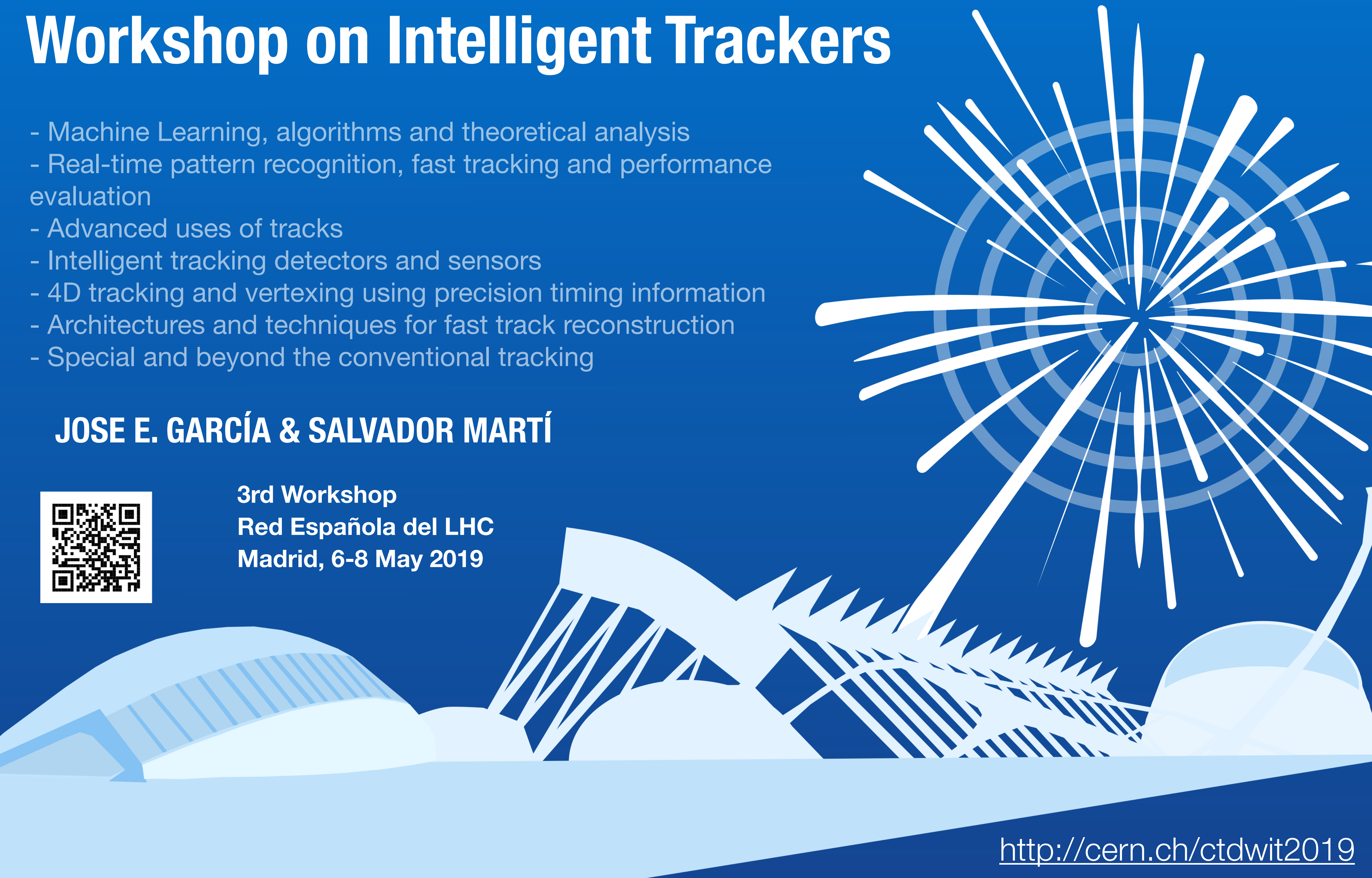
Connecting The Dots AND Workshop on Intelligent Trackers

- Machine Learning, algorithms and theoretical analysis
- Real-time pattern recognition, fast tracking and performance evaluation
- Advanced uses of tracks
- Intelligent tracking detectors and sensors
- 4D tracking and vertexing using precision timing information
- Architectures and techniques for fast track reconstruction
- Special and beyond the conventional tracking

JOSE E. GARCÍA & SALVADOR MARTÍ

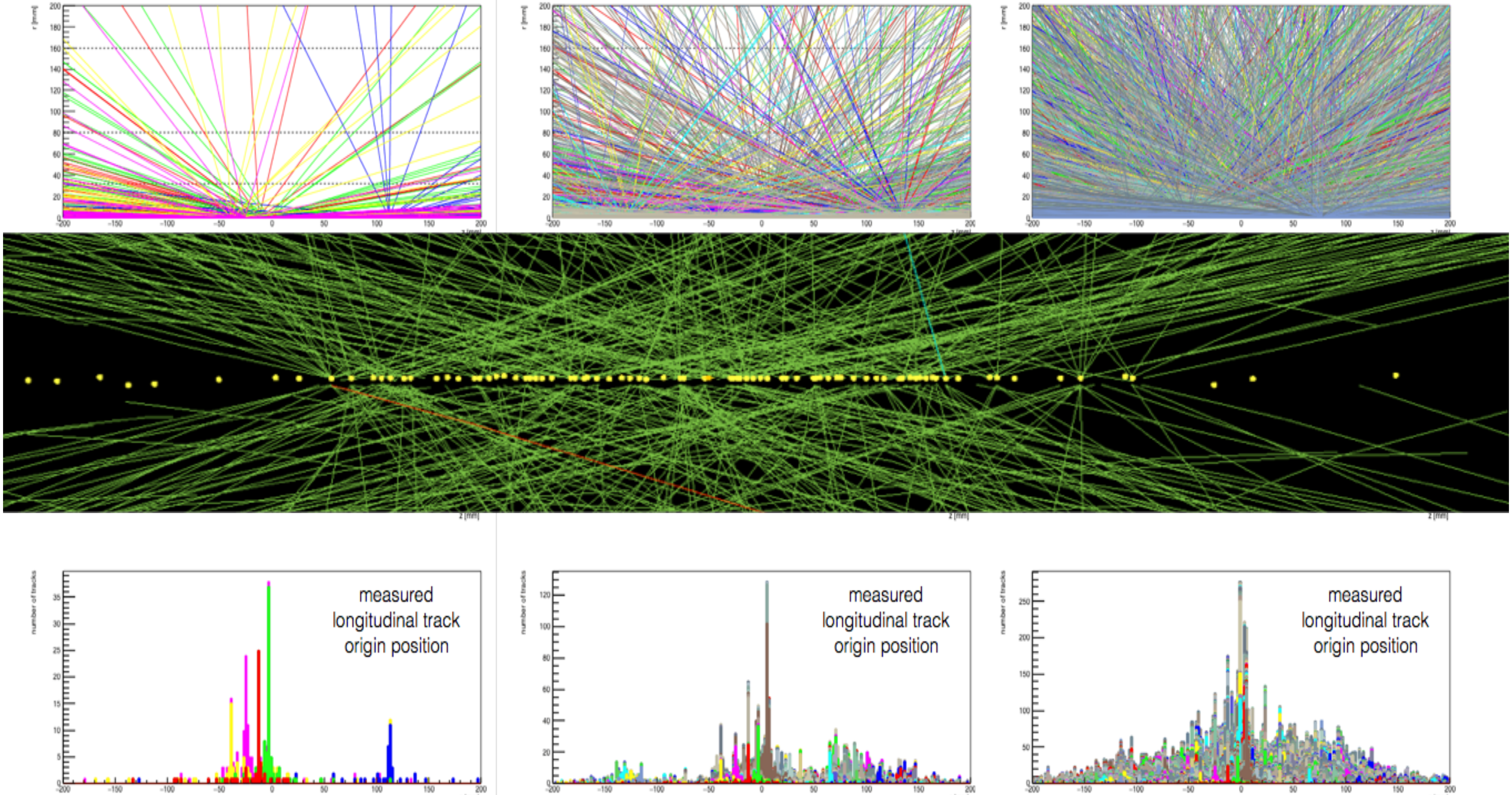
3rd Workshop
Red Española del LHC
Madrid, 6-8 May 2019



