

PicoCal ML reconstruction

Aniol Lobo¹, Jean-François Marchand¹

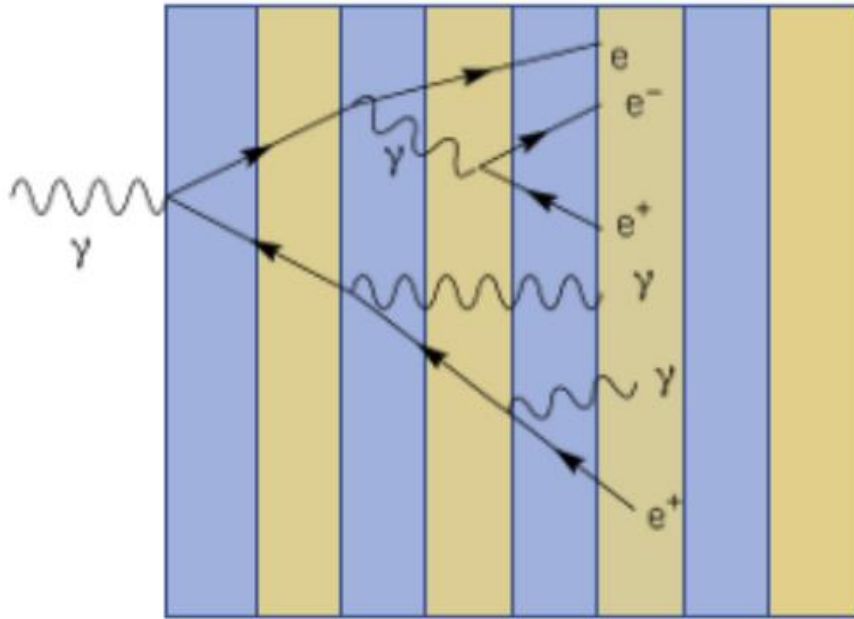
¹LAPP

Club data September 2026



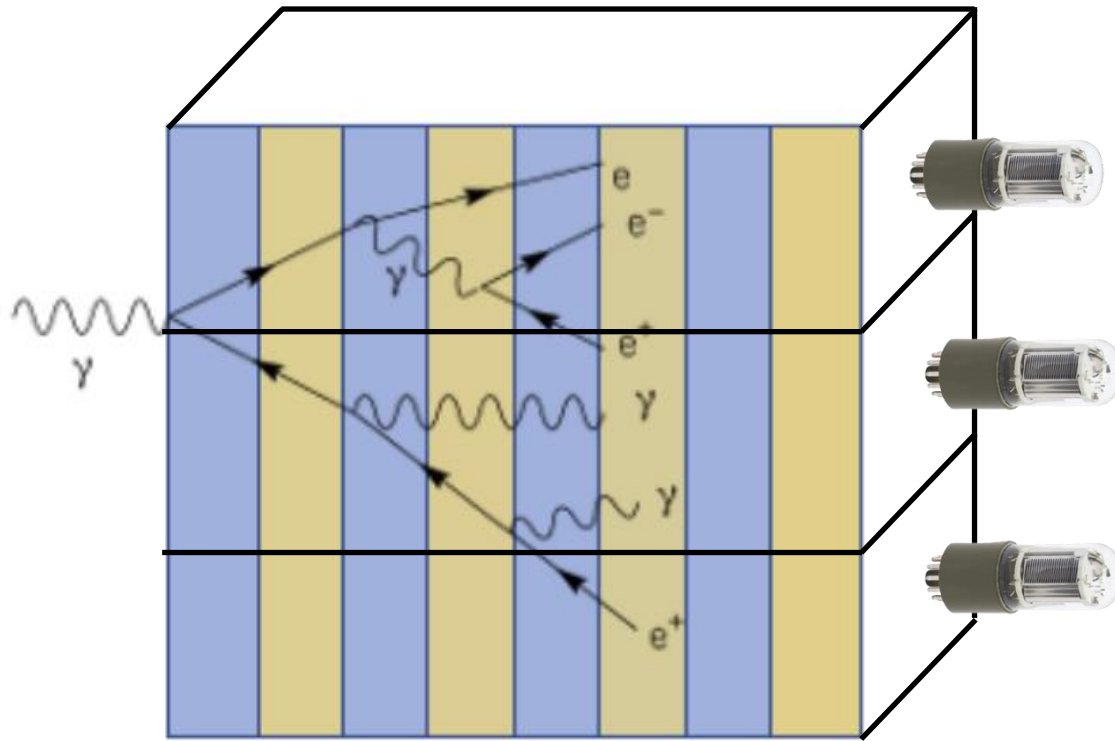
Calorimeters

- Calorimeters measure the energy and position of incoming particles



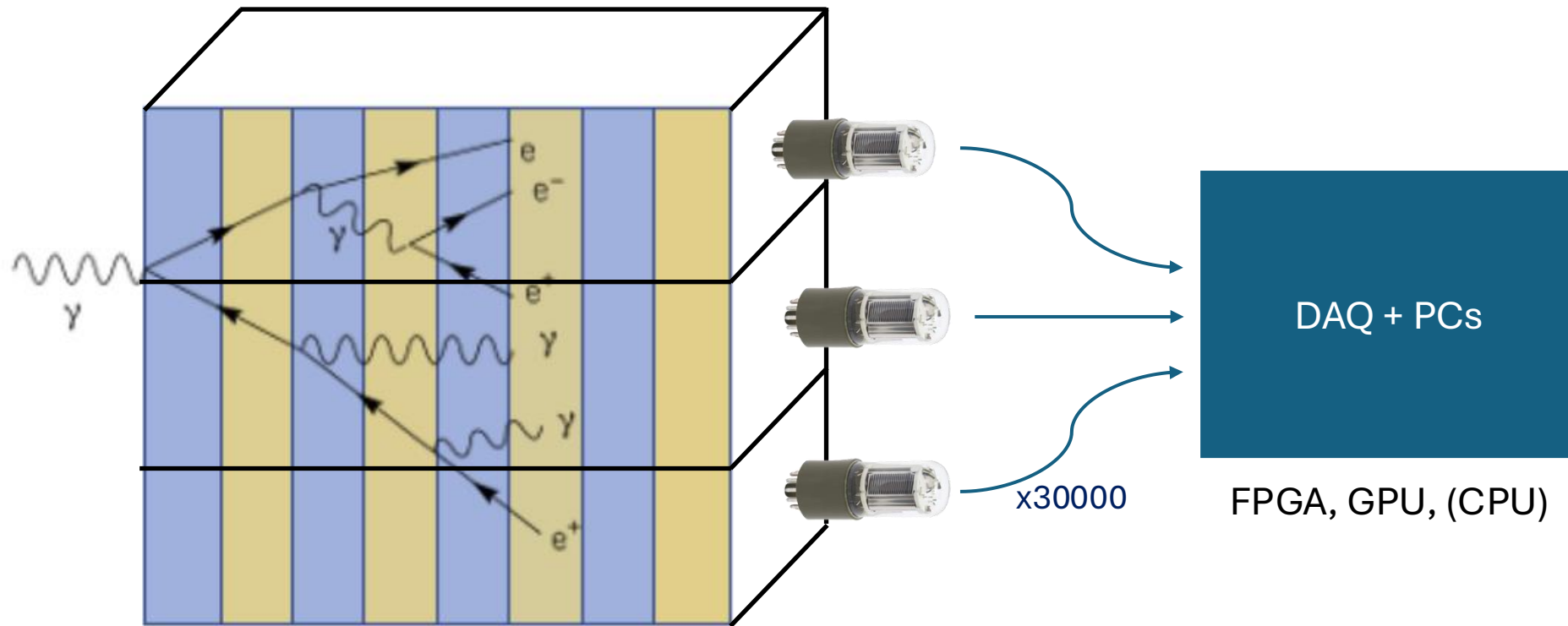
Calorimeters

- Calorimeters measure the energy and position of incoming particles



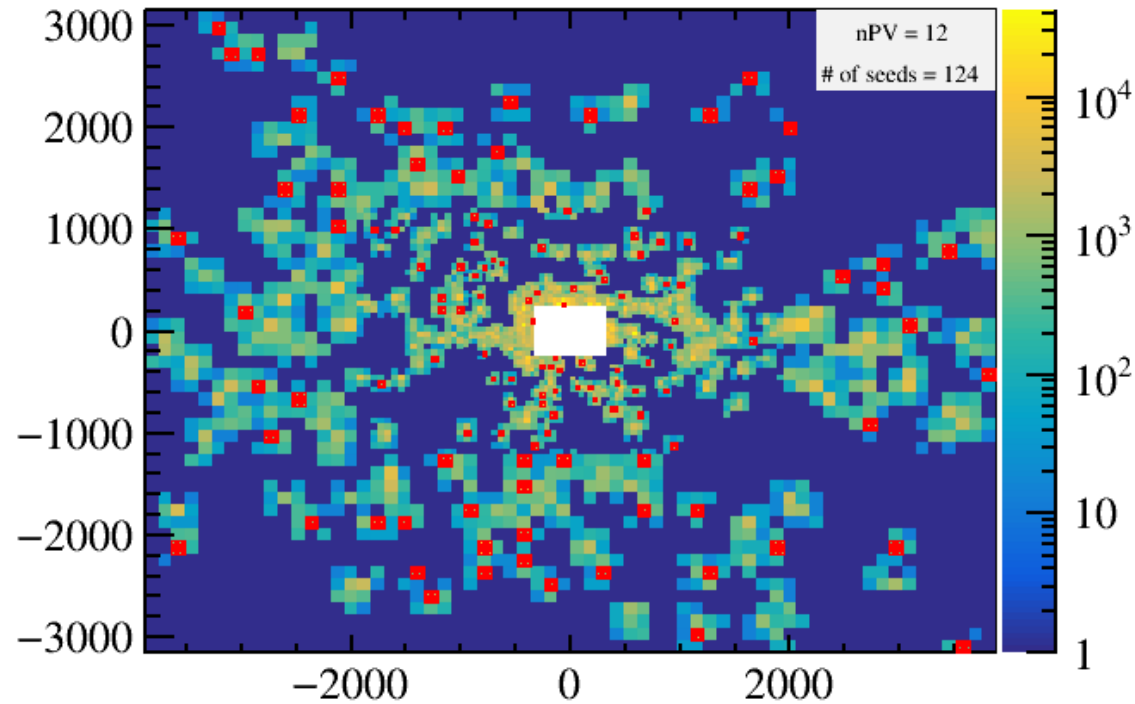
Calorimeters

- Calorimeters measure the energy and position of incoming particles



