Denoising graph superresolution toward improved collider event reconstruction

Nilotpal Kakati,* Etienne Dreyer, and Eilam Gross
Weizmann Institute of Science, Rehovot, Israel

(Received 23 March 2026; revised 10 August 2026; accepted 16 September 2026; published 28 October 2026)

In preparation for Higgs factories and energy-frontier facilities, future colliders are moving toward high-granularity calorimeters to improve reconstruction quality. However, the cost and construction complexity of such detectors is substantial, making software-based approaches like superresolution an attractive complementary tool. This study explores integrating superresolution techniques into an LHC-like reconstruction pipeline to effectively enhance calorimeter granularity and suppress noise. We find that this software preprocessing step significantly improves reconstruction quality at fixed detector granularity, narrowing the performance gap to a finer calorimeter without closing it. To demonstrate its impact, we propose a novel transformer-based particle-flow model that offers improved particle reconstruction quality and interpretability. Our results serve as a proof of concept, demonstrating the potential of superresolution to enhance reconstruction in collider environments.

DOI: 10.1103/2bjm-w5yq

I. INTRODUCTION

The energy and precision frontiers of high-energy physics are driven by technological advances both in generating particle beams and in recording and reconstructing particle collisions. Detectors at current and future collider facilities are designed to extract precise information about outgoing particles using extensive arrays of sensors, including calorimeter layers. Calorimeters measure energy by instigating a shower of secondary particles, distributing and absorbing the incident particle's energy as it propagates through successive layers of dense material. To help determine particle direction and type, as well as separate nearby particles, calorimeters are segmented into a gridlike array of cells.

Although calorimeter performance is generally characterized by energy resolution, spatial resolution (inversely related to granularity) assumes a crucial role in dense collision environments. At the Large Hadron Collider (LHC) [1] proton-proton collisions yield dense environments rich in hadronic jets and numerous simultaneous collisions (pileup), with a nearly tenfold increase expected in future high-luminosity runs [2]. To reconstruct jet properties accurately, particle-flow (PFlow) algorithms [3,4] leverage tracking information in tandem with calorimeter and other measurements. However, low-granularity calorimeters have limited imaging capabilities, constraining the performance of pattern reconstruction in PFlow algorithms such as assigning energy deposits to tracks, identifying neutral hadrons, and employing software compensation techniques [5]. These challenges have positioned *particle-flow calorimetry* [6] as a leading design paradigm for future detectors, enabling the 3% jet energy resolution required at future electron-positron Higgs factories to distinguish hadronic W and Z boson decays [7].

In practical applications, calorimeter granularity is constrained by cost as well as mechanical and electrical engineering challenges, leading to less than optimal granularity. Superresolution (SR) techniques, which have been extensively studied in the field of image processing [8–14], offer the potential to mitigate the limitations of the spatial resolution intrinsic to the calorimeter design by imposing physical priors learned from simulation. By reconstructing high-resolution (HR) data from low-resolution (LR) inputs, these techniques could, in essence, recover substructure features obscured by coarse segmentation, thereby maximizing the performance of a given detector design.

In this study, we explore the integration of superresolution techniques within the core pipeline of particle reconstruction in collider experiments. The initial stage of any collider experiment involves reconstructing low-level particles or particlelike objects, referred to as PFlow objects. By applying superresolution to calorimeter data, we aim to enhance the accuracy of PFlow objects and, thereby, that of higher-level objects used in physics analyses, such as jets, leptons, and photons. We anticipate our approach to be especially useful for reconstructing

*Contact author: nilotpal.kakati@weizmann.ac.il

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International 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.

boosted topologies such as top jets, boosted Higgs decays, and hadronically decaying tau leptons.

A. Related work

Calorimeter SR was first explored in [15], which demonstrated a proof of concept using a neural network-based method with a simplified detector and physics setup. Since then, SR has found various applications in particle physics, such as enhancing jet constituent images with generative adversarial networks for improved jet property estimation [16], employing normalizing flows for upsampling calorimeter showers for fast simulation [17], and producing high-resolution photon shower images [18]. In the broader context, a large number of generative models have been developed for fast calorimeter shower simulation (see [19] for a review). Recent work [20] also highlights the potential of SR for enhancing event reconstruction in neutrino telescopes.

B. Key contributions

Our approach offers two enhancements to calorimeter data to benefit the downstream reconstruction task. The first is the suppression of electronic noise that distorts recorded energy deposits, which we perform by training our model to denoise the energy in each cell. Second, our approach improves spatial resolution by predicting a high-granularity graph of calorimeter cells, allowing the reconstruction of finer details such as nearby particles. Representing the calorimeter data as a graph is a natural choice for handling irregular geometry, nontrivial spatial correlations, and the sparsity in the recorded signal. Superresolution on graph data, where the number of nodes gets upscaled, is not an entirely new concept [21,22] but is uncommon in the literature. To assess the improvement in reconstruction, we introduce a novel transformer-based PFlow model, based on predicting hypergraphs, as introduced in [23]. This model benefits from modernized architecture and provides improved interpretability, allowing the impact of superresolution to be analyzed transparently.

We summarize our contributions as follows:

(i) demonstrating superresolution with a full particle reconstruction chain,
(ii) denoising energy deposits,
(iii) graph-based architecture, and
(iv) novel transformer-based particle-flow algorithm.

II. DATASETS

To train the superresolution model, we require the same calorimeter shower data in both low and high granularity (i.e., resolution). This was achieved using the COCOA package [24], which provides an LHC-like generic calorimeter model simulated with Geant4 [25–27]. COCOA includes three electromagnetic (ECAL) and three hadronic (HCAL) concentric layers, each with configurable material, granularity and noise levels. The package emulates sampling fractions of $f = 0.07$ and $f = 0.025$ in the ECAL and HCAL, respectively, by using an effective molecule that represents the mixture of active (liquid Ar and plastic scintillator) and passive (Pb and Fe) materials. This is done by randomly dropping a fraction f of the shower segments and then scaling up the response in each cell by $1/f$. To focus solely on the calorimeter response, all upstream material, such as the inner tracker and solenoid iron, was removed, and the magnetic field was turned off, allowing showers from charged particles to propagate radially.

In this study, we apply superresolution exclusively in the pseudorapidity (η)-azimuthal angle (ϕ) plane, keeping the calorimeter layers unchanged. We use two datasets: the first is a simplified set where a single electron is fired with $p_T \in [50, 51]$ GeV, $\eta \in [-0.01, 0.01]$, and $\phi \in [-\pi, \pi]$. The second dataset involves multiple particles with angular separation on the scale of the calorimeter cells. The particles are arranged around a principal axis chosen uniformly in $\eta \in [-2.5, 2.5]$ and $\phi \in [-\pi, \pi]$. In three out of four cases, an electron is fired along the principal axis with $p_T \in [20, 50]$ GeV. In addition, N photons are fired with $p_T \in [5, 25]$ GeV, where N is chosen randomly from 0 to 3 (and required to be nonzero in case of no electron). Each photon is displaced from the principal axis by $\pm\text{Gaus}(4 \cdot \delta, \delta)$, where the sign is chosen randomly and δ is the angular width of a cell in the high-resolution ECAL1. For the single-electron dataset, the cell resolution is downsampled by factors, $f_\eta = f_\phi = 2$, and for the multiparticle dataset, by $f_\eta = f_\phi = 4$. Table I details the noise levels and granularity used for each calorimeter layer in this study.

The particles interact with the high-resolution detector, initially with zero noise. The high-resolution cells are then combined to form low-resolution cells, to which noise is applied to simulate electronic noise. This procedure is illustrated in Fig. 1. Subsequently, the built-in topological clustering algorithm in COCOA, which resembles that of ATLAS, is used to cluster calorimeter hits based on energy-over-noise ratio thresholds. Only clustered cells with positive energy values are retained. Since our focus is

TABLE I. Noise levels and granularity of the three electromagnetic calorimeter layers in both low and high resolution. The noise is only included for the low-granularity cells.

Layer		ECAL1	ECAL2	ECAL3
σ_{noise} [MeV]	(Low)	13	34	41
η, ϕ segmentation	(Low)	128	128	64
single electron	(High)	256	256	128
η, ϕ segmentation	(Low)	64	64	32
multiparticle	(High)	256	256	128

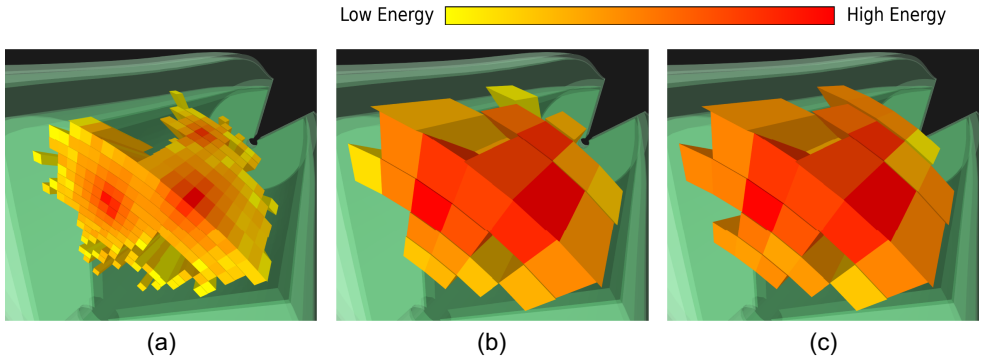


FIG. 1. Phoenix event display [28] of one electron and two photons in COCOA, demonstrating the dataset creation procedure. Here we see (a) the truth high-resolution configuration, (b) the truth low-resolution configuration, and (c) the measured low-resolution configuration which also contains noise.

on electromagnetic showers from electrons and photons, we limit our study to the ECAL layers.

