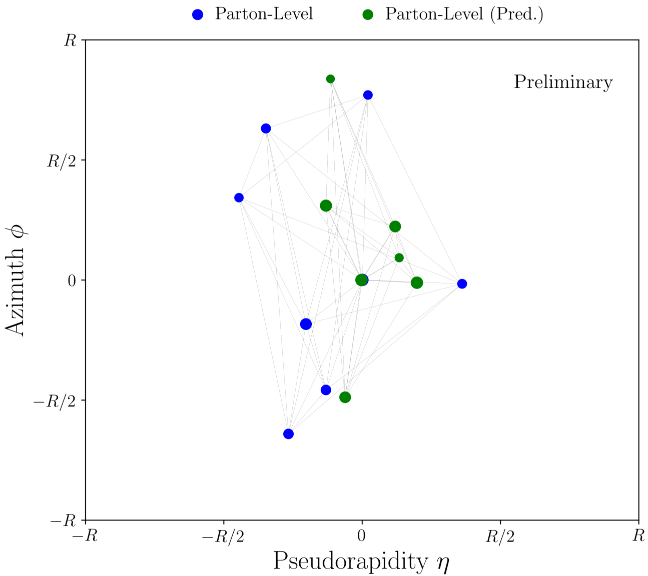
- Pythia 8.312, pp $\sqrt{s} = 14$ TeV.
- Anti-k_T, $R = 0.8$.
- $1000 < p_T < 2000$ GeV.

Results: Predictions on Unseen Data



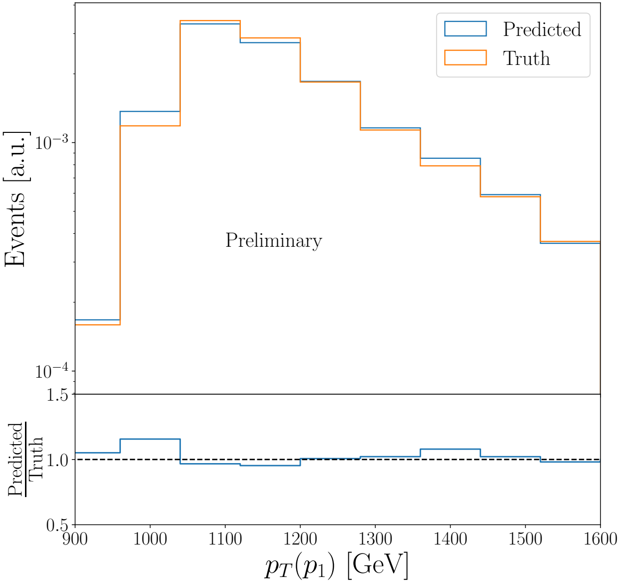
- Pythia 8.312, pp $\sqrt{s} = 14$ TeV.
- Anti-k_T, $R = 0.8$.
- $1000 < p_T < 2000$ GeV.

Results: Predictions on Unseen Data



- Pythia 8.312, pp $\sqrt{s} = 14$ TeV.
- Anti-k_T, $R = 0.8$.
- $1000 < p_T < 2000$ GeV.

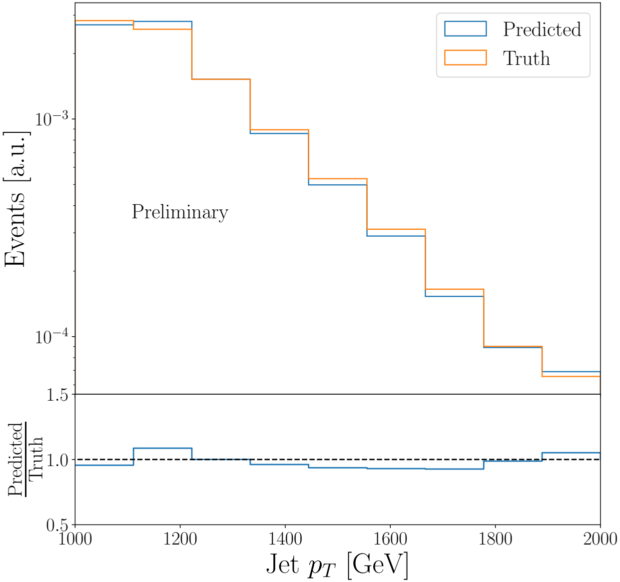
Results: Leading Parton p_T Spectra



Comparison of predicted and ground leading parton p_T spectra.

