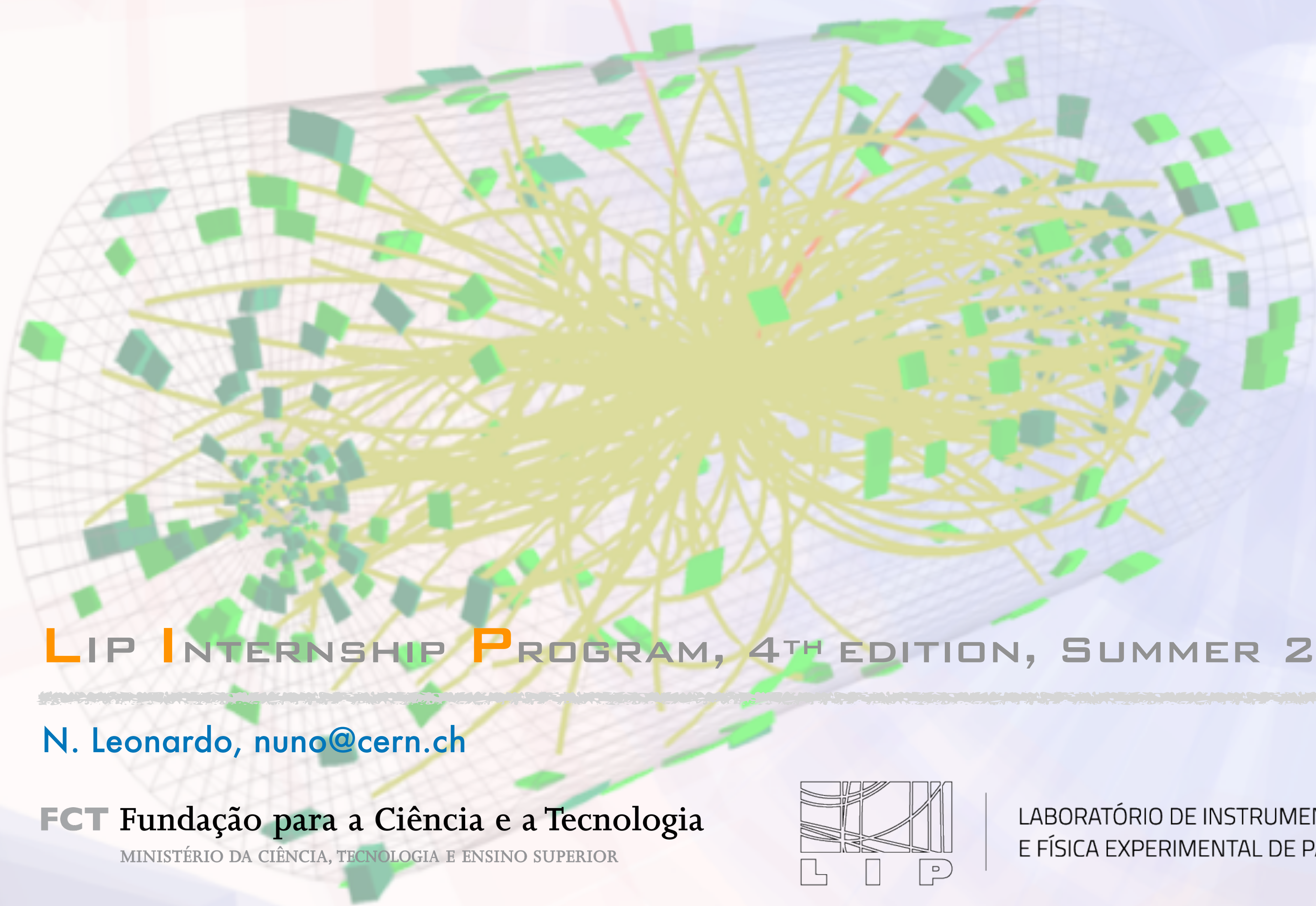
INTERNSHIP PROGRAM

Tutorial on

Data Analysis



LIP INTERNSHIP PROGRAM, 4TH EDITION, SUMMER 2020

N. Leonardo, nuno@cern.ch

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

LIP
LABORATÓRIO DE INSTRUMENTAÇÃO
E FÍSICA EXPERIMENTAL DE PARTÍCULAS

introduction

- this week you’ve seen already many nice physics results in the form of plenty of colourful looking plots
- you’ve seen how (stable) particles are detected and their trajectories (and kinematics) measured and used to reconstruct our (unstable) particles of interest
- on the last two afternoons you’ve had an introduction to HEP’s baseline programming environment and tools, Linux, C++, ROOT
- now we’ll try to connect the dots by performing a data analysis

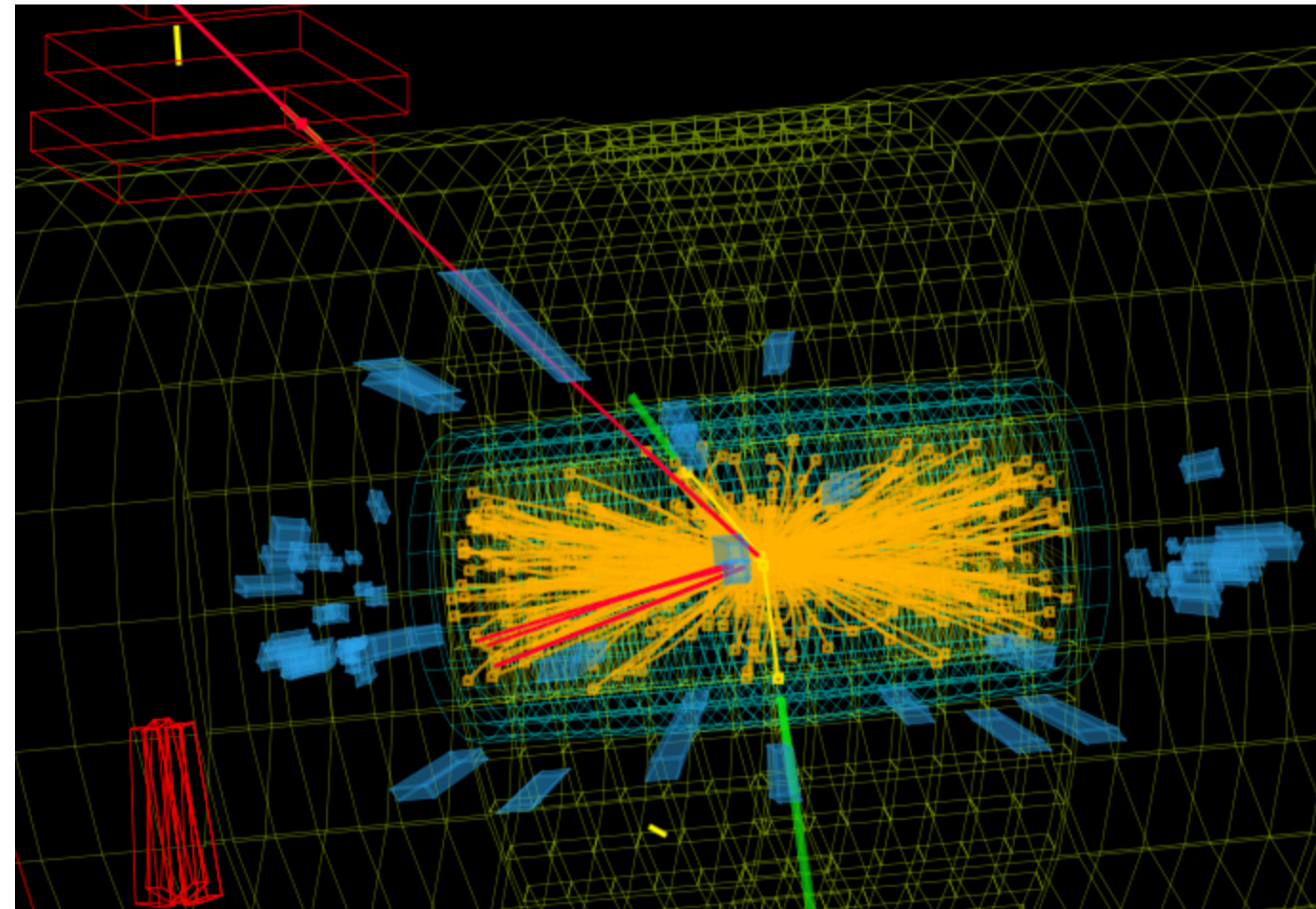
goals

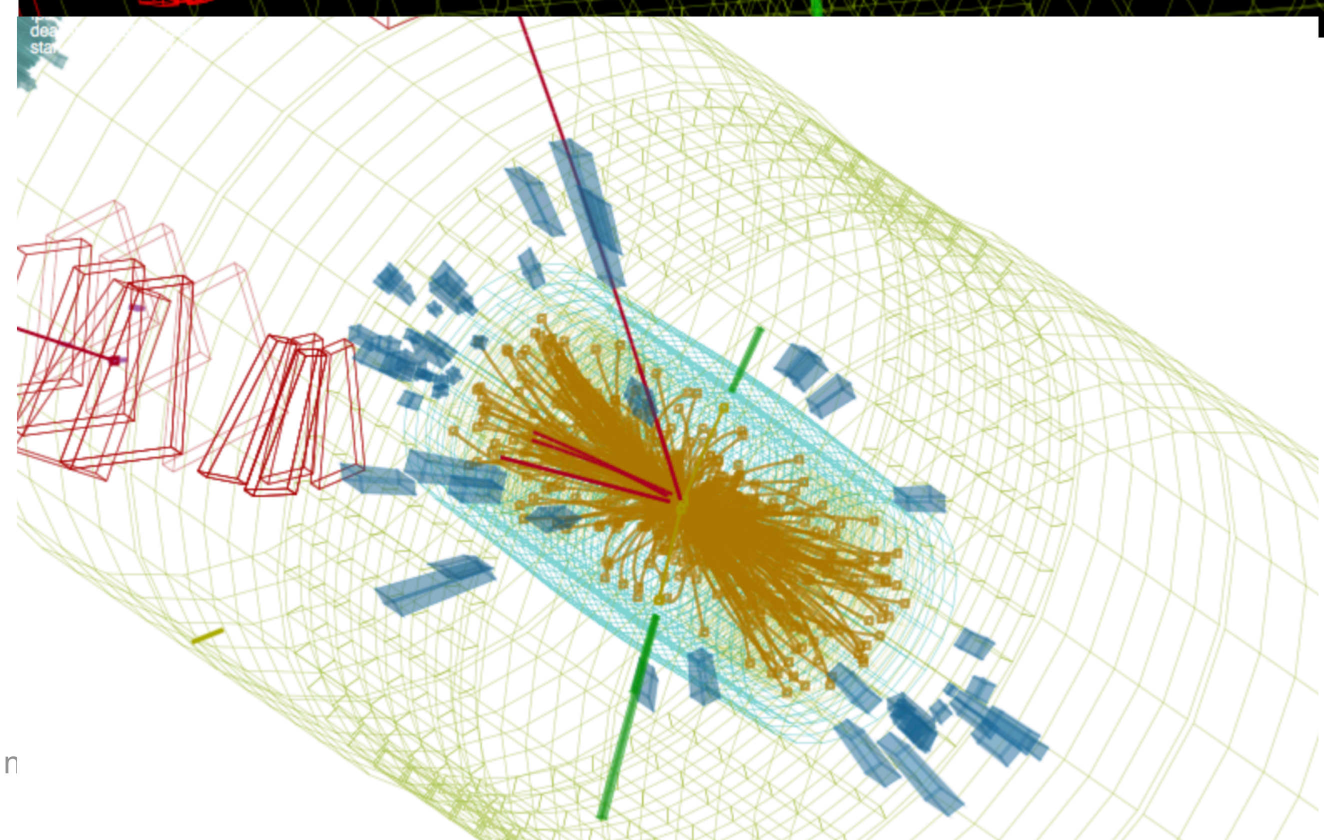
perform a simple data analysis

- visualise the data
- manipulate data ntuples
- produce, process, and display data histograms
  - select different physics signals
  - plot kinematic distributions
  - inspect trigger effects
  - selection criteria and efficiency
- extract physics parameters from data
  - by performing a likelihood fit
  - statistical errors
- systematic errors !

