Featured in Physics

Contrastive Metric Learning for Point Cloud Segmentation in Highly Granular Detectors

Max Marriott-Clarke,[1,2,*] Lazar Novakovic,[3] Elizabeth Ratzer,[1]
Robert J. Bainbridge,[1] Loukas Gouskos,[3,4] and Benedikt Maier[1]
[1]*Blackett Laboratory, Imperial College London, London, United Kingdom*
[2]*Experimental Physics Department, European Organization for Nuclear Research (CERN), Geneva, Switzerland*
[3]*Department of Physics & Astronomy, Brown University, Providence, Rhode Island, USA*
[4]*Brown Center for Theoretical Physics and Innovation (BCTPI), Brown University, Providence, Rhode Island, USA*

(Received 30 March 2026; accepted 16 June 2026; published 28 July 2026)

We propose a clustering approach for point cloud segmentation based on supervised contrastive metric learning (CML). Rather than predicting cluster assignments or object-centric variables, the method learns a latent representation in which points belonging to the same object are embedded nearby, while unrelated points are separated. Clusters are then reconstructed using a density-based readout in the learned metric space, decoupling representation learning from cluster formation and enabling flexible inference. The approach is evaluated on simulated data from a highly granular calorimeter, where the task is to separate highly overlapping particle showers represented as sets of calorimeter hits. A direct comparison with object condensation (OC) is performed using identical graph neural network backbones and equal latent dimensionality, isolating the effect of the learning objective. The CML method produces a more stable and separable embedding geometry for both electromagnetic and hadronic particle showers, leading to improved local neighborhood consistency, a more reliable separation of overlapping showers, and better generalization when extrapolating to unseen multiplicities and energies. This translates directly into higher reconstruction efficiency and purity, particularly in high-multiplicity regimes, as well as improved energy resolution. In mixed-particle environments, CML maintains strong performance, suggesting stable learning of the shower topology across particle types, while OC exhibits significant degradation. These results demonstrate that similarity-based representation learning combined with density-based aggregation is a promising alternative to object-centric approaches for point cloud segmentation in highly granular detectors.

DOI: 10.1103/tsc6-gn3r

I. INTRODUCTION

Modern particle detectors increasingly rely on high-granularity sensor technologies that provide detailed spatial and timing measurements of energy deposits. The resulting data naturally form point clouds with irregular geometry and variable size. A central reconstruction task is the segmentation of these point clouds into groups of measurements originating from individual particles. In high-granularity calorimeters such as the CMS High Granularity Calorimeter (HGCAL) [1], this task is particularly challenging due to the frequent spatial and energetic overlap of particle showers, requiring algorithms capable of resolving complex and highly correlated hit patterns.

Graph neural networks (GNNs) have emerged as a powerful framework for point cloud reconstruction in particle physics [2–4]. A widely used learned clustering approach built on this framework is object condensation (OC) [5], in which the network predicts object-centric latent variables that guide cluster formation. By integrating clustering into the training objective, OC achieves strong performance across a range of reconstruction tasks [5–8]. However, this formulation tightly couples representation learning to a specific clustering procedure. In dense environments, where multiple nearby showers compete for representative points, this coupling can lead to ambiguities in both the learned clustering coordinates and the assignment of hits to objects.

In this work, we propose an alternative clustering paradigm based on contrastive metric learning (CML) [9]. Rather than predicting object-level variables, the method learns a latent embedding in which hits from the same particle shower are placed nearby, while hits from different showers are separated. Clustering is performed only after training and acts as a readout of the learned representation. This decoupling allows the embedding geometry to be optimized directly for pairwise compatibility, without imposing constraints associated with a particular clustering mechanism.

A key advantage of this formulation is that the contrastive objective depends only on relative relationships between hits, rather than absolute object-level properties [9]. In object-centric approaches, the network must learn representative points and clustering scales that are implicitly tied to the detailed morphology of particle showers. As a result, mismodeling in simulation can propagate directly into the learned

*Contact author: max.marriott-clarke@cern.ch

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.

clustering variables. In contrast, CML requires only that hits from the same shower be more similar than those from different showers, which may make the learned representation less sensitive to variations in shower shape, energy response, and event composition, though we do not demonstrate this empirically in the present work.

To extract clusters from the learned embedding, we introduce a density-based readout procedure tailored to metric spaces. This method identifies representative points from local neighborhood structure and forms clusters without requiring explicit object-level predictions, ensuring consistency with the geometry induced by the contrastive objective.

The proposed approach is evaluated on simulated electromagnetic (EM) and hadronic (HAD) showers in a detector model inspired by the CMS HGCAL. A direct comparison with OC is performed using identical GNN backbones and matched embedding dimensionality to isolate the impact of the learning objective. Representation quality is assessed using embedding-geometry metrics, while reconstruction performance is evaluated using physics-based observables including efficiency, purity, and energy resolution.

We find that CML produces a more structured embedding geometry, with improved separability particularly in the tails of the distance distributions where reconstruction failures typically occur. This leads to significant improvements in clustering performance in high-multiplicity environments with strong shower overlap. These results demonstrate that learning a similarity-based representation provides a robust alternative to object-centric clustering methods for high-granularity detectors and more generally for dense point cloud segmentation problems.

II. METHODS

A. Model architecture

Both learning objectives employ an identical GNN backbone to permit a direct comparison. Differences between methods are restricted to the task-specific output heads required by the respective loss functions.

Each calorimeter hit is represented by a five-dimensional input feature vector

$$(x, y, z, E, L),$$

where (x, y, z) denote the hit position, E is the deposited energy, and L is the detector layer index. Further details of the dataset and input representation are given in Sec. III. Events are treated as variable-size point clouds with no graph connectivity provided as input. Truth shower identifiers are stored separately and are used only for defining the learning objectives and evaluation metrics. No preprocessing or feature normalization is applied to the inputs, and we find this has negligible impact on performance.

The features are first projected into a learned latent space using a two-layer multilayer perceptron (MLP) with hidden dimension 64 and exponential linear unit (ELU) activations, producing a 64-dimensional representation.

The encoded hit features are subsequently processed by three DynamicEdgeConv layers [10], each constructed using a k-nearest-neighbor graph ($k = 24$) in the current latent feature space using Euclidean distance. A learnable edge function is applied and aggregated using the permutation-invariant "max" operation. Because the graph connectivity is recomputed after every DynamicEdgeConv layer, the neighborhood structure evolves during training, allowing the network to jointly learn feature representations and a data-driven similarity metric between hits. Constructing the graph in latent feature space reduces sensitivity to the geometric constraints imposed by the detector layout and enables interactions between hits across layers, which is important for capturing the longitudinal development of calorimeter showers.

After the shared DynamicEdgeConv backbone, the network separates into task-specific output heads. In both cases, features are processed by a two-layer MLP ($64 \to 32$) with ELU activations and dropout ($p = 0.1$), applied identically across both heads. For the contrastive objective, a final linear layer projects the 32-dimensional features to a 16-dimensional embedding used for metric learning. For object condensation, two parallel linear heads are applied to the same 32-dimensional features, predicting a per-hit condensation score $\beta_i \in (0, 1)$ and 16-dimensional clustering coordinates c_i. The embedding dimensionality is matched between the two methods to ensure a comparable latent-space dimensionality. Both methods were also evaluated with a reduced four-dimensional embedding; the corresponding results are presented in Sec. V.

Since all backbone layers (the input MLP, the three DynamicEdgeConv layers, and the shared output MLP), activation functions, and optimization hyperparameters are shared, differences in embedding structure and reconstruction performance between the two methods can be primarily attributed to the choice of learning objective. The two methods are trained entirely independently; the shared backbone refers to identical architectural design only, not joint optimization.

