Machine Learning for High Energy Physics

part 2

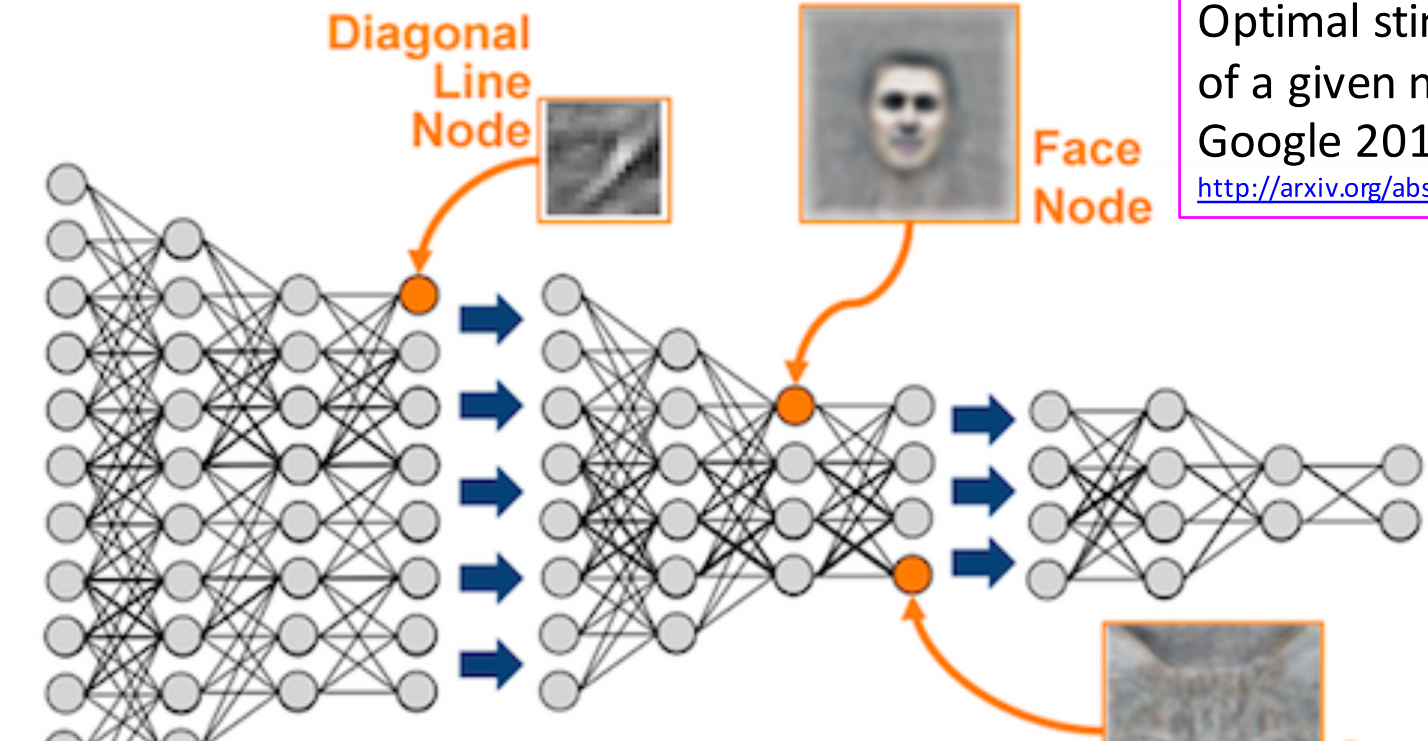
David Rousseau

LAL-Orsay, CNRS/IN2P3, Université Paris-Sud/Paris-Saclay

rousseau@lal.in2p3.fr , twitter: @dphmrou

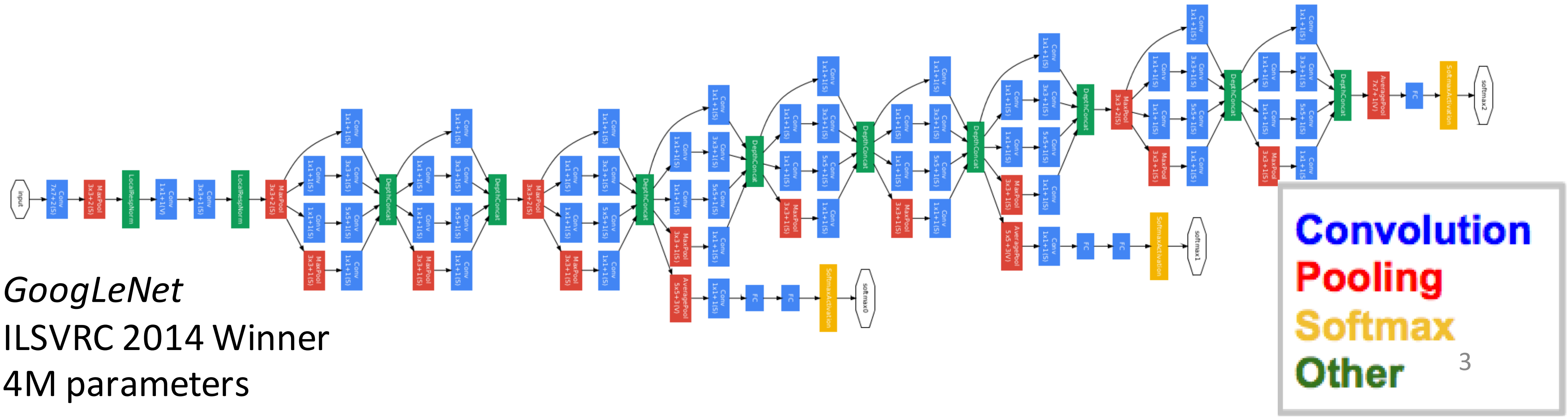
Deep learning

Deep learning



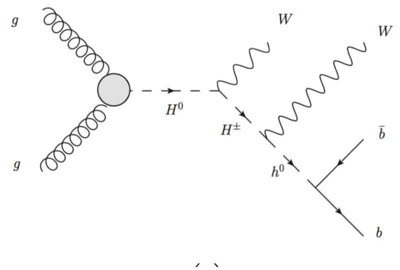
Optimal stimulus
of a given neuron
Google 2012
http://arxiv.org/abs/1112.6209

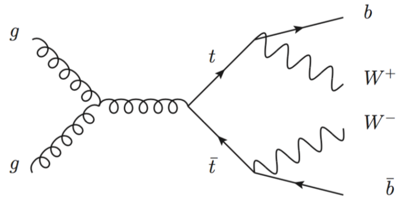
~1950
1990 handwritten digits
2000 deep learning explosion

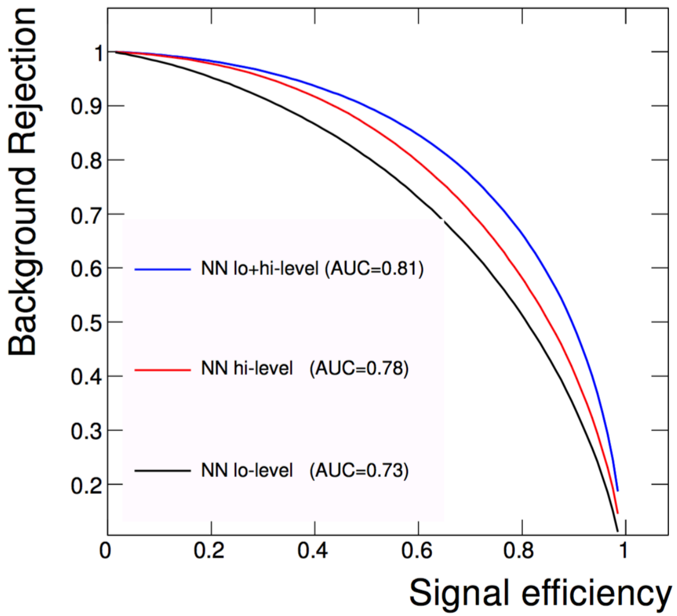


GoogLeNet
ILSVRC 2014 Winner
4M parameters

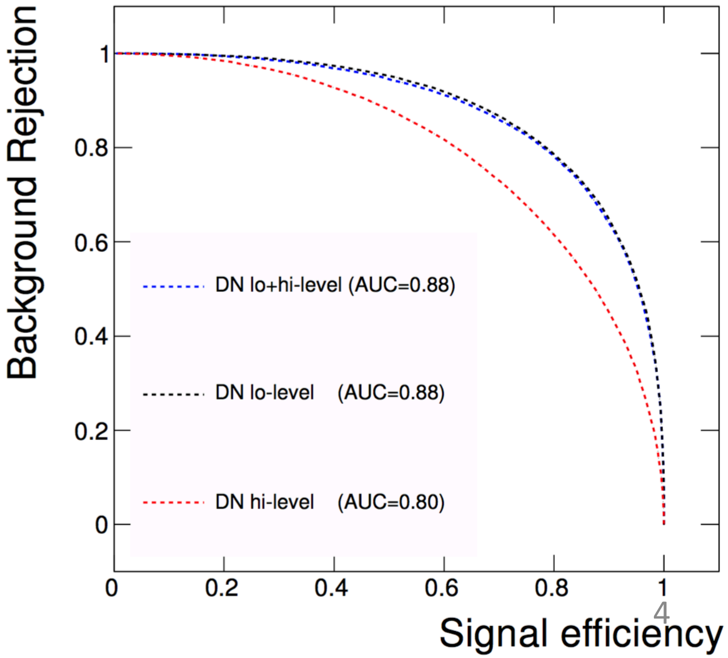
[1402.4735](1402.4735) Baldi, Sadowski, Whiteson





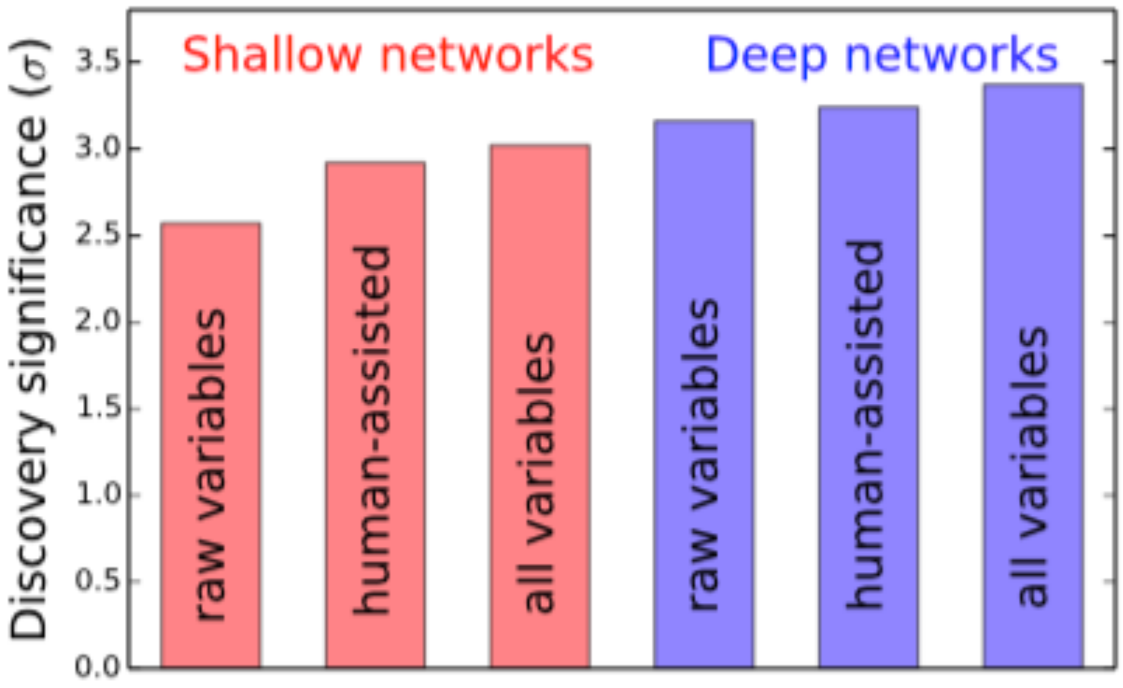


- Looking for heavy Higgs: H^0 ➔WWbb vs tt➔WWbb
- Simplified LHC simulation (Delphes)
- Low-level variables
  - 4-vectors
- High level variables
  - Invariant masses
- ➔DNN (Deep Neural Network) works better
- ➔DNN does not need high level
- ➔➔DNN learns the physics ???



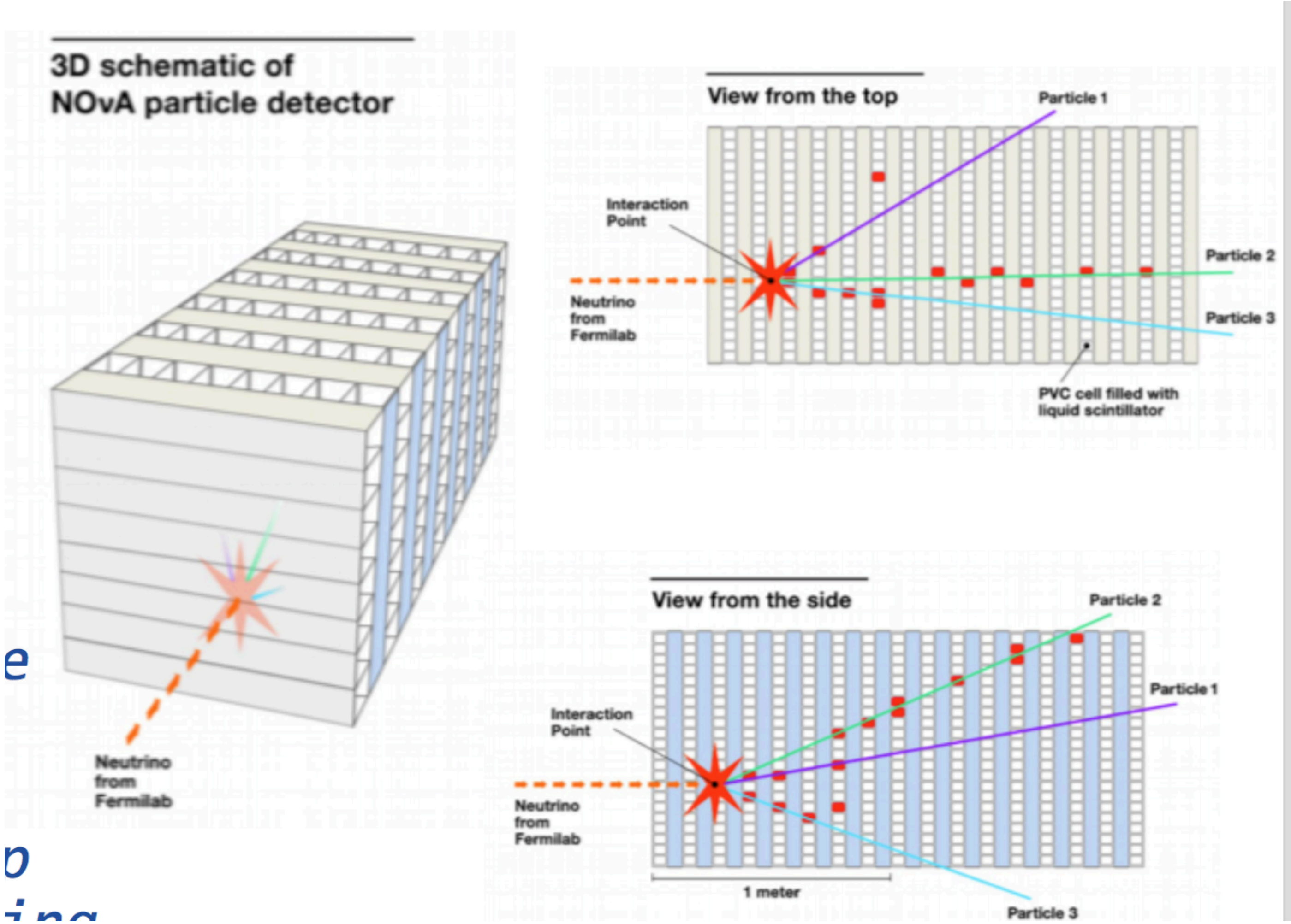
1410.3469 Baldi Sadowski Whiteson

- Analysis H tautau : H➔tautau vs Z➔tautau
  - Low level
  - High level (=human assisted)