III. METHODS

A. Continuous normalizing flows

Continuous normalizing flow [29] is a type of generative model that transforms a simple base probability distribution $p_0(x_0)$ into a more complex target distribution $p_1(x_1)$ through a continuous sequence of invertible transformations. This transformation is governed by the following neural ordinary differential equation:

$$\frac{dx}{dt} = v_\theta(x,t), \tag{1}$$

where v_θ is the vector field parametrized by a neural network and $t \in [0,1]$. Flow matching, as proposed in [30], is a technique for training continuous normalizing flows that aims to match the learned flow $v_\theta(t,x)$ induced by the differential equations of the model with a target flow $u_t(x)$. The flow-matching objective can be defined as follows:

$$L_{FM}(\theta) = E_{t,p_t(x)} \|v_\theta(t,x) - u_t(x)\|^2. \tag{2}$$

The objective, however, is intractable because the target flow u_t, that generates the desired probability paths between p_0 and p_1, is typically not available in closed form. As a solution, [30] also proposed a conditional flow-matching objective where u_t and p_t can be constructed in a sample-conditional manner:

$$L_{CFM}(\theta) = E_{t,q(z),p_t(x|z)} \|v_\theta(t,x) - u_t(x|z)\|^2, \tag{3}$$

where $t \in [0,1]$, $z \sim q(z)$, and $x \sim p_t(x|z)$. L_{CFM} is fully tractable and the authors showed that optimizing it is equivalent to optimizing the L_{FM} objective. The probability path p_t and the target flow u_t can be defined in various ways [30,31]. For our studies, we use the formulation from the original paper [30]:

$$\begin{aligned} p_t(x|z) &= \mathcal{N}(x|tx_1, (t\sigma - t + 1)^2), \\ u_t(x|z) &= \frac{x_1 - (1-\sigma)x}{1-(1-\sigma)t}. \end{aligned} \tag{4}$$

This defines a straight path between the standard normal and a Gaussian distribution of width σ around x_1. We chose σ to be 10^{-4} following the original paper. We used the TORCHDIFFEQ [32] library with the Dormand-Prince (dopri5) solver for generation.

B. Superresolution model

The core concept of our superresolution model is inspired by the work [14]. To allow for irregular cell geometry and to cope with the sparsity of the recorded energy deposits, we represent the data as graphs. Each low-resolution graph is expanded into a high-resolution graph, where each low-resolution cell is divided into $f_\eta \times f_\phi$ high-resolution cells. For each high-resolution cell, the target is its truth energy value, while the corresponding low-resolution energy serves as the conditional input, as illustrated in Fig. 2. Although nearest-neighbor edges among nodes in the high-resolution graph would, in general, be motivated for larger graphs, we model them as fully connected for simplicity.

Our model employs a transformer-based architecture to predict the vector field v_t, as illustrated in Fig. 3. The process begins by upsampling the low-resolution graph through repetition, which serves as the conditional input. Subsequently, a copy of the high-resolution graph, initialized with random values x_0, is created. These two high-resolution graphs are then stacked, providing each node with both the upsampled information and the randomly initialized values. This combined input is fed into a transformer network, generating a neighborhood-aware representation that aids in understanding the energy deposition profile of the incident particles. Following this, a multilayer perceptron (MLP) predicts the vector field v_t for each node. To advance the graph, we integrate v_t over a

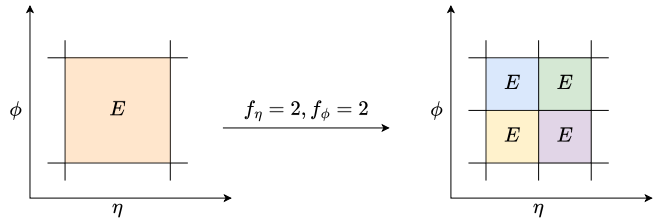


FIG. 2. Creation of high-resolution cell in our approach. Each low-resolution cell is split in η and ϕ by the superresolution factor f_η and f_ϕ, respectively. They get their new angular coordinates and the parent low-resolution energy (E) is copied to serve as the conditional inputs. Here, we see the procedure for $f_\eta = 2$ and $f_\phi = 2$. Each cell is a node in the graph, but, for illustrative purpose, we are denoting them as images.

small time interval Δt, yielding $x_{0+\Delta t}$. This updated graph $x_{0+\Delta t}$ is then used as input, along with the conditional information, for the subsequent transformer iteration, and the process of predicting v_t is repeated. Finally, at $t = 1$, the output x_1 represents the denoised, superresolved graph.

The hyperparameters used during training are summarized in Table II. For the single-electron study, the training, validation, and test datasets consisted of 1 000 000, 3000, and 10 000 events, respectively. In the multiparticle study, we utilized 250 000 events for training, 3000 for validation, and 10 000 for testing.

C. Ensemble sampling

Generation with continuous normalizing flow is inherently a stochastic process. While this randomness is often beneficial in applications like image generation, where variation is desired, it poses a challenge in the superresolution of calorimeter data, where there is a specific deterministic target. Ideally, the network's output should fluctuate around the true target if functioning correctly. To address this, we employ an ensemble approach. Unlike typical ensemble methods that involve training multiple copies of the model, our approach simply involves passing the same conditional input through the same model multiple times to produce the ensemble output, which is then averaged to obtain the final prediction. This technique helps reduce the impact of fluctuations and brings the output closer to the true target. However, it is important to acknowledge that due to the stochastic noise inherently added by the calorimeter, recovering the exact true energy is not possible; we can only approximate it.

D. Particle-flow model

Machine learning-based particle-flow models for LHC-like environments have been extensively explored [23,34–42]. Among these, *HGPflow* [23,42] distinguishes itself by explicitly learning the energy assignment matrix

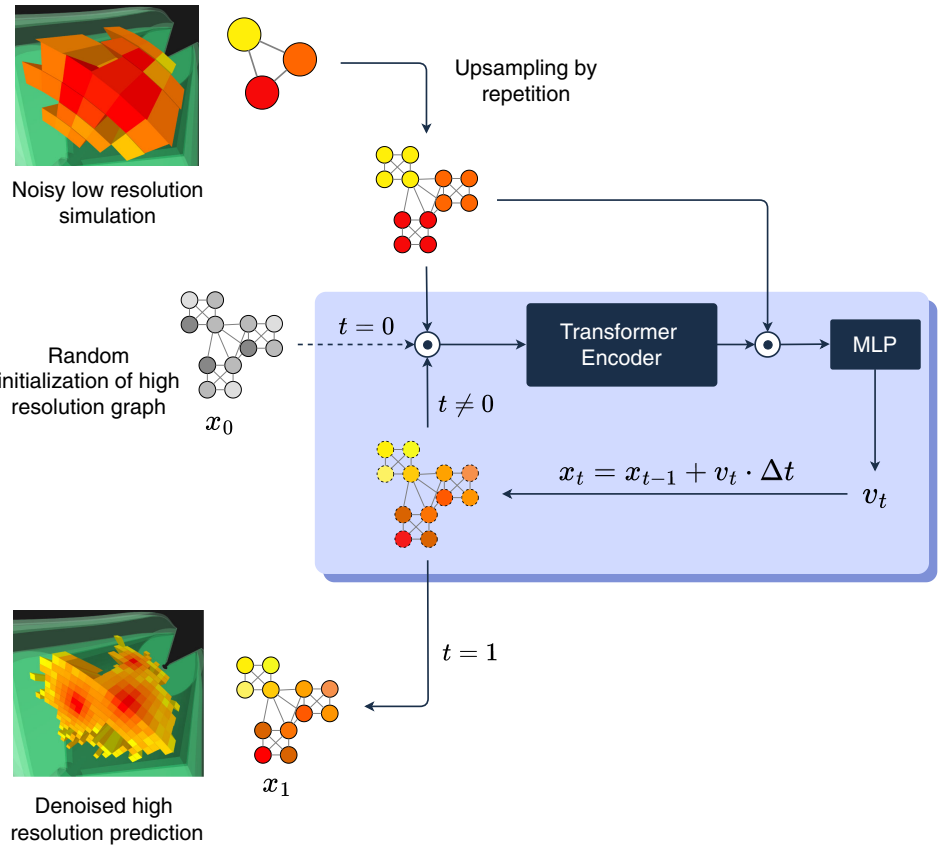


FIG. 3. The superresolution architecture. ⊙ indicates concatenation.

TABLE II. Hyperparameters used in the superresolution model.

Hyperparameter	
Batch size	25
Optimizer	AdamW [33]
Learning rate	0.001
Number of epochs	100
Number of time steps	25
Number of transformer layers	4
Number of trainable parameters	4.2M

between inputs and outputs rather than treating the reconstruction as a black-box regression. This approach offers a high degree of interpretability, as it allows for the direct inspection of how recorded energy in the detector is assigned to reconstructed particles. To better understand the impact of superresolution, we aimed for a similarly interpretable model. Here, we present a novel particle-flow architecture that builds on the principles of HGPflow but replaces the *iterative learning* approach with a *supervised attention* mechanism.

PFlow algorithms usually combine track and calorimeter inputs. However, since we are primarily interested in studying the impact of calorimeter SR, we do not include tracks in the input. To indicate this, we refer to our model as PFlow$^-$, since the architecture itself can be used in the same capacity as HGPflow.