B. Contrastive metric learning

CML is a class of methods [11–13] that learns a representation space in which samples sharing the same semantic identity are embedded nearby, while unrelated samples are separated. Rather than predicting discrete labels or cluster assignments, the objective directly shapes the geometry of the latent space, making the embedding itself the primary target of optimization. This paradigm is well suited to point cloud segmentation in high-granularity detectors, where reconstruction depends on identifying compatible sets of hits rather than assigning them to predefined object templates. Such objectives have been widely used in representation learning and clustering tasks [12,14,15], where downstream performance depends on the separability of learned features.

We apply supervised contrastive (SupCon) learning at the hit level within individual events. For an event containing N hits with embeddings $\{\mathbf{z}_i\}_{i=1}^{N}$ and corresponding shower labels $\{y_i\}_{i=1}^{N}$, positive pairs are defined as hits originating from the same simulated particle shower, while negative pairs correspond to hits from different showers within the same event. Comparisons across events are excluded since separate events do not share physically meaningful relationships.

The embeddings are ℓ_2 normalized and compared using cosine similarity $s_{ij} = \mathbf{z}_i^{\top}\mathbf{z}_j$, constraining all representations to lie on the unit hypersphere. This encourages hits from

the same shower to form compact regions in angular space while separating hits from different showers, thereby learning a discriminative similarity metric without requiring explicit cluster centers or assignment variables. This formulation is advantageous in calorimeter reconstruction, where showers are spatially extended and frequently overlap, making object boundaries inherently ambiguous. Since the DynamicEdgeConv graph is constructed in the current latent feature space at each layer, the neighborhood structure used for message passing evolves alongside the learned representation. For CML, this creates a coupling between the contrastive objective and the graph topology: As the embedding learns to place shower-compatible hits nearby, the graph increasingly connects hits from the same shower, reinforcing the representation. This feedback between representation learning and graph construction may contribute to the stability of the CML embeddings observed in Sec. V.

The loss is given by the supervised contrastive objective [13]:

$$\mathcal{L}_{\text{SupCon}} = -\frac{1}{|\mathcal{A}|}\sum_{i\in\mathcal{A}}\frac{1}{|\mathcal{P}(i)|}\sum_{p\in\mathcal{P}(i)}\ln\frac{\exp\left(\mathbf{z}_i^\top\mathbf{z}_p/\tau\right)}{\sum_{j\neq i}\exp\left(\mathbf{z}_i^\top\mathbf{z}_j/\tau\right)}, \quad (1)$$

where each hit i acts as an anchor (i.e., a reference point against which all other hits in the event are compared), $\mathcal{A}$ denotes the set of anchors with at least one positive sample, $\mathcal{P}(i)$ is the set of positive hits for anchor i (i.e., hits sharing the same-shower label), and τ is a temperature parameter that controls the concentration of the embedding distribution on the hypersphere. Smaller τ sharpens the similarity distribution and encourages tighter clusters. In this work, we use $\tau = 0.1$. The 16-dimensional output of the final linear projection layer constitutes the embedding used for all metric learning and evaluation; no intermediate representation is used.

Rather than sampling pairs, all hits within the event contribute to the loss: Positive pairs appear explicitly in the numerator, while all negative pairs contribute through the partition function in the denominator, actively shaping the gradient. Calorimeter events typically contain many weak negatives and relatively few informative hard negatives, such that random sampling often yields low-information pairs. Aggregating over all pairs ensures that informative relationships consistently contribute to the optimization. As shown in the SupCon formulation [13], the gradient naturally emphasizes hard positives and hard negatives, providing implicit hard pair mining. Although contrastive methods often rely on large batch sizes to obtain sufficient negatives, the large number of hits within each event naturally provides a rich negative set, allowing each event to act as an effective training batch.

C. Object condensation baseline

The OC [5] is a learned clustering method that reconstructs particle candidates from detector hits using per-hit latent variables. The approach has been applied across a range of high-energy physics tasks [5,7,8,16], and is adopted here as a baseline.

For each hit i, the network predicts clustering coordinates $\mathbf{c}_i \in \mathbb{R}^d$ and a condensation score $\beta_i \in (0, 1)$. The coordinates define a latent clustering space, while β_i identifies representative hits. Hits with large β act as condensation points around which compatible hits are grouped. In this work, the clustering space dimension is $d = 16$, matching the contrastive embedding dimension to ensure a comparable latent-space dimensionality.

The condensation score β is mapped to a positive charge

$$q_i = \operatorname{arctanh}^2(\beta_i) + q_{\min}, \quad (2)$$

which determines the interaction strength in the clustering space. For each object k, the hit with the largest charge defines the condensation point

$$\alpha_k = \arg\max_{i\in k} q_i, \quad (3)$$

where k indexes the set of hits belonging to a given truth shower and K denotes the total number of showers in the event.

Hits belonging to the same object are attracted to the condensation point,

$$\mathcal{L}_{\text{att}} = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{|k|}\sum_{i\in k} q_i q_{\alpha_k}\left\|\mathbf{c}_i - \mathbf{c}_{\alpha_k}\right\|^2, \quad (4)$$

while hits from different objects are repelled within a finite radius,

$$\mathcal{L}_{\text{rep}} = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{|\bar{k}|}\sum_{i\notin k} q_i q_{\alpha_k}\max\left(0,\ 1 - \left\|\mathbf{c}_i - \mathbf{c}_{\alpha_k}\right\|\right). \quad (5)$$

A regularization term encourages condensation points to attain high confidence,

$$\mathcal{L}_{\beta} = \frac{1}{K}\sum_{k=1}^{K}\left(1 - \beta_{\alpha_k}\right). \quad (6)$$

The total event loss is

$$\mathcal{L}_{\text{OC}} = s_{\text{att}}\mathcal{L}_{\text{att}} + s_{\text{rep}}\mathcal{L}_{\text{rep}} + s_{\text{coward}}\mathcal{L}_{\beta}. \quad (7)$$

In our implementation, $s_{\text{att}} = s_{\text{rep}} = s_{\text{coward}} = 1.0$ and $q_{\min} = 0.1$, following the default configuration of Ref. [5]. A limited scan of the loss weights was performed and found to produce negligible changes in reconstruction performance relative to the default configuration. The model learns latent variables that define clustering behavior, while the actual grouping of hits is performed at inference time using these learned quantities.

For direct comparison, the OC model uses the same DynamicEdgeConv backbone described in Sec. II A, differing only in the output heads predicting β_i and $\mathbf{c}_i$. Training is performed eventwise with all hits assigned to their corresponding truth shower, consistent with the supervised setting used for contrastive learning. In contrast to CML, which learns a global similarity metric, OC directly optimizes object-centric clustering variables in the latent space. In this formulation, clustering structure is explicitly encoded through object-level attractors rather than emerging from pairwise representations.

D. Clustering

Clustering is performed after training and treated as a readout of the learned representation rather than part of the

TABLE I. Clustering thresholds. The parameter δ_{agg} denotes the agglomerative clustering distance threshold, while t_β and t_d correspond to thresholds used in both the density-based readout and object condensation inference.