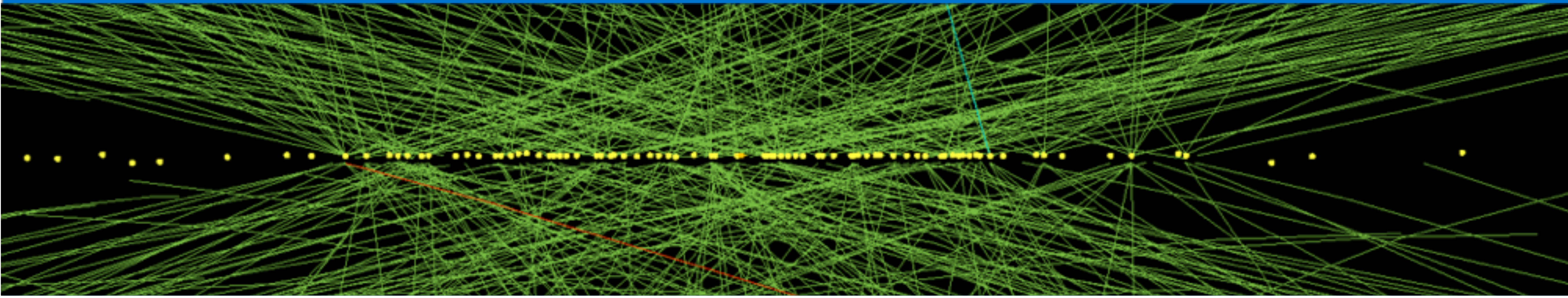
http://cern.ch/ctdwit2019

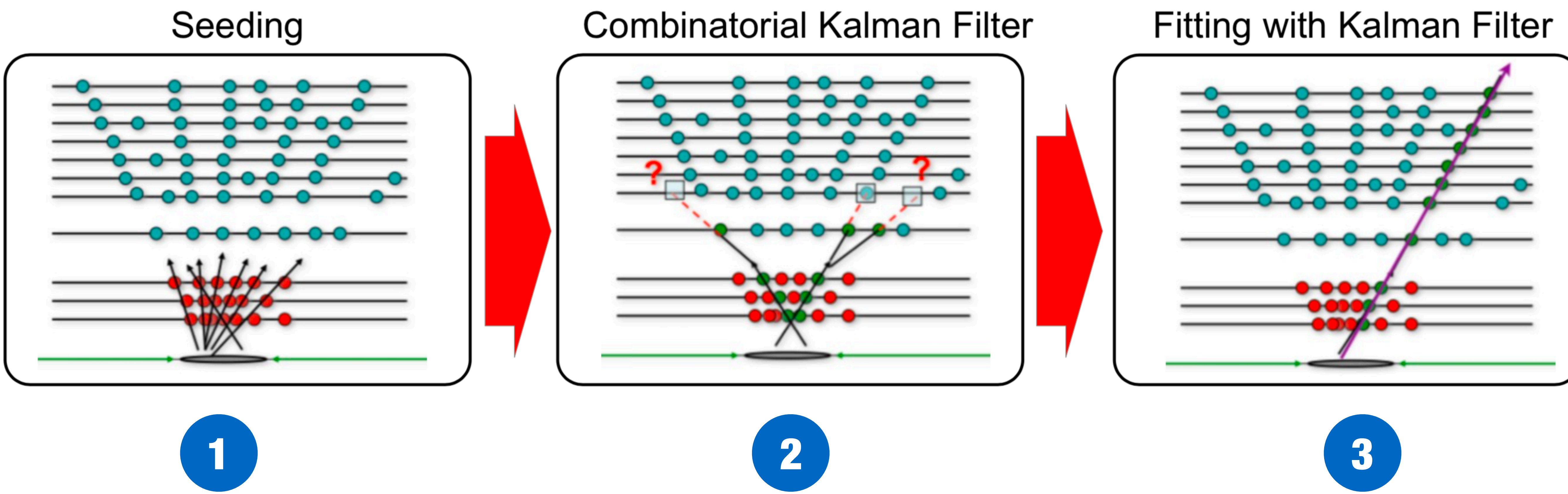
PILE-UP

5 p-p collisions	40 p-p collisions	200 p-p collisions
LHC early Run-1 2010	LHC early Run-2 2015/16	HL-LHC conditions