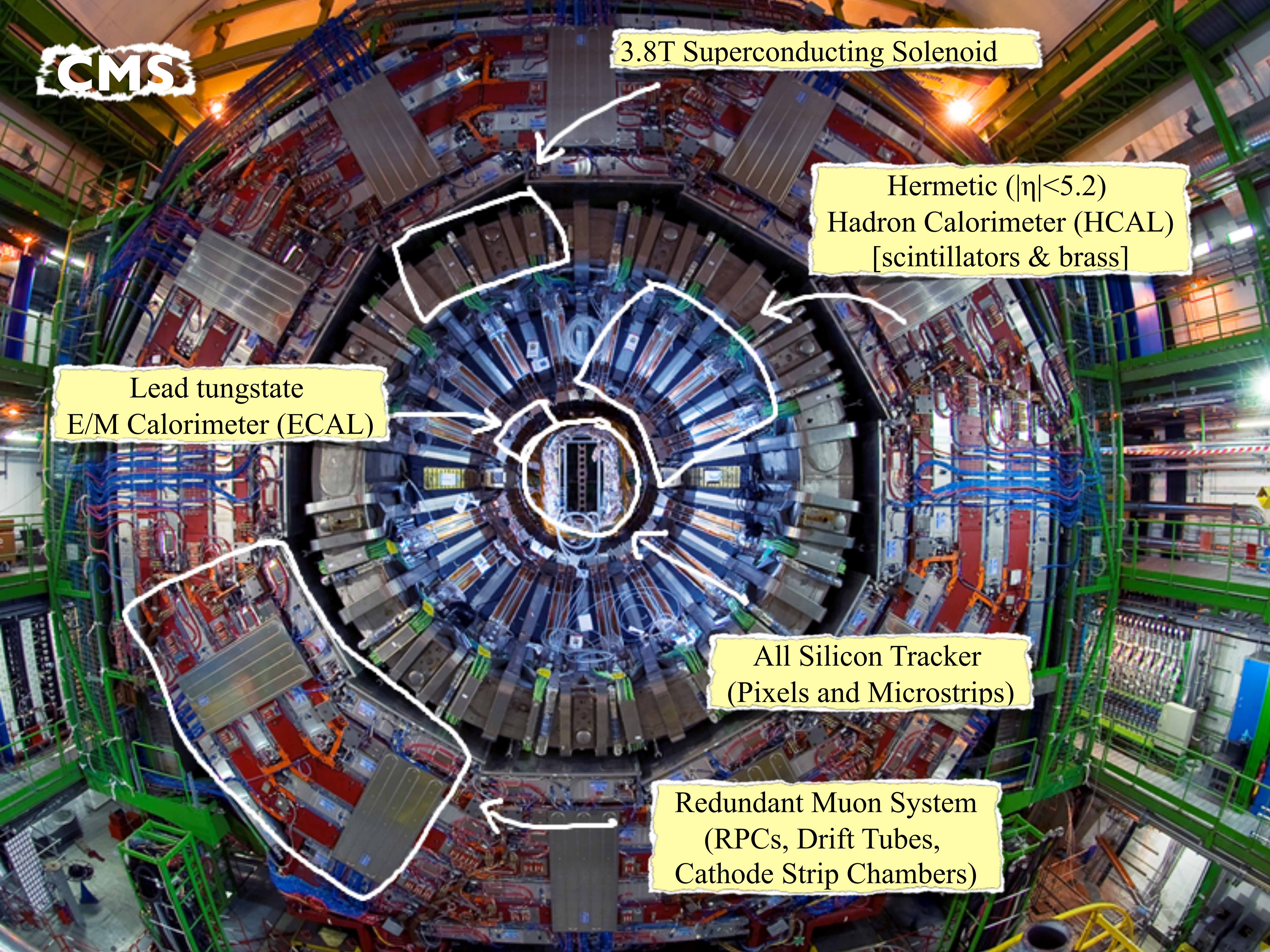
LHC Open Data

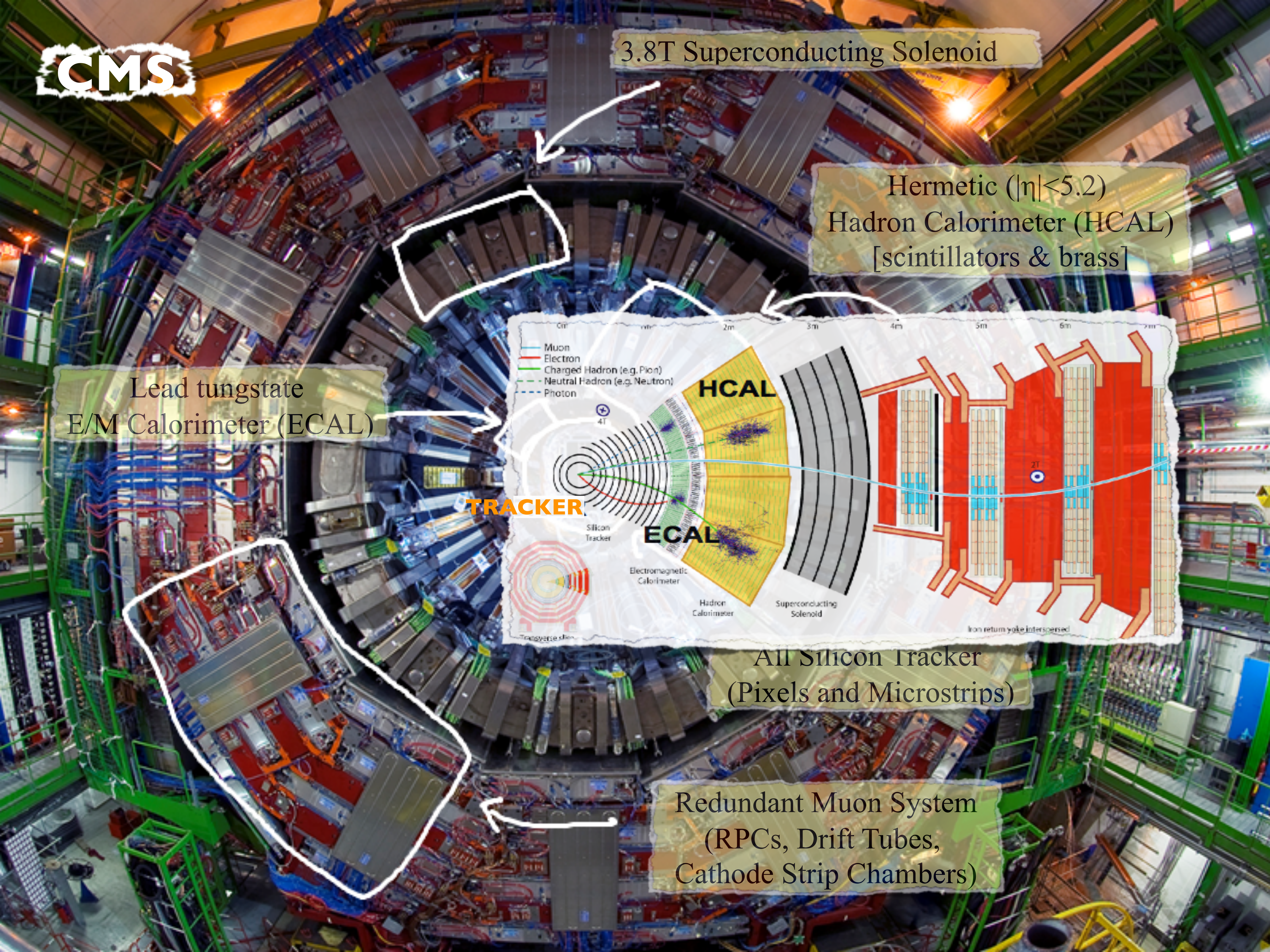
- the LHC collaborations make good chunks of their data publicly available
  - http://**opendata.cern**.ch/
- along with tools & software & examples
- for data visualisation and analysis
- from event reconstruction algorithms to machine learning challenges
- via virtual machines (with no need to install different software packages)
- few pointers
  - http://opendata.cern.ch/visualise/events/cms
  - http://www.i2u2.org/elab/cms/event-display/
- you're invited to **explore the LHC data** also on your own leisure





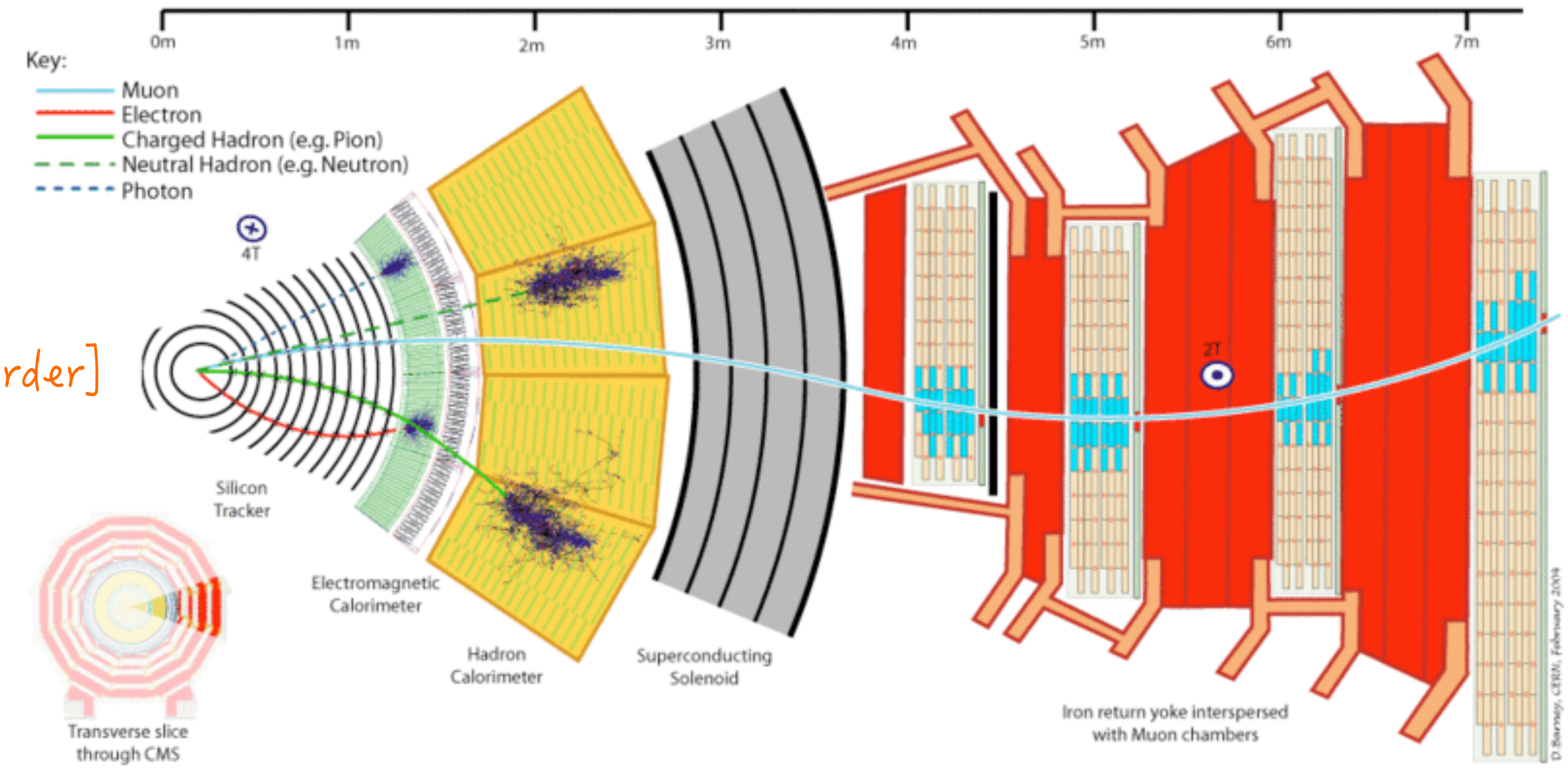
Detector & Event Reconstruction & Visualisation


CMS
3.8T Superconducting Solenoid
Hermetic (|η|<5.2)
Hadron Calorimeter (HCAL)
[scintillators & brass]
Lead tungstate
E/M Calorimeter (ECAL)
All Silicon Tracker
(Pixels and Microstrips)
Redundant Muon System
(RPCs, Drift Tubes,
Cathode Strip Chambers)

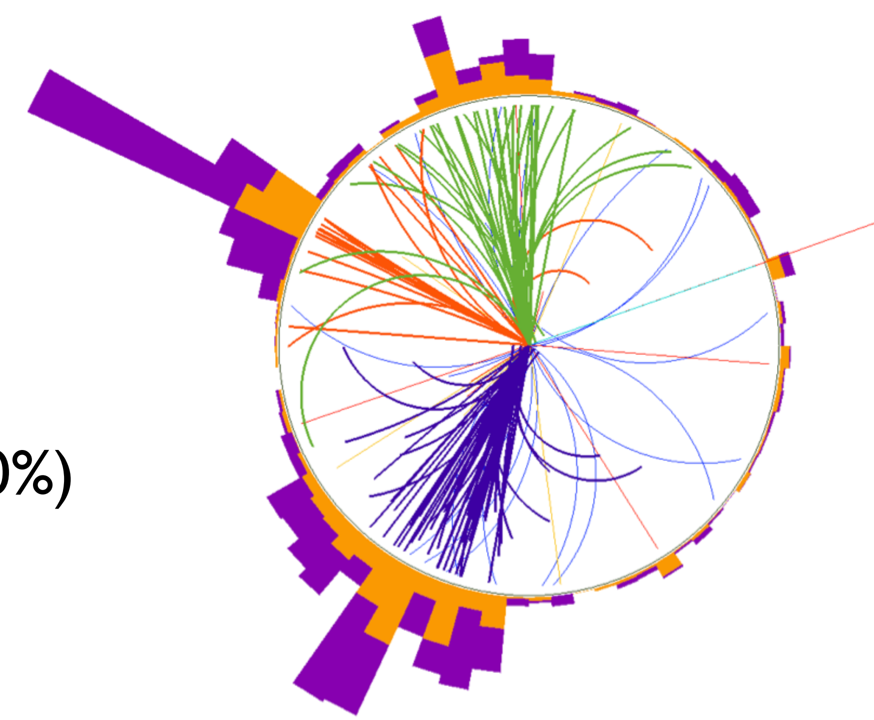

CMS
3.8T Superconducting Solenoid
Hermetic (|η|<5.2)
Hadron Calorimeter (HCAL)
[scintillators & brass]
Lead tungstate
E/M Calorimeter (ECAL)
Muon
Electron
Charged Hadron (e.g. Pion)
Neutral Hadron (e.g. Neutron)
Photon
4T
HCAL
ECAL
TRACKER
Silicon Tracker
Electromagnetic Calorimeter
Hadron Calorimeter
Superconducting Solenoid
2T
All Silicon Tracker
(Pixels and Microstrips)
Redundant Muon System
(RPCs, Drift Tubes,
Cathode Strip Chambers)

particle identification

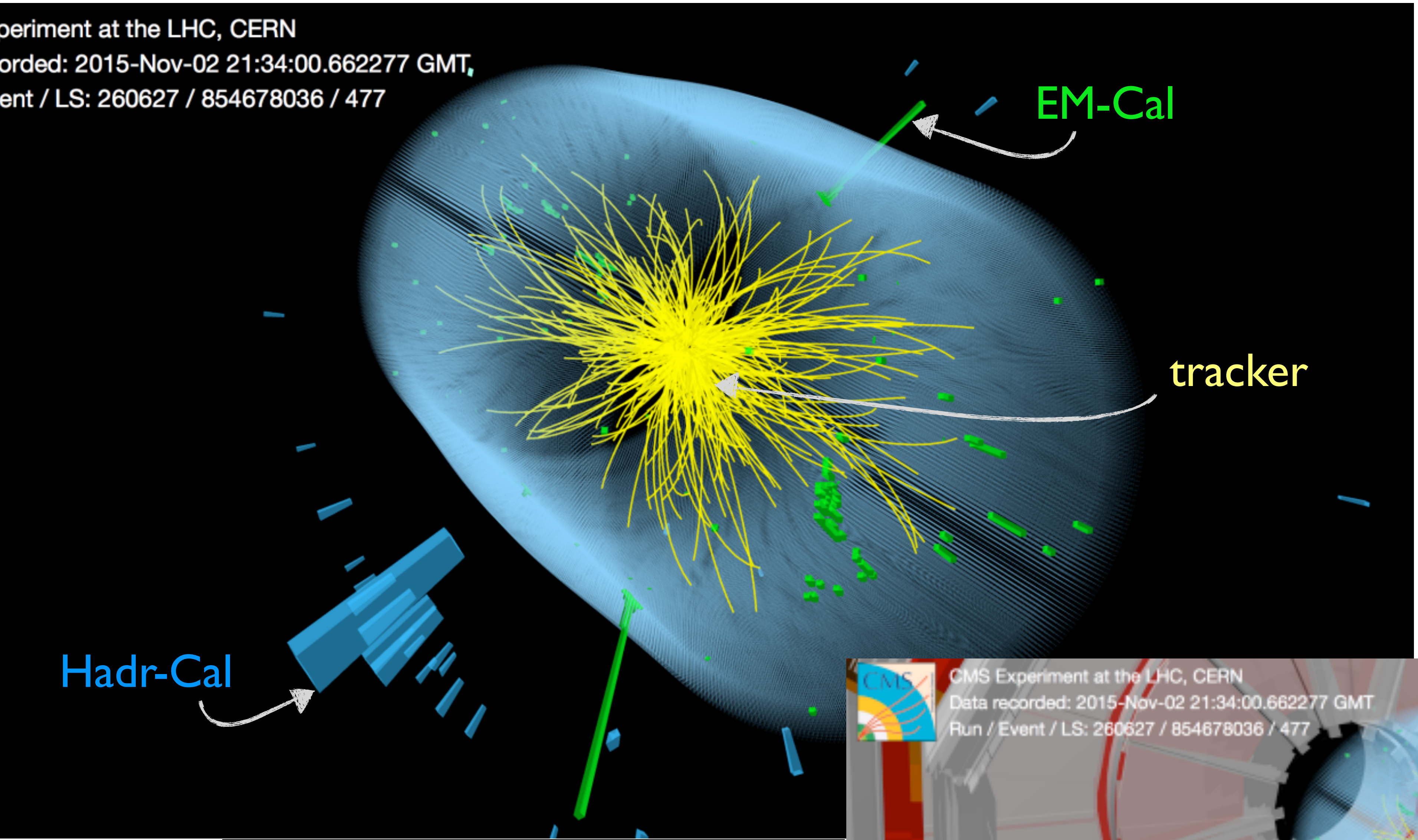
[leading order]



