# Open Data Detector

Dimitra Amperiadou, Erica Brondolin, Noemi Calace, Paul Gessinger, Andreas Salzburger, **Anna Zaborowska**

CERN

ML4Jets2023



This work benefited from support by the CERN Strategic R&D Programme on Technologies for Future Experiments (CERN-OPEN-2018-006)

**Goals:** Open-access detector for algorithmic development and benchmarking;

Release of the large-scale public dataset;

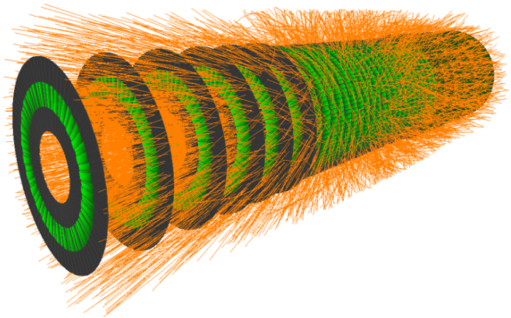


Comments: anna.zaborowska@cern.ch

Tracking machine learning challenge (2018)

Definition of an HL-LHC-like detector.

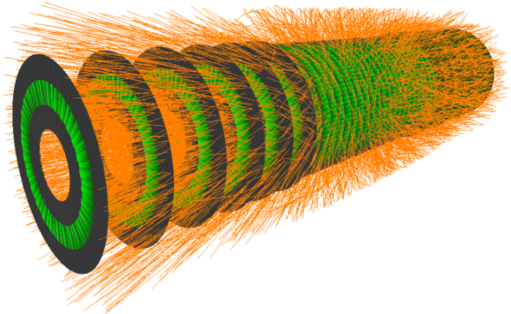
Release of the datasets.



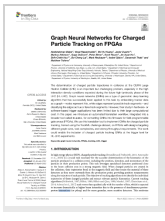
Tracking machine learning challenge (2018)

Definition of an HL-LHC-like detector.

Release of the datasets.



A very rich scientific outcome - and a dataset that is **still in use**.



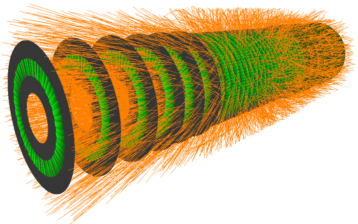


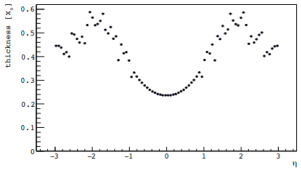

...

# The TrackML shortcomings

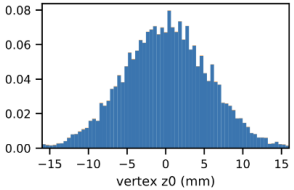
TrackML dataset:

- restricted to a very optimistic Inner Tracking Detector
- generated with fast simulation

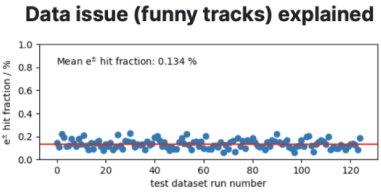





Too optimistic material estimates




Too narrow beam spot for a realistic (HL-)LHC like experiment




Loopers in the dataset,
wrong scattering formula for electrons

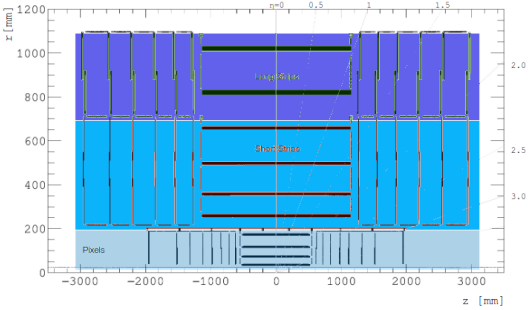
## The evolution: ODD tracker

TrackML detector re-implemention using DD4hep.

