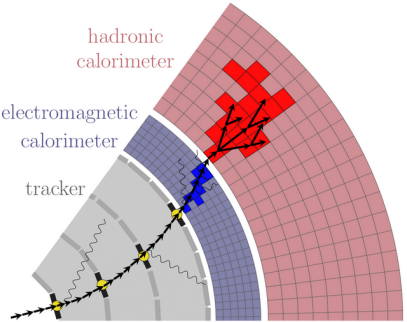
Status of developments on Fast Simulation

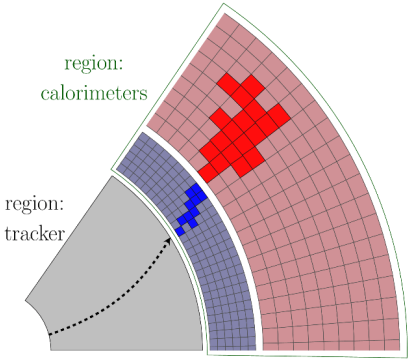
Anna Zaborowska
Dalila Salamani

Geant4 Collaboration Week, 26.09.2023

Fast simulation @ HEP

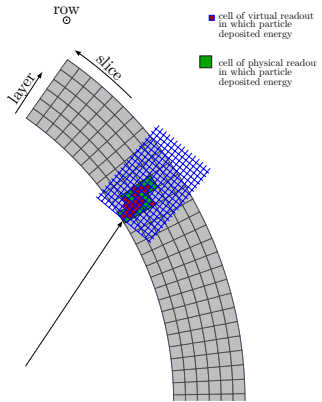
Approximation of full simulation, with focus on calorimeters as the most time consuming detectors.





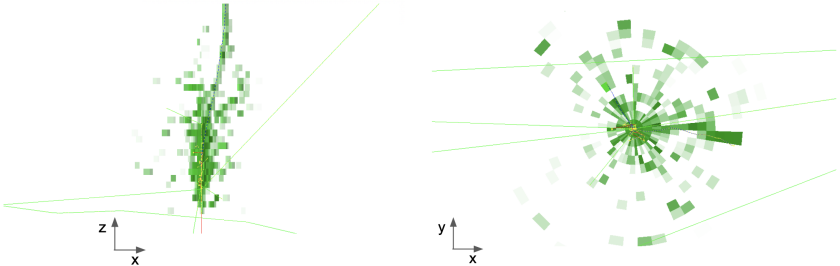
Geant4 example Par04

- Extended example Par04 shows how to use Machine Learning (ML) models within GEANT4.
- Distributed with a Variational Autoencder (VAE) model of showers used in fast simulation
- Demonstrates how to incorporate inference libraries: ONNX runtime, pyTorch, lwtnn.
- In the coming release also allows to run it on GPU (a choice done by UI command).
- It scores energy along shower axis, performs validation of shower observables.
- Recent additions of physical detector readout for performance benchmarking.



Geant4 example Par04: datasets

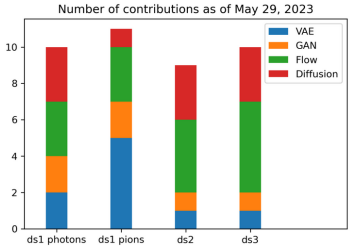
- Full simulation source of datasets:
  - on zenodo: High Granularity Electromagnetic Calorimeter Shower Images
    DOI 10.5281/zenodo.6082201
  - used as dataset 2 and 3 in CaloChallenge:
    - dataset 2: 6'480 voxels DOI 10.5281/zenodo.6366271
    - dataset 3: 40'500 voxels DOI 10.5281/zenodo.6366324



Calo Challenge calochallenge.github.io

Michele Faucci Giannelli, Gregor Kasieczka, Claudius Krause, Ben Nachman, Dalila Salamani, David Shih and Anna Zaborowska

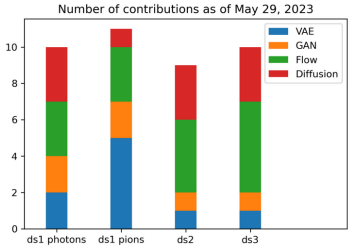
- Lots of on-going effort with ML models for fast shower simulation, but with **no place for comparison or discussions** of different methods.
- This challenge was released with three datasets with increasing dimension of input: first one comes from ATLAS open data, **the 2nd and 3rd from Par04**.