- [higher order corrections] objects are finally reconstructed using information from different detector subsystems combined in a particle flow algorithm
  - electrons radiate via bremsstrahlung
  - photons may convert to e^+e^- pairs in the tracker
  - jet (q,g) energy is formed of charged/neutral hadrons (65%/10%) and photons (25%): calorimeter and tracker info exploited
  - missing E_T requires 'full event' reconstruction



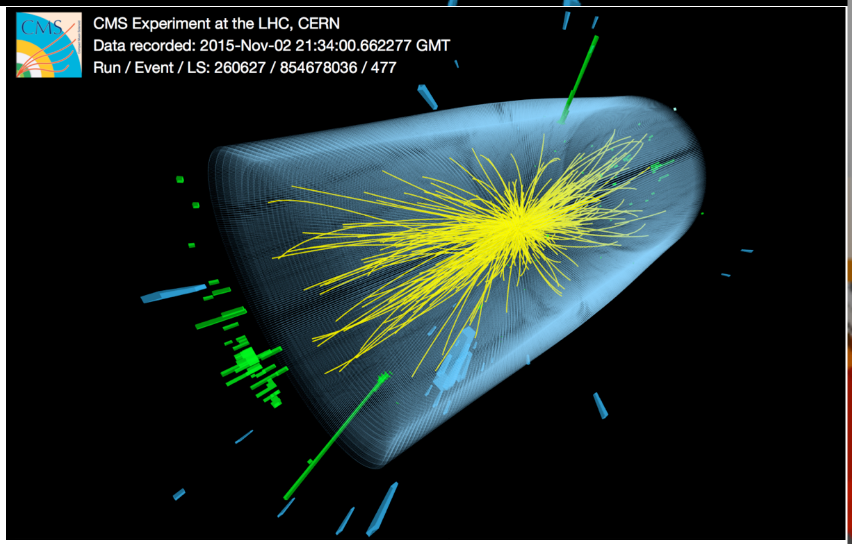
di-photons

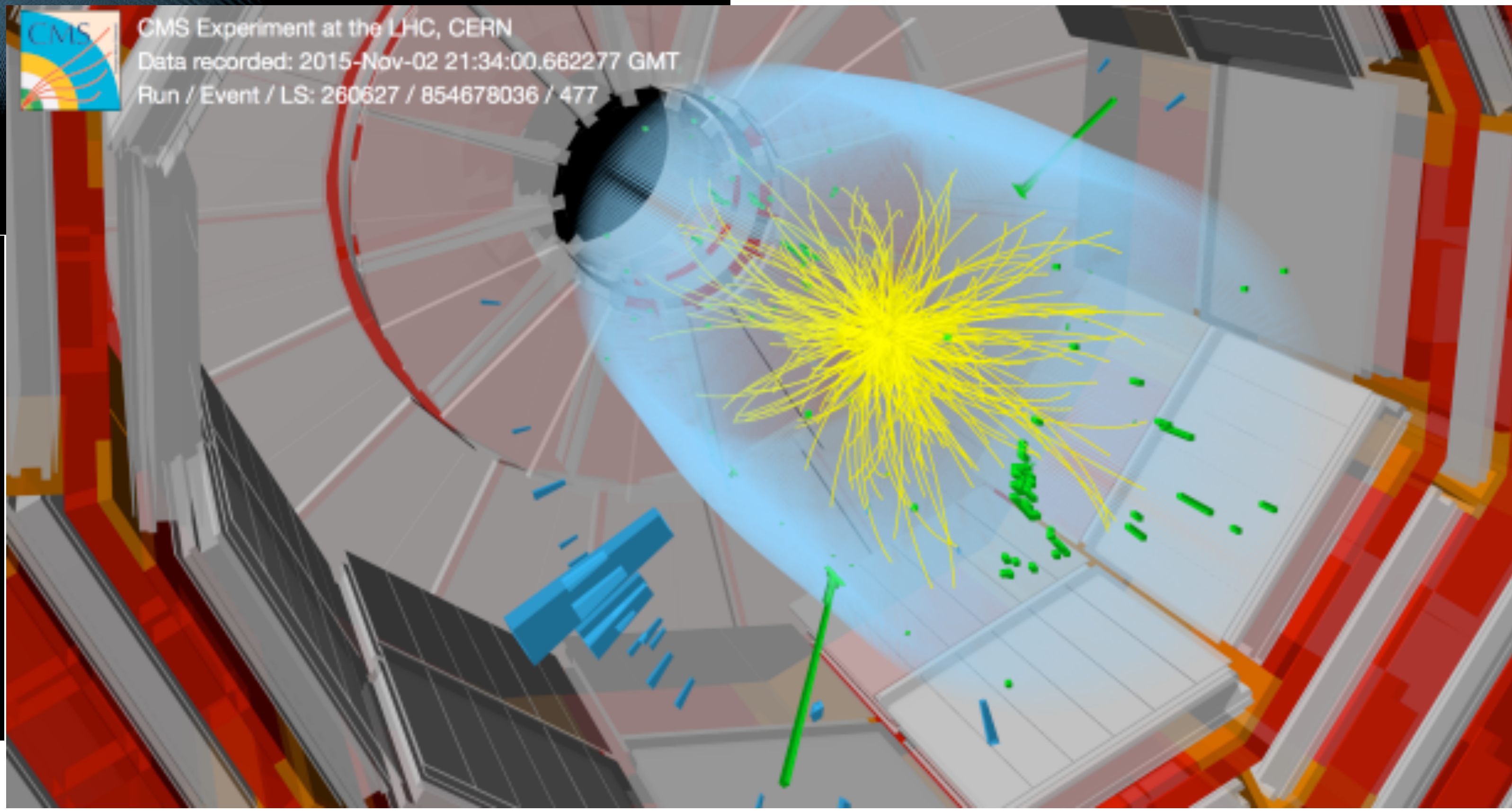


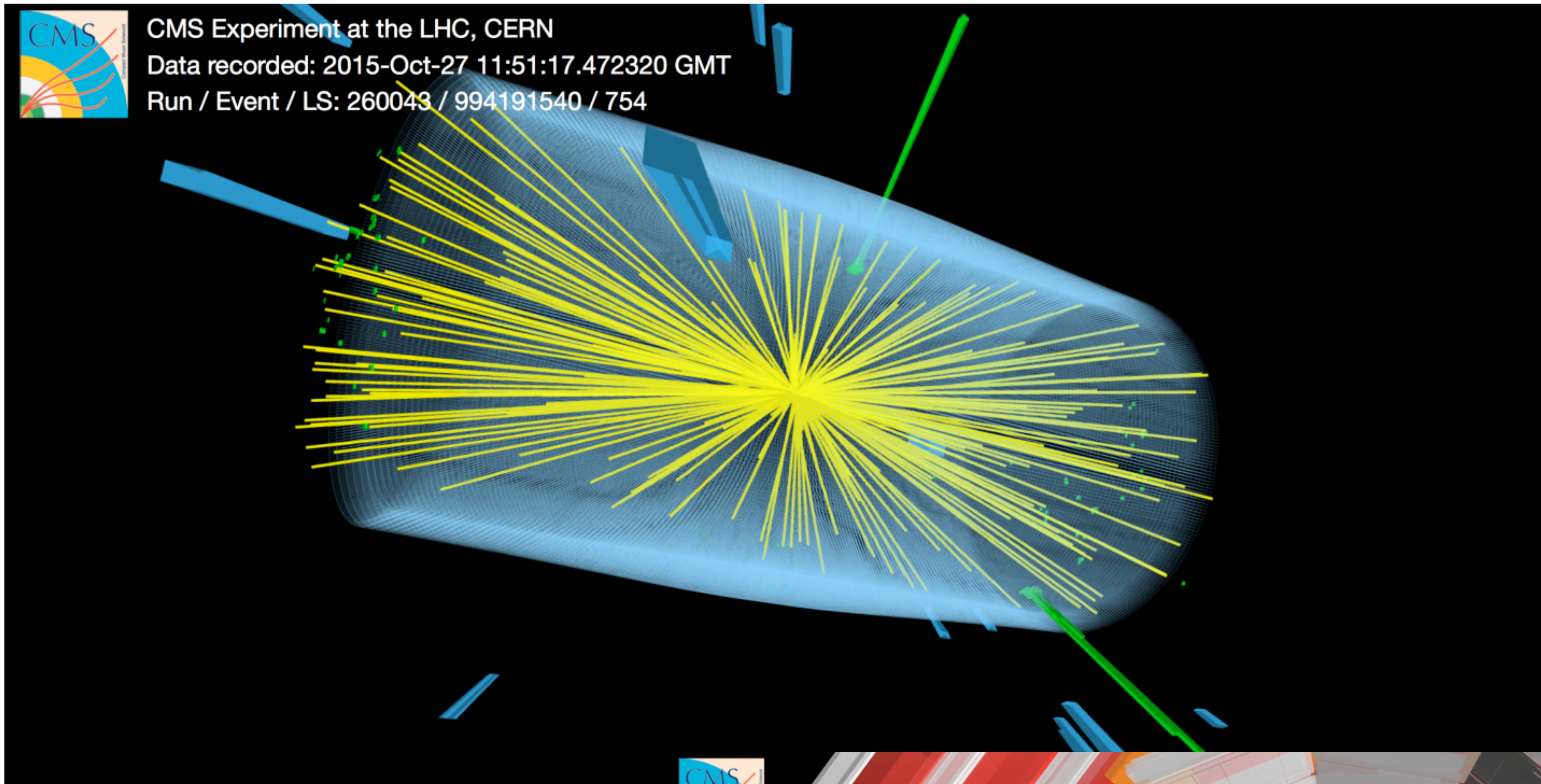
$X \rightarrow \gamma\gamma$

$m_{\gamma\gamma}$ ~ 750 GeV

CMS-PHO-EVENTS-2015-007

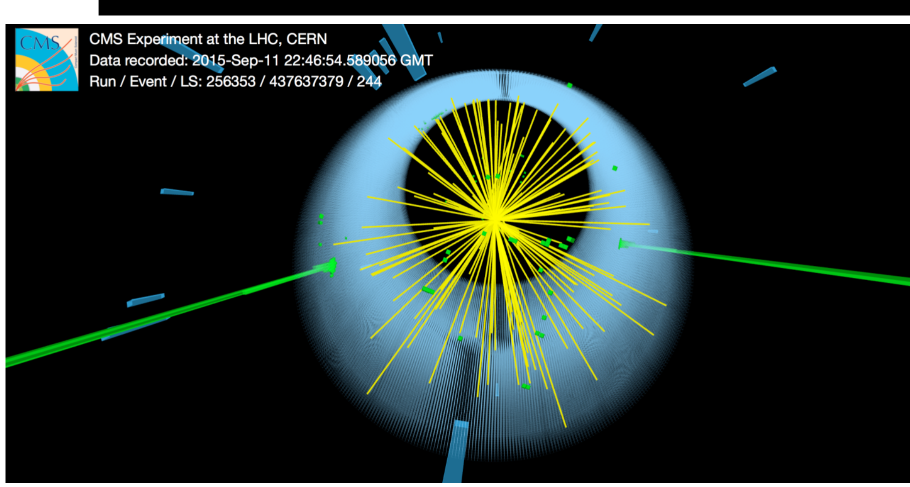


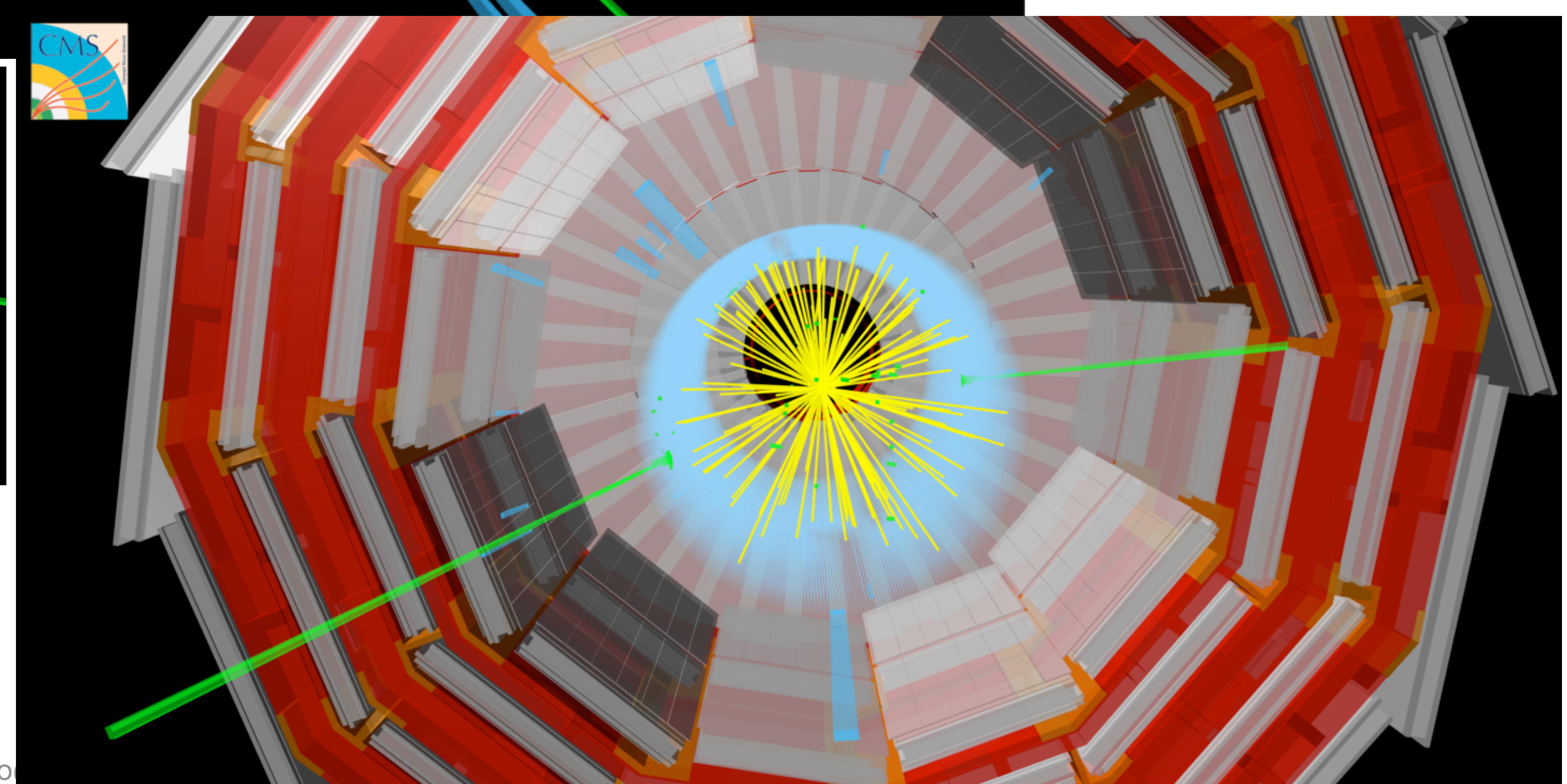


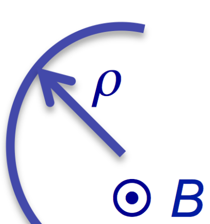


?