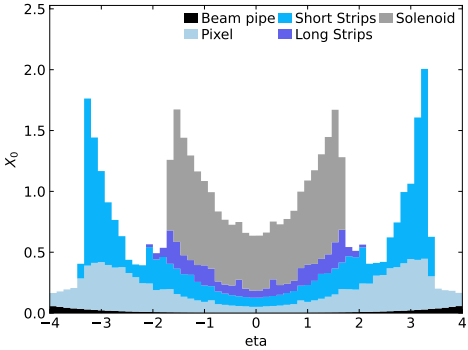
Sub-systems:

- Innermost Pixel System
- Short Strip System
- Long Strip System

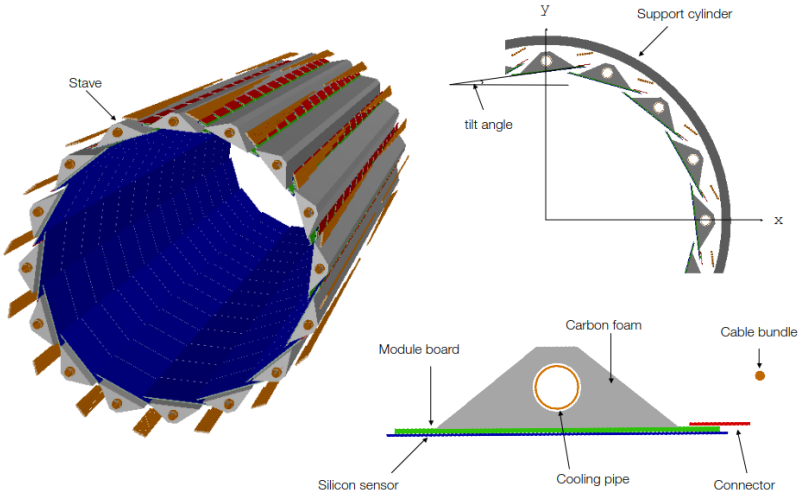
- hermetic
- minimum 12 hit system,
- Silicon sensors
- $|\eta| < 3$




Geant4 hit locations of the ODD detector.



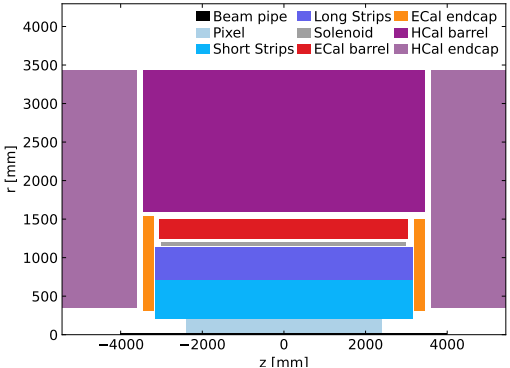
# Higher level of detail

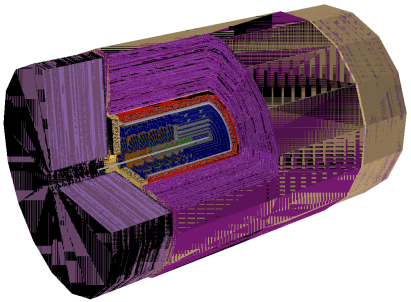


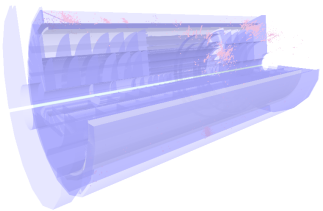

Innermost Pixel detector barrel, stave and module details of the ODD Tracker.
Displayed with DD4hep/ROOT.

- Aim at description at a similar level of detail as trackers for LHC experiments.
- Increases complexity of the created dataset.

# Complete system





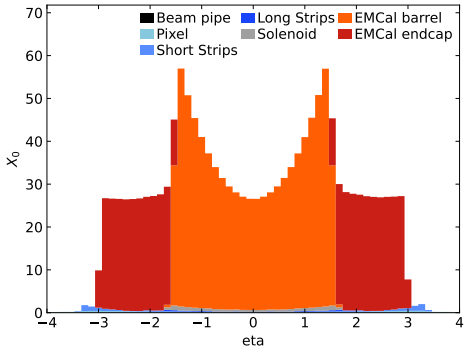
Extension of the detector description to the other sub-systems:

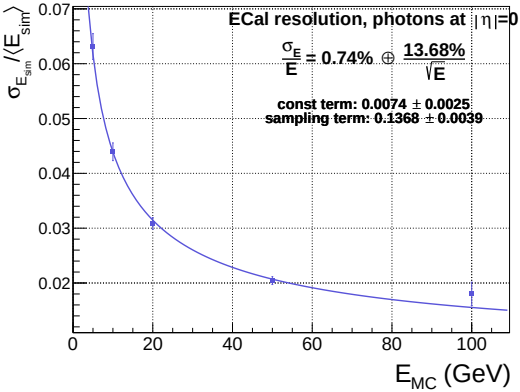
- Electromagnetic calorimeters
- Hadronic calorimeters
- Muon system [work-in-progress]

Barrel + endcap system covering $|\eta| < 3$.

# ODD electromagnetic calorimeter

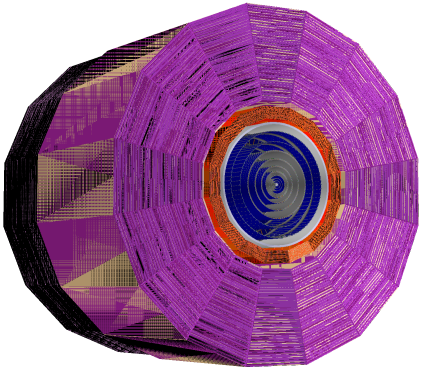
- Silicon-based calorimeter inspired by CMS HGCal, CLD, SiD, ... detectors
- Hexadecagon (16-sided polygon) in cross section.
- 48 sampling layers with 1.9 mm W, 0.5 mm Si, and readout (PCB, glue, air, ...) with 5.1 mm square cells.

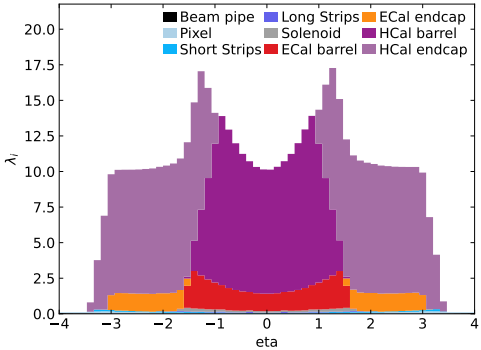


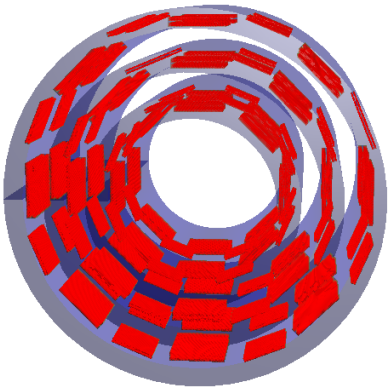
# ODD hadronic calorimeter

- Scintillator-based calorimeter inspired by CMS HGCal, AHCAL, CLD, SiD, ... detectors
- Hexadecagon in cross section.
- 30 sampling layers with 30 mm Fe, 3 mm Sci, and readout (PCB, glue, air, ...) with 30 mm square cells.



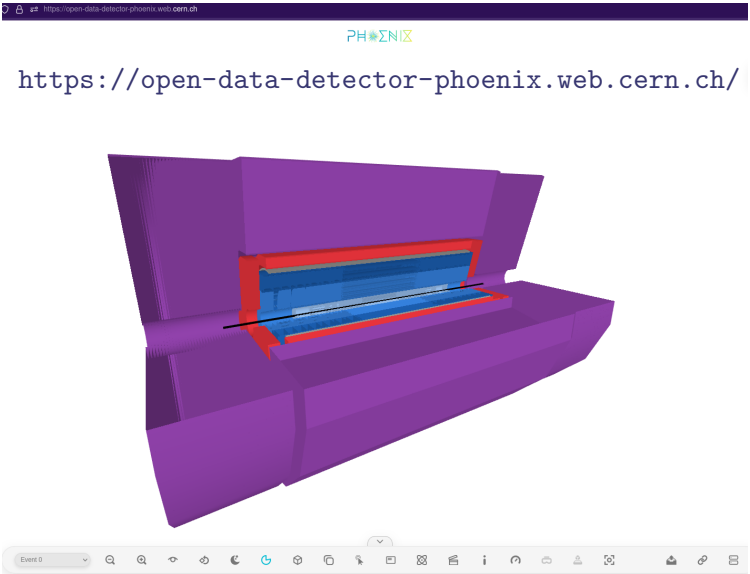


# ODD muon spectrometer



- Work in progress
- 3 layers of chambers with tubes
- Flexible construction factory
- Allows to complement tracker system for combined muon reconstruction studies

