The High Energy Physics Tracking Machine Learning challenge

David Rousseau (LAL) (rousseau@lal.in2p3.fr),

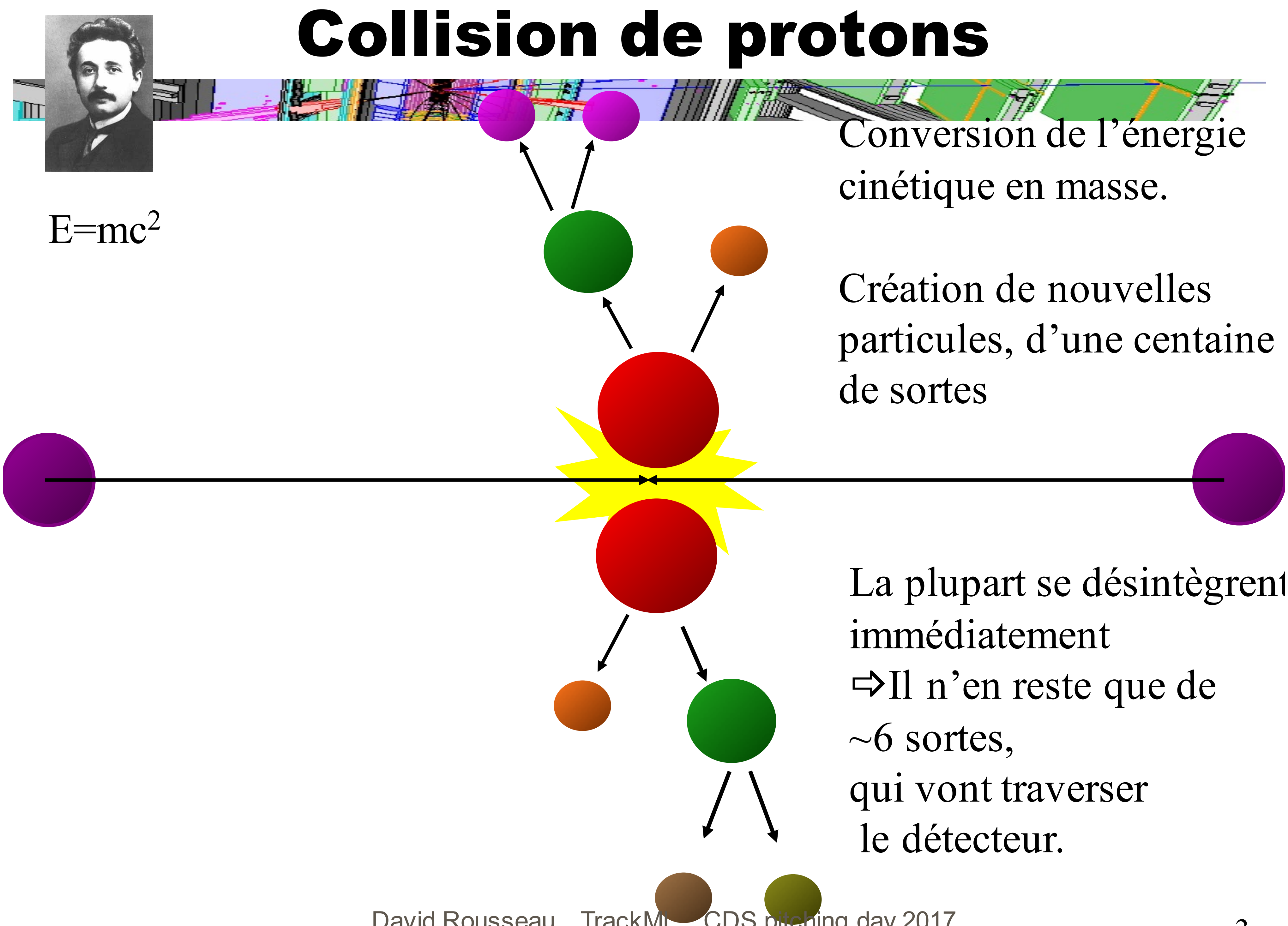
Cécile Germain (LAL/LRI) , Isabelle Guyon (Chalearn/LRI)

with Paolo Calafiura, Steven Farrell, Heather Gray (LBNL-Berkeley), Jean-Roch Vlimant (CalTech), Vincenzo Innocente, Andreas Salzburger (CERN), Tobias Golling, Moritz Kiehn, Sabrina Amrouche (U Geneva), Vava Gligorov (LPNHE-Paris), Mikhail Hushchyn, Andrey Ustyuzhanin (Yandex) ...

CDS pitching day 8th Nov 2017

LHC purpose in a nutshell

Collision de protons



E=mc^2

Conversion de l’énergie cinétique en masse.

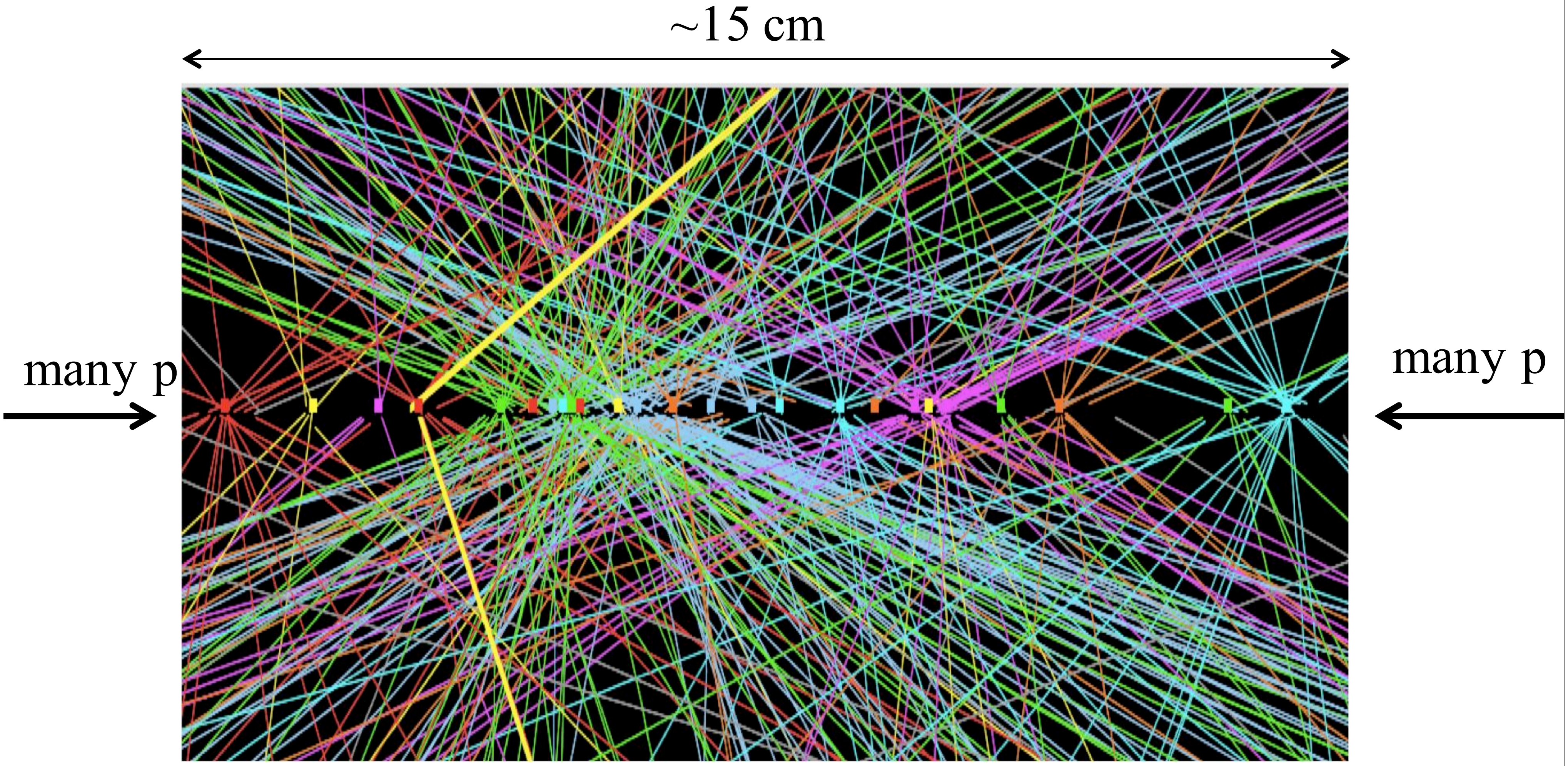
Création de nouvelles particules, d’une centaine de sortes

