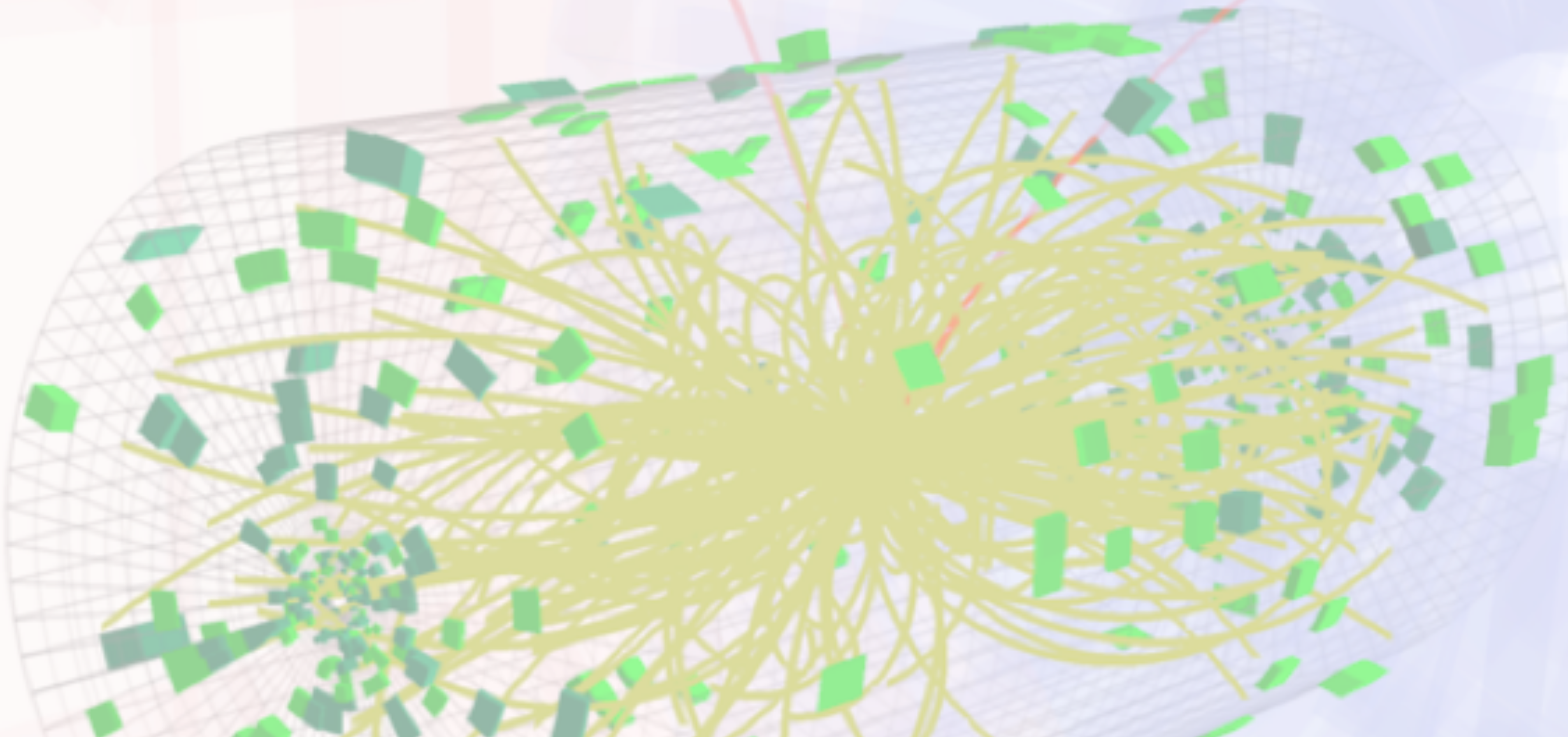
Tutorial on

Data Analysis

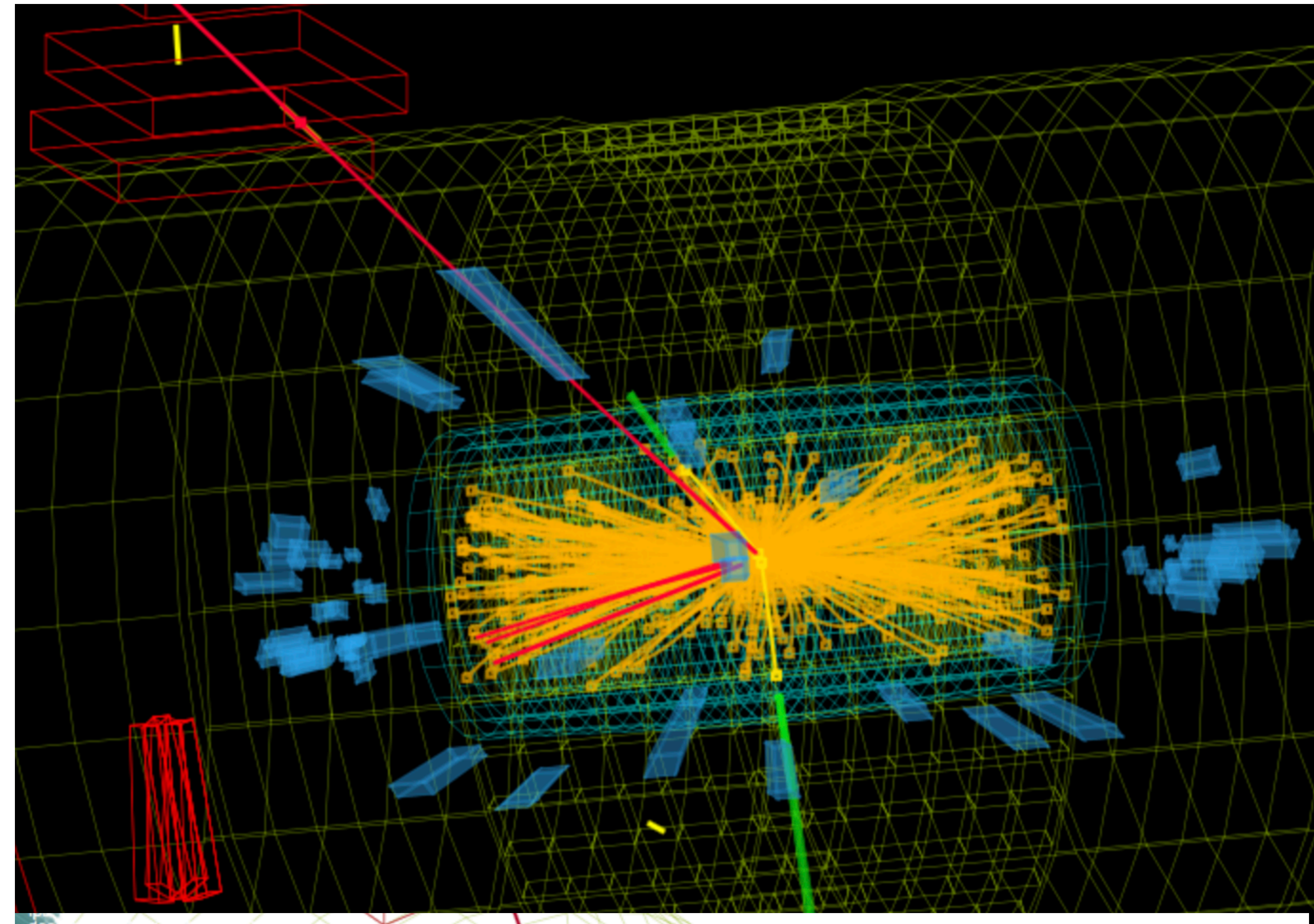


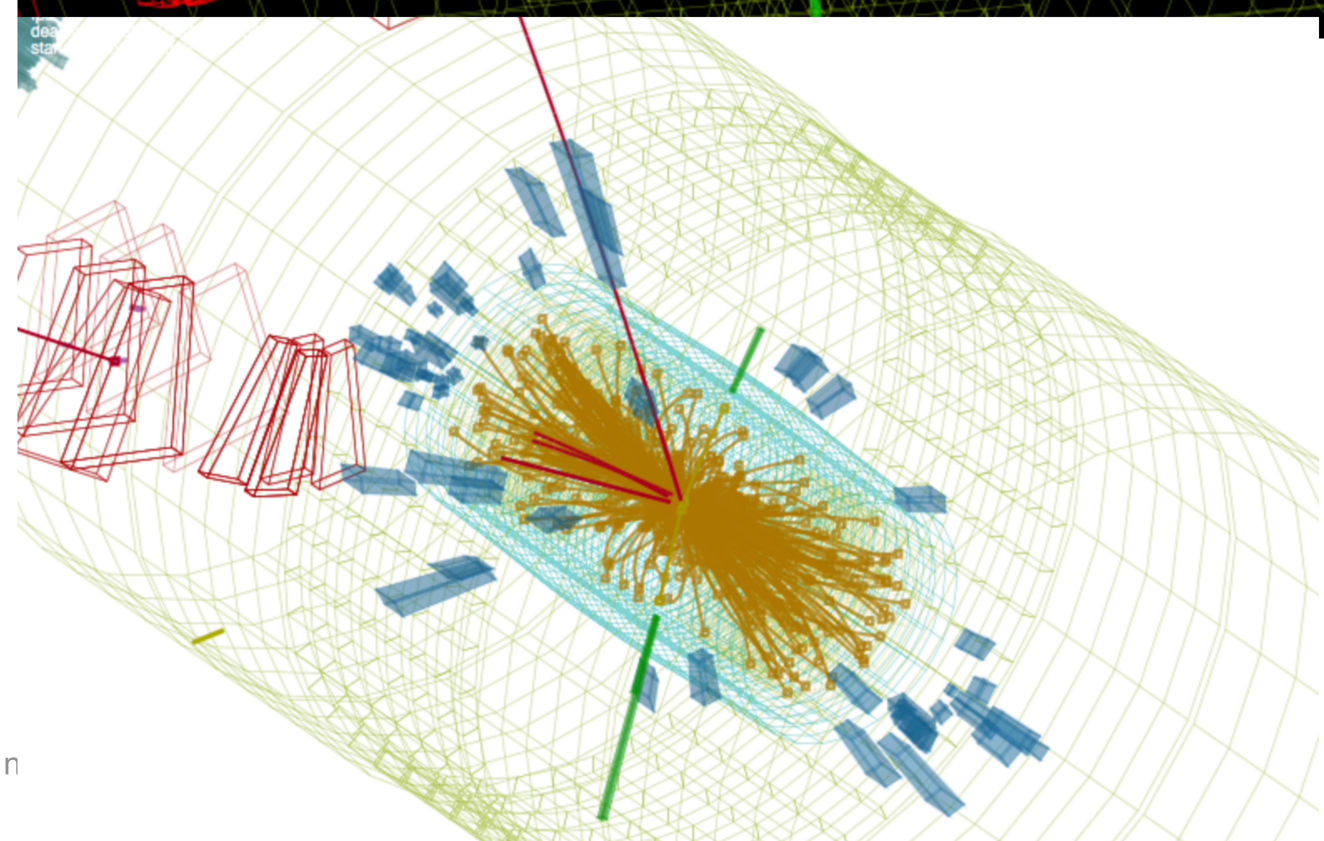
LIP Internship Program, 5th edition, Summer 2021

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

LIP
LABORATÓRIO DE INSTRUMENTAÇÃO
E FÍSICA EXPERIMENTAL DE PARTÍCULAS

LHC Open Data





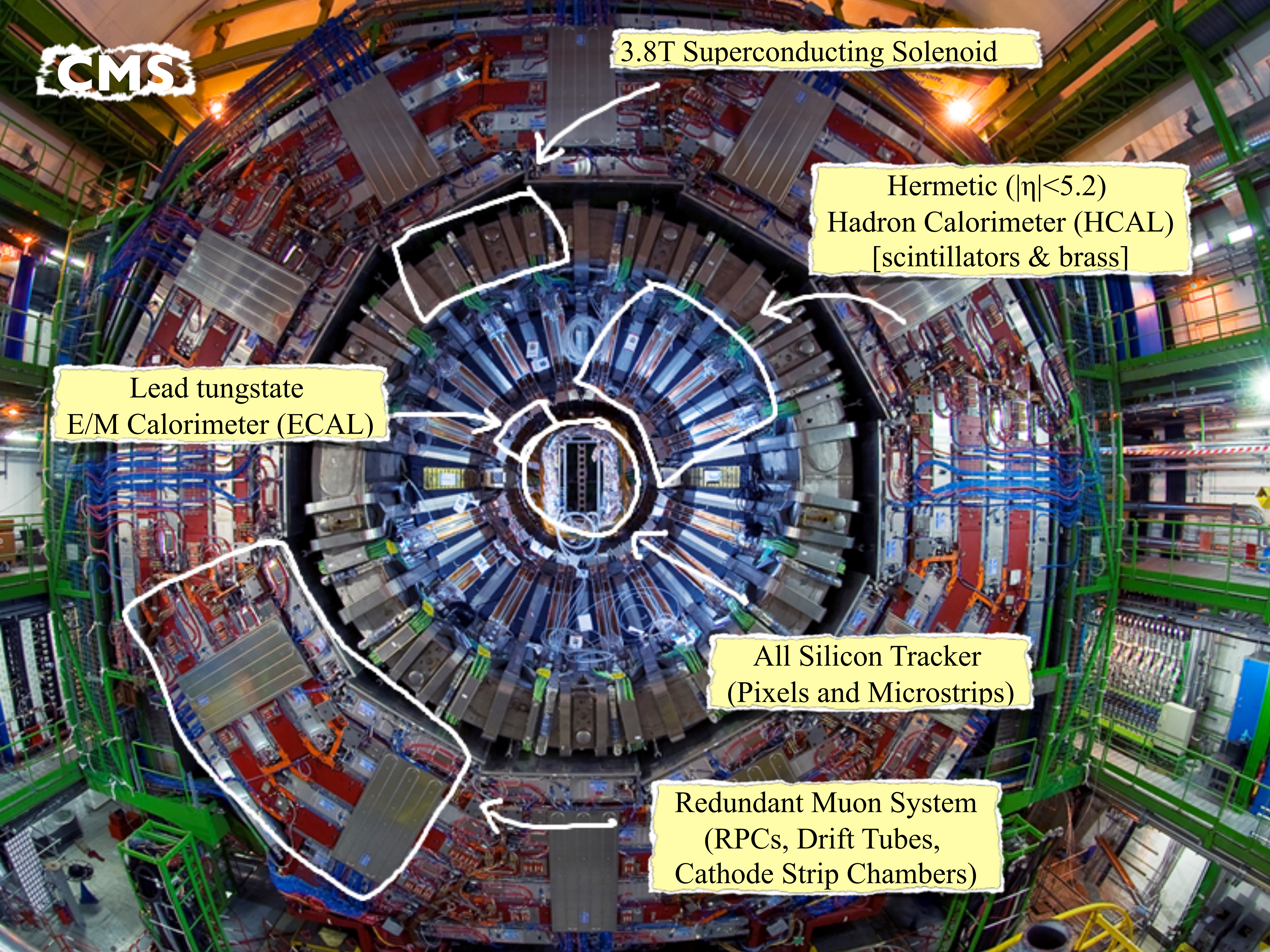
- the LHC collaborations make good chunks of their data publicly available
  - http://**opendata.cern**.ch/
- along with tools & software & examples
- for data visualisation and analysis
- from event reconstruction algorithms to machine learning challenges
- via virtual machines (with no need to install 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