Model	Method	δ_{agg}	t_β	t_d
EM	CML	8	0.4	0.45
	OC	6.75	0.1	0.25
HAD	CML	10	0.4	0.45
	OC	11	0.1	0.35
Mixed	CML	9.5	0.4	0.45
	OC	7.5	0.1	0.3

learning objective. All clustering is performed independently for each event.

a. Agglomerative clustering. As a common baseline, agglomerative clustering [17] is applied directly to the learned embeddings produced by both CML and OC. We use Ward linkage with a Euclidean distance metric applied to the ℓ_2-normalized embeddings; on the unit hypersphere, Euclidean and cosine distances are monotonically related. The method does not require a fixed number of clusters and naturally accommodates events with variable particle multiplicity. The number of clusters is determined by applying a distance threshold δ_{agg} to the Ward linkage dendrogram: Clusters separated by a distance greater than δ_{agg} are not merged. The value of δ_{agg} is optimized on the auxiliary dataset and reported in Table I. Applying the same clustering algorithm to both embeddings enables comparison of the intrinsic quality of the learned representations, independent of method-specific inference procedures.

b. Object condensation inference. For OC, clusters are obtained using the inference procedure associated with the OC loss [5]. Candidate condensation points are selected by thresholding the predicted scores $\beta_{\text{OC},i} > t_\beta$ and sorting in decreasing β_{OC}. A greedy separation step enforces a minimum Euclidean distance t_d between selected points in the learned clustering coordinate space $\mathbf{c}_i \in \mathbb{R}^{16}$. Hits within radius t_d of a condensation point are assigned to it, and any remaining hits are assigned to the nearest condensation point. The values of t_β and t_d are given in Table I.

c. Density-based readout of metric embeddings. Contrastive embeddings do not predict representative points. We therefore define a clustering readout operating directly in the embedding space, deriving candidate centers from local neighborhood structure rather than from a learned score.

For each hit, a local density estimate is obtained from the distance to its kth nearest neighbor $d_k(i)$. In this work, we use the same value of k for the density estimate as for the DynamicEdgeConv graph construction during training, although these two choices are conceptually independent. This distance is mapped to a score

$$\beta_{\text{CML},i} = \exp\left(-\frac{d_k(i)}{\tau}\right), \tag{8}$$

where τ is the temperature parameter used in the contrastive loss. This mapping is consistent with the similarity scaling used during training, assigning large scores to densely populated regions of the embedding.

Candidate centers are selected by thresholding $\beta_{\text{CML},i} > t_\beta$ and enforcing a minimum separation t_d. Clusters are formed by assigning hits within distance t_d to each center, followed by nearest-center assignment for remaining hits. Here, $\beta_{\text{CML},i}$ plays a role analogous to the condensation score in OC, but is computed from the embedding rather than predicted by a network head.

Unlike standard density-based methods such as DBSCAN [18] and HDBSCAN [19], which rely on fixed density thresholds, the proposed readout operates on local neighborhood structure in the learned embedding, making it compatible with spatially varying density across showers.

d. Summary. Three clustering strategies are evaluated: shared agglomerative clustering applied to both embedding spaces, the native OC inference procedure, and the proposed density-based readout for contrastive embeddings. This design isolates the effect of the learned representation from the choice of clustering algorithm.

III. EXPERIMENTAL SETUP

A. Dataset

The datasets used in this study are produced with a standalone Geant4 [20] simulation closely resembling the HGCAL detector [1].

The calorimeter model comprises 50 longitudinal layers: 28 electromagnetic (CE-E) followed by 22 hadronic (CE-H). The transverse cell size is $1 \times 1\,\text{cm}^2$ over a $100 \times 100\,\text{cm}^2$ area, ensuring full shower containment. Energy deposits are calibrated in minimum-ionizing-particle units, with the simulated hit response consistent with test-beam observations [21,22].

Three training datasets are considered: an electromagnetic (EM, electrons) sample, a hadronic (HAD, charged pions) sample, and a mixed sample containing both particle types. The EM dataset contains 2–10 electrons per event with primary energies uniformly distributed between 30 and 400 GeV. The HAD dataset contains 2–7 pions per event in the same energy range. The mixed dataset contains an equal fraction of electrons and charged pions with 2–7 particles per event and primary energies between 30 and 250 GeV. For each configuration, 200 000 events are used for training and 50 000 for validation.

Models are evaluated on independent test datasets containing between 1 and 30 particles per event with primary energies ranging from 30 to 600 GeV. Separate EM and HAD test samples are used. Models trained on the EM and HAD datasets are evaluated on their respective particle types, while the model trained on the mixed dataset is evaluated on both.

In addition to the training, validation, and primary test datasets, an auxiliary optimization dataset is generated for each training configuration with the same multiplicity and energy distributions as the corresponding training sample. These events are not used during network training and are used only to determine the clustering hyperparameters at inference time. Optimal clustering thresholds are selected by maximizing reconstruction performance on this dataset. For models trained on the EM and HAD samples, thresholds are determined separately. For the mixed model, a single set of thresholds is determined on the mixed optimization dataset

TABLE II. Spatial overlap metrics and average hit counts computed in detector coordinate space (x, y, z) at representative multiplicities. Metrics are defined in Sec. IV A and applied directly to the hit positions without any learned transformation. The large point cloud sizes and high contamination values at $N = 30$ illustrate the intrinsic difficulty of the segmentation task.

Dataset	$N_{\text{particles}}$	$R@1$	$R@10$	$C@10$	$\langle N_{\text{hits}} \rangle$
	2	0.980	0.996	0.022	2 180
EM	15	0.784	0.953	0.246	15 406
	30	0.711	0.924	0.334	26 493
	2	0.942	0.992	0.072	3 148
HAD	15	0.629	0.910	0.426	24 461
	30	0.509	0.846	0.549	43 661

and applied to both test samples. The resulting threshold values are summarized in Table I.

When multiple particles deposit energy in the same cell, their contributions are merged into a single hit, as in realistic detector readout. Each hit is assigned the label of the particle contributing the largest fraction of its deposited energy. The fractional contribution of the dominant particle is recorded to define hit purity, which is close to unity across the dataset. This quantity is not used as an input feature.

All particles are generated within a cone of fixed opening angle, with a half angle of 5° around the detector axis. Consequently, increasing the event multiplicity leads to a systematic increase in local particle density within the calorimeter volume. This provides a controlled mechanism to vary the degree of shower overlap, allowing reconstruction performance to be probed across regimes of increasing ambiguity and hit-level confusion.

Table II reports the average number of hits per event and the degree of spatial overlap between showers in raw detector coordinates, computed using the same neighborhood metrics defined in Sec. IV A, but applied directly to the (x, y, z) hit positions rather than learned embeddings. The point clouds are large: At $N = 2$, EM events contain on average 2180 hits and HAD events 3148 hits, rising to 26 493 and 43 661, respectively, at $N = 30$, reflecting the broader lateral and longitudinal extent of hadronic showers. The spatial overlap metrics confirm that the reconstruction task is intrinsically challenging. At $N = 30$, the nearest-neighbor recall $R@1$ drops to 0.711 for EM and 0.509 for HAD in detector space, with contamination $C@10$ reaching 0.334 and 0.549, respectively, indicating that for hadronic showers at high multiplicity roughly half of the ten nearest spatial neighbors of a given hit originate from a different shower. These values motivate the need for a learned representation that can disentangle showers that are not separable in input space.

B. Training details

All models are trained using the Adam optimizer [23] with an initial learning rate of 3×10^{-4} and a batch size of 16 events. The learning rate is reduced by a factor of 0.5 every 25 epochs, and training is performed for up to 100 epochs. Identical optimization settings, data splits, and initialization procedures are used for both CML and OC models, with differences arising only from the learning objective and corresponding output heads.

Model selection is based on the validation loss evaluated after each epoch on an independent validation set. The checkpoint with the lowest validation loss is retained, and early stopping is applied once no further improvement is observed.

Training is performed using mixed-precision arithmetic with gradient scaling. All models are trained on a single GPU (NVIDIA Tesla V100, 32 GB, or NVIDIA H100, 80 GB), with consistent results observed across both hardware configurations.

