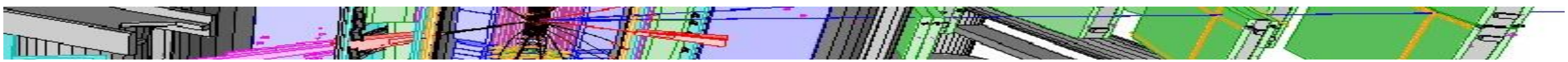
TrackML : the roller coaster of organizing a HEP challenge on Kaggle and Codalab -



David Rousseau

LAL-Orsay

rousseau@lal.in2p3.fr @dhpmrou

EPS-HEP Conference 10-17 Jul 2019, Gand, BElgique

Who and How

- Organisation: Jean-Roch Vlimant (Caltech), Vincenzo Innocente, Andreas Salzburger (CERN), Sabrina Amrouche, Tobias Golling, Moritz Kiehn (Geneva University), David Rousseau, Yetkin Yilmaz (LAL-Orsay), Paolo Calafiura, Steven Farrell, Heather Gray (LBNL), Vladimir Vava Gligorov (LPNHE-Paris), Laurent Basara, Cécile Germain, Isabelle Guyon, Victor Estrade (LRI-Orsay), Edward Moyse (University of Massachussets), Mikhail Hushchyn, Andrey Ustyuzhanin (Yandex, HSE)

5-6 FTE year

Platforms: kaggle Phase 1 | CodaLab Phase 2

accuracy

throughput

TrackML sponsors

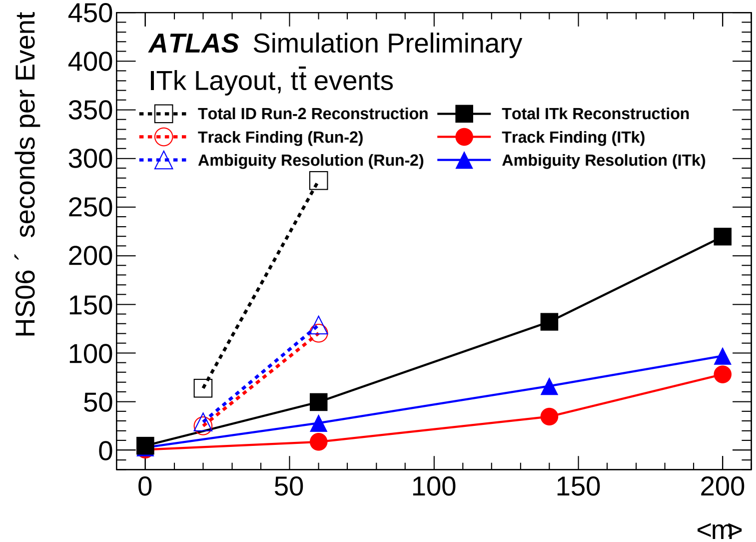
CERN
kaggle
NVIDIA
UNIVERSITÉ DE GENÈVE
Paris-Saclay Center for Data Science
INSTITUTE DATAIA Data Science, Intelligence & Society
CHA LEARN
iris hep Institute for Research & Innovation in Software for High Energy Physics
CERN openlab
Inria INVENTORS FOR THE DIGITAL WORLD
PyTorch
cnrs
erc
Fermilab
COMMON GROUND
inqnet Alliance for Quantum Technologies
UNIVERSITÉ PARIS SUD
université PARIS-SACLAY

Tracking crisis

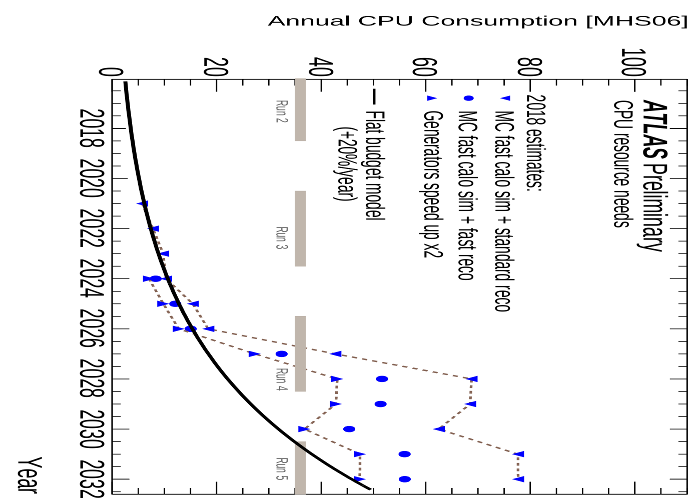
- Tracking (in particular pattern recognition) dominates reconstruction CPU time at LHC
- HL-LHC (phase 2) perspective : increased pileup :Run 1 (2012): <>~20, Run 2 (2015): <>~50,Phase 2 (2025): <>~200
- CPU time quadratic/exponential extrapolation

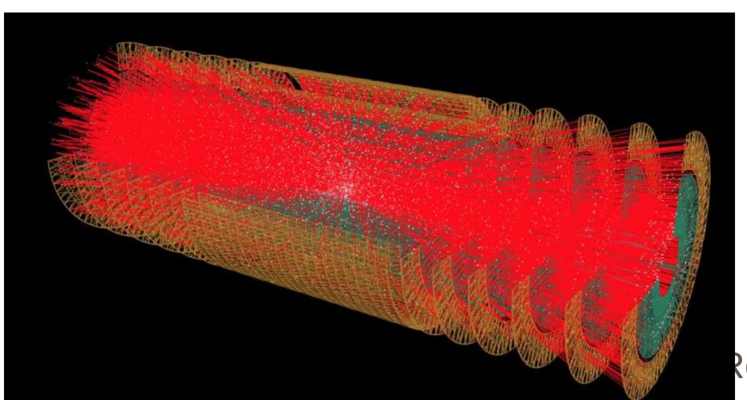
- On-going Large effort within HEP to optimise software and tackle micro and macro parallelism.
- >20 years of LHC tracking development. Everything has been tried?
  - Maybe yes, but maybe algorithm slower at low lumi but with a better scaling have been dismissed ?
  - Maybe no, brand new ideas from ML

