Learning IRC-Safe Jet Clustering with Geometric Algebra Transformers

Gregor Kržmanc
ETH Zürich
Zürich, Switzerland
gregor.krzmanc@cern.ch

Roberto Seidita
ETH Zürich
Zürich, Switzerland
rseidita@phys.ethz.ch

Annapaola de Cosa
ETH Zürich
Zürich, Switzerland
adecosa@phys.ethz.ch

Abstract

Jet clustering is a key algorithm in particle physics used to detect sprays of particles produced as a result of the hadronization of highly energetic quarks and gluons produced in particle colliders. We present a novel machine learning-based jet clustering algorithm based on the Lorentz-equivariant geometric algebra transformer trained with object condensation loss to cluster particle flow candidates around the originating quarks. We propose an additional loss term that prevents the model from relying on features of the hadronization process not well described by theoretical models (IRC safety). The models are evaluated on events resulting from quantum chromodynamics (QCD) processes, as well as a possible dark sector signature in the form of semi-visible jet signal events. Our results demonstrate an improvement in jet clustering across a wide range of parameters of dark sector models compared to the standard in the field (anti-kt), regardless of whether the models were trained on QCD background or semi-visible jet signal events. Comparison of the performance on generator particles before and after hadronization demonstrates a high degree of IRC safety of the models. The IRC safety loss can be extended to existing algorithms for processing high-energy physics data from particle colliders, thereby facilitating a more accurate reconstruction of the underlying physics processes.

1 Introduction

The Large Hadron Collider (LHC) at CERN collides bunches of particles at high energies [1]. The experiments store a subset of the recorded collision events for downstream physics analyses, selected by a two-level trigger system [2]. The final-state particles interact with different components of the detectors, and the resulting signals are then input to the Particle-Flow algorithm [3], which determines the final set of reconstructed particles (PF candidates) on which analysis is finally performed. The CMS and ATLAS experiments at the LHC have, among other results, confirmed the existence of the Higgs boson [4, 5], as well as considerably constrained parameters of a wide range of theories beyond the standard model (BSM).

Quarks and gluons are produced abundantly during proton-proton collisions. Such particles rapidly undergo a process known as hadronization, producing a large number of particles called hadrons. This leaves a distinct signature in the detector, known as a jet, i.e., a narrow cone of hadrons originating from a gluon or a quark. The momentum of the originating quark or gluon is spread among the resulting hadrons, making accurate jet finding from reconstructed particles a key ingredient to physics analysis [6]. A jet clustering algorithm should be infrared and collinear (IRC) safe: firstly, the algorithm's outcome shouldn't change if an arbitrary number of low-energy particles are added to the event (infrared safety). Secondly, if a particle is split into multiple collinear particles, the resulting jets should also not change (collinear safety). This is necessary to accurately compare experimental

results with theoretical predictions, which are known to become unreliable in the aforementioned regimes.

This work focuses on a subset of BSM theories in which semi-visible jets (SVJ) can occur. Such jets consist of both visible and invisible hadrons. As such, reconstructing these jets is more challenging than reconstructing standard model (SM) jets. We attempt to construct a clustering model that outperforms the classical anti-kt baseline while maintaining a high degree of IRC safety.

2 Strongly coupled dark matter searches at the CMS experiment

In Hidden Valley theories, additional neutral dark matter particles are introduced that can interact with the Standard Model (SM) particles [7, 8]. In strongly coupled dark matter theories, the hidden sector couples to the SM via a heavy mediator particle, which can decay into two dark quarks. The resonant production of a heavy boson mediator Z' is shown in Figure 1(a). We consider two parameters of such models: the mediator mass $m_{Z'}$ and invisible fraction r_{inv}. The invisible fraction is defined as the fraction of stable dark hadrons: $r_{inv} = \langle \frac{N_{stable}}{N_{stable}+N_{unstable}} \rangle$, as the unstable dark bound states decay promptly to SM quarks [9]. The CMS experiment [10] searched for such events using data from 2015 to 2018, and was able to exclude the existence of mediators with masses ranging from 1.5 TeV to 5.1 TeV at the 95% confidence level [9]. At lower mediator masses, the current triggering criteria significantly decrease the sensitivity of the search [11].