Figure 4 illustrates the high-level view of the architecture. The model processes detector data from cells using a transformer network to generate a neighborhood-aware representation. The transformer's output is processed by two branches. One branch predicts the particle cardinality using an MLP, which treats this prediction as a classification task rather than regression. While predicting cardinality as a global metric is not the most robust approach, and many existing models [23,38] handle this as a padding problem (predicting a flag for each particle to indicate if it's real or padded), we opted for a simpler classification approach given our lower cardinality and simpler environment.

The second branch focuses on kinematic prediction. It achieves this by learning the fractional energy distribution of each cell over the particles as shown in Eq. (5), where E_{ai} represents the energy deposited by particle a in cell i. This distribution, referred to as the incidence matrix in HGPflow [23], was originally defined for topological clusters rather than individual cells. By using cells directly, we retain more detailed information, as topological clusters aggregate the features of multiple cells, potentially losing fine-grained details. Working with cells increases the complexity, since the cardinality can be an order of magnitude or more above that of clusters. In exchange it eliminates the dependency on topoclustering, which is not guaranteed to be an optimal algorithm. We implement this approach using an attentionlike framework, where particles and cells are processed through two MLPs to obtain the key (K) and query (Q), respectively. The resulting attention weights are computed as usual [Eq. (6)] and supervised to match the target incidence matrix. Given that we operate on sets, permutation invariance of particles is essential, and we achieve this using Hungarian matching [43]:

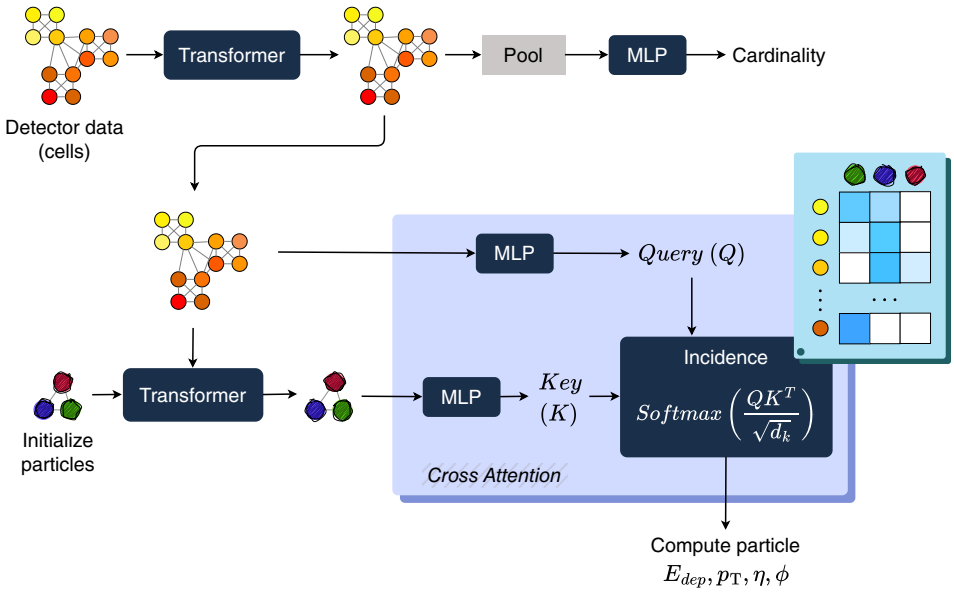


FIG. 4. The PFlow$^-$ architecture. The core concept is similar to [23], but instead of learning the incidence matrix through iterative refinements, we predict it as a standard cross attention.

$$\text{target incidence,} \quad I_{ai} = \frac{\text{energy deposited by particle } a \text{ in cell } i}{\text{total energy deposited in cell } i} = \frac{E_{ai}}{\sum_a E_{ai}}, \tag{5}$$

$$\text{predicted incidence} = \text{attention weights} = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_K}}\right). \tag{6}$$

Once the energy-based incidence matrix is determined, the *proxy* kinematics of the reconstructed particles can be calculated from the associated cells using the approach outlined in [23]:

$$\hat{E}_a = \sum_i (I_{ai} \cdot E_i); \qquad \hat{p}_{\mathrm{T}_a} = \frac{\hat{E}_a}{\cosh \hat{\eta}_a} (\text{zero mass}),$$
$$\hat{\eta}_a = \frac{1}{\hat{E}_a} \sum_i (I_{ai} \cdot E_i \cdot \eta_i); \qquad \hat{\phi}_a = \frac{1}{\hat{E}_a} \sum_i (I_{ai} \cdot E_i \cdot \phi_i). \tag{7}$$

In HGPflow [23], these proxy properties are further refined using MLPs, a step we omit in our default configuration. Keeping the kinematics an explicit function of the incidence matrix and the cell coordinates ensures that any difference in performance between the two resolutions can be attributed to the input cells rather than to a learned correction. This choice makes the comparison transparent, but it also ties the angular resolution to the granularity of the input, since the predicted position can only be expressed as an energy weighted combination of cell centers. To quantify the effect of this restriction, we additionally train a variant in which the proxy kinematics are refined by a correction network, following [23]. The correction is implemented as a residual on top of Eq. (7). It takes the full per-particle representation as input, and is initialized to zero, so that training begins from the uncorrected estimator. We apply it to both the low- and the high-resolution models and report the results.

We trained two identical models using the same set of hyperparameters: one with the low-resolution measured cells and the other with the high-resolution predicted cells. The particle-flow models were trained on the superresolution predictions for the 250 000 training events, and all results are reported on a separate sample of 10 000 events that was not used to train either the superresolution model or the particle-flow models. For the incidence loss, we followed a procedure similar to that in [23], pairing truth and predicted particles using the Hungarian algorithm with the Kullback-Leibler divergence between their incidence values as the metric. Cardinality loss was handled using cross-entropy, and the sum of the incidence and cardinality losses was used for optimization. The same procedure was used for the variants with the correction network, with the addition of per-particle kinematic terms for the angular coordinates and the energy, without which the correction network receives no gradient. Their weights were chosen so that each term contributes comparably at the start of training, and these models were initialized from the corresponding uncorrected model and trained for a further 60 epochs. The hyperparameters used for training are summarized in Table III.

TABLE III. Hyperparameters used in the particle-flow model.

Hyperparameter	
Batch size	30
Optimizer	AdamW [33]
Learning rate	0.001
Number of epochs	100
Number of transformer encoder layers	3
Number of transformer decoder layers	4
Number of trainable parameters	333K (+19K)

IV. RESULTS

A. Single electron

Figure 5 provides a qualitative view of superresolution applied to a single-electron event. The predicted high-resolution cell energies closely align with the true values, and the time evolution of these predicted energies is also shown. Given that the target definition involves a correction to the input energy, the $t = 0$ state is not entirely random but instead closely reflects the conditional input. The network also manages to suppress noise as seen from the figure.

To compare our generative architecture with a deterministic one, we trained a variant with a direct regression objective rather than a flow-matching objective. The regression model has an identical setup except that the time step and current state embedding inputs are dropped. Figure 6 shows qualitatively how the regression model's output compares to the flow model output. While the flow model attempts to generate the local fluctuations in energy, the regression model produces a smooth energy distribution. For reference, we include an evaluation of the shower mean computed over all truth HR examples, which behaves similarly to the regression model output.

To quantitatively assess the impact on reconstruction, we estimate the electron's energy by summing the energies of all cells in the event. We examine the energy residuals using both the measured low-resolution cell energies and the predicted high-resolution cell energies, with the target being the sum of the true high-resolution cell energies. Figure 7(a)

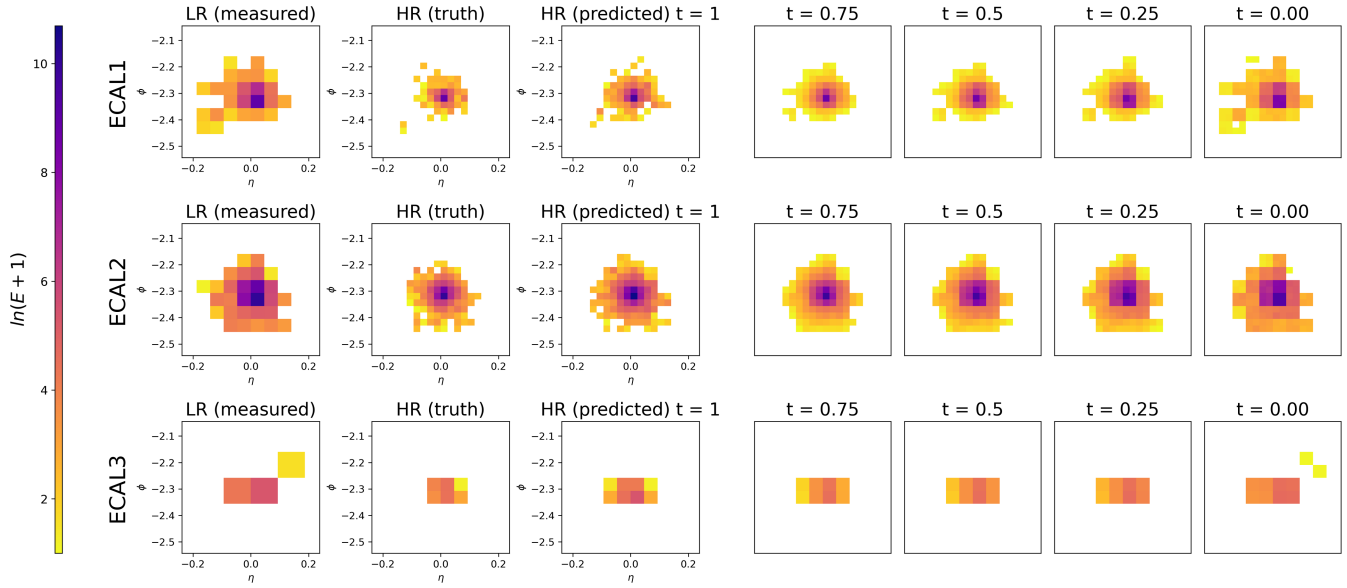


FIG. 5. Superresolution applied to a single-electron event from the test set. The three rows depict the superresolution process for the three ECAL layers. For each layer, the measured LR cell energy distribution, the target HR cell energy distribution, and the predicted HR are shown. The predicted HR cells closely match the truth HR ones. Additionally, the time evolution of the cell energy distribution is also illustrated.