$m_{??}$ ~ 800 GeV

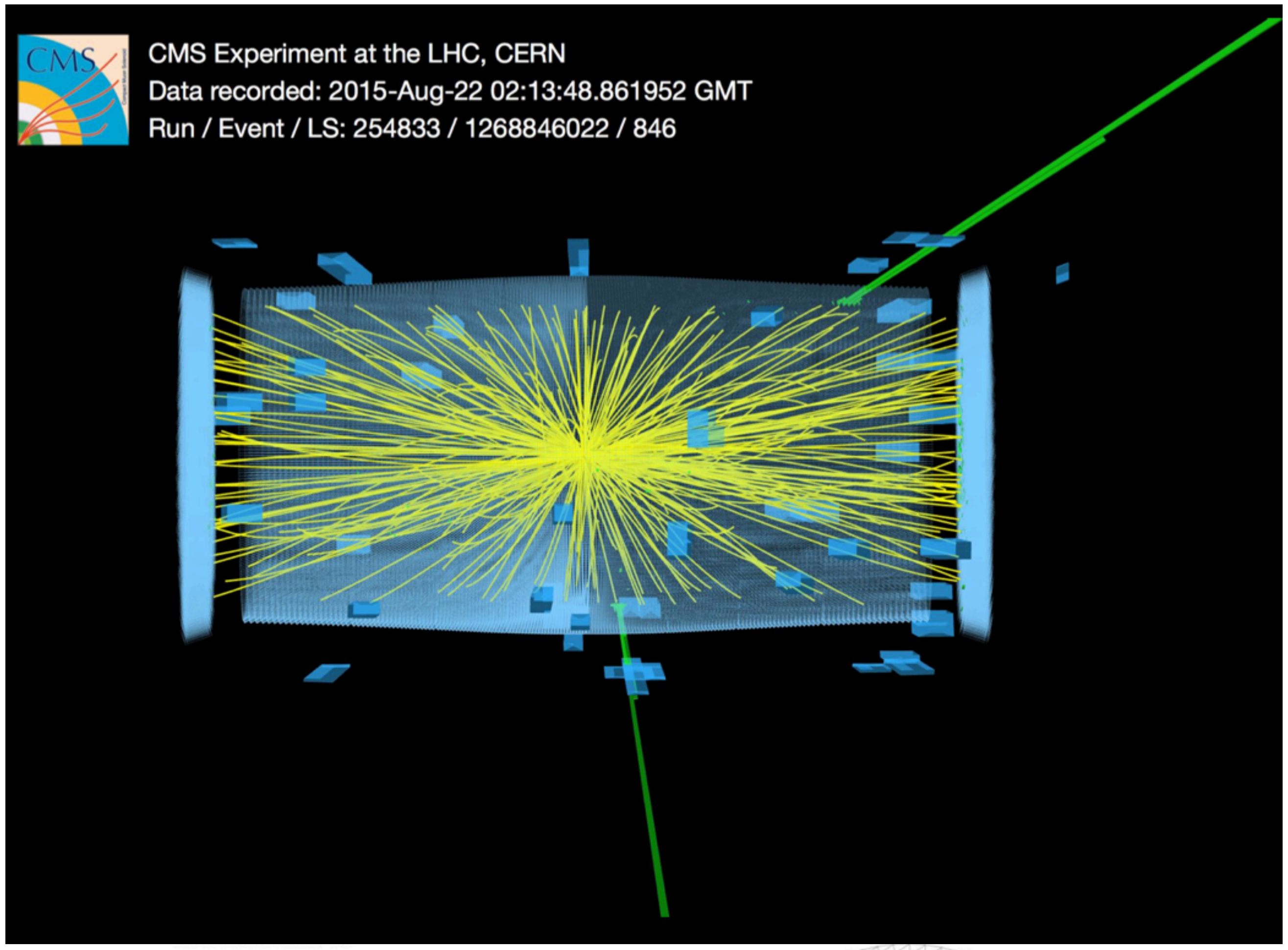




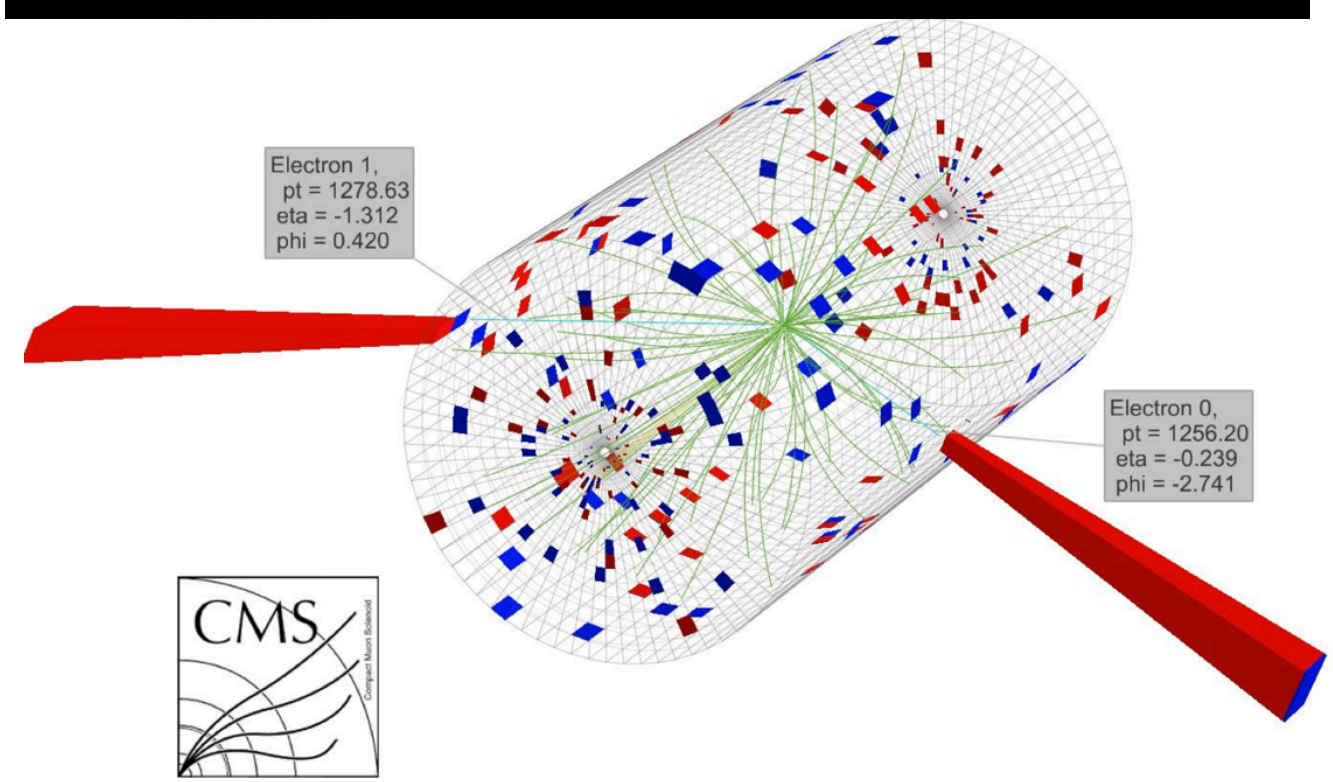


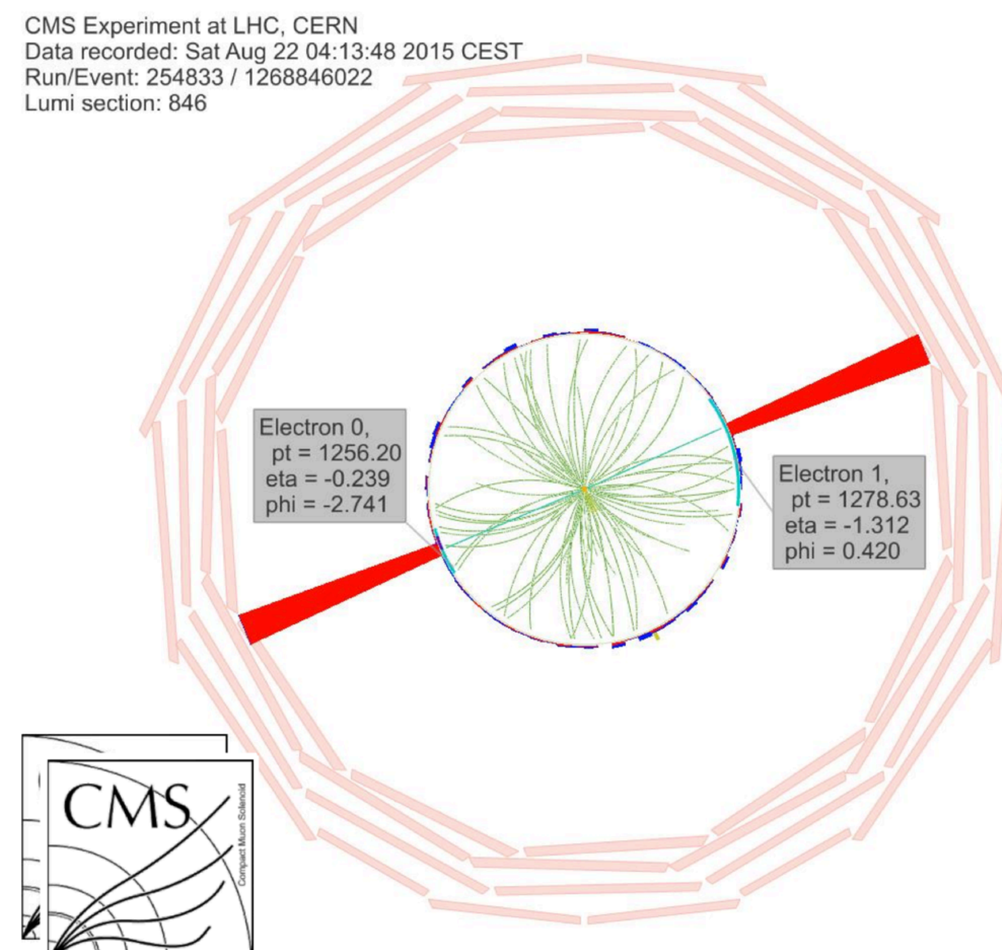
$$\rho = \frac{p}{ZeB}$$

a di-electron event



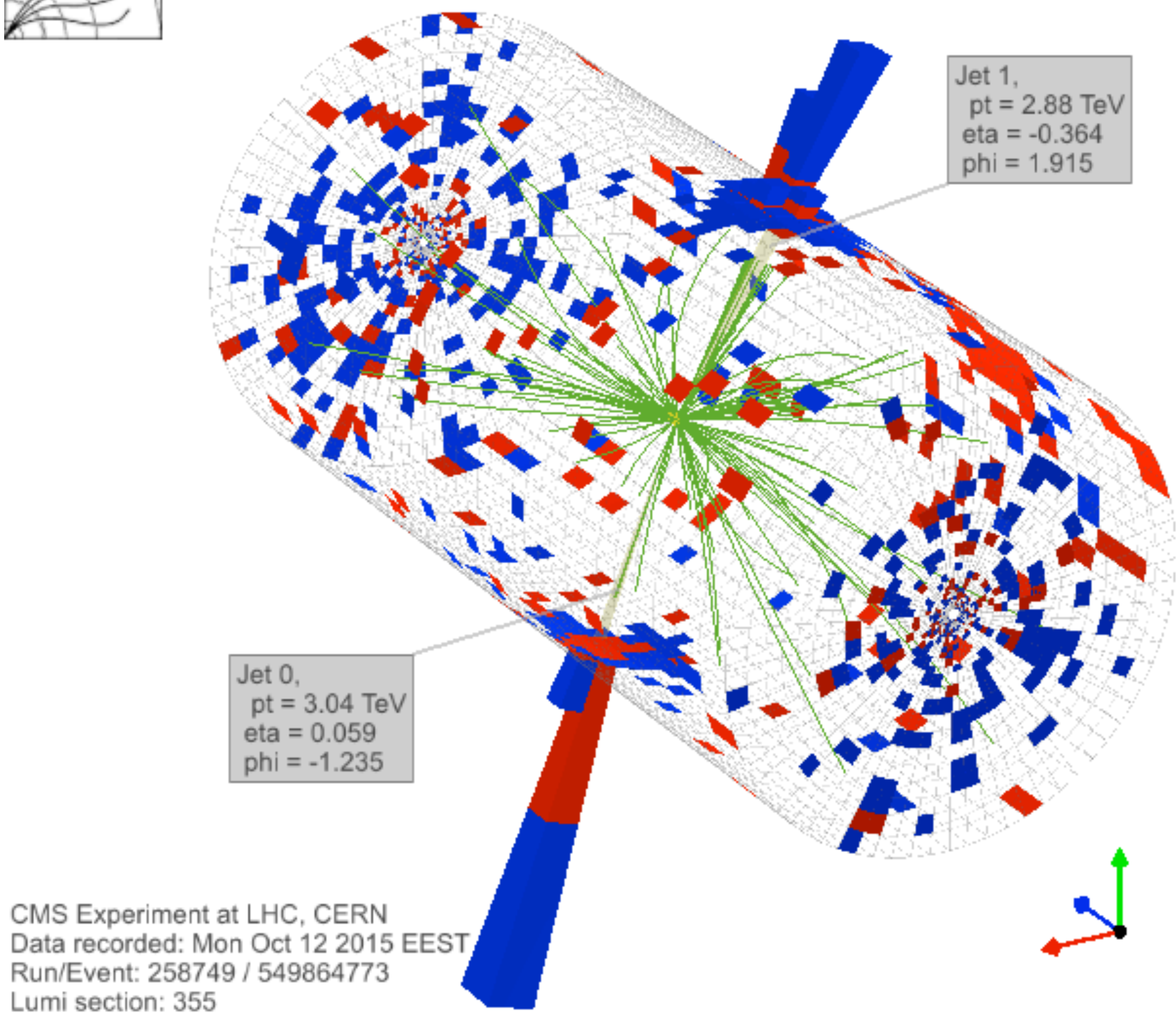
Event Display of a Candidate Electron-Positron Pair with an Invariant Mass of 2.9 TeV