3 Dataset

We simulate s-channel production of dark quarks via a Z' vector mediator using different sets of parameters $m_{Z'}$ and $r_{inv.}$. The physics processes are simulated with Pythia 8.307 [12], and the detector simulation is done using Delphes 3.5.0 [13]. Additionally, we also simulate a background dataset of standard model QCD events, in which quarks and gluons hadronize. For more details on the simulation, see Appendix A. For each model, we store three particle sets: the reconstructed PF candidates (termed PFCands), visible final-state generator particles (GL), i.e., before simulating their interaction with the detector, and visible parton-level generator particles (PL), i.e., before the formation of a bound state (hadronization). The GL and PL sets are stored in order to evaluate the IRC safety of the models by comparing the performance on GL and PL with the IRC-safe baseline AK.

4 Anti-kt baseline

The anti-kt (AK) algorithm [14] works iteratively by picking the highest-p_T particle, and then keeps adding the closest particles to the jet one-by-one according to the metric $d_{ij} = \min(p_{T,i}^{-2}, p_{T,j}^{-2})\frac{\Delta_{ij}^2}{R^2}$. This continues until the closest particle is further away than $d_{iB} = p_{T,i}^{-2}$. Then, the selected particles are removed from the event, and the process is repeated until no more particles remain. The distance depends firstly on the angular distance between two particles $\Delta_{ij} = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, where ϕ is the azimuthal angle in the plane perpendicular to the beam axis, and $y = \frac{1}{2}\ln\frac{E+p_z c}{E-p_z c}$, where p_z is the projection of particle momentum onto the beam axis and c is the speed of light. Secondly, it depends on the largest of the particles' transverse momenta and the hyperparameter R, set to 0.8 for this work.

5 Model

Multivectors are objects that contain components of both scalar and vector nature, as well as higher-order components. Geometric algebra transformer (GATr) processes multivector tokens instead of only vectors or scalars, and is designed to respect the symmetries of Euclidean space by respecting E(3) equivariance. The attention mechanism used in the GATr uses the equivariant inner product between the multivectors. [15]

GATr could, in principle, be built over any geometric algebra. Therefore, we use the Lorentz-equivariant geometric algebra transformer (L-GATr) [16] to process four-vectors embedded in the

Lorentz geometric algebra, as it is particularly suitable for processing sets of particles produced in collisions. The novelty of this work is mainly in the loss and not the architecture itself; therefore, comparison between different transformer architectures is beyond the scope of this paper. We encode the particle four-momentum (a four-component object that represents a vector in the Lorentz geometric algebra) in the vector part of the multivectors. Additionally, we encode the particle charge as an additional scalar.

5.1 Object condensation loss

We base our clustering model on object condensation [17], followed by performing the HDBSCAN algorithm [18] in the clustering space output by the model. Assuming there are N particles and K objects in an event, the object condensation loss function is

$$\mathcal{L}_{OC} = \frac{1}{N} \sum_{j=1}^{N} q_j \sum_{k=1}^{K} \left(M_{jk} \check{V}_k(x'_j) w_{attr.} + (1 - M_{jk}) \hat{V}_k(x'_j) w_{repul.} \right) \tag{1}$$

where the mask M_{jk} is set to 1 if particle j belongs to ground truth jet k and 0 otherwise, and x'_i are the virtual coordinates of the particles regressed by the model[1]. We set $w_{attr.} = 0.1$ and $w_{repul.} = 1$. The attractive loss is defined as $\check{V}_k(x) = |x - x_\alpha|^2 q_{\alpha k}$ and the repulsive loss is defined as $\hat{V}_k(x) = \max(0, 1 - |x - x_\alpha|) q_{\alpha k}$, where $q_{\alpha k} = \max_i q_i M_{ik}$. We set $q_i = p_{T,i}$, motivated by the IRC safety requirement.[2] This way, hard particles (with higher p_T) would tend to be the cluster centers, and softer, low-energy particles would cluster around them. Note that in the initially proposed object condensation loss [17, 19], the parameter q_i is learnable. Additionally, a loss term $\mathcal{L}_{coord} = 0.1 \frac{1}{N} \sum_j \frac{x_j}{|x_j|} - \frac{x'_j}{|x'_j|}$ is added such that the output coordinates of the model are close to the original input coordinates of the particles.

