GLOW: A Unified Particle Flow Transformer

Dmitrii Kobylianskii[2,*], **Samuel Van Stroud**[1,*], **Kwok Yiu Wong**[1], **Max Hart**[1], **Etienne Dreyer**[2], **Eilam Gross**[2], **Gabriel Facini**[1], **Tim Scanlon**[1]

[1]Centre for Data Intensive Science and Industry, University College London
[2]Faculty of Physics, Weizmann Institute of Science
[*]Authors contributed equally.

Abstract

We present GLOW, a transformer-based particle flow model that combines incidence matrix supervision from HGPflow with a MaskFormer architecture. Evaluated on CLIC detector simulations, GLOW achieves state-of-the-art particle flow reconstruction performance, improving jet energy resolution by 15%. Together with prior work, this demonstrates that a single unified transformer architecture can effectively address diverse reconstruction tasks in particle physics.

1 Introduction

Modern particle physics detectors produce sparse, multi-modal data that must be reconstructed into sets of particles. Particle flow (PFlow) algorithms address this challenge by combining reconstructed charged particle tracks from tracking detectors with calorimeter energy clusters to determine each particle's type, energy, and direction. Traditional algorithms like Pandora [1, 2] rely on expert-designed clustering, matching, and subtraction rules to avoid double counting and maximise resolution; these methods tend to be detector-specific and require extensive manual tuning for new geometries or operating conditions.

We frame particle flow reconstruction as an object-detection task and introduce GLOW, a transformer-based set-to-set PFlow model that achieves state-of-the-art reconstruction performance. The results demonstrate the effectiveness of unified transformer architectures for particle physics reconstruction.

2 Related Work

HGPflow [3–5] treats PFlow as a hypergraph prediction problem with supervised *incidence matrices* that associate detector objects (tracks or clusters) to output particles. For clusters, the matrix defines energy-conserving fractional assignments, from which *proxy* particle kinematics are calculated as weighted sums of associated cluster properties before refinement by a regression network.

MLPF [6–9] frames the problem as a direct, supervised mapping from inputs to particles. Rather than explicit set-to-set reconstruction, MLPF enforces one-to-one associations between input elements and particles, suppressing duplicates when a track and cluster originate from the same charged particle at the cost of resolution performance.

MaskFormers [10, 11], originally developed for image segmentation, have recently been adapted to HEP tasks such as vertexing and tracking [12, 13], where local attention enables scaling to high-occupancy detector data. MaskFormers process unordered input sets to simultaneously identify objects, assign constituents, and estimate object-level properties, making them well-suited for PFlow reconstruction, though this application has not been previously explored.

Machine Learning and the Physical Sciences Workshop, NeurIPS 2025.

Object condensation [14, 15] uses contrastive loss functions to cluster calorimeter energy deposits in latent space. Reconstructed particles are then obtained by aggregating the latent representations around a set of condensation points. This approach is suitable for scenarios where each input element can be unambiguously assigned to one particle.

3 Model

We introduce GLOW: a unified model combining incidence-based supervision from HGPflow, which enforces energy-consistent associations between detector objects and particles, with MaskFormer's query-based architecture, which naturally handles heterogeneous inputs and enables efficient, permutation-invariant decoding of variable-sized particle sets for PFlow reconstruction (Figure 1). After a six-layer self-attention encoder, learned queries predict unordered particle sets with flexible output cardinality via a four-layer masked cross-attention decoder [11]. Attention masks are dynamically adjusted based on learned similarities between particle queries and input detector objects. Following HGPflow, the model predicts an incidence matrix I_{ia} representing the fraction of object i's energy attributed to particle a, i.e. $I_{ia} = E_{ia}/E_i$, where E_i is the object's total energy. This ensures energy conservation by construction.