- Accurate prediction of $p_T(p_1)$.
- Almost all data $\varepsilon < 5\%$.

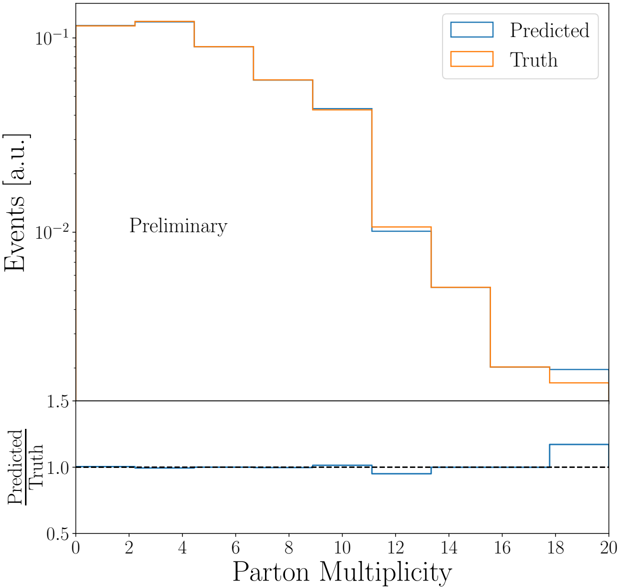
Results: Parton Jet p_T Spectra



Comparison of predicted and ground jet p_T spectra.

- Accurate prediction of jet p_T.
- Almost all data $\varepsilon < 5\%$.

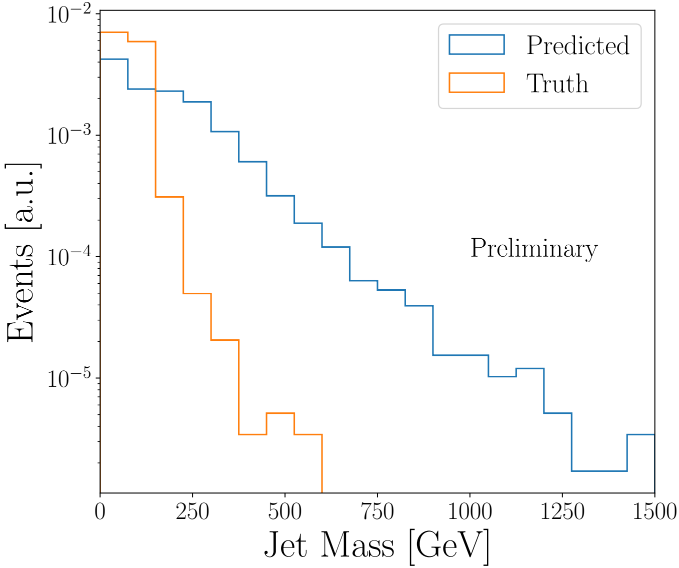
Results: Parton Multiplicity



Comparison of predicted and ground parton multiplicity spectra.

- Accurate prediction of multiplicity.
- Almost all data $\varepsilon < 5\%$.

Results: Parton Jet Mass Spectra



Comparison of predicted and ground jet mass spectra.

- Poor prediction of jet mass.
- Distribution is shifted and significantly more spread out.

Discussion and Conclusion

Summary

- We present a first look at using generative neural networks to reconstruct parton-level jets after hadronization.
- Our method captures the entire parton-level jet.
- Jury is still out on substructure observables.

Future Work

- Investigate the predictions for more jet substructure observables.
- Better ML models? Custom loss functions?
- Study the inclusion of detector effects, underlying events, and pileup.