We empirically find that adding approx. 500 particles with a low momentum of $p_T^{GP} = 0.01$ GeV spread uniformly in direction improve performance (without modifying the jets due to the low p_T of these particles). The models using this trick are marked with LGATr_GP. This in turn hints that the clustering step itself may be a limiting factor, and that architectures which directly learn per-particle instance assignments (such as MaskFormer-inspired instance segmentation methods) could be better suited to this task.

5.2 IRC safety loss

To ensure the infrared and collinear (IRC) safety of the learned event representations, we introduce an additional loss term that penalizes changes to the clustering space when *IRC transformations* (augmentations) are applied to the model inputs: using **collinear splitting** $T_{||}$, the 1-5 particles sampled from the 5 highest-p_T particles are split into 2-5 collinear particles each. Using **soft particle addition** T_{soft}, 500 particles with $p_T = 0.01$ GeV each are added to the event, spread randomly in different directions. The IRC safety loss attempts to minimize the mean squared error of the shift in the virtual coordinates that occurs when the event is augmented with T_{soft} or $T_{||}$. We consider two variants of the IRC safety loss: **LGATr_GP_IRC_S**, using only $T_{||}$, and **LGATr_GP_IRC_SN**, using both T_{soft} and $T_{||}$.

6 Results and discussion

We consider a jet candidate matched with a dark quark if its closest dark quark is closer than $R = 0.8$ in angular distance and no other jets have already been matched to it. We compute precision (number of matched dark quarks over the number of jets output by the model), recall (number of matched dark quarks over the number of dark quarks), and F_1 score (harmonic mean of precision and recall). Figures 1 (e) and (f) show that gains over the baseline can be achieved regardless of whether the model was trained on signal or background, demonstrating that the model is signal-agnostic. The

---

[1]See Appendix A.1 for the definition of the ground truth jets.

[2]p_T is the projection of the momentum vector onto the plane perpendicular to the beam axis.

Table 1: Comparison of models across PFCands, GL, and PL in terms of the F_1 score, evaluated on the signal dataset $m_{Z'} = 700$ GeV, $r_{inv.} = 0.7$. The models were trained on the QCD event dataset. The jets with $p_T < 100$ GeV are rejected.

Method	PFCands	GL	PL
AK8 (baseline)	0.386	0.362	0.374
LGATr	0.383	0.346	0.366
LGATr_GP	**0.400**	0.363	0.370
LGATr_GP_IRC_S	**0.400**	**0.364**	**0.375**
LGATr_GP_IRC_SN	0.395	**0.364**	0.368

results in Table 1 show that our models retain comparable performance with the baseline before and after hadronization (PL and GL), demonstrating the IRC safety of our models.

The IRC safety loss term requires further investigation. For example, it remains unclear why including both noise particles and collinear splittings in the IRC safety loss (GP_IRC_SN) degrades performance compared to only including the split particles (GP_IRC_S). The model might be simply learning an effective transverse momentum cutoff, thereby disregarding particles with momenta below that threshold. The weight of the IRC loss was set to a fixed, arbitrary number (100); to avoid this, methods such as the Modified Differential Multiplier Method [20] could be utilized. The proposed IRC loss can be easily extended to other models in high-energy physics where IRC safety is needed, such as jet tagging models.

7 Conclusion

We have presented a machine learning-based method for jet clustering using Geometric Algebra Transformers that preserves the IRC safety while outperforming the baseline anti-kt. Additionally, we have proposed a novel IRC safety loss term that ensures the consistency of the outputs of the neural network in case the hard particles are split. Our work contributes to using machine learning to develop IRC-safe jet algorithms.