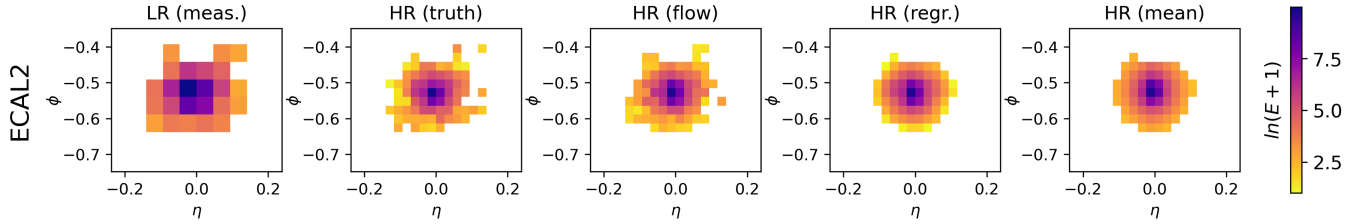


FIG. 6. Comparison of flow and regression model outputs for superresolution on a single-electron event from the test set.

illustrates these residual distributions. As discussed in Sec. III C, the predicted high-resolution energies result from ensemble sampling, with an ensemble size of 10 in this instance. The figure shows that the high-resolution residual is much closer to zero and exhibits a narrower spread. The peak shift in the low-resolution distribution is due to the asymmetric noise contribution, as we retain only the cells with positive energy. The network effectively adapts to this asymmetric noise, demonstrating an improvement in energy resolution. The performance of the regression model is also shown by the dashed line in Fig. 7(a). Its tendency to predict the marginal distribution is reflected in the wider residual distribution. The negative shift reflects its attempt at removing positively biased noise without the capability of producing local fluctuations.

We also examine how the performance depends on the ensemble size. As expected, the position of the residual distribution remains unchanged regardless of the ensemble size, and the width of the distribution varies. Figure 7(b) illustrates how the 1σ width of the residual distribution decreases as the ensemble size increases, with the improvement saturating above an ensemble size of 20. For all our studies, both for the single-electron sample and the multiparticle sample, we opted for an ensemble size of 10, to keep the computational costs lower while still benefiting from the advantages of ensemble sampling.

In addition, we analyzed the energy correlator observables C_2, C_3, and D_2 [44,45] at the cell level for both high-resolution and low-resolution data. Unlike the total energy, which is a sum over cells and therefore unchanged by downsampling, these observables depend on the spatial distribution of the energy and are sensitive to granularity. The upper row of Fig. 8 presents the resulting distributions, with lower values indicating more substructure. The predicted high-resolution cells closely match the true distribution and offer significant improvements over the noisy low-resolution data across all three substructure observables considered. The lower row shows the per-event residuals, which for the predicted high-resolution cells are narrower than the spread of the observables themselves, so the prediction follows the individual event rather than reproducing the ensemble distribution.

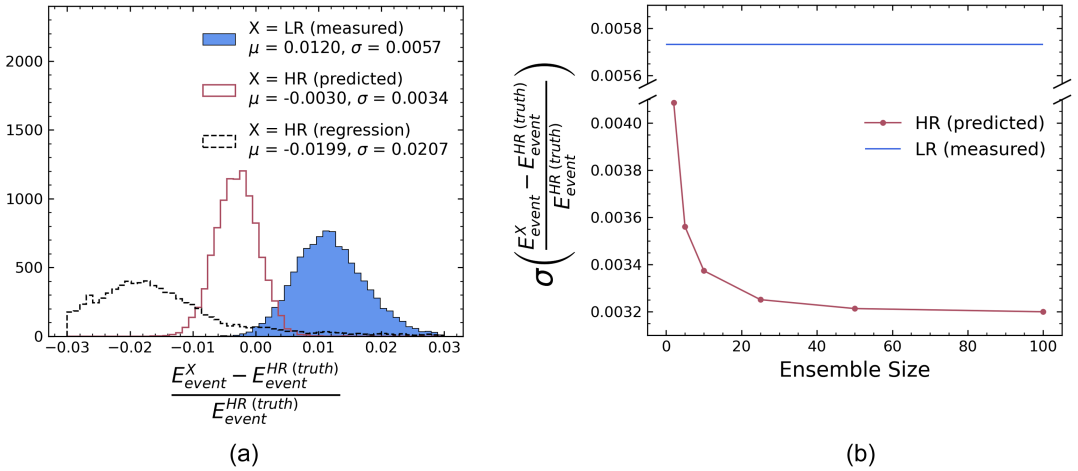


FIG. 7. (a) The event-level relative energy residual for the LR measured cell energy and the HR predicted cell energies with an ensemble size of 10. (b) The one-σ width of the event-level residual distributions for the predicted cell energies as a function of the ensemble size.

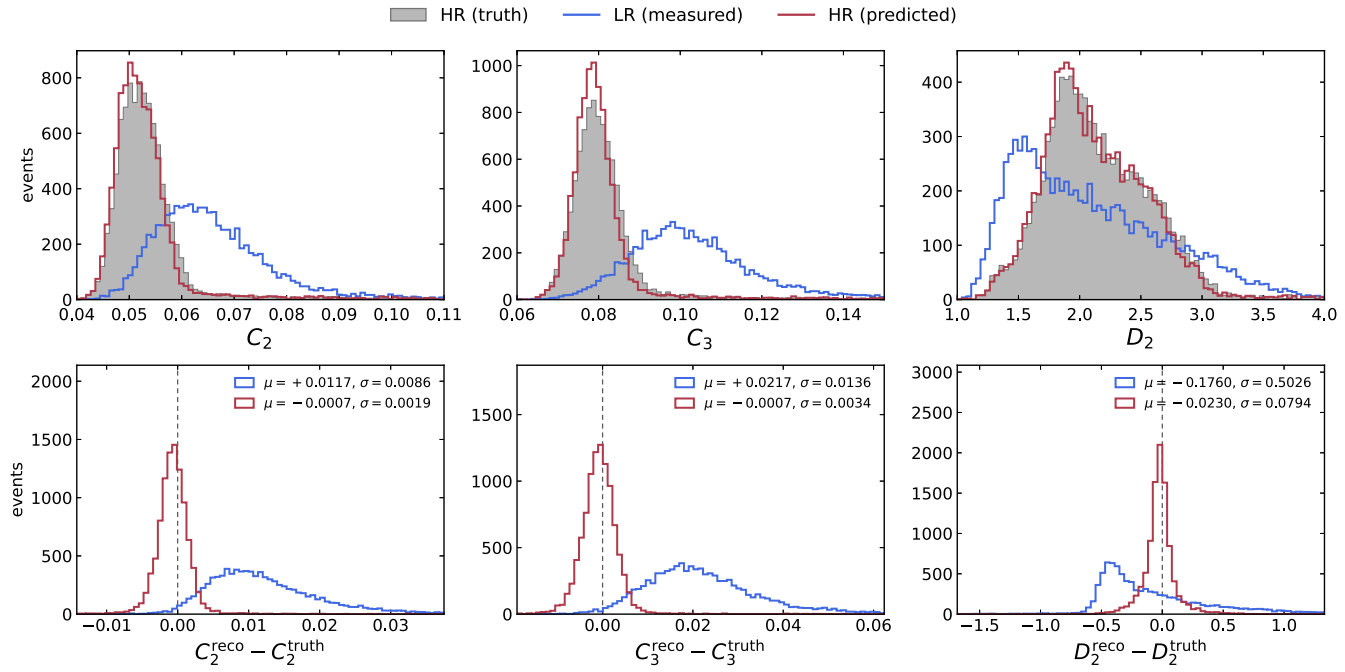


FIG. 8. Substructure distributions for the truth high-resolution, measured low-resolution and predicted high-resolution cell energies (upper row), and the corresponding per-event residuals with respect to the truth high-resolution value (lower row). All sets of cells are computed on the same events. Here μ and σ denote the median and the interquartile range divided by 1.349.

B. Multiple particles

For the multiparticle sample, the most exciting aspect of the study is the impact on PFlow performance. We begin by examining the cardinality performance of the two networks, as summarized in Table IV. Across the board, the network trained with predicted high-resolution cells shows either comparable or improved performance relative to the one trained with low-resolution measured cells. Both networks almost perfectly identify single-particle events, but as the true cardinality increases, their performance degrades, though at a slower rate for the network trained with high-resolution cells. When the true cardinality reaches its maximum, both networks exhibit similar performance, as this represents an edge-case scenario.