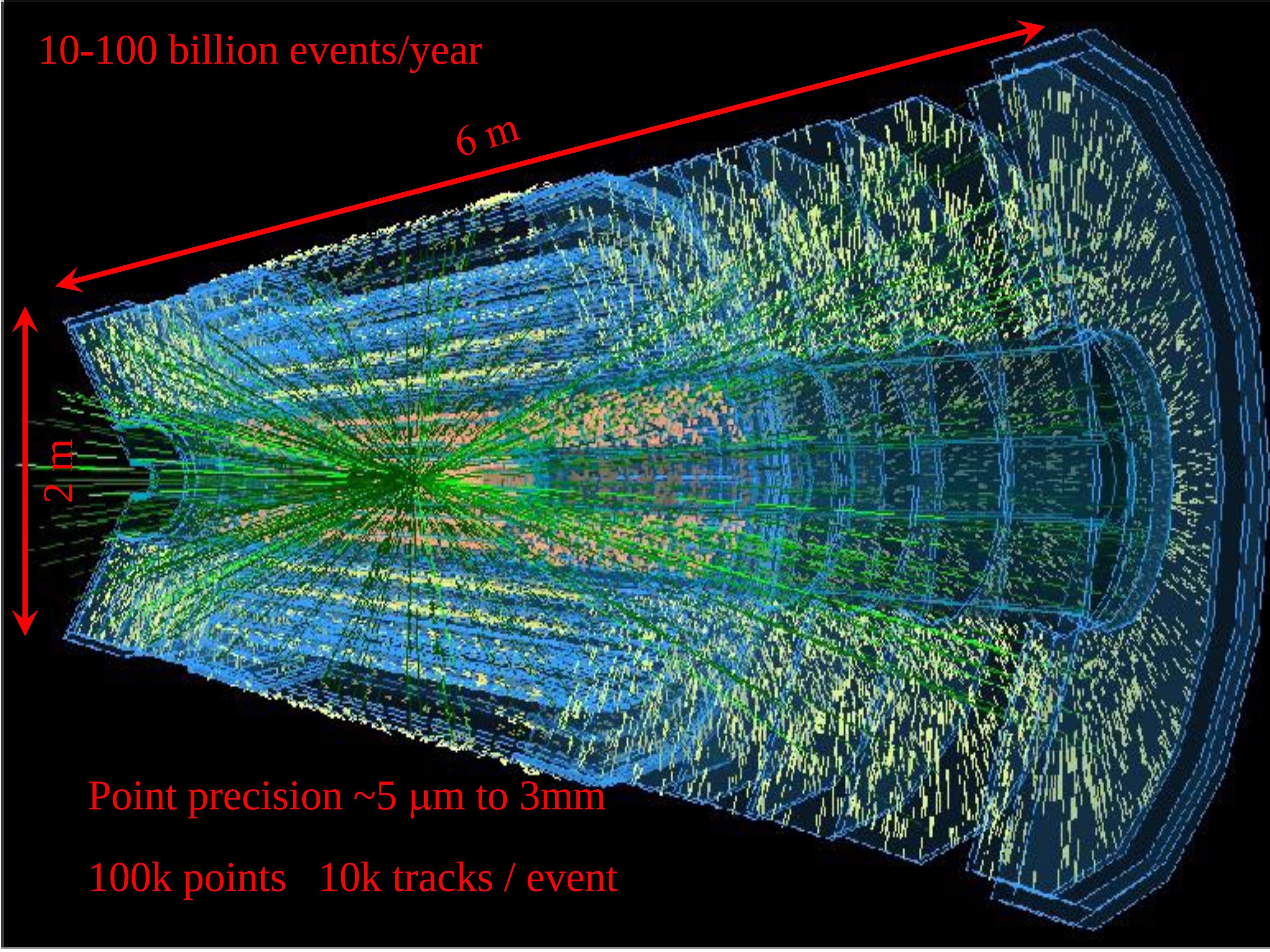
- ➔challenge TrackML !!