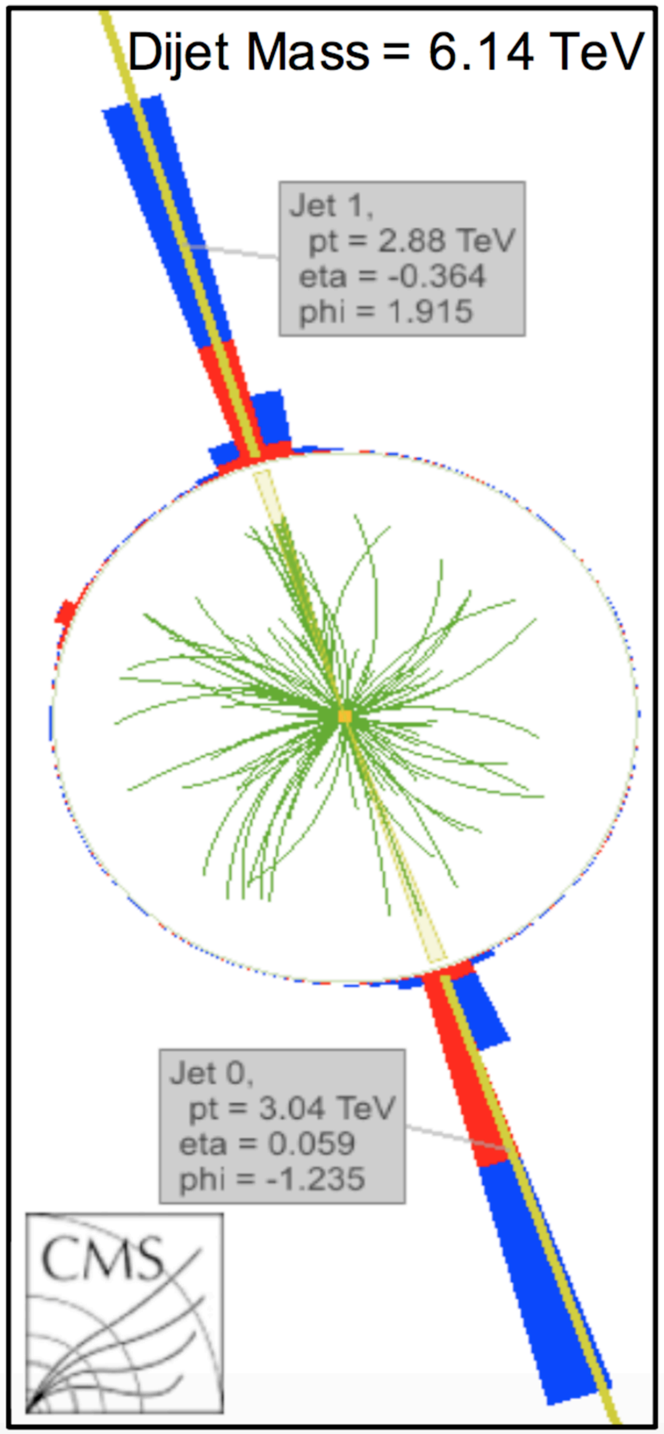




di-jets

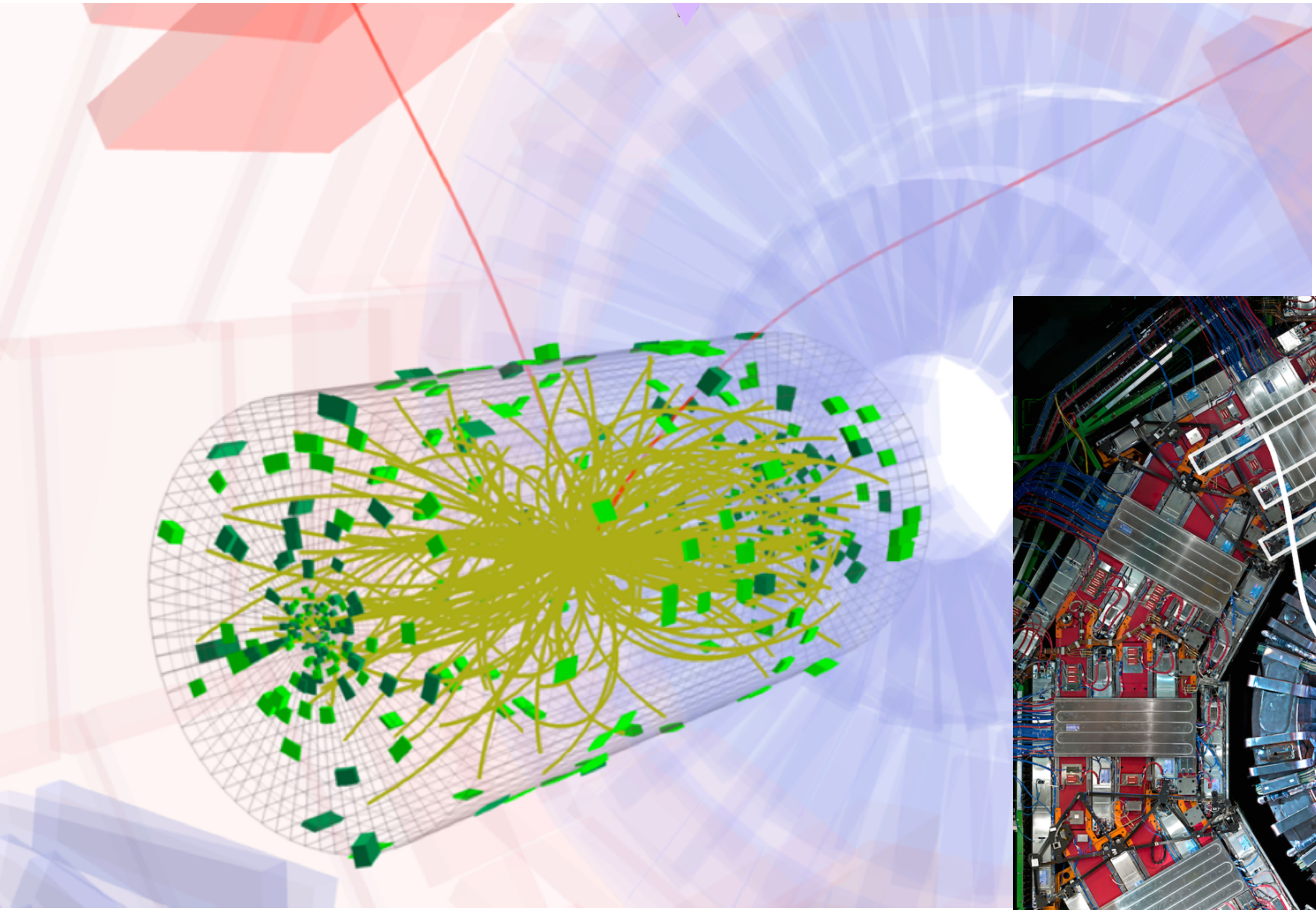
X → j j

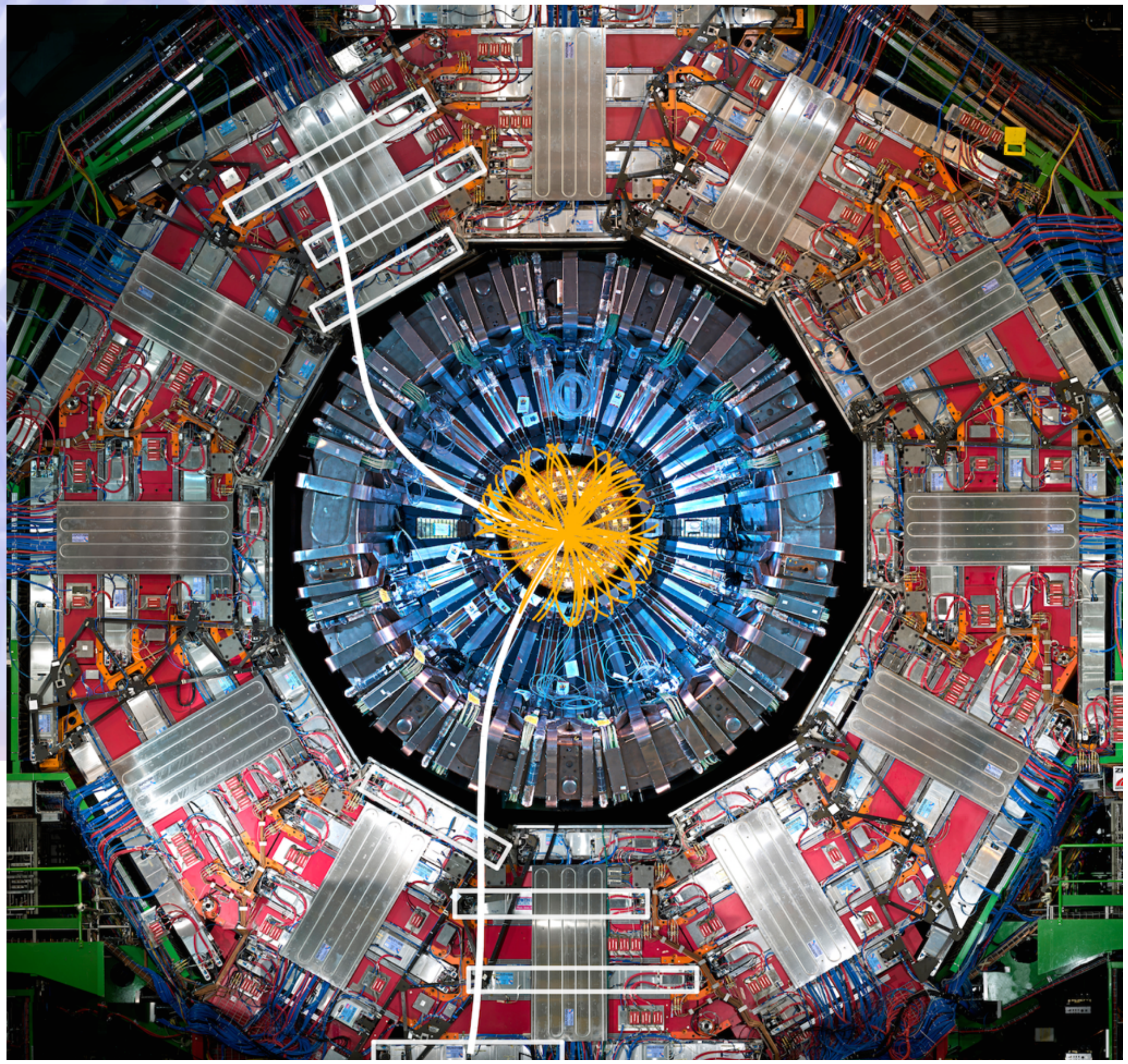

Jet 1,
pt = 2.88 TeV
eta = -0.364
phi = 1.915
Jet 0,
pt = 3.04 TeV
eta = 0.059
phi = -1.235
CMS Experiment at LHC, CERN
Data recorded: Mon Oct 12 2015 EEST
Run/Event: 258749 / 549864773
Lumi section: 355
Dijet Mass: 6.14 TeV


Dijet Mass = 6.14 TeV
Jet 1,
pt = 2.88 TeV
eta = -0.364
phi = 1.915
Jet 0,
pt = 3.04 TeV
eta = 0.059
phi = -1.235
CMS