La plupart se désintègrent immédiatement

⇨Il n’en reste que de ~6 sortes,

qui vont traverser le détecteur.

Bunch collision



Situation actuelle : 20aine de collision parasites

HL-LHC : facteur 10

Spécial Festival d'Avignon

La 66e fête du théâtre démarre le 7 juillet

Le Monde

Science : la matière dévoilée

Le boson de Higgs, particule manquante pour expliquer l'Univers, vient d'être découvert

Les physiciens du CERN de Genève ont prouvé son existence

Libération

Physique des particules

La masse est dite

Le Cern a réussi à mettre en évidence le boson de Higgs qui résout une énigme fondamentale et ouvre une nouvelle étape scientifique. PAGES 2-5

Les derniers feux des pharaons

Au musée Jacquemart-André, à Paris, une exposition passionnante s'attarde sur la période tardive de l'antiquité égyptienne, souvent oubliée. PAGES 14-15

Suicides chez France Télécom : l'ancien patron mis en examen

Didier Lombard, qui dirigeait l'opérateur téléphonique lors de la vague de suicides ayant touché l'entreprise en 2008 et 2009, est visé par une enquête de la justice pour harcèlement moral PAGE 14

A nos lecteurs

En raison d'un mouvement de grève dans les imprimeries, ce numéro de Libération n'est disponible que sous sa forme électronique. Toutes nos excuses à nos lecteurs.

U.S. Edition ▾

The New York Times

Wednesday, July 4, 2012 Last Update: 4:00 AM ET

IGITAL SUBSCRIPTION: 4 WEEKS FOR

Pool photo by Denis Balibouse

New Particle Could Be Physics' Holy Grail

OPINION

EDITORIA

Too Qui Health

The Oba forcefull Republic the refor

MARKE

Britain

FTSE 100

5,673.04

−14.69

−0.26%

GET QUO

EL PAÍS

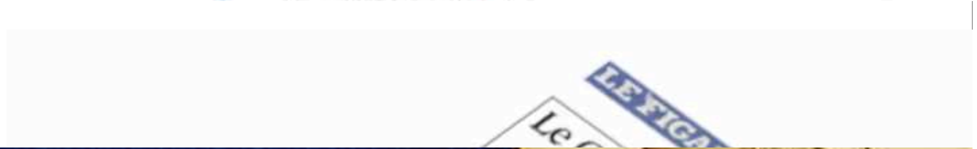
ESTÁ PASANDO

DIRECTO

INTERNACIONAL POLÍTICA ECONOMÍA CULTURA SOCIEDAD DEPORTES

Registrarse

Hallada "la más sólida evidencia" de la existencia del bosón de Higgs



LE FIGA

2013 NOBEL PRIZE IN PHYSICS

François Englert

Peter W. Higgs

© The Nobel Foundation. Photo: Lovisa Engblom.

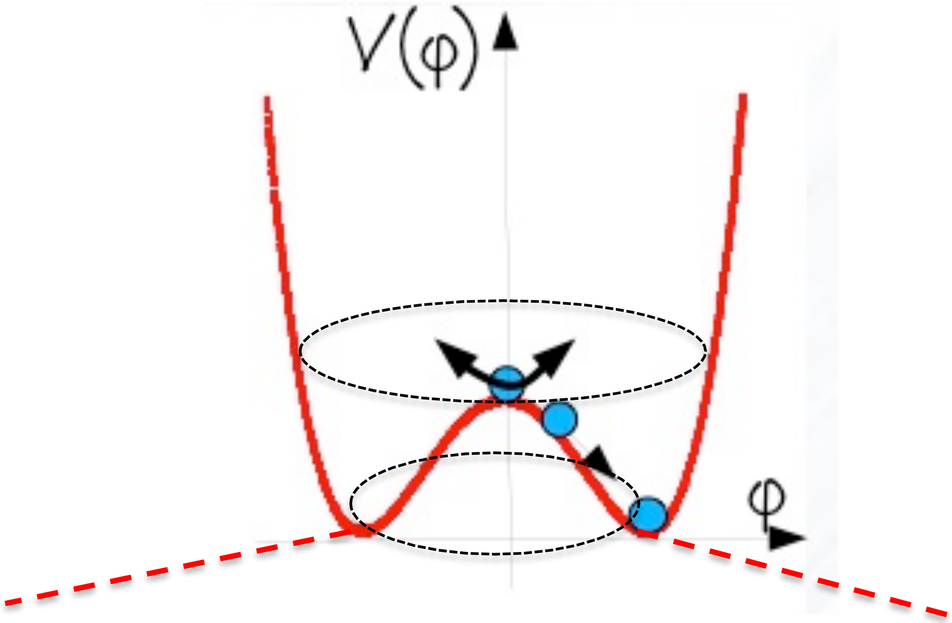
Future of LHC beyond Higgs boson discovery