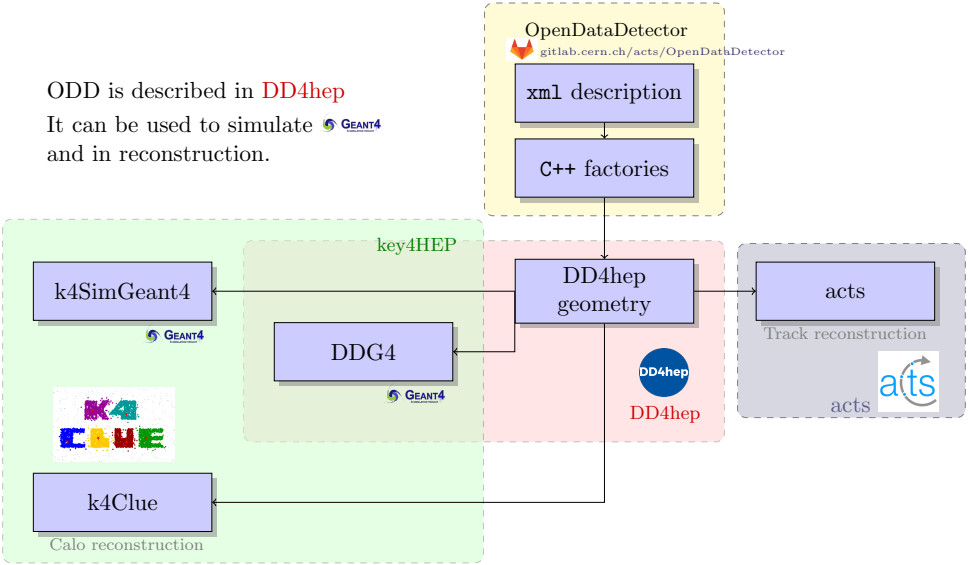
# Visualization



# How to use ODD

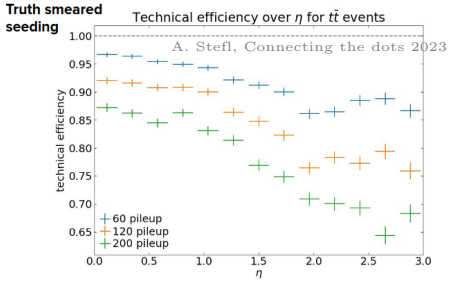
ODD is described in DD4hep

It can be used to simulate Geant4 and in reconstruction.



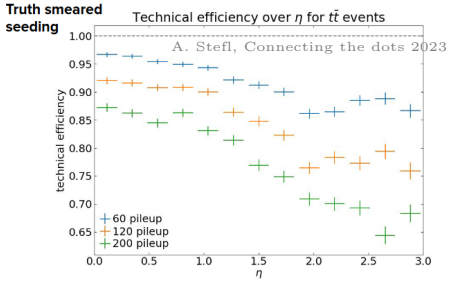
# ODD dataset potential

- ODD Tracker is heavily used in development of ACTS track reconstruction toolkit.



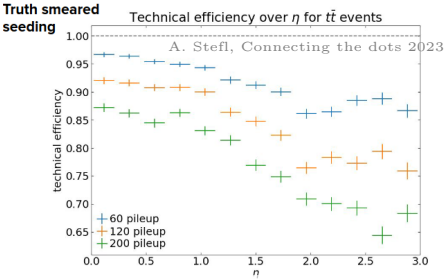
# ODD dataset potential

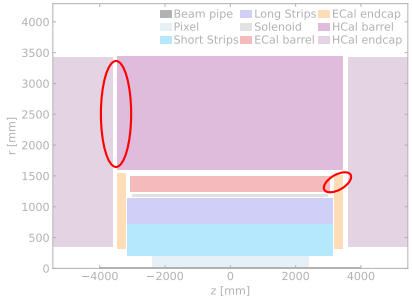
- ODD Tracker is heavily used in development of ACTS track reconstruction toolkit.
- A realistic detector is essential to allow for state-of-the art R&D and testing.