TRACKING

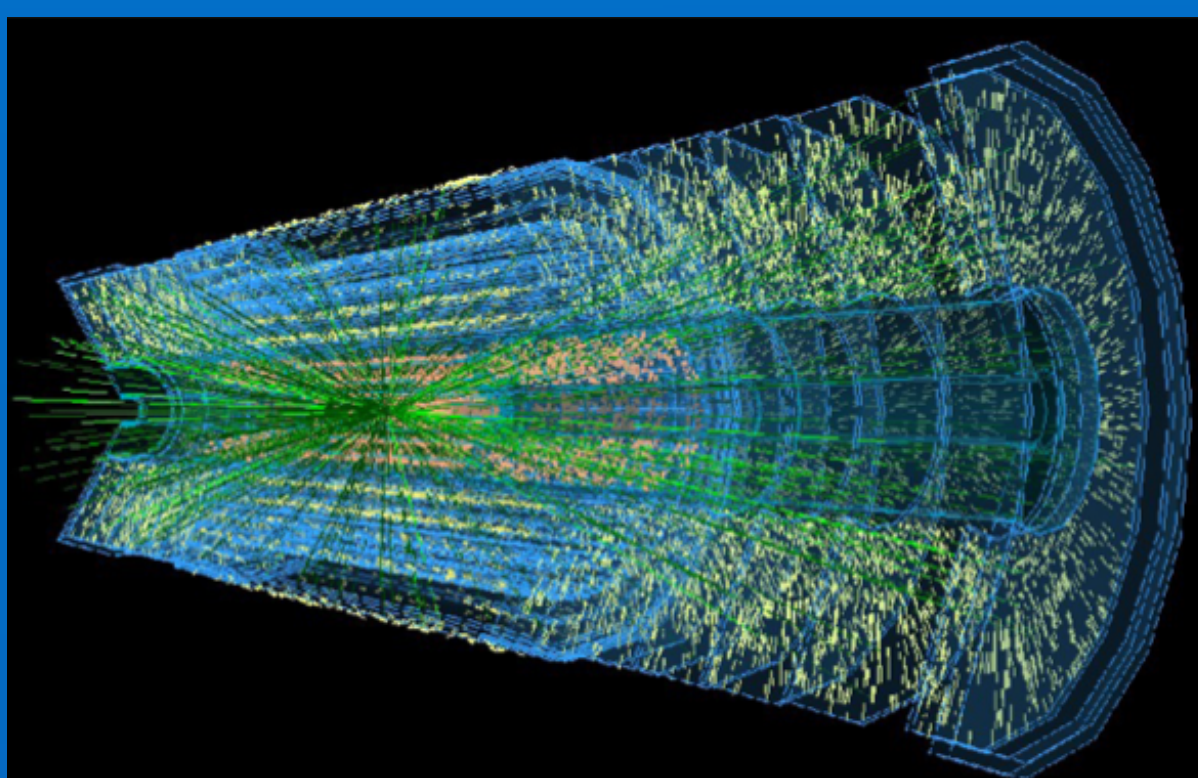


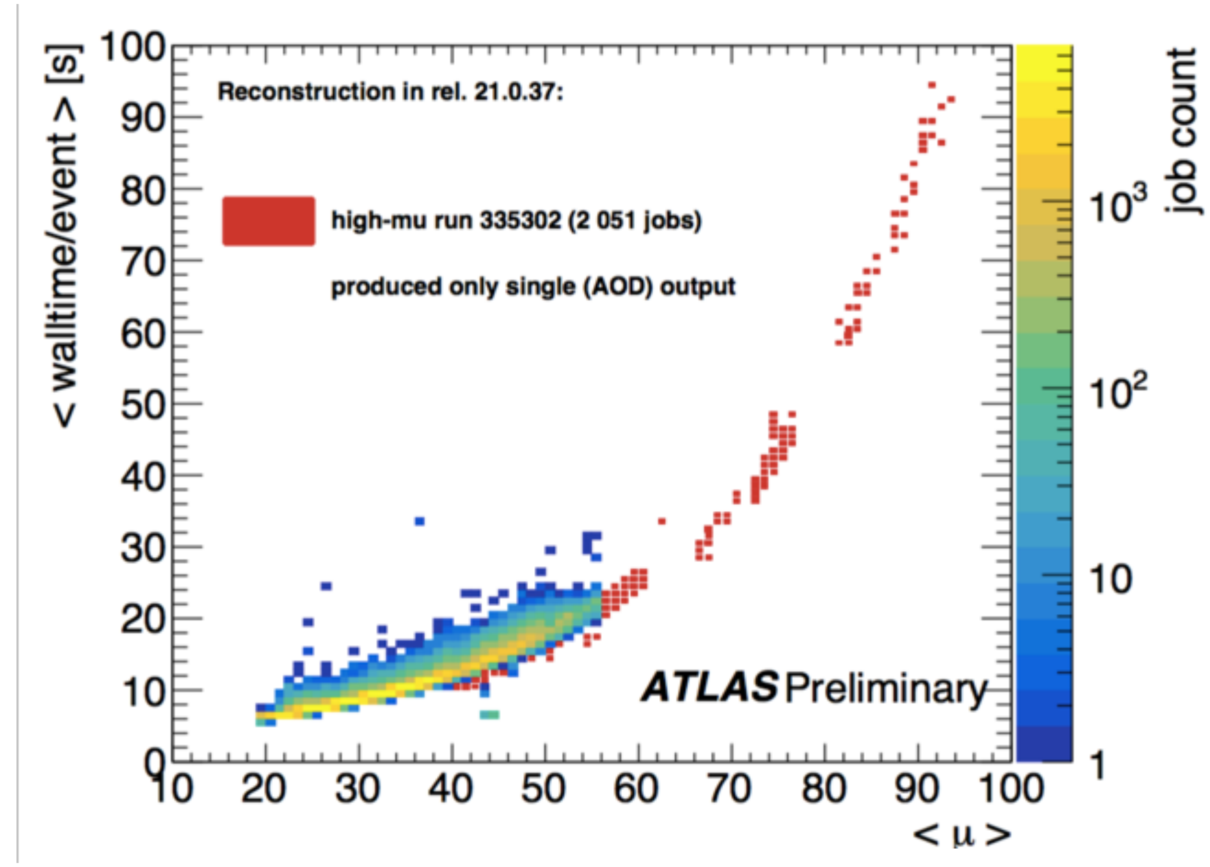


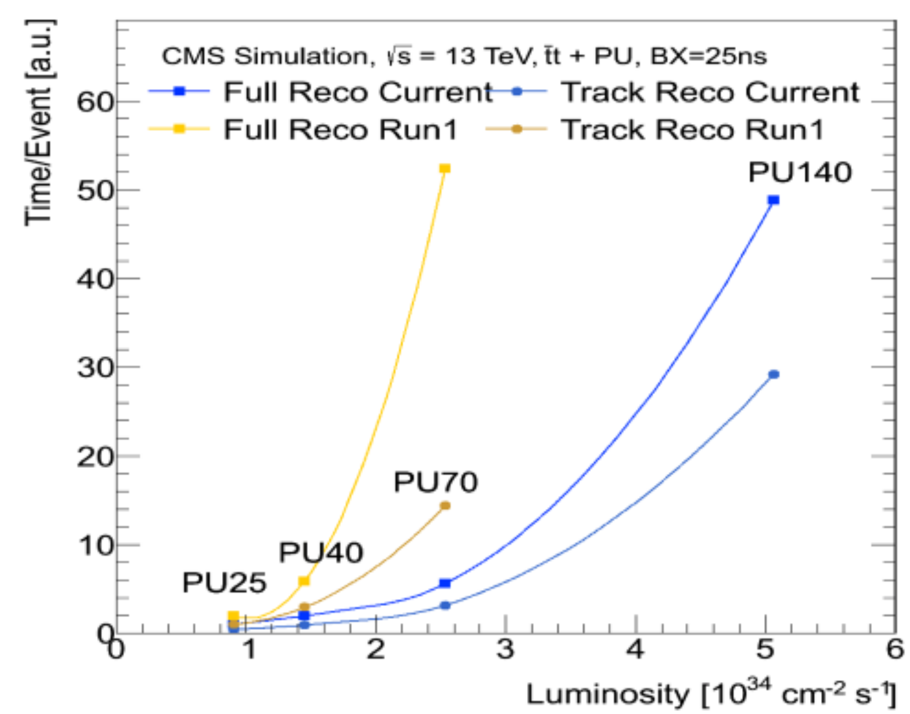
Large hit combinatorics in both seeding and stepping pattern recognition

TRACKING CRISIS

- Tracking (in particular pattern recognition) dominates reconstruction CPU time at LHC
- High Luminosity-LHC perspective : increased rate of parasitic collisions from 40 (2017) to 200
- CPU time of current software quadratic/ exponential extrapolation (difficult to quote any number)
- Current software give sufficiently good results in terms of accuracy, but x10 too slow







MOTIVATION OF NEW TECHNIQUES

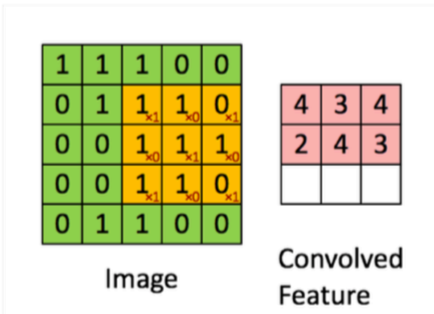
- LHC experiments future computing budget flat (at best) (LHC experiments use 300k CPU cores on the LHC world wide computing grid)
- Installed CPU power per $ == € == CHF expected increase factor less than 10 in 2025
- Experiments plan on increase of amount of data recorded (by a factor ~10)
  - High Luminosity reconstruction has to be as fast as current reconstruction despite factor 10 in complexity
  - Requires very significant software CPU improvement, factor ~10
- Large effort to optimize current software
  - Also development of dedicated hardware for fast tracking
- More 20 years of LHC tracking development. Everything has been tried!
  - Maybe yes, but maybe algorithm slower at low luminosity but with a better scaling have been dismissed ?
  - Maybe no, brand new ideas from ML
- Need to engage a wide community to tackle this problem

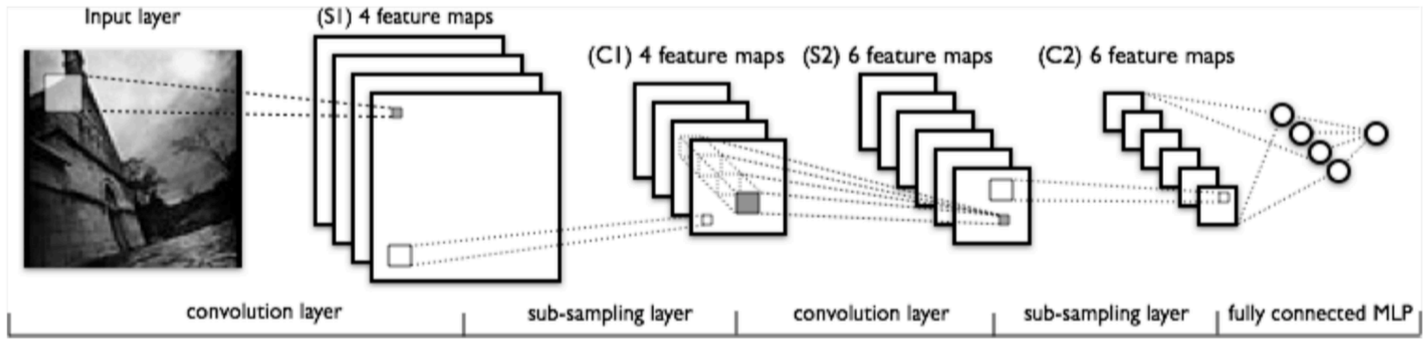
WORKSHOP HIGHLIGHTS

HIGHLIGHTS (I)

Convolutional Neural Networks for Track Seeding

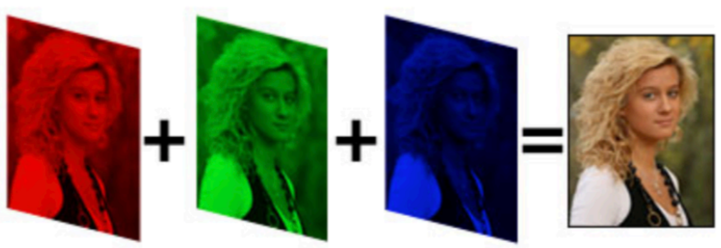
Convolutional Neural Networks are a specialized kind of neural networks for processing data that has a grid-like structure, such as 2D images. The building block of a CNNs is a layer that uses discrete convolution in place of general matrix multiplication.