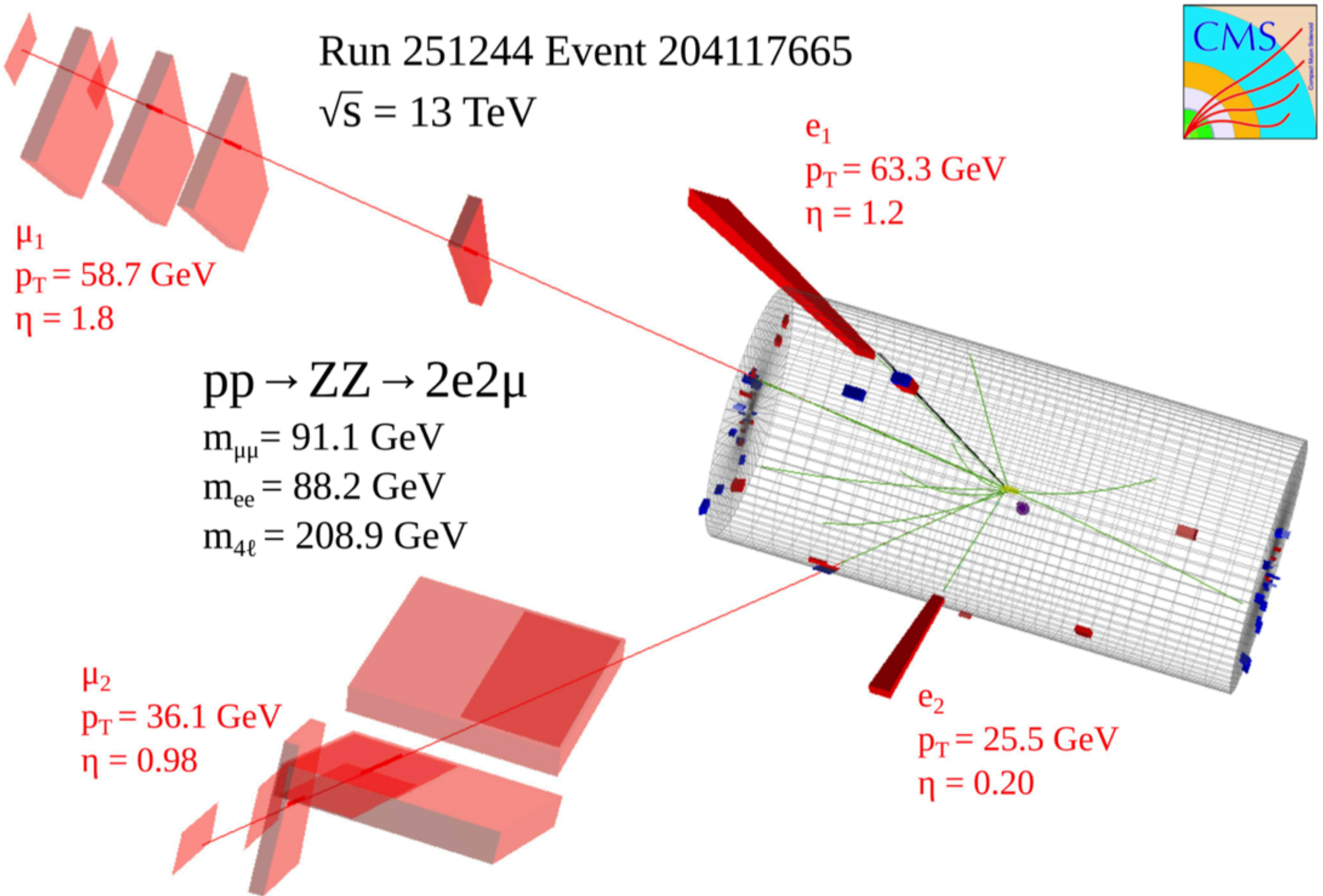
a di-muon event

$X \rightarrow \mu\mu$





a $\mu^+\mu^-e^+e^-$ event

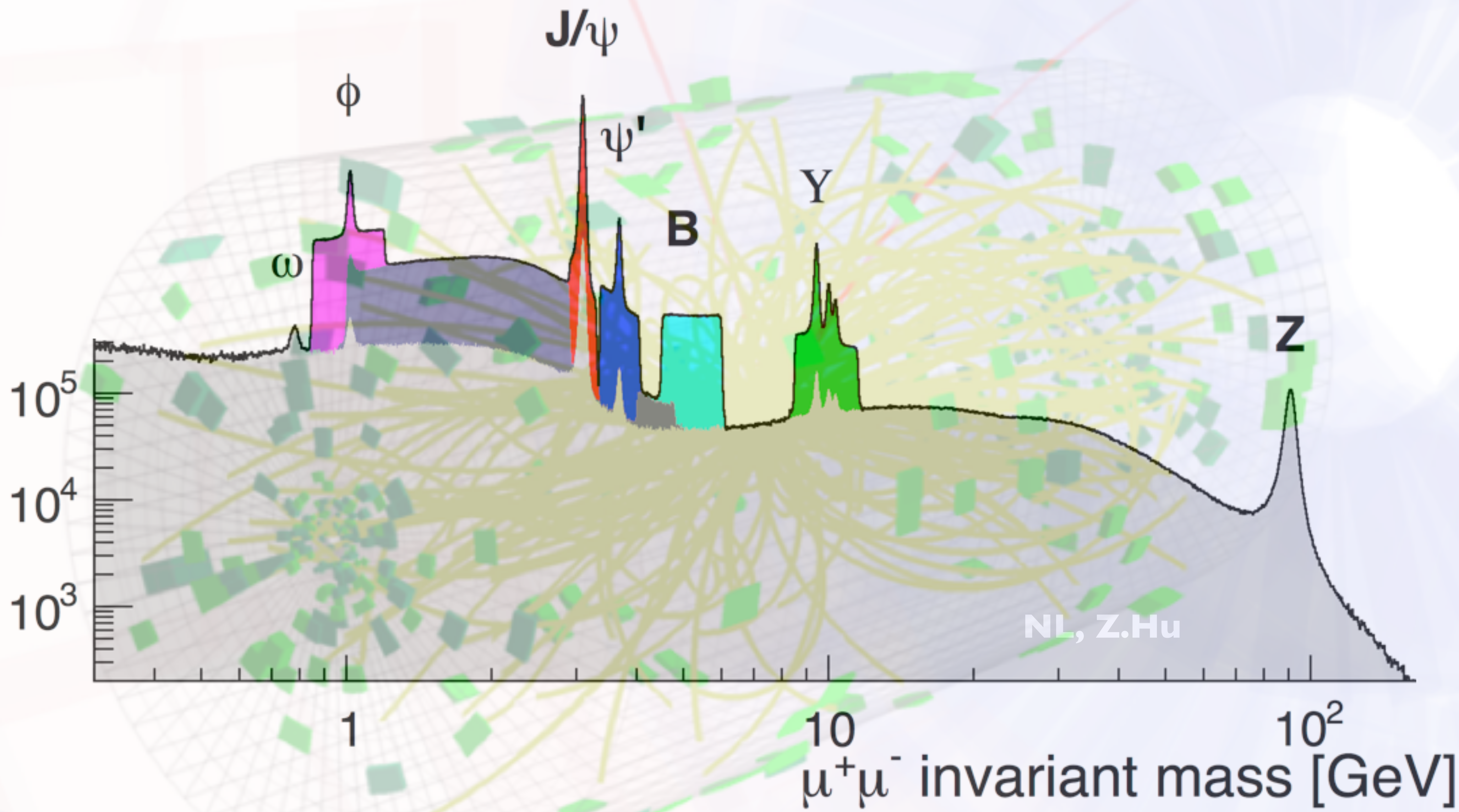


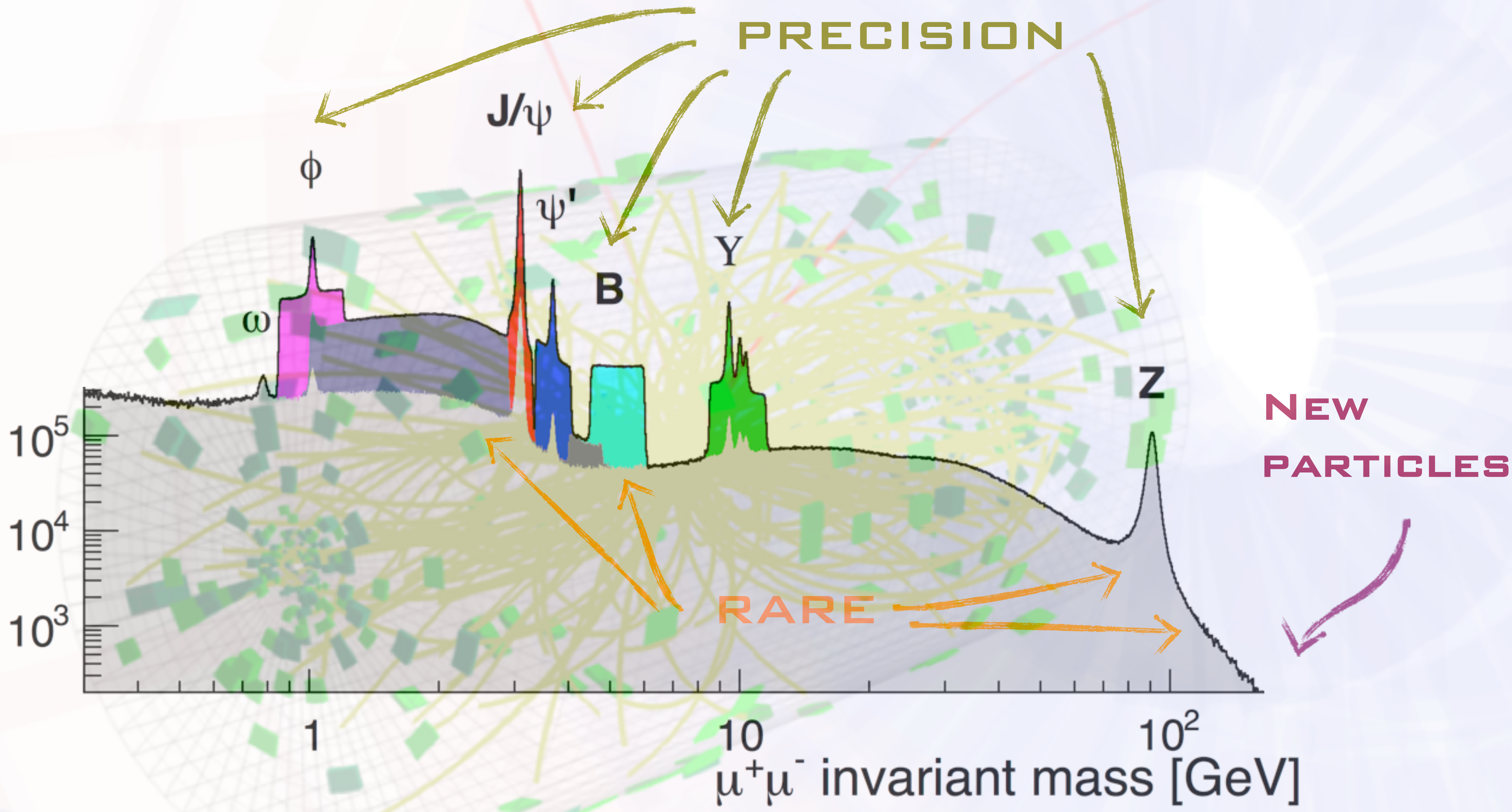
➡️ processes are explored with many (more) final state particles

the di-muon analysis

the di-muon spectrum (X →μμ)

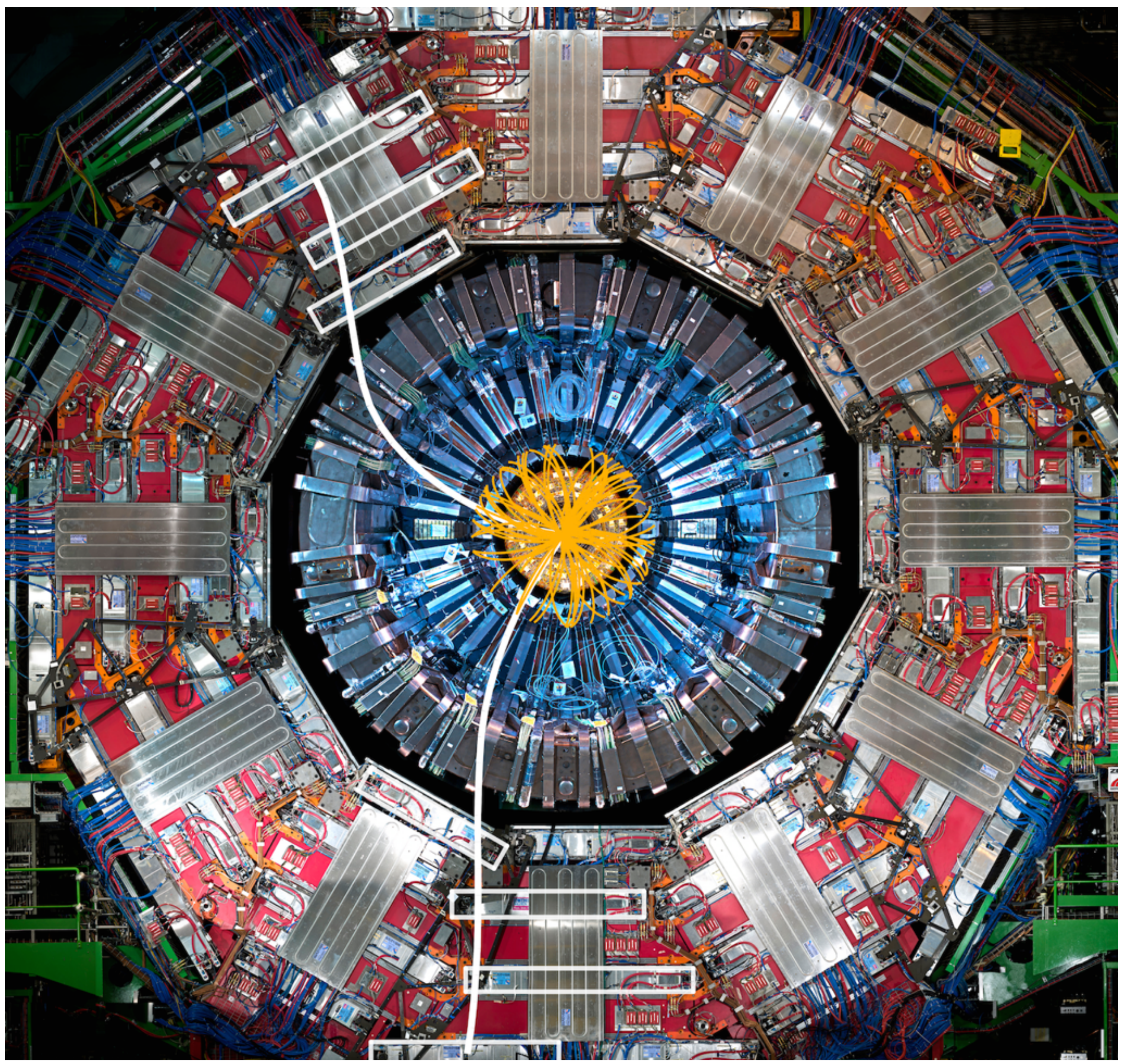
50 years of particle physics in one plot!

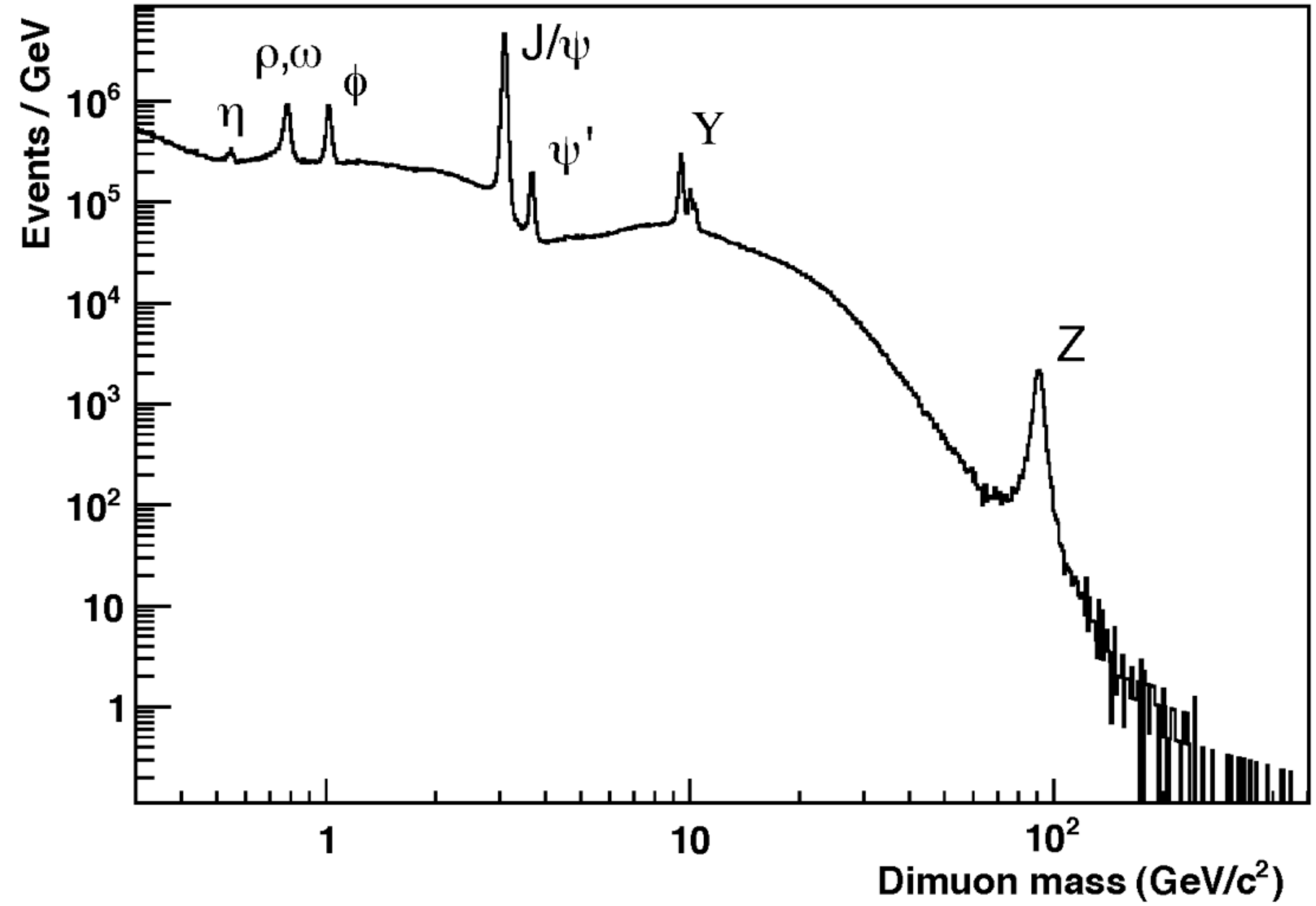



PRECISION
J/ψ
ϕ
ψ'
B
Υ
ω
Z
NEW
PARTICLES
RARE
10^5
10^4
10^3
1
10
10^2
$\mu^+\mu^-$ invariant mass [GeV]

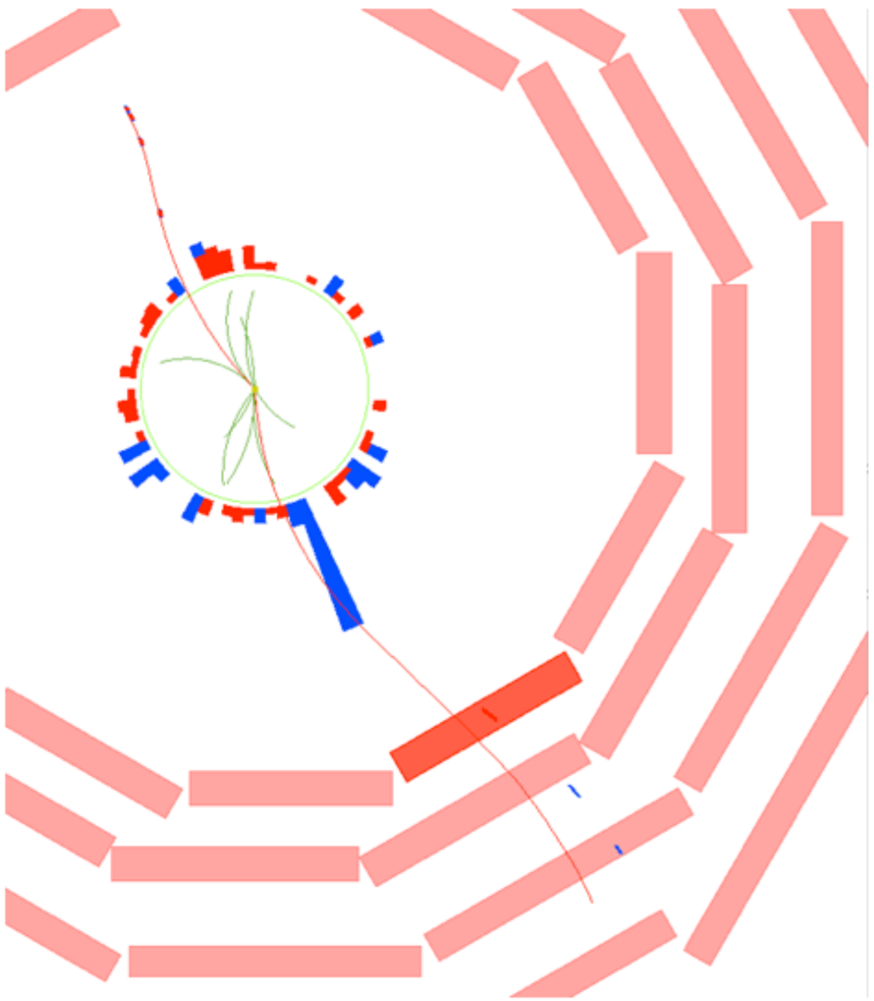
from detector to physics ...