Calo Challenge calochallenge.github.io

Michele Faucci Giannelli, Gregor Kasieczka, Claudius Krause, Ben Nachman, Dalila Salamani, David Shih and Anna Zaborowska

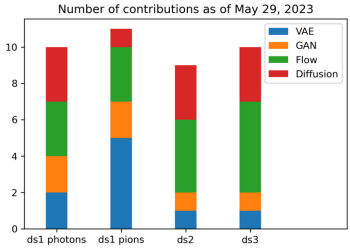
- Lots of on-going effort with ML models for fast shower simulation, but with **no place for comparison or discussions** of different methods.
- This challenge was released with three datasets with increasing dimension of input: first one comes from ATLAS open data, **the 2nd and 3rd from Par04**.
- Set of **common datasets** and **common validation metrics** allows to productively discuss different approaches.



Calo Challenge calochallenge.github.io

Michele Faucci Giannelli, Gregor Kasieczka, Claudius Krause, Ben Nachman, Dalila Salamani, David Shih and Anna Zaborowska

- Lots of on-going effort with ML models for fast shower simulation, but with **no place for comparison or discussions** of different methods.
- This challenge was released with three datasets with increasing dimension of input: first one comes from ATLAS open data, **the 2nd and 3rd from Par04**.
- Set of **common datasets** and **common validation metrics** allows to productively discuss different approaches.

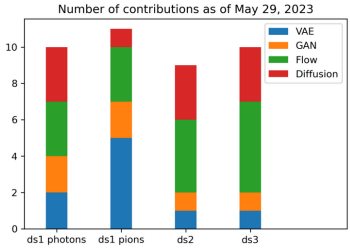


- No winner is expected, it's a compilation of alternative solutions, often with different architectures.

Calo Challenge calochallenge.github.io

Michele Faucci Giannelli, Gregor Kasieczka, Claudius Krause, Ben Nachman, Dalila Salamani, David Shih and Anna Zaborowska

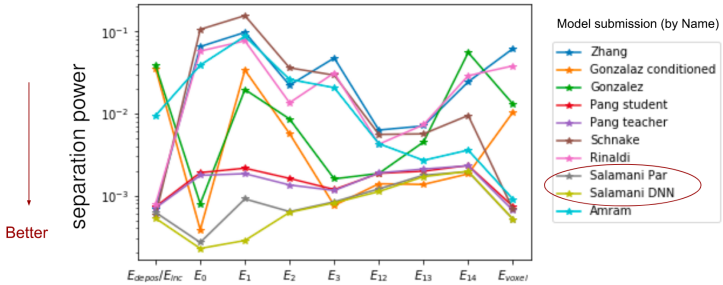
- Lots of on-going effort with ML models for fast shower simulation, but with **no place for comparison or discussions** of different methods.
- This challenge was released with three datasets with increasing dimension of input: first one comes from ATLAS open data, **the 2nd and 3rd from Par04**.
- Set of **common datasets** and **common validation metrics** allows to productively discuss different approaches.



- No winner is expected, it's a compilation of alternative solutions, often with different architectures.
- Calo Challenge Workshop held on 30-31 May 2023 in Frascati gathered many contributors and was a place for fruitful discussions, and first benchmarks.
- Final benchmarking and paper write-up is in progress.

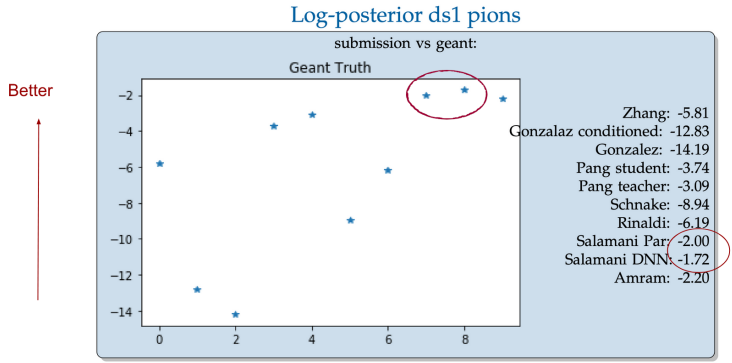
Calo Challenge: VAE model

Variational autoencoder (VAE) model developed for ATLAS pion dataset by Dalila Salamani.
It seems to be very attractive, both comparing shower observables (histograms),



Calo Challenge: VAE model

Variational autoencoder (VAE) model developed for ATLAS pion dataset by Dalila Salamani.
It seems to be very attractive, both comparing shower observables (histograms), as well as the binary classifier (single cell comparison).
Currently being implemented within ATLAS.



