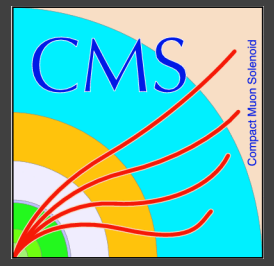
HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

DeGeSim

Deep Generative models for fast and precise physics Simulation



JSC - Jülich Supercomputing Centre

Dirk Krücker
DESY CMS group
Hamburg, 7.9.2020

The Place

DESY is a research center in Hamburg/Germany

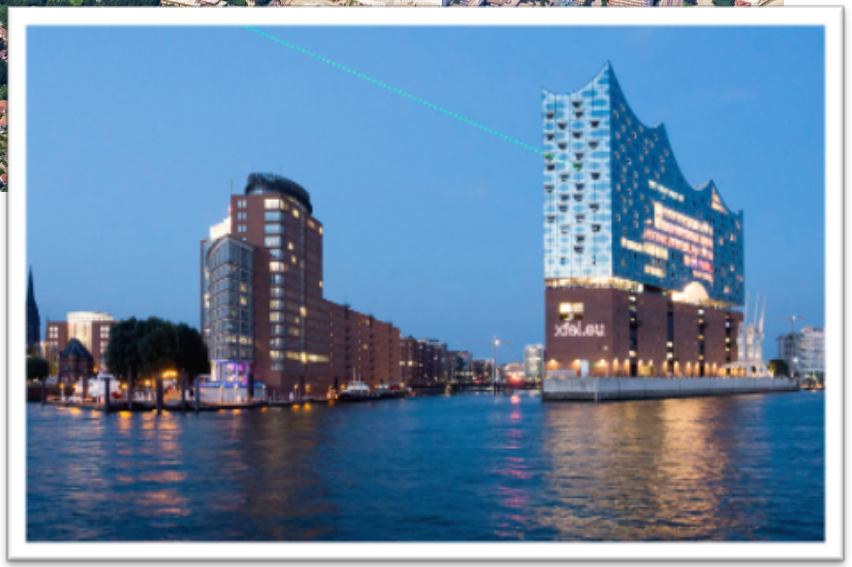
DESY is involved in several HEP experiments

- **CMS** group ~100
- ATLAS group ~100
  - Belle, etc
- We are a large Tier 2 center
- Home of the German National Analysis Facility (NAF) providing computing resources for the German particle physics community
- Running summer schools, e.g. Machine Learning for particle physics

Beside that, we are also a large laboratory for photon science, and we run the linear accelerator for the European XFEL

DESY is member of the Helmholtz Association, an association of 19 German research centers

A nice place to live and to work

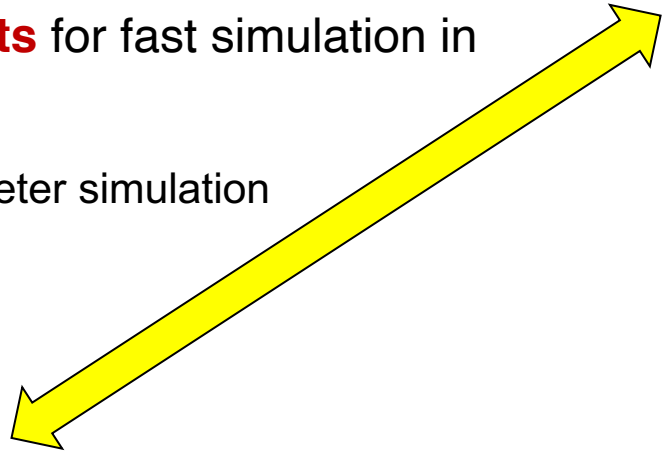


HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

Helmholtz.ai

Helmholtz.ai provides support for the application of Machine Learning/Deep Learning/Artificial Intelligence