# ODD dataset potential

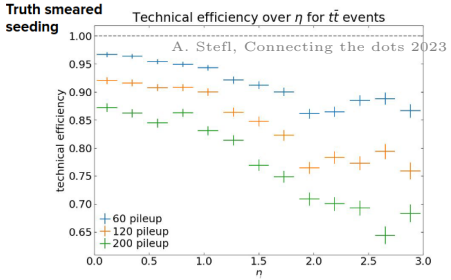
- ODD Tracker is heavily used in development of ACTS track reconstruction toolkit.
- A realistic detector is essential to allow for state-of-the art R&D and testing.
- The complexity for calorimetry is represented by a realistic layout and the barrel/endcap transition region.

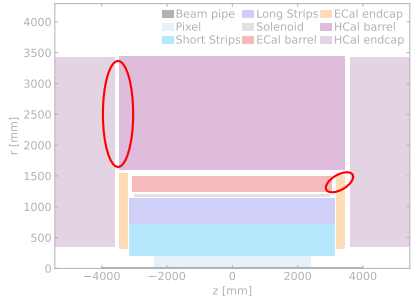


# ODD dataset potential

- ODD Tracker is heavily used in development of ACTS track reconstruction toolkit.
- A realistic detector is essential to allow for state-of-the art R&D and testing.
- The complexity for calorimetry is represented by a realistic layout and the barrel/endcap transition region.
- The dataset, once published:
  - will replace the TrackML for track reconstruction studies
  - calorimetry can be used for fast shower simulation studies, as well as calorimetry reconstruction
  - will offer a possibility to use combined information from tracker-calo(-muon), e.g. for particle-flow studies.

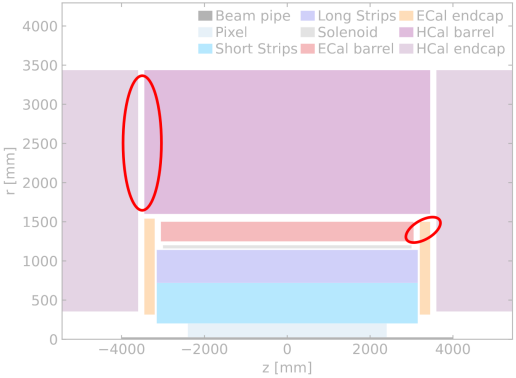


# ODD dataset for fast simulation

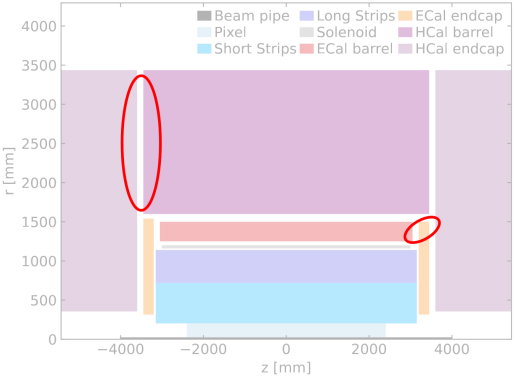
- Different particles: $e^{\pm}$, $\gamma$, $\pi^{\pm}$.