Calo Challenge: transformer-based model

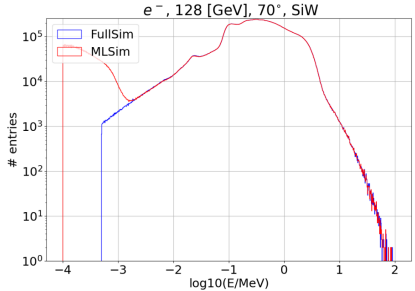
Also developed in EP-SFT by Piyush Raikwar, in cooperation with CERN openlab and IBM
Transformer architecture is powerful, proposed in 2017 (10.48550/arXiv.1706.03762) and gained lots of attention recently (chatGPT, DALL-E, ...)

DALL-E output for "Geant4 members admiring the best view in Sapporo"

Calo Challenge: transformer-based model

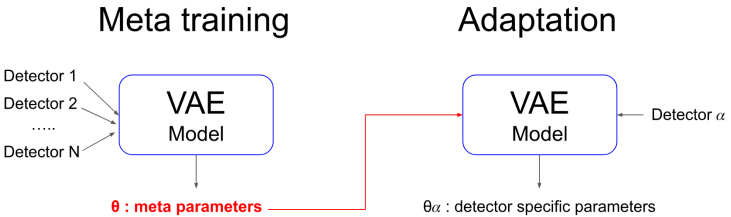
Also developed in EP-SFT by Piyush Raikwar, in cooperation with CERN openlab and IBM
Transformer architecture is powerful, proposed in 2017 (10.48550/arXiv.1706.03762) and gained lots of attention recently (chatGPT, DALL-E, ...)

- The main objective is to better model the cell energy distribution
- Final goal is to train a model on showers from various detectors to allow it to learn representations, with quick adaptation phase to each experiment



not yet a final result form a generative model, a interim prototype

MetaHEP



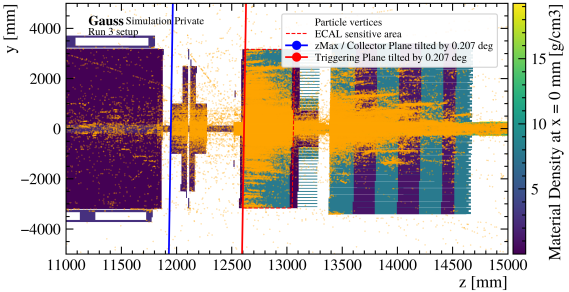
Finalised publication
MetaHEP: Meta learning for fast shower simulation of high energy physics experiments
10.1016/j.physletb.2023.138079.

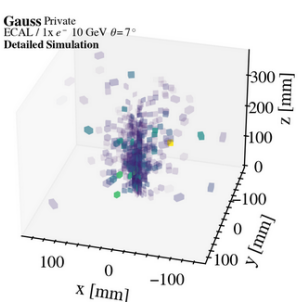
- Covered at the last Geant4 Collaboration Week by Dalila
- Extensive instruction on how to use MetaHEP for any detector described at our website: g4fastsim.web.cern.ch/
- Documentation based on Par04 and the first experiment-framework implementation: in key4HEP simulation toolkit k4SimGeant4
- Inference part of Par04 was also integrated into DD4hep by DESY group (Frank Gaede et al) as DDML

Implementation within Gaussino for the LHCb

work done by Michal Mazurek

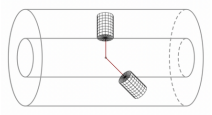
- Implementation of Par04 approach in Gaussino, use on LHCb's calorimetry;
- Many cavaets discovered on the way, as expected in realistic conditions;
- Implementation ready to be used, first model (VAE) is being tested;
- Very general implementation, allows to use MetaHEP, or any models submitted to CaloChallange;





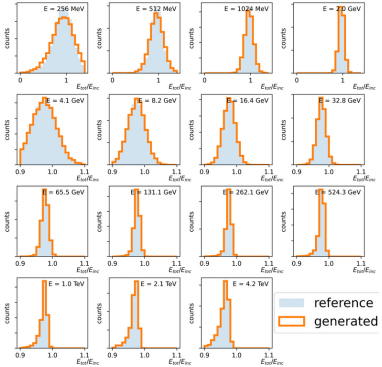
First look on MetaHEP for ATLAS

MetaHEP : Adaptation



$R_x P_x N = 18_x 50_x 45$

Artificially placing the ATLAS voxels into $R_x P_x N$ voxels and padding with 0



CaloChallenge ATLAS Dataset 1 photons

Naturally, a proper “cylinder” should come directly from simulation, but looks promising!