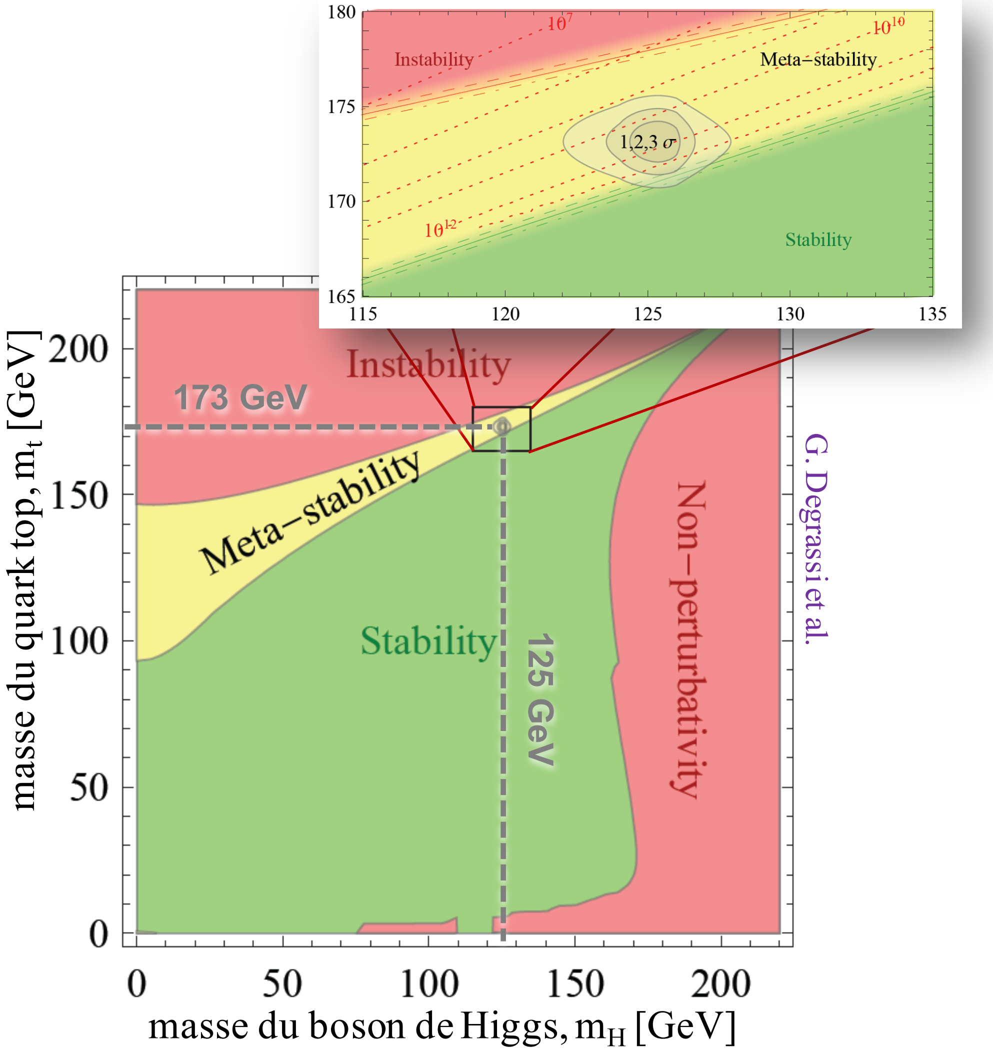
CDS pitching day 9th Nov 2016

L'Univers est-il stable ?