8 Reproducibility

A live demo of the model is available at https://huggingface.co/spaces/gregorkrzmanc/jetclustering. The trained models are available for download at https://huggingface.co/gregorkrzmanc/jetclustering/tree/main. The used datasets of both signal and background events are available for download at https://huggingface.co/datasets/gregorkrzmanc/jetclustering and the code is available at https://github.com/gregorkrz/jetclustering.

Acknowledgements

(a) Signal process Feynman diagram

(b) Input space sample

(c) Clustering space sample

(d) Number of particles per event

(e) F_1 score on background
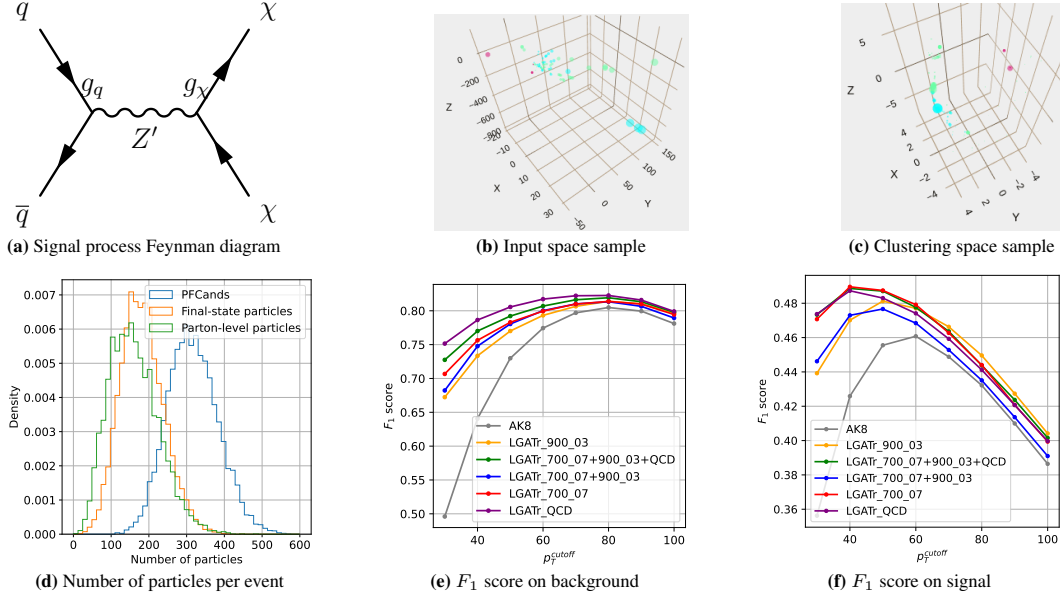

(f) F_1 score on signal

Figure 1: Figure (a) displays the Feynman diagram of the signal process. The final-state particles can be captured by the detector, as shown in Figure (b). Our model outputs a clustering space where the particles belonging to the same jet are close together, as seen in Figure (c). The signal dataset statistics for the signal hypothesis with $m_{Z'} = 900$ GeV, $r_{inv.} = 0.3$ are displayed in Figure (d). Figures (e) and (f) display how the F_1 score changes if the cutoff p_T is lowered from 100 GeV on both signal ($m_{Z'} = 700$ GeV, $r_{inv.} = 0.7$) and background for the model trained using the GP_IRC_S loss as well as the baseline AK8. Different colors in Figures (e) and (f) represent different training datasets, e.g., 900_03 means the signal dataset with $m_{Z'} = 900$ GeV, $r_{inv.} = 0.3$, and QCD means the background dataset. Our method outperforms the baseline in terms of the F_1 score at all p_T^{cutoff} values. See Appendix D for the plots of the precision and recall components of the F_1 score.

References

[1] *CMS celebrates the end of LHC Run 2 | CMS Experiment*. URL: https://cms.cern/news/end-of-LHC-Run2 (visited on 03/18/2025).