Similar plots from CMS

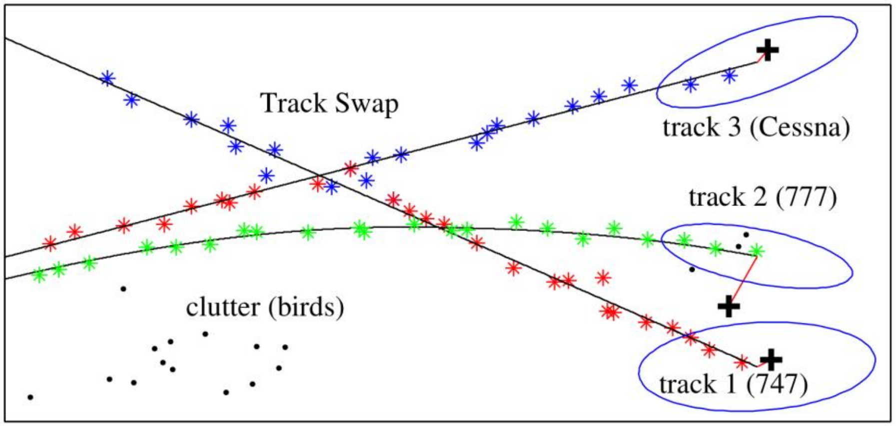


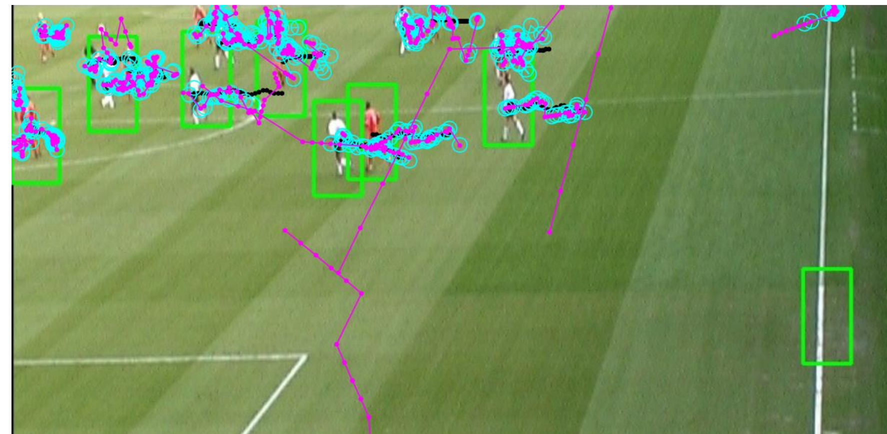


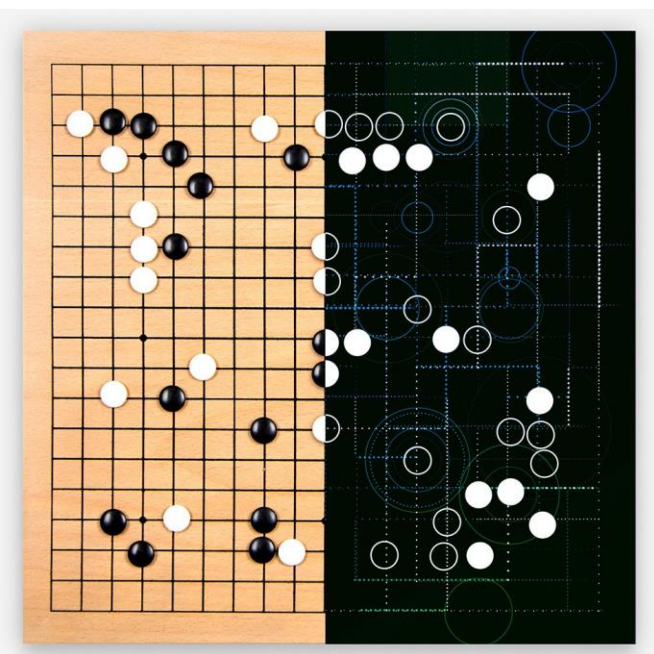

10-100 billion events/year
6 m
2 m
Point precision ~5 µm to 3mm
100k points 10k tracks / event