La stabilité du vide dépend des masses du boson de Higgs et du quark top

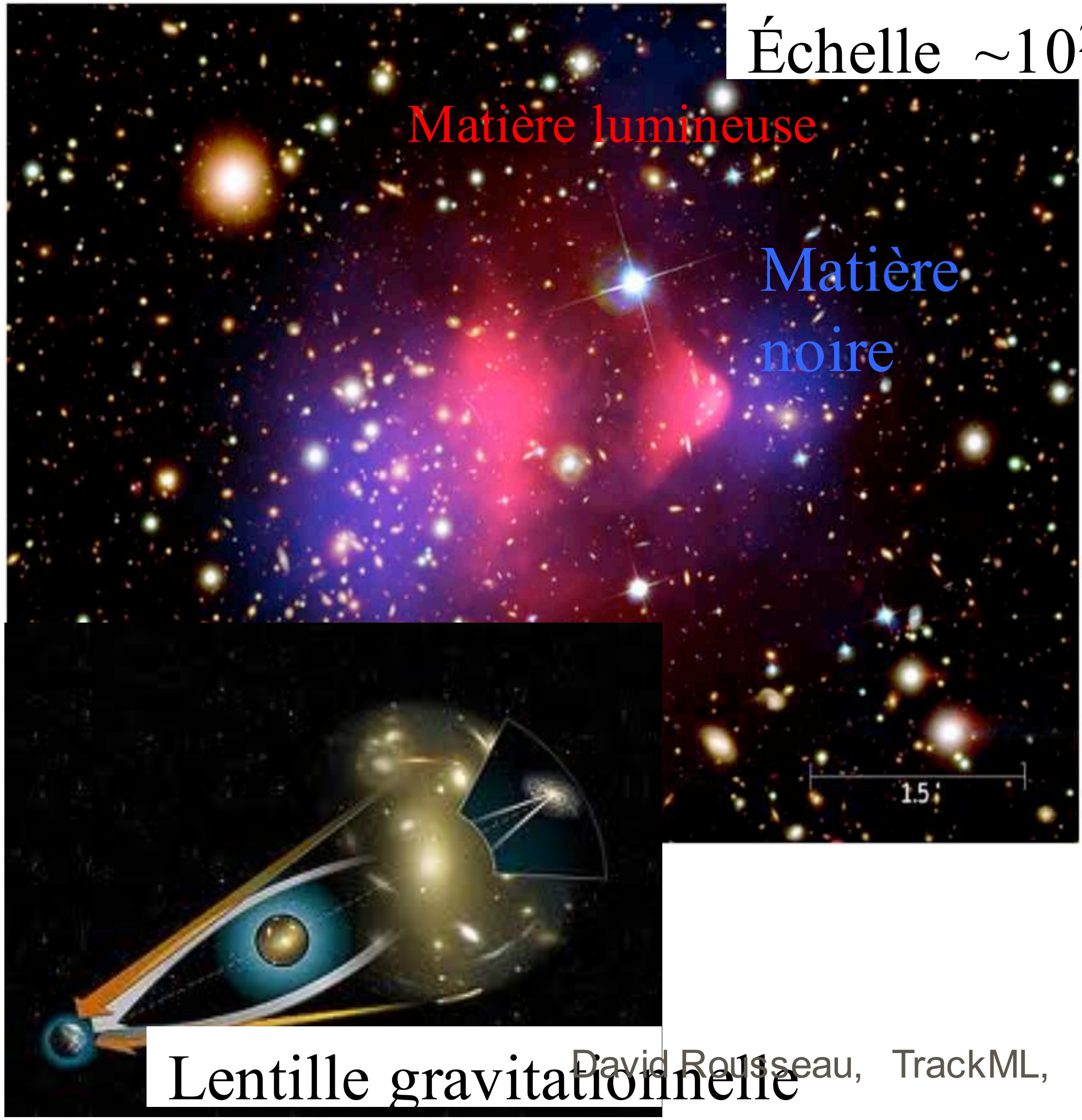
Notre Univers vit au bord du pre





“Physique des deux infinis

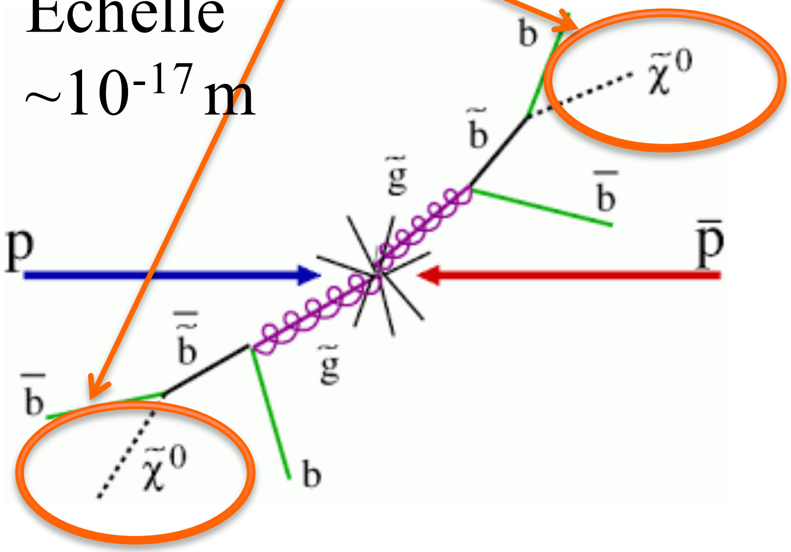
