Open Data Detector

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

CERN

ML4Jets 2023





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.

And an an and an an and an an				
	And a state of the		Protect	
Image: Source in the source intervent i	INCOME INTO A	- B		
And a	Common Tanàna Indonesi Antoni			
		The Real Providence		Graph Neural Networks for C Particle Tracking on FPGAs
	H. Kasadatay, Astar Jugebby, Sanko and Santay, Astar Jugebby, Sanko and Sanko, Astar Santay, Sanko and Sanko, Astar Santay, Sanko Sanko, Sanko, Sanko, Sanko, Sanko Sanko, Sanko, San	agen" tangan bargi " tanan bargi". Y datan bilangi " tanan bargi". Indang		References from: the framework, in the same set in the actual states, same actual data from the framework actual interaction of the company of the framework actual interaction.
	COCCUT			
Distance	A source of the spin or of the spin or of the spin of the spin of the spin or of the spin	The discretized is the local consequence of the discretized of the di		And and a second s
And a state of the second seco	Labo Alman and an and a second second second and a second second second associations and a second second association and a second second second second association and a second second second second association and a second second second second second association and a second second second second second association and a second second second second second second association association association association association association association association association association association association association association association association association association asociation association association association asocia	Paragenet Plant, Nacional et Vallante, Mariago Cardingh Senator Lancest, Carlingha, Scillante, Jian Cardingh Senator, Nacional Senator, Jian Cardinal Senator, Nacional Senator, Jian Sanasara Handi, Mathanana, Hariako, Mariago Sanasara Senator, Stario, Sanasara Sanasara Senator, Sanasara Senator, Sanasara Senator, Sanasara Sanasara Senator, Sanasara Senator, Sanasara Sanasara Senator, Sanasara Senator, Sanasara Senator, Sanasara Senator, Sanasara Sanasara Senator, Sanasara Senator, Sanasara Senator, Sanasara Senator, Sanasara Sanasara Senator, Sanasara	Letter Letter	 A start of personal start of the start of t
	this also in print	6 h m		

Overgent Porticity Tracking via Mape Cl	authorgindersation Nationalis
lagelisher" Anana Yus' interlage" in Robuster' Fre Stat	Resident Relation and been
Autor rise and August minute and	
Binal and the source of the processing of the burgers in particular of the transmission products in the matter of the processing of the product of the trans- tion of the product of the product of the trans- tion of the product of the product of the trans- tion of the product of the product of the trans- tion of the product of the product of the burgers of the product of the product of the burgers of the product of the	which cits pand muchanish. Oth as vi up provide priority is an index and in the priority manufacture of the priority is and the priority memory of the priority is and the priority is and memory of the priority is and the priority is and the priority is and the priority is and the priority is an index and the priority is and the priority is and the priority is and the priority is and the priority is and the priority is and the priority is an index and the priority is and the priority is an index and the priority is and the priority is and the comparison of the priority is and the priority is an index and the priority is and the priority priority of the comparison of the priority is and the priority of the end of the priority is and the priority of the priority of the end of the priority is and the priority of the end of the priority is and the priority of the end of the priority of the end of the priority is and the priority of the end of the priority of the end of the priority is and the priority of the end of the priority of the end of the priority is and the priority of the end of of the e
Agende Organisation for the forst of the	
Histofficial Negal partice trading a month A many physics construction deliverability with billing (1, 5, and in monements 4, 4, 4 and physics trading (1, 5, and in monements 4, 4, 4 and physics trading (1, 6) in monements (1, 1) and physics trading (1, 1) in monements (1, 1) and physics trading (1, 1) and (1, 1) approach (1, 1) are particly band as the	configurated findings that the 0 set has been in such trace distributive values of the trace deplete and source compared prior branched being that the 0.000 states are used of tragging in the 0.000 states are used of tragging in the trace of the trace of the trace of the based of the trace of the trace of the trace of the trace of prior trace of the trace of the trace of the trace of prior trace of the trace of the trace of the trace of prior trace of the trace of the trace of the trace of prior trace of the trace of the trace of the trace of the trace of the trace of the trace of the trace of the trace of the trace of the trace of the trace of the trace of the trace of the trace of the trace of the trace of the trac
	addresses are being deathing hard a p
	Second St. Andreast. Economic day tooling with a strain tool of Advant 40 Access of Second Strains
Al (C) A present LB (C) Preside () an other () a sparser () A () A preserving () A preserving () A () A () A () A preserving () A () A preserving () A	Seering OD, Andreger, Karnerer A, Sarperstein, et M. J. (K. 40, 5, 4). In Article and the seering sequence of the control of the CM in the second proposalization field in a spectrum second and the second second second test of a spectrum second second second second second test of a spectrum second second second second second test of a spectrum second seco

Rai Ray, Lin (1997), R. 18 Ray, Ann. (2017), Thingson (Rev. (2017), R. 19)	Tax Discovery of
Apte ANI: Episodelipies	
MLPF) efficient machine-borroof pe ming graph record activities	rtide-flow reconstruction
Josep Fold **, John Fasch * John Rad Wessel */ **Ander Schutz ** Schutz * Schutz ** Schutz ** Schutz ** Schutz **Schutz ** Schutz ** Schutz ** Schutz ** Schutz ** **Schutz ** Schutz ** Schutz ** **Schutz ** Schutz ** **Schutz **	Ausen Aufer (Ause Ausen) April Bir Max Ausen
NAME OF TAXABLE PARTY.	