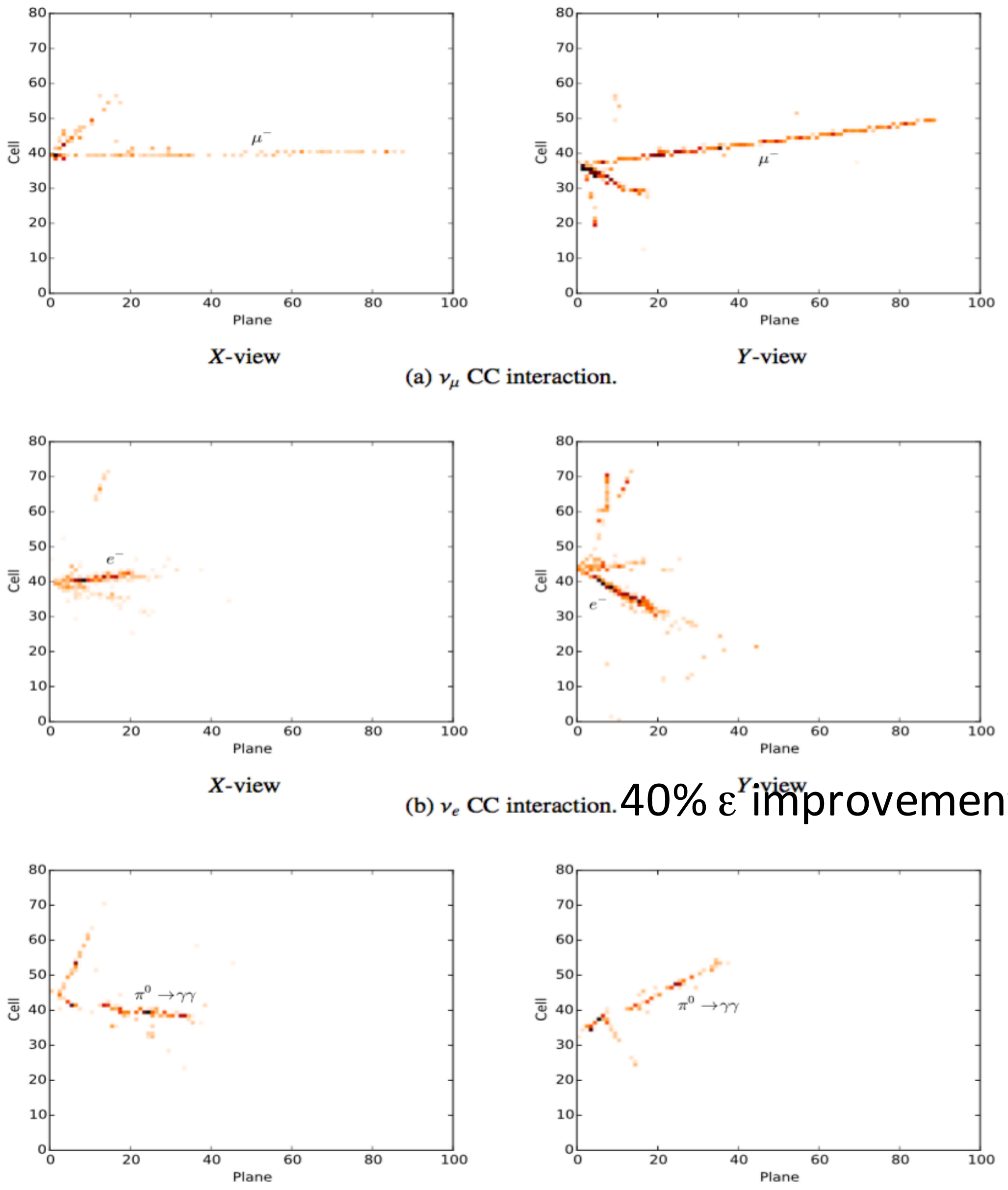
- DNN works better
- But now needs the high level variables
- Things gets complicated…
- Also need millions of events (we typically have few thousands)
- For now, no physics results with DNN except…

Deep Learning success : NOVA (neutrino)



NOVA (2)

arXiv 1604.01444 Aurisano et al



X-view *Y*-view
(a) ν_μ CC interaction.

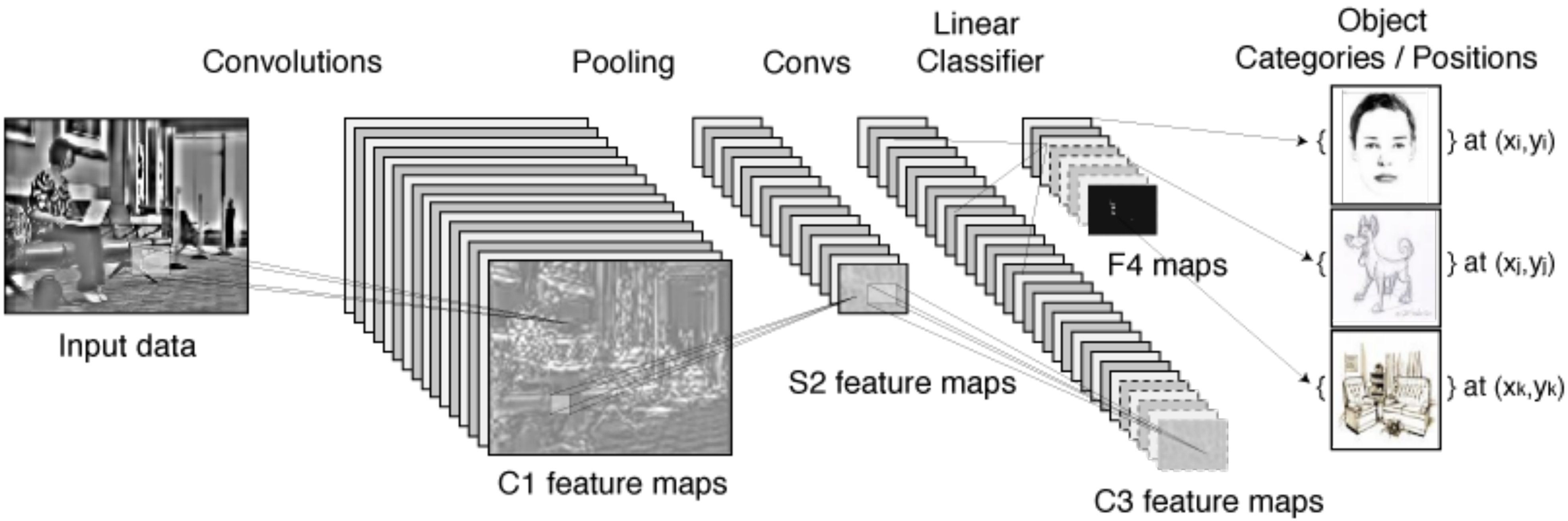
X-view *Y*-view
(b) ν_e CC interaction.

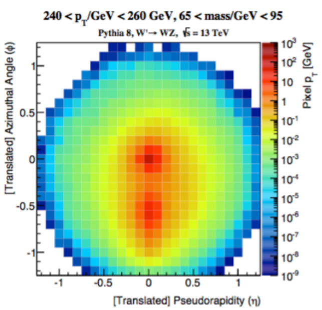
X-view *Y*-view
(c) NC interaction.

40% ε improvement

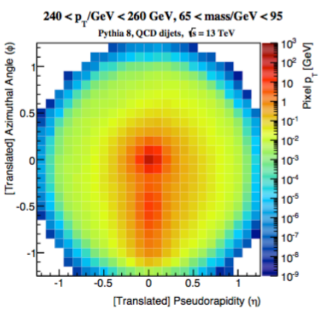
Neutrino interaction classification
Using Convolutionnal Neural Network (GoogleNet)
Actually used in physics results 1703.03328 and 1706.04592

Convolutionnal Neural Network (CNN)





Higgs Lourd



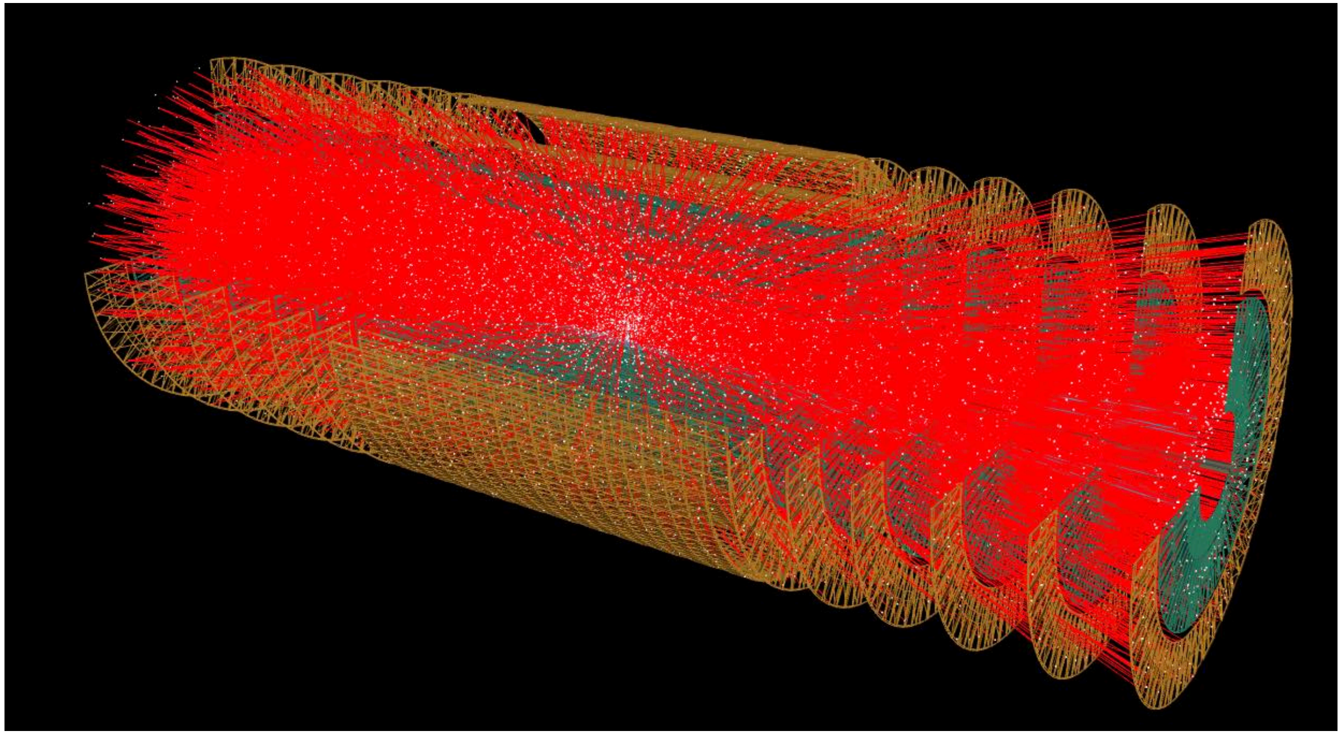
Bruit de fond

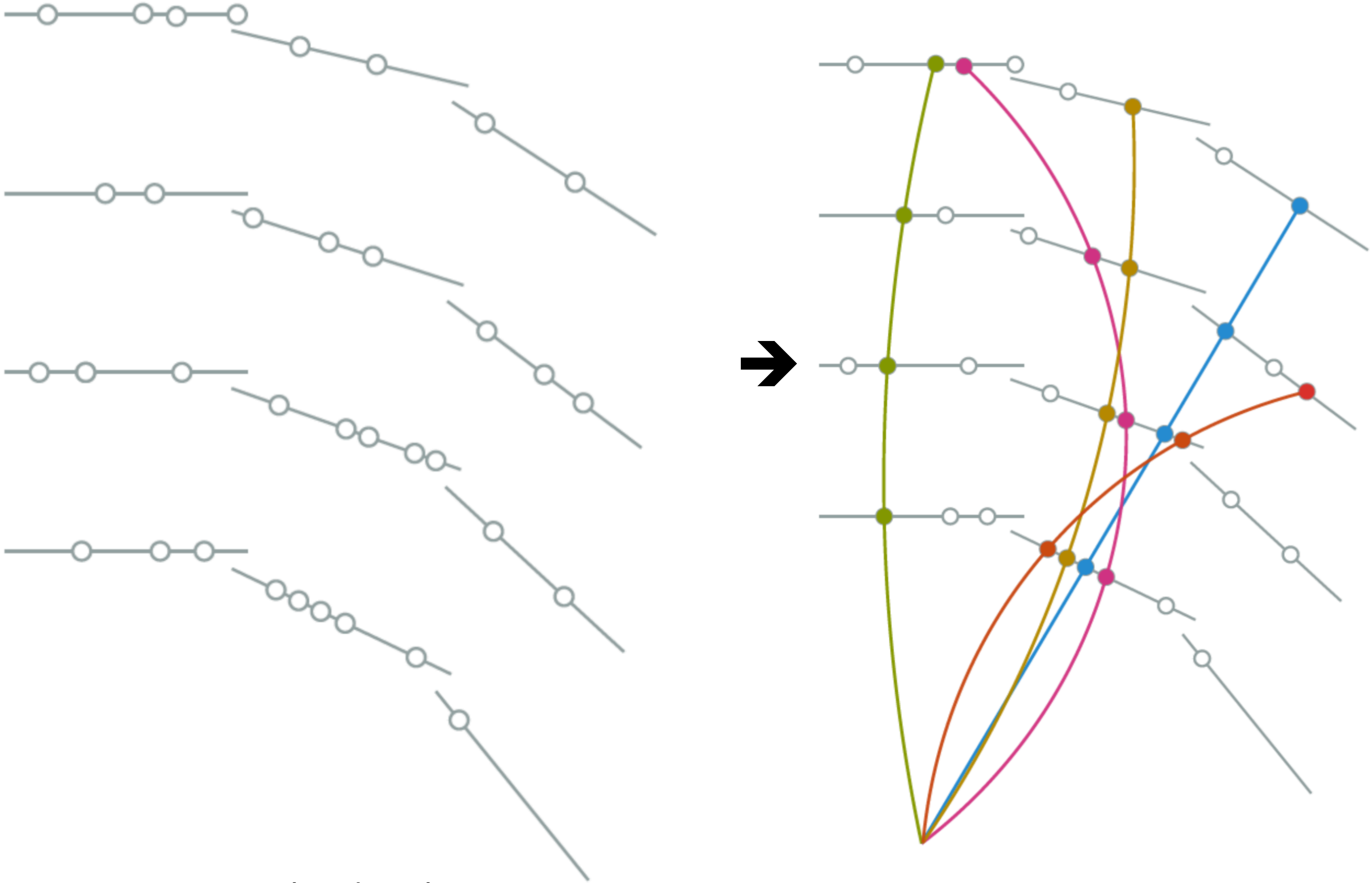
Tracking challenge

https://sites.google.com/site/trackmlparticle/
https://www.kaggle.com/c/trackml-particle-identification
https://twitter.com/trackmllhc