- Helmholtz likes the idea to **connect** different areas within the Helmholtz association and provides funding for interesting projects between groups
- We got founding for research on **Generative Deep Neural Nets** for fast simulation in particle physics
  - I am interested in fast calorimeter simulation



JÜLICH FORSCHUNGSZENTRUM

JSC - Jülich Supercomputing Centre

- FZ Jülich will provide supercomputing resources resources (HPC with GPUs)

Picture Copyright: Forschungszentrum Jülich / Ralf-Uwe Limbach

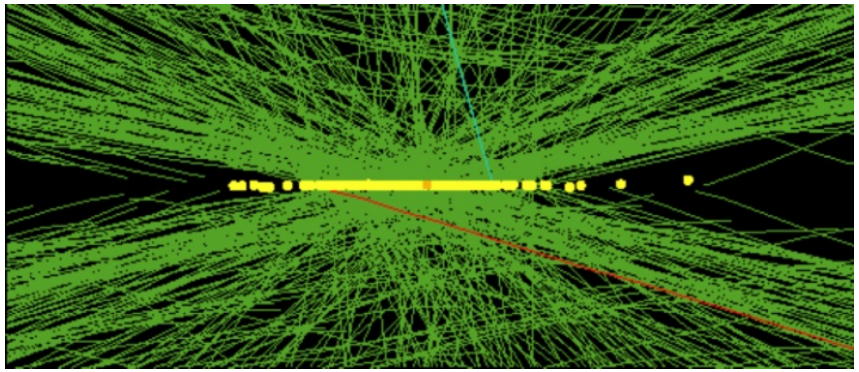
The project

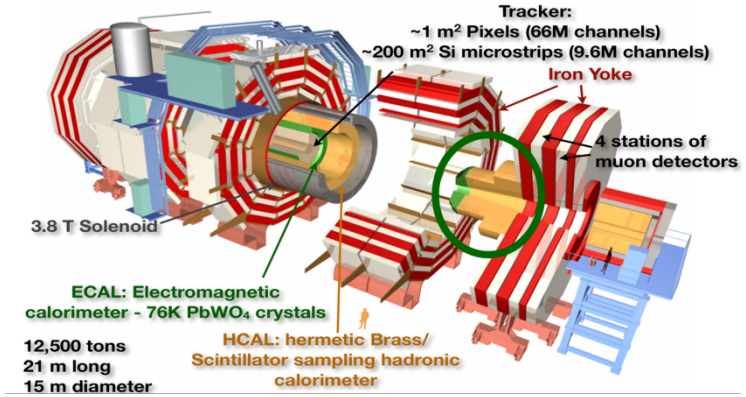
DeGeSim - Deep Generative Models e.g. GANs

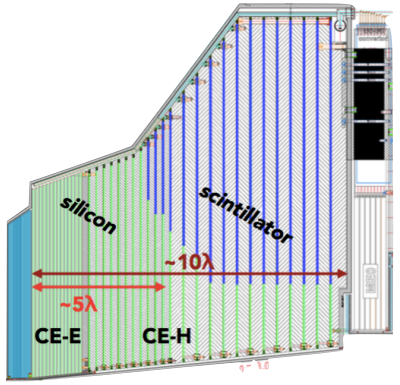
- The CMS branch of the project is looking for a clever PhD to work on Calorimeter Simulation with Deep Learning techniques aiming for HGCal (special interest in Hadron calorimeter simulation)

Massive need for simulation and smart reconstruction algorithms for HL-LHC (2027) and beyond:

- Phase II Upgrade: pile-up → 200
  → novel fine granular detectors
  CMS-HGCal: **6 M channels** in 5dim
  (space+energy+time)