IV. EVALUATION

A. Embedding geometry metrics

The learned representations are evaluated independently of any clustering procedure by quantifying geometric properties of the embedding space within each event. Such metrics are commonly used to assess representation quality in metric learning settings [24,25]. All quantities are computed using cosine similarity on ℓ_2-normalized embeddings and averaged over events.

*a. Recall@k (*R@*k).* For each hit, the k nearest neighbors in embedding space are identified. A hit is considered correctly retrieved if at least one neighbor originates from the same truth shower. The fraction of such hits defines $R@k$. We report $R@1$ and $R@10$, which probe whether compatible hits are placed as the closest neighbor and within the local neighborhood, respectively.

*b. Contamination@10 (*C@*10).* Local neighborhood purity is quantified as the fraction of the ten nearest neighbors that originate from a different truth shower. Unlike Recall@k, which provides a binary criterion, this metric captures the continuous fraction of incorrect neighbors. Lower contamination indicates that local neighborhoods correspond more closely to physically consistent showers.

c. Pairwise separability. Positive (same-shower) and negative (different-shower) hit pairs are sampled and their cosine similarities compared. The area under the receiver operating characteristic (AUC) curve measures the probability that a positive pair has higher similarity than a negative pair, providing a global measure of separability in the embedding space.

d. Intra- and intershower distances. To characterize the geometric structure of individual showers, the analysis is restricted to the energetic core of each shower by retaining the highest-energy hits that together account for 95% of the total shower energy. This suppresses low-energy peripheral deposits while preserving the physically relevant structure.

Cosine distances $d = 1 - \mathbf{z}_i^\top \mathbf{z}_j$ are computed between hits. Intrashower compactness is quantified by the 99th percentile of same-shower pairwise distances, denoted by d_{intra}^{Q99}, capturing the largest separations within a shower core. Intershower separation is measured as the median distance between hits from different showers, providing a robust estimate of typical separation between distinct objects.

B. Reconstruction metrics

The performance of reconstructed particle candidates is evaluated using standard high-energy physics metrics following the CMS TICL validation procedure [26]. These quantities assess the physical quality of reconstructed objects rather than the learned representation.

a. Efficiency. A simulated particle is considered reconstructed if at least 70% of its deposited energy is contained within a single reconstructed object. The reconstruction efficiency is defined as

$$\varepsilon = \frac{N_{\text{matched sim}}}{N_{\text{sim}}}.$$

b. Purity. Purity quantifies the extent to which a reconstructed object originates from a single simulated particle. This is evaluated using the Reco-to-Sim score $S(r \to s)$, defined for a reconstructed object r and simulated particle s as

$$S(r \to s) = \frac{\sum_{\text{hits}} \max(0,\ f_{\text{reco}} - f_{\text{sim}})^2 E^2}{\sum_{\text{hits}} f_{\text{reco}}^2 E^2}, \tag{9}$$

where f_{reco} and f_{sim} are the fractional energy contributions of a hit to the reconstructed object and simulated particle, respectively, and E is the hit energy. The score is zero for a perfectly pure object and increases with contamination. A reconstructed object is considered pure if there exists at least one simulated particle s for which $S(r \to s) < 0.2$. The purity is then defined as

$$P = \frac{N_{\text{pure reco}}}{N_{\text{reco}}}.$$

c. Multiplicity ratio. To quantify the tendency of the clustering to split or merge objects, the ratio

$$R_N = \frac{N_{\text{reco}}}{N_{\text{sim}}}$$

is used, where $R_N > 1$ indicates object splitting and $R_N < 1$ indicates merging.

d. Energy resolution. For each simulated particle, the reconstructed object with the largest shared energy is selected as the match. Objects with $S(r \to s) > 0.5$ are discarded as impure. The energy response is defined as

$$r = \frac{E_{\text{reco}}}{E_{\text{sim}}},$$

where E_{reco} is the sum of the energies of the hits assigned to the reconstructed object and E_{sim} is the true particle energy.

Within bins of E_{sim}, the mean response μ is used to define a calibrated response $r' = r/\mu$. The energy resolution is then defined as the effective sigma σ_{eff}, given by half the minimum interval containing 68.3% of the r' distribution. For Gaussian distributions, σ_{eff} is equivalent to the standard deviation, while remaining robust to non-Gaussian tails and outliers [27].

As a reference, we also compute the resolution of an ideal pattern-recognition algorithm that perfectly assigns all hits to their truth shower, with no merging or splitting. The energy of each reconstructed object is taken as the sum of the energies of all hits assigned to the corresponding truth shower. This provides a lower bound on the achievable resolution given the detector response and the hit-level energy calibration, and is independent of any clustering algorithm.

V. RESULTS

A. Embedding geometry

Embedding-level performance is evaluated separately for the EM and HAD datasets using the metrics defined in Sec. IV A. Results are reported at representative event multiplicities to probe behavior as shower overlap increases.

TABLE III. Embedding-geometry metrics for EM and HAD showers at representative multiplicities. CML consistently achieves higher Recall@k, lower contamination@10, and higher AUC than OC, with the performance gap increasing as multiplicity increases. Bold values indicate the better-performing method for each metric at each multiplicity.

			$R@k$			
Dataset	$N_{\text{particles}}$	Method	1	10	$C@10$	AUC
EM	2	CML	**0.941**	**0.992**	**0.067**	**0.956**
		OC	0.935	0.989	0.072	0.944
	15	CML	**0.691**	**0.932**	**0.339**	**0.950**
		OC	0.667	0.920	0.363	0.923
	30	CML	**0.572**	**0.886**	**0.467**	**0.945**
		OC	0.536	0.871	0.503	0.918
HAD	2	CML	**0.938**	**0.992**	**0.077**	**0.945**
		OC	0.922	0.989	0.092	0.927
	10	CML	**0.744**	**0.943**	**0.286**	**0.927**
		OC	0.698	0.927	0.330	0.889
	20	CML	**0.628**	**0.900**	**0.407**	**0.917**
		OC	0.573	0.876	0.460	0.877

Across both datasets, CML consistently outperforms OC, achieving higher Recall@k, lower contamination@10, and higher AUC at all tested multiplicities (Table III). The performance gap is modest at low multiplicity but increases steadily with shower density, indicating that CML preserves more stable local neighborhood structure under increasing overlap.

Figure 1 shows the corresponding embedding-geometry distributions. The top row presents the median intershower separation $d_{\text{inter}}^{\text{median}}$, while the middle row shows the intrashower distance tail $d_{\text{intra}}^{\text{Q99}}$. For CML, both quantities remain narrowly distributed across all multiplicities, indicating a stable embedding geometry with consistent distance scales both within and between showers. In contrast, OC produces much broader distributions, reflecting substantially larger event-to-event variability.

The bottom row shows the separation margin

$$\Delta = d_{\text{inter}}^{\text{median}} - d_{\text{intra}}^{\text{Q99}},$$

which directly measures whether showers remain geometrically separable. At $N = 2$, the margin is broad for both methods, as with only two showers per event absolute distances are unconstrained; however, the CML margin remains centered at positive values, confirming separability even in this regime. At higher multiplicities, where genuine shower overlap is present, CML maintains a narrow margin distribution with a positive peak for EM showers and a near-zero peak for HAD showers, indicating a well-defined and stable clustering scale for both particle types.