Issue

- High Luminosity LHC in 2025
- Increase of number of proton collision per event
- ➔pile-up of parasitic collision ~200 (compared to ~50 now)
- ➔very complex events
- In particular for trajectography
- Factor 10 speed-up to be found





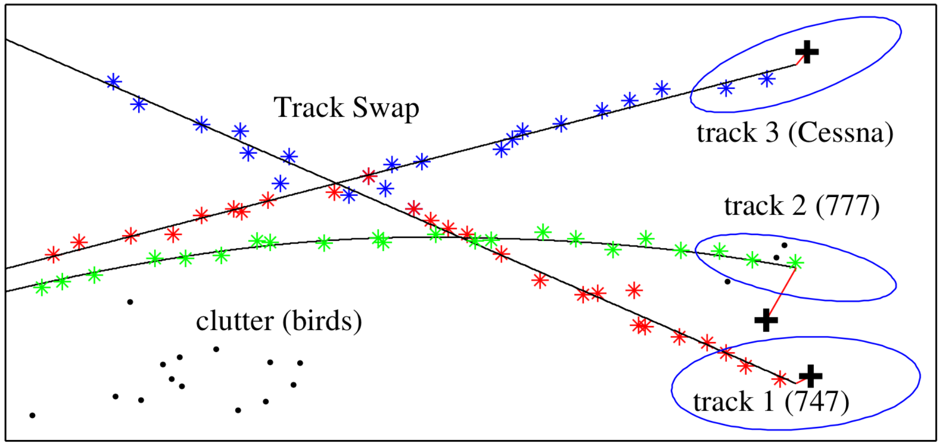
Connecting the dots but

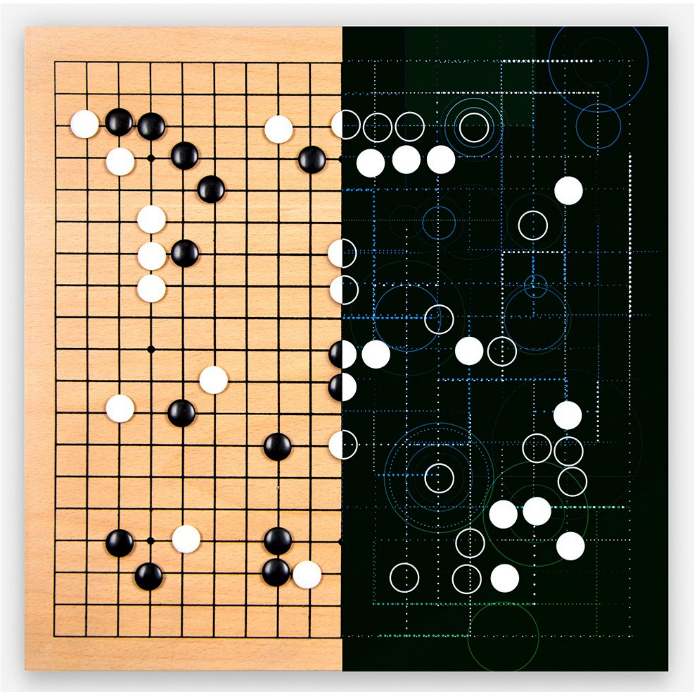
- 3 dimensions
- 10’000 tracks x 10 points

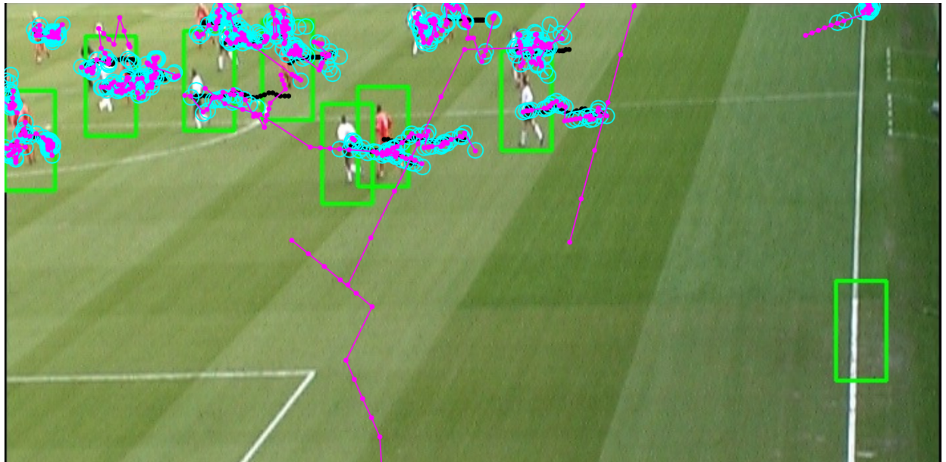
Why is it difficult?

- 100’000 to group into 10’000 tracks of 10 points
  - ➔~$10^{450'000}$ combinations
  - ⇒brute force has (really) no chance
- Precision of the points : ~50μm on a volume ~40 m^3
  - ➔3 10^{14} voxels!
  - 2D projection ➔2 10^9 pixels !
  - ⇒ image recognition algorithm have (really) no chance
- Not a classical problem

Tracking outside HEP

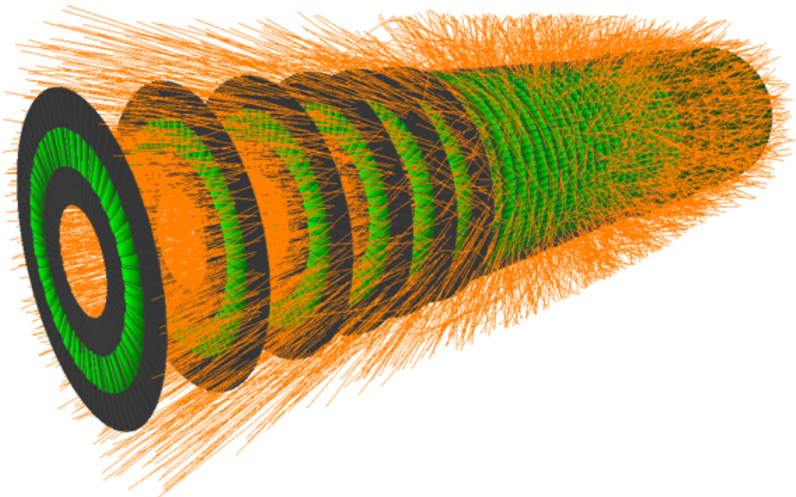






Tracking Machine Learning challenge

https://twitter.com/trackmllhc



pitch : put the points (x,y,z) on the web

ask people to connect the dots

Objective : expose new algorithms

- Accuracy phase on Kaggle, only the accuracy: May to August 2018
- Throughput phase on-going till 12th March 2019:accuracy and speed

Dataset

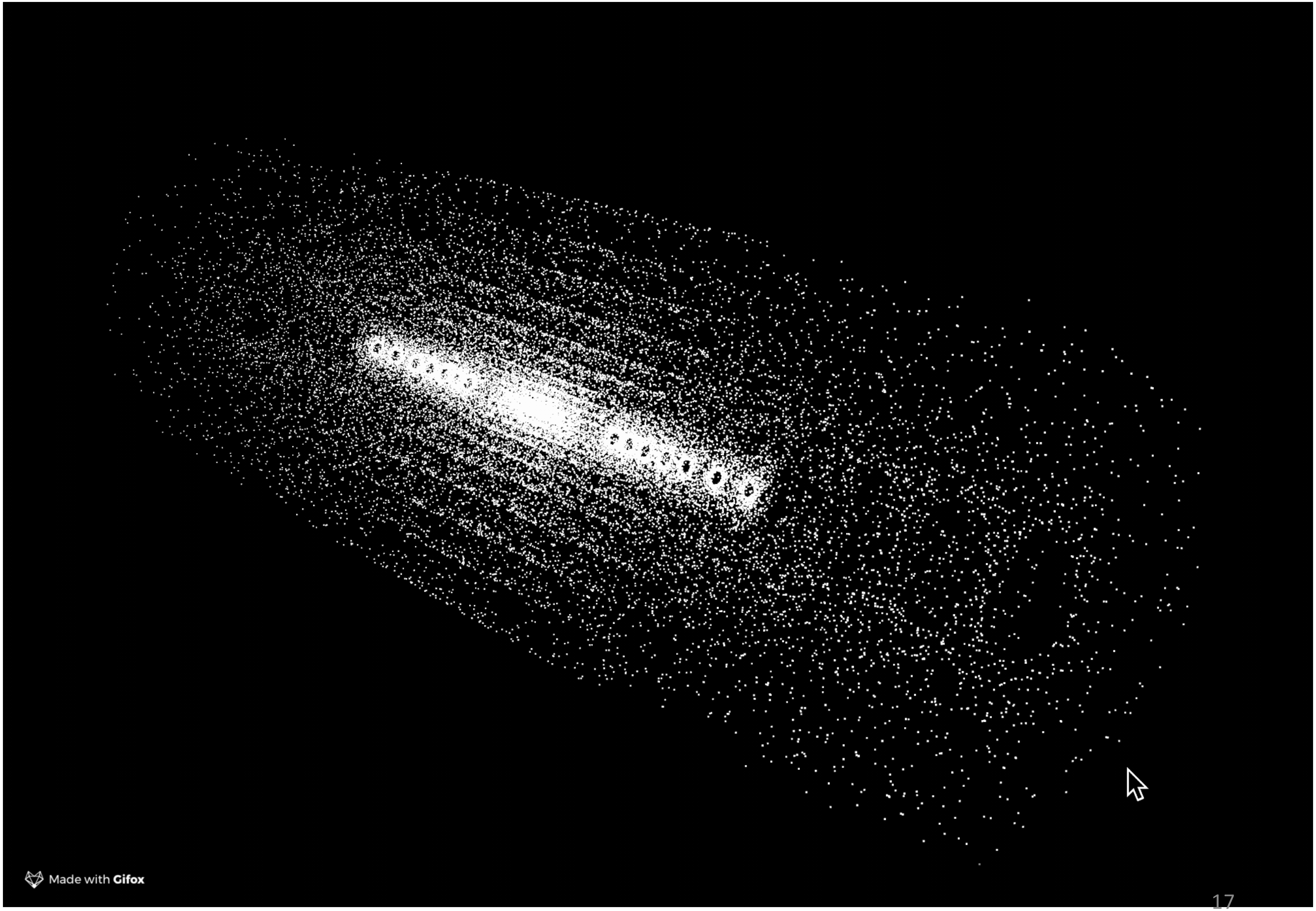
- Hit file

(measured position mm)

	hit_id	x	y	z	volume_id	layer_id	module_id
0	1	-64.409897	-7.163700	-1502.5	7	2	1
1	2	-55.336102	0.635342	-1502.5	7	2	1
2	3	-83.830498	-1.143010	-1502.5	7	2	1
3	4	-96.109100	-8.241030	-1502.5	7	2	1
4	5	-62.673599	-9.371200	-1502.5	7	2	1

❑ Truth file (true position mm particle momentum GeV)

	hit_id	particle_id	tx	ty	tz	tpx	tpy	tpz	weight
0	1	0	-64.411598	-7.164120	-1502.5	250710.000000	-149908.000000	-956385.00000	0.000000
1	2	22525763437723648	-55.338501	0.630805	-1502.5	-0.570605	0.028390	-15.49220	0.000010
2	3	0	-83.828003	-1.145580	-1502.5	626295.000000	-169767.000000	-760877.00000	0.000000
3	4	297237712845406208	-96.122902	-8.230360	-1502.5	-0.225235	-0.050968	-3.70232	0.000008
4	5	418835796137607168	-62.659401	-9.375040	-1502.5	-0.281806	-0.023487	-6.57318	0.000009