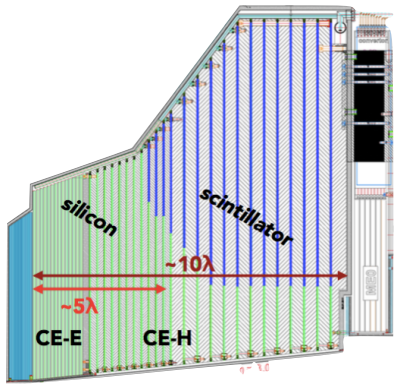
The Phase II CMS endcap calorimeter: HGCal

Fast simulation for CMS HGCal

Massive need for simulation and smart reconstruction algorithms for HL-LHC and beyond:

- Investigate Deep Learning methods (like GANs, VAEs) to produce fast simulation with high precision for particle tracking, electromagnetic and hadronic showers

Idea first presented in:
L. Oliveira, M. Paganini and B. Nachman
https://arxiv.org/abs/1705.02355



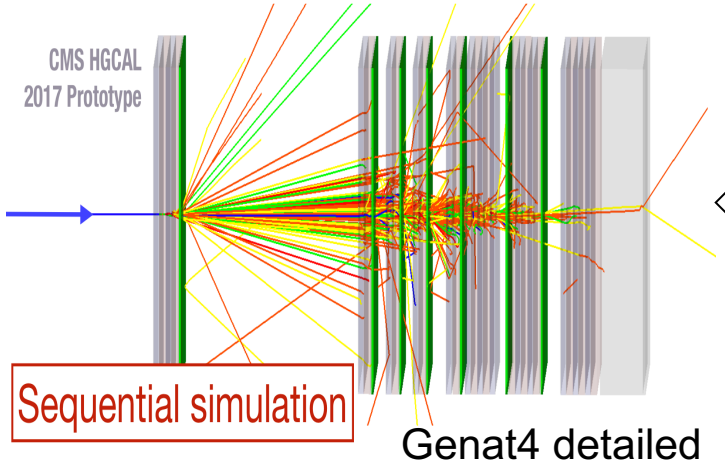
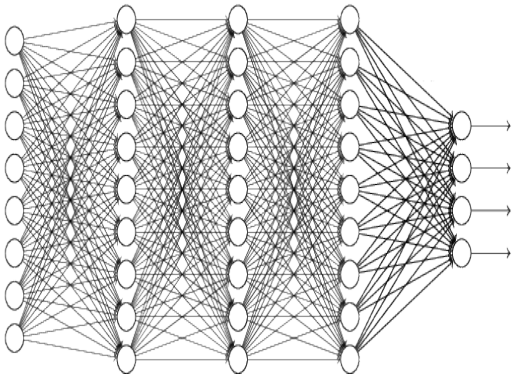


image: Torben Quast

Genat4 detailed simulation (**slow**)



Deep Neural Network **(fast)**

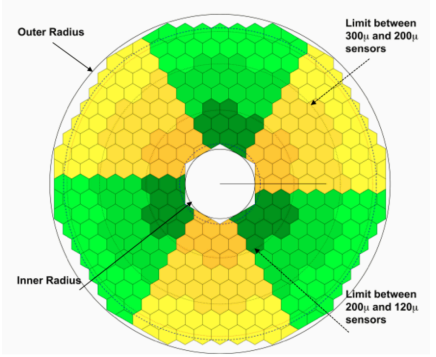
Fast simulation for CMS HGCal

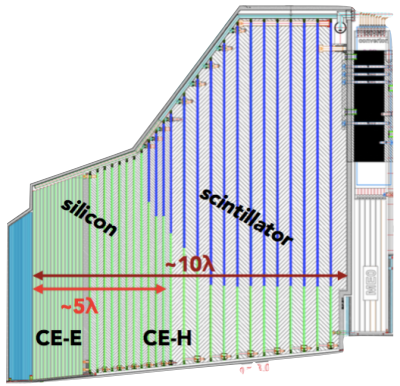
Massive need for simulation and smart reconstruction algorithms for HL-LHC and beyond:

- Investigate Deep Learning methods (like GANs, VAEs) to produce fast simulation with high precision for particle tracking, electromagnetic and hadronic showers

Easy idea but many practical challenges

- Conditioning on input particle
  - HGCal complicated hexagonal geometry
- Physical distributions must often be enforced
- GANs are famous for difficult training
- Is Genat4 optimal for hadronic showers
  - Additional training on real data
- etc





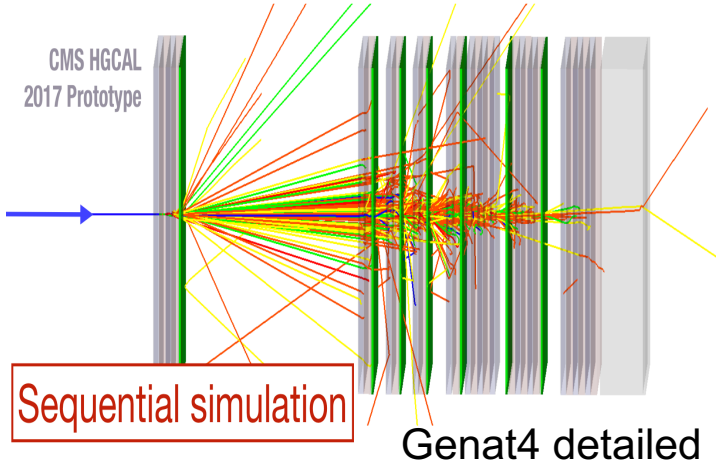
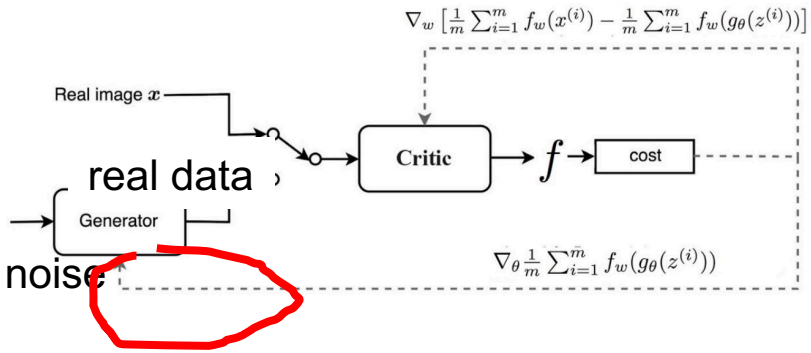


image: Torben Quast

Genat4 detailed simulation (**slow**)



E.g. GAN **(fast)**

The job

Start time: we aim for September the 1st – duration 3 years

- **Based at DESY** visits to FZ Jülich (JSC) and CERN
- PhD in physics from University **Aachen**
  (prof. Kerstin Borras)
- This will not be a regular physics PhD but mainly data science!
  You will probably not have the time to do a CMS physics data analysis
  - Not author of CMS
- Looking for a candidate with strong interest in computing/deep learning with a good math background (not just a hacker ☺) and understanding of calorimeters

- We will try to hire the CMS and ATLAS student roughly at the same time
- There is already an ongoing project, AMALEA, at DESY looking into GANs for the Calice calorimeter (same detector design as HGCal) for the ILD
  - There is test beam data

⟹ To get the critical mass of people working on Generative Deep Learning

Appendix

CMS and ATLAS

Deep Generative models for fast and precise physics Simulation

- **CMS** Fast simulation
  - Detailed GEANT4 simulations are computing intensive
    - CMS can be several minutes for one event already now
  - **HL-LHC** will increase the amount of data by at least an order of magnitude but computing resources will not scale appropriate
  - Complex systems are demanding
    - **HGCAL**: 6 Million readout channels
- **ATLAS** precise simulation
  - Pile-up is largely soft QCD physics
    - Difficult to simulate
  - Learn a statistical model from directly from data
    - Apply the same/similar generative DL approaches