Next, we evaluate the performance of the computed particle p_T, η, and ϕ values derived from the incidence predictions. Ideally, the kinematic model would receive the predicted cardinality as input during inference. However,

TABLE IV. Cardinality prediction with the network trained on the LR cells with measured energy, compared with the network trained on the HR cells with predicted energy.

	Accuracy	
Truth cardinality	LR (measured) (%)	HR (predicted) (%)
1	99.38	**99.41**
2	88.52	**91.02**
3	77.62	**81.97**
4	81.32	**81.42**

to isolate the performance of the kinematics predictor from the cardinality predictor, we instead use the true cardinality as input. To evaluate the kinematics at the particle level, we pair the true and predicted particles per event through Hungarian matching, using the Kullback-Leibler divergence loss between their incidence values as the cost matrix. The p_T, η, and ϕ residuals after matching are shown in Fig. 9. Both models perform well in modeling the kinematic values, but the model trained on predicted high-resolution cells shows a marked improvement in angular resolution.

While both PFlow models have 400k parameters, one may argue that the model trained on high-resolution cells benefits implicitly from the 4.2M trained parameters in the SR model. To check whether a significantly larger PFlow model could have the same effect, we trained a variant with 4.6M parameters on the low-resolution cells. This model corresponds to the dashed line in Fig. 9, enabling a direct comparison. While some improvement is indeed seen for kinematics, the same limitation in angular resolution persists.

When a particle interacts with calorimeter cells, it often deposits a significant portion of its energy in a single cell, making that cell "hotter" than the others. Consequently, the η and ϕ values computed from the incidence matrix tend to gravitate toward those of the hottest cells. This results in the predicted particles not spreading evenly across the continuous space but instead forming diffused clumps at the midpoints of the highest resolution cells. In our case these are the cells in the ECAL1 layer. The effect is shown directly in Appendix A, where the predicted η and ϕ follow a staircase as a function of their true values, with steps of one low-resolution cell. This clustering effect is the reason

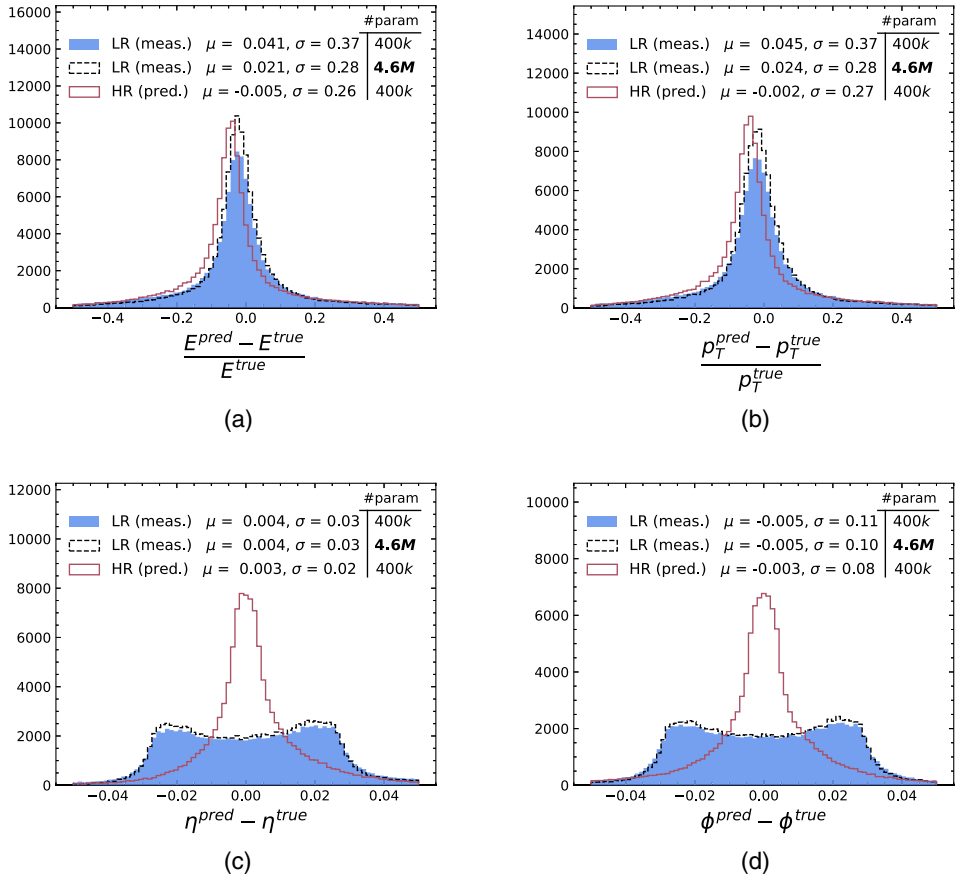


FIG. 9. Particle-flow performance. (a) The relative energy residuals, (b) the relative transverse momentum residuals, (c) the pseudorapidity residuals, and (d) the azimuthal angle residuals between the truth particles and the reconstructed particles with PFlow model trained on measured low-resolution cells and the predicted high-resolution cells. The dashed line indicates the performance of an additional PFlow model trained on the low-resolution cells but with over 10× more parameters.

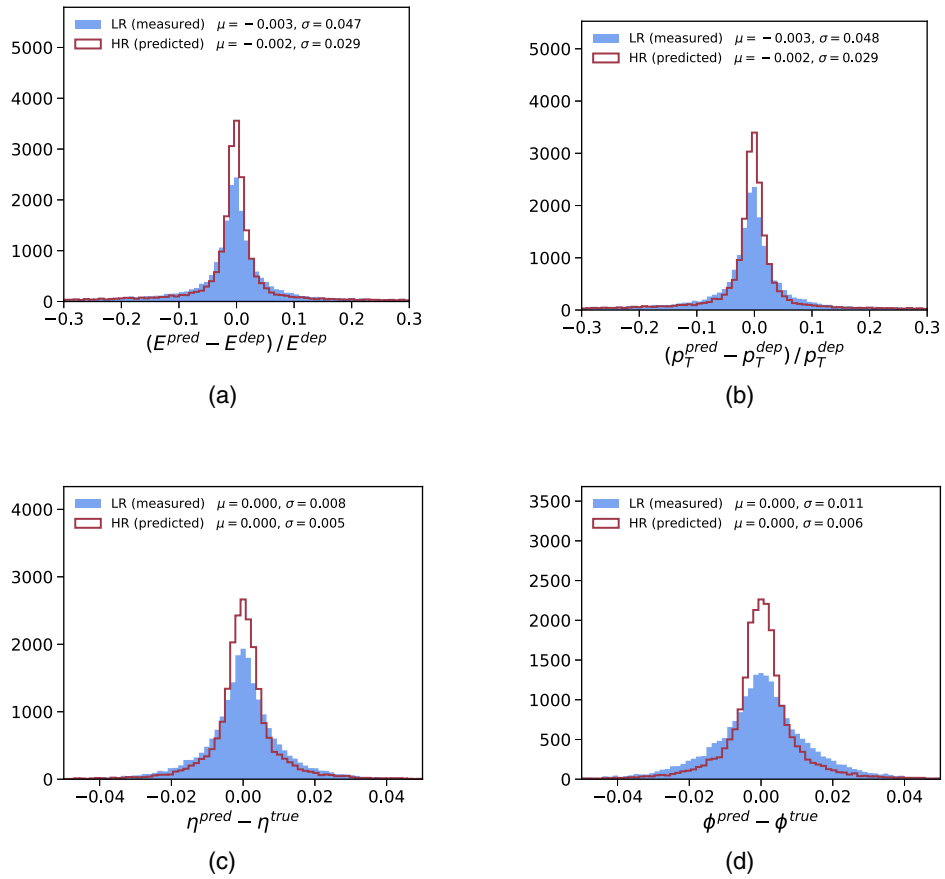


FIG. 10. Particle-flow performance *with the correction network*. (a) The relative energy residuals, (b) the relative transverse momentum residuals, (c) the pseudorapidity residuals, and (d) the azimuthal angle residuals between the truth particles and the reconstructed particles, for the PFlow models trained on measured low-resolution cells and on predicted high-resolution cells. The energy and transverse momentum residuals are computed with respect to the deposited energy, which is the quantity the models are trained to predict, and the corresponding transverse momentum. Both models are identical to those in Fig. 9 apart from the correction network applied to the proxy kinematics.

we observe the slightly bimodal structure in Figs. 9(c) and 9(d) for the low-resolution model. Although the high-resolution model should theoretically display a similar bimodal pattern, because of its finer resolution, we do not see this effect in the figures.

The structure described above is a property of the uncorrected estimator rather than of the input granularity, and it is removed once the correction network is included. Figure 10 shows the same residuals for the two models with the correction network. The low-resolution model improves substantially and the bimodal structure disappears, reaching an angular resolution of $\sigma(\Delta\eta) = 0.008$. This is well below the value of 0.027 expected from the cell size alone, which confirms that the correction network recovers subcell information from the energy sharing between neighboring cells. The high-resolution model also improves, and the difference between the two is reduced but does not vanish. The model trained on predicted high-resolution cells retains a factor of 1.5 in η, 1.9 in ϕ and 1.6 in energy. Since both models can now access subcell information, the remaining difference is attributable to the input cells rather than to the form of the estimator. The two comparisons answer different questions: the first quantifies what superresolution offers to a fixed reconstruction, and the second what it offers when the reconstruction is itself free to correct for the granularity of its input.