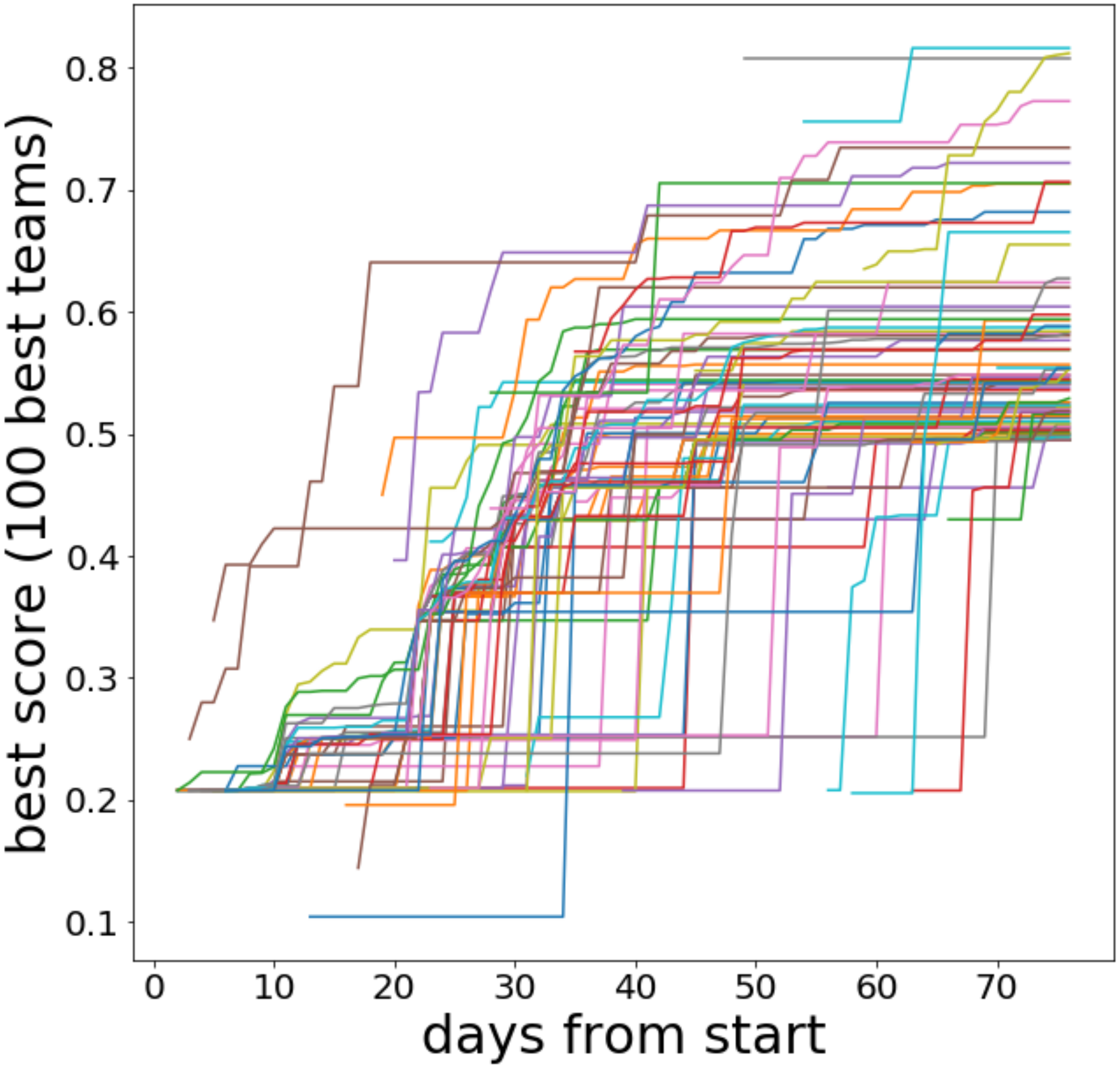

Made with Gifox

Visualisation spin-off

- Visit at CERN Tobias Isenberg visualisation scientist at LRI-Orsay with PhD student Xiyao Wang
- Will use TrackML dataset to experiment with visualisation/interaction with Microsoft' Hololens

Leaderboard

#	△1w	Team Name	Kernel	Team Members	Score	Entries	Last
1	—	Edwin Steiner			0.8161	2	13d
2	▲2	demelian			0.8118	16	9h
3	▼1	outrunner			0.8070	1	1mo
4	▼1	yuval r			0.7726	19	5h
5	—	Mickey			0.7345	10	18d
6	—	Zidmie			0.7220	11	10d
7	▲3	On est les champions !!!!!			0.7065	10	2d
8	▼1	icecuber			0.7054	3	1mo
9	▼1	Vicens Gaitan			0.7049	13	6d
10	▼1	Félicitations à la France			0.6818	42	6d
11	—	Victor Nedel'ko			0.6653	4	10d
12	—	John Sweeney			0.6551	16	5d
13	▲3	trian2018			0.6274	23	21h
14	▼1	Seb B			0.6240	20	15d
15	▼1	Grzegorz Sionkowski			0.6200	22	25d
16	▼1	Andrea Lonza			0.6043	7	1mo



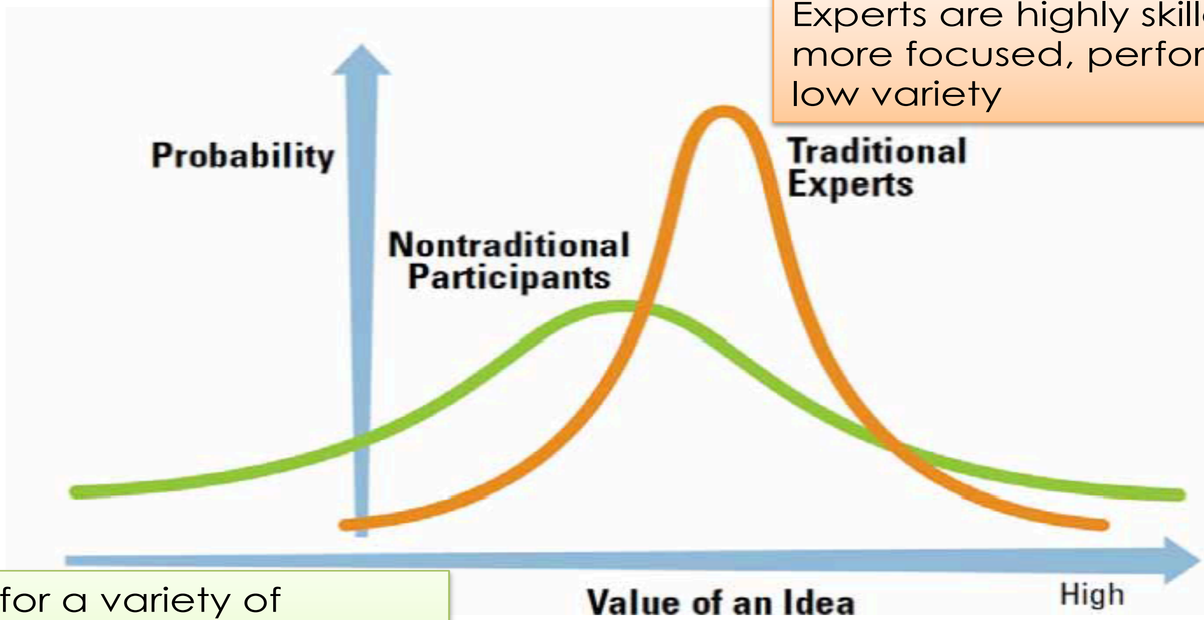
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
best score (100 best teams)
0
10
20
30
40
50
60
70
days from start

Why challenges work ?

MOTIVATION OF ORGANIZING CONTESTS: EXTREME VALUE

Courtesy : Lakhani 2014

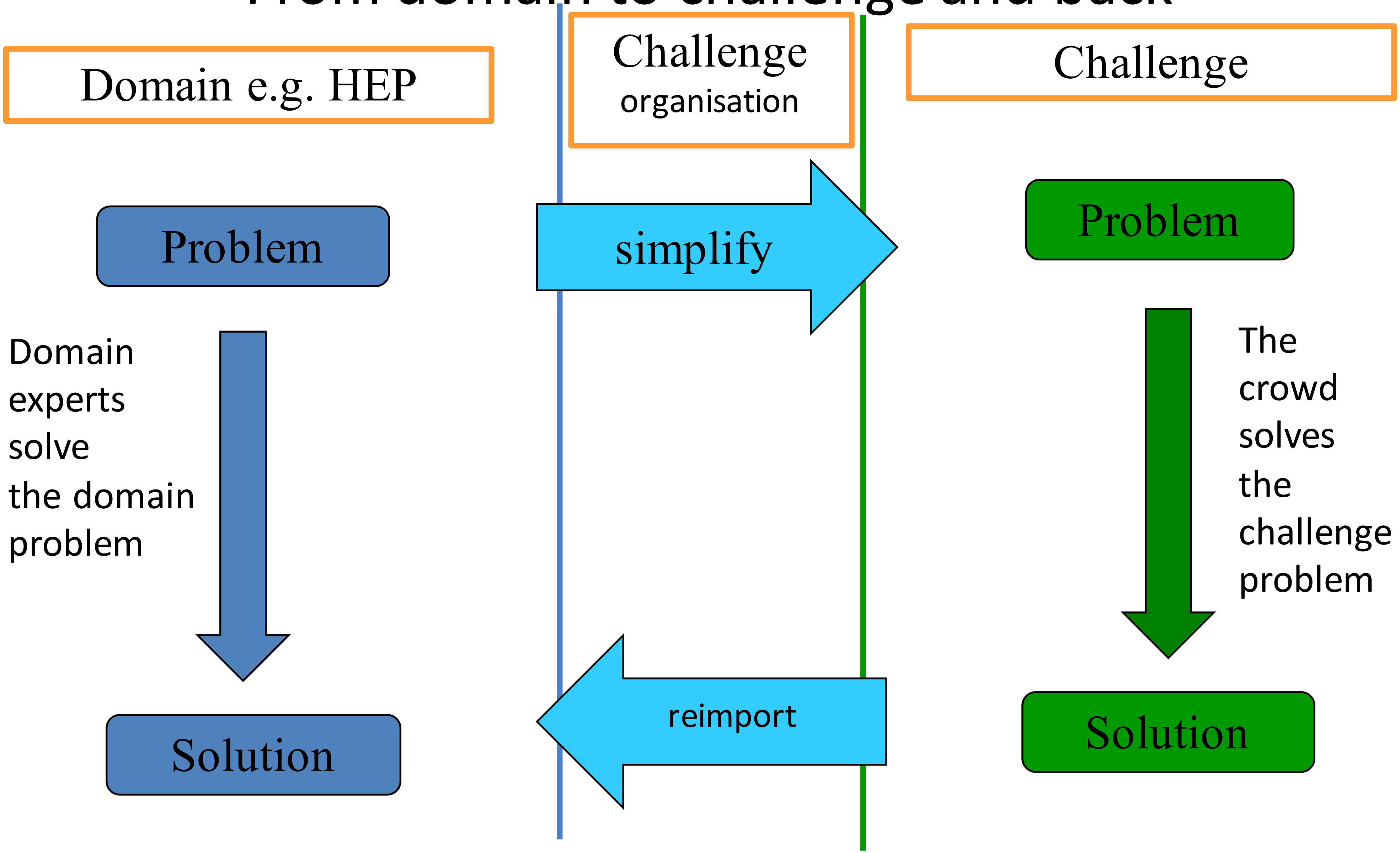
Experts are highly skilled, trained - > more focused, performed solution, low variety



OI is suitable for a variety of nonconvential surprising ideas that are « far » from traditional expertise - > high volatility

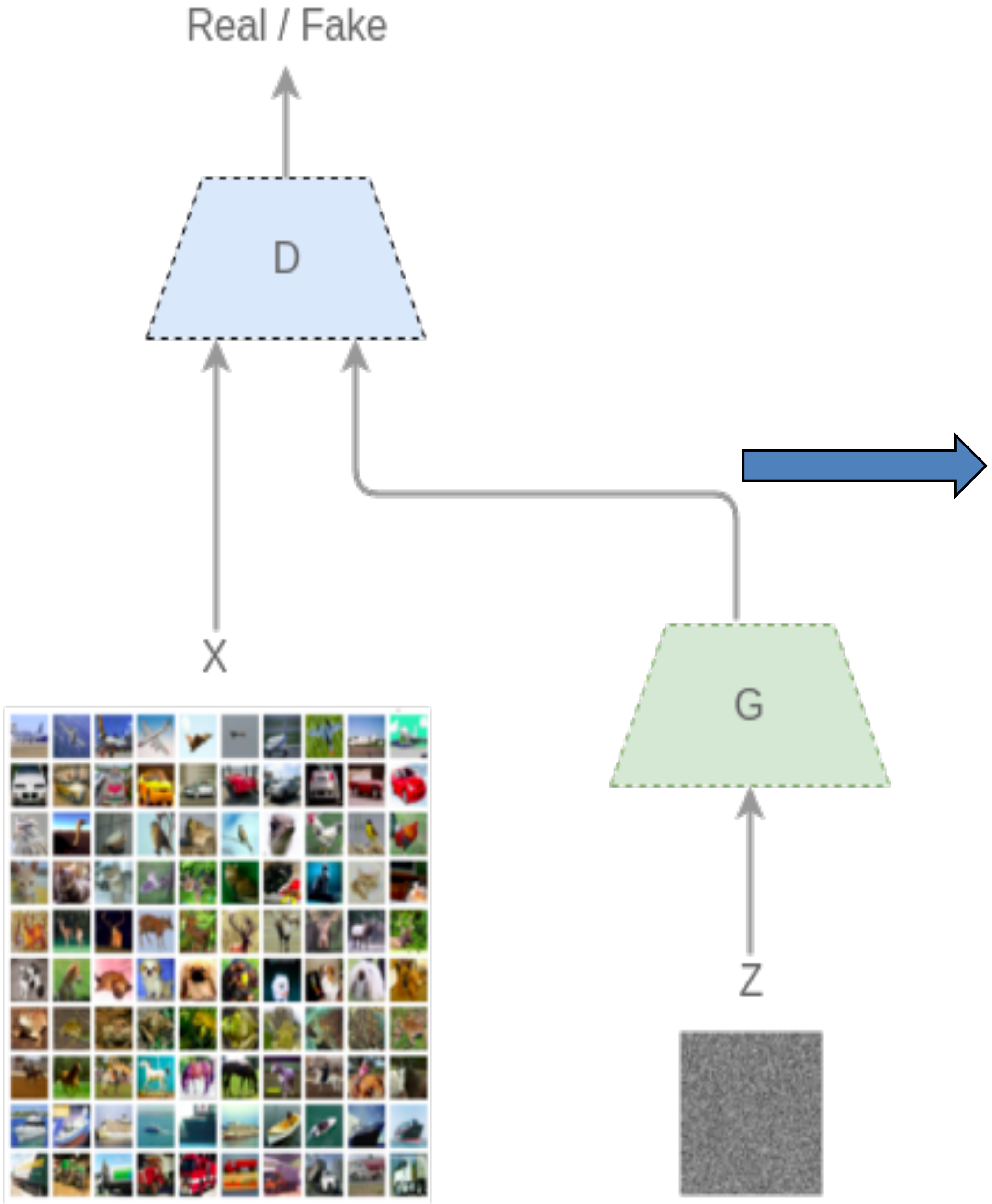
Not just ML, but a general trend:
Open Innovation