?





di-muon 'invariant mass' ?



particle identification
- signal in muon chambers

→ it's a muon!

➠ $m = m(\mu) \sim 106 MeV/c^2$

particle trajectory
- muon chambers but especially the silicon tracker

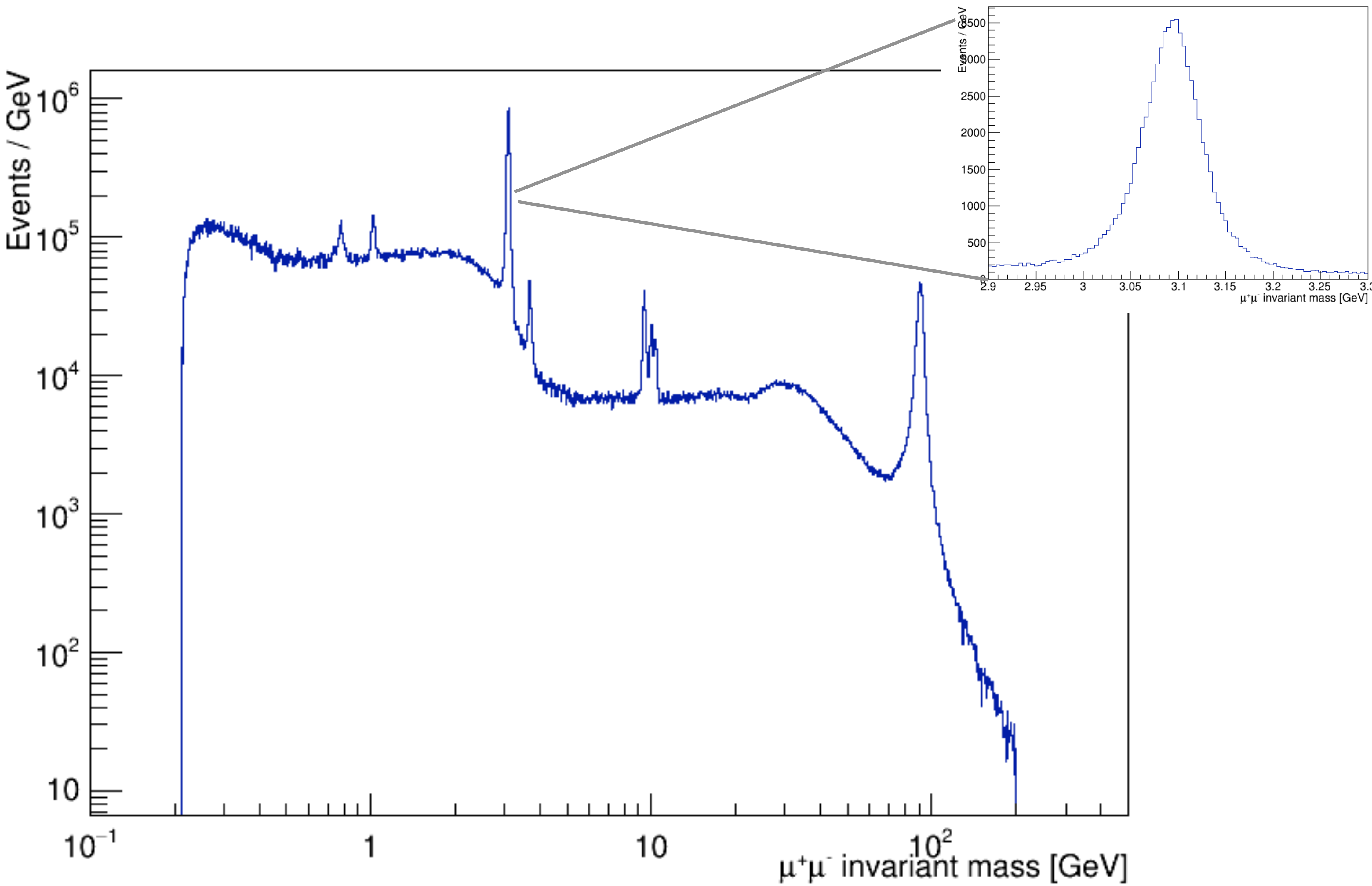
➠ linear momentum, $\underline{p} \equiv (p_x, p_y, p_z)$

➠ form 4-momentum of each muon: $\mathbf{P}_\mu \equiv (E, p_x, p_y, p_z)$

➠ that of the di-muon pair $\mathbf{P}_{\mu\mu} = \mathbf{P}_{\mu 1} + \mathbf{P}_{\mu 2} = \mathbf{P}_{\mathbf{X} \rightarrow \mu\mu}$

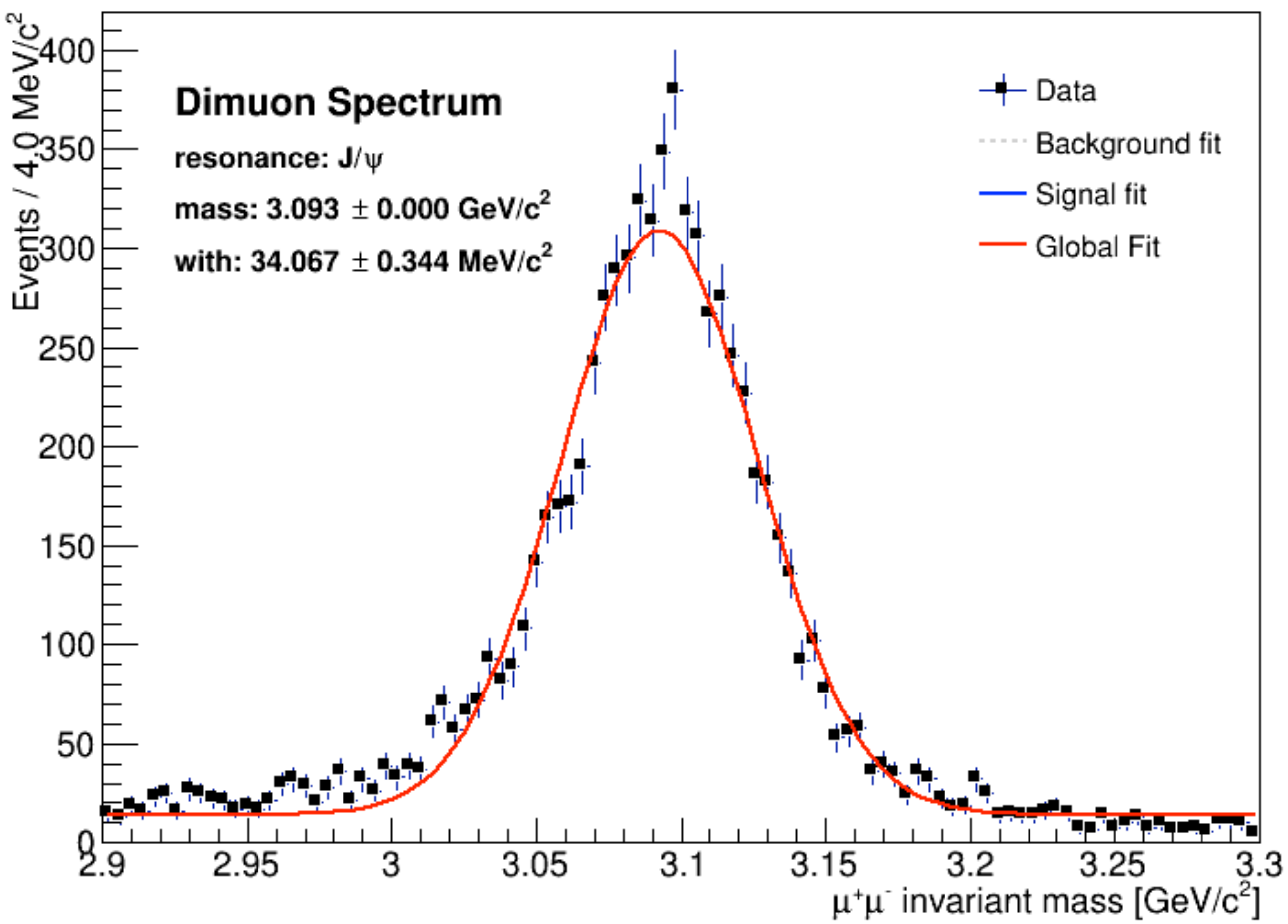
➠ invariant mass $\mathbf{P}_{\mu\mu} \cdot \mathbf{P}_{\mu\mu} = \mathbf{M}_{\mu\mu}^2 = (\mathbf{M}_\mathbf{X})^2$

the reconstructed di-muon spectrum



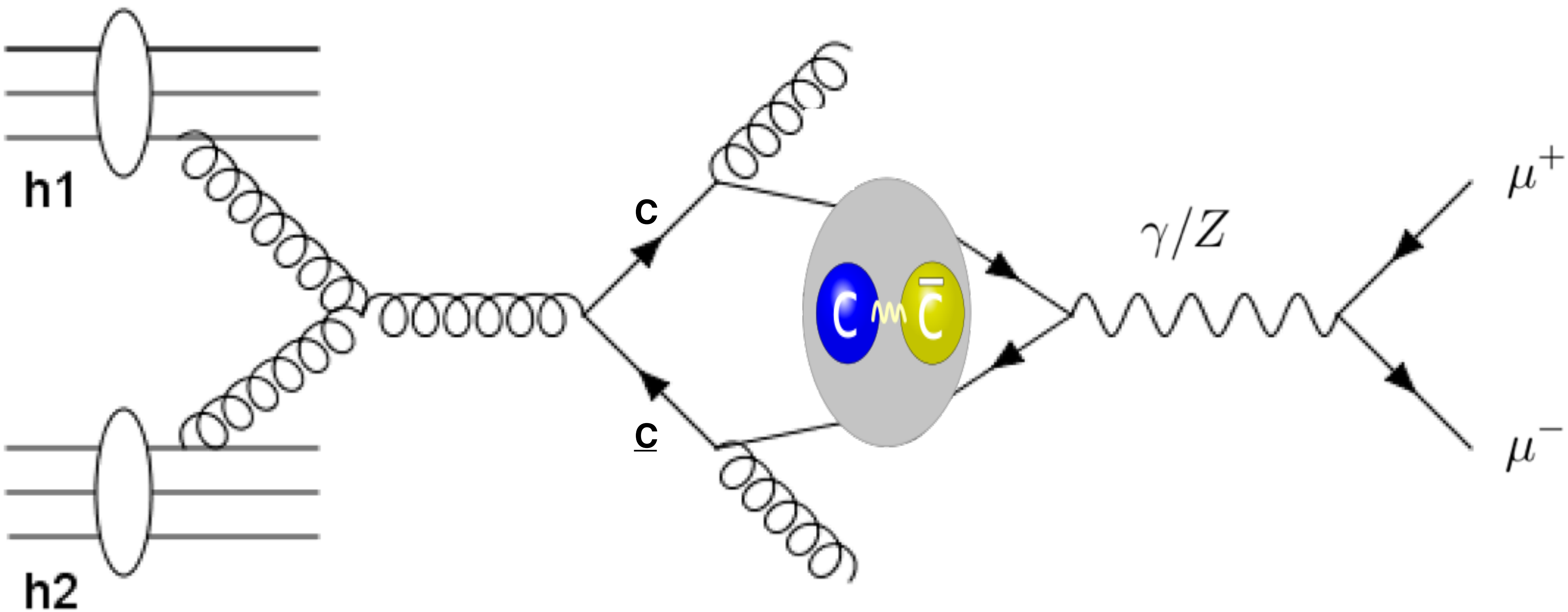
feature: variable bin widths, resolution-dependent, properly normalized, doubly-log scales

fit the data



- establish a **fit model**
  - signal; Gaussian
  - background: polynominal
- extract **signal parameters**
  - yield ($N \pm \sigma_N$), mass ($m \pm \sigma_m$)
- estimate **systematic errors**
  - does the choice of fit model affect the measured results ?
  - quantify the systematic variations by employing different models
- quote **final measurements**
  - $N \pm \sigma_{stat} \pm \sigma_{syst}$
- inspect quality of fit
  - can model be improved?
  - hint: final state radiation ($\mu \rightarrow \mu\gamma$) may distort shape

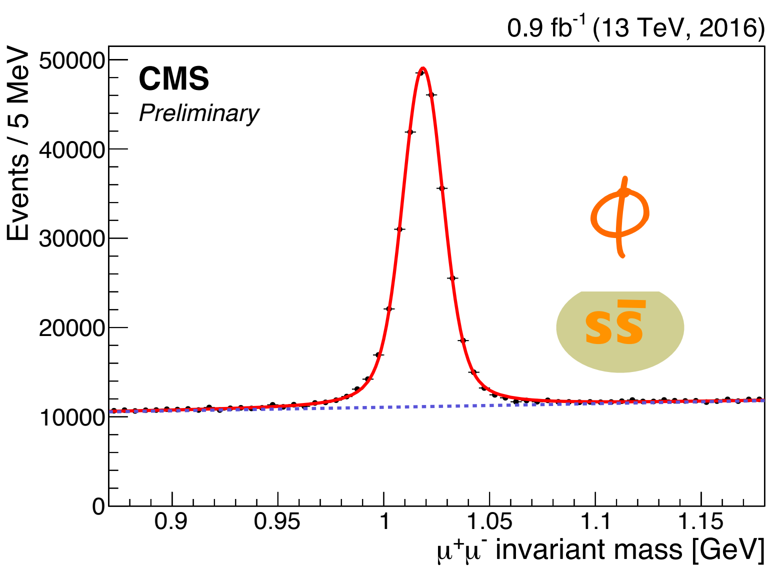
what's the physics process ?