Superresolution both suppresses noise and increases granularity, and the comparison above does not separate the two. To disentangle them, we sum the predicted high-resolution cells within each parent low-resolution cell. These cells carry the noise suppression of the superresolution model at the original granularity. A third particle-flow model with the correction network was trained on them, and its residuals are shown in Fig. 13 in Appendix B. The denoised cells recover most of the improvement in energy but only a small part of the improvement in η. The model trained on the predicted high-resolution cells still retains a factor of 1.4 in η, 1.5 in ϕ and 1.2 in energy over them. This split follows from the observables. The energy of a particle is a sum over cells and is unchanged by the downsampling

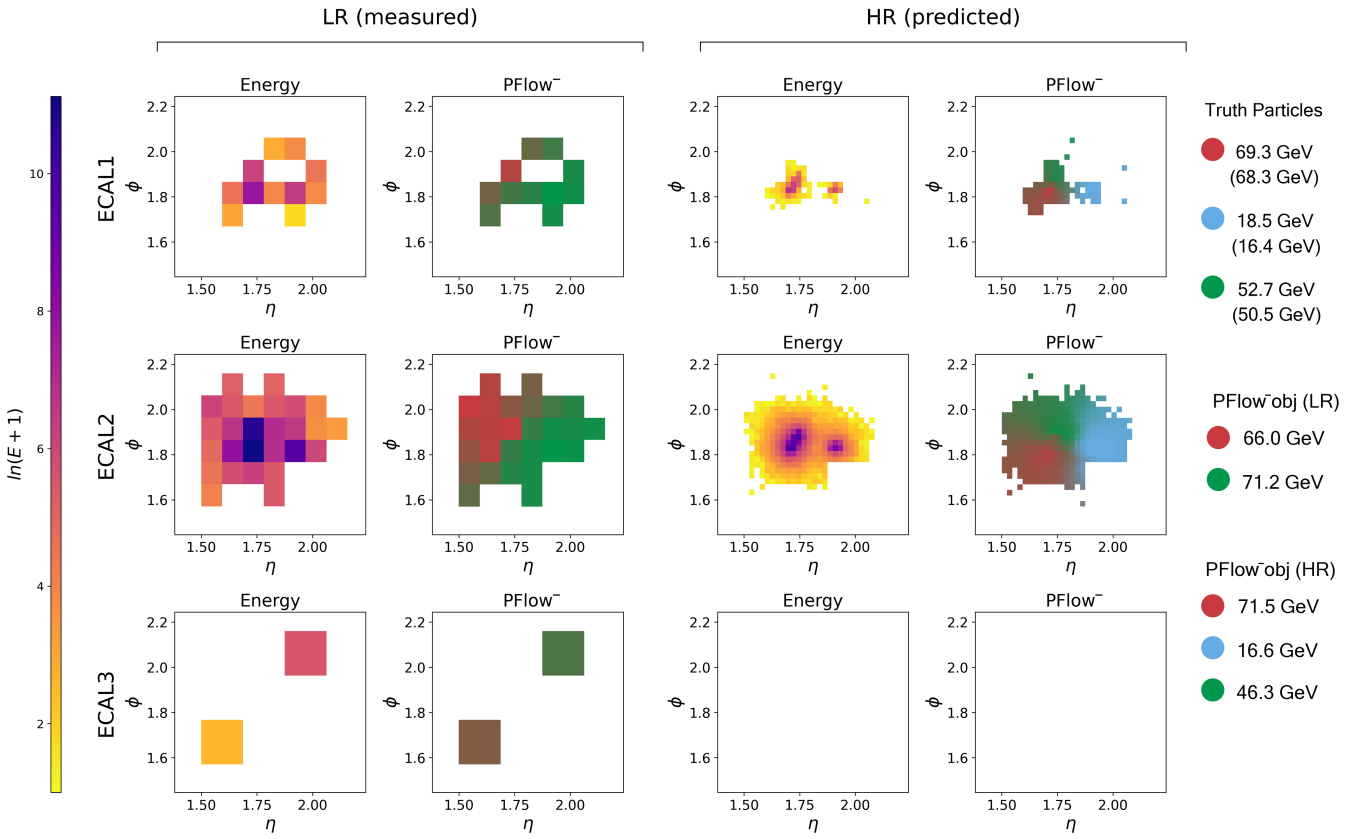


FIG. 11. Event display illustrating the application of superresolution on particle flow, as well as the interpretability of the particle reconstruction model. Each cell is assigned a fractional weight based on the predicted attention or incidence values, which determine the cell colors. The legend shows the truth and predicted particles with their energies. The true deposited energy for the truth particles is shown in parentheses. The particle-flow model trained on the predicted high-resolution cells correctly identifies all three particles with very accurate energy estimations, while the model trained on low-resolution measured cells fails to disentangle the event into three particles and predicts only two particles.

that defines the low-resolution cells, so it can only be improved by noise suppression. The angular coordinates depend on how the energy is shared among cells, and for these the finer segmentation carries most of the gain. The denoised cells are derived from the superresolution model rather than from a denoiser trained at low resolution, so this separation is indicative rather than exact.

Figure 11 presents a detailed event display showcasing the particle-flow model in action. In this case, the cardinality predictor is run first, and its output is fed into the kinematic predictor. Although the entire process occurs at the graph level, for clarity, we present the results in the $\eta - \phi$ plane as images.

V. DISCUSSION

The results from the studies above indicate that superresolution is a viable and promising approach for improving reconstruction. In the following paragraphs, we address the question of how these improvements should be understood, discuss the path toward deployment in real experiments, and present areas of future work.

A. Superresolution or hallucination

A central question in this context is understanding the origin of the "extra" information that enhances low-resolution inputs into high-resolution outputs. The key insight lies in the patterns learned from the training data, particularly the nature of electromagnetic (EM) showers. Although stochastic on a case-by-case basis, the showers sample from a statistical prior $p(\mathrm{HR})$ that the network can learn from the high-resolution training examples. The model generates high-resolution features from the conditional distribution $p(\mathrm{HR}|\mathrm{LR})$, where the mapping is inherently one to many due to the stochastic nature of EM showers. While the model can approximate the manifold of consistent solutions arbitrarily well, the data processing inequality imposes a fundamental ceiling on its ability to recover the specific "truth" instance lost during measurement. Extra features thus stem from the information stored in model parameters, yielding samples that are distinct from the truth but physically consistent with its underlying prior. The same ceiling implies that superresolution cannot improve upon an optimal reconstruction from the low-resolution cells, and

we do not claim that it does. What it provides instead is a representation: the subcell information already present in the low-resolution cells is expressed in the standard high-granularity form, where algorithms developed for fine-granularity detectors can use it directly. The relevant question for a reconstruction chain is what a given downstream model achieves in practice, and the comparison in Fig. 10 shows that the improvement persists when both models are free to correct for the granularity of their input.

B. Model dependence

Training on EM showers from Standard Model particles will not equip the SR model to handle showers drawn from an anomalous prior. Examples include hypothetical particles with long lifetimes, exotic couplings, fractional electric charge, or magnetic monopoles. Such scenarios present a challenge to standard reconstruction techniques as well as SR which are tuned to Standard Model signatures. These limitations are a reminder that SR cannot be considered a replacement for actual high-granularity calorimeters.

C. Data-simulation alignment

A related corollary is that training on simulation can lead to degraded or unexpected performance on real data. This alignment problem affects machine learning models in general and has also been addressed in collider physics applications [46,47]. To mitigate the effect, future applications of this technique must include rigorous validation strategies to quantify this bias. While direct comparisons are difficult with experimental data, it is possible to develop small, high-granularity prototype modules to validate the network's outputs. In addition, techniques such as data augmentation can be employed to help reduce the model's dependence on features specific to the simulation [48].

D. Superresolution as an auxiliary task

Moreover, SR should be considered not just as an independent task but as an auxiliary to the broader reconstruction processes. By guiding the network to focus on relevant features, superresolution can complement traditional reconstruction algorithms, which explains why LR PFlow performance lags behind HR PFlow. In this paper, we explored SR as a preprocessing task, but it could also be embedded as a component of the PFlow training objective. The integration of SR techniques in this way has the potential to inform future detector designs and enhance the development of reconstruction algorithms, ultimately leading to more accurate and reliable data interpretation in collider experiments.

E. Toward deployment

To demonstrate a proof of concept, we have restricted our study to a simplified physics and detector setup. Several developments are needed to apply the SR technique to real experiments. First, the model needs to be trained on both hadronic and EM showers. Incorporating hadronic showers will introduce greater complexity due to the increased fluctuations inherent in these processes. Progressing toward full, realistic collision events will entail a significant rise in particle multiplicity due to jets and pileup. Including tracks as an input to the SR model may aid its performance in highly dense environments. To manage the model's memory demands, limiting graph connectivity and possibly clustering in angular space will be advantageous. Another important effect we have ignored presently is that of a magnetic field which alters the incident angle of particles and, in some detectors, their daughter particles.

F. Ideas for future work

While we have focused on superresolution in the angular space, it would be interesting to try this in the longitudinal direction, where the number of calorimeter layers is upscaled. Uncertainty quantification is another idea that could improve the interpretability of the SR outputs and flag instances of poor generation. Uncertainty quantification for generative models is an active area of research, with recent proposals targeting Bayesian inference for diffusion and flow models [49]. Finally, while we have focused on generic particle reconstruction, future work could explore the potential of SR to enhance unfolding in specific topologies [50].