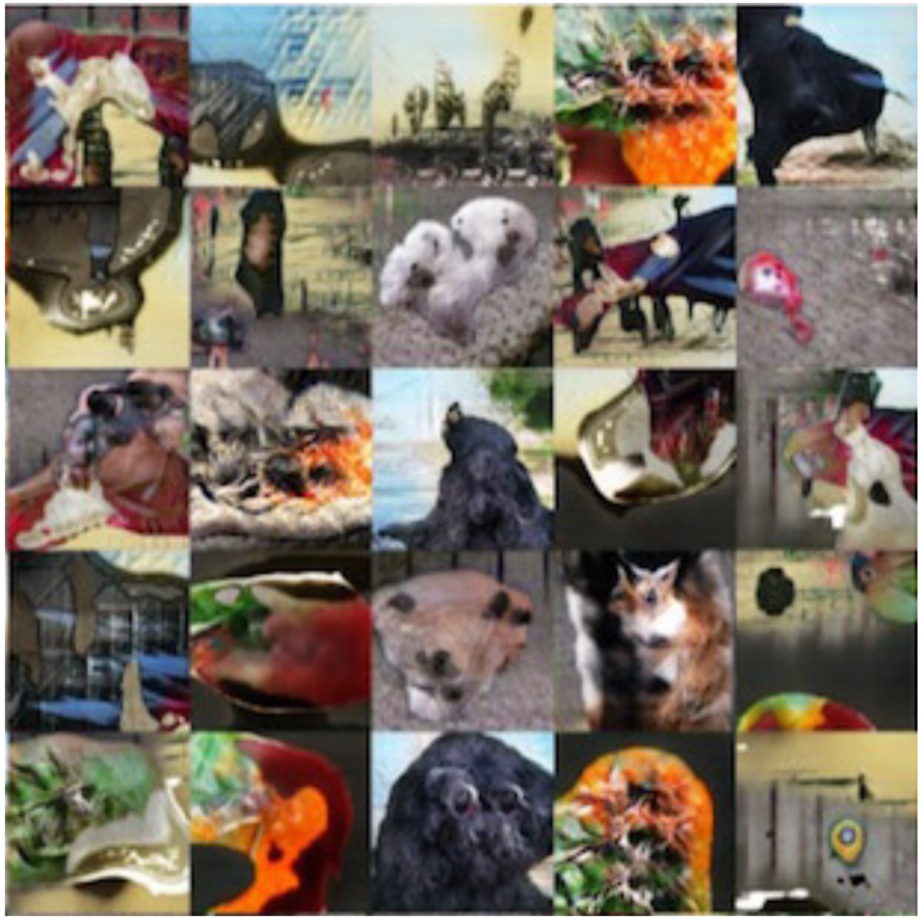
From domain to challenge and back



Generative Adversarial Network

Generative Adversarial Network





Condition GAN

Text to image

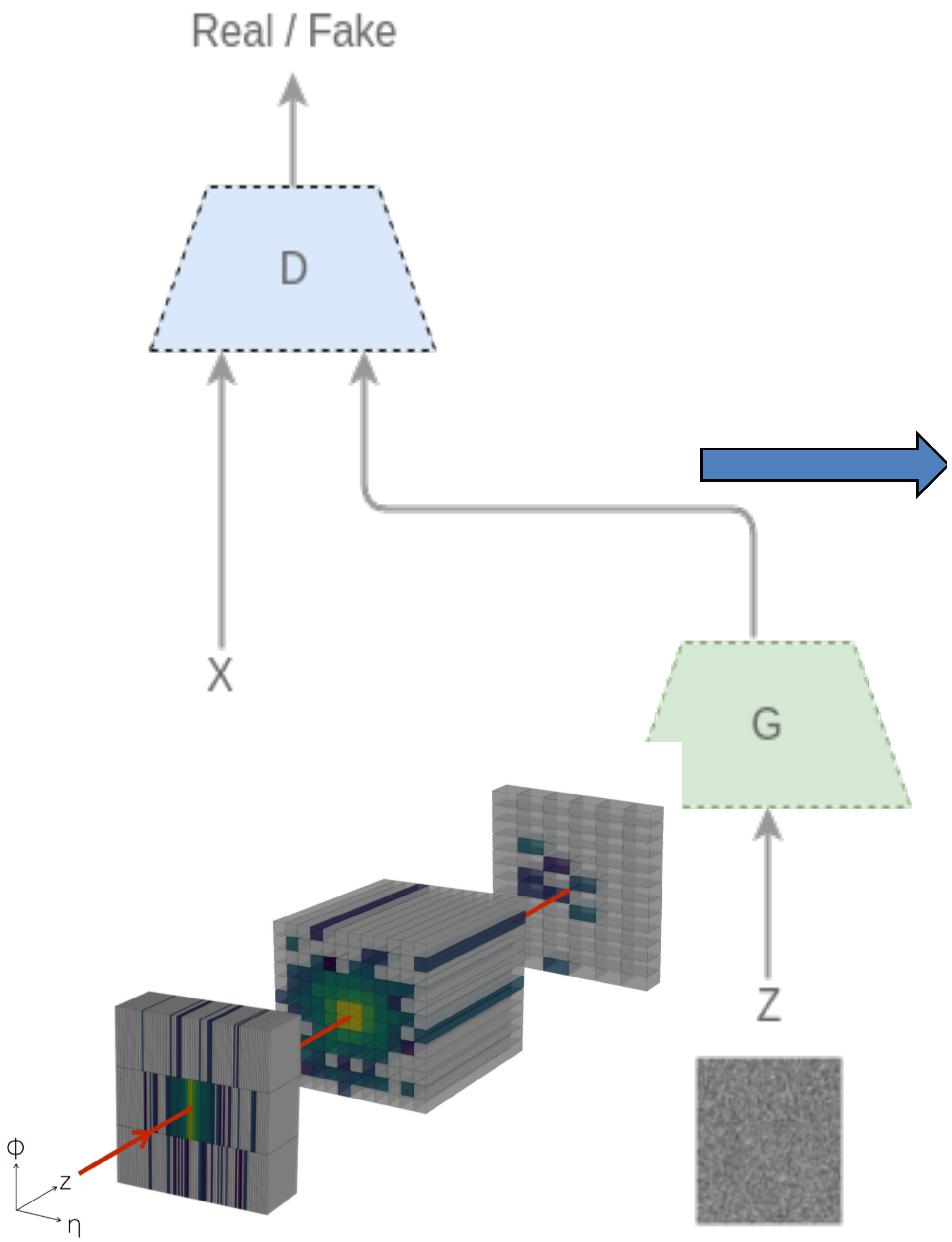
this small bird has a pink breast and crown, and black primaries and secondaries.

this magnificent fellow is almost all black with a red crest, and white cheek patch.

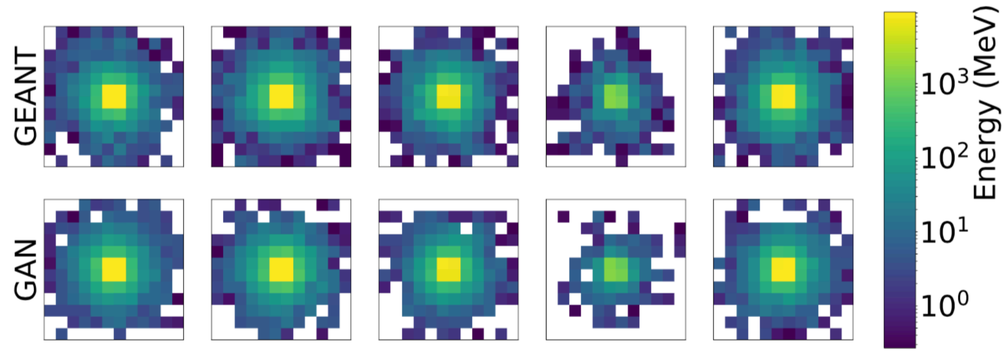
the flower has petals that are bright pinkish purple with white stigma

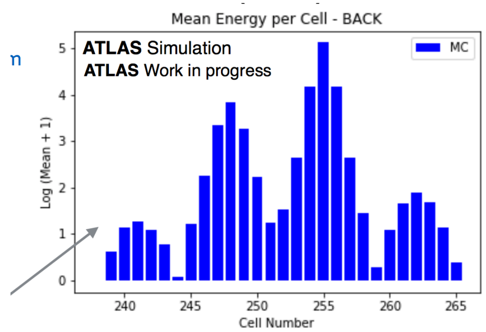
this white and yellow flower have thin white petals and a round yellow stamen

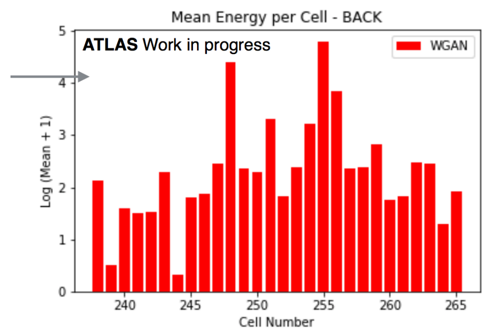
GAN for detector simulation



Gain en temps de calcul x1000







Beware of surprises

- https://twitter.com/goodfellow_ian/status/937406530743287808

Scientific Method

A typical scientific publication

Phys.Rev.Lett. 114 (2015)191803

Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments

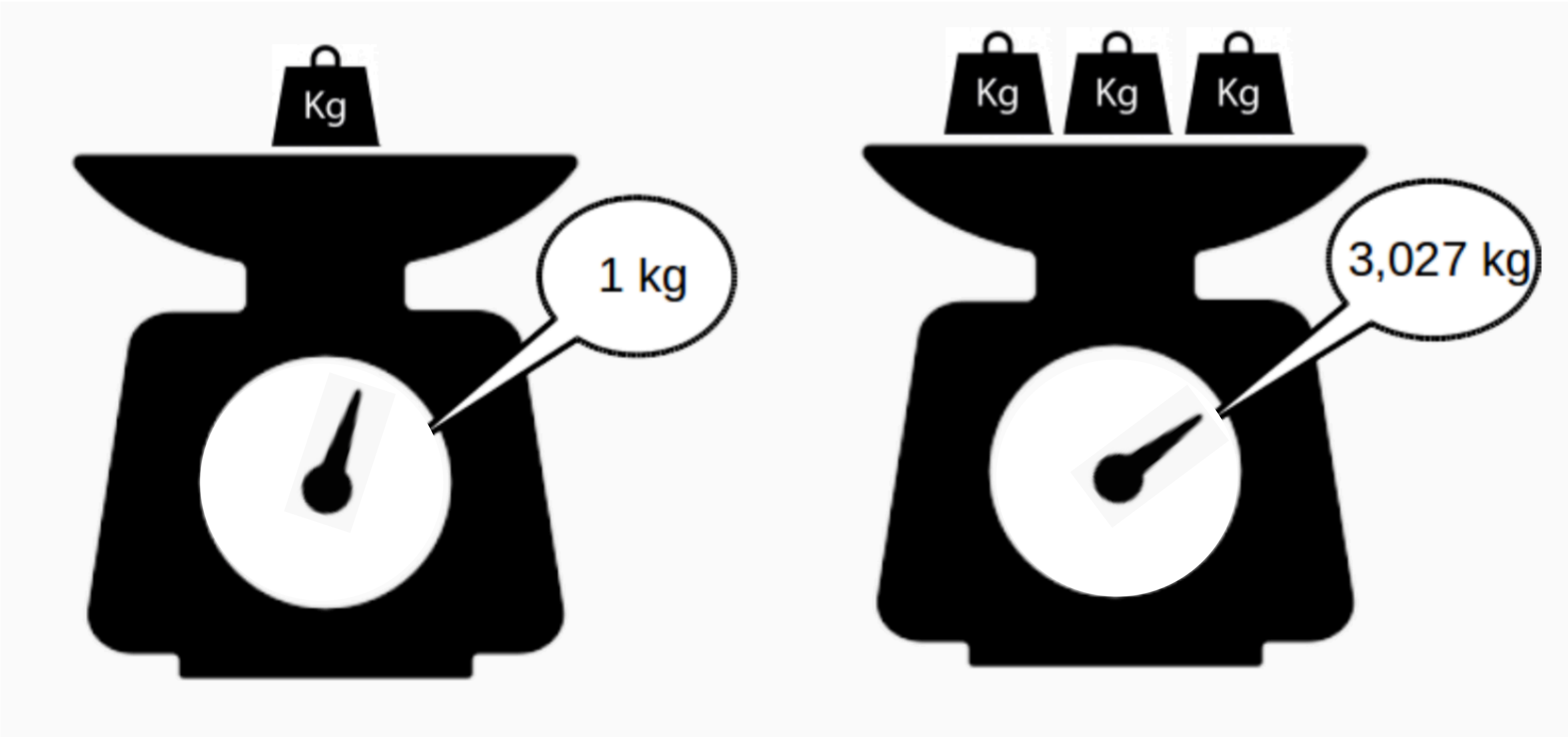
(ATLAS Collaboration)[†]
(CMS Collaboration)[‡]

(Received 25 March 2015; published 14 May 2015)

A measurement of the Higgs boson mass is presented based on the combined data samples of the ATLAS and CMS experiments at the CERN LHC in the $H \to \gamma\gamma$ and $H \to ZZ \to 4\ell$ decay channels. The results are obtained from a simultaneous fit to the reconstructed invariant mass peaks in the two channels and for the two experiments. The measured masses from the individual channels and the two experiments are found to be consistent among themselves. The combined measured mass of the Higgs boson is $m_H = 125.09 \pm 0.21\,(\text{stat}) \pm 0.11\,(\text{syst})$ GeV.

Systematic uncertainties
Everything we do not know

Experimental bias



About 500 terms for the systematical uncertainty

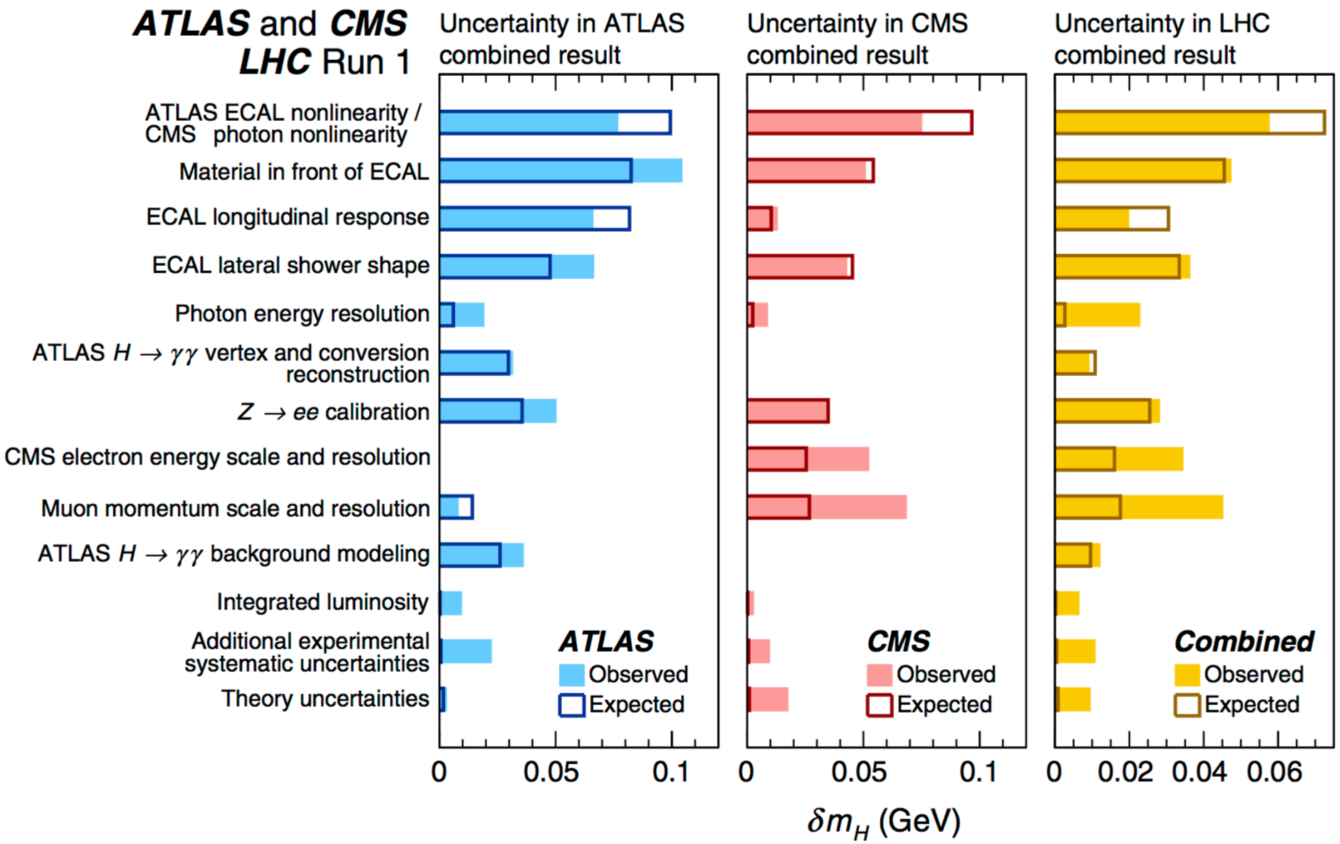


FIG. 3 (color online). The impacts δm_H (see text) of the nuisance parameter groups in Table I on the ATLAS (left), CMS (center), and combined (right) mass measurement uncertainty. The observed (expected) results are shown by the solid (empty) bars.