In contrast, OC produces broad and often negative margin distributions. For EM showers, the margin spans both positive and negative values, implying strong overlap between intra- and intershower distances. For HAD showers, the margin is systematically negative and becomes increasingly broad

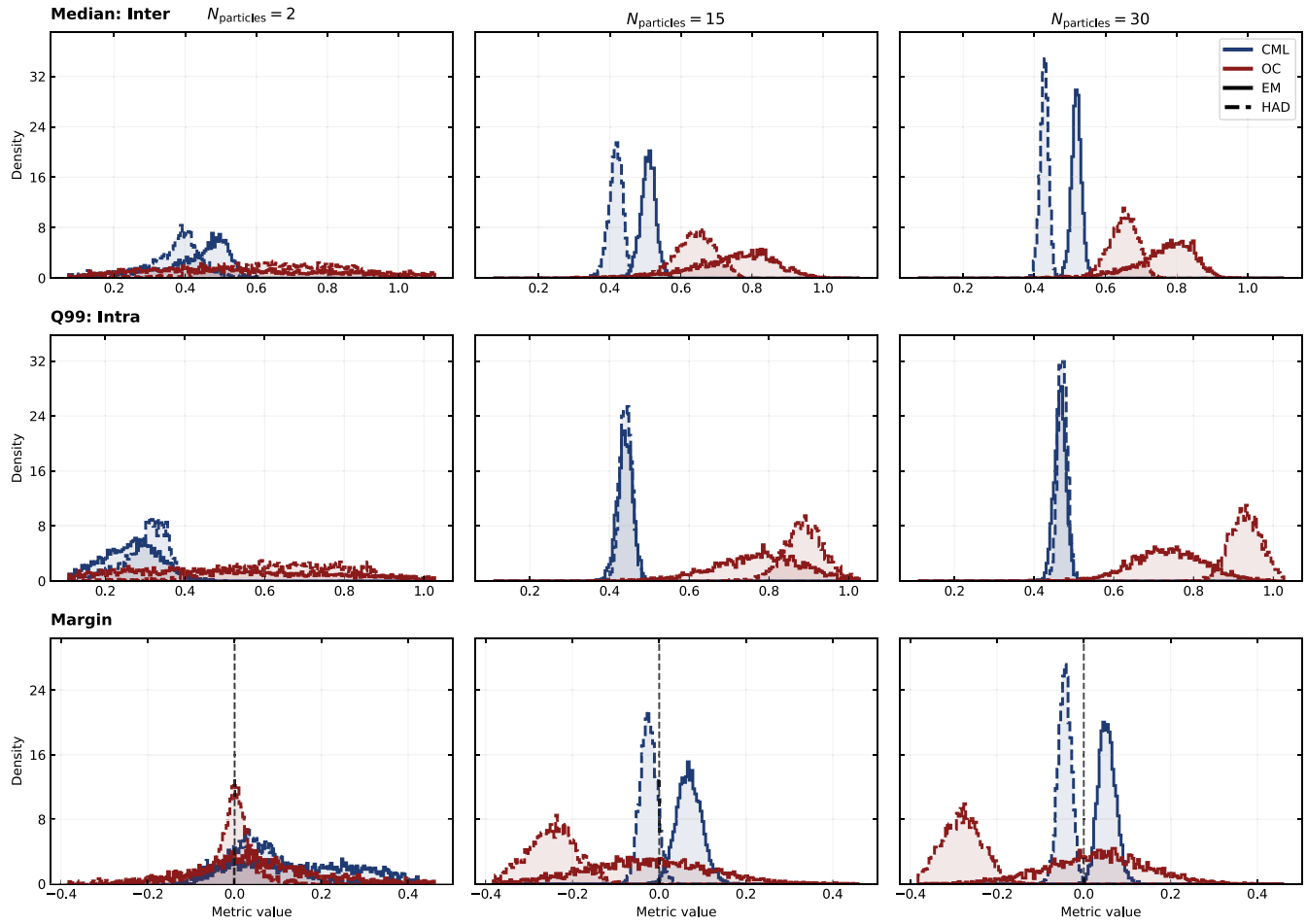


FIG. 1. Event-level embedding-geometry distributions for EM and HAD showers at increasing generated particle multiplicity ($N_{\text{particles}}$). Top: median intershower separation $d_{\text{inter}}^{\text{median}}$. Middle: intrashower distance tail $d_{\text{intra}}^{\text{Q99}}$. Bottom: separation margin $\Delta = d_{\text{inter}}^{\text{median}} - d_{\text{intra}}^{\text{Q99}}$. Narrow positive or near-zero margins indicate a well-defined clustering scale, while broad or negative margins indicate increasing ambiguity between showers.

with multiplicity. In both cases, no single distance threshold can separate showers consistently across events.

Overall, the advantage of CML arises primarily from controlling the tails of the distance distributions rather than shifting their central values. By maintaining compact intrashower structure together with stable intershower separation, CML produces embeddings that remain reliably clusterable even in dense environments.

B. Reconstruction performance

Reconstruction performance is evaluated using the physics metrics defined in Sec. IV B. Across all datasets and reconstruction configurations, CML outperforms OC, with the largest gains appearing at high multiplicity. These differences follow directly from the embedding geometry: The narrow margin distributions of CML define a stable clustering scale, whereas the broad and often negative margins of OC make clustering increasingly ambiguous.

a. Energy dependence. Figure 2 shows reconstruction efficiency and purity as a function of the primary particle energy. For the dedicated EM model, CML achieves near-perfect efficiency (∼99%) and substantially higher purity (∼80%–82%) than OC, which remains below 85% efficiency and 72% purity. For HAD showers, efficiencies are more similar across methods (∼70%–76%), but CML retains a clear purity advantage of ∼10%–15%.

The mixed-trained model reveals a stronger separation. For EM showers, OC degrades severely, with efficiencies dropping to ∼30%–40%, whereas CML retains efficiencies above 80% and substantially higher purity. For HAD showers, efficiencies remain comparable, but CML again achieves higher purity.

Across all cases, performance remains stable beyond the training range, indicating that the dominant differences arise from the learned representation rather than from any strong dependence on primary energy. This is consistent with the embedding geometry, for which the separation margin is controlled mainly by local shower structure rather than absolute shower energy.

b. Multiplicity dependence. Figure 3 shows performance as a function of the particle multiplicity. This is the regime in which the separation between methods is most pronounced.

At low multiplicity, all methods perform similarly. As multiplicity increases, clear differences emerge. For EM showers