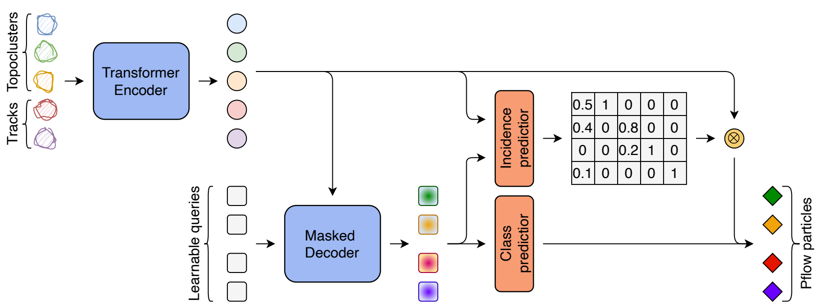

Figure 1: GLOW architecture overview. Dedicated mask prediction and regression modules are not depicted.

The input consists of fully reconstructed tracks and calorimeter topoclusters [16], each represented as a feature vector. A fixed number of learnable query embeddings represent particles. At each decoder layer, cross-attention operates between queries and input objects, with attention masks derived from learned similarity scores. Each particle query simultaneously predicts: (i) a soft assignment over inputs (approximating the ground-truth incidence matrix), (ii) particle-type labels (photon, electron, hadron), and (iii) kinematic quantities (energy, p_T, η, ϕ), computed first as incidence-weighted sums and then refined by a regression head. The training loss combines mask prediction, incidence matrix supervision, classification, and regression objectives. The Hungarian algorithm [17, 18] matches model predictions to ground truth targets for permutation-invariant training. This approach enables calibrated particle flow object reconstruction from charged tracks and calorimeter clusters, ensuring differentiability and scalability to future high-granularity detectors.

4 Experimental Setup

We use a dataset [19] of one million simulated $e^+e^- \to$ dijet events at $\sqrt{s} = 380$ GeV from the Compact Linear Collider (CLIC) detector [20, 21]. This dataset reflects the expected high-multiplicity, high-resolution conditions of future lepton colliders. The model input consists of fully reconstructed tracks and calorimeter clusters, produced by PandoraPFA [1], which applies traditional track finding, clustering, and iterative energy subtraction. Following [5], only particles that interact with the detector are included as targets—those with at least one associated track or nonzero calorimeter deposit within $|\eta| < 4$. Converted daughter particles are retained, but targets are defined at the parent level. Each particle is associated with detector objects through the energy-based incidence matrix detailed in [5].

GLOW has 12M parameters and was trained for 200 epochs over 23 hours on Isambard-AI [22] using two H100 GPUs. Inference is optimised for numerical precision rather than throughput, and reaches approximately 5k events/s per H100. Code is available on GitHub [23].

5 Results

For a fair architectural comparison, we retrained a publicly available MLPF model (originally trained on 21.6M events across multiple processes [6]) on our dataset, using the original preprocessing from Ref. [24]. The model architecture and hyperparameters were left unchanged. HGPflow results are taken from Ref. [5]. We evaluate performance on 20k independent dijet events using event- and jet-level metrics. Predictions for all models are processed using the same evaluation pipeline.

5.1 Event-level performance

Figure 2 shows event-level reconstruction performance across four key metrics. The top panels display residual distributions for missing transverse momentum $p_T^{\text{miss}} = |\sum_i \vec{p}_{T,i}|$ (left) and scalar transverse momentum sum $H_T = \sum_i |\vec{p}_{T,i}|$ (right), while the bottom panels show differences between predicted and true particle counts for charged (left) and neutral (right) particles. GLOW produces the most centered and narrow distributions for both momentum quantities, achieving a median of 0.42 and IQR of 1.43 for p_T^{miss} and a median of -0.005 and IQR of 0.061 for H_T, outperforming both HGPflow and MLPF. For particle counting, GLOW and HGPflow perform similarly. MLPF tends to underestimate the number of constituents. This likely arises from MLPF's target definition, which assigns each topocluster to a single leading particle and can therefore ignore subleading neutrals.

