Application of Distance-Weighted Graph Networks to Real-life Particle Detector Output

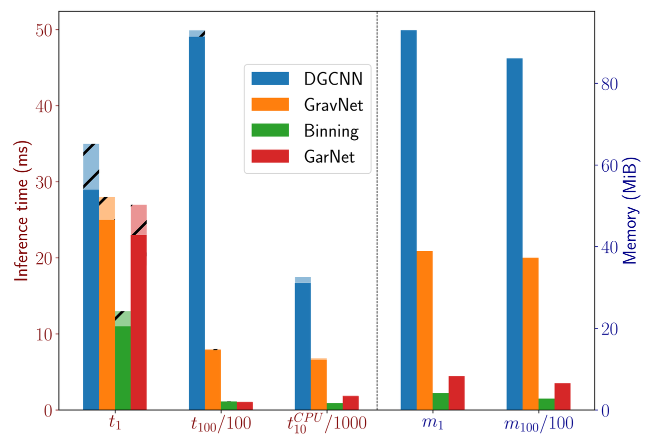
Yutaro Iiyama (CERN) on behalf of the CMS Collaboration

Machine Learning and the Physical Sciences, NeurIPS 2019

Distance-weighted graph-based neural networks take point-wise structured input (vertices) and translate the semantic affinity between the vertices into geometric proximity in high-dimensional latent spaces. Two distance-weighted graph network architectures, GravNet and GarNet, are applied to the simulated detector readout of the future endcap calorimeter of the CMS experiment.

GravNet and GarNet

- GravNet and GarNet: Distance-weighted graph neural network architectures introduced in Ref [1]
- Self-contained and stackable in a larger network
- Input $B \times V \times F_{in}$ → output $B \times V \times F_{out}$
  B = batch size, V = (maximum) number of vertices, F_{in}, F_{out} = number of input and output vertex features
- GravNet involves a $V \times V$ adjacency matrix
  → Memory-intensive, but provides higher accuracy
- GarNet is light on compute resource and fast



GravNet for accuracy-critical tasks, GarNet for latency- and resource-critical tasks.

Layers step-by-step

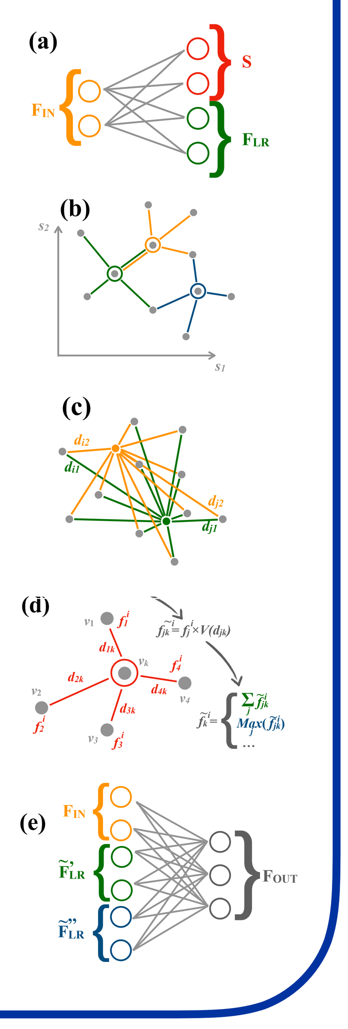
(a) F_{in} features are converted into S "spatial" and F_{LR} "intrinsic" attributes of the vertices via dense layers.

(b) GravNet: S features are interpreted as Cartesian coordinates in an S-dimensional latent space. Each vertex forms an edge with N nearest neighbor vertices by the Euclidean distance to form a graph.

(c) GarNet: S features are interpreted as one-dimensional distance to abstract "aggregator" nodes. All vertices are connected to all aggregators.

(d) Each vertex (GravNet) or aggregator (GarNet) collects the max and mean of the features of the connected vertices. Features are transformed by a common dense layer, weighted by a decreasing function of the distance. In GarNet, aggregators return the collected features back to the vertices with the same weight function. Collected features are appended to the intrinsic features of the vertices.