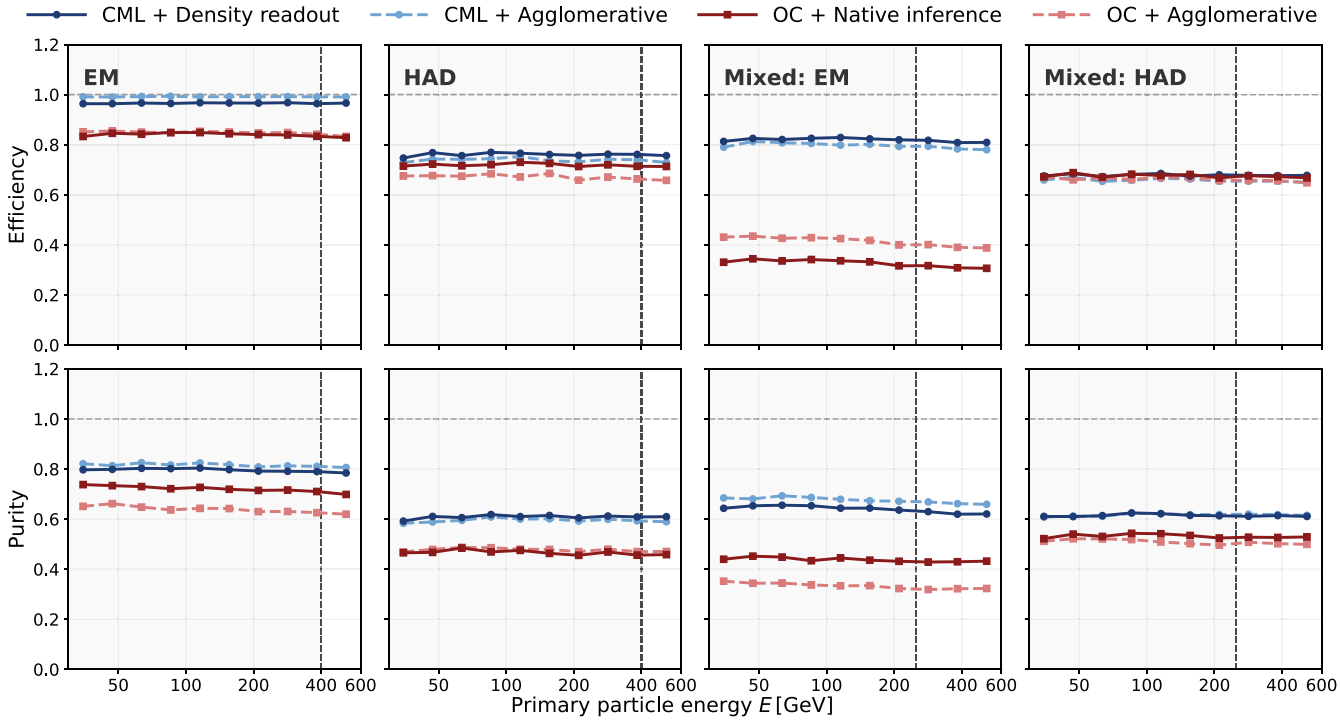


FIG. 2. Reconstruction efficiency (top) and purity (bottom) as functions of primary particle energy for EM and HAD showers. Columns show models trained on EM, trained on HAD, and a mixed-trained model evaluated separately on EM and HAD showers. The vertical dashed line indicates the upper boundary of the training energy range. The dominant differences are largely independent of energy, indicating that performance is controlled primarily by the learned representation rather than by the absolute shower energy.

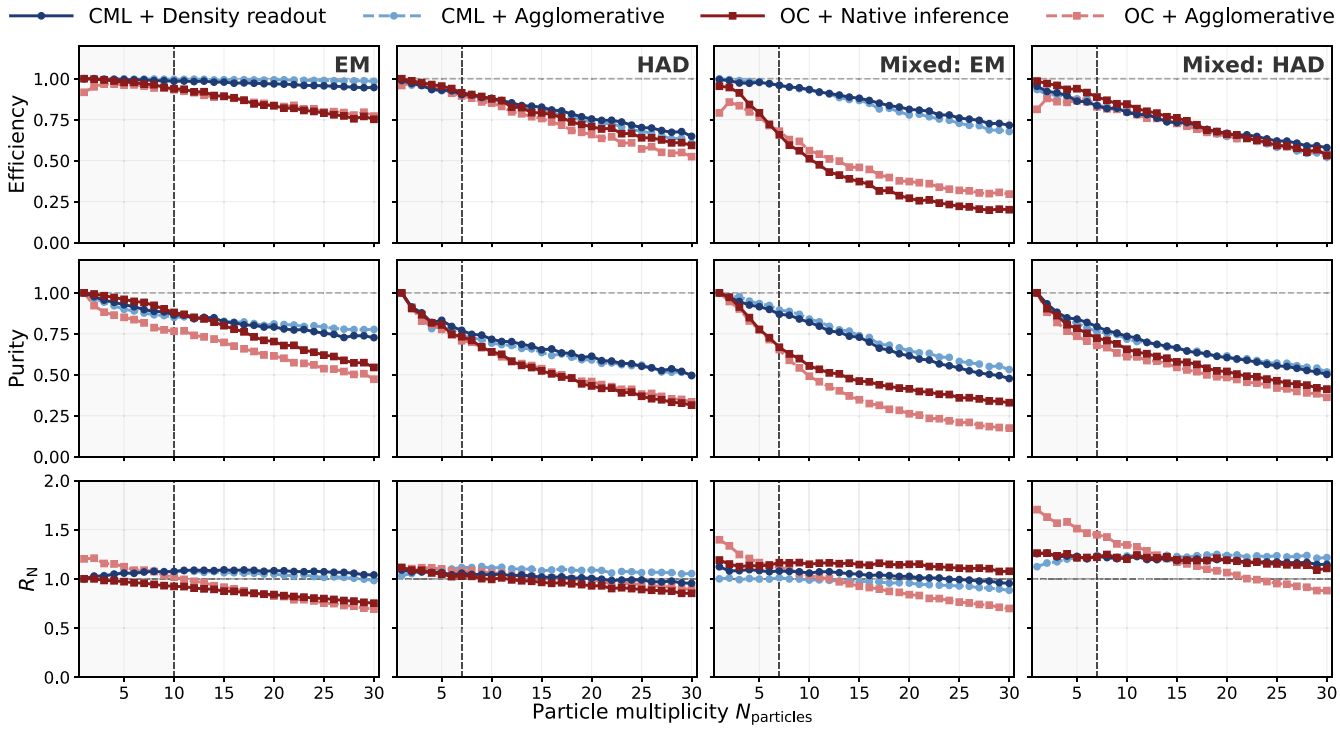


FIG. 3. Reconstruction efficiency (top), purity (middle), and number ratio (bottom) as functions of particle multiplicity N for EM and HAD showers. Columns show models trained on EM, trained on HAD, and a mixed-trained model evaluated separately on EM and HAD showers. The vertical dashed line indicates the upper boundary of the training multiplicity range. The performance gap between CML and OC increases strongly with multiplicity, showing that clustering stability in dense environments is determined by the underlying embedding geometry.