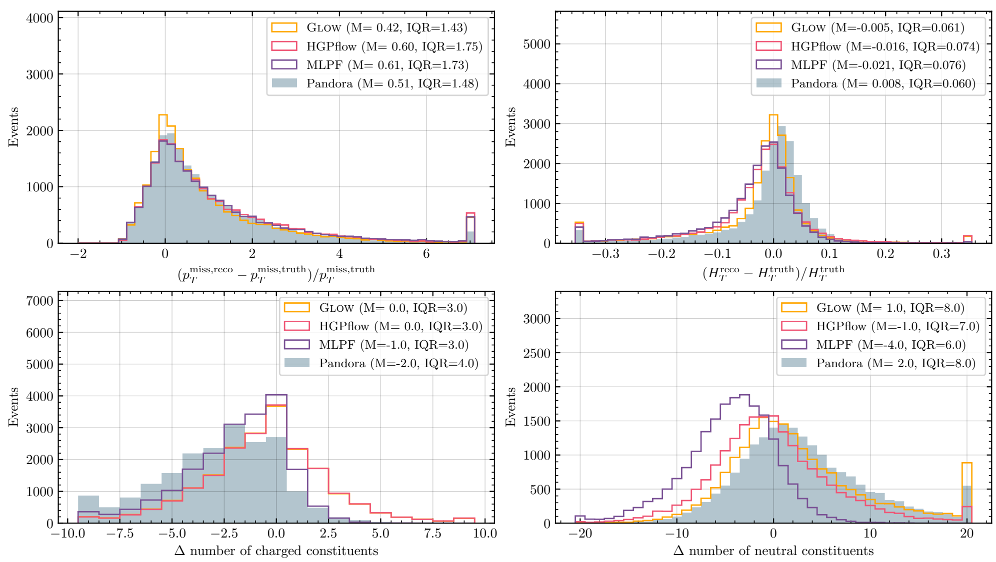

Figure 2: Residual distributions for missing transverse momentum p_T^{miss} (top left), scalar momentum sum H_T (top right), and counts for charged (bottom left) and neutral (bottom right) particles. The legend includes median and interquartile ranges for each model.

5.2 Jet-level performance

We reconstruct jets using FastJet with the generalized k_T algorithm ($R = 0.7$), keeping up to two leading jets with $p_T > 10$ GeV. Reconstructed jets are matched to truth jets within $\Delta R < 0.1$. Fig. 3 shows jet energy and momentum residuals along with angular separation. Both GLOW and HGPflow outperform MLPF and Pandora. GLOW produces the narrowest residual distributions and the smallest angular separations to truth jets. For jet p_T, GLOW achieves an IQR of 0.049, compared to 0.059 for HGPflow and 0.067 for MLPF. For jet energy, GLOW obtains an IQR of 0.052, versus 0.062 for HGPflow and 0.070 for MLPF. Fig. 4 demonstrates how jet energy resolution varies with truth jet energy. GLOW maintains the best median accuracy (within $\pm 2\%$) and consistently achieves a relative improvement of 15% in jet energy resolution over HGPflow across all energy ranges. MLPF shows comparable performance below 50 GeV but degrades at higher energies, as expected from the higher rate of overlapping particles.