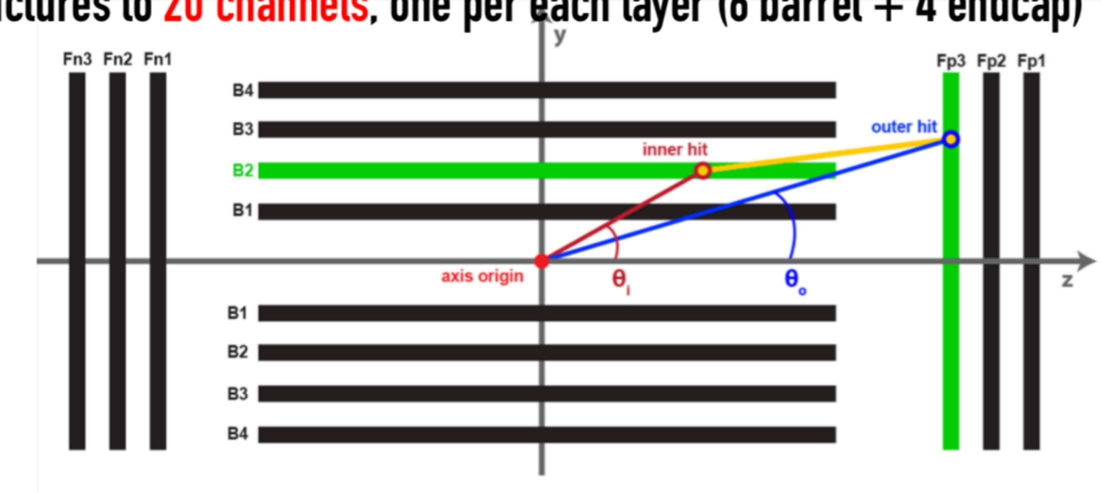




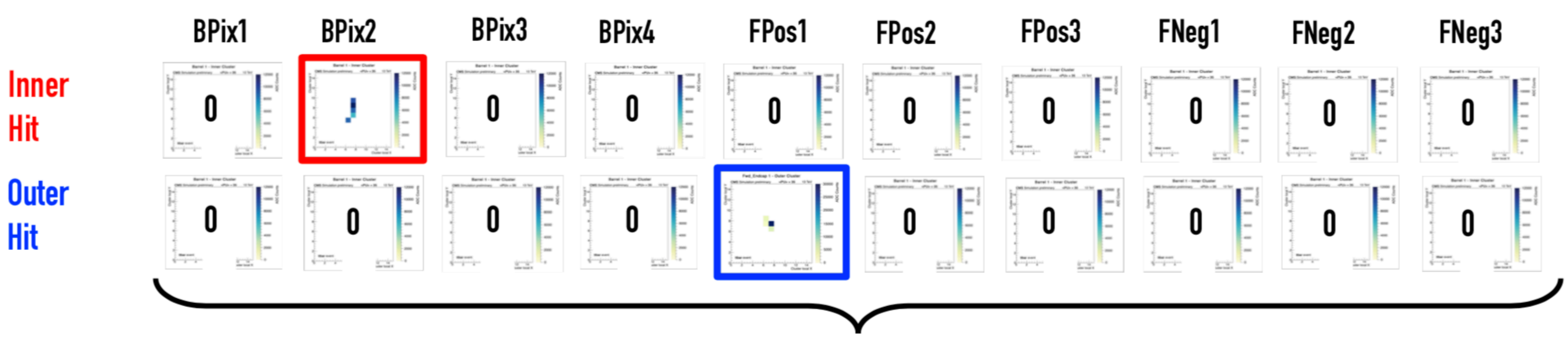
Make of use of the layer structure of the detector to extend each single doublet from two pictures to 20 channels, one per each layer (6 barrel + 4 endcap)

Channels: a common picture is usually a superimposition of different color levels. E.g. RGB levels.





E.g. A doublet on *BPIX2* and *FPIX_POS1*

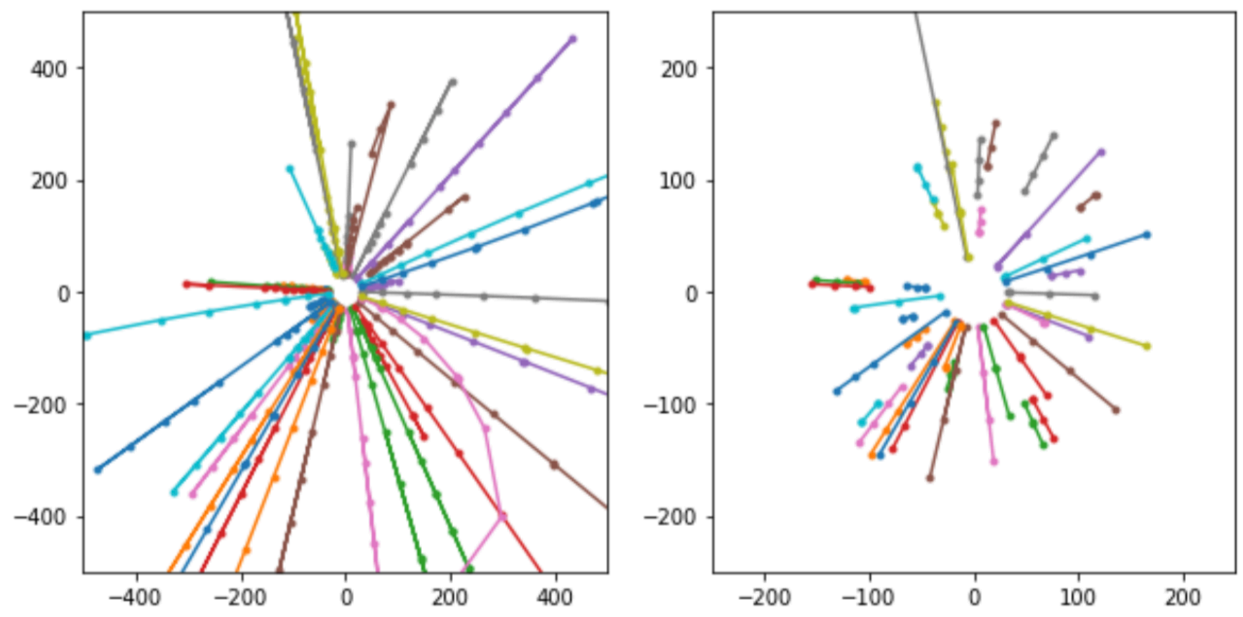


20 channel image as input

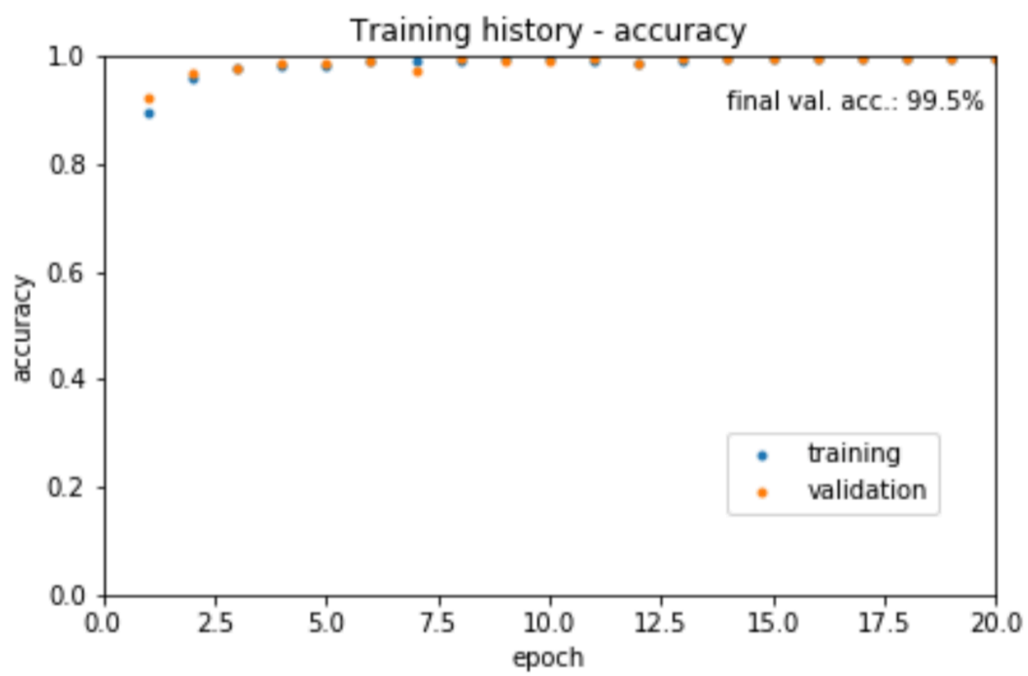
HIGHLIGHTS (II)

