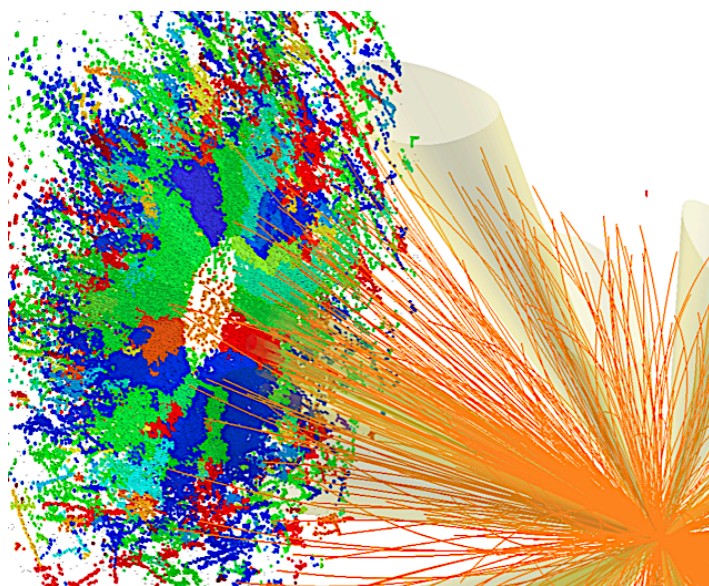


Unpacking Particle Showers with Machine Learning

A new AI-powered event-processing algorithm is being readied to cope with the torrent of data that will stream from the Large Hadron Collider upgrade.

By **Samuel Jarman**

With a novel machine-learning approach, a team of researchers in the US and the UK has shown how densely packed, deeply intertwined showers of particles can be distinguished within the deluge of collisions created by the Large Hadron Collider (LHC). The approach could provide a valuable analytical tool kit for the High-Luminosity LHC (HL-LHC), an upgrade to the collider, currently scheduled to launch in 2030 [1].



A simulation of the collision products that will reach the High-Granularity Calorimeter.

Credit: Adapted from P. Baesso *et al.*, “The High-Luminosity Upgrade of the CMS Detector”

The LHC’s parent organization, CERN, began planning to upgrade the collider’s luminosity even before the LHC discovered the Higgs boson in 2012. By boosting the density of particle bunches within the beam, the HL-LHC will spawn around 40 million collisions per second, including 4 to 5 times more simultaneous particle collisions than the current LHC can produce.

Among the instruments designed to handle this higher collision rate is the High-Granularity Calorimeter (HGCAL), which will be installed in the LHC’s CMS detector and will begin operating starting in 2030 at the earliest. With around 6 million readout channels arranged in a honeycomb pattern, HGCAL will detect showers of the secondary particles produced as newly created particles decay almost immediately after each collision.

At the level of resolution HGCAL provides, CERN’s researchers hope that each of these secondary particles could be traced back to the unstable particles that begat them. In this way, HGCAL could detect over its lifetime a steady stream of particles including top quarks and W and Z bosons, along with some 380 million Higgs bosons, compared with around 55 million detected so far. The anticipated bounty could enable the most rigorous tests to date of the standard model of particle physics.

However, these improvements also present a new set of challenges. “HGCAL’s high resolution, or granularity, turns shower reconstruction into a complex, high-dimensional point-cloud problem,” explains PhD student Max Marriott-Clarke from Imperial College London. “Showers overlap heavily in space and energy and untangling them is a serious algorithmic challenge.”

This kind of deeply complex task is routinely handled by machine learning. At the LHC, researchers have been using an approach called object condensation, which reconstructs particle showers by learning the identifying features of each shower. Detector hits can then be grouped together into individual showers. But the showers produced by HL-LHC will be so densely packed that object condensation will struggle.

To address the problem, Marriott-Clarke and his collaborators flipped the process around by using contrastive machine learning, which instead learns by drawing out the contrast between similar and dissimilar data points. In the context of HGCal, the algorithm learns which detector hits are likely to belong to the same shower, pulling similar hits closer together, while pushing dissimilar ones farther apart. “It decouples learning what a shower looks like from deciding how to group hits,” Marriott-Clarke says. As a result, the clustering is more flexible and more stable.

To test the approach, the researchers started by generating simulations of overlapping particle showers, closely resembling the data expected in HGCal. Without showing these simulations to their algorithm beforehand, they then used contrastive machine learning to sort the detected particles according to the showers they originated from. This enabled them to reconstruct each shower individually—tracking each of their particles to single proton–proton collisions.

When compared with object condensation, the new method consistently outperformed the existing approach at separating overlapping showers, reconstructing them more accurately and with more precise energy determinations.

This advantage increased in the most crowded overlapping

collision events, closely matching the conditions expected at the HL-LHC. The contrastive method also remained reliable when trained on different types of particles, successfully reconstructing both electromagnetic showers, produced by electrons, and hadronic showers, produced by charged pions, as well as a mixture of the two.

Building on these results, the team will now aim to reproduce this performance in CMS simulations of the particle showers that HGCal is expected to capture. If the method continues to perform well under these complex conditions, the researchers hope it could be applied more generally to other highly granular detectors at the LHC, with some further adjustments to account for the unique architectures of each detector.

“The matching of individual particles to detector signatures is a central challenge in collider physics, where the properties of nature are inferred from detector readouts,” comments Vincius Mikuni of Nagoya University in Japan, whose group is developing algorithms in the intersection between AI and physics. “The authors show that their method is more robust than current strategies and can generalize to particle energies and multiplicities not seen during training, making it a promising candidate for particle reconstruction at future experimental facilities.”

Samuel Jarman is a science writer based in the UK.

REFERENCES

1. M. Marriott-Clarke *et al.*, “Contrastive metric learning for point cloud segmentation in highly granular detectors,” *PRX Intell.* **1**, 013004 (2026).