Acknowledgements

- I thank Professor Rithya Kunnawalkam Elayavalli for their invaluable support and guidance.
- This work is supported in part by DOE Office of Science Award DE-SC0024660 and a Vanderbilt Immersion Grant.

Thank you! Questions?

Backup Slides

Preprocessing: Centering the Jets

- The (η, ϕ) coordinates of jet constituents are centered based on the jet (η, ϕ) using the E-scheme jet axis:

$$\overline{\eta} = \frac{\sum\limits_{i \in \text{jet}} \eta_i p_{T,i}}{\sum\limits_{i \in \text{jet}} p_{T,i}}, \quad \overline{\phi} = \frac{\sum\limits_{i \in \text{jet}} \phi_i p_{T,i}}{\sum\limits_{i \in \text{jet}} p_{T,i}}$$

$$\eta_i \to \eta_i - \overline{\eta}, \quad \phi \to \phi_i - \overline{\phi}$$

Comparing Jets: The Energy Mover's Distance Metric

EMD Metric (PRL 123.041801)

- Quantifies the distance between two jets.
- The minimum "energy" required to rearrange a jet $\mathcal{G}$ to $\mathcal{G}'$.

$$\mathcal{E}(\mathcal{G},\mathcal{G}') = \min_{\{f_{ij}\geq 0\}} \sum_{i=1}^{M}\sum_{j=1}^{M'} f_{ij}\left(\frac{\Delta R_{ij}}{R}\right) + \left|\sum_{i=1}^{M} E_i - \sum_{j=1}^{M'} E'_j\right|,$$

$$\sum_{j=1}^{M'} f_{ij} \leq E_i, \quad \sum_{i=1}^{M} f_{ij} \leq E'_j, \quad \sum_{i=1}^{M}\sum_{j=1}^{M'} f_{ij} = E_{\min},$$

$\mathcal{E}(\widehat{\mathcal{P}},\mathcal{P})$ gives a discrepancy measure between reconstructed graphs $\widehat{\mathcal{P}}$ and the ground truth $\mathcal{P}$.

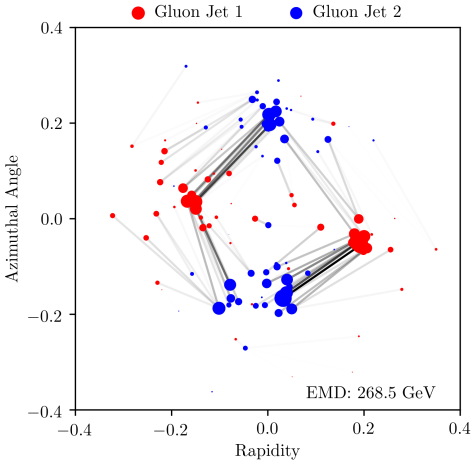
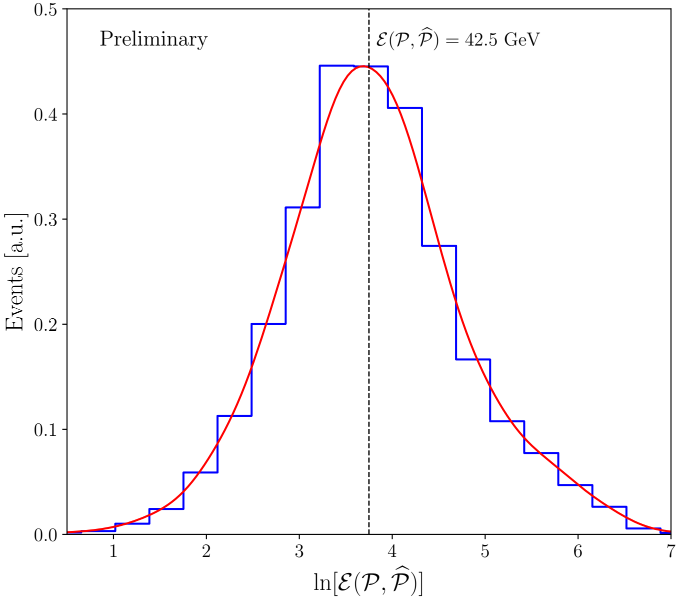


Figure 5: EMD between two gluon jets.