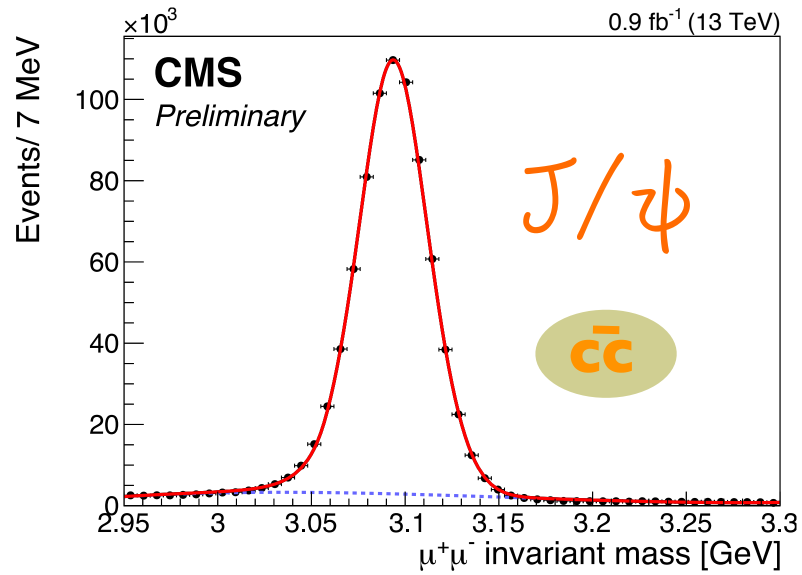


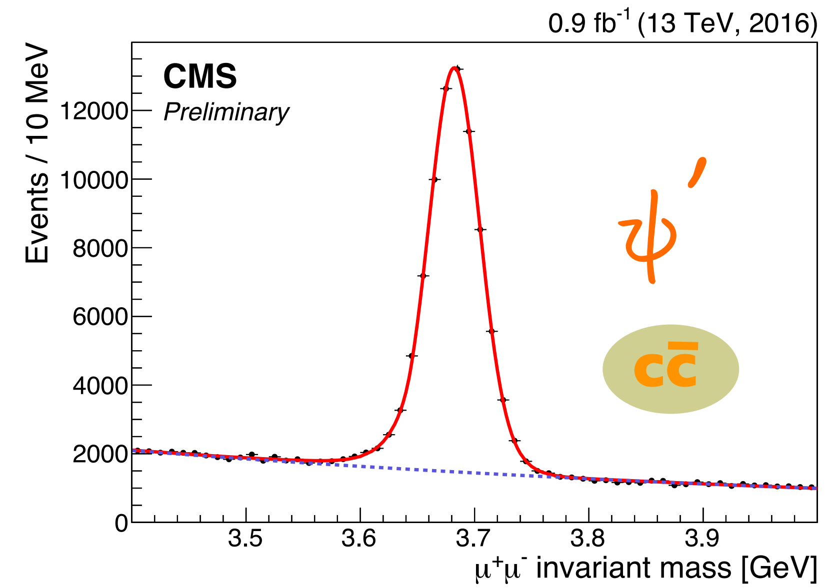
production: strong force

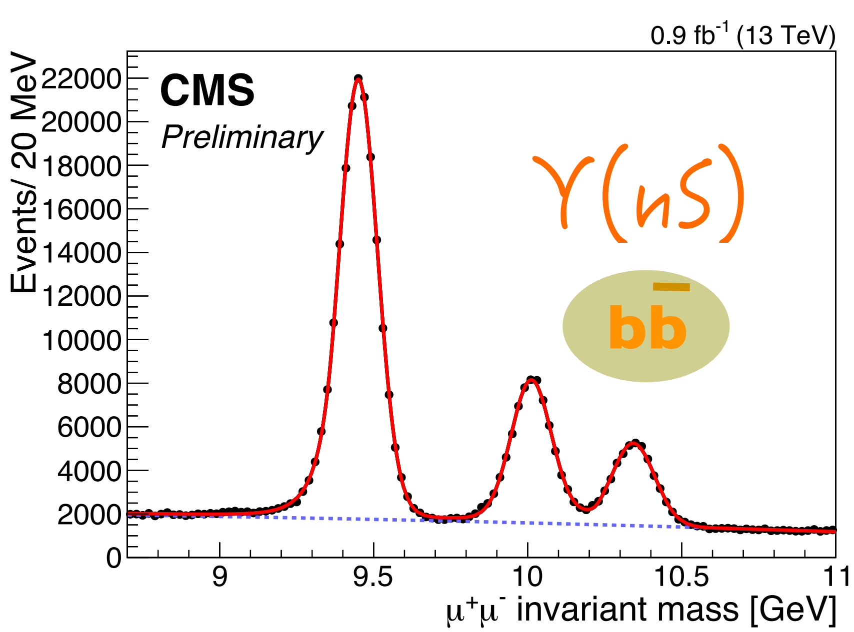
decay: electroweak force

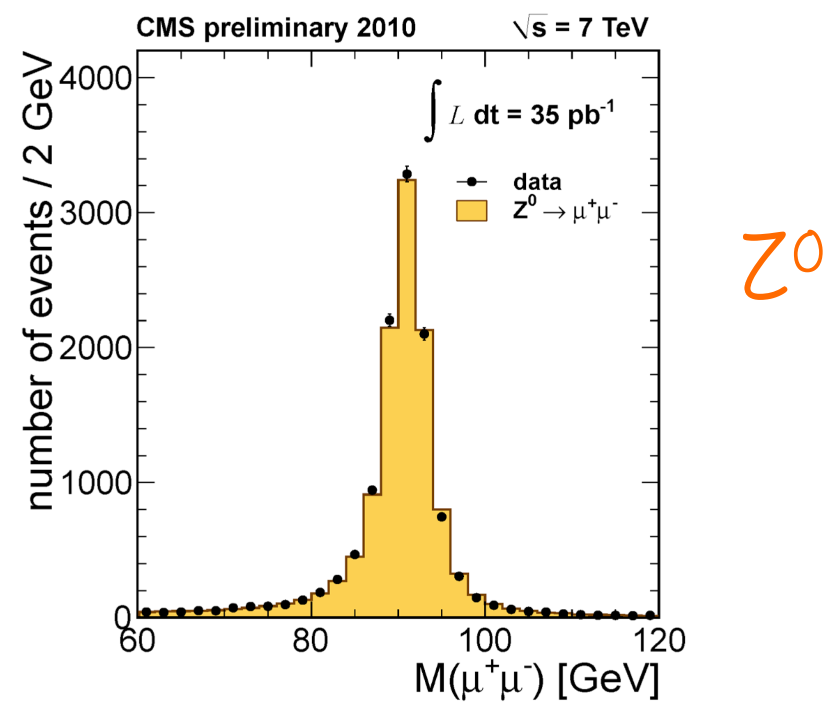
what are the peaks?











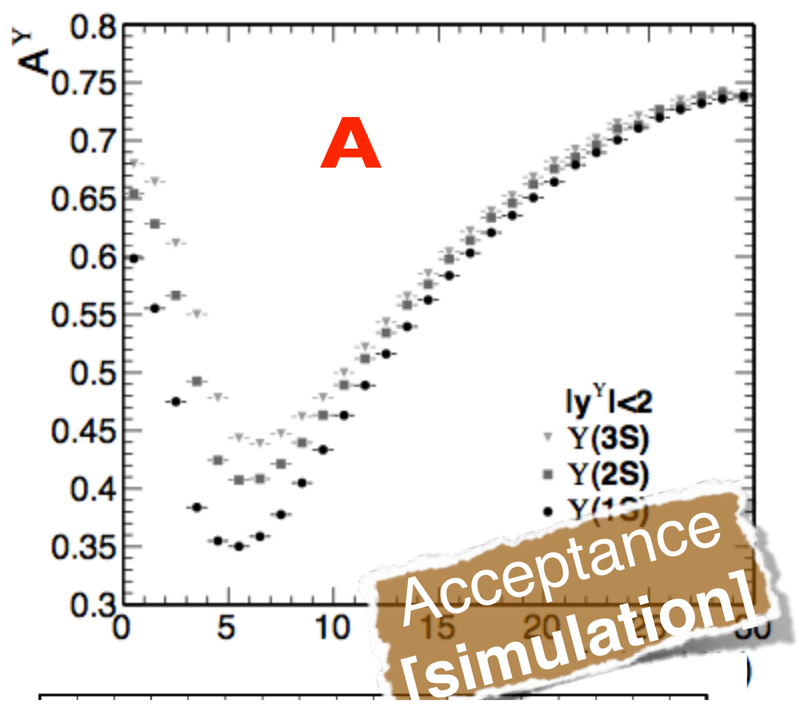
Check their measured properties from: http://pdglive.lbl.gov

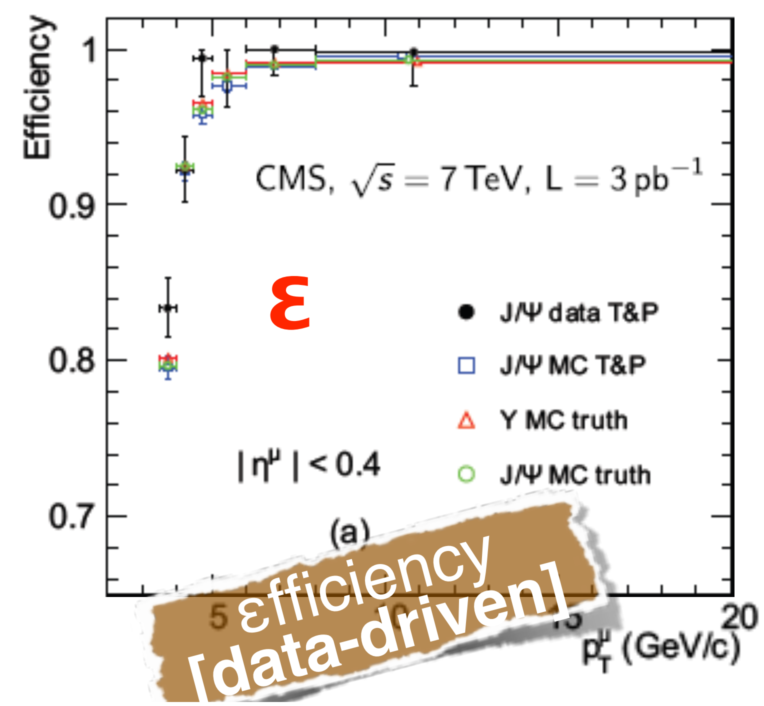
production cross section

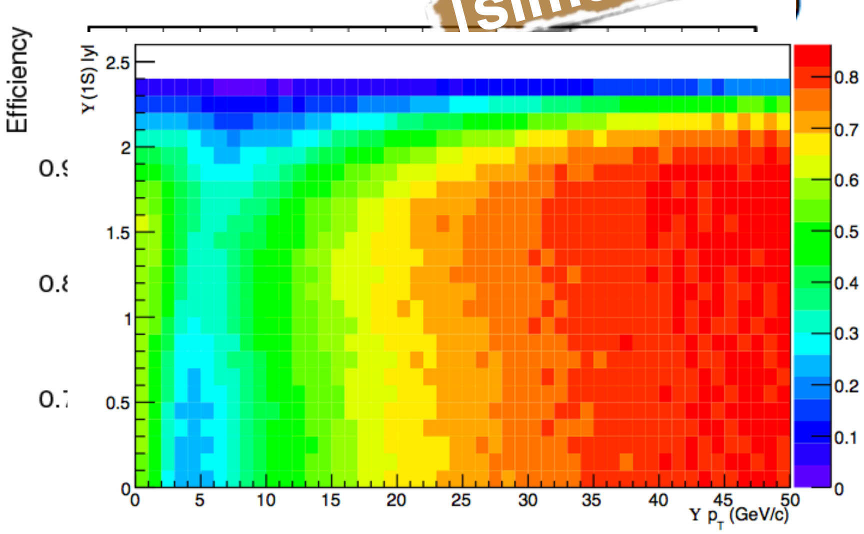
"N=L.σ"

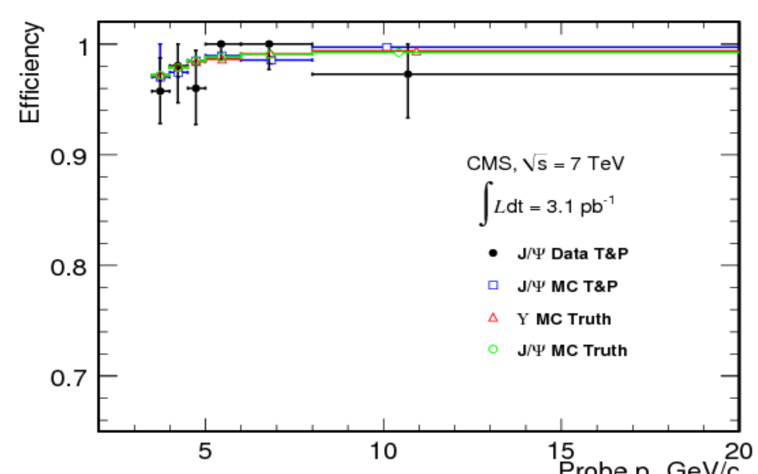
an effective area of interaction
unit: barn, 1b = 10^{-28} m² = 100fm²

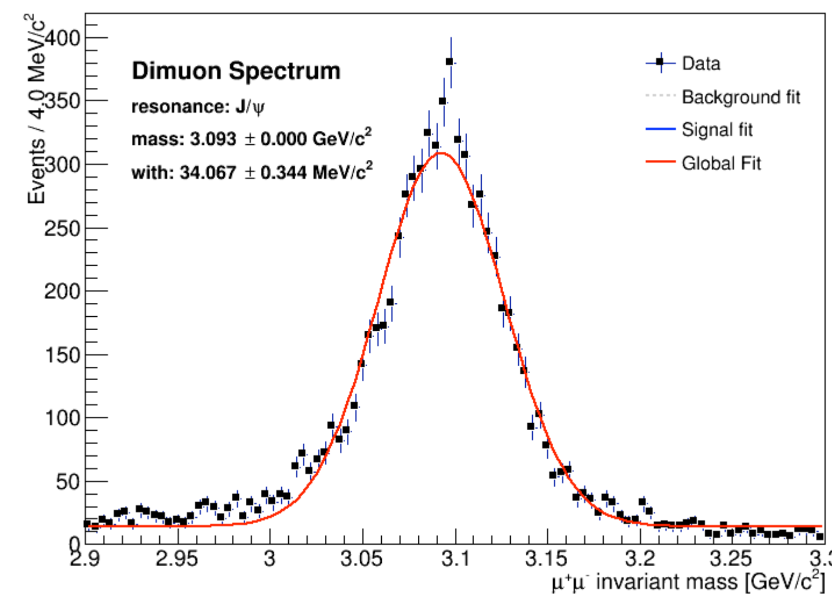
$$\frac{d^2\sigma(Q\overline{Q})}{dp_T dy}\mathcal{B}\left(Q\overline{Q}\rightarrow\mu^+\mu^-\right)=\frac{N_{fit}(Q\overline{Q})}{\mathcal{L}\cdot\mathcal{A}\cdot\epsilon\cdot\Delta p_T\cdot\Delta y}$$











- N: fitted signal yield
- A: detector acceptance from simulation
- ε: detector reconstruction and trigger efficiencies (simulation or data-driven)
- L: integrated sample luminosity

overview of your analysis exercise

check exercise instructions !

1. Pick a peak
   - you have several to choose from ;)
2. Place selection cuts
   - inspect the muon kinematic distributions
   - let's require a p_T threshold on each muon at 10 GeV/c
   - how are the signal yields affected? what's the selection efficiency?
3. Extract the fit result
   - signal yield, signal mass
   - statistical uncertainties included
4. Systematic effects
   - implement different models for signal and background
   - repeat the fit and extract the systematic uncertainties
5. extra: perform a differential measurement
   - produce yield plot as a function of p_T and rapidity

☑ congratulations: you've grasped the ingredients of a physics measurement, the production cross-section of your chosen particle!