[2] A. Hayrapetyan et al. "Performance of the CMS high-level trigger during LHC Run 2". en. In: *Journal of Instrumentation* 19.11 (Nov. 2024), P11021. ISSN: 1748-0221. DOI: 10.1088/1748-0221/19/11/P11021. URL: https://iopscience.iop.org/article/10.1088/1748-0221/19/11/P11021 (visited on 03/18/2025).

[3] C. M. S. Collaboration. "Particle-flow reconstruction and global event description with the CMS detector". In: *Journal of Instrumentation* 12.10 (Oct. 2017). arXiv:1706.04965 [physics], P10003–P10003. ISSN: 1748-0221. DOI: 10.1088/1748-0221/12/10/P10003. URL: http://arxiv.org/abs/1706.04965 (visited on 05/05/2025).

[4] The CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012). arXiv:1207.7235 [hep-ex], pp. 30–61. ISSN: 03702693. DOI: 10.1016/j.physletb.2012.08.021. URL: http://arxiv.org/abs/1207.7235 (visited on 05/12/2025).

[5] The ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012). arXiv:1207.7214 [hep-ex], pp. 1–29. ISSN: 03702693. DOI: 10.1016/j.physletb.2012.08.020. URL: http://arxiv.org/abs/1207.7214 (visited on 03/18/2025).

[6] Jesse Thaler. "Separated at Birth: Jet Maximization, Axis Minimization, and Stable Cone Finding". In: *Physical Review D* 92.7 (Oct. 2015). arXiv:1506.07876 [hep-ph], p. 074001. ISSN: 1550-7998, 1550-2368. DOI: 10.1103/PhysRevD.92.074001. URL: http://arxiv.org/abs/1506.07876 (visited on 12/24/2024).

[7] Matthew J. Strassler and Kathryn M. Zurek. "Echoes of a Hidden Valley at Hadron Colliders". In: *Physics Letters B* 651.5-6 (Aug. 2007). arXiv:hep-ph/0604261, pp. 374–379. ISSN: 03702693. DOI: 10.1016/j.physletb.2007.06.055. URL: http://arxiv.org/abs/hep-ph/0604261 (visited on 04/15/2025).

[8] Timothy Cohen, Mariangela Lisanti, Hou Keong Lou, and Siddharth Mishra-Sharma. "LHC searches for dark sector showers". en. In: *Journal of High Energy Physics* 2017.11 (Nov. 2017), p. 196. ISSN: 1029-8479. DOI: 10.1007/JHEP11(2017)196. URL: https://doi.org/10.1007/JHEP11(2017)196 (visited on 04/15/2025).

[9] CMS Collaboration. "Search for resonant production of strongly coupled dark matter in proton-proton collisions at 13 TeV". In: *Journal of High Energy Physics* 2022.6 (June 2022). arXiv:2112.11125 [hep-ex], p. 156. ISSN: 1029-8479. DOI: 10.1007/JHEP06(2022)156. URL: http://arxiv.org/abs/2112.11125 (visited on 02/19/2024).

[10] The Cms Collaboration et al. "The CMS experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08004–S08004. ISSN: 1748-0221. DOI: 10.1088/1748-0221/3/08/S08004. URL: https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08004 (visited on 04/15/2025).

[11] Cesare Cazzaniga and Annapaola de Cosa. "Leptons lurking in semi-visible jets at the LHC". In: *The European Physical Journal C* 82.9 (Sept. 2022). arXiv:2206.03909 [hep-ph], p. 793. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-022-10775-2. URL: http://arxiv.org/abs/2206.03909 (visited on 04/15/2025).

[12] Christian Bierlich et al. *A comprehensive guide to the physics and usage of PYTHIA 8.3*. arXiv:2203.11601 [hep-ph]. Mar. 2022. DOI: 10.48550/arXiv.2203.11601. URL: http://arxiv.org/abs/2203.11601 (visited on 05/13/2025).

[13] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi. "DELPHES 3, A modular framework for fast simulation of a generic collider experiment". In: *Journal of High Energy Physics* 2014.2 (Feb. 2014). arXiv:1307.6346 [hep-ex], p. 57. ISSN: 1029-8479. DOI: 10.1007/JHEP02(2014)057. URL: http://arxiv.org/abs/1307.6346 (visited on 05/13/2025).