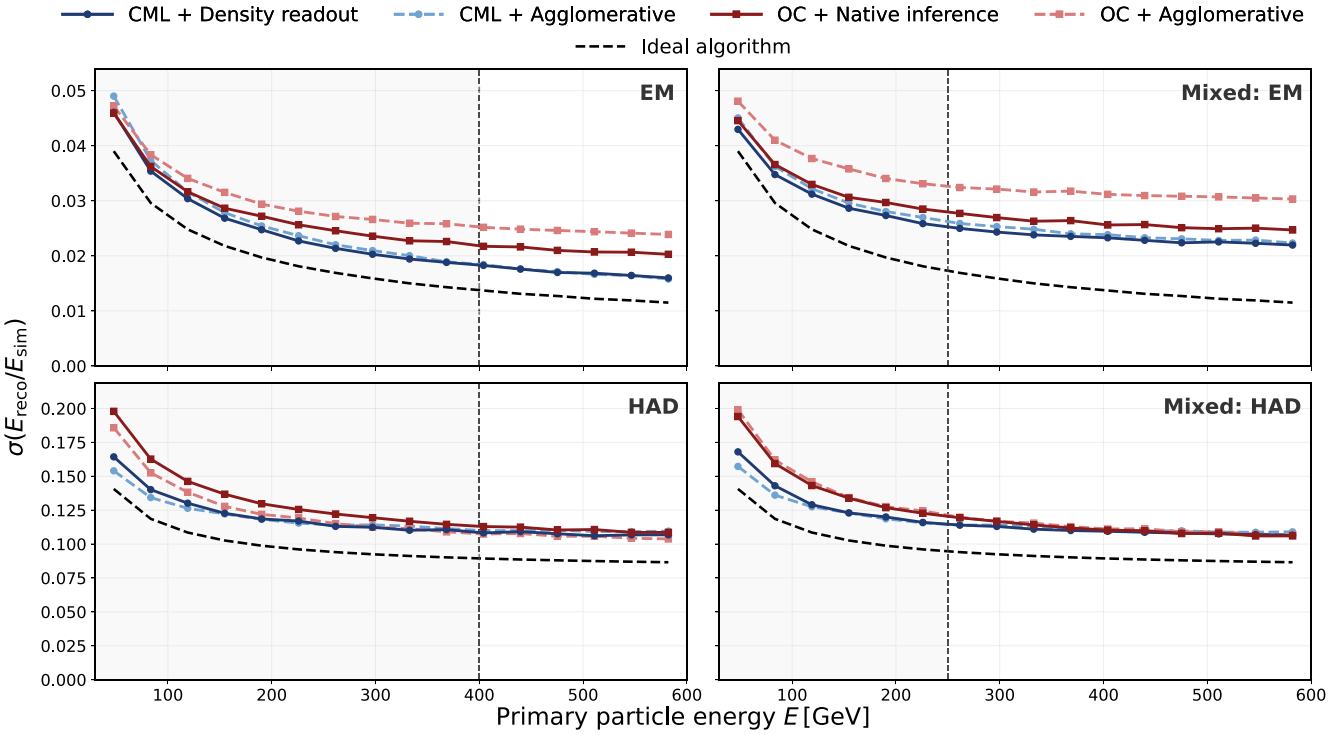


FIG. 4. Energy resolution as a function of primary particle energy for EM and HAD showers. Panels show models trained on EM, a mixed-trained model evaluated on EM, trained on HAD, and a mixed-trained model evaluated on HAD. The vertical dashed line marks the upper boundary of the training energy range. The black dashed curve denotes the ideal pattern-recognition limit, defined as the resolution obtained by a perfect clustering algorithm with no merging or splitting, and represents a lower bound on the achievable resolution given the detector response. Improvements in CML resolution follow directly from its improved clustering purity and reduced merging. Results are averaged over the full test dataset spanning multiplicities from 1 to 30 particles per event.

at $N = 30$, CML maintains high efficiency (∼95%–98%) and purity (∼73%–78%), while OC degrades to ∼75% efficiency and ∼47%–55% purity. The number ratio remains close to unity for CML but drops significantly for OC, indicating substantial merging.

A similar trend is observed for HAD showers. At $N = 30$, CML improves purity by nearly 20 percentage points over OC while also providing a modest but systematic efficiency gain. Again, CML maintains a more stable R_N, indicating improved control of merging in dense environments.

The mixed-trained model shows the strongest contrast. For EM showers at $N = 30$, CML remains functional with efficiencies of ∼70%, whereas OC collapses to ∼20%–30%. For HAD showers, efficiencies remain comparable, but CML retains a clear purity advantage.

These trends can be understood directly from the embedding margins. For CML, the margin distributions remain narrow and concentrated near the separation boundary across multiplicities, so a single clustering scale continues to separate showers even as overlap increases. For OC, the behavior depends strongly on shower type. In EM events, the margin becomes extremely broad, so no global threshold can separate showers consistently; this explains the large degradation in R_N and purity, particularly for agglomerative clustering. In HAD events, the OC margin is less broad but systematically negative, so clustering is somewhat more stable but still intrinsically ambiguous, leading to persistent misclustering and reduced purity. The mixed setting makes this contrast most explicit: CML produces similar margin distributions for EM and HAD showers and therefore supports a common threshold, whereas OC does not, causing the EM performance of the mixed model to collapse while the HAD performance remains closer to that of the HAD-only model.

c. Energy resolution. Figure 4 shows the corresponding energy resolution. For EM showers, CML consistently achieves the best resolution and remains closest to the ideal pattern-recognition limit. At high energy, $\mathcal{O}(600\ \mathrm{GeV})$, CML reaches ∼1.6%, compared with ∼2.0% for OC with native inference and ∼2.4% for OC with agglomerative clustering.

For HAD showers, the separation is smaller but remains systematic. The CML method provides the largest improvement at low energy, where the resolution is dominated by clustering quality rather than intrinsic detector response (e.g., 15.4% versus 18%–20% at 50 GeV), while at higher energies the curves approach a constant term and the gap narrows.

These results are the direct consequence of the clustering behavior. The stable separation margins of CML reduce both merging and fragmentation, leading to more accurate assignment of energy deposits to reconstructed objects. In contrast, the broad and negative margins of OC lead to systematic misassignment of energy, particularly in dense environments, and therefore to degraded resolution.

d. Robustness to simulation variations. To assess the sensitivity of both methods to simulation mismodeling, we evaluate the HAD model under three systematic variations: a 1.2× scaling of the hadronic interaction length, a change of Geant4

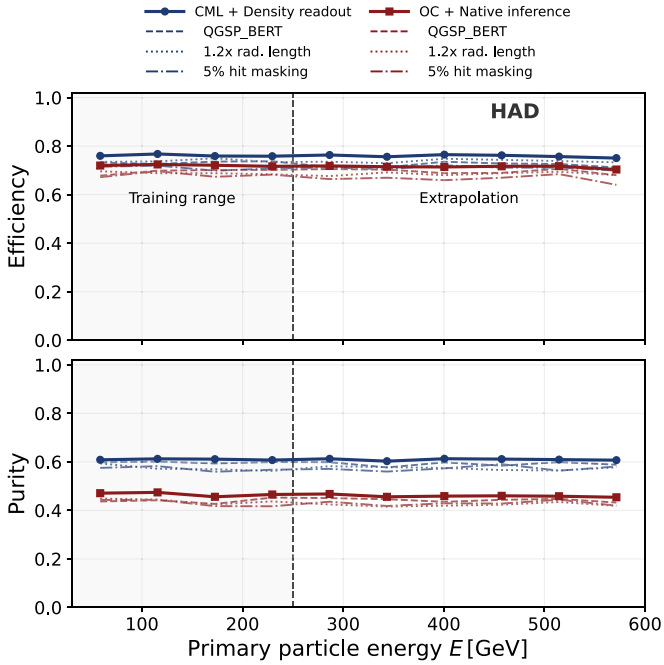


FIG. 5. Reconstruction efficiency (top) and purity (bottom) for the HAD model as functions of primary particle energy under three systematic variations: 1.2× hadronic interaction length scaling, QGSP_BERT physics list, and 5% random hit masking. The nominal performance is shown for reference. Neither method shows significant degradation, and the performance gap between CML and OC is preserved under all variations.

physics list from FTFP_BERT to QGSP_BERT, and random masking of 5% of hits. Figure 5 shows reconstruction efficiency and purity for CML and OC under these variations. In all cases, both methods show only modest degradation relative to the nominal, and the performance gap between CML and OC is preserved across all tested variations. This indicates that the relative advantage of CML is stable under the types of simulation mismodeling considered here.

e. Dependence on embedding dimensionality. Table IV shows the mean reconstruction performance of the mixed-trained model for 16- and 4-dimensional embeddings, evaluated on an independent mixed sample. Reducing the embedding dimension leads to a modest degradation for all methods, but the relative ordering remains unchanged: CML continues to outperform OC in both purity and efficiency across all configurations. This shows that the observed advantage is not specific to the 16-dimensional setting. The dimension of 16 was selected following a preliminary sweep over 4, 16, 128, and 512 dimensions; performance saturates at 16 for both methods, with degradation at higher dimensionalities consistent with distance concentration in high-dimensional spaces. The four-dimensional result is included as a lower bound relevant to OC-based approaches, where lower-dimensional clustering coordinates are common.

TABLE IV. Mean reconstruction purity and efficiency for the mixed-trained model evaluated on the combined EM and HAD test sample for 16- and 4-dimensional embeddings. The relative ordering between methods is unchanged when the embedding dimensionality is reduced, showing that the CML advantage is not driven by latent-space size alone.