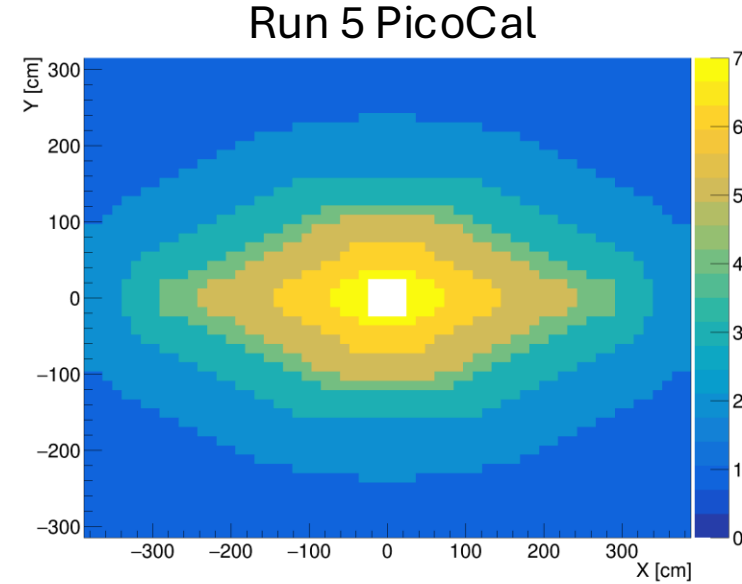
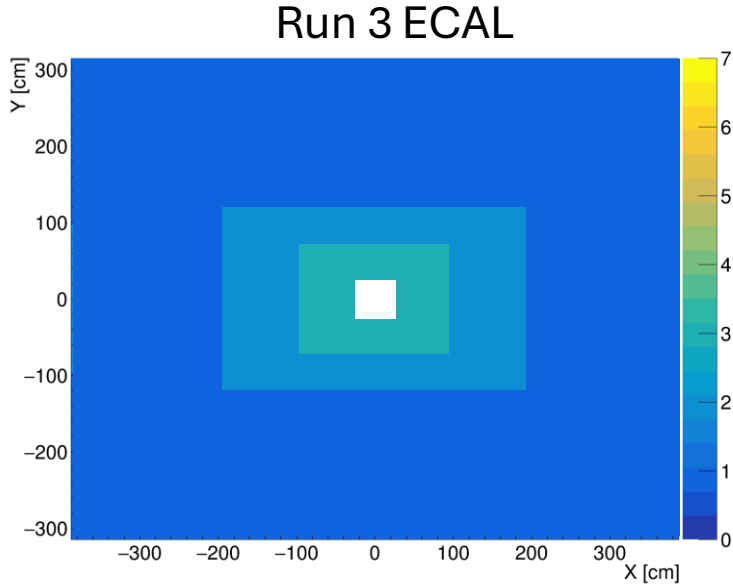
Calorimeters

- You can imagine the data as a 2D picture where each pixel is the energy recorded



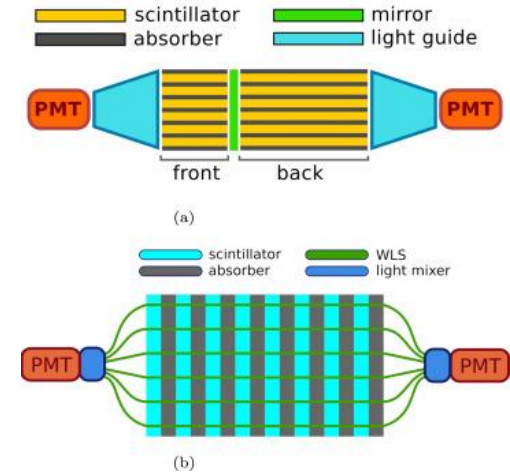
PicoCal

- PicoCal is the LHCb Upgrade II calorimeter
 - Better granularity.
 - Timing.
 - Double sided read-out.

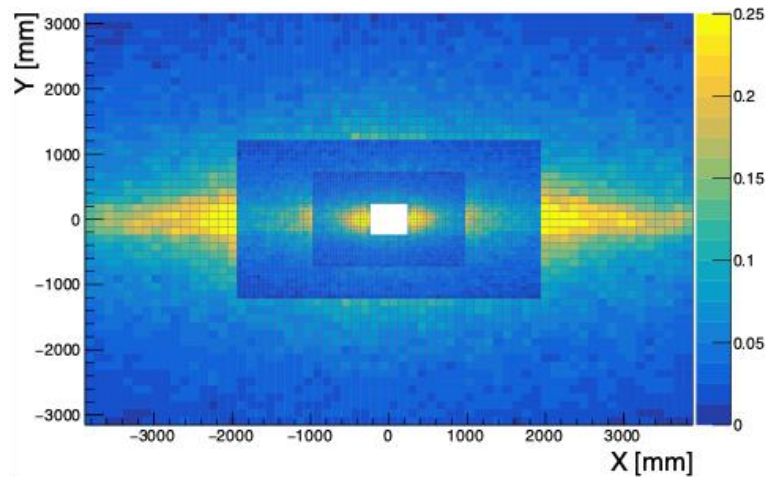


PicoCal

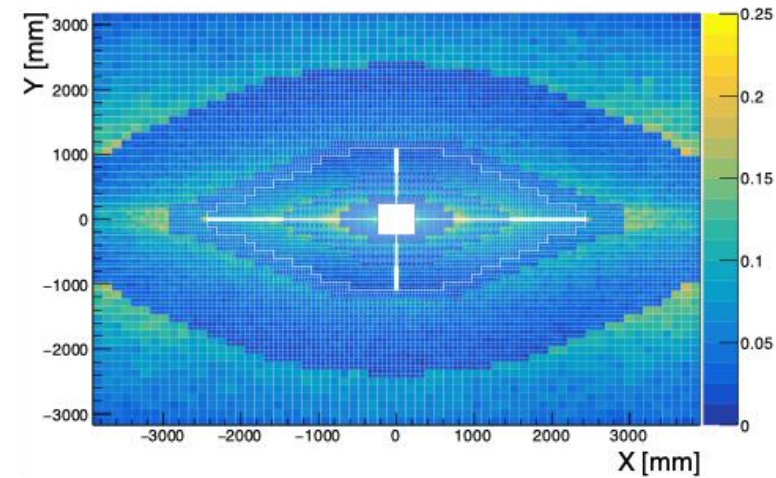
- PicoCal is the LHCb Upgrade II calorimeter
 - Better granularity.
 - Timing.
 - Double sided read-out.