Track Seeding with Deep Neural Networks

- Assume seeding starts at beginning of tracks
- Try training on only first four hits of each track
  -> worsens performance
  Take random hits on full track instead



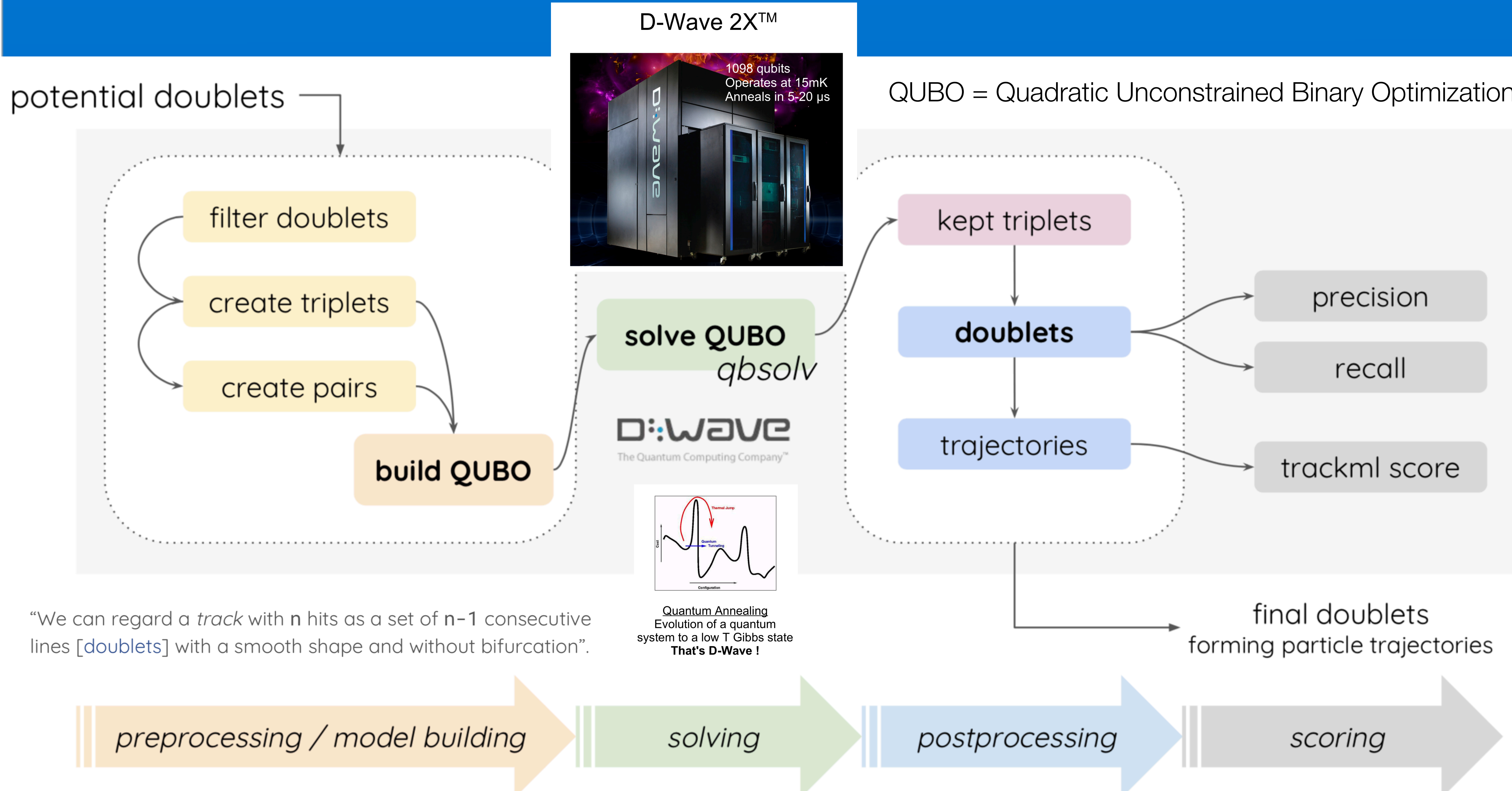
- Introduce **3 seed classes:**
  good/medium/bad
  with 0/1/2 wrong hits



Mostly high energy tracks
(lines are even straighter than half-circle experiment)
-> extremely high model accuracy

HIGHLIGHTS (III)

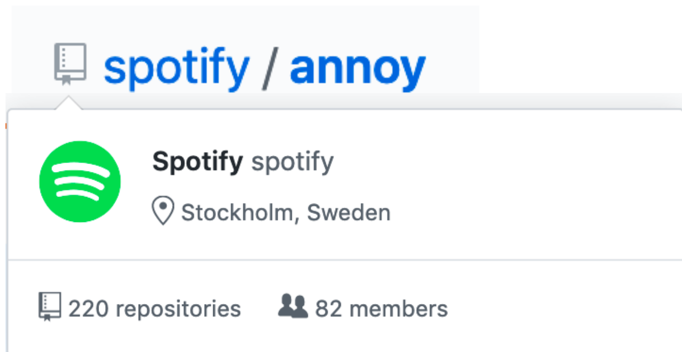
Embrace the combinatorics considering all possible branches of track candidates, and solve the complex optimization problem with quantum annealing



"We can regard a *track* with **n** hits as a set of **n-1** consecutive lines [doublets] with a smooth shape and without bifurcation".

HIGHLIGHTS (IV)

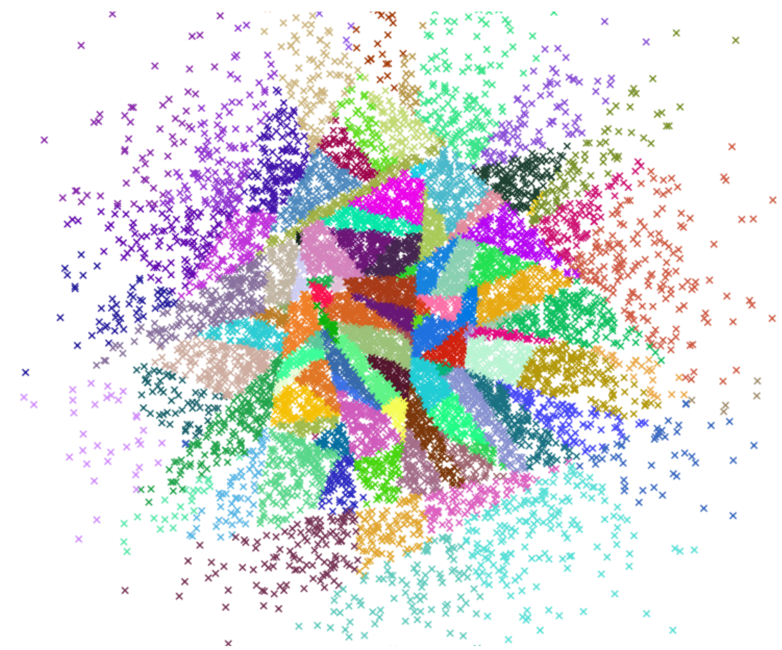
Reduce combinatorial complexity : define regions



Approximate Nearest Neighbors

build passing | windows - OK | pypi v1.15.2

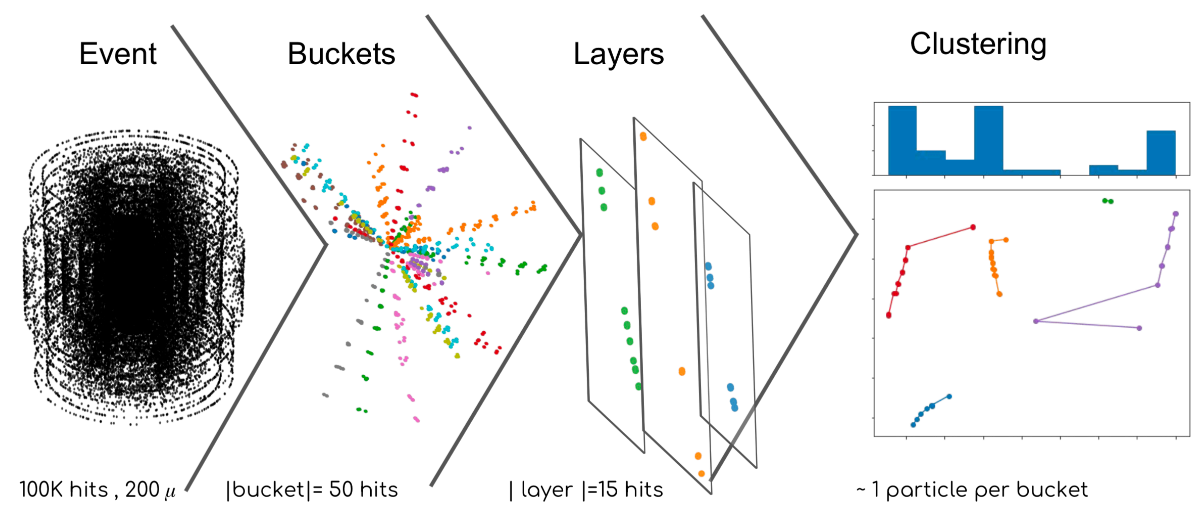
Annoy (Approximate Nearest Neighbors Oh Yeah) is a C++ library with Python bindings to search for points in space that are close to a given query point. It also creates large read-only file-based data structures that are mmapped into memory so that many processes may share the same data.