>>>89-90: Where is this measurement documented in this note?
**That has now been finished and published, and references will be updated in the notes
JHEP 09 (2014) 079 arXiv:1407.5532

>>>116: Doesn't "3 GeV" contradict your previous statement that we cover a pT range "from 1 GeV..."
**No, those muons with pT~1 GeV and p~3 GeV are at high eta. But this is all irrelevant, as we only use muons with pT>4 GeV.

>>>116: I'm not sure what you mean when you say that they "require" the combined reconstruction. I would just say "Muons used in this analysis are reconstructed using a statistical combination of an MS track and ID track."
**Only the muons that have a successful statistical combination of an MS track and ID track are used in this analysis, the text is updated.

>>>128: what is fully efficient for signal candidates? It's not clear what the word "which" refers to.
**This related to the vertex quality criteria applied to the fit.

>>>142: Please mention the total number of di-muon candidates here.
**This is added to the text, 7.8M for 2011, and ~65M for 2012.

>>>171: could you provide more information about how the psi is assigned to a primary vertex? I'm surprised by your statement that "few" events contain multiple vertices; I thought pileup was a significant issue at 7 TeV.
** Not really. The only relevance the primary vertex has is to measure Lxy, which is measured in the transverse direction only, and hence is not changing much from one collision vertex to another. However, the determination of the primary vertex position depends on whether the two muon tracks were used in its fit or not, hence we need to know which vertex was it. But at 7 TeV there was not much ambiguity, the vertex which jpsi came from was almost always the main primary one. Studies from the 2011 Jpsi Phi analysis -- where vertex choice may have been an issued-- showed that there was no impact in the few cases of an incorrect choice of vertex.

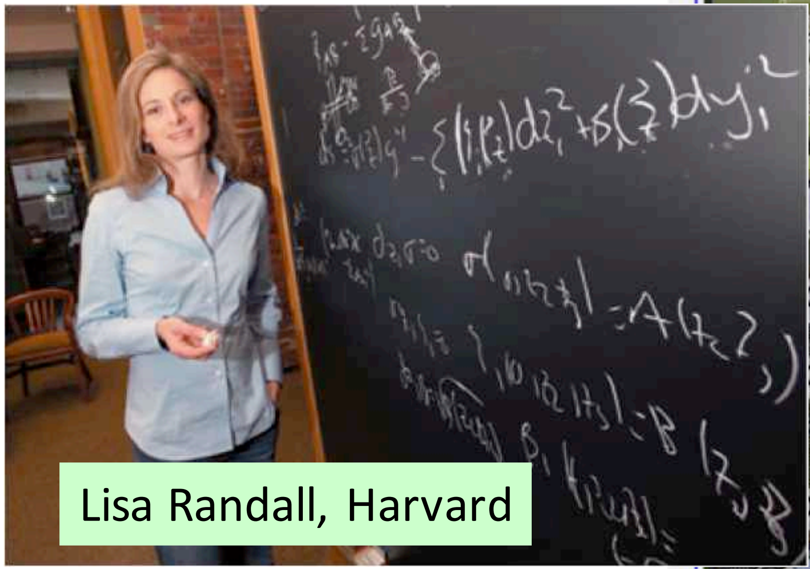
>>>Fig 1: Do you understand the eta dependence of this plot? What is the z-axis -- number of events per bin? The bins are extremely small -- what size bins did you choose?
** Yes. This is the scatter plot of dimuon candidates. The x-axis is the absolute rapidity "y" of the dimuon candidate, with the structure roughly reflecting the (smeared) geometry of the muon chambers, with dips near/around the cracks and edges.
The z-axis is just the candidate yield, the bin-width in |y| is 5e-3 and in pT is 320 MeV

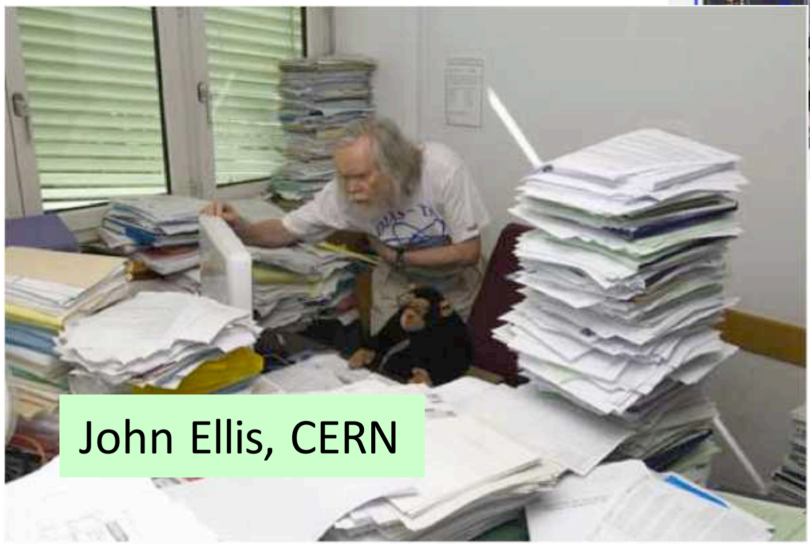
>>>228: The phi* definition is unclear to me, in particular what the "psi production" is. Do you mean the psi momentum vector?
**phi* is the angle between the psi production plane (defined by psi momentum and colliding proton

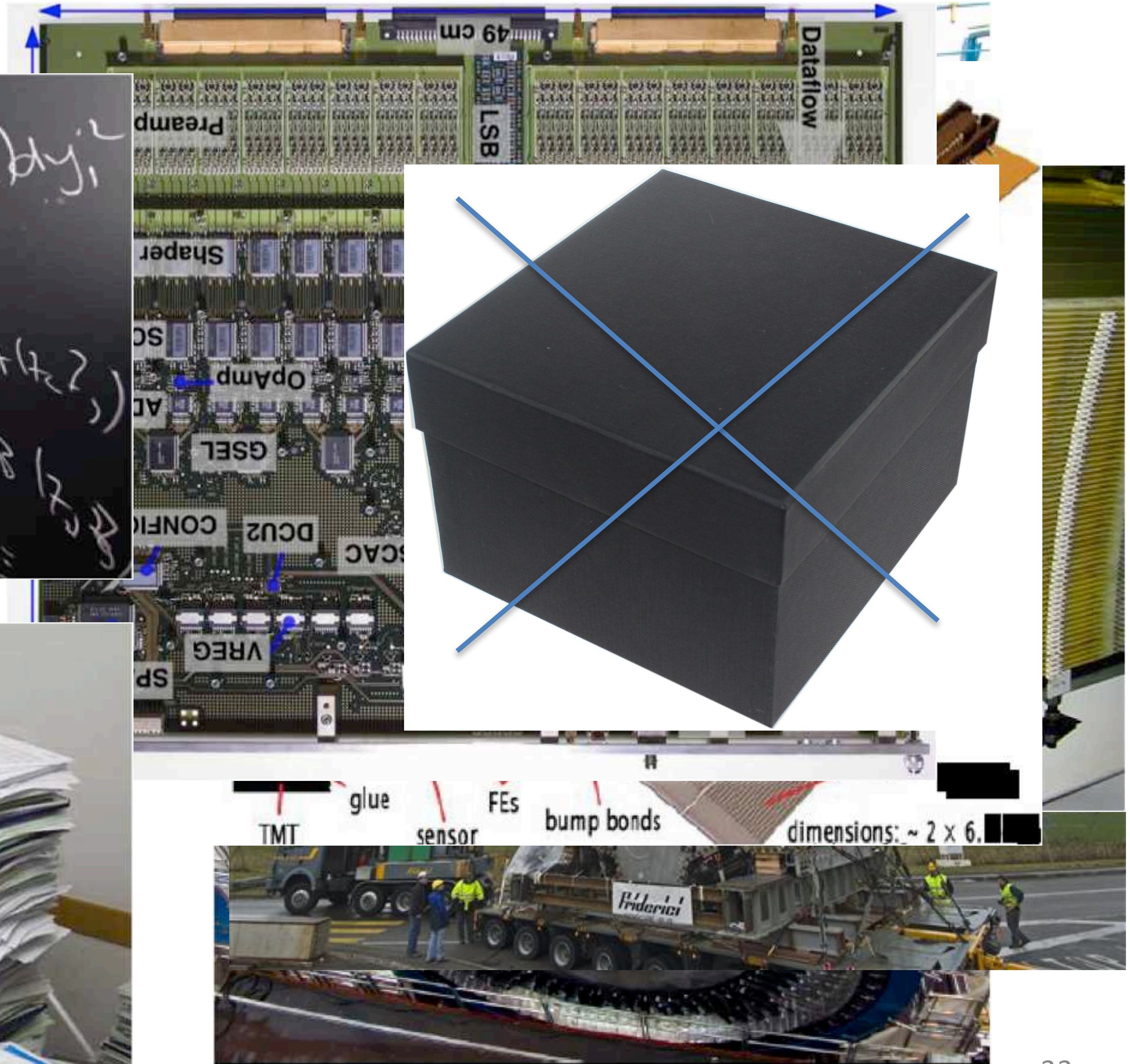
ue

200 pages and weeks of discussion
In the collaboration

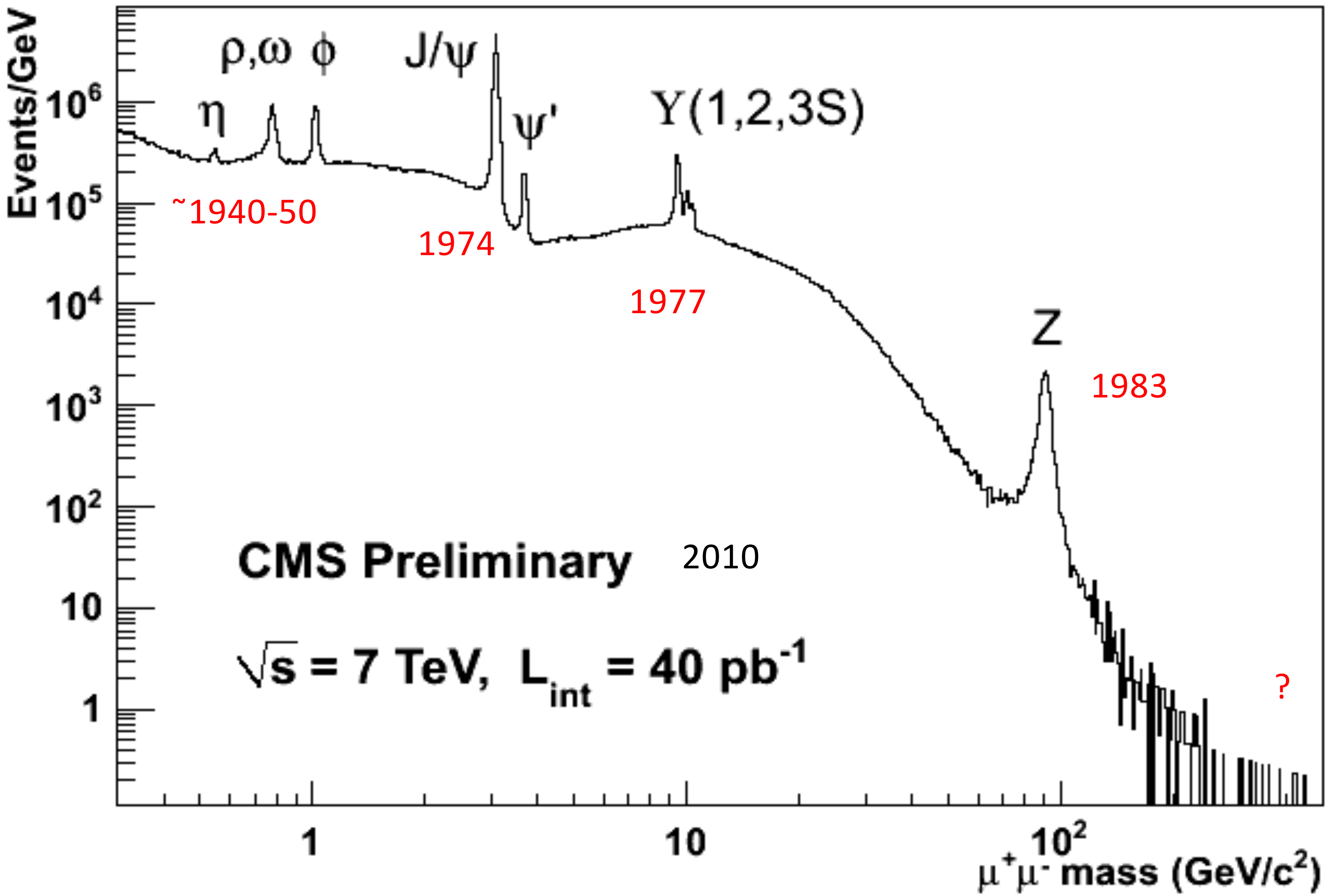
Trust but verify, from theory to experiment


Lisa Randall, Harvard


John Ellis, CERN



Re-discovery of particles



ML and systematical errors

- Paper conclusion end with:
  - measurement = m ± σ(stat) ± σ(syst)
  - σ(syst) systematical uncertainty
- We need to minimise the total error:

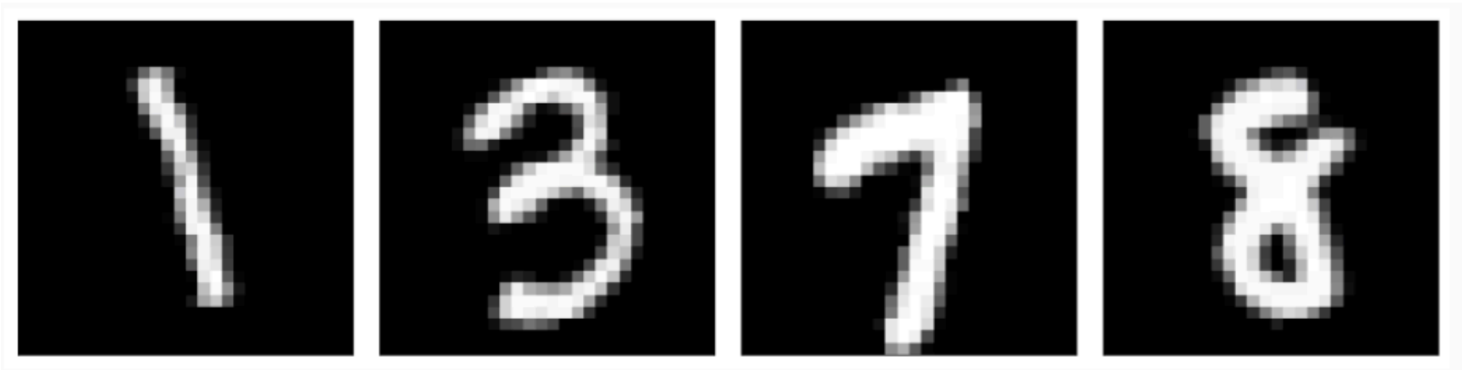
$$\sigma(\text{stat}) \pm \sigma(\text{syst})$$

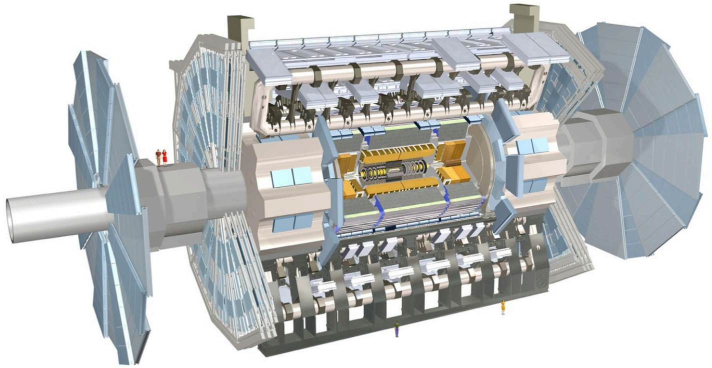
- Standard ML minimises σ(stat)

❑ How to tell ML to minimise σ(stat) ± σ(syst) ?

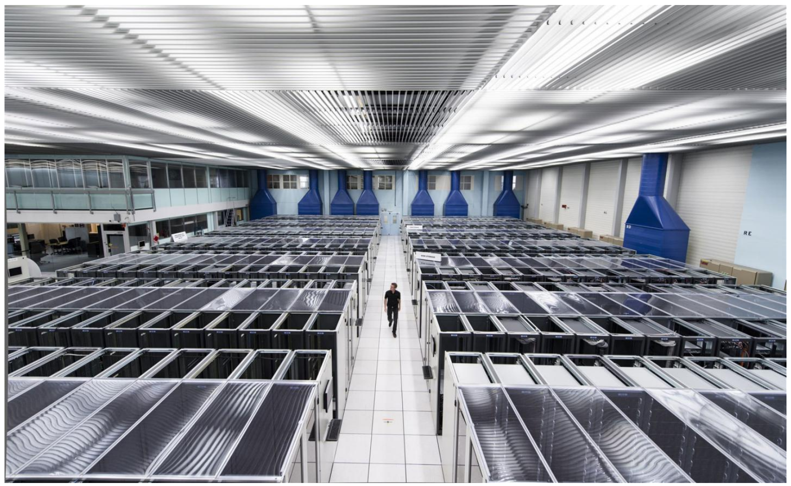
Systematical effect

Example of impact of the angle on handwritten digits





Data

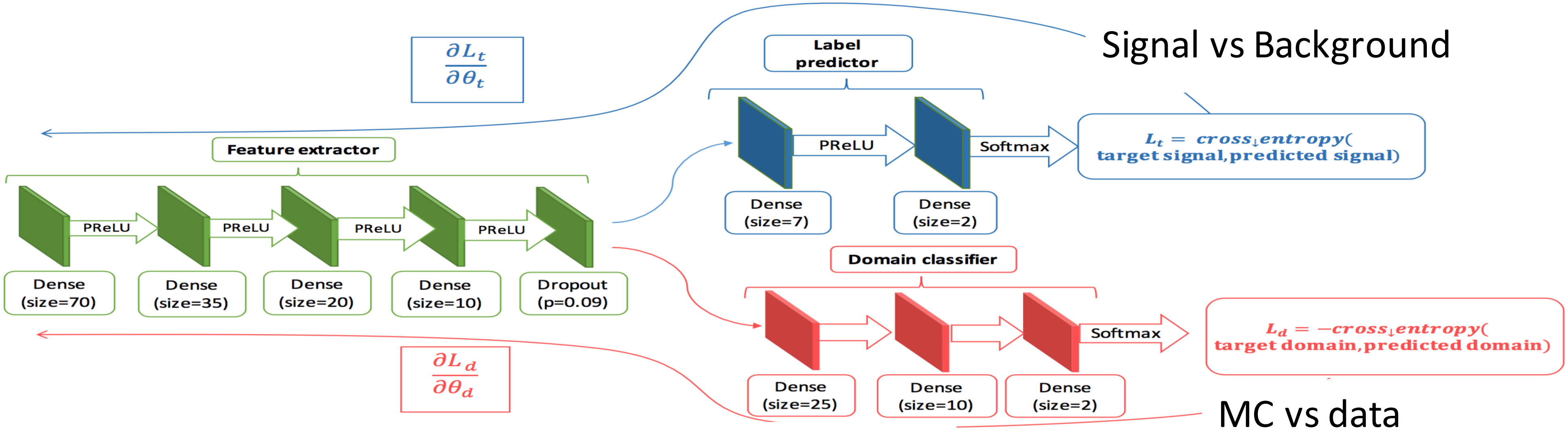


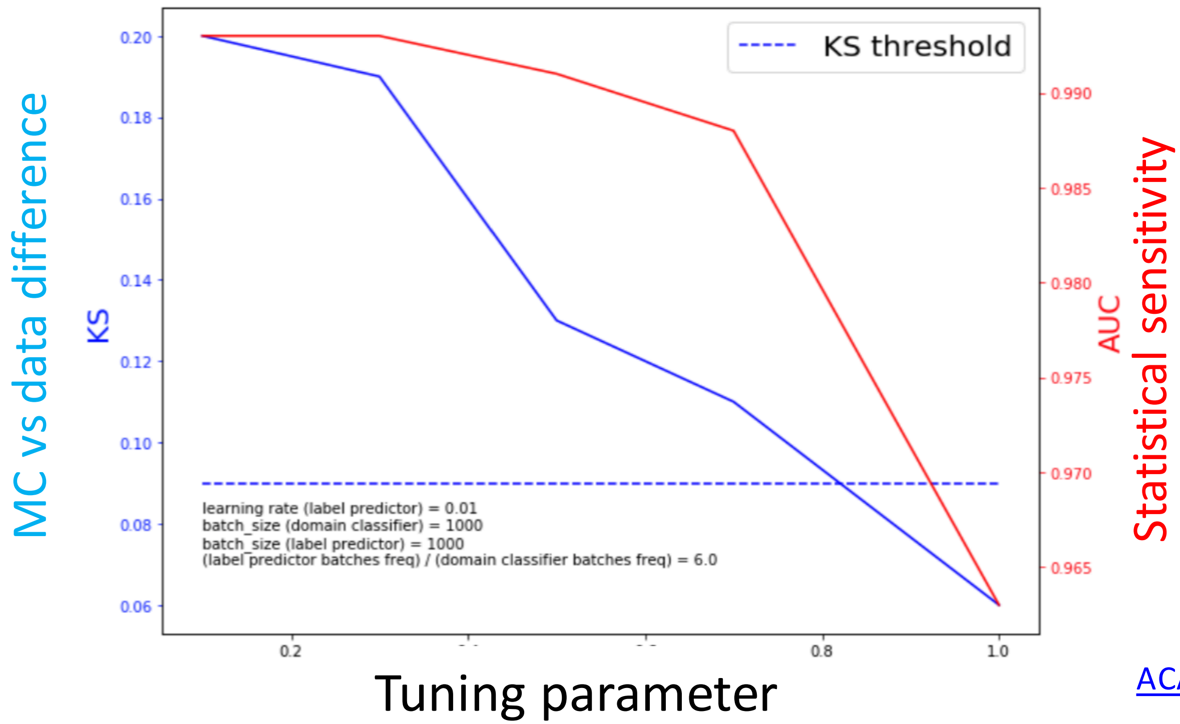
Simulation

Can we trust the simulation?

Adversarial training

Inspired from 1505.07818 Ganin et al :