Run 3 ECAL



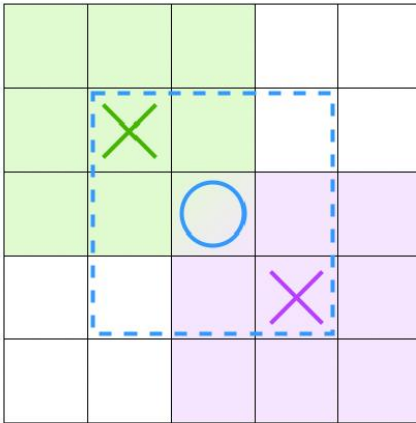
Run 5 PicoCal



Calo reconstruction

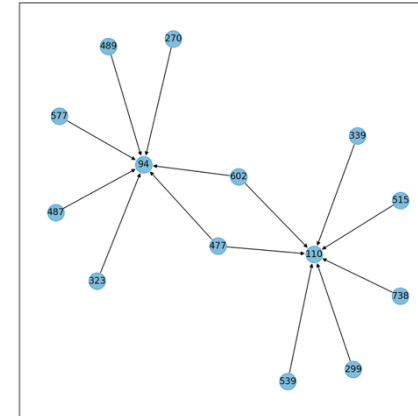
- Calorimeter reconstruction in LHCb has been evolving along years.
- Simple in HLT1, a bit more complex in HLT2.

HLT1



3x3 + simple overlap correction

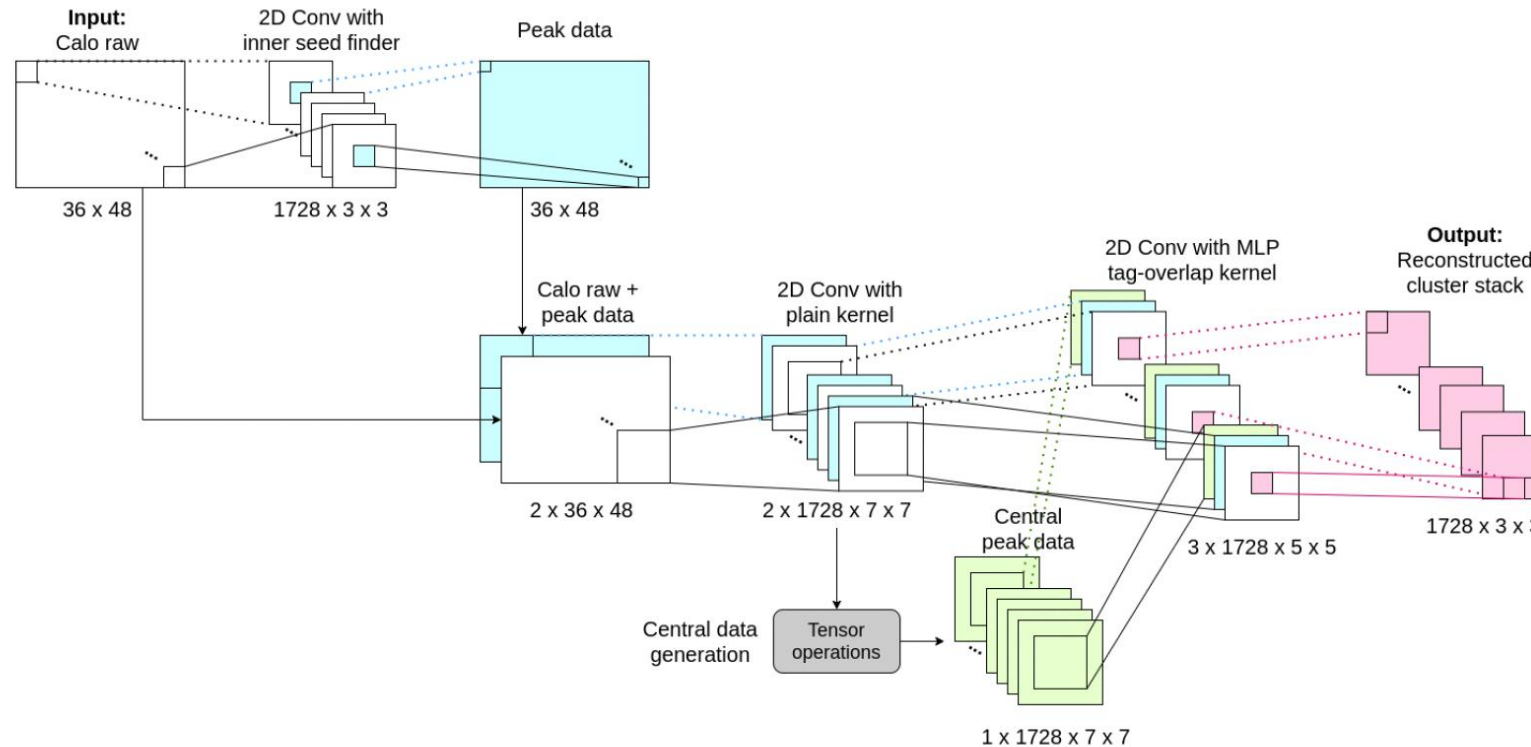
HLT2



Graph clustering

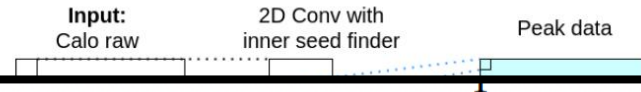
Calo ML reconstruction

- First LHCb calo ML explorations described in Núria's thesis



Calo ML reconstruction

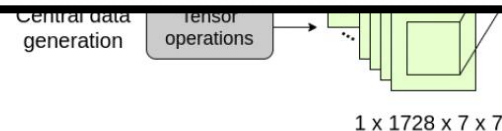
- First LHCb calo ML explorations described in Núria's thesis



time obtained is almost constant with respect to the event's complexity. Although the very interesting results, further developments of this approach were stopped due to the lack of an efficient inference engine for neural networks inside the LHCb framework. Further discussion



Therefore it is important to highlight that deep learning models will only be suited for HEP trigger-like applications if they are small, simplified and well optimized. This can only be achieved when the insights of the problem are well understood and a network is modeled

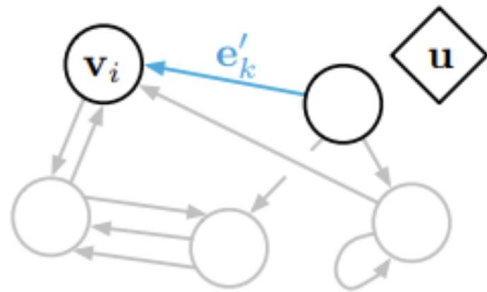


Calo ML reconstruction: GNN



- Building on the work done by our colleagues.