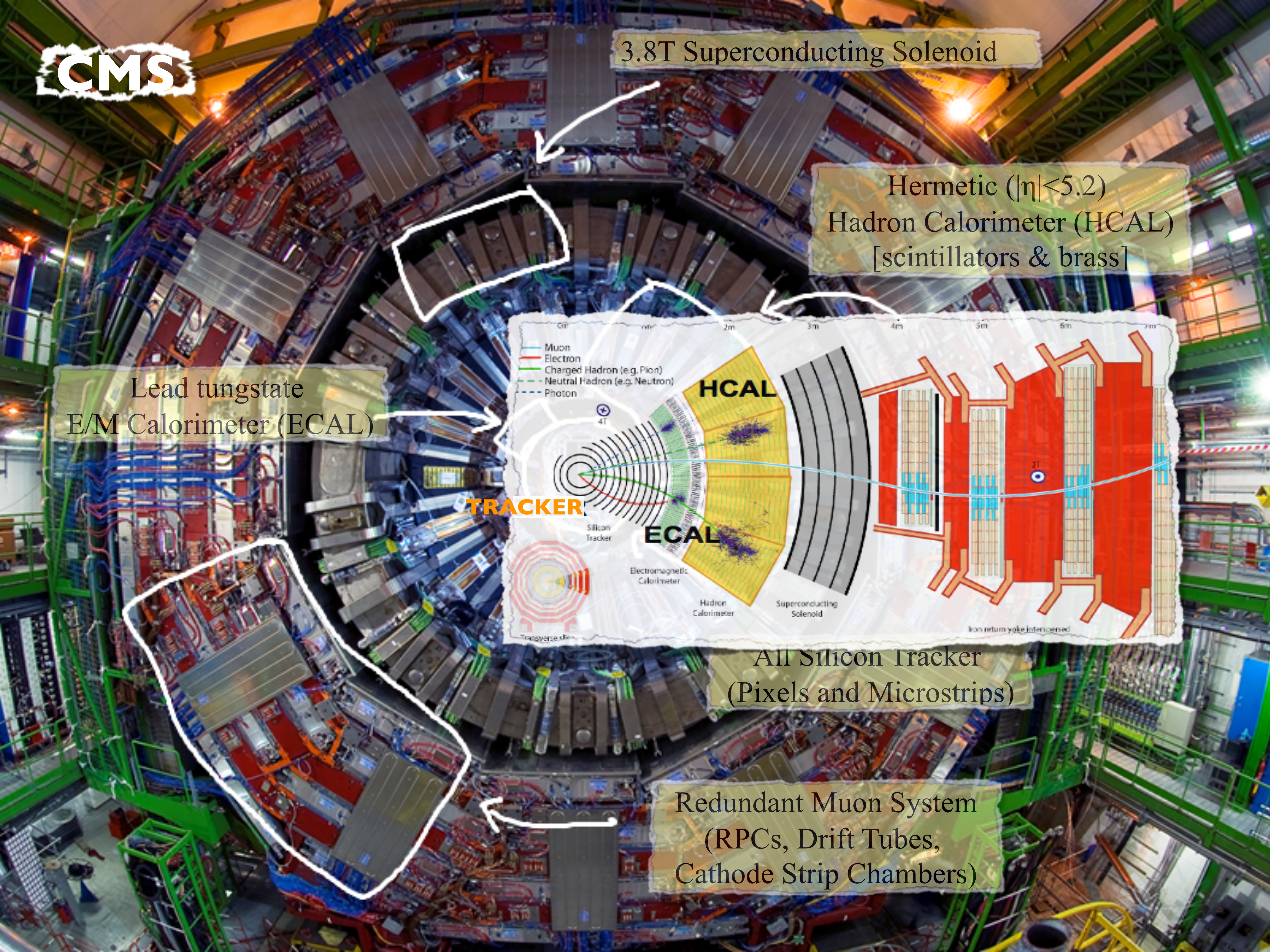
goals

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 detector/trigger effects
- extract physics parameters from data
  - measure signal yields by performing a likelihood fits
  - inspect statistical and systematic errors
  - perform a differential measurement
  - (bonus) explore (more) advanced fitting techniques

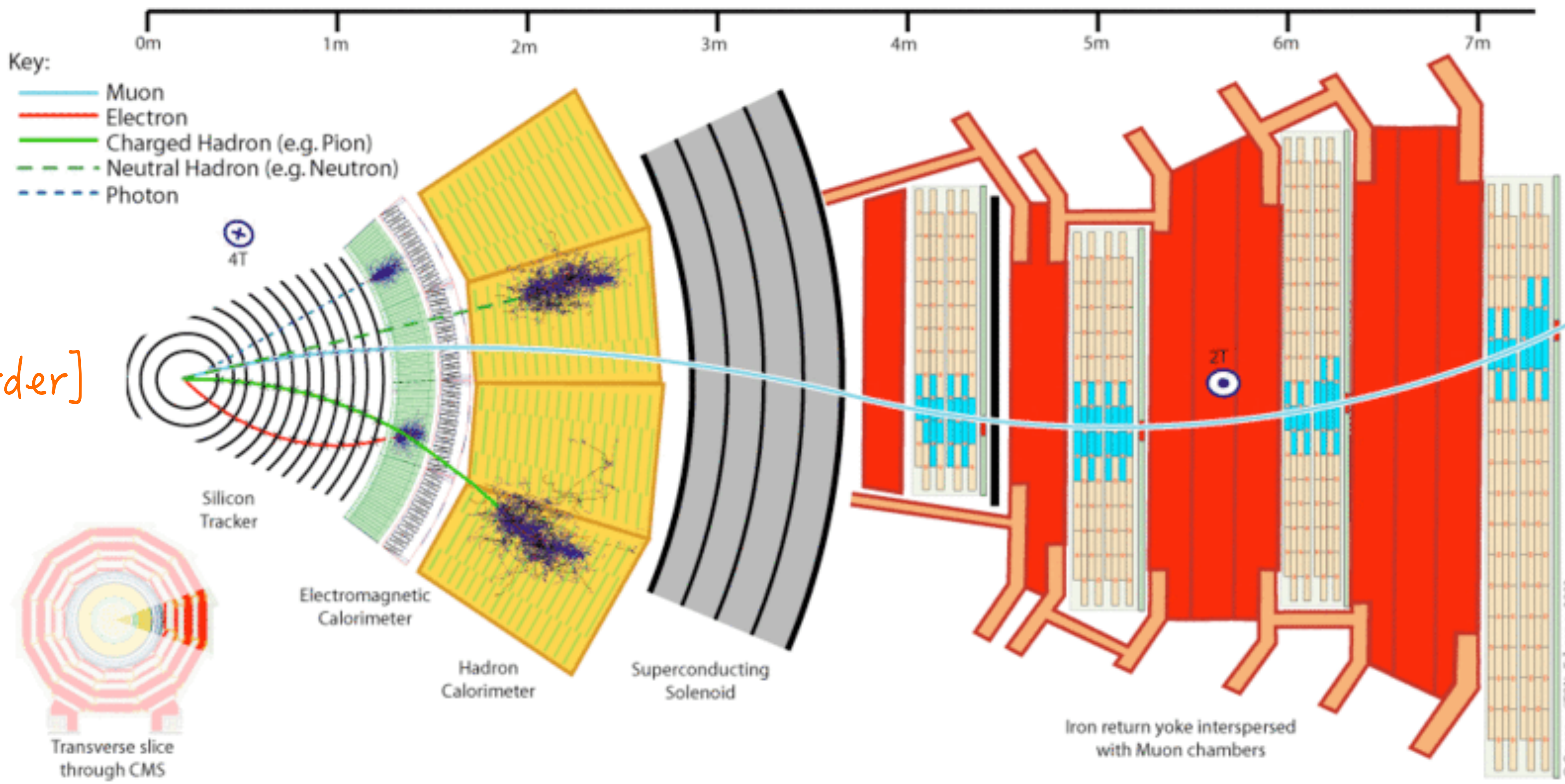
Detector & Event Reconstruction & Visualisation