Échelle ~10^{22} m



Lentille gravitationnelle

Matière noire

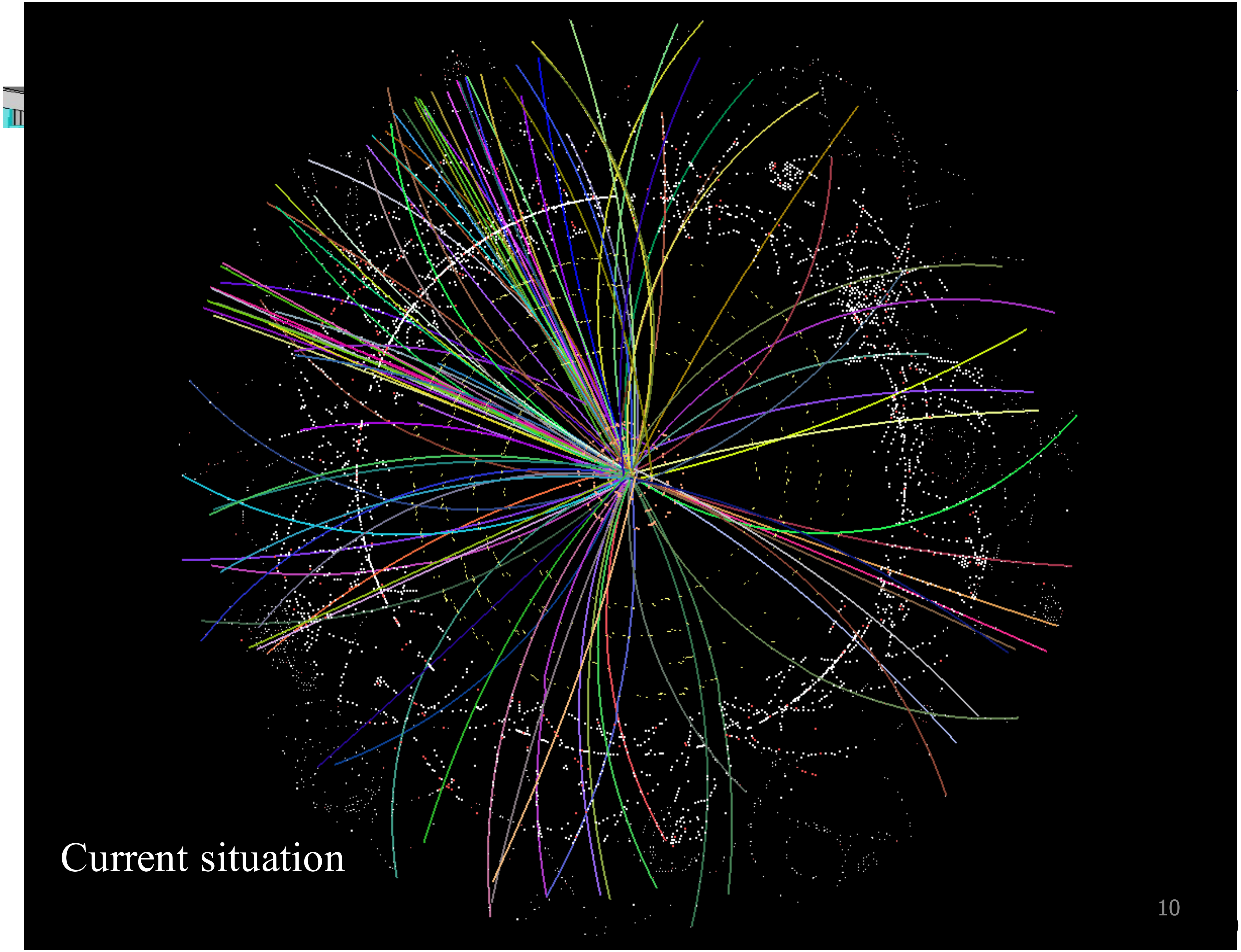
Échelle ~10^{-17} m



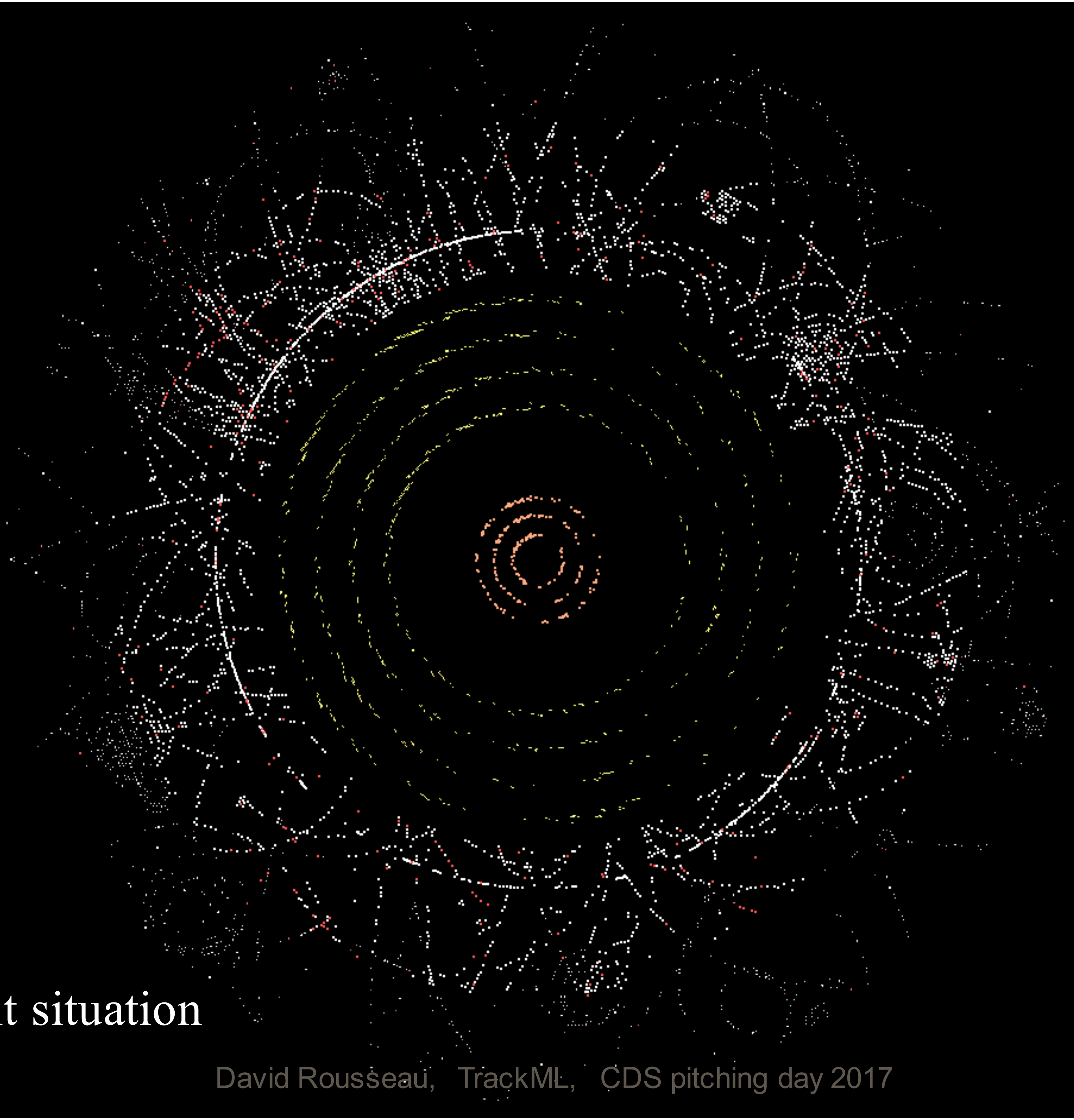
How ?
➔HL-LHC, increase LHC Luminosity by 10 in 2025