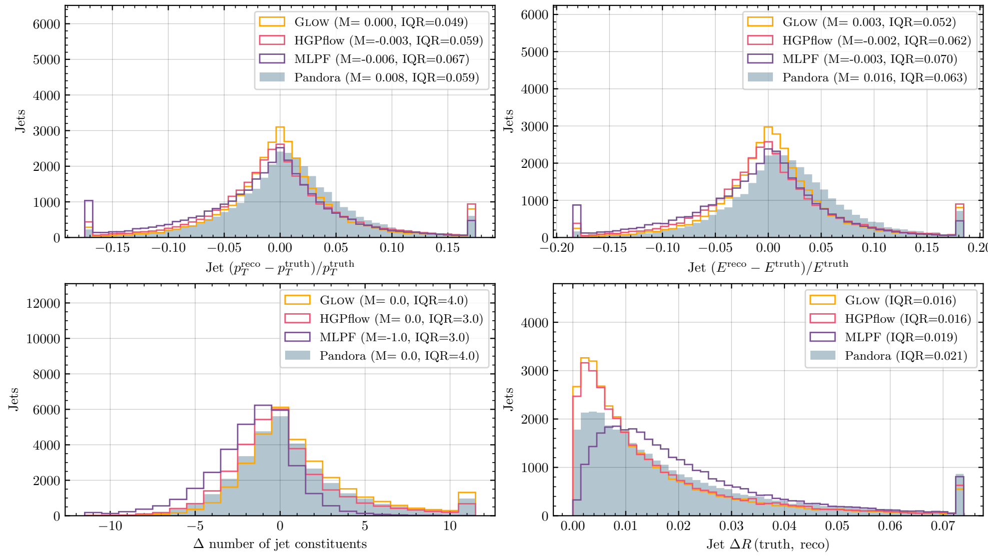

Figure 3: Relative residuals for jet transverse momentum p_T (top left), energy E (top right), number of jet constituents (bottom left), and angular separation ΔR between reconstructed and truth jets (bottom right). The legend includes median and interquartile ranges for each model.

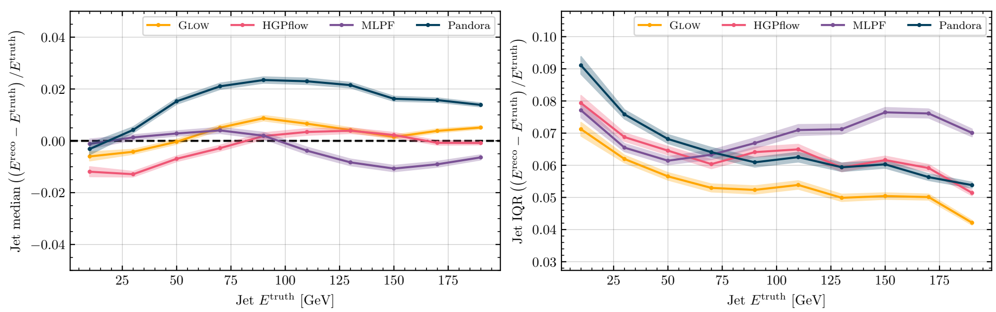

Figure 4: Median (left) and interquartile range (right) of jet energy relative residual distributions as a function of truth jet energy[1].

6 Conclusion

We present GLOW, a transformer-based particle flow reconstruction model that combines a MaskFormer architecture with incidence matrix supervision from HGPflow. GLOW improves on HGPflow by replacing iterative hypergraph refinement with a masked decoder in a fully transformer-based architecture. By avoiding backpropagation through time, training and inference are accelerated, while multi-head cross-attention provides greater expressiveness than incidence-weighted message passing. The use of standard transformer blocks also eases deployment with ONNX or TensorRT. Compared with MLPF, the physics-based incidence matrix formulation allows topoclusters to contribute to multiple particles while preserving energy consistency through fractional assignments.

Evaluated on CLIC detector simulations, GLOW achieves state-of-the-art reconstruction performance across event- and jet-level metrics, improving jet energy resolution by approximately 15% relative to HGPflow while reducing biases to within $\pm 2\%$. Together with prior work, we demonstrate that a unified encoder-decoder transformer can effectively handle core reconstruction tasks in particle physics, providing a flexible and scalable foundation for reconstruction at future colliders.

[1] The $\pm 1\sigma$ error bands are computed under the assumption of normal distributions using $\sigma_{\text{Median}} = 0.93 \cdot \text{IQR}/\sqrt{N}$ and $\sigma_{\text{IQR}} = 1.16 \cdot \text{IQR}/\sqrt{N}$.