Classical parameterisation

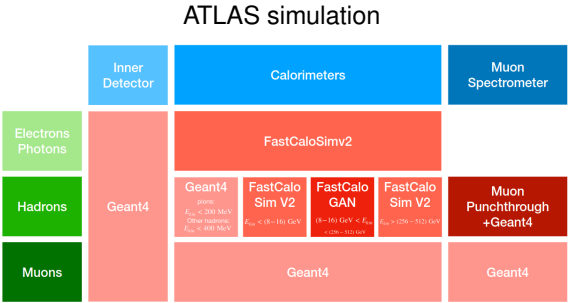


Figure 17: The configuration of the different tools used for AtlFast3, which depends on the particle type, the detector and the particle energy.

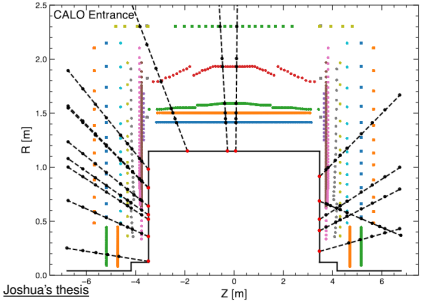
10.1007/s41781-021-00079-7

- Most R&D ML-driven;
- But ML still not widely used in production;

FastCaloSimV2-based classical shower simulation

work done by Joshua Falco Beirer

- ATLAS moves to Geant4 fast-sim hooks;
- Extraction of classical parameterisation to a package only Geant4-dependant;
- Getting it ATLAS independent starting with extrapolation to calo layers;
- Next moving to extraction of parameterisation to be ATLAS geometry independent;

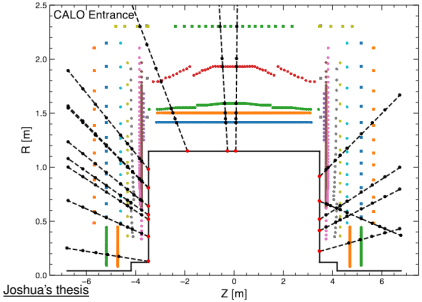


Joshua's thesis

FastCaloSimV2-based classical shower simulation

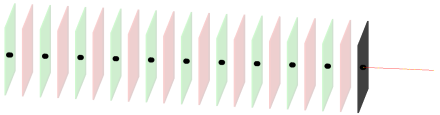
work done by Joshua Falco Beirer

- ATLAS moves to Geant4 fast-sim hooks;
- Extraction of classical parameterisation to a package only Geant4-dependant;
- Getting it ATLAS independent starting with extrapolation to calo layers;
- Next moving to extraction of parameterisation to be ATLAS geometry independent;



Joshua's thesis

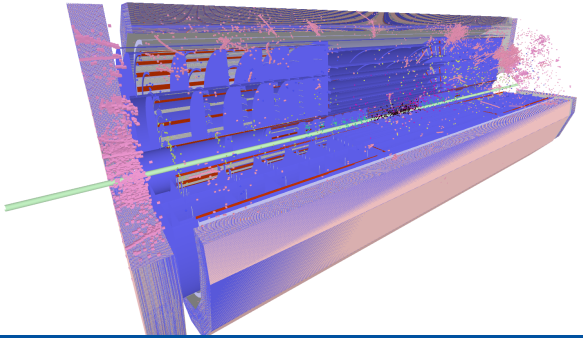
- First tests to validate against on ATLAS geometry, then move towards some future detector (or Open Data Detector)

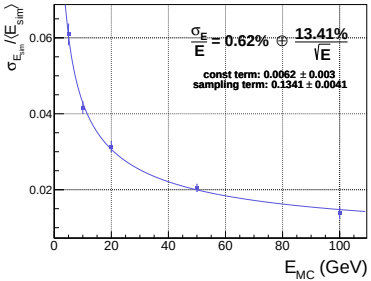


Open Data Detector

Open Data Detector is a benchmark detector for algorithmic studies (reconstruction, fast simulation, ...)

Will be the source for the next datasets, offering more realism and complexity, as well as different structure of data.





Summary and Outlook

On-going work on finalisation of many of the topics:

- Calo Challenge write-up and final benchmark results;
- Open Data Detector finalisation and publication of datasets;
- Application on ML models for LHCb calorimetry;

Summary and Outlook

On-going work on finalisation of many of the topics:

- Calo Challenge write-up and final benchmark results;
- Open Data Detector finalisation and publication of datasets;
- Application on ML models for LHCb calorimetry;

More long-term plans:

- Similar implementation of MetaHEP for other experiments;
- Classical parameterisation a standalone package available for experiments;
- Development of ML models based on transformers, applicable within above-mentioned implementations to frameworks;
- Support to community with efforts like Calo Challenge, Open Data Detector, making sure the link to Geant4 is maintained and facilitate use by experiments.