[14] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. "The anti-k_t jet clustering algorithm". en. In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063–063. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2008/04/063. URL: http://stacks.iop.org/1126-6708/2008/i=04/a=063?key=crossref.2ffef838d94c2b899ee27950c18ee8d2 (visited on 10/27/2024).

[15] Johann Brehmer, Pim de Haan, Sönke Behrends, and Taco Cohen. *Geometric Algebra Transformer*. arXiv:2305.18415 [cs]. Nov. 2023. DOI: 10.48550/arXiv.2305.18415. URL: http://arxiv.org/abs/2305.18415 (visited on 05/01/2025).

[16] Jonas Spinner, Victor Bresó, Pim de Haan, Tilman Plehn, Jesse Thaler, and Johann Brehmer. *Lorentz-Equivariant Geometric Algebra Transformers for High-Energy Physics*. arXiv:2405.14806 [physics]. Oct. 2024. DOI: 10.48550/arXiv.2405.14806. URL: http://arxiv.org/abs/2405.14806 (visited on 05/07/2025).

[17] Jan Kieseler. "Object condensation: one-stage grid-free multi-object reconstruction in physics detectors, graph and image data". In: *The European Physical Journal C* 80.9 (Sept. 2020). arXiv:2002.03605 [physics], p. 886. ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-020-08461-2. URL: http://arxiv.org/abs/2002.03605 (visited on 03/18/2025).

[18] Claudia Malzer and Marcus Baum. "A Hybrid Approach To Hierarchical Density-based Cluster Selection". In: arXiv:1911.02282 [cs]. Sept. 2020, pp. 223–228. DOI: 10.1109/MFI49285.2020.9235263. URL: http://arxiv.org/abs/1911.02282 (visited on 05/01/2025).

[19] Shah Rukh Qasim, Jan Kieseler, Yutaro Iiyama, and Maurizio Pierini. "Learning representations of irregular particle-detector geometry with distance-weighted graph networks". en. In: *The European Physical Journal C* 79.7 (July 2019), p. 608. ISSN: 1434-6044, 1434-6052. DOI: 10.1140/epjc/s10052-019-7113-9. URL: http://link.springer.com/10.1140/epjc/s10052-019-7113-9 (visited on 03/21/2025).

[20] John C. Platt and Alan H. Barr. "Constrained Differential Optimization". In: 1988. URL: https://authors.library.caltech.edu/records/05yxn-c4z11 (visited on 05/19/2025).

[21] Leland McInnes, John Healy, and Steve Astels. "hdbscan: Hierarchical density based clustering". In: *The Journal of Open Source Software* 2.11 (Mar. 2017). DOI: 10.21105/joss.00205. URL: https://doi.org/10.21105%2Fjoss.00205.

[22] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet User Manual". In: *Eur. Phys. J. C* 72 (2012). _eprint: 1111.6097, p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2.

[23] Johann Brehmer, Víctor Bresó, Pim de Haan, Tilman Plehn, Huilin Qu, Jonas Spinner, and Jesse Thaler. *A Lorentz-Equivariant Transformer for All of the LHC*. arXiv:2411.00446. Nov. 2024. URL: http://arxiv.org/abs/2411.00446 (visited on 11/13/2024).

A Dataset details

The direction and magnitude of the particle momentum are described by the azimuthal angle ϕ, pseudorapidity $\eta = -\ln\tan(\frac{\theta}{2})$ (where θ is the angle from the beam axis), and transverse momentum p_T.

The clustering models are trained and tested on multiple signal hypotheses with $r_{inv.} \in \{0.3, 0.5, 0.7\}$ and $m_{Z'} \in \{700, 800, 900, 1000, 1100, 1200\}$ GeV, with 10,000 events per signal hypothesis. Additionally, a testing dataset of 10,000 and a training dataset of 100,000 QCD events (quark and gluon jets) were simulated using the same software, with minimum parton p_T set to 100 GeV.