Tracking outside HEP

- …is very different



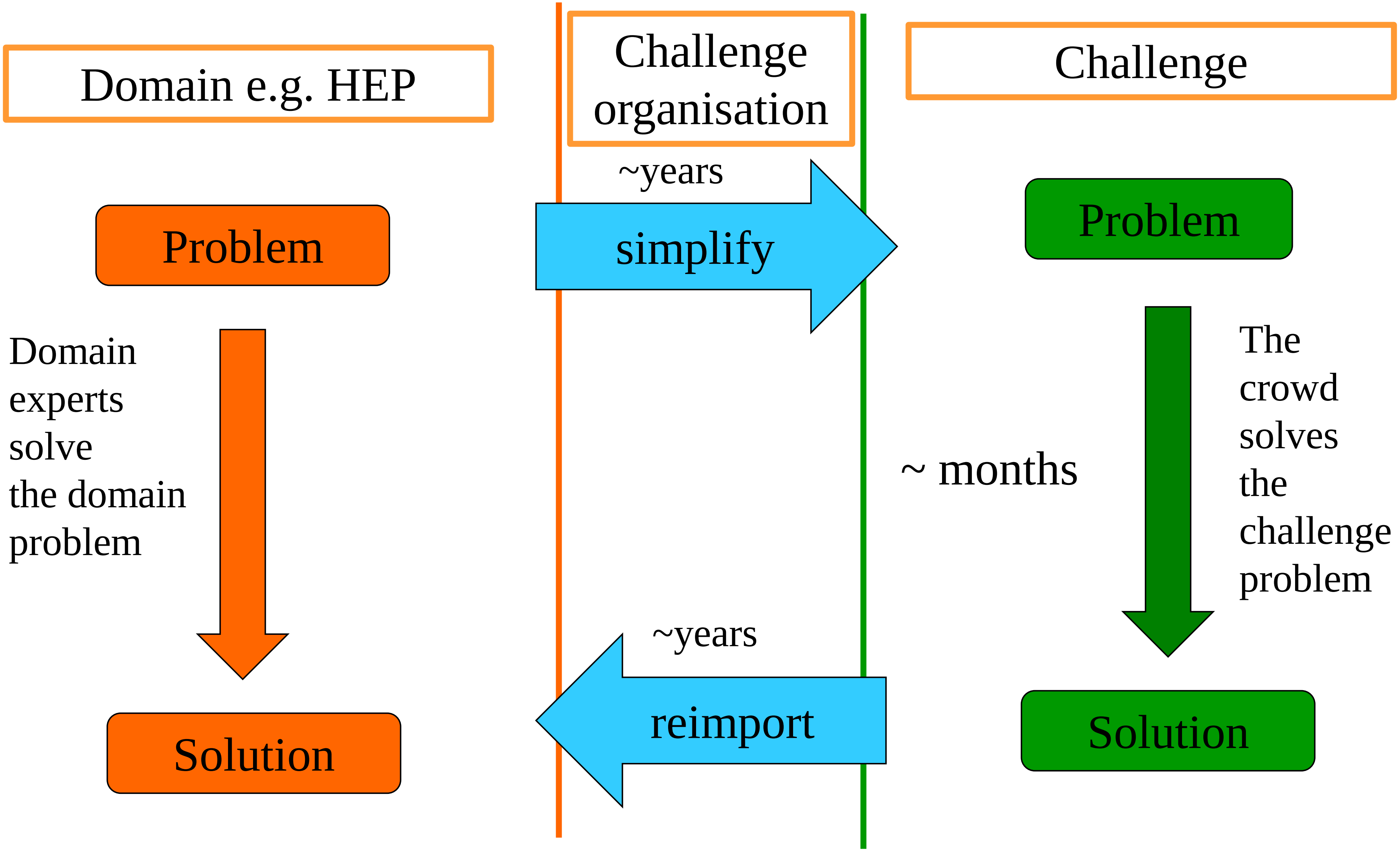




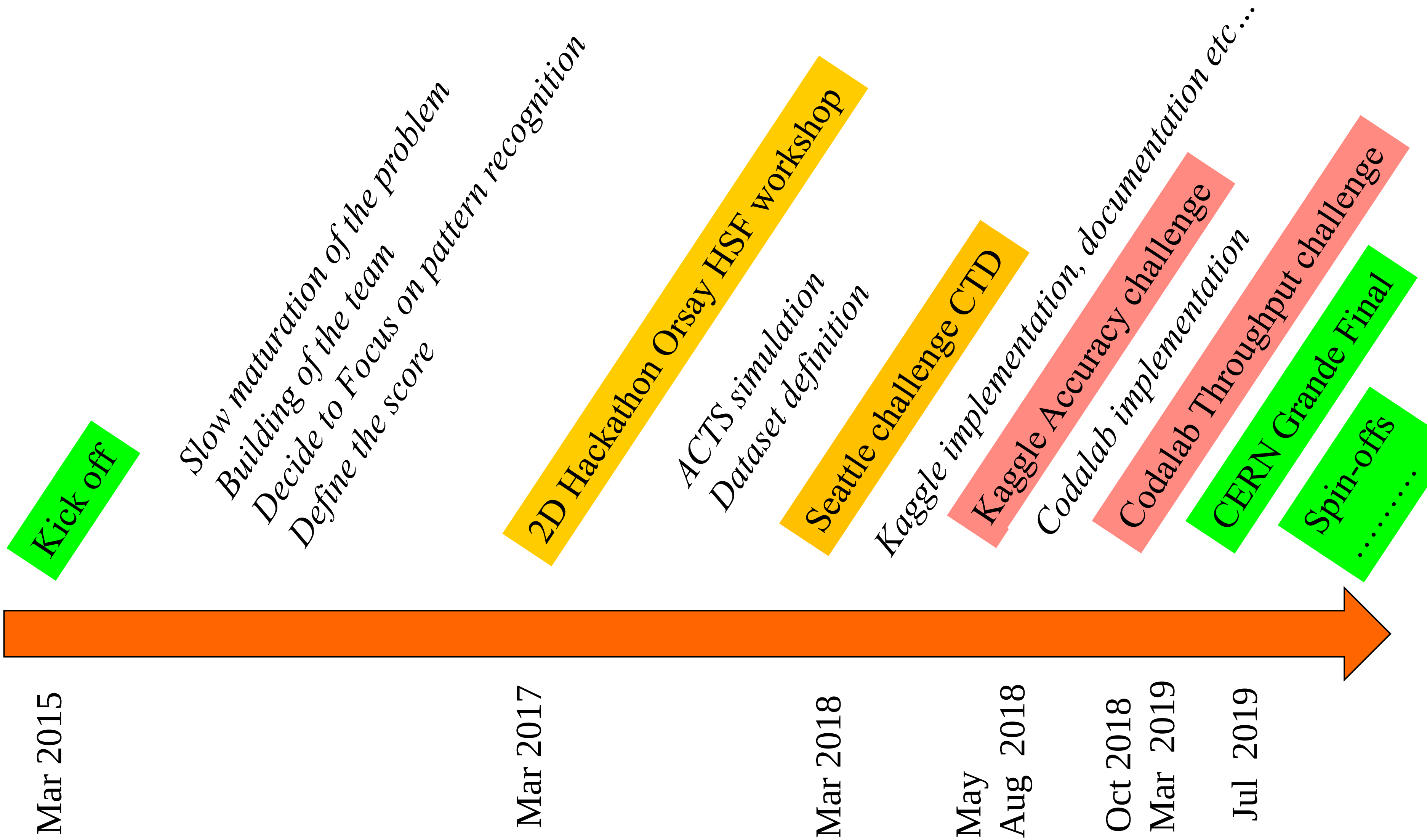
TrackML in a nutshell

- Accurate simulation engine (ACTS https://gitlab.cern.ch/acts/acts-core) to produce realistic events
  - Ttbar events with 200 pileup
  - Silicon detector with barrels and disks (simplified HL-LHC ATLAS or CMS Si detector)
  - One file with list of 3D points
  - Ground truth : one file with point to particle association
  - Ground truth auxiliary : true particle parameter (origin, direction, curvature)
  - Typical events with ~200 parasitic collisions (~10.000 tracks/event)
- Large training sample 10k events, 0.1 billion tracks, 1 billion points, ~100GByte
- Accuracy phase (May to August 2018) on Kaggle
  - Participants are given the test sample (with usual split for public and private leaderboard) and run the evaluation to find the tracks
  - They should upload the tracks they have found
  - A track is a list of 3D points
  - Score : fraction of points correctly grouped together
  - Evaluation on test sample with per-mille precision on 100 event
- Throughput phase Sep to Mar 2019 on Codalab
  - Participants submit their code to solve the same probmem
  - Strong CPU incentive