BLUE updated by BLACK (not utilizing GREY)



(a) Edge update

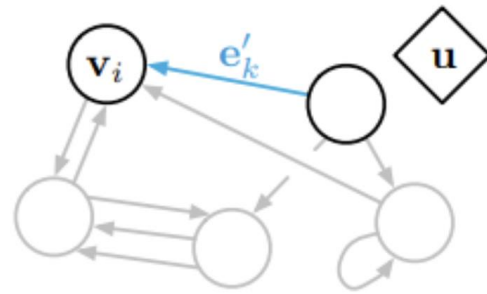
Update edges using
connected nodes and globals

Calo ML reconstruction: GNN



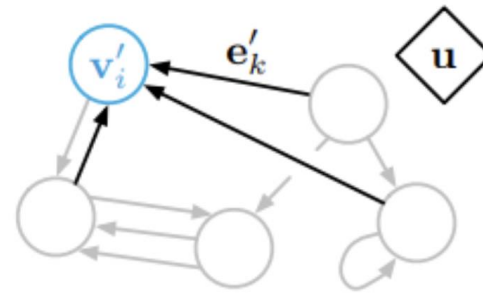
- Building on the work done by our colleagues.

BLUE updated by BLACK (not utilizing GREY)



(a) Edge update

Update edges using
connected nodes and globals



(b) Node update

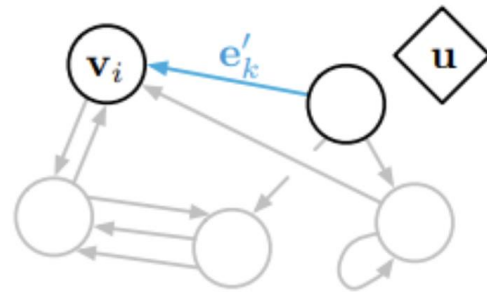
Update nodes
using (new) edges
and globals

Calo ML reconstruction: GNN



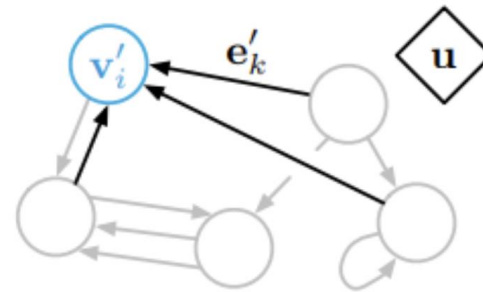
- Building on the work done by our colleagues.

BLUE updated by BLACK (not utilizing GREY)