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)

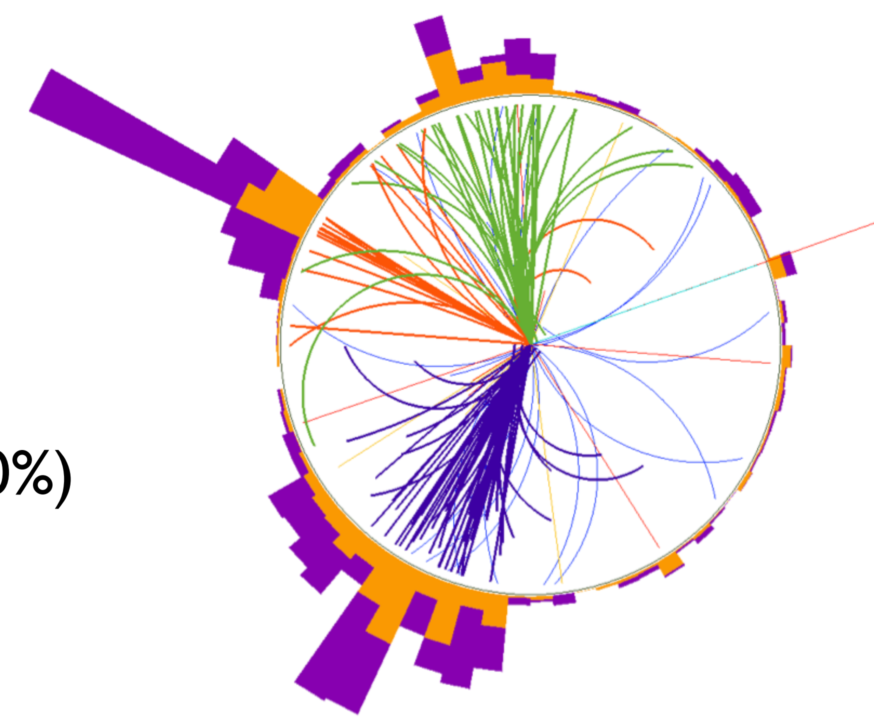

CMS
3.8T Superconducting Solenoid
Hermetic (|η|<5.2)
Hadron Calorimeter (HCAL)
[scintillators & brass]
Lead tungstate
E/M Calorimeter (ECAL)
All Silicon Tracker
(Pixels and Microstrips)
Redundant Muon System
(RPCs, Drift Tubes,
Cathode Strip Chambers)
Muon
Electron
Charged Hadron (e.g. Pion)
Neutral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

particle identification

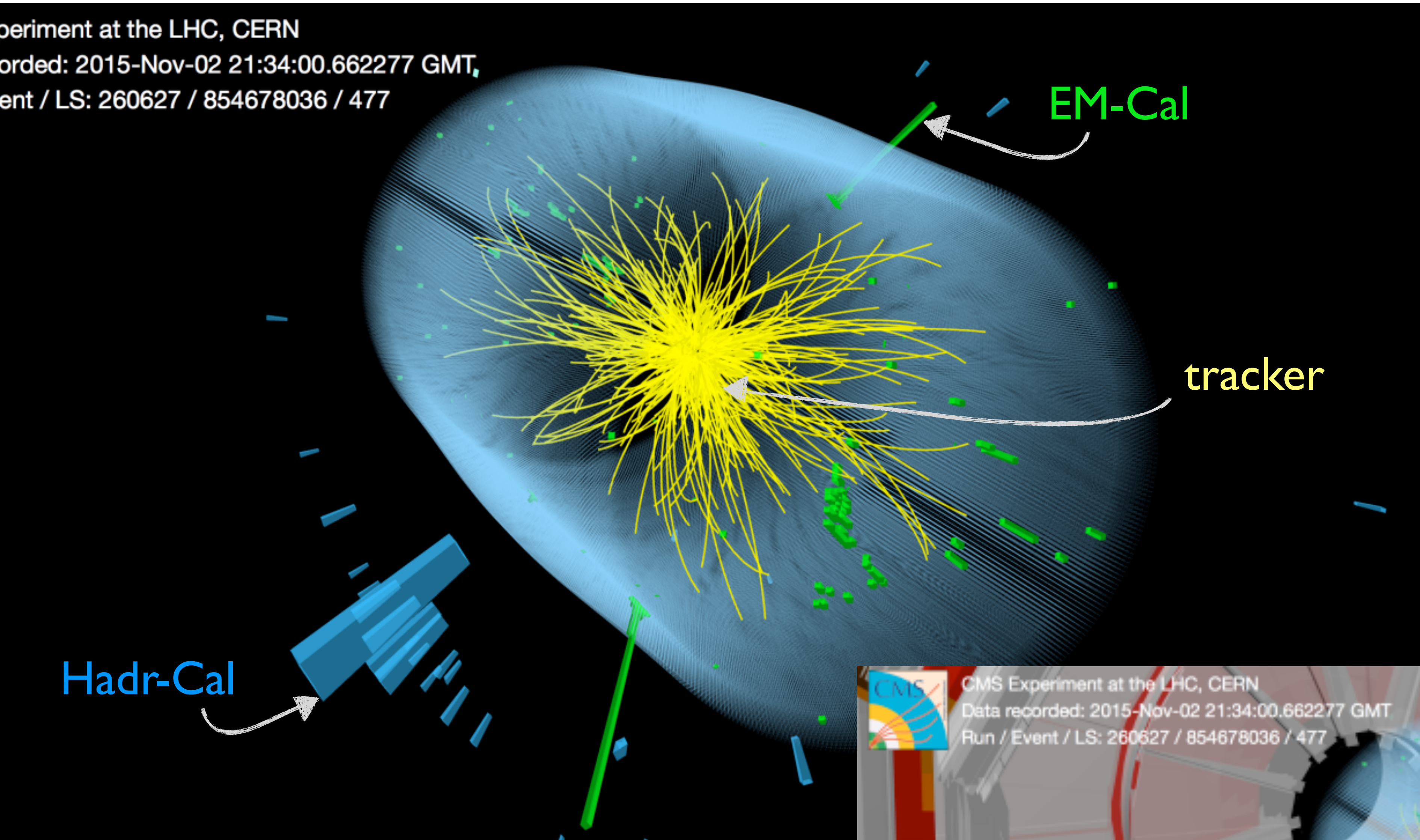
[leading order]