From domain to challenge and back

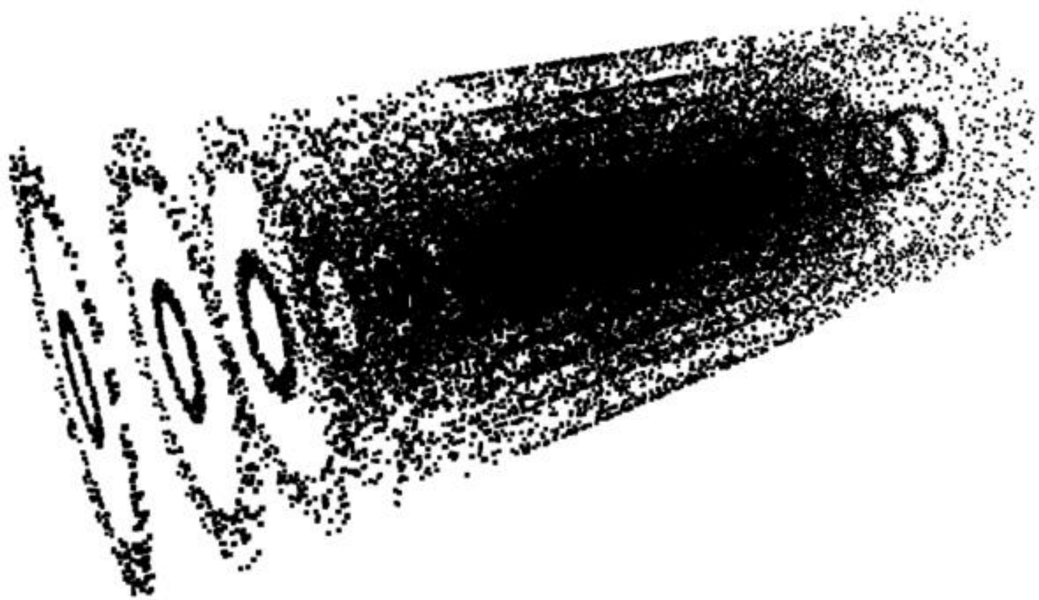


TrackML timeline



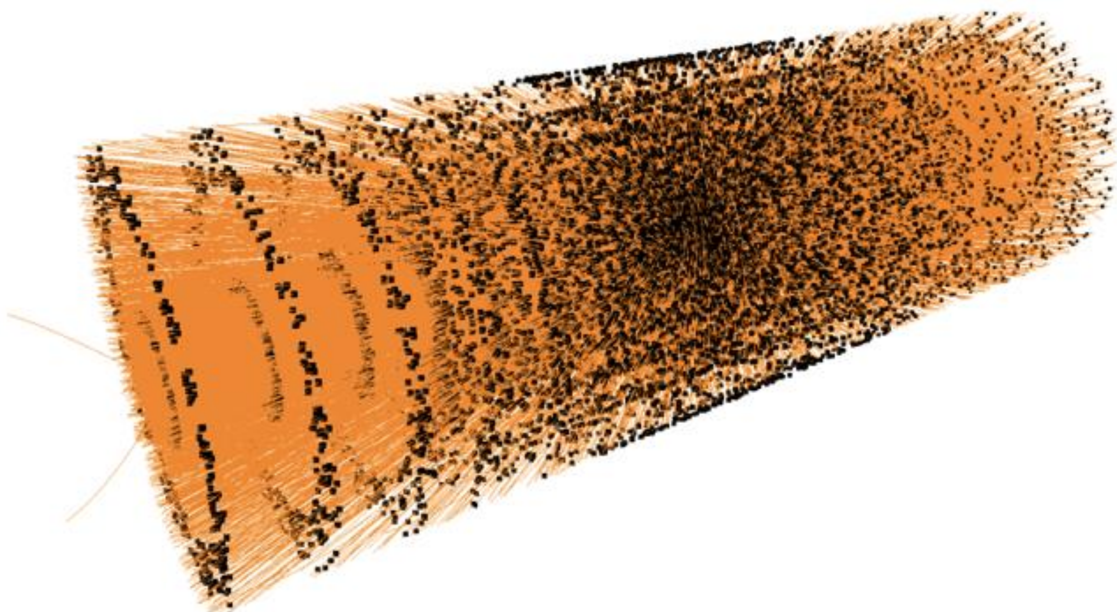
Dataset

3D points



Dataset

3D points ➔ tracks

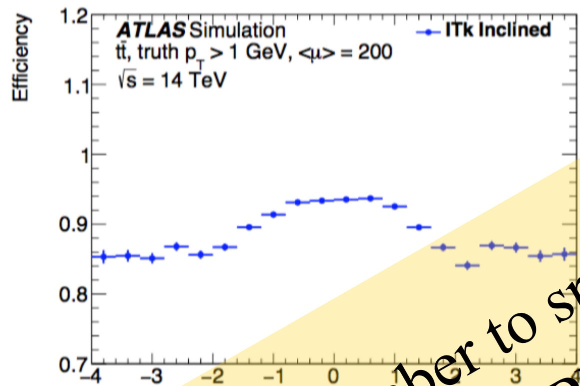


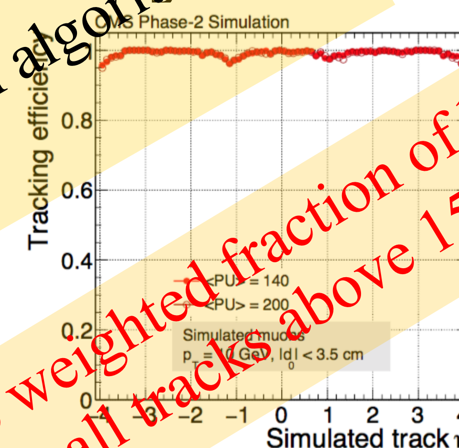
Score

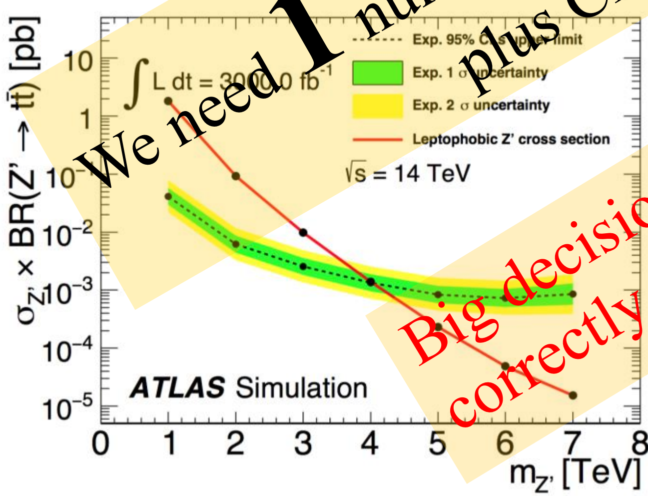
- [2017 CMS tracker Technical Design Report](#) : Chapter 6 expected performance 31 pages 58 figures
- [ATLAS Si strip Technical Design Report](#) Chapter 4 ITk Performance and Physics Benchmark Studies 54 pages 80 figures

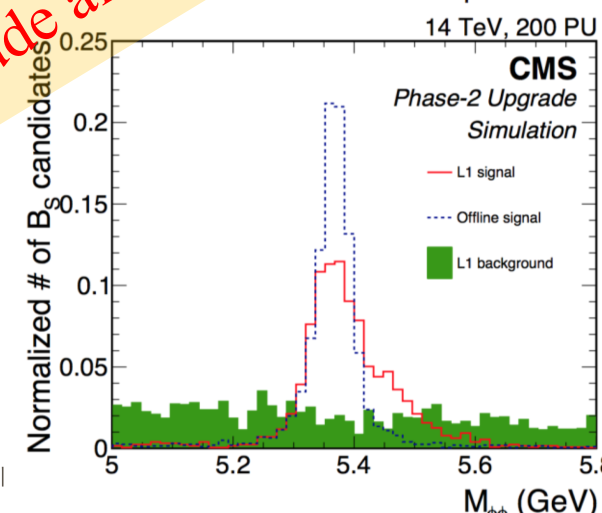
We need **1** number to specify how good an algorithm is!
plus CPU time