Results: EMD Metric Distribution

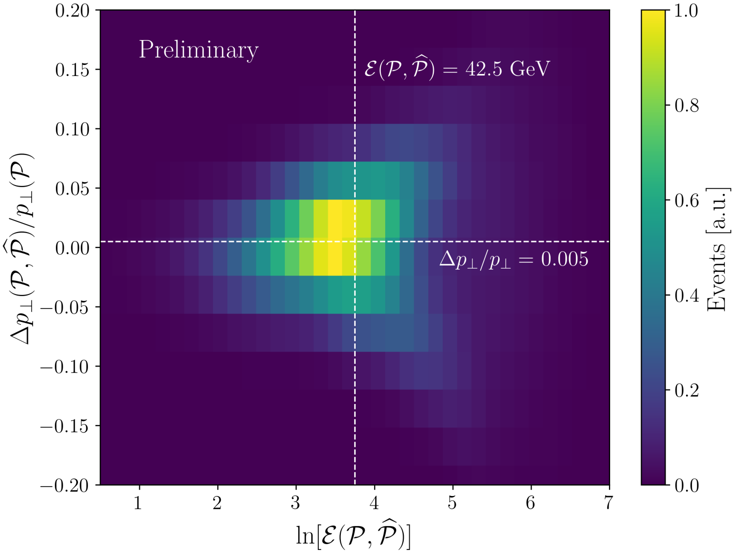


Predicted jets close to ground truth (Pythia)!

Benchmark EMDs:

- Good: $\ln \mathcal{E} \leq 4$
  - Jets are similar.
- Fair: $4 \leq \ln \mathcal{E} \leq 5.5$
  - Jets are fairly similar.
- Bad: $\ln \mathcal{E} \geq 5.5$
  - Jets are disparate.

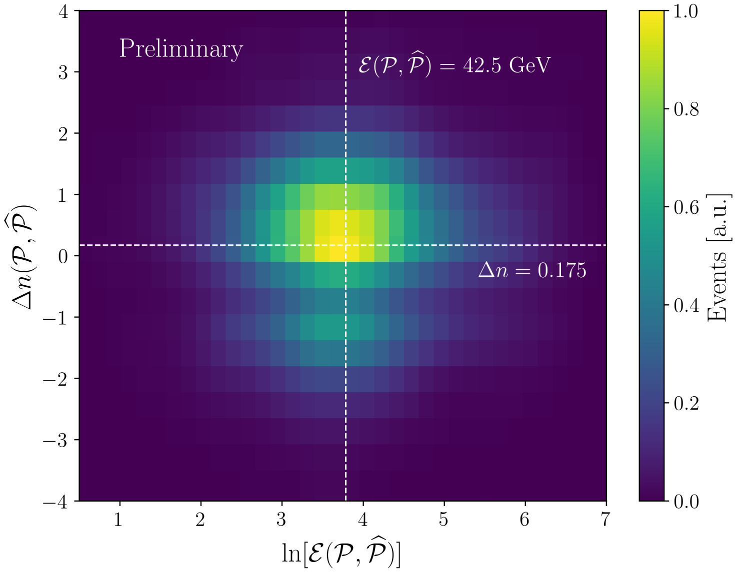
Results: Predictions for Parton Jet p_T



Correlation of EMD and the fractional difference in jet p_T.

- Accurate prediction of p_T.
- Peak at $\Delta p_T / p_T \approx 0$.
- Almost all data: $|\Delta p_T / p_T| < 0.1$.

Results: Predictions for Parton Multiplicities



Correlation of EMD and the difference in prediction and ground truth parton multiplicities.

- Accurate prediction of multiplicities.
- Average particle multiplicity is ~ 22.
- Peak at $\Delta n = 0$.
- Almost all data: $|\Delta n| < 2$.