Model	Dimension	Clustering	Purity	Efficiency
CML	16	Agglomerative	0.722	0.932
	16	Density	0.710	0.945
	4	Agglomerative	0.690	0.887
	4	Density	0.637	0.905
OC	16	Agglomerative	0.469	0.849
	16	Density	0.520	0.838
	4	Agglomerative	0.512	0.813
	4	Density	0.500	0.866

VI. DISCUSSION AND CONCLUSIONS

We have presented a clustering framework for point cloud segmentation in high-granularity calorimeters based on CML. Rather than learning object-centric clustering variables, the method learns a representation in which hits from the same shower are placed nearby and hits from different showers are separated, with clustering applied only as a readout of the learned geometry. This decoupling allows the representation to be optimized for pairwise compatibility while retaining flexibility in the choice of inference procedure.

The central result of this work is that the learned embedding geometry directly determines clustering performance. The CML approach produces narrow separation-margin distributions that remain positive for EM showers and only slightly negative for HAD showers, indicating a stable and well-defined clustering scale even in dense environments. In contrast, OC yields substantially broader and often negative margin distributions, particularly for EM showers, implying strong overlap between intra- and intershower distances and therefore intrinsically ambiguous clustering decisions. These geometric differences explain the observed reconstruction behavior: CML consistently achieves higher purity, higher efficiency, more stable R_N, and improved energy resolution, with the largest gains appearing at high multiplicity where shower overlap is most severe.

The mixed-training results provide the strongest evidence for stability across heterogeneous particle types. The CML method maintains similar separation scales for EM and HAD showers, allowing a single clustering threshold to operate effectively across both particle types. In contrast, OC learns different geometric structure for EM and HAD showers. This leads to a pronounced degradation for EM showers in particular, indicating that the object-centric formulation is less able to accommodate heterogeneous shower topologies within a single model.

Taken together, these results show that, for highly granular calorimeter reconstruction, learning a stable similarity geometry is more effective than learning explicit object-centric clustering variables. More broadly, they suggest that contrastive metric learning provides a robust alternative for dense point cloud segmentation problems in which object boundaries are ambiguous, overlap is common, and inference must remain stable under changing event complexity. We propose to test this strategy with ultrarealistic simulations of the HG-

CAL detector in high-pileup conditions, as implemented in the CMS software stack.

ACKNOWLEDGMENTS

We thank Sunanda Banerjee for the help in creating a realistic dataset corresponding to particle showers in a highly granular detector. The neural networks in this study have been trained on the Imperial College RCS HPC cluster and the Oscar cluster at Brown University. L.G. and L.N. were supported by the DOE, Office of Science, Office of High Energy Physics Early Career Research program under Award No. DE-SC0026288. B.M. acknowledges the support of Schmidt Sciences.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

[1] CMS Collaboration, *The Phase-2 Upgrade of the CMS Endcap Calorimeter*, Technical Report (CERN, Geneva, 2017), https://cds.cern.ch/record/2293646.

[2] X. Ju *et al.*, Graph neural networks for particle reconstruction in high energy physics detectors, arXiv:2003.11603.

[3] L. Ehrke, J. A. Raine, K. Zoch, M. Guth, T. Golling, Topological reconstruction of particle physics processes using graph neural networks, Phys. Rev. D **107**, 116019 (2023).

[4] J. Duarte and J.-R. Vlimant, Graph neural networks for particle tracking and reconstruction, in *Artificial Intelligence for High Energy Physics*, edited by P. Calafiura, D. Rousseau, and K. Terao (World Scientific, Singapore, 2022), pp. 387–436.

[5] J. Kieseler, Object condensation: One-stage grid-free multi-object reconstruction in physics detectors, graph, and image data, Eur. Phys. J. C **80**, 886 (2020).

[6] G. Matousek and A. Vossen, AI-assisted object condensation clustering for calorimeter shower reconstruction at CLAS12, Nucl. Instrum. Methods Phys. Res. A **1082**, 170990 (2026).

[7] 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).

[8] S. Gardner, R. Tyson, D. Glazier, and K. Livingston, Object condensation for track building in a backward electron tagger at the EIC, J. Instrum. **19**, C05052 (2024).

[9] R. Hadsell, S. Chopra, and Y. LeCun, Dimensionality reduction by learning an invariant mapping, in *Proccedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* (IEEE Computer Society, 2006), pp. 1735–1742.

[10] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, Dynamic graph CNN for learning on point clouds, ACM Trans. Graph. **38**, 146 (2019).

[11] A. van den Oord, Y. Li, and O. Vinyals, Representation learning with contrastive predictive coding, arXiv:1807.03748.

[12] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, A simple framework for contrastive learning of visual representations, in *Proceedings of the 37th International Conference on Machine Learning (ICML 2020)* (PMLR, 2020), Vol. 119, pp. 1597–1607.

[13] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, Supervised contrastive learning, in *Advances in Neural Information Processing Systems* (NeurIPS, 2020), Vol. 33, pp. 18661–18673.

[14] P. H. Le-Khac, G. Healy, and A. F. Smeaton, Contrastive representation learning: A framework and review, IEEE Access **8**, 193907 (2020).

[15] Y. Li, P. Hu, Z. Liu, D. Peng, J. T. Zhou, and X. Peng, Contrastive clustering, in *Proceedings of the AAAI Conference on Artificial Intelligence* (AAAI, 2021), Vol. 35, pp. 8547–8555.

[16] K. Lieret, G. DeZoort, D. Chatterjee, J. Park, S. Miao, and P. Li, High pileup particle tracking with object condensation, arXiv:2312.03823.

[17] D. Müllner, Modern hierarchical, agglomerative clustering algorithms, arXiv:1109.2378.

[18] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, A density-based algorithm for discovering clustersin large spatial databases with noise, in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (AAAI Press, Menlo Park, CA, 1996), pp. 226–231.

[19] L. McInnes, J. Healy, and S. Astels, hdbscan: Hierarchical density based clustering, J. Open Source Software **2**, 205 (2017).

[20] S. Agostinelli *et al.*, Geant4—A simulation toolkit, Nucl. Instrum. Methods Phys. Res., Sect. A **506**, 250 (2003).

[21] N. Akchurin, Response of a CMS HGCAL Silicon-pad electromagnetic calorimeter prototype to 20–300 GeV positrons, Tech. Rep. (CERN, Geneva, 2021), https://cds.cern.ch/record/2798347.

[22] B. Acar *et al.* (CMS Collaboration), Performance of the CMS High Granularity Calorimeter prototype to charged pion beams of 20–300 GeV/c, arXiv:2211.04740.

[23] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, *Proceedings of the 3rd International Conference on Learning Representations* (ICLR, San Diego, CA, 2015).

[24] D. Zhang, Y. Li, and Z. Zhang, Deep metric learning with spherical embedding, *Advances in Neural Information Processing Systems* (NeurIPS, 2020), Vol. 33, pp. 18772–18783.

[25] K. Musgrave, S. Belongie, and S.-N. Lim, A metric learning reality check, *Computer Vision, ECCV 2020*, Lecture Notes in Computer Science (Springer, Cham, 2020), Vol. 12370, pp. 681–699.

[26] CMS Collaboration, The iterative clustering (TICL) (v5a) reconstruction at the CMS phase-2 high granularity calorimeter endcap, 2024, https://cds.cern.ch/record/2920448.

[27] CMS Collaboration, Energy calibration and resolution of the CMS Electromagnetic Calorimeter in pp collisions at $\sqrt{s} = 7$ TeV, J. Instrum. **8**, P09009 (2013).