Big decision : score is ~ « the weighted fraction of hits correctly associated ». Include all tracks above 150MeV









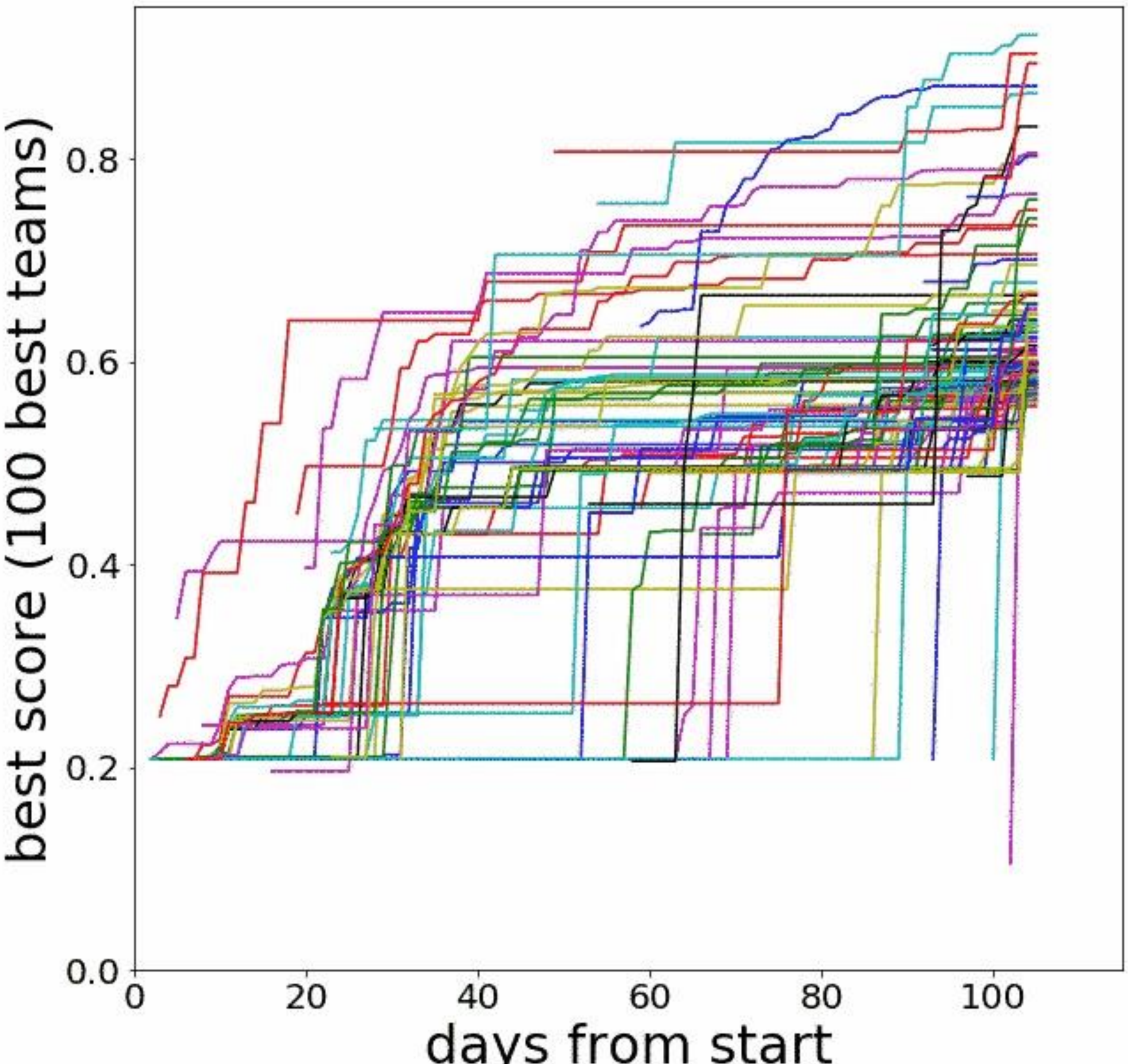
Real life vs challenge

1. Wide type of physics events
2. Full detailed Geant 4 / data
3. Detailed dead matter description
4. Complex geometry (tilted modules, double layers, misalignments…)
5. Hit merging
6. Allow shared hits
7. Output is hit clustering, track parameter and covariance matrix
8. Multiple metrics (see TDR's)

1. One event type (ttbar)
2. ACTS (MS, energy loss, hadronic interaction, solenoidal magnetic field, inefficiency)
3. Cylinders and slabs
4. Simple, ideal, geometry (cylinders and disks)
5. No hit merging
6. Disallow shared hits
7. Output is hit clustering
8. Single number metrics

Simpler, but not too simple!

Evolution of leaderboard



Final Leaderboard

1	—	Top Quarks	0.92182	10	19d
2	—	outrunner	0.90302	9	18d
3	HEP	Sergey Gorbunov	0.89353	6	18d
4	HEP	demelian	0.87079	35	1mo
5	—	Edwin Steiner	0.86395	5	18d
6	—	Komaki	0.83127	22	18d
7	—	Yuval & Trian	0.80414	56	18d
8	—	bestfitting	0.80341	6	18d
9	—	DBSCAN forever	0.80114	23	18d
10	—	Zidmie & KhaVo	0.76320	26	18d
11	—	Andrea Lonza	0.75845	15	18d
12	—	Finnies	0.74827	56	18d
13	—	Rei Matsuzaki	0.74035	12	18d
14	—	Mickey	0.73217	10	2mo
15	—	Vicens Gaitan	0.70429	19	1mo
16	—	Robert	0.69955	3	21d
17	—	Yuval-CPMP tribute band	0.69364	20	20d
18	—	N. Hi. Bouzu	0.67573	9	22d
19	—	Steins;Gate	0.66763	12	19d
20	▲1	Victor Nedel'ko	0.66723	4	2mo