- "Many millions of songs"
- **< 0.1 ms** to get **n** similar songs

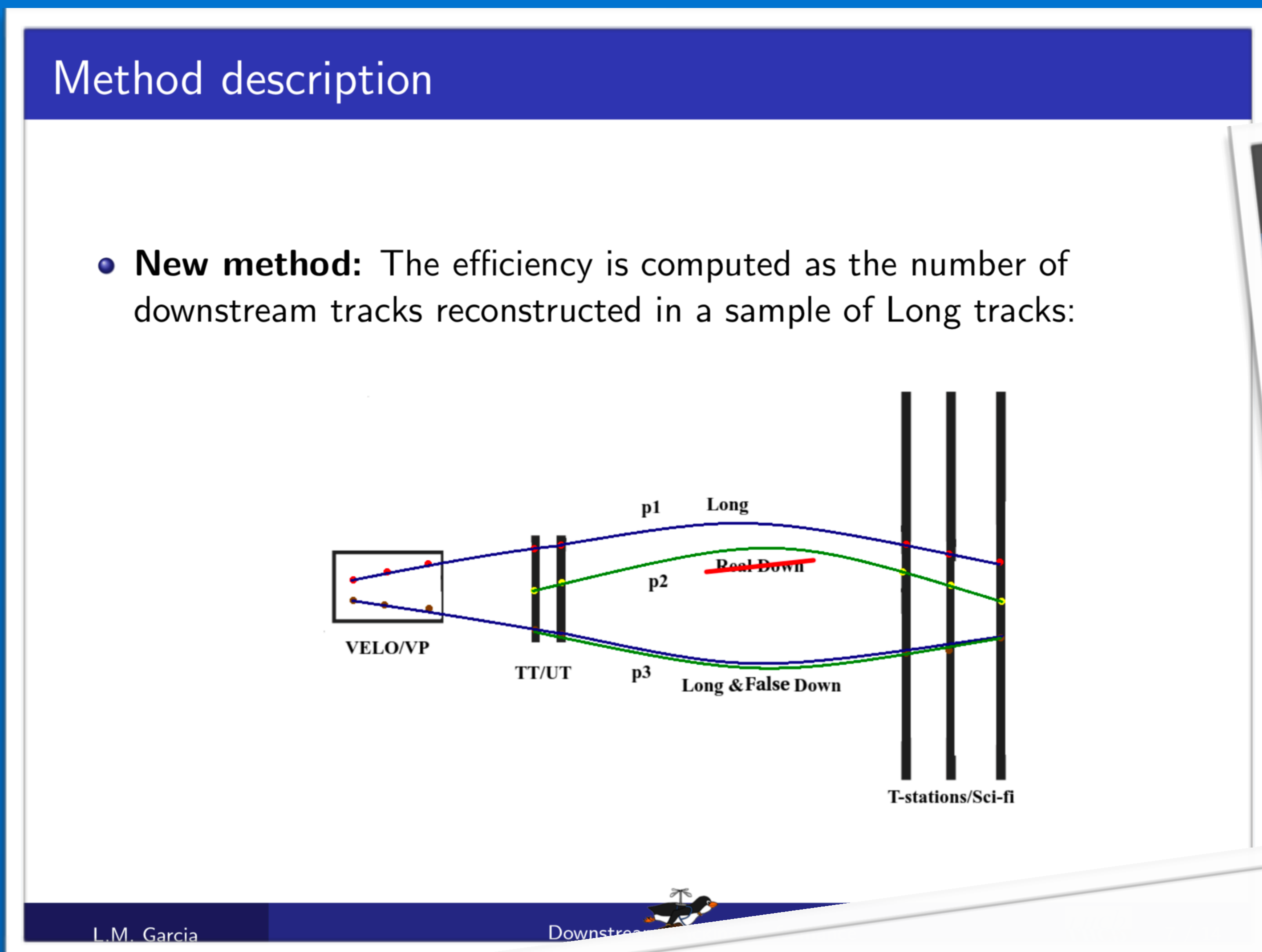
- **Application to Tracking:**

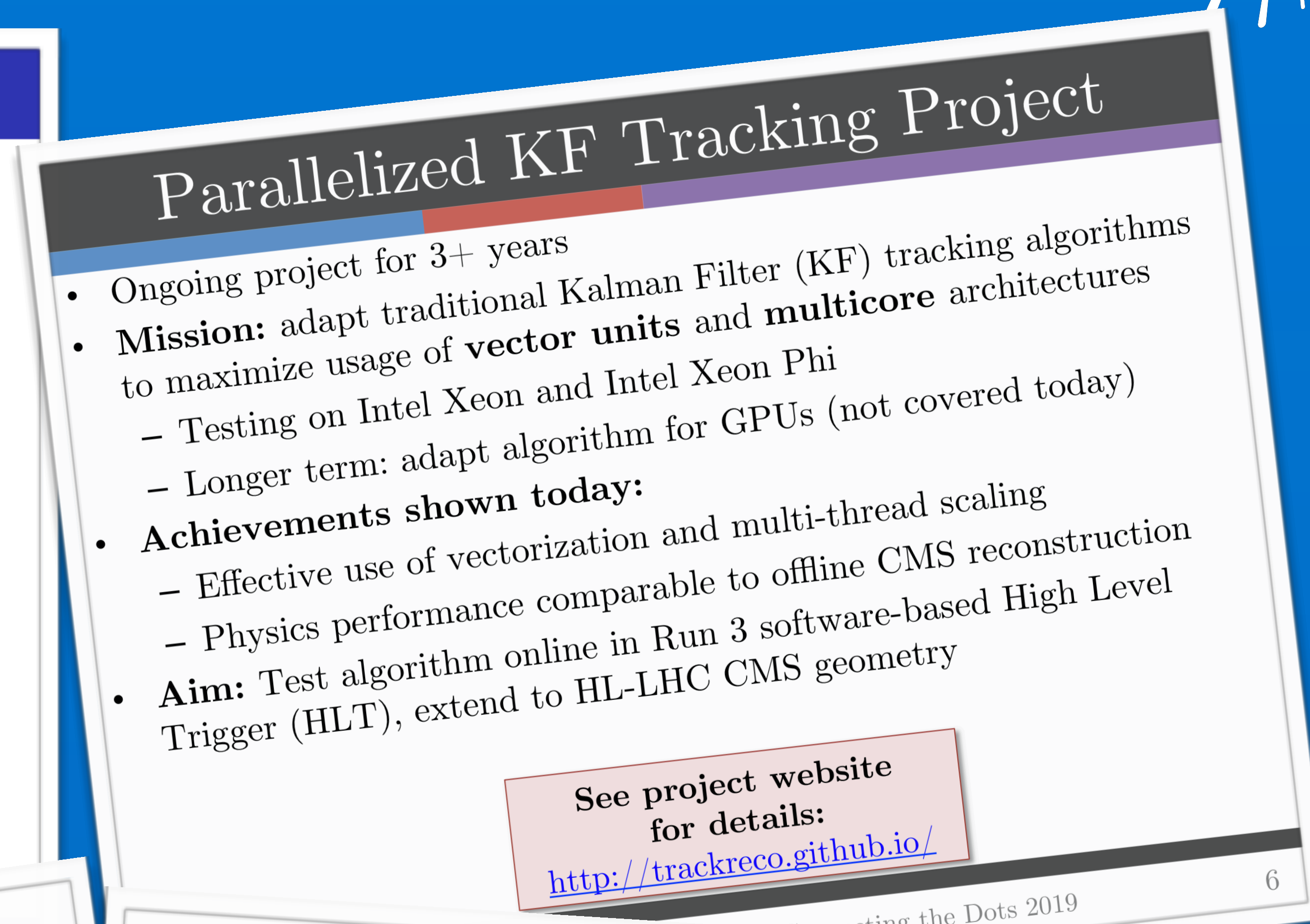
  Supervised model trained on buckets evaluating if N hits belong together