Tracking challenge

CDS pitching day 9th Nov 2016



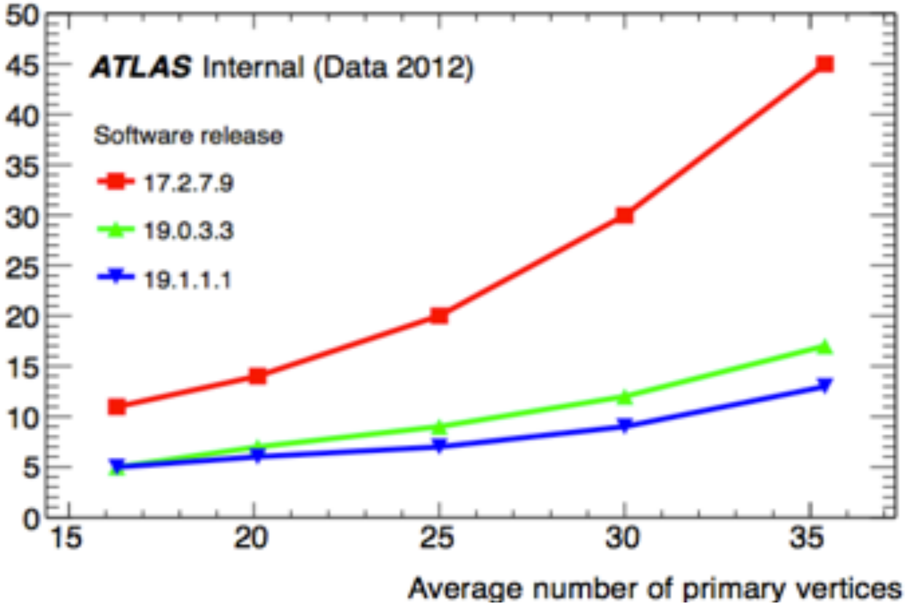
Current situation

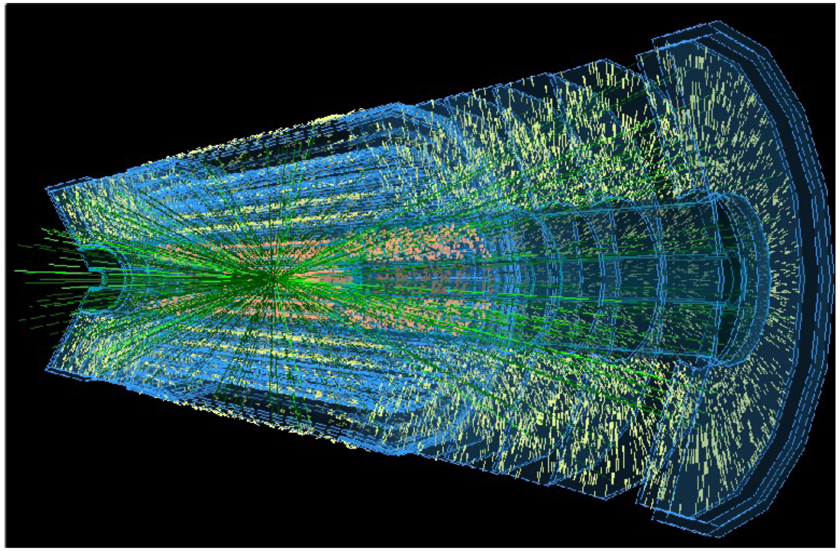


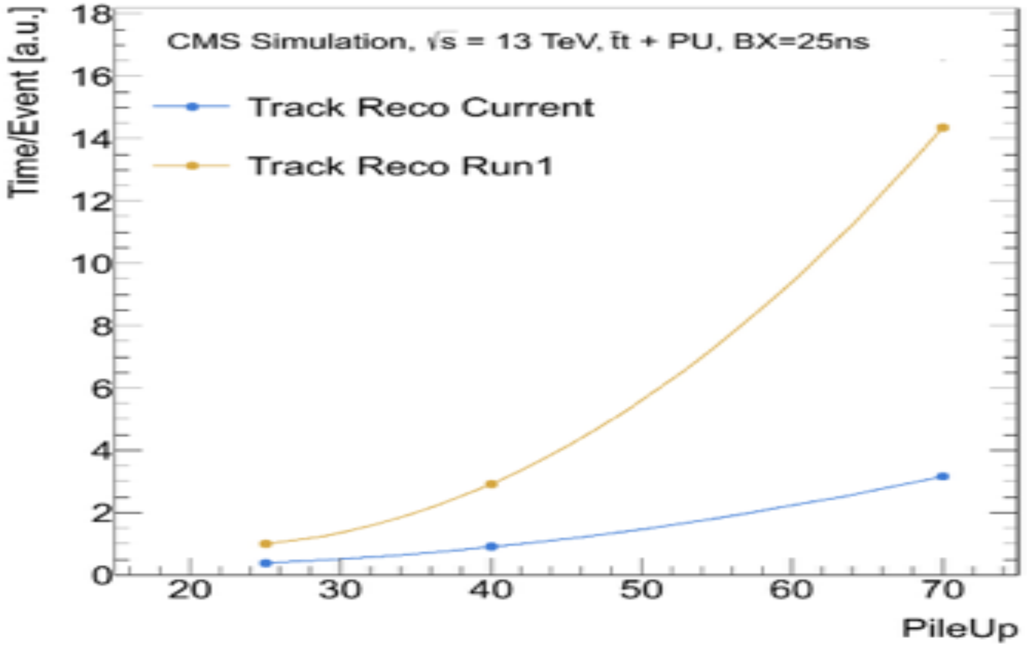
Current situation

Motivation 1

- Tracking (in particular pattern recognition) dominates reconstruction CPU time at LHC
- HighLumi-LHC perspective : increased rate of parasitic collisions
  - Run 1 (2010-2012): <>~20
  - Run 2 (2015-2018): <>~30
  - Phase 2 (2025): <>~150
- CPU time of current software quadratic/exponential extrapolation (difficult to quote any number)
- (but current software give reasonably good results, but too slow)





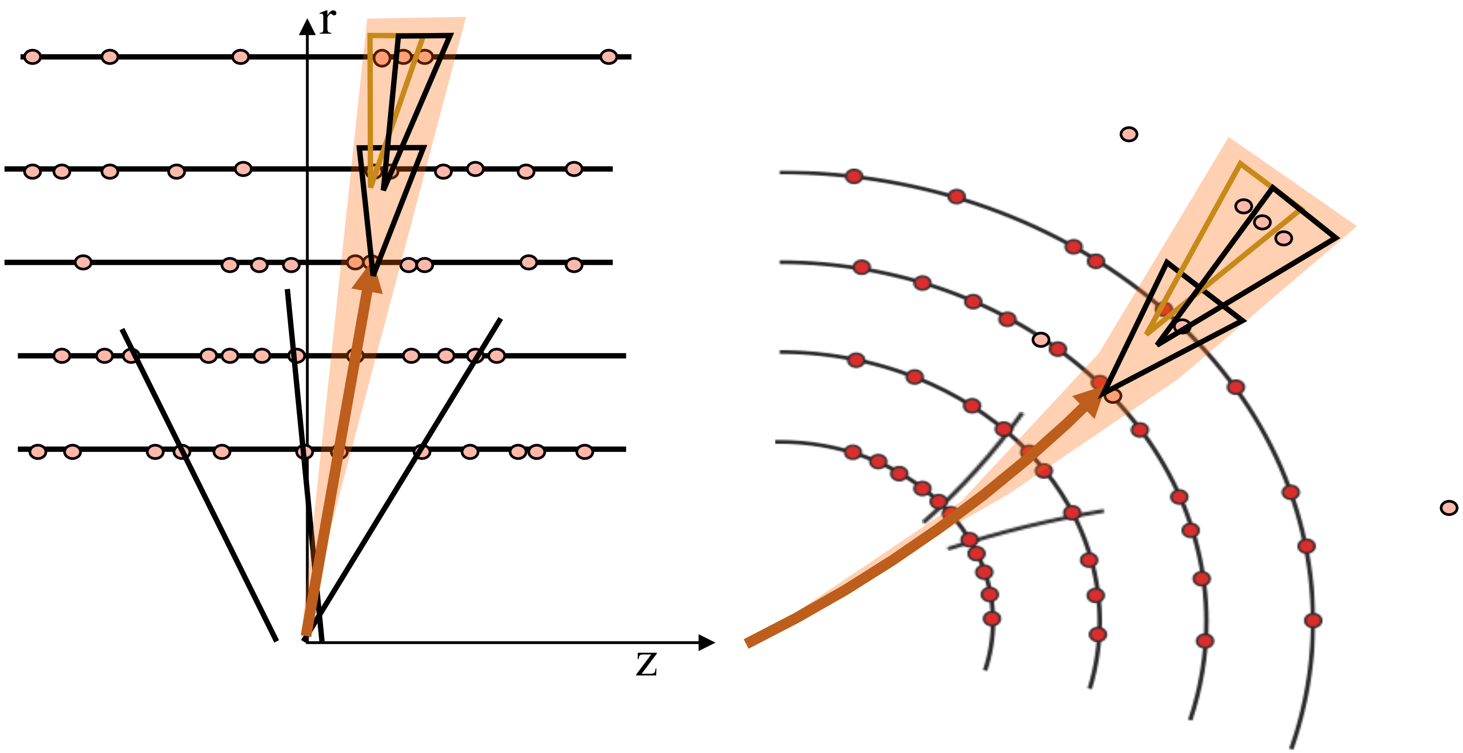


Motivation 2

- LHC experiments future computing budget flat (at best) (LHC experiments use 300.000 CPU cores worldwide)
- Installed CPU power per $==€==CHF expected increase factor <10 in 2025
- Experiments plan on increase of amount of data recorded (by a factor ~10)
- ➔HighLumi reconstruction to be as fast as current reconstruction despite factor 10 in complexity
- ➔requires very significant software CPU improvement, factor ~10
- Large effort within HEP to optimise software and tackle micro and macro parallelism, likely not enough
- >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 (i.e. Convolutional NN)
- Need to engage a wide community to tackle this problem