- [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



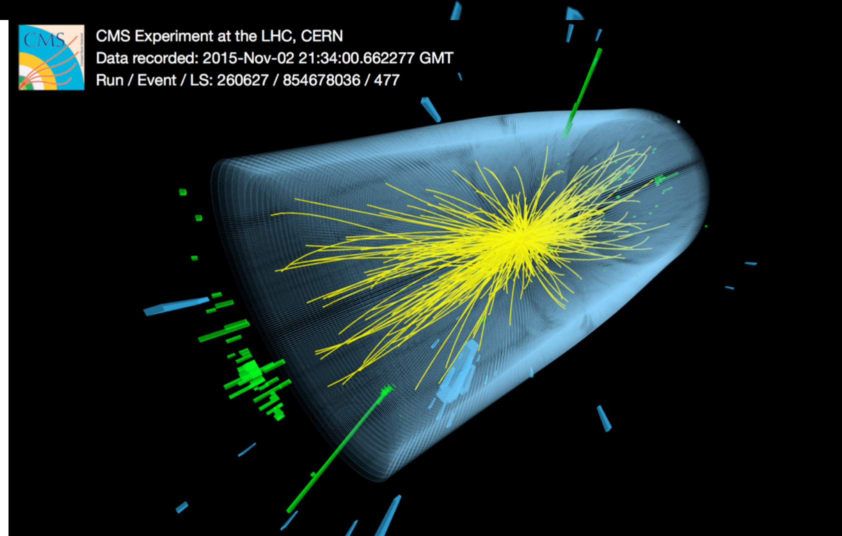
di-photons

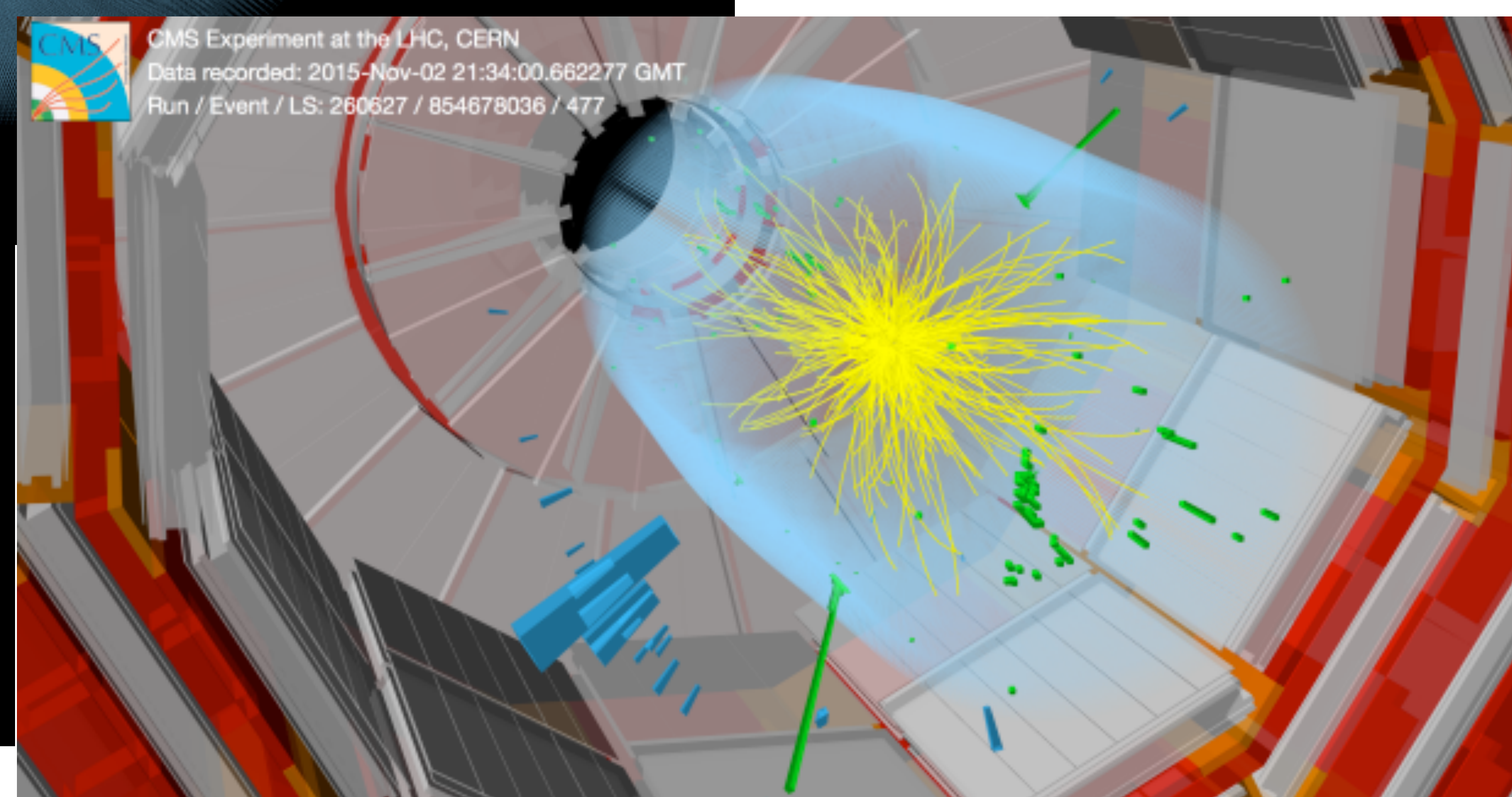


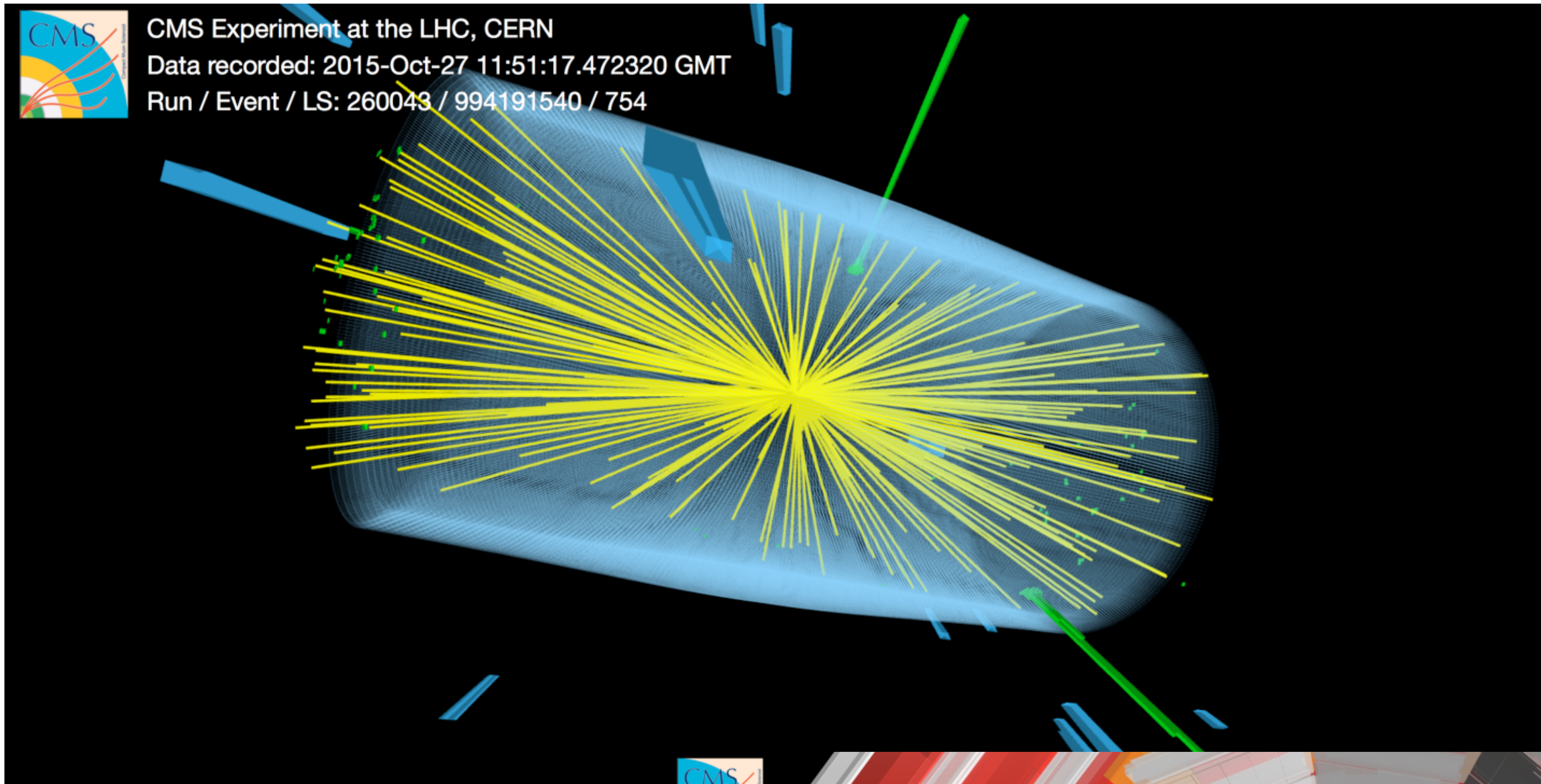
$X \rightarrow \gamma\gamma$