(e) Concatenation of the input and internal intrinsic features are passed through a dense layer to produce the F_{out} output features for each vertex.



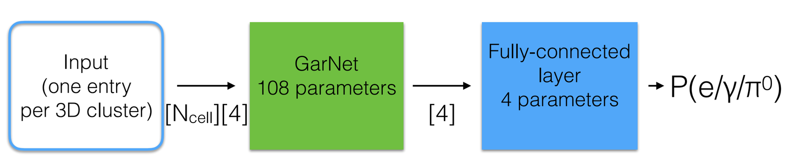
GarNet on FPGA for Triggering

- GarNet is implemented as FPGA firmware (Verilog / VHDL) using the HLS4ML framework [3] (Xilinx Vivado backend)
- Proof-of-concept: simple architecture applied on a simple problem (e/γ vs hadron classification of pre-made L1T clusters)
- Sub-μs latency per layer achieved with more room for optimization
- Classification performance comparable to Keras implementation run on GPU

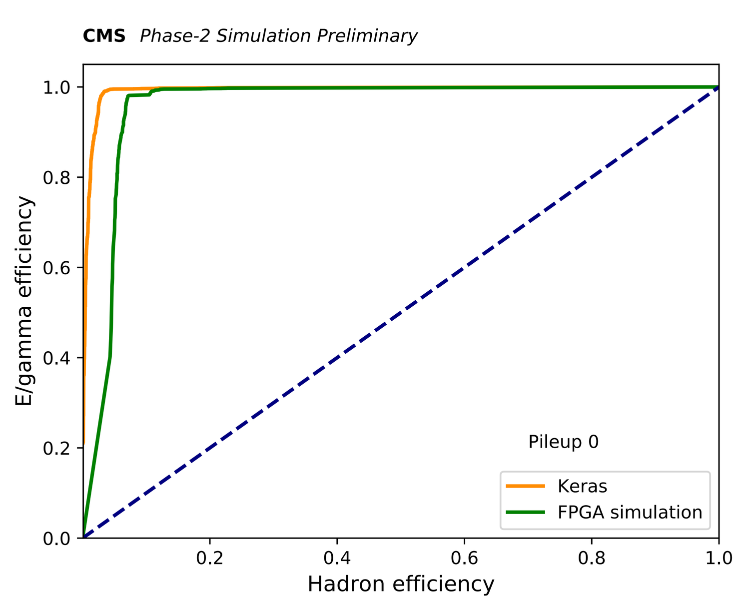
hls 4 ml

Synthesis report
On Xilinx Kintex UltraScale XCKU115
Clock frequency 200 MHz

Latency	55-100 clocks
Interval:	30-93 clocks
DSP	2.4k / 5.5k (44%)
LUT	87.5k / 663k (13%)
Block RAM	2.3MB / 75.9MB (3%)