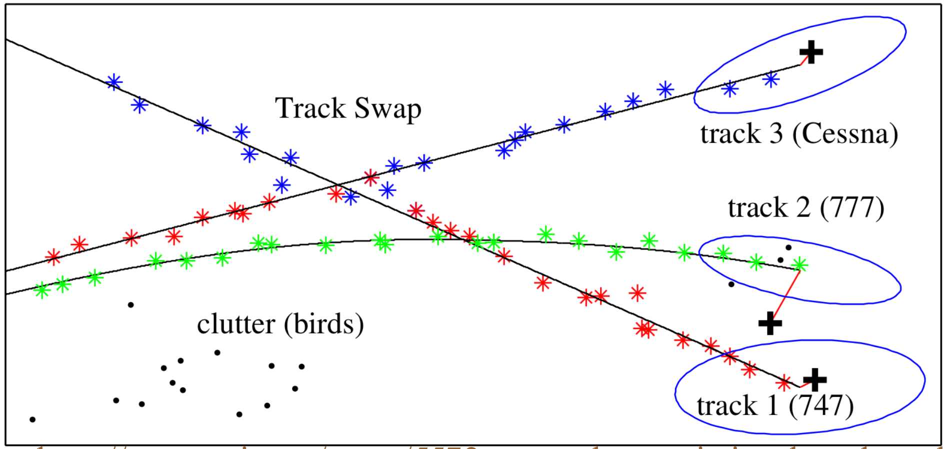
Curent Algorithm

- Pattern : connect 3D points into tracks
- Essentially combinatorial approach
- Tracks are (not perfect) helices pointing (approximately) to the origin
- Challenge : explore completely new approaches
- (not part of the challenge : given the points, estimate the track parameters)

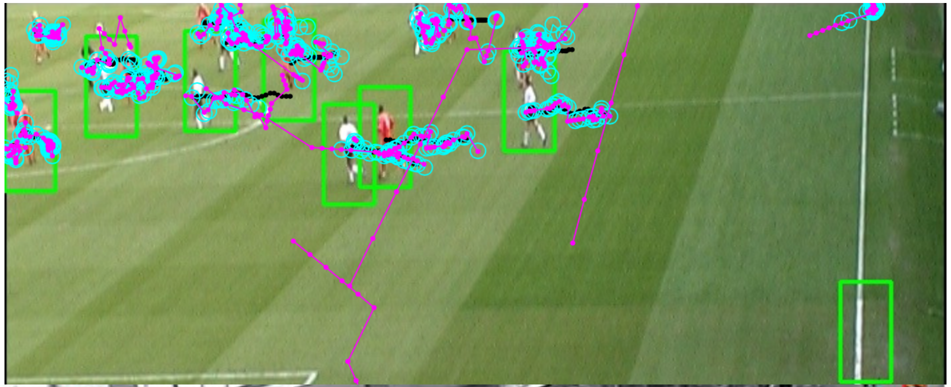


Pattern recognition

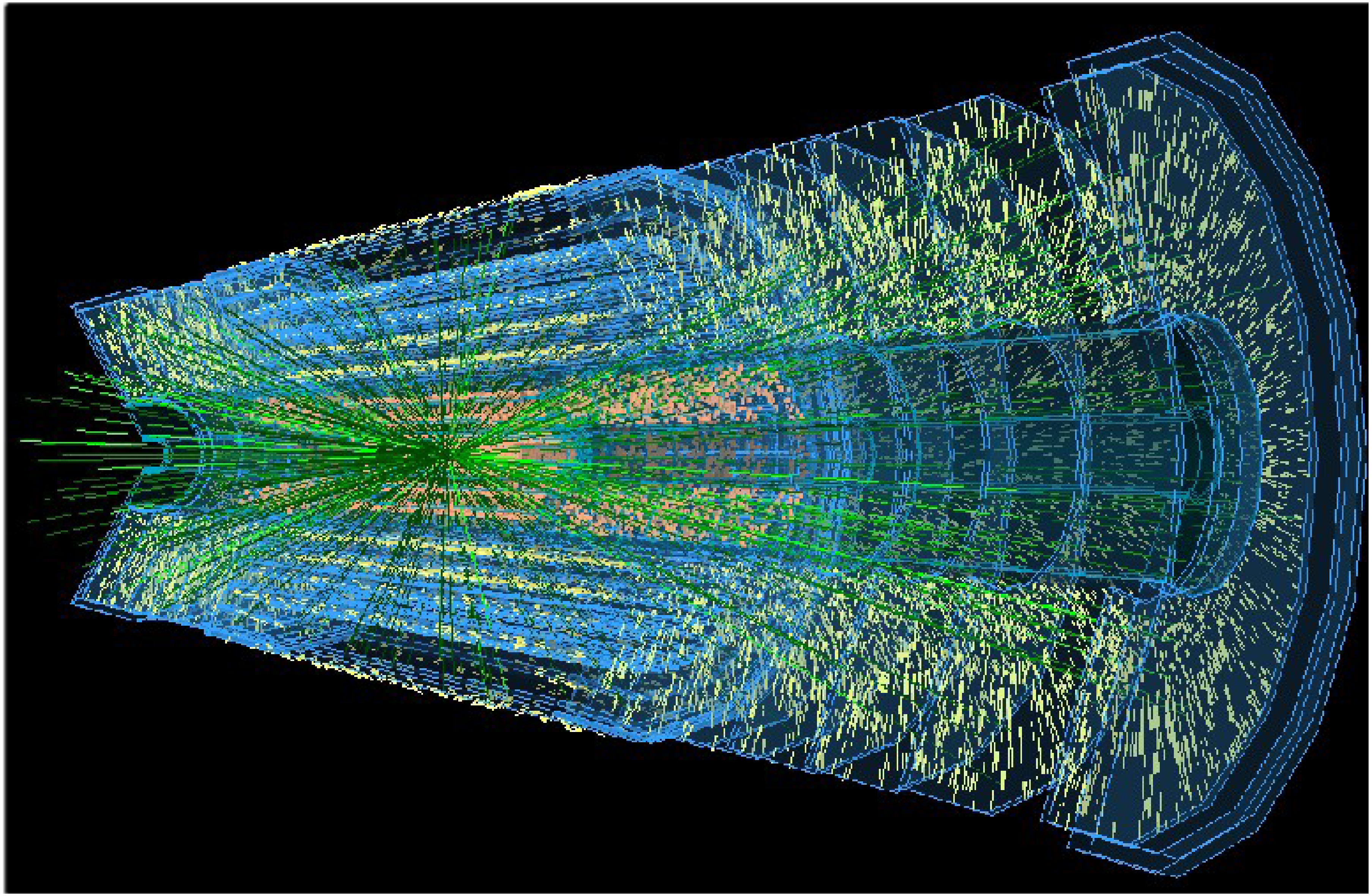
- Pattern recognition, tracking, is a very old, very hot topic in Artificial Intelligence : examples→