San Alam J. C. (2000 million Republic operation and an analysis)	THE BUILDENA C
Pagin tech Caprison Rate	
Performance of a geometric deep he particle tracking	rateg pipeline for HE-12HC
$\label{eq:second} \begin{array}{l} \label{eq:second} \left\{ \begin{array}{l} \mbox{second} & secon$	Bolder (Kasal) Rat (Solar) Rata Ratal) in Anna (Kasal) Rata (Rata) Rata (Kasal) (Kasal) Rata (Kasal) Rata (Kasal) (Kasal) (Kasal) (Kasal) Rata (Kasal) (Kasal) (Kasal) (Kasal) (Kasal) Rata (Kasal)
Roman Laboration Acception in Appendix ACC	
NIMI: In COLUMN (1) In card particular interpretation of the second s	A set of the set of th
A A A A A A A A A A A A A A A A A A A	

PHILOSOPHICAL	Output to the output to the output to the
TRANSACTIONS A	showith my fee channel marticle
subscrapelliding agreements	tracking
	tracking .
G 🗶	8.M.Gaj ¹⁷
leview 🛈 🚠	Nysia Jopateuri, Johnsky of Lehteria, Belefey,
a state in the second second	9.908.24
the provine dealing to head	"Aper, Serona, Lawren Bolking Showing, " Show Wall
FOR Marines, Mar. Autor. 1 Stat. 1 (1994)	Address screening and
rents.	0 186,180-002 28 4%
dip./blog/2.09016,30100	THE REPORT OF A DESCRIPTION OF A DESCRIP
	and such below a right a praced contracts
kond Separate M	second Babbanesty Loss Tables Califier
upted 75 miler XD	in the next decide and close more so at better colliders. A har deploying in the second-points of
a subdate di ta share bar	second of developed data and collector data is the
one is here is which they	polices acception algorithms and in-following the
	togetime archarged particles. The fall of grantees
April Boost	considers and it many firmula a rade of sould
INTERNAL ADDRESS OF ADDRESS	downlopment and hours might provide a solution to
have been a second s	of standard sciencilies for chernel surface automation
when separting further methods	morphisms Aphanargo physics
don acquite, partie raches	This article is part of the thoma issue Quantum
and a	annenges a parts privat.
who for an expendence	
i i i i i i	1. Introduction
and suffer productions	The high-energy physics coparisons at the long Radius Callaber (2011) incomed part conside Gaussia Inclusional have incomented for held of particle
	physics, must makely through the multing discovery
	of the Higgs bases (1.2 by He #1.45-02 and CM
	of operation when and an or other the
	By Anarting comparing definings of the long-Annerty
	equiled him the upperts of the LHC, the High
	homosty LW, (H, (H) is the set deals an Anny solution in Julies the 10.100). The physics
	7 King and a start have been as a second s
	Casta Castan Aldala Long Harlandarana arterat
	lighth, which percent annumber on product the segment active and
THE ROTAL SOCIETY	test a total

The TrackML shortcomings



- 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

Data issue (funny tracks) explained



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
- $|\eta| < 3$





Higher level of detail



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

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

5/16

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 $|\eta| < 3$.

6

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, \ldots 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

2 🔒 🐲 https://open-data-detector-phoenix.web.cern.ch

PH₩ΣNIX

https://open-data-detector-phoenix.web.cern.ch/



10/16

How to use ODD



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



- 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 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.



12/16

- 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:
 - $\circ~$ will replace the TrackML for track reconstruction studies
 - $\circ~$ calorimetry can be used for fast shower simulation studies, as well as calorimetry reconstruction
 - $\circ~$ will offer a possibility to use combined information from tracker-calo(-muon), e.g. for particle-flow studies.



12/16

ODD dataset for fast simulation

• Different particles: e^{\pm} , γ , π^{\pm} .

ODD dataset for fast simulation



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

ODD dataset for fast simulation



- Different particles: e^{\pm} , γ , π^{\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	$59\pm$ 8	111 ± 16	409 ± 80
10	382 ± 21	$993\pm$ 48	4210 ± 266
100	2112 ± 58	5560 ± 189	42073 ± 800
1000	9410 ± 186	18796 ± 405	417317 ± 3835



 * voxelization with 0.5 $\mathrm{X}_0\!\times\!0.25~\mathrm{R}_M\!\times\!0.125$ 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)
I	

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

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