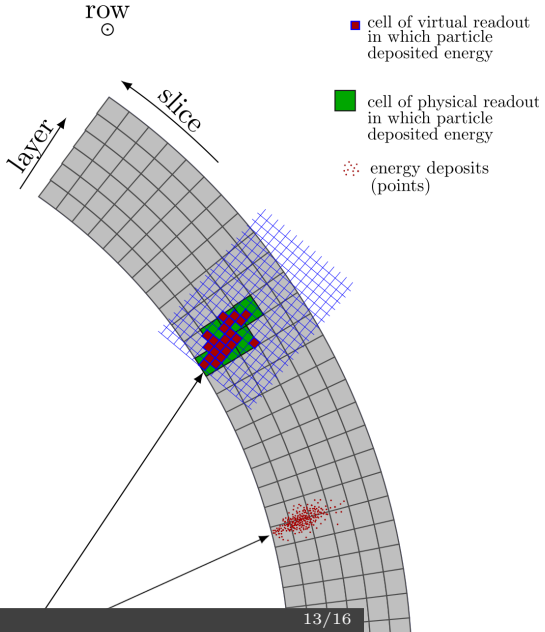
# ODD dataset for fast simulation




- Different particles: $e^{\pm}$, $\gamma$, $\pi^{\pm}$.
- Full detector region, including the challenging gap between barrel and endcaps.

# ODD dataset for fast simulation




- Different particles: $e^{\pm}$, $\gamma$, $\pi^{\pm}$.
- Full detector region, including the challenging gap between barrel and endcaps.
- Different representation of data (voxels and deposits)

## Data representation

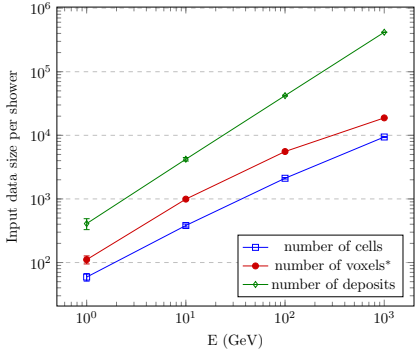
Voxelization of higher granularity than the cell readout (this is also in CaloChallange datasets). Q: What is the right size?

## Data representation

Voxelization of higher granularity than the cell readout (this is also in CaloChallange datasets). Q: What is the right size?

Many models turn towards point clouds and while it's unfeasible to use all simulation deposits, clusterization can be applied. Q: Drop to what size?

| E (GeV) | $N_{cells}$ | $N_{voxels}$ | $N_{deposits}$ |
|---|---|---|---|
| 1 | 59± 8 | 111± 16 | 409± 80 |
| 10 | 382± 21 | 993± 48 | 4210± 266 |
| 100 | 2112± 58 | 5560± 189 | 42073± 800 |
| 1000 | 9410 ± 186 | 18796± 405 | 417317 ±3835 |




* voxelization with 0.5 $X_0 \times 0.25$ $R_M \times 0.125$ rad voxel size

## Expected timeline

| | |
|---|---|
| 2021 | ODD tracker implementation |
| November 2022 | Draft version of the ODD electromagnetic calorimeter |
| October 2023 | Finalisation of ECal and HCal |
| today | Tracker + Calorimetry |
| beginning of 2024 | + Muon spectrometer |
| February 2024 | Proper digitisation implementation for trackers |
| March 2024 | Final validation of calorimetry |
| April 2024 | First prototype dataset (tracker + calorimetry) |

## Summary

Open Data Detector is the next generation of a HL-LHC–like detector for algorithmic studies.

Current implementation includes a full tracking system and calorimetry and can be used in simulation or reconstruction.

## Summary

Open Data Detector is the next generation of a HL-LHC–like detector for algorithmic studies.

Current implementation includes a full tracking system and calorimetry and can be used in simulation or reconstruction.

Once full validation is completed, we will publish the dataset in EDM4HEP (`root`) and/or `hdf5` formats.

## Summary

Open Data Detector is the next generation of a HL-LHC–like detector for algorithmic studies.

Current implementation includes a full tracking system and calorimetry and can be used in simulation or reconstruction.

Once full validation is completed, we will publish the dataset in EDM4HEP (`root`) and/or `hdf5` formats.

To meet the community needs, two further variants will be offered (but w/o the datasets):

- with a drift chamber in place of strips (for ee colliders).
- extending tracker up to $|\eta| < 4$ (for HL-LHC–like studies: ATLAS/CMS).

# BACKUP

## Previous talks

Connecting the dots 2023, *Reconstruction performance with ACTS and the Open Data Detector* , A. Stefl

CHEP 2023, *The Open Data Detector Project*, A. Salzburger

ACAT 2021, *The Open Data Detector - Tracking and Vertexing* , P. Gessinger