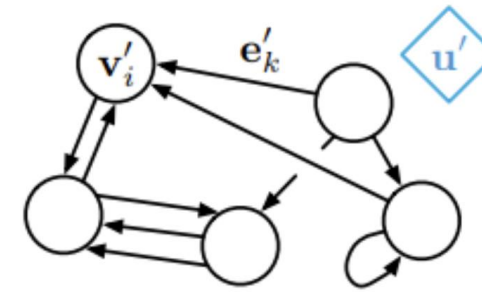
(a) Edge update

Update edges using
connected nodes and globals



(b) Node update

Update nodes
using (new) edges
and globals

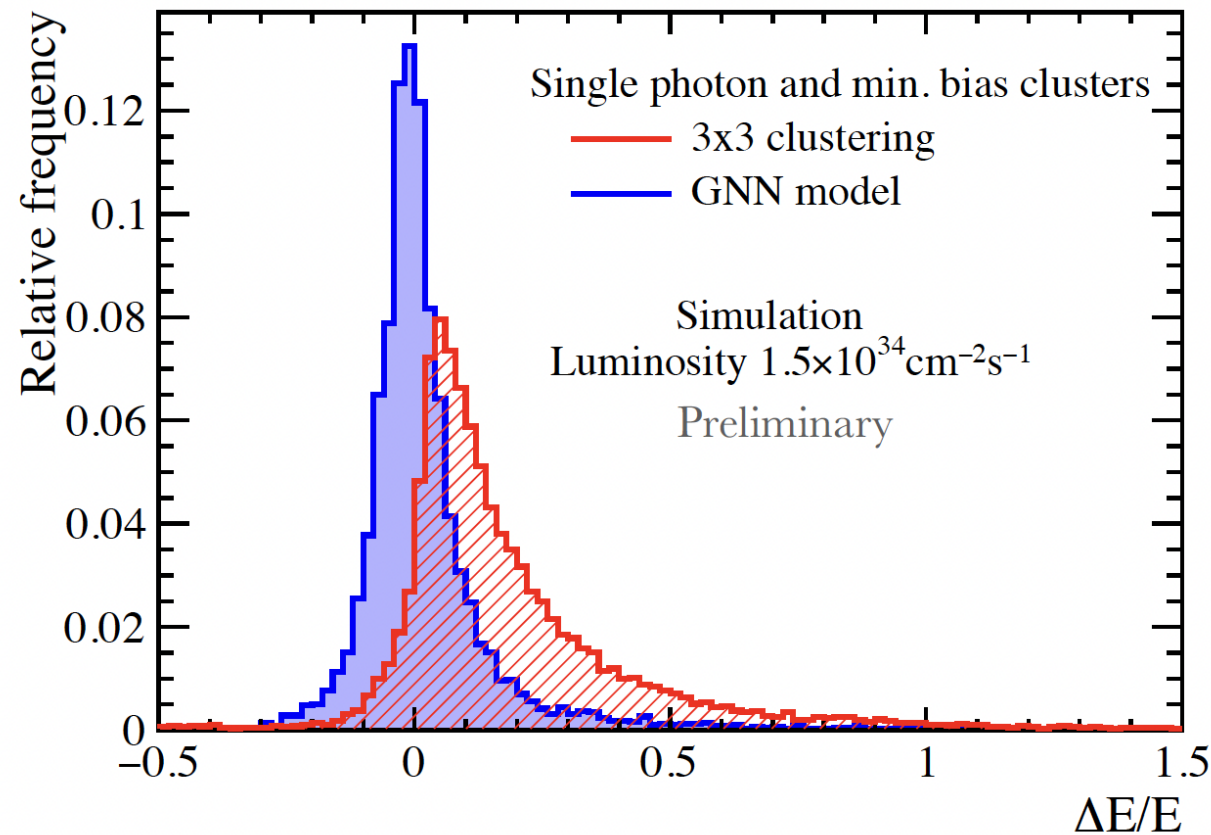


(c) Global update

Update globals
using (new) nodes
and (new) edges

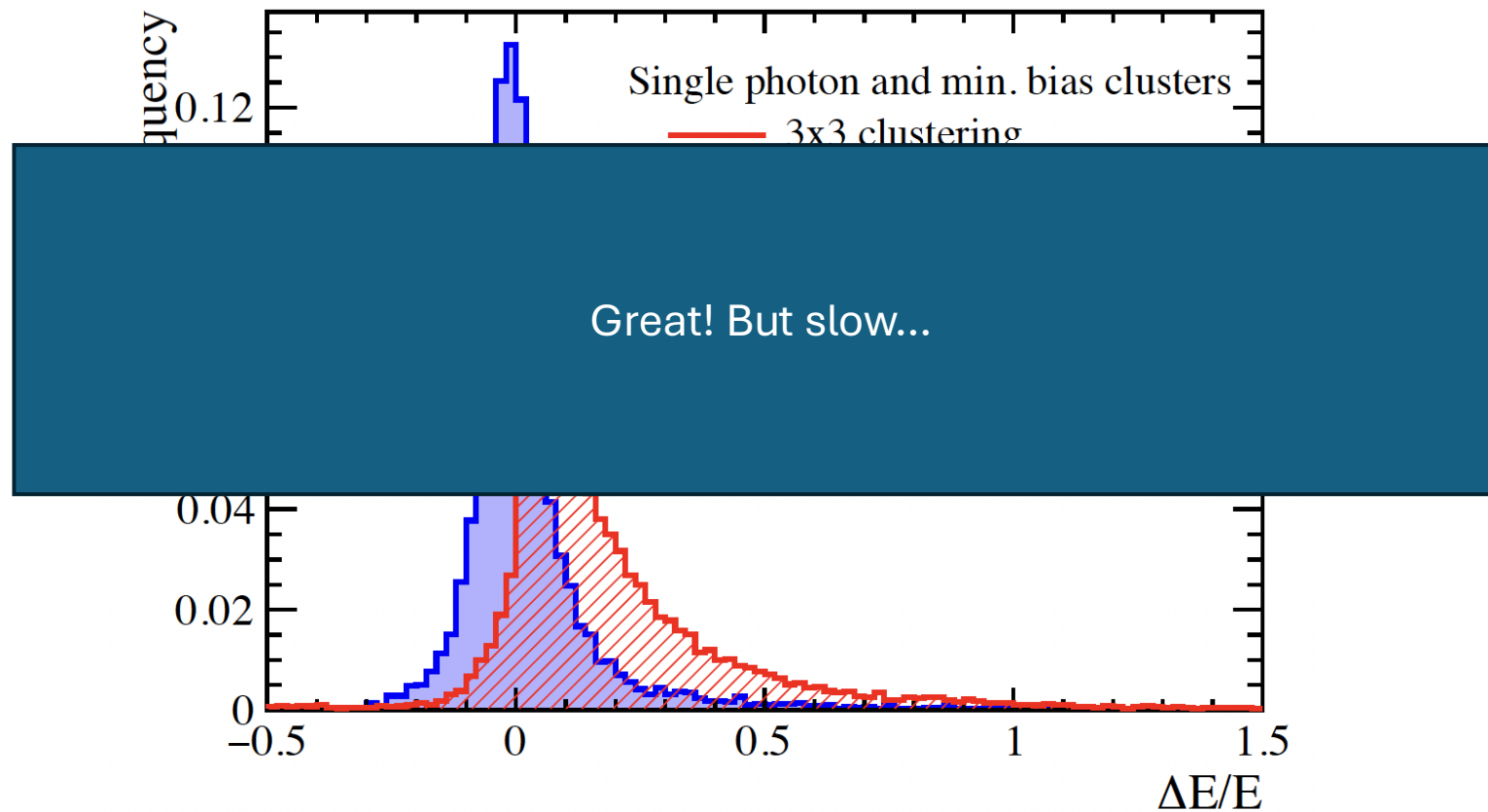
Calo ML reconstruction: GNN

- Building on the work done by our colleagues.



Calo ML reconstruction: GNN

- Building on the work done by our colleagues.



Calo ML reconstruction: GNN

- Inspiration

Learning representations of irregular particle-detector geometry with distance-weighted graph networks

Shah Rukh Qasim^{1,2} (14beesqasim@seecs.edu.pk), Jan Kieseler¹ (jan.kieseler@cern.ch), Yutaro Iiyama¹ (yutaro.iiyama@cern.ch), and Maurizio Pierini¹ (maurizio.pierini@cern.ch)

<https://arxiv.org/abs/1902.07987>

DISTANCE-WEIGHTED GRAPH NEURAL NETWORKS ON FPGAs FOR REAL-TIME PARTICLE RECONSTRUCTION IN HIGH ENERGY PHYSICS

<https://arxiv.org/abs/2008.03601>

Calo ML reconstruction: GNN

- Inspiration

GRAPH ATTENTION NETWORKS

Petar Veličković*

Department of Computer Science and Technology
University of Cambridge
petar.velickovic@cst.cam.ac.uk

Guillem Cucurull*

Centre de Visió per Computador, UAB
gcucurull@gmail.com

Arantxa Casanova*

Centre de Visió per Computador, UAB
ar.casanova.8@gmail.com

Adriana Romero

Montréal Institute for Learning Algorithms
adriana.romero.soriano@umontreal.ca

Pietro Liò

Department of Computer Science and Technology
University of Cambridge
pietro.lio@cst.cam.ac.uk

Yoshua Bengio

Montréal Institute for Learning Algorithms
yoshua.umontreal@gmail.com

ABSTRACT

We present graph attention networks (GATs), novel neural network architectures that operate on graph-structured data, leveraging masked self-attentional layers to address the shortcomings of prior methods based on graph convolutions or their approximations. By stacking layers in which nodes are able to attend over their neighborhoods' features, we enable (implicitly) specifying different weights to different nodes in a neighborhood, without requiring any kind of costly matrix operation (such as inversion) or depending on knowing the graph structure upfront. In this way, we address several key challenges of spectral-based graph neural networks simultaneously, and make our model readily applicable to inductive as well as transductive problems. Our GAT models have achieved or matched state-of-the-art results across four established transductive and inductive graph benchmarks: the *Cora*, *Citeseer* and *Pubmed* citation network datasets, as well as a *protein-protein interaction* dataset (wherein test graphs remain unseen during training).

<https://arxiv.org/pdf/1710.10903>

Attention Is All You Need

Ashish Vaswani*

Google Brain
avaswani@google.com

Noam Shazeer*

Google Brain
noam@google.com

Niki Parmar*

Google Research
nikip@google.com

Jakob Uszkoreit*

Google Research
usz@google.com

Llion Jones*

Google Research
llion@google.com

Aidan N. Gomez*[†]

University of Toronto
aidan@cs.toronto.edu

Łukasz Kaiser*

Google Brain
lukaszkaizer@google.com

Illia Polosukhin*[‡]