VI. CONCLUSION

In this work, we demonstrated, for the first time, that superresolution techniques can be effectively incorporated into a simplified LHC-like particle-flow reconstruction chain. We introduced a novel graph-based model for superresolution that leverages the powerful flow-matching framework alongside a transparent and interpretable particle-flow model. However, the key takeaway is not the specific models employed; rather, it is that superresolution can indeed enhance reconstruction. Our approach faithfully replicates high-resolution cells event by event, leading to significantly improved modeling of cell substructure variables. Moreover, utilizing these high-resolution cells for particle flow substantially enhances both particle cardinality and angular resolution. This improvement persists when both particle-flow models are equipped with a correction network for the proxy kinematics, showing that it does not follow from the restricted form of the default kinematic estimator. We believe these findings motivate further exploration of SR in reconstruction pipelines. Detector designs are evaluated through the reconstruction performance they deliver in simulation, so a software step that improves that performance at fixed granularity bears directly on the performance-to-cost trade-off, without standing in for the measurement a finer detector would make.

ACKNOWLEDGMENTS

We thank Dmitrii Kobilianskii for fruitful discussions on flow-matching implementation and Kyle Cranmer and Yihui Ren for their valuable insights. E. G. also acknowledges support from ISF Grant No. 2871/19, the BSF-NSF Grant No. 2020780, and the Weizmann Institute of Science and Mohamed bin Zayed University of Artificial Intelligence (MBZUAI) collaboration grant. We also thank the Benoziyo Center for High Energy Physics for supporting our research.

DATA AVAILABILITY

To facilitate reproducibility, the code used for the analyses presented in this paper is available on GitHub [51]. The dataset generated and analyzed during this study can be accessed through Zenodo [52].

APPENDIX A: QUANTIZATION OF THE PROXY KINEMATICS

The proxy kinematics are obtained as incidence weighted averages of fixed cell coordinates, so the predicted position of a particle is a weighted mean of the centers of the cells that receive its energy. When most of the energy is collected by a single cell, the prediction is drawn toward the center of that cell. Figure 12 shows the predicted η and ϕ against their true values for the model trained on the low-resolution cells. The predictions follow a staircase whose steps correspond to the width of a low-resolution cell in the first two ECAL layers, 0.094 in η and 0.098 in ϕ. This is the origin of the bimodal structure seen in Figs. 9(c) and 9(d). The effect follows from the form of the estimator rather than from the input cells, and it is absent once the correction network is included.

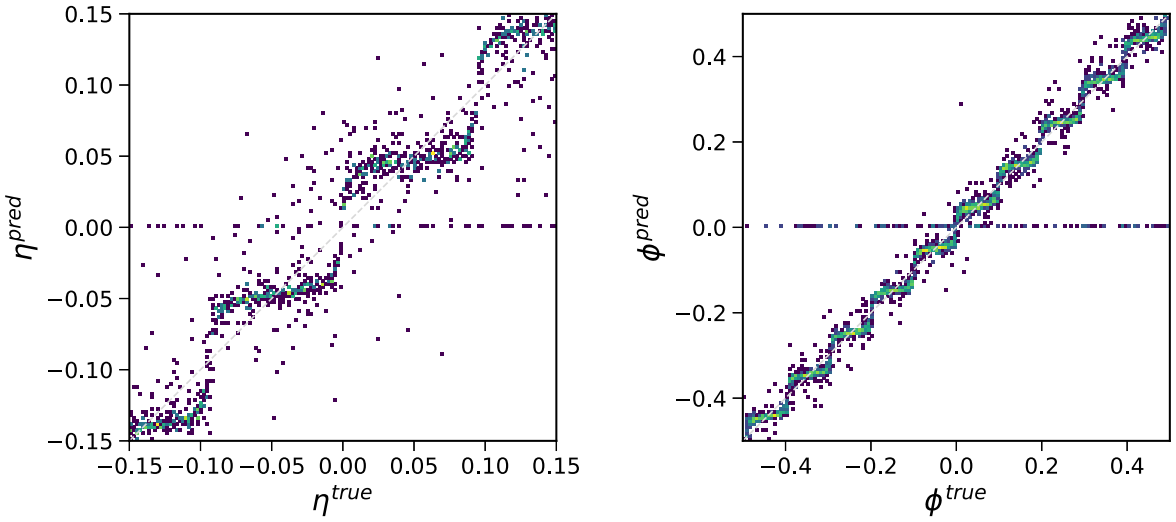


FIG. 12. Predicted against true pseudorapidity (left) and azimuthal angle (right) for the particle-flow model trained on the measured low-resolution cells, without the correction network. The predictions are quantized in steps of one low-resolution cell. The dashed line indicates equality.

APPENDIX B: SEPARATING DENOISING FROM GRANULARITY

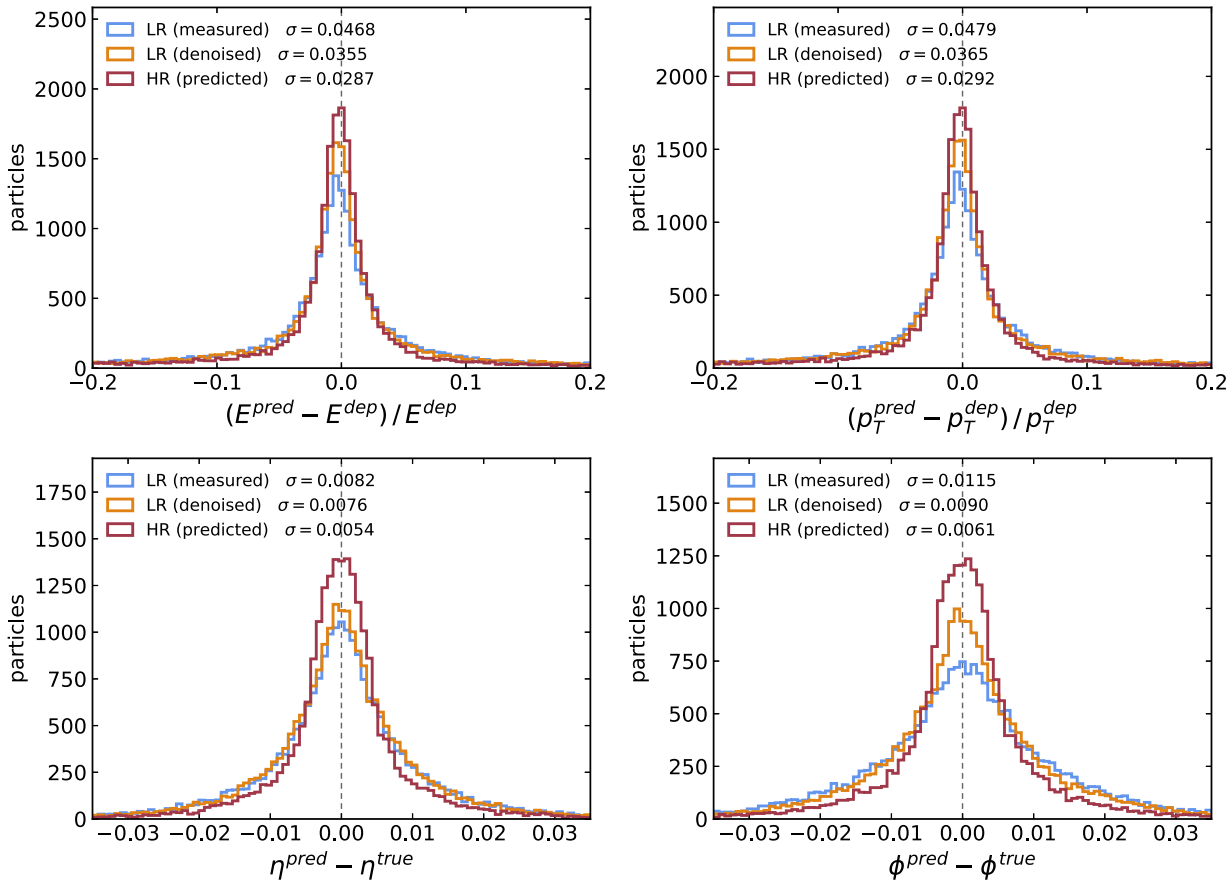


FIG. 13. Particle-flow performance with the correction network for three sets of input cells: (top left) the relative energy residuals, (top right) the relative transverse momentum residuals, (bottom left) the pseudorapidity residuals, and (bottom right) the azimuthal angle residuals. The three models are identical apart from their inputs. The denoised low-resolution cells are obtained by summing the predicted high-resolution cells within each parent low-resolution cell, so they carry the noise suppression of the superresolution model at the original granularity, and the difference between them and the predicted high-resolution cells reflects the granularity of the representation alone.

[1] Lyndon Evans and Philip Bryant, LHC machine, J. Instrum. **3**, S08001 (2008).

[2] I. Zurbano Fernandez *et al.*, High-Luminosity Large Hadron Collider (HL-LHC): Technical design report, Technical Report No. CERN-2020-010, 2020.

[3] ATLAS Collaboration, Jet reconstruction and performance using particle flow with the ATLAS Detector, Eur. Phys. J. C **77**, 466 (2017).

[4] CMS Collaboration, Particle-flow reconstruction and global event description with the CMS detector, J. Instrum. **12**, P10003 (2017).

[5] Huong Lan Tran, Katja Krüger, Felix Sefkow, Steven Green, John Marshall, Mark Thomson, and Frank Simon, Software compensation in particle flow reconstruction, Eur. Phys. J. C **77**, 698 (2017).