Only public LB to private LB rank change →

Experience with first phase

- 630 participants
- Some only downloaded provided solutions, but >100 did provide original code (or tuning of existing code)
- Lots of exchange on the forum
  - People googling courses on HEP tracking…
  - Exchanging ideas, and even code…
  - …up to a certain point (score <=50%)
- A variety of algorithms with various role for ML

e.g. Participant Data Analysis

We provided a data visualisation notebook: but participants did much better within two days:

See link

Data Exploration:

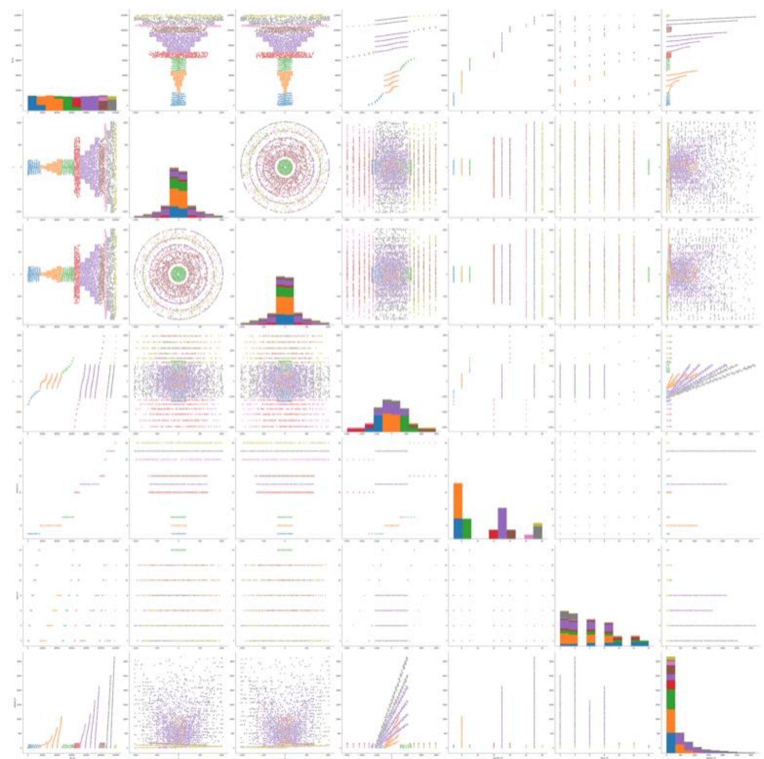
Data

Comments

The (`hit_id` , `x`) and (`hit_id` , `y`) pairs show us how different volumes are layered.

This figure idea is taken from Joshua Bonatt's notebook.

Code

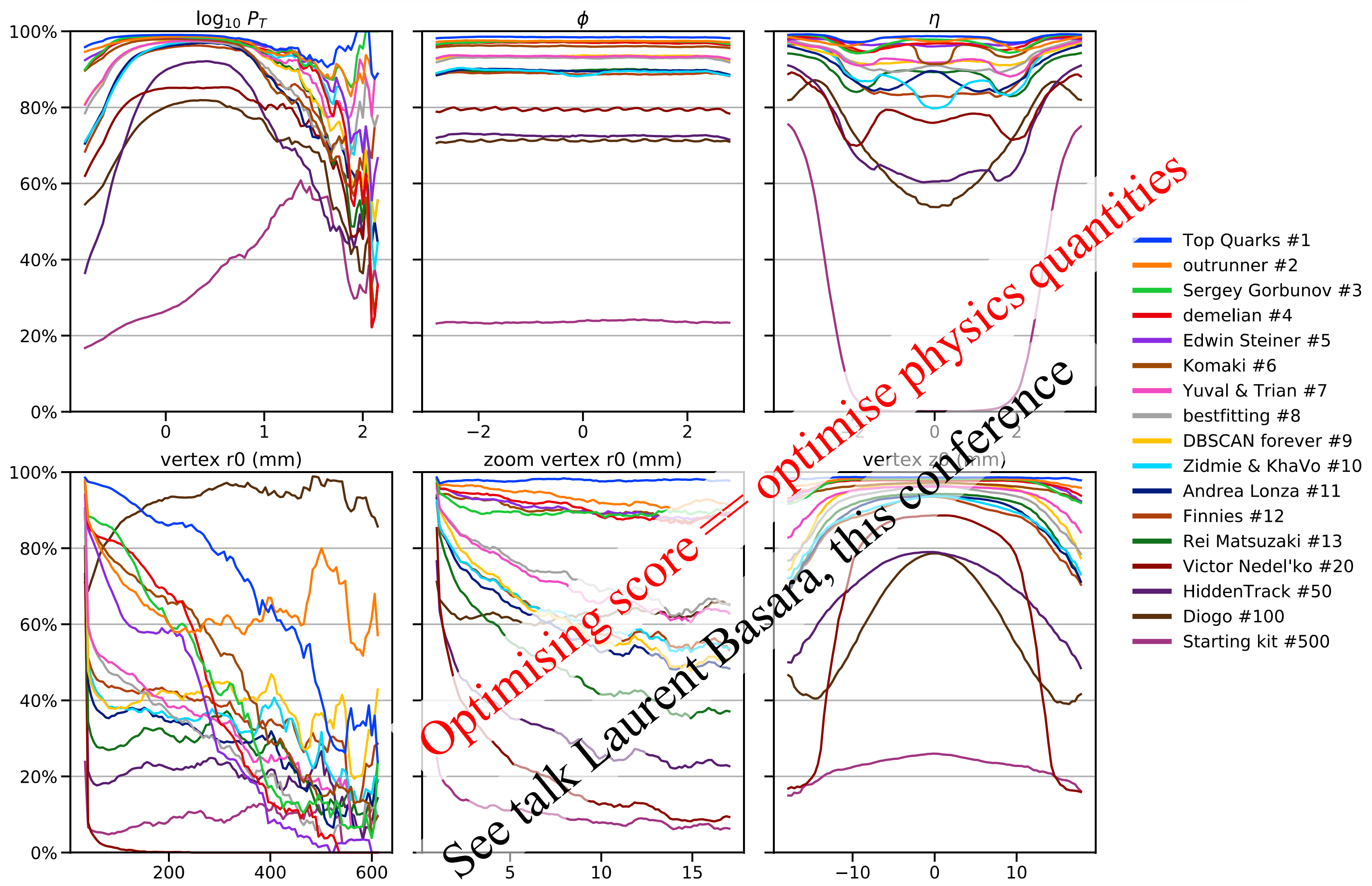


Particle Data

The particle data help us understand each particle's initial position, momentum, and charge, which we can join with the event truth data set to get the particle's final position and momentum. This is needed to identify the tracks that each particle generated.