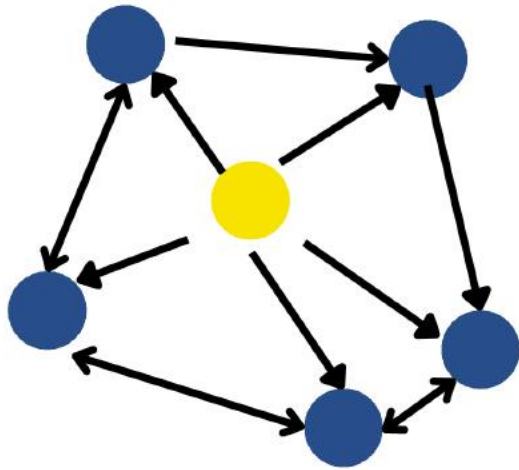
illia.polosukhin@gmail.com

<https://arxiv.org/pdf/1706.03762>

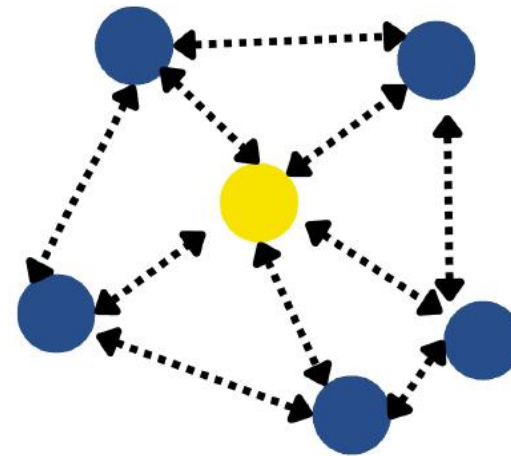
Calo ML reconstruction: GNN

- La Salle came up with new ideas

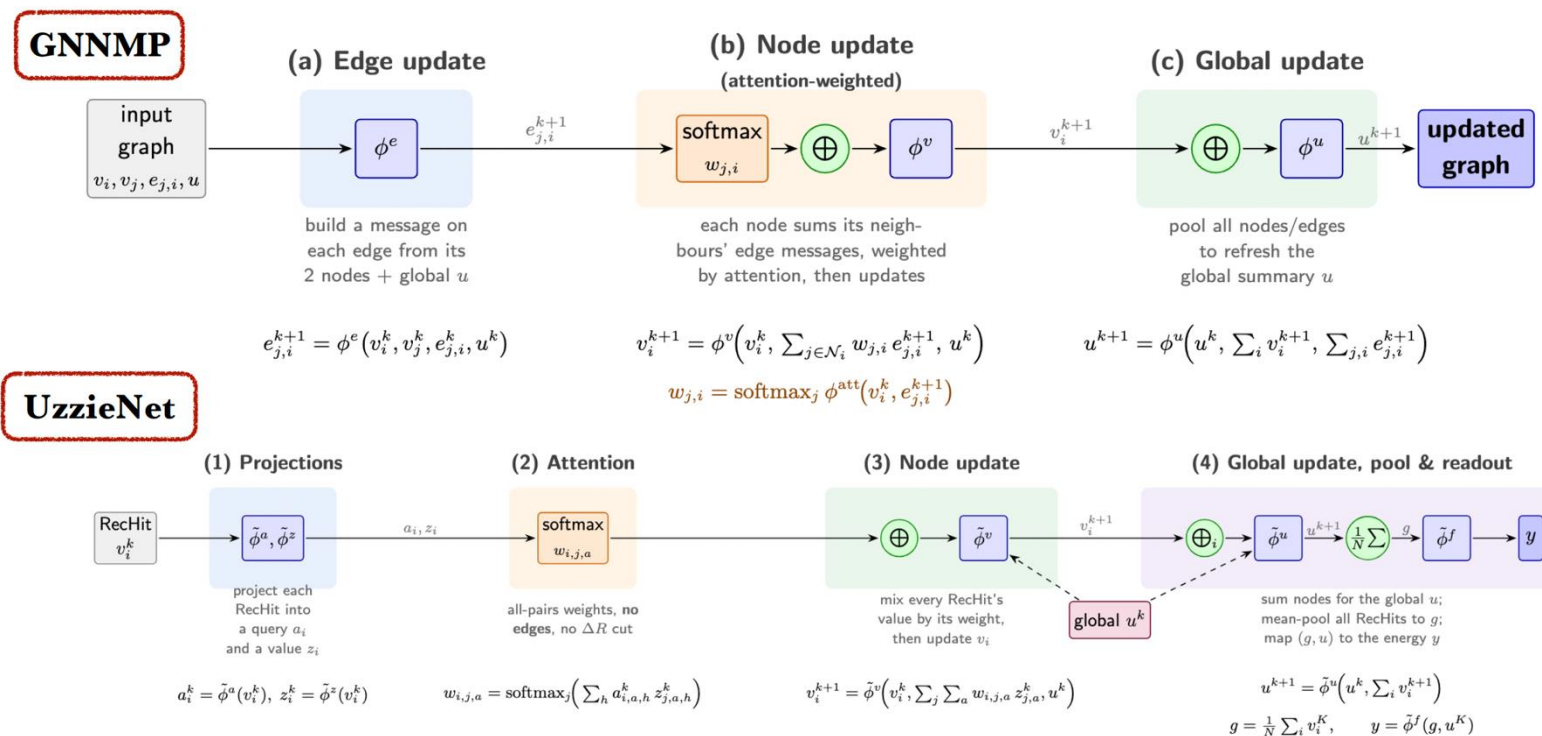
GNNMP



UzzieNet

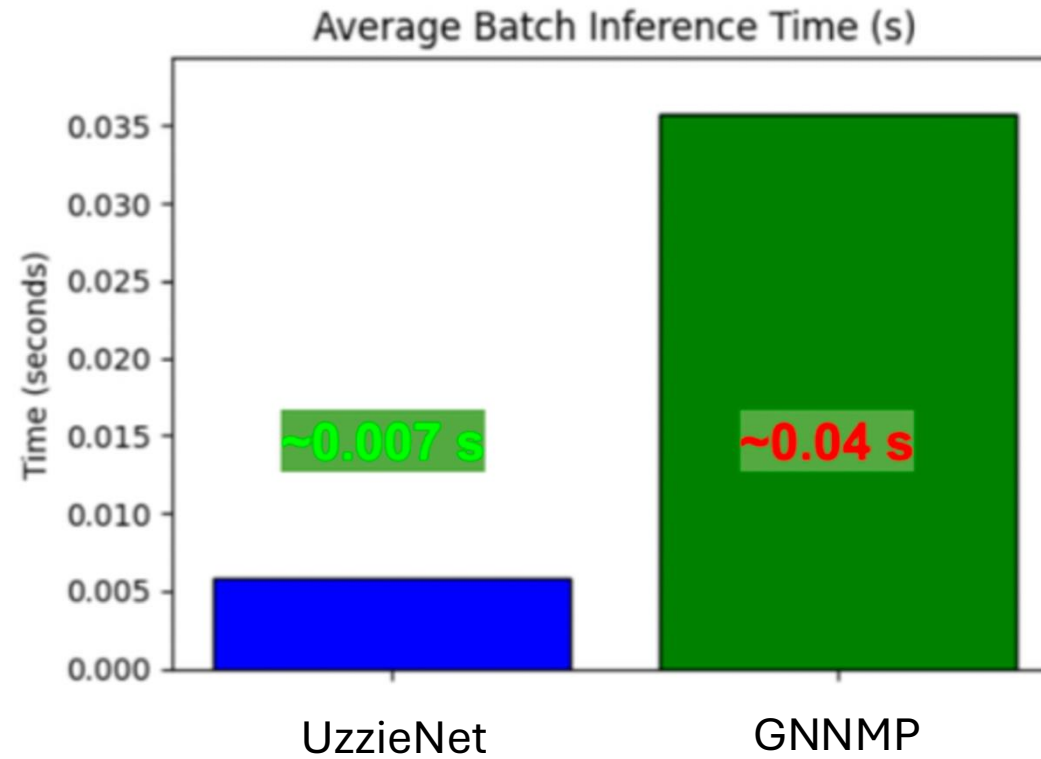


Calo ML reconstruction: GNN



Calo ML reconstruction: GNN

- Fast

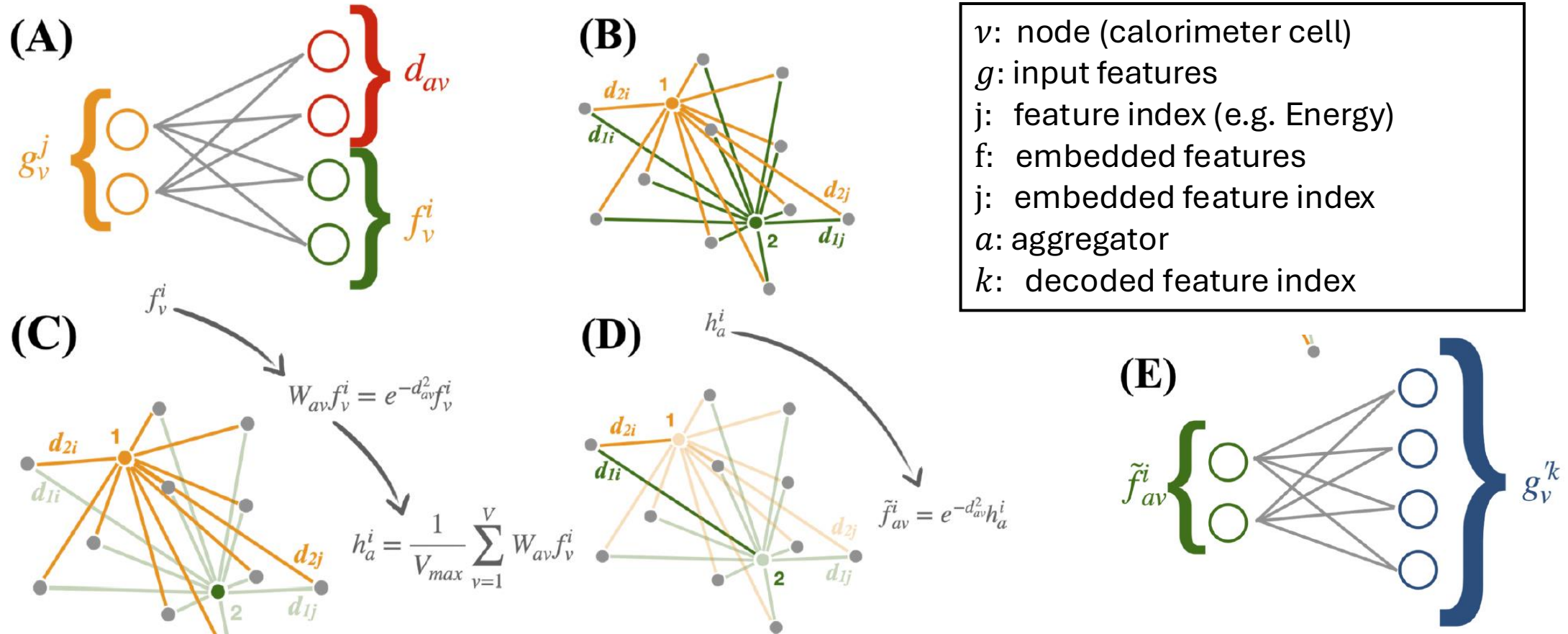


Calo ML reconstruction: GNN

- Our focus has been implementation and optimization
 - Architecture
 - Input engineering