Only particles with $p_T > 0.5$ GeV and $|\eta| < 2.4$ are kept.

The visible parton-level generator particles (PL) are the particles with Pythia status between 51 and 59 that belong to the SM (i.e., no dark matter).

A.1 Ground truth construction

We set the ground truth for the clustering in the following way: The particles within $R_{GT} = 0.8$ angular distance away from the dark quarks are set to belong in the same cluster. Particles further away than R_{GT} from the closest dark quark are treated as noise.

B Model hyperparameters

The LGATr models have 10 blocks, 16 multivector channels, and 64 scalar channels, and in the end have ~ 1.3 million learnable parameters.

The HDBSCAN clustering parameters are `min_cluster_size=2`, `min_samples=1`, `epsilon=0.3`. The HDBSCAN Python library [21] is used to perform clustering in virtual space, and FastJet [22] is used to perform anti-kt clustering.

B.1 Lorentz symmetry breaking in the model

The measurement process at the LHC is not Lorentz-invariant, mainly due to the fixed direction of the beam axis and varying reconstruction efficiencies for particles hitting the detector at different angles. Therefore, it may be beneficial to break the Lorentz symmetry. As proposed in [23], we break the Lorentz symmetry by additionally encoding the direction of the beam axis in the inputs.

We additionally artificially break the Lorentz symmetry by encoding the particle energy E and transverse momentum p_T as additional scalar channels. Note that this information is already encoded in the four-vector passed to the model; this way, we provide the model with another option to break Lorentz symmetry. Note that the loss function is also not Lorentz-invariant due to the $q_i = p_{T,i}$ setting as well as the computation of distances in clustering space only from the spatial part of the four-vectors. We leave the construction of a fully Lorentz-invariant loss function for future work.

C Training

The models were trained and evaluated on a single NVIDIA GeForce GTX 1080 Ti GPU. The initial training of the model takes around 1.5 days, and further training using the complete IRC safety loss takes around a day with a batch size of 20 events. Evaluating 10,000 events takes around 10 minutes. These runtimes could likely be reduced through further code optimization and more efficient implementation.

Figures 2 and 3 display how the contributions of different loss terms to the total loss evolve with training.

D Precision, recall, and F_1 score plots

Figures 4 and 5 display the precision, recall, and F_1 scores for signal and background depending on the p_T^{cutoff} and the training dataset.
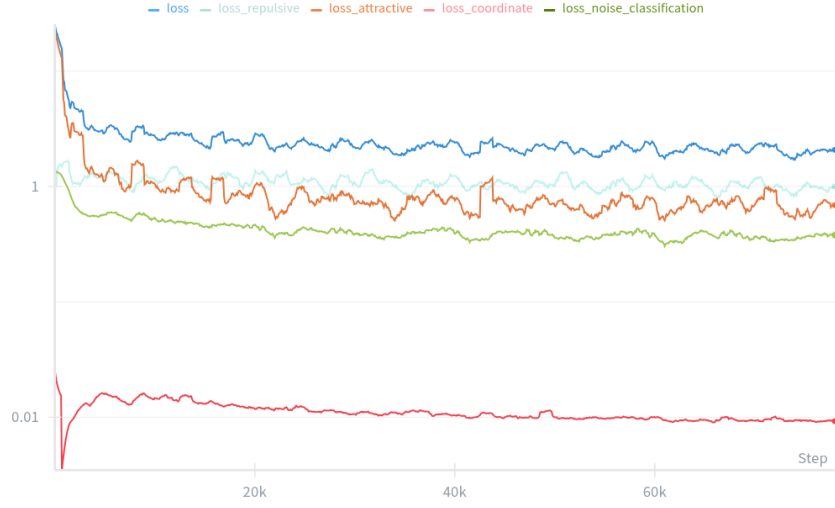
Figure 2: Display of the evolution of different components of the training loss with the training step for the initial training of the LGATr model (without the IRC safety loss). The data are smoothed using a rolling average with a window size of 16.
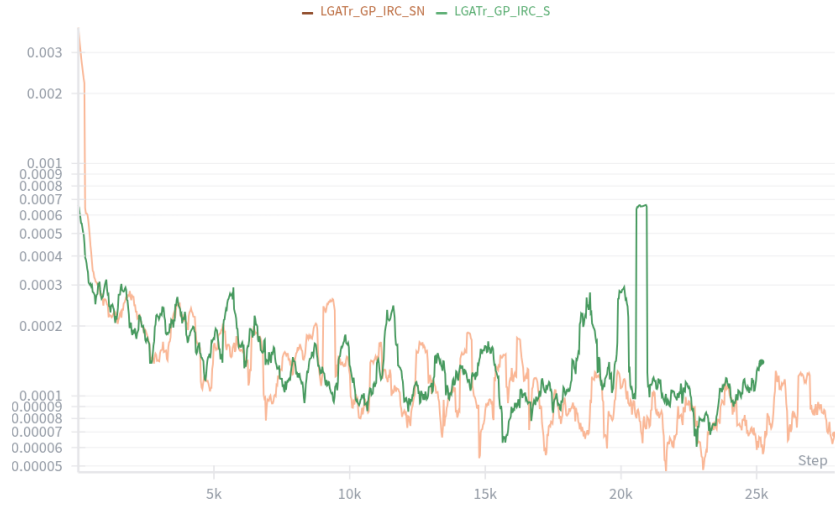