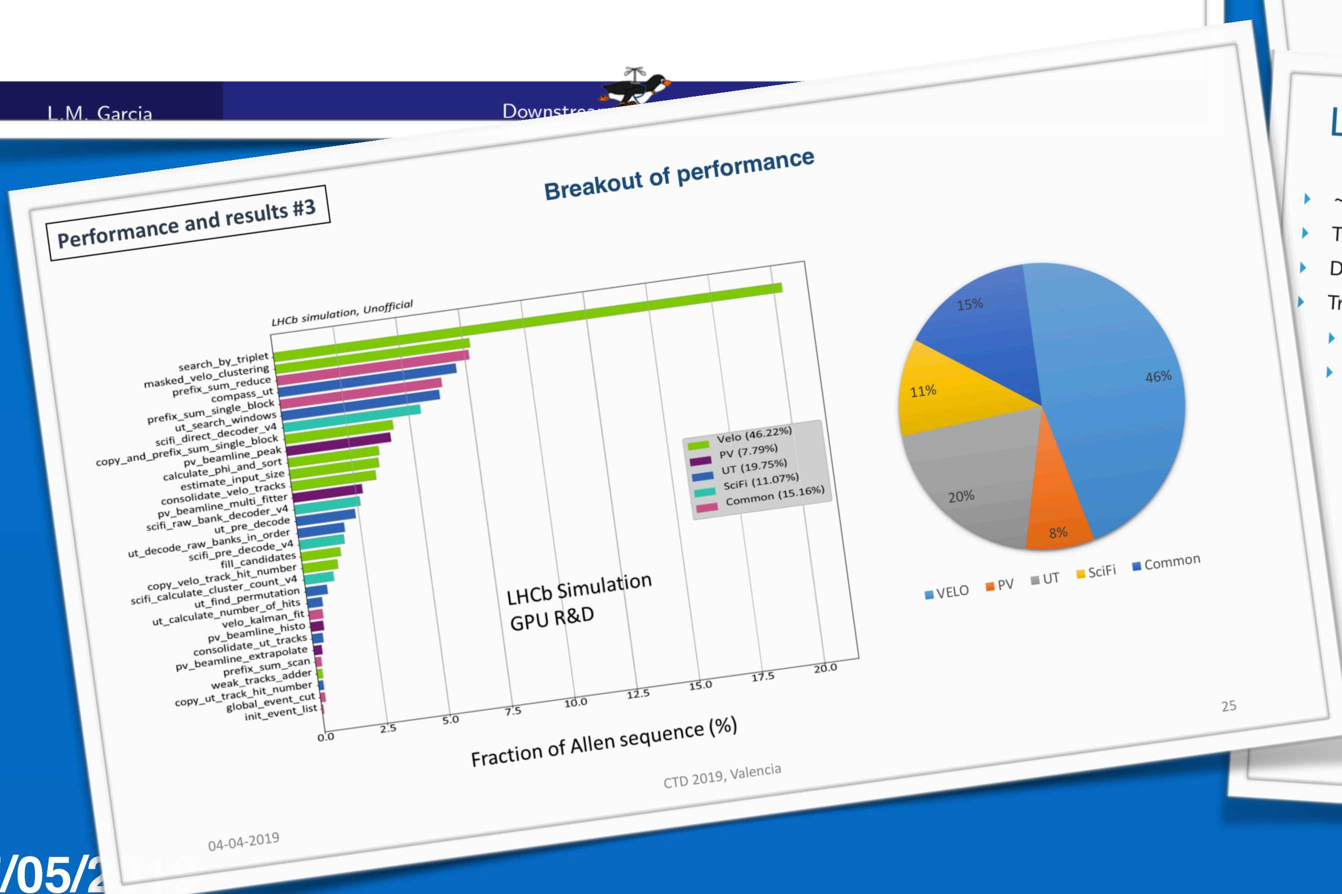


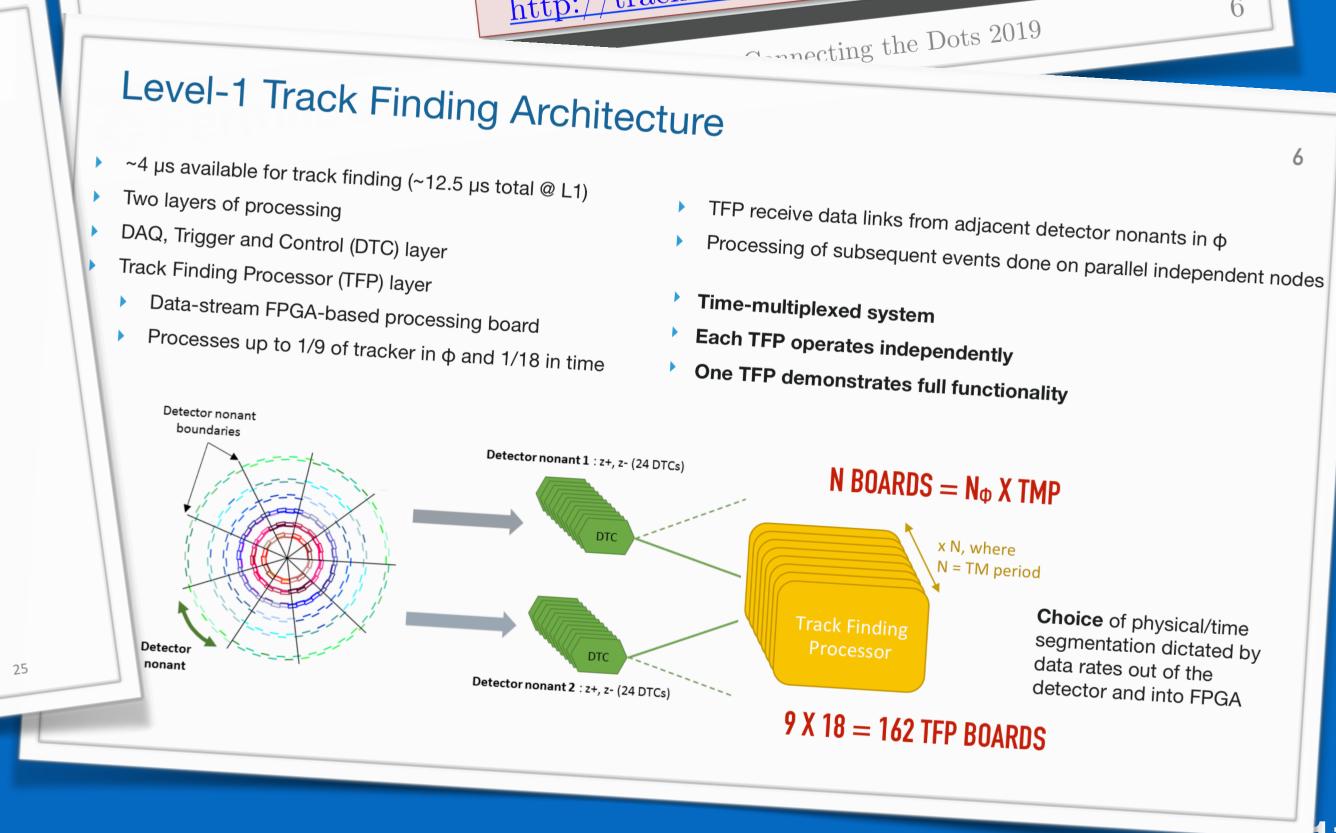
HIGHLIGHTS (V)

Trigger









HIGHLIGHTS (VI)

The challenge in 2 phases

Phase 1: **accuracy phase**

kaggle

Launch
6 months ago

Close
2 months ago

Apr 30, 2018

2 months ago
Rules
De

Aug 14, 2018

Phase 2: **throughput phase**

CodaLab

▶ Current	Next	End
Development	Final	Competition Ends
Sept. 7, 2018, midnight UTC	Nov. 5, 2018, 11:59 p.m. UTC	Nov. 12, 2018, 11:59 p.m. UTC