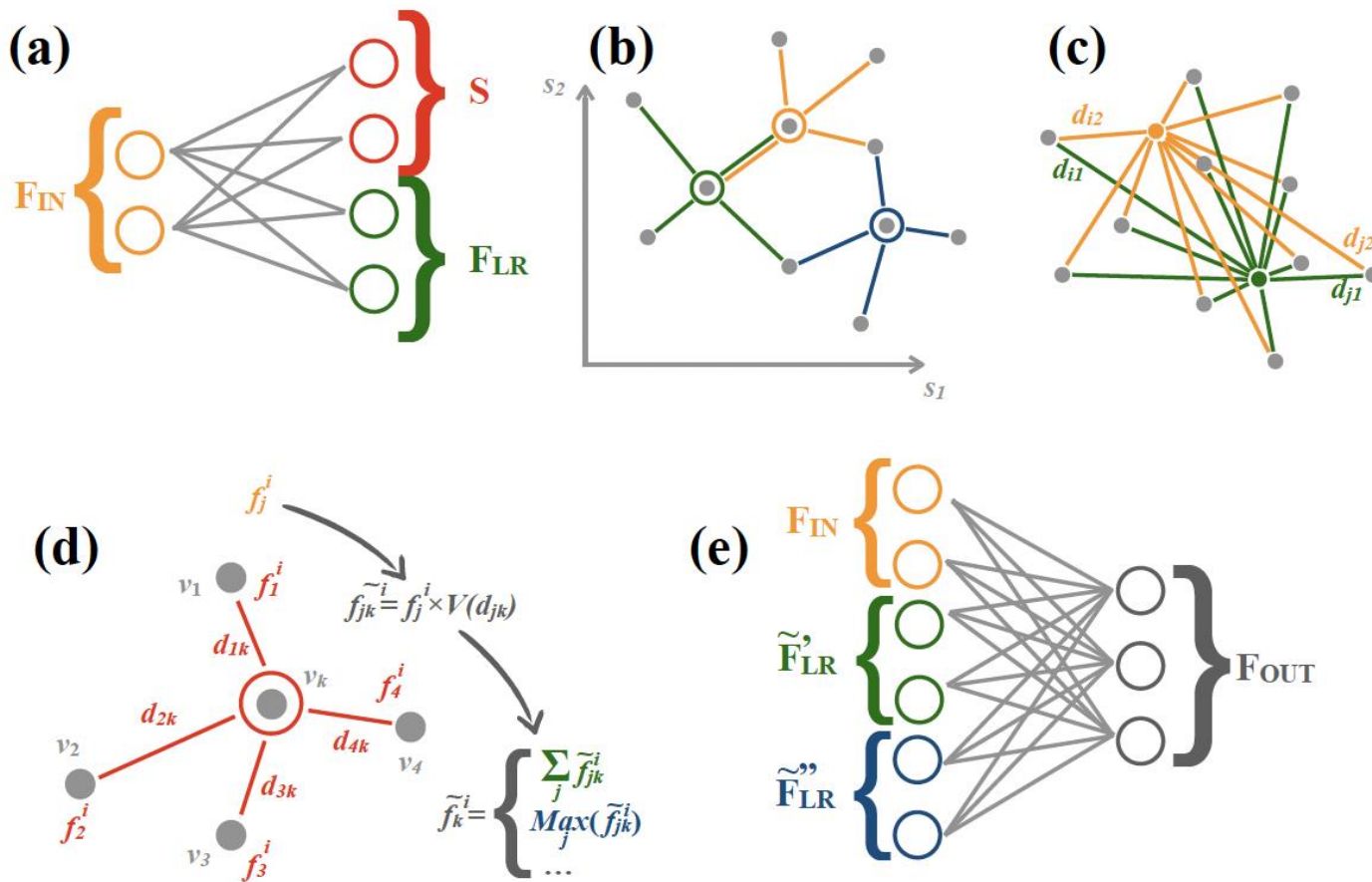
PicoCal GNN optimization

- Architecture: GarNet II



PicoCal GNN optimization

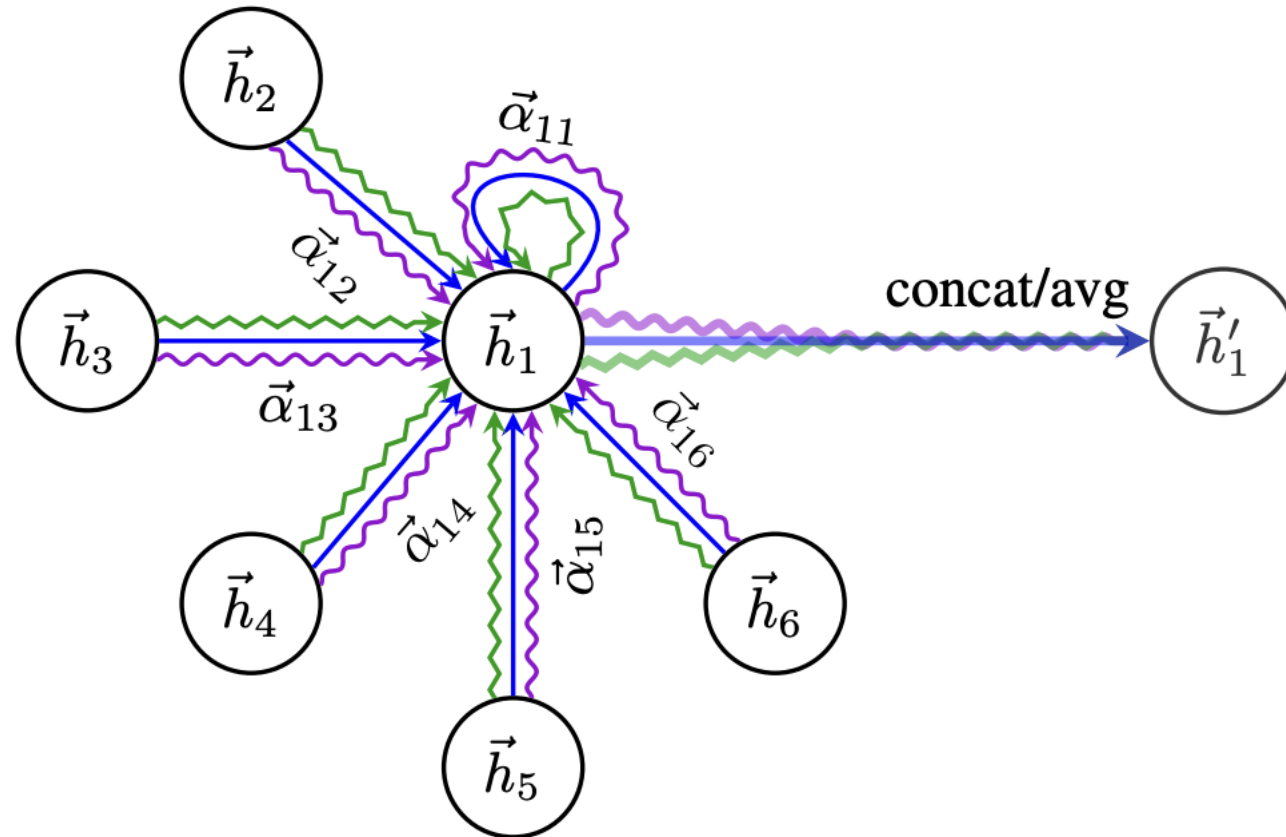
- Architecture: GarNet I



- **GarNet model:** The original vertex features are concatenated with the mean of the vertex features and then passed on to one dense layer with 32 nodes and tanh activation before entering 11 subsequent GAR-NET layers. These layers contain $S = 4$ aggregators, to which $F_{LR} = 20$ features are passed, and $F_{OUT} = 32$ output nodes. The output of each layer is passed to the next and added to a vector containing the concatenated outputs of each GARNET layer. The latter is finally passed to a dense layer with 48 nodes and ReLU activation.