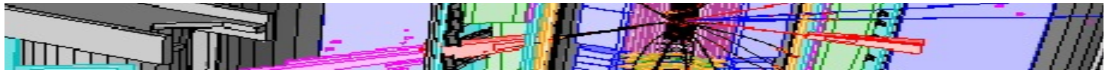
http://papers.nips.cc/paper/5572-a-complete-variational-tracker.pdf



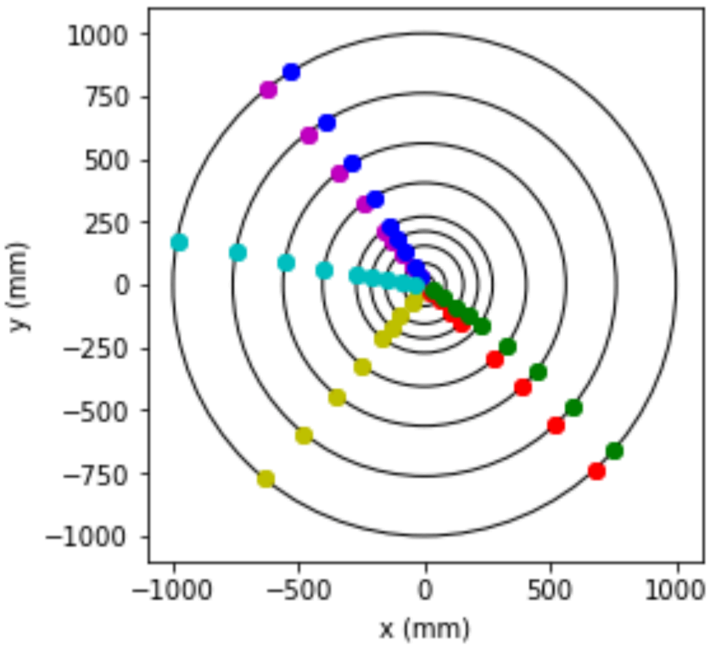
- Note that these are real-time applications, with CPU constraints
- Worry about efficiency, "track swap",…
- But no on-the-shelf algorithm will solve our problem

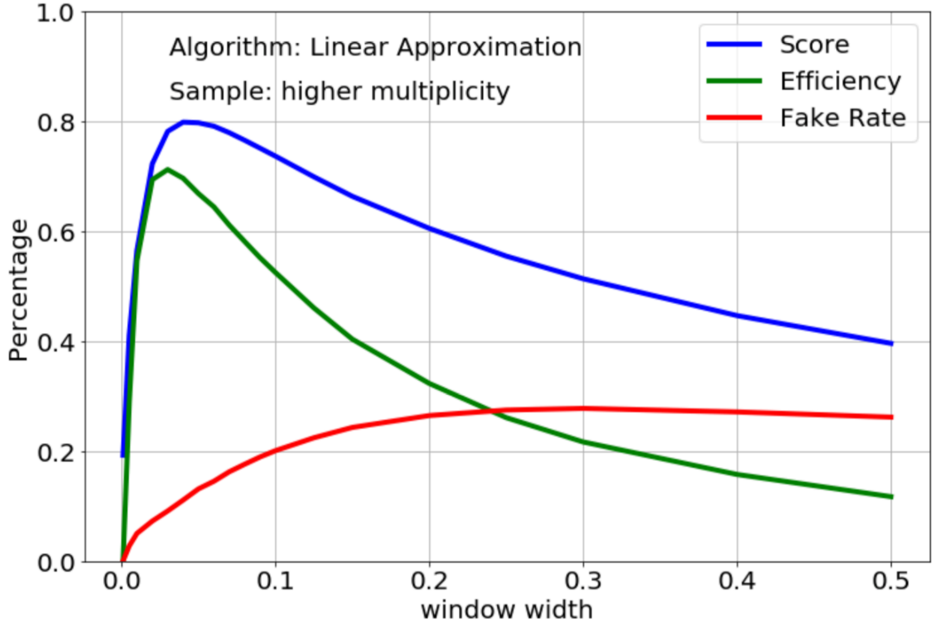


TrackMLRamp



- A simplified tracking challenge setup on RAMP with CDS help (Yetkin Yilmaz Balazs' post-doc 3months, setting up and submission analysis)
- A (non completely trivial) 2D simulation with 10 tracks instead of 3D/10.000 tracks
- Run as a 40 hours hackathon during CTDWIT 6-9th March 2017 LAL-Orsay
- Allowed to validate robustness a scoring variable and show richness of possible algorithms: combinatorial (HEP baseline), conformal mapping, MCTS, LSTM
- Published in proceedings EPJ Web Conf., 150 (2017) 00015





TrackML : current thinking

- We now have a dataset (sorry it took so long)
  - Use ACTS (A Common Tracking Software) to generate fast simulation of a generic Silicon detector at HL-LHC (cylinder and disks)
  - battlefield tested ATLAS software moved to public gitlab@cern
  - ➔simplified simulation but not too simple (otherwise a simple Hough transform would probably work)
  - "cheap" but realistic events which do not "belong" to any collaboration (ATLAS, CMS,…)
- Dataset:
  - 3D points and truth track parameters for n events
  - Typical events with ~200 parasitic collisions (~10.000 tracks/event)
  - Large training sample 1 million events, 100 billion tracks ~1TeraByte
  - Also thinking of allowing participants to generate their dataset
- Participants are given the test sample. They should upload the tracks they have found
  - A track is a list of points belonging to it
  - We don't ask for track parameters, nothing will beat Kalman filter
  - Figure of merit built from efficiency, fake rate, CPU time
- We have decided to run in two phases
  - Phase 1 : focus only on accuracy, no CPU incentive
    - Discussing with Kaggle next week
    - To run in Winter 2018
  - Phase 2 : focus on CPU, preserving accuracy
    - More tricky, require the challenge platform to run the algorithm within controlled environment
    - To run in Summer 2018

What with CDS?

- We're not looking really for new collaboration on preparing the challenge itself but still:
  - 3-months of an engineer (preferably from CDS core) to finalise the challenge operation, especially phase 2
  - …tricky, need to run submitted software in a controlled environment, and to handle many submissions
- Put more emphasis on post challenge analysis and mid/long term collaboration
  - use the CDS channels to advertise the challenge
  - master internship for post-challenge analysis
  - Build in/post-challenge collaboration with CDS scientists on innovative approaches to the tracking problem as revealed by the challenge.
  - possibly collaboration on the visualization (Tobias Isenberg INRIA/Saclay expressed interest, thanks to pitching day 2016)