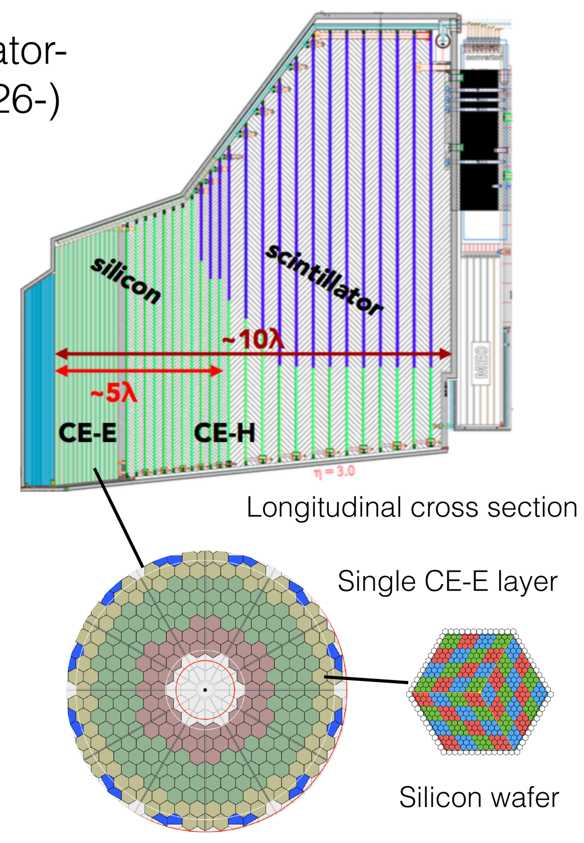
Classification model used for the test

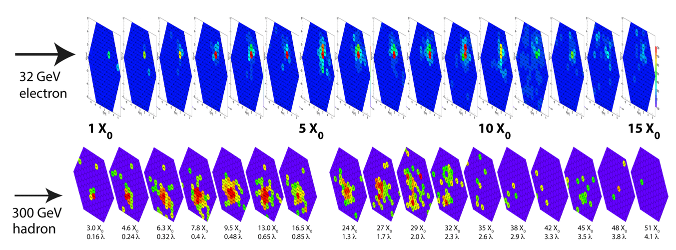


ROC curves for the e/γ versus hadron classification of HGCAL L1T clusters. Constituent trigger cells of the clusters in single-photon and single-pion events with 0 pileup are used as inputs to the model. Fake rate is higher in the HLS implementation due to the lower numerical precision used in the calculation to keep the FPGA resource usage to a manageable level.

High-Granularity Calorimeter

- CMS experiment plans to install a novel silicon + scintillator-based endcap calorimeter (HGCAL) [2] for HL-LHC (2026-)
- Pseudorapidity coverage 1.5 < |η| < 3.0
- 52 layers corresponding to max. ~10 interaction lengths
  CE-E thickness ~25 radiation lengths
- 6 million hexagonal Si sensors
  600 m² total sensor area, 0.5-1 cm² per sensor
- Occupancy maximum ~60% at 200 pileup
- Hexagonal (irregular) arrangement of sensors and hit sparseness motivate graph neural network approach for data processing
- Reconstruction goals
  - Form clusters of energy depositions from individual particles (O(100) per endcap @ PU200)
  - Accurate energy assignment to clusters
  - Particle identification
- HGCAL participates in level-1 trigger (L1T)
  - Need to perform rough clustering, energy estimate, and particle ID within 5 μs
  - 4 or 9 silicon pads are grouped into one trigger cell