PicoCal GNN optimization

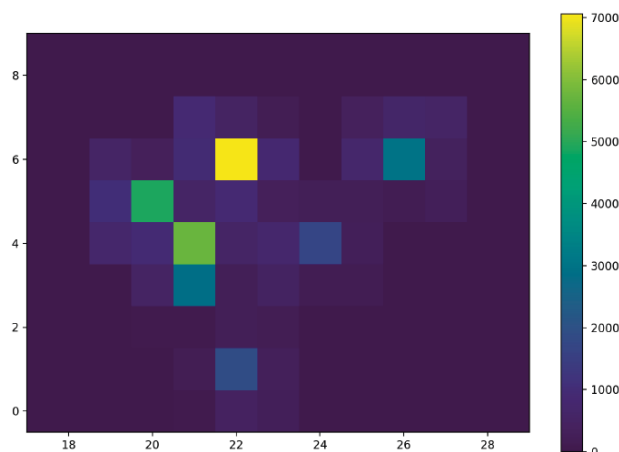
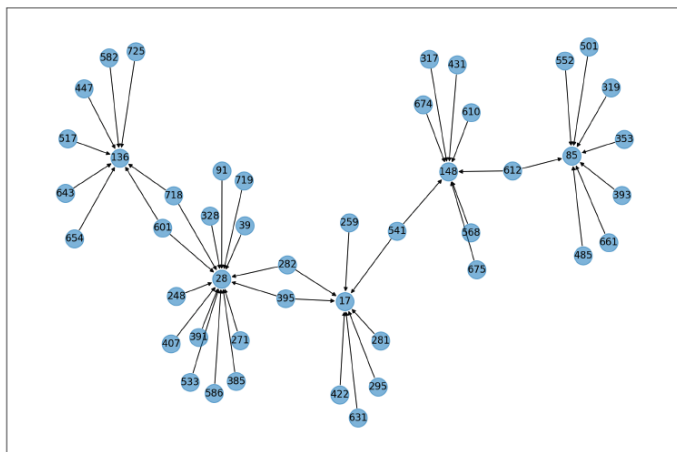
- Architecture: Attention



PicoCal GNN optimization

- Architecture:
 - Quantization
 - Pruning
 - Distillation
 - ?

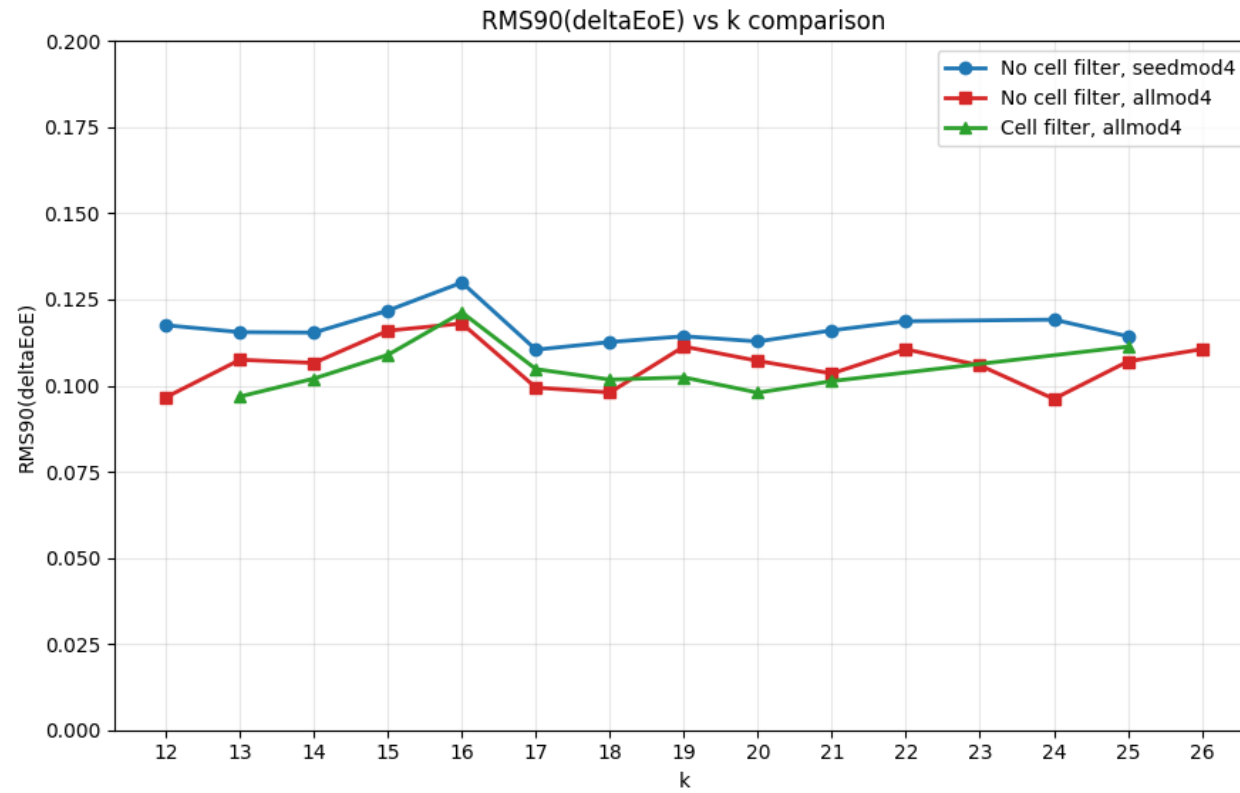
- Input engineering



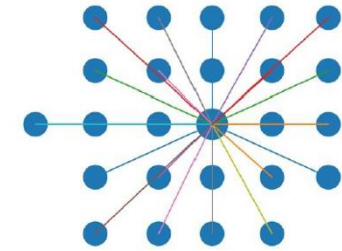
26

PicoCal GNN optimization

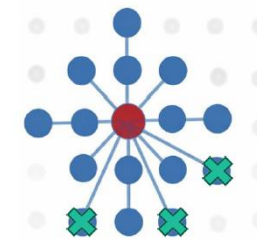
- Input engineering



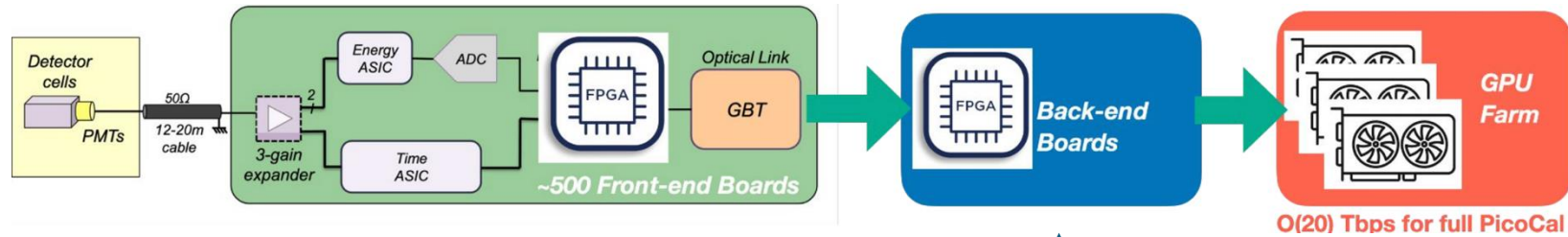
kNN after filtering



kNN w/o filtering



Bonus: exploiting the hardware chain



Can we load some steps of the reconstruction in the FPGAs?

- Seed finding and graph building can probably run on the FPGAs.
- ML in FPGAs is more challenging but it's a very active field in the community.

Conclusions

- PicoCal ML reconstruction has demonstrated great potential to handle high-occupancy scenarios with great physics performance.
- Throughput remains the main difficulty towards their deployment.
- Active work in the area has demonstrated that smart architecture design leads to impressive throughput optimization.

Backup