Figure 3: IRC safety loss term value vs. step, for two versions of the IRC safety loss. The IRC loss is weighted 100 times more than the other loss terms during optimization. The data are smoothed using a rolling average with a window size of 16.
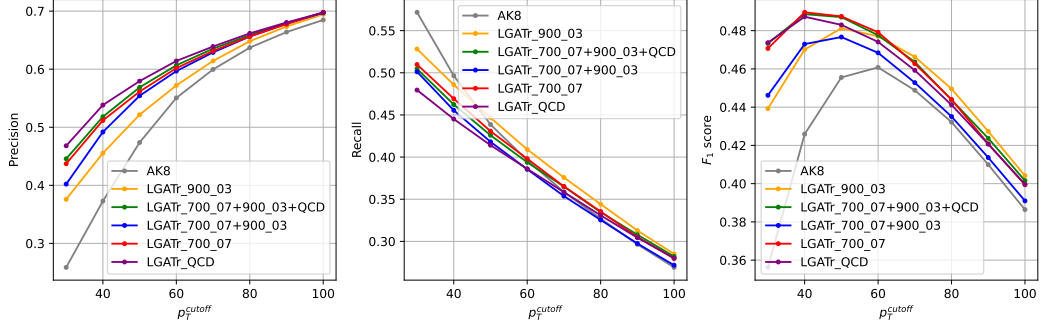
Figure 4: Precision, recall, and F_1 scores for the signal dataset ($m_{Z'} = 700$ GeV, $r_{inv.} = 0.7$), depending on the jet threshold p_T.
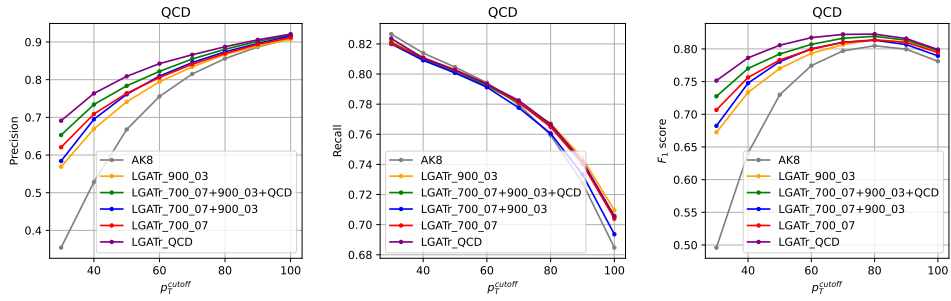

Figure 5: Precision, recall, and F_1 scores for the QCD background dataset, depending on the jet threshold p_T.