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

CMS-PHO-EVENTS-2015-007

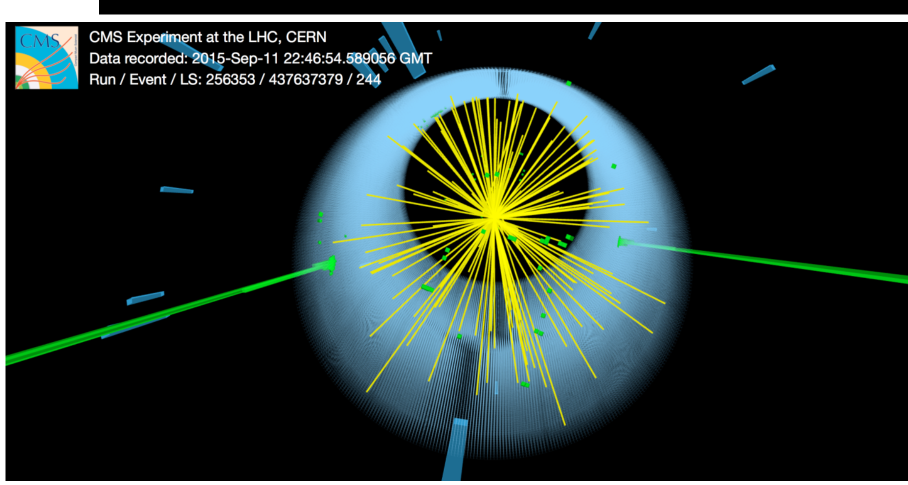


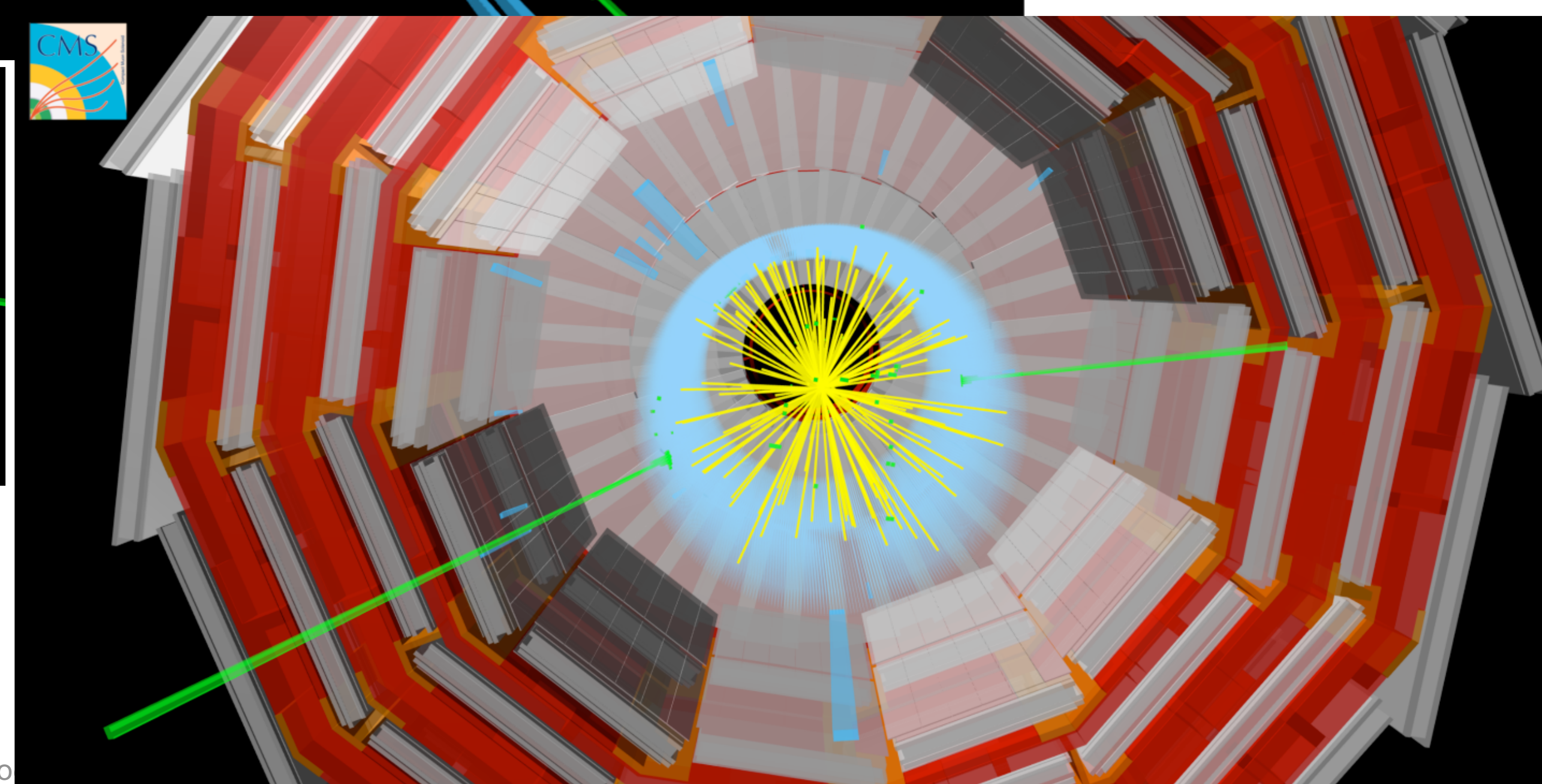


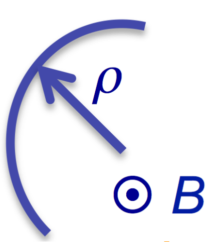


?