The data look like this:

Efficiency all



Throughput Phase

Now participants submit their software…
… and are evaluated on accuracy AND speed !

Launched 6th Sep 2018 until 12th March 2019 on Codalab

Throughput platform

- Kaggle initially told us they would also provide the speed estimate…
- …but they suddenly declined.
- …so we did it ourself on Codalab, with U Paris-Sud resources.
- Specific difficulties:
  - Speed measurement reproducibility no better than 3% (even on dedicated machine)
  - Many hacks anticipated (e.g. dumping the data in the log file…)
  - More hacks for sure…
- ➔decision : remeasure speed at the end of the competition many times on a dedicated machine
  - ➔it worked
- Providing for competition with accurate online time measurement is an open problem (Kaggle is working on it, given the demand, see e.g. « the Airbus Ship Detection challenge »)

Throughput phase LB

RESULTS

Private leaderboard

#	User	Entries	Date of Last Entry	score	accuracy_mean	accuracy_std	computation time (sec)	computation speed (sec/event)	Duration
1	sgorbuno HEP	9	03/12/19	1.1727 (1) 1.16	0.944 (2) 0.943	0.00 (14)	28.06 (1)	0.56 (1) 0.60	64.00 (1)
2	fastrack HEP	53	03/12/19	1.1145 (2) 1.12	0.944 (1) 0.944	0.00 (15)	55.51 (16)	1.11 (16) 1.00	91.00 (6)
3	cloudkitchen ~HEP	73	03/12/19	0.9007 (3) 0.897	0.928 (3) 0.927	0.00 (13)	364.00 (18)	7.28 (18) 7.41	407.00 (8)
4	cubus	8	09/13/18	0.7719 (4) 0.770	0.895 (4) 0.895	0.01 (9)	675.35 (19)	13.51 (19) 13.7	724.00 (9)
5	Taka	11	01/13/19	0.5930 (5)	0.875 (5)	0.01 (12)	2668.50 (23)	53.37 (23)	2758.00 (13)
6	Vicennial	27	02/24/19	0.5634 (6)	0.815 (6)	0.01 (10)	1270.73 (20)	25.41 (20)	1339.00 (10)
7	Sharad	57	03/10/19	0.2918 (7)	0.674 (7)	0.02 (4)	1902.20 (22)	38.04 (22)	1986.00 (12)
8	WeizmannAI	5	03/12/19	0.0000 (8)	0.133 (11)	0.01 (11)	88.08 (17)	1.76 (17)	124.00 (7)
9	harshakoundinya	2	03/12/19	0.0000 (8)	0.085 (13)	0.01 (6)	49.22 (8)	0.98 (8)	86.00 (3)
10	iWit	6	03/10/19	0.0000 (8)	0.082 (15)	0.01 (8)	48.23 (3)	0.96 (3)	85.00 (2)
				0.0000					

Where did ML people go ?

- 100 participants registered on Codalab but only 10 submitted non trivial code. Why ? Our guesses:
  - Kaggle visibility vs Codalab visibility.
  - On Kaggle people win points across competition, can access « Grand master status », etc… very valuable on their CV
  - « Professional » kagglers move from one challenge to the next. No interest in long term involvement
    - (still we had some praises like « most interesting challenge I had ever done »)
  - Codalab is a research platform
  - No GPU (while ML code « naturally » run on GPU)
  - C++ vs python : python was allowed but people realise they had to write in C++ for speed. Many ML people do not know C++
  - Not completely trivial effort to properly wrap code for submission

HEP wins at the end

- The podium are HEP experts. Was it worth it ?
- Definitely : best solutions in <1 s to be compared to >10 s for ATLAS or CMS (order of magnitude comparison)
- HEP people liked the gamification of the problem.
  - Also one is ALICE, one is ATLAS, one is Computing Center management.
- The dataset will be released on CERN Open Data portal for future development
  - Already used in research papers e.g. tracking with quantum computing (see talk in CERN Grand Finale workshop)
- On going work to integrate the best ideas (of both phases) in future algorithms for ATLAS and CMS

Visualisation spin-off

- Visit at CERN Tobias Isenberg visualisation scientist at LRI-Orsay with PhD student Xiyao Wang
- They are using TrackML dataset to experiment with visualisation/interaction with Microsoft’ Hololens (see talk in CERN Grand Finale workshop)

TrackML Conference talks

- Connecting The Dots 2015 Seattle
- Connecting The Dots 2016 Vienna
- CHEP 2016 Okinawa
- Connecting The Dots / Intelligent Trackers 2017 Orsay
- NeurIPS 2017 Los Angeles CiML workshop
- Connecting The Dots 2018 Seattle
- CHEP 2018 Sofia
- WCCI 2018 Rio de Janeiro
- ICHEP 2018 Seoul
- IEEE NSSMIC 2018 Sidney
- IEEE eScience 2018 Amsterdam
- NeurIPS 2018 Montreal Competition workshop
- ACAT 2019 Saas-Fe
- Connecting The Dots 2019 Valencia
- EPS 2019 Ghent
- CHEP 2019 Adelaïde
- …and much more workshops and seminars….

Useful links

- See also Laurent Basara's talk in Detector and Data Handling session Friday 12:45, about the algorithms exposed
- Contact : trackml.contact@gmail.com https://sites.google.com/site/trackmlparticle Twitter : @trackmllhc
- Accuracy phase @ Kaggle : https://www.kaggle.com/c/trackml-particle-identification
  - ➔chapter in the NeurIPS 2018 Competition book arXiv:1904.06778 final version just released
- Throughput phase @ Codalab : https://competitions.codalab.org/competitions/20112
  - Write-up being finalized
- CERN Grand Finale workshop 1-2 Jul 2019 : https://indico.cern.ch/event/813759/