In the money   Gold   Silver   Bronze

#	△pub	Team Name	Kernel	Team Members	Score	Entries	Last
1	—	Top Quarks			0.92182	10	2mo
2	—	outrunner			0.90302	9	2mo
3	—	Sergey Gorbunov			0.89353	6	2mo
4	—	demelian			0.87079	35	2mo
5	—	Edwin Steiner			0.86395	5	2mo
6	—	Komaki			0.83127	22	2mo
7	—	Yuval & Trian			0.80414	56	2mo
8	—	bestfitting			0.80341	6	2mo

RESULTS

#	User	Entries	Date of Last Entry	score ▲	accuracy_mean ▲	accuracy_std ▲	computation time (sec) ▲	computation speed (sec/event) ▲	Duration ▲
1	sgorbuno	9	03/12/19	1.1727 (1)	0.944 (2)	0.00 (14)	28.06 (1)	0.56 (1)	64.00 (1)
2	fastrack	53	03/12/19	1.1145 (2)	0.944 (1)	0.00 (15)	55.51 (16)	1.11 (16)	91.00 (6)
3	cloudkitchen	73	03/12/19	0.9007 (3)	0.928 (3)	0.00 (13)	364.00 (18)	7.28 (18)	407.00 (8)
4	cubus	8	09/13/18	0.7719 (4)	0.895 (4)	0.01 (9)	675.35 (19)	13.51 (19)	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)

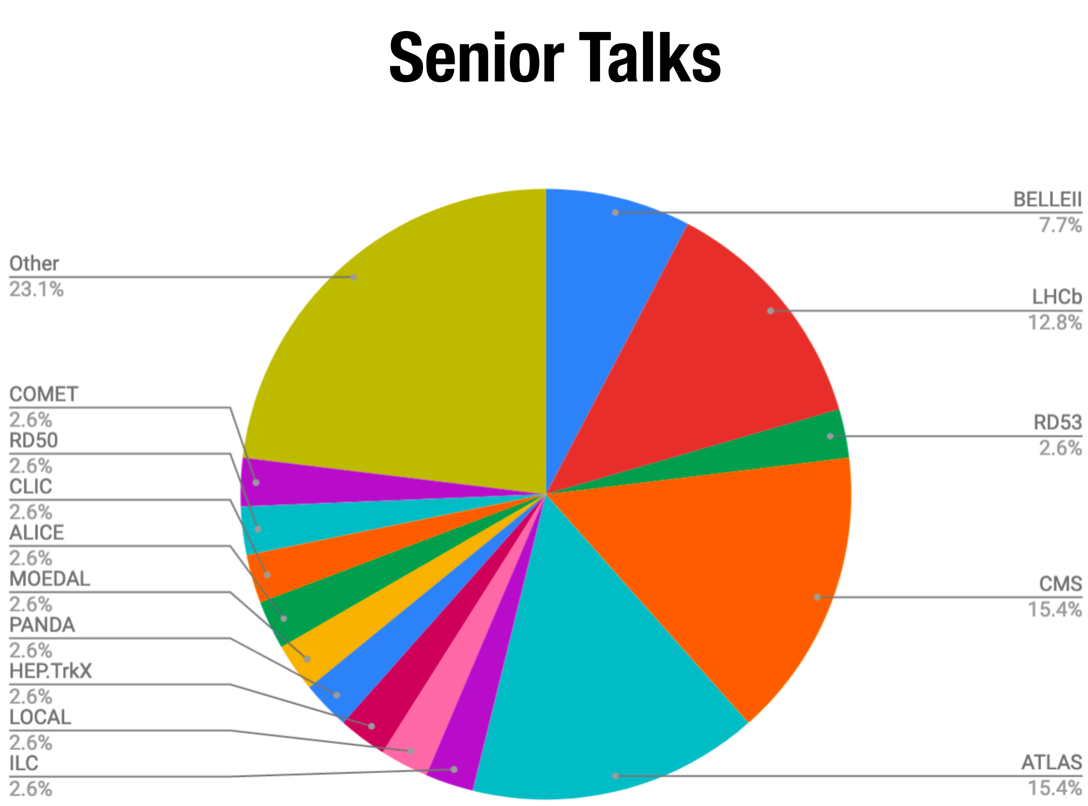
Tracking Machine Learning Challenge

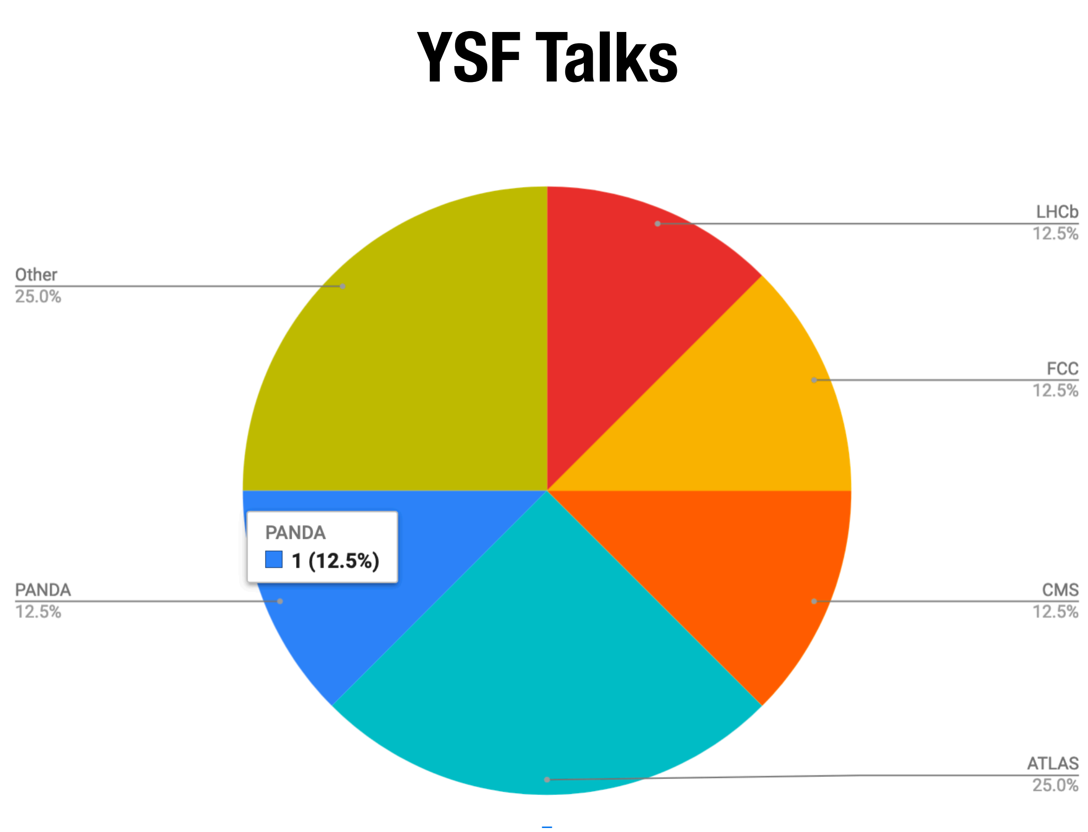
Summary from Phase 1 & Phase 2

A. Salzburger (CERN)
@SaltyBurger

WORKSHOP SUMMARY

- Attendance
  - 90 people (30 PhD students)
- Talks:
  - 35 Senior, 10 YSF, 12 Posters, 1 Botanic Garden + 1 Discussion session
  - Female 11, Male 34
  - 5 by Spanish Institutes





WORKSHOP SUMMARY

Intelligent Trackers

- RD50 and RD53 overview
- GPU/FPGA based tracking for trigger/online
  - both classical and ML based, for HLT or L1
- FTK shows first tracking results, design of FPGA L1Track in CMS is progressing
- ML R&D for tracking is massive, moving into production readiness of first applications

Connecting the Dots

- QUBO quantum computing is reconstructing HEP events
- Music searches inspire HEP tracking
- Tracking and Vertexing Challenges were a success
  - Tracking ML results are fascinating
  - Tracking ML dataset used widely in results presented in talks, it fulfills its purpose
- Tuning of classical tracking is still relevant for upgrades and non-conventional tracking (LLP)
  - classical tracking on CPUs still pays out (vectorising, data structures, …)
- Contributions beyond LHC and BELLE II: PANDA, COMET
  - LHC Run 2 and readiness for Run 3
  - Belle II Phase 3 has started with good detector performance
- The future (FCC and CLIC) is challenging

Questions?