$m_{??}$ ~ 800 GeV

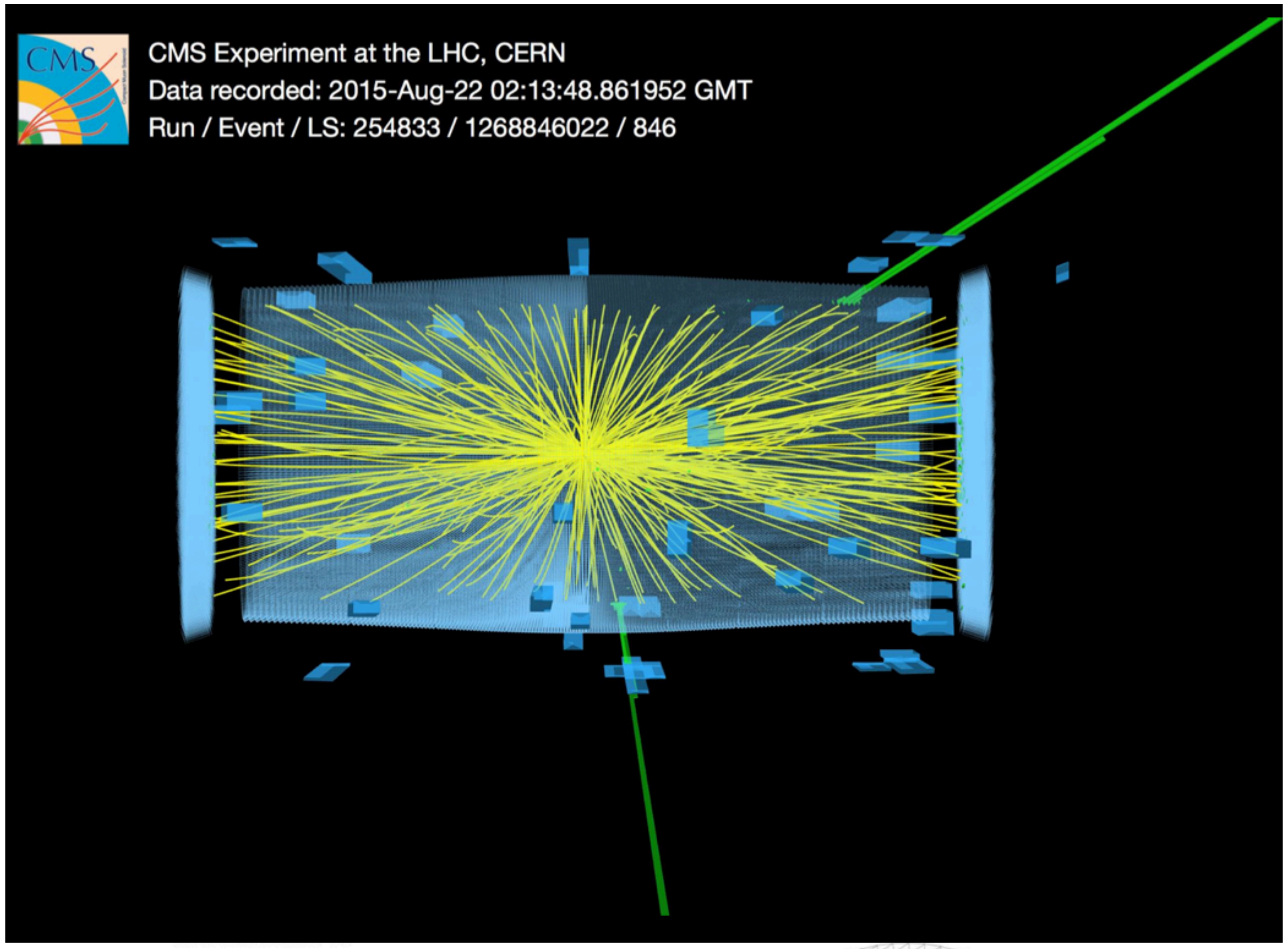




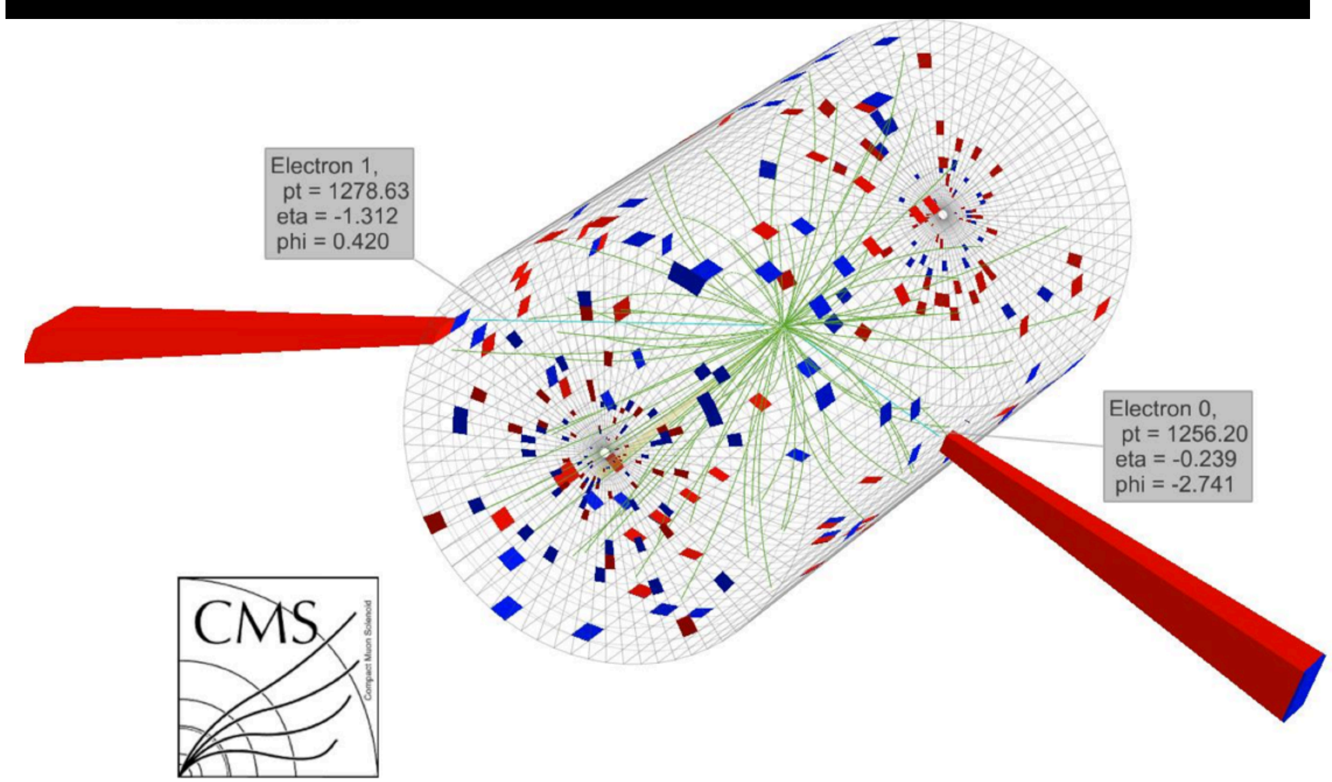


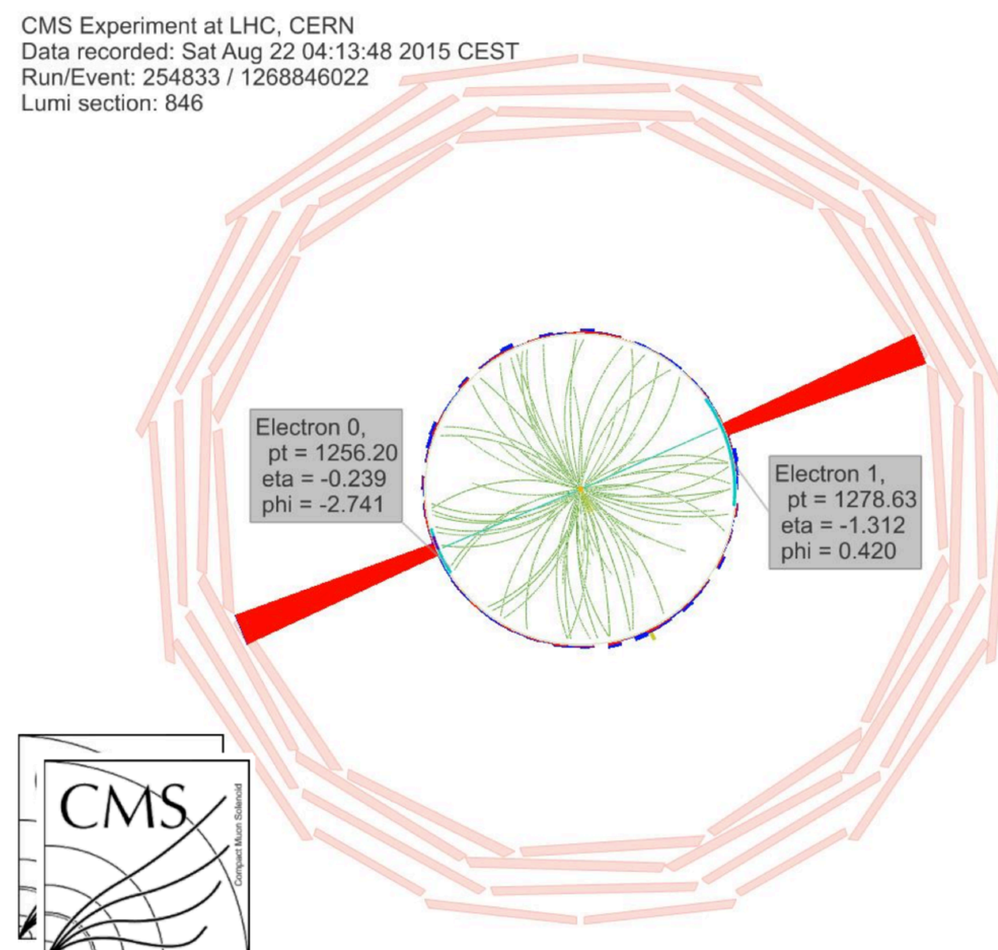
$$\rho = \frac{p}{ZeB}$$