[6] M. A. Thomson, Particle flow calorimetry and the PandoraPFA algorithm, Nucl. Instrum. Methods Phys. Res., Sect. A **611**, 25 (2009).

[7] Martin Aleksa, Franco Bedeschi, Roberto Ferrari, Felix Sefkow, and Christopher G. Tully, Calorimetry at FCC-ee, Eur. Phys. J. Plus **136**, 1066 (2021).

[8] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, Image super-resolution using deep convolutional networks, IEEE Trans. Pattern Anal. Mach. Intell. **38**, 295 (2016).

[9] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee, Accurate image super-resolution using very deep convolutional

networks, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 1646–1654.
[10] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee, Enhanced deep residual networks for single image super-resolution, in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2017), pp. 1132–1140.
[11] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi, Photo-realistic single image super-resolution using a generative adversarial network, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 105–114.
[12] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu, Image super-resolution using very deep residual channel attention networks, in *Computer Vision—ECCV 2018*, edited by Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (Springer International Publishing, Cham, 2018), pp. 294–310.
[13] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy, ESRGAN: Enhanced super-resolution generative adversarial networks, in *Computer Vision—ECCV 2018 Workshops*, edited by Laura Leal-Taixé and Stefan Roth (Springer International Publishing, Cham, 2019), pp. 63–79.
[14] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J. Fleet, and Mohammad Norouzi, Image super-resolution via iterative refinement, IEEE Trans. Pattern Anal. Mach. Intell. **45**, 4713 (2023).
[15] Francesco Armando Di Bello, Sanmay Ganguly, Eilam Gross, Marumi Kado, Michael Pitt, Lorenzo Santi, and Jonathan Shlomi, Towards a computer vision particle flow, Eur. Phys. J. C **81**, 107 (2021).
[16] Pierre Baldi, Lukas Blecher, Anja Butter, Julian Collado, Jessica N. Howard, Fabian Keilbach, Tilman Plehn, Gregor Kasieczka, and Daniel Whiteson, How to GAN higher jet resolution, SciPost Phys. **13**, 064 (2022).
[17] Ian Pang, David Shih, and John Andrew Raine, Calorimeter shower superresolution, Phys. Rev. D **109**, 092009 (2024).
[18] Johannes Erdmann, Aaron van der Graaf, Florian Mausolf, and Olaf Nackenhorst, SR-GAN for SR-gamma: Super resolution of photon calorimeter images at collider experiments, Eur. Phys. J. C **83**, 1001 (2023).
[19] Claudius Krause *et al.*, CaloChallenge 2022: A community challenge for fast calorimeter simulation, Rep. Prog. Phys. **88**, 116201 (2025).
[20] Felix J. Yu, Nicholas Kamp, and Carlos A. Argüelles, Enhancing events in neutrino telescopes through deep-learning-driven superresolution, Phys. Rev. D **111**, L041301 (2025).
[21] Yanting Yin, Yajing Wu, Xuebing Yang, Wensheng Zhang, and Xiaojie Yuan, Super resolution graph with conditional normalizing flows for temporal link prediction, IEEE Trans. Knowledge Data Eng. **36**, 1311 (2024).
[22] Nishant Rajadhyaksha and Islem Rekik, Diffusion-based graph super-resolution with application to connectomics, in *Predictive Intelligence in Medicine* (Springer Nature Switzerland, Cham, 2023), pp. 96–107.
[23] F. A. Di Bello, E. Dreyer, S. Ganguly, E. Gross, L. Heinrich *et al.*, Reconstructing particles in jets using set transformer and hypergraph prediction networks, Eur. Phys. J. C **83**, 596 (2023).
[24] F. A. Di Bello, A. Charkin-Gorbulin, K. Cranmer, E. Dreyer, S. Ganguly *et al.*, Configurable calorimeter simulation for AI applications, Mach. Learn. Sci. Technol. **4**, 035042 (2023).
[25] S. Agostinelli *et al.*, GEANT4–a simulation toolkit, Nucl. Instrum. Methods Phys. Res., Sect. A **506**, 250 (2003).
[26] John Allison *et al.*, Geant4 developments and applications, IEEE Trans. Nucl. Sci. **53**, 270 (2006).
[27] J. Allison *et al.*, Recent developments in Geant4, Nucl. Instrum. Methods Phys. Res., Sect. A **835**, 186 (2016).
[28] E. Moyse, F. Ali, E. Cortina, R. M. Bianchi, and B. Couturier, The Phoenix event display framework, EPJ Web Conf. **251**, 01007 (2021).
[29] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K. Duvenaud, Neural ordinary differential equations, in *Advances in Neural Information Processing Systems*, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Curran Associates, Inc., Red Hook, NY, 2018), Vol. 31.
[30] Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le, Flow matching for generative modeling, in *The Eleventh International Conference on Learning Representations* (2023).
[31] Alexander Tong, Kilian Fatras, Nikolay Malkin, Guillaume Huguet, Yanlei Zhang, Jarrid Rector-Brooks, Guy Wolf, and Yoshua Bengio, Improving and generalizing flow-based generative models with minibatch optimal transport, Trans. Mach. Learn. Res. (2024), https://openreview.net/forum?id=CD9Snc73AW.
[32] Ricky T. Q. Chen, torchdiffeq, https://github.com/rtqichen/torchdiffeq.
[33] Ilya Loshchilov and Frank Hutter, Decoupled weight decay regularization, in *International Conference on Learning Representations* (2019).
[34] J. Kieseler, Object condensation: One-stage grid-free multi-object reconstruction in physics detectors, graph, and image data, Eur. Phys. J. C **80**, 1 (2020).
[35] S. R. Qasim, J. Kieseler, Y. Iiyama, and M. Pierini, Learning representations of irregular particle-detector geometry with distance-weighted graph networks, Eur. Phys. J. C **79**, 608 (2019).
[36] S. R. Qasim, K. Long, J. Kieseler, M. Pierini, and R. Nawaz, Multi-particle reconstruction in the high granularity calorimeter using object condensation and graph neural networks, EPJ Web Conf. **251**, 03072 (2021).
[37] S. R. Qasim, N. Chernyavskaya, J. Kieseler, K. Long, O. Viazlo, M. Pierini, and R. Nawaz, End-to-end multi-particle reconstruction in high occupancy imaging calorimeters with graph neural networks, Eur. Phys. J. C **82**, 753 (2022).
[38] J. Pata, J. Duarte, J. R. Vlimant, M. Pierini, and M. Spiropulu, MLPF: Efficient machine-learned particle-flow reconstruction using graph neural networks, Eur. Phys. J. C **81**, 1 (2021).
[39] J. Pata, J. Duarte, F. Mokhtar, E. Wulff, J. Yoo, J. R. Vlimant, M. Pierini, and M. Girone, Machine learning for particle flow reconstruction at CMS, J. Phys. Conf. Ser. **2438**, 012100 (2022).

[40] F. Mokhtar, J. Pata, J. Duarte, E. Wulff, M. Pierini, and J. R. Vlimant, Progress towards an improved particle flow algorithm at CMS with machine learning, in *ACAT 2022* (2023).
[41] Joosep Pata, Eric Wulff, Farouk Mokhtar, David Southwick, Mengke Zhang, Maria Girone, and Javier Duarte, Improved particle-flow event reconstruction with scalable neural networks for current and future particle detectors, Commun. Phys. **7**, 124 (2024).
[42] Nilotpal Kakati, Etienne Dreyer, Anna Ivina, Francesco Armando Di Bello, Lukas Heinrich, Marumi Kado, and Eilam Gross, HGPflow: Extending hypergraph particle flow to collider event reconstruction, Eur. Phys. J. C **85**, 847 (2025).
[43] H. W. Kuhn, The Hungarian method for the assignment problem, Naval research logistics quarterly **2**, 83 (1955).
[44] Andrew J. Larkoski, Gavin P. Salam, and Jesse Thaler, Energy correlation functions for jet substructure, J. High Energy Phys. 06 (2013) 108.
[45] Andrew J. Larkoski, Ian Moult, and Duff Neill, Power counting to better jet observables, J. High Energy Phys. 12 (2014) 009.
[46] ATLAS Collaboration, A continuous calibration of the ATLAS flavour-tagging classifiers via optimal transportation maps, Eur. Phys. J. C **85**, 1272 (2025).
[47] CMS Collaboration, Identification of tau leptons using a convolutional neural network with domain adaptation, J. Instrum. **20**, P12032 (2025).
[48] Yingyi Zhong, Wen'an Zhou, and Zhixian Wang, A survey of data augmentation in domain generalization, Neural Process. Lett. **57**, 34 (2025).
[49] Siqi Kou, Lei Gan, Dequan Wang, Chongxuan Li, and Zhijie Deng, BayesDiff: Estimating pixel-wise uncertainty in diffusion via bayesian inference, in *The Twelfth International Conference on Learning Representations (ICLR)* (2024), https://openreview.net/forum?id=YcM6ofShwY.
[50] Nathan Huetsch *et al.*, The landscape of unfolding with machine learning, SciPost Phys. **18**, 070 (2025).
[51] Nilotpal Kakati, SuperResolutionHEP, https://github.com/nilotpal09/SuperResolutionHEP (2025).
[52] Nilotpal Kakati, Etienne Dreyer, and Eilam Gross, Dataset for denoising graph super-resolution towards improved collider event reconstruction, Zenodo, 10.5281/zenodo.15582324 (2025).