Acknowledgments

We gratefully acknowledge the support of the UK's Science and Technology Facilities Council (STFC). S.V.S is supported by ST/X005992/1. M.H., K.Y.W, N.P. are supported by the STFC UCL Centre for Doctoral Training in Data Intensive Science (ST/W00674X/1) and by departmental and industry contributions. G.F and T.S. receives support from the STFC (ST/W00058X/1) and T.S. is supported by the Royal Society (URF/R/180008).

E.G., E.D., and D.K. are supported by the Minerva Stiftung through funding from the German Federal Ministry of Education and Research under grant number 715027, and the Weizmann Institute for Artificial Intelligence grant program.

We also extend our thanks to UCL for the use of their high-performance computing facilities, with special thanks to Edward Edmondson for his expert management and technical support.

The authors acknowledge the use of resources provided by the Isambard-AI National AI Research Resource (AIRR). Isambard-AI is operated by the University of Bristol and is funded by the UK Government's Department for Science, Innovation and Technology (DSIT) via UK Research and Innovation; and the Science and Technology Facilities Council [ST/AIRR/I-A-I/1023].

References

[1] J. S. Marshall and M. A. Thomson. "The Pandora Software Development Kit for Pattern Recognition". In: *Eur. Phys. J. C* 75.9 (2015), p. 439. DOI: 10.1140/epjc/s10052-015-3659-3. arXiv: 1506.05348 [physics.data-an].

[2] M. A. Thomson. "Particle Flow Calorimetry and the PandoraPFA Algorithm". In: *Nucl. Instrum. Meth. A* 611 (2009), pp. 25–40. DOI: 10.1016/j.nima.2009.09.009. arXiv: 0907.3577 [physics.ins-det].

[3] Francesco Armando Di Bello et al. "Reconstructing Particles in Jets Using Set Transformer and Hypergraph Prediction Networks". In: *The European Physical Journal C* 83.7 (July 11, 2023), p. 596. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-023-11677-7. URL: https://doi.org/10.1140/epjc/s10052-023-11677-7.

[4] Nilotpal Kakati, Etienne Dreyer, and Eilam Gross. "Denoising Graph Super-Resolution towards Improved Collider Event Reconstruction". In: (Sept. 2024). DOI: 10.48550/arXiv.2409.16052. arXiv: 2409.16052 [hep-ex]. URL: http://arxiv.org/abs/2409.16052.

[5] Nilotpal Kakati et al. "HGPflow: extending hypergraph particle flow to collider event reconstruction". In: *The European Physical Journal C* 85.8 (Aug. 2025). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-025-14443-z. URL: http://dx.doi.org/10.1140/epjc/s10052-025-14443-z.

[6] Farouk Mokhtar et al. "Fine-Tuning Machine-Learned Particle-Flow Reconstruction for New Detector Geometries in Future Colliders". In: *Physical Review D* 111.9 (May 29, 2025), p. 092015. ISSN: 2470-0010, 2470-0029. DOI: 10.1103/PhysRevD.111.092015. URL: https://link.aps.org/doi/10.1103/PhysRevD.111.092015.

[7] Joosep Pata et al. "Improved Particle-Flow Event Reconstruction with Scalable Neural Networks for Current and Future Particle Detectors". In: *Communications Physics* 7.1 (Apr. 10, 2024), pp. 1–12. ISSN: 2399-3650. DOI: 10.1038/s42005-024-01599-5. URL: https://www.nature.com/articles/s42005-024-01599-5.

[8] Joosep Pata et al. "Machine Learning for Particle Flow Reconstruction at CMS". In: *Journal of Physics: Conference Series* 2438.1 (Feb. 1, 2023), p. 012100. ISSN: 1742-6588, 1742-6596. DOI: 10.1088/1742-6596/2438/1/012100. URL: https://iopscience.iop.org/article/10.1088/1742-6596/2438/1/012100.