ACAT 2017 Ryzhikov and Ustyuzhanin

Style Transfert

CycleGAN

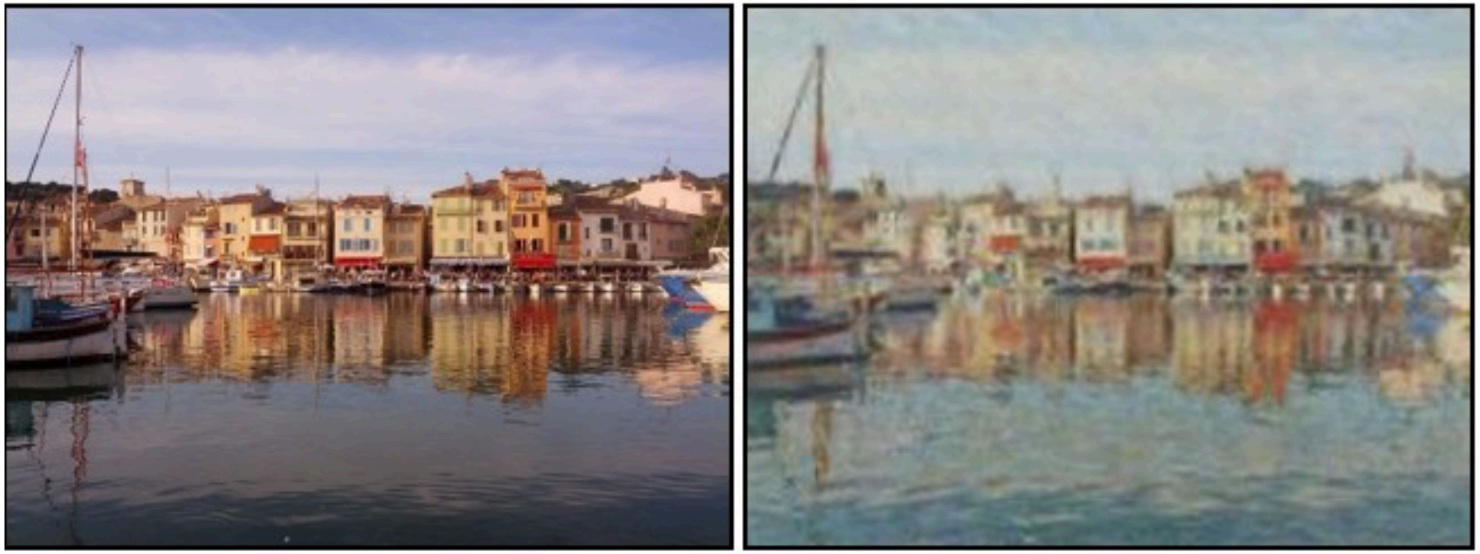


photo ⟶ Monet

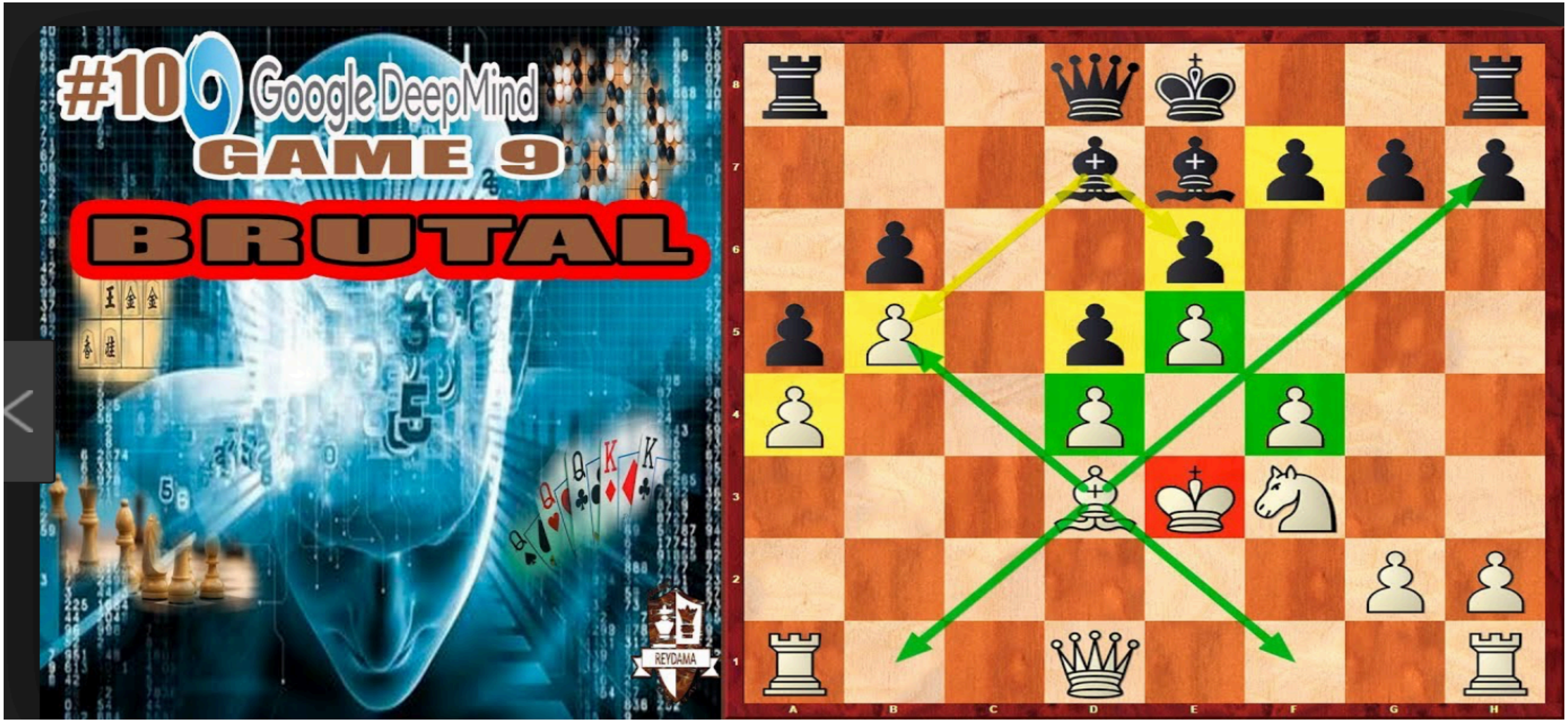
Monet ⟶ photo

CycleGAN

CycleGAN

Reinforcement learning

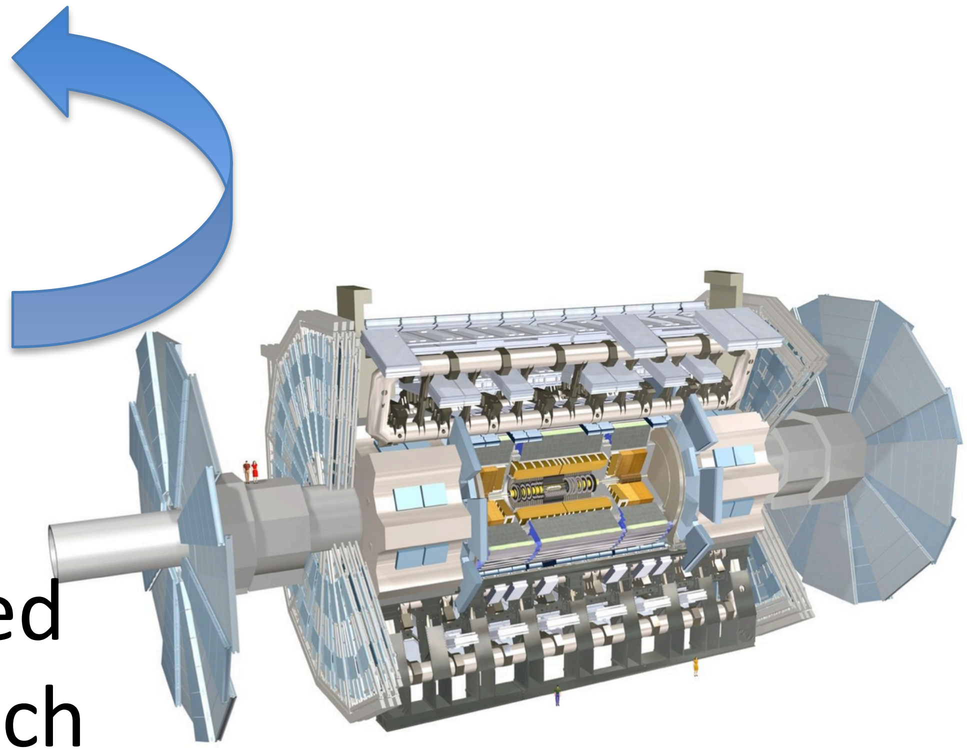
Reinforcement learning



Alpha Zero : starting only from the rules, learn on its own by playing against himself in a few days to play (separately) Go, chess, or japanese chess, and beats everyone, men and computers

HEP Application?

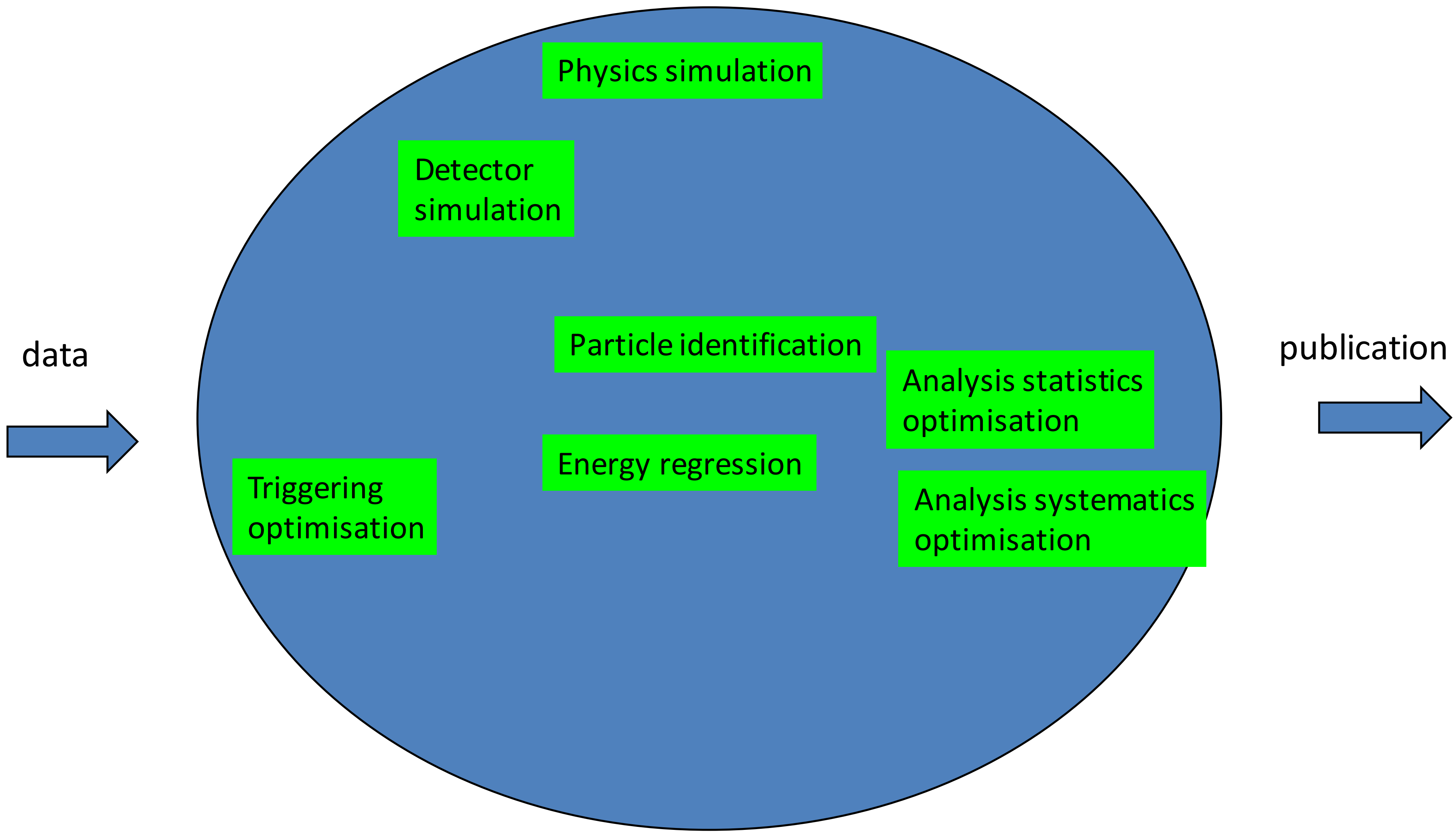
- Experiment design
- Data taking
- Success/failure?



➔In practice, could be used to optimise triggering, which can easily be virtualised

Finally…

Machine Learning playground



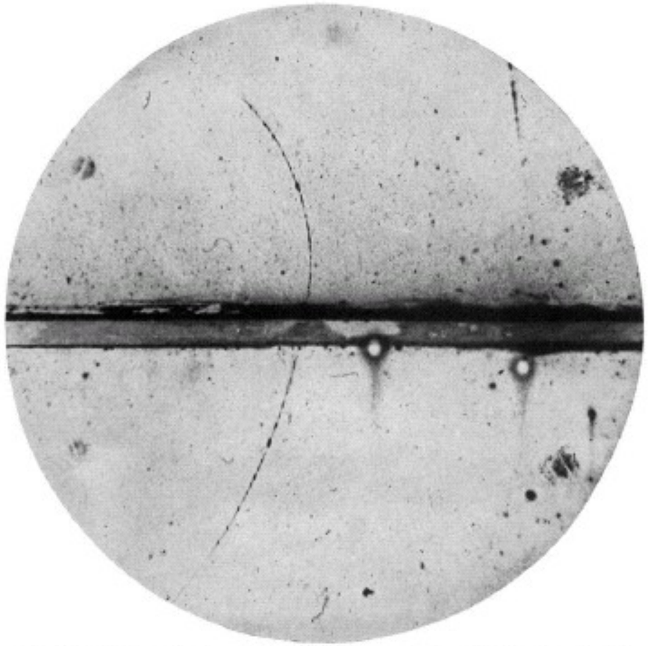
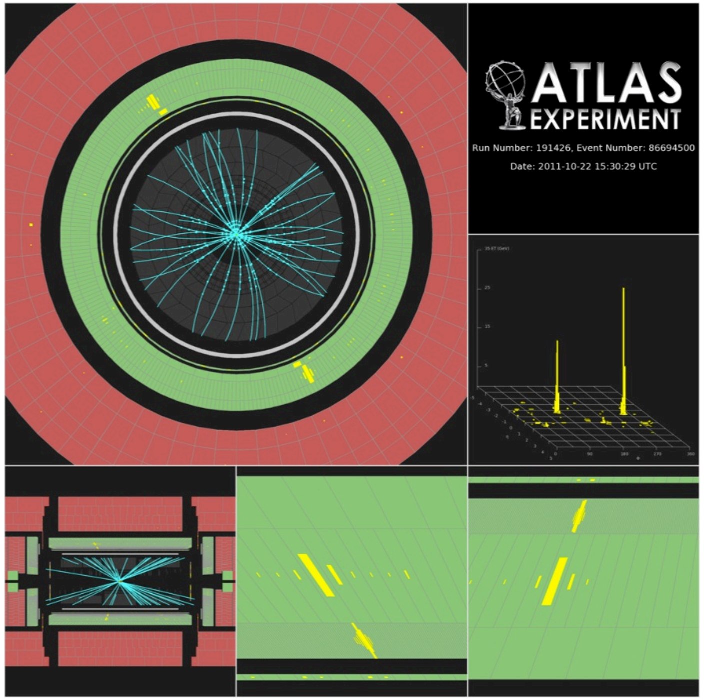
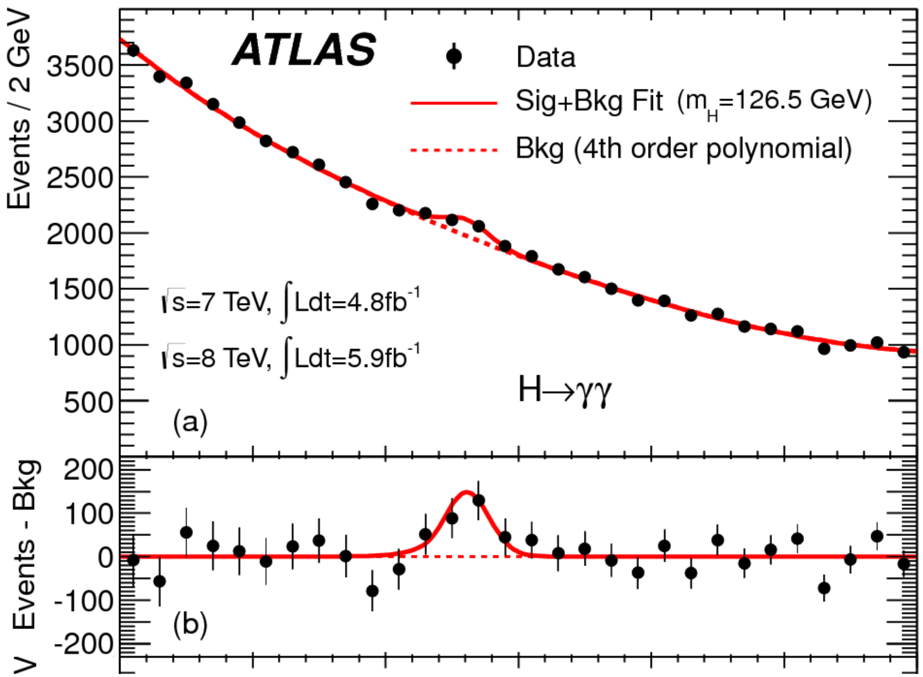
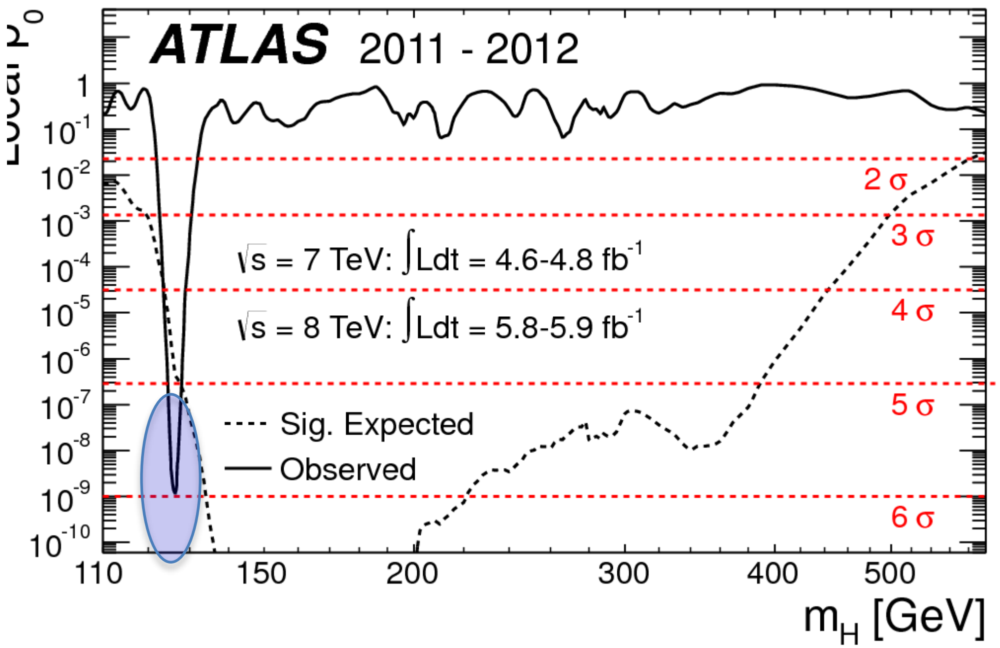


FIG. 1. A 63 million volt positron ($H\rho = 2.1 \times 10^5$ gauss-cm) passing through a 6 mm lead plate and emerging as a 23 million volt positron ($H\rho = 7.5 \times 10^4$ gauss-cm). The length of this latter path is at least ten times greater than the possible length of a proton path of this curvature.







Conclusion

- Machine Learning/Artificial Intelligence : lots of promises for HEP (and other science actually)
- Powerful algorithms…
- … on more and more playgrounds
- However not trivial, not plug and play..
- …will take time