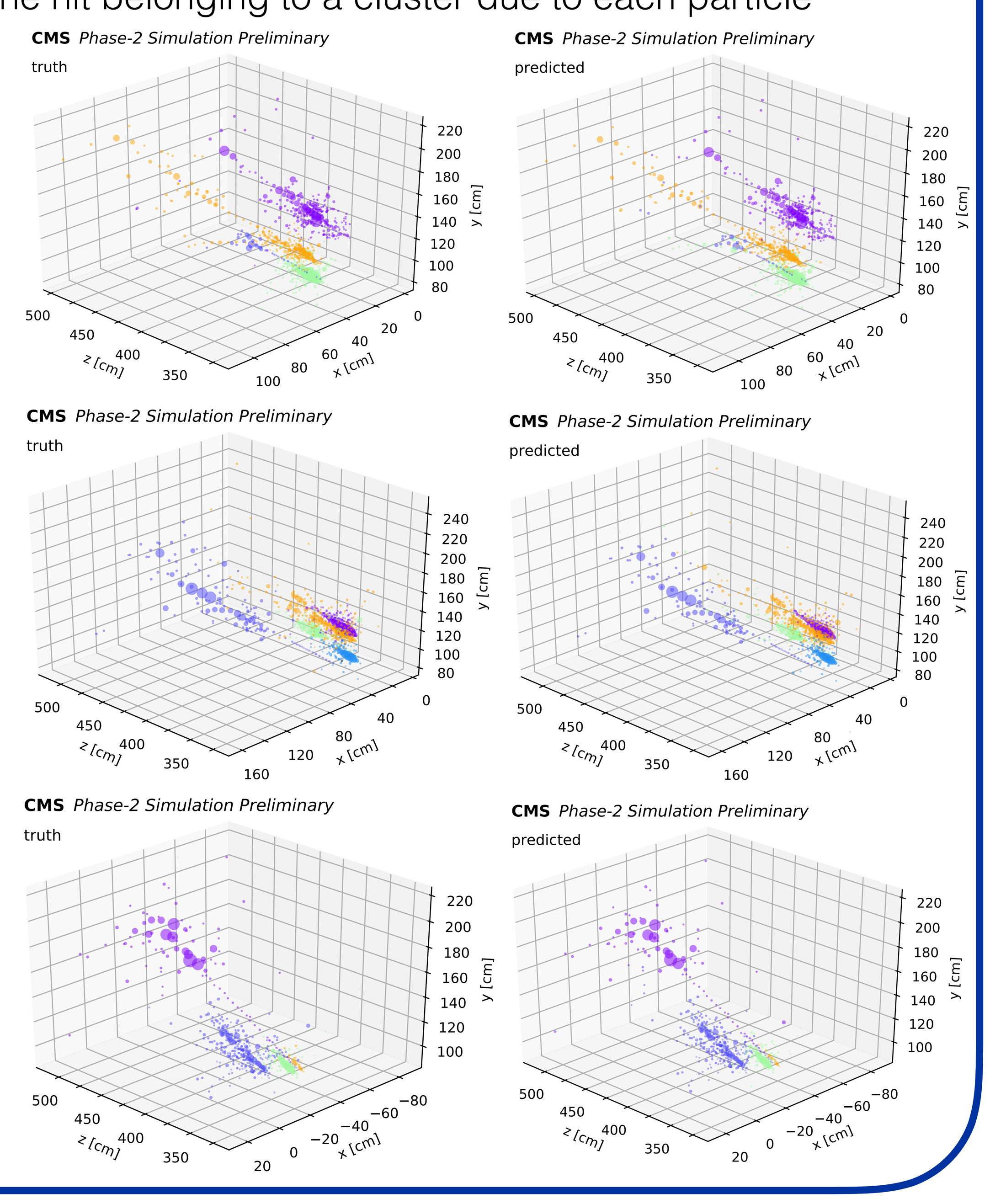




Event displays from 16- and 17-layer beam tests

Clustering with GravNet

- Using a model with 3 GravNet layers interleaved with additional message-passing layers and fully-connected layers (166k parameters)
- Input data: simulation of 5 particles shot within a narrow (ΔR < 0.5) cone
  - Particle species: randomly chosen from μ±, γ, e, and π±
  - Particle energies: randomly chosen from 10-100 GeV
  - Hits are pre-clustered within each HGCAL layer before fed into the network
- Output: per hit, probability of the hit belonging to a cluster due to each particle
- High accuracy achieved even for highly nontrivial cluster shapes
- Algorithm for predicting arbitrary number of clusters currently under development
- Clustering under non-zero pileup & over full HGCAL also under development



Displays of 3 example events. Points represent the HGCAL hits, with size of the points proportional to the recorded energy at the sensor. Colors correspond to the source particle. The left plot of each row represents the ground truth, and the right plot shows the model prediction where hits are colored by the highest-probability cluster. The network architecture is based on GravNet, with some modifications with respect to the original implementation.

References

[1] 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*, vol. 79, no. 7, p. 608, 2019.
[2] CMS Collaboration, “The Phase-2 Upgrade of the CMS Endcap Calorimeter,” Tech. Rep. CERN- LHCC-2017-023. CMS-TDR-019, CERN, Geneva, Nov 2017.
[3] J. Duarte et al., “Fast inference of deep neural networks in FPGAs for particle physics,” *JINST*, vol. 13, no. 07, p. P07027, 2018.

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement no772369).