[9] Joosep Pata et al. "MLPF: Efficient Machine-Learned Particle-Flow Reconstruction Using Graph Neural Networks". In: *The European Physical Journal C* 81.5 (May 2, 2021), p. 381. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-021-09158-w. URL: https://doi.org/10.1140/epjc/s10052-021-09158-w.

[10] Bowen Cheng, Alexander G. Schwing, and Alexander Kirillov. "Per-Pixel Classification is Not All You Need for Semantic Segmentation". In: (2021). arXiv: 2107.06278 [cs.CV].

[11] Bowen Cheng et al. "Masked-attention Mask Transformer for Universal Image Segmentation". In: (2022). arXiv: 2112.01527 [cs.CV].

[12] Samuel Van Stroud et al. "Secondary Vertex Reconstruction with MaskFormers". In: *The European Physical Journal C* 84.10 (Oct. 8, 2024), p. 1020. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-024-13374-5. URL: https://doi.org/10.1140/epjc/s10052-024-13374-5.

[13] Samuel Van Stroud et al. *Transformers for Charged Particle Track Reconstruction in High Energy Physics*. 2024. arXiv: 2411.07149 [hep-ex]. URL: https://arxiv.org/abs/2411.07149.

[14] Jan Kieseler. "Object condensation: one-stage grid-free multi-object reconstruction in physics detectors, graph, and image data". In: (2020). arXiv: 2002.03605 [physics.data-an]. URL: https://arxiv.org/abs/2002.03605.

[15] Syed R. Qasim et al. "End-to-end multi-particle reconstruction in high occupancy imaging calorimeters with graph neural networks". In: *Eur. Phys. J. C* 82.8 (2022), p. 753. DOI: 10.1140/epjc/s10052-022-10665-7. URL: https://doi.org/10.1140/epjc/s10052-022-10665-7.

[16] ATLAS Collaboration. "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1". In: *Eur. Phys. J. C* 77 (2017), p. 490. DOI: 10.1140/epjc/s10052-017-5004-5. arXiv: 1603.02934 [hep-ex].

[17] H. W. Kuhn. "The Hungarian method for the assignment problem". In: *Naval research logistics quarterly* 2.1-2 (1955), pp. 83–97.

[18] Stefan Guthe and Daniel Thuerck. "Algorithm 1015: A Fast Scalable Solver for the Dense Linear (Sum) Assignment Problem". In: *ACM Trans. Math. Softw.* 47.2 (Apr. 2021). ISSN: 0098-3500. DOI: 10.1145/3442348. URL: https://doi.org/10.1145/3442348.

[19] Joosep Pata et al. *Simulated datasets for detector and particle flow reconstruction: CLIC detector*. Version 1.1. https://doi.org/10.5281/zenodo.8260741. Zenodo, Aug. 2023. DOI: 10.5281/zenodo.8260741. URL: https://doi.org/10.5281/zenodo.8260741.

[20] CLIC collaboration. "CLICdet: The post-CDR CLIC detector model". In: (2017). https://cds.cern.ch/record/2254048.

[21] Dominik Arominski et al. "A detector for CLIC: main parameters and performance". In: (Dec. 2018). https://arxiv.org/abs/1812.07337. arXiv: 1812.07337 [physics.ins-det].

[22] Simon McIntosh-Smith, Sadaf R. Alam, and Christopher Woods. "Isambard-AI: a leadership class supercomputer optimised specifically for Artificial Intelligence". In: (2024). DOI: 10.48550/arXiv.2410.11199. URL: https://doi.org/10.48550/arXiv.2410.11199.

[23] Samuel Van Stroud. *hepattn: Attention for High Energy Physics*. Version 0.2.0. DOI: 10.5281/zenodo.16880940. URL: https://github.com/samvanstroud/hepattn/releases/tag/clic-paper.

[24] Joosep Pata et al. *jpata/particleflow: v2.3.0*. 2025. DOI: 10.5281/ZENODO.14930299. URL: https://zenodo.org/doi/10.5281/zenodo.14930299.