a di-electron event



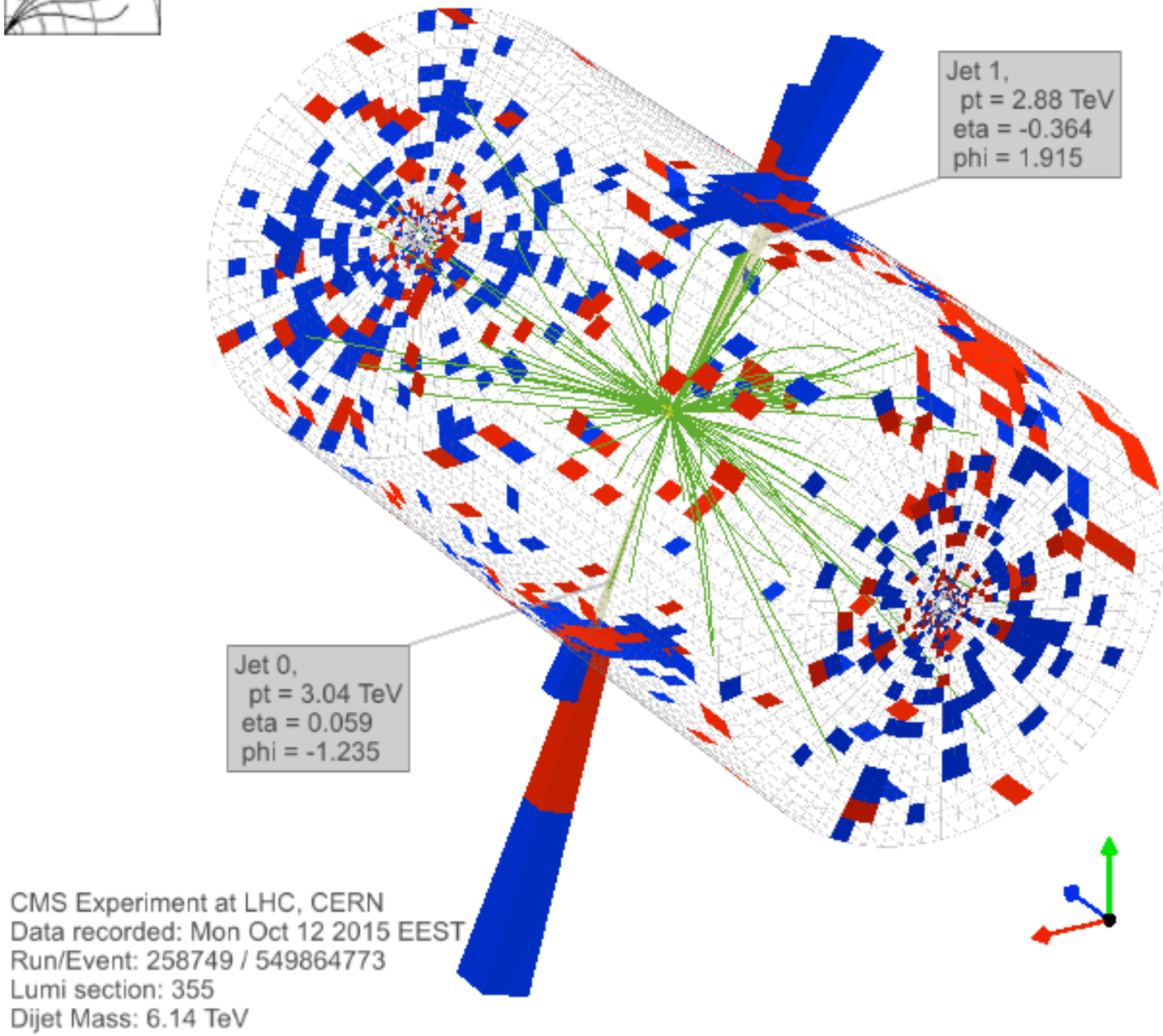
Event Display of a Candidate Electron-Positron Pair with an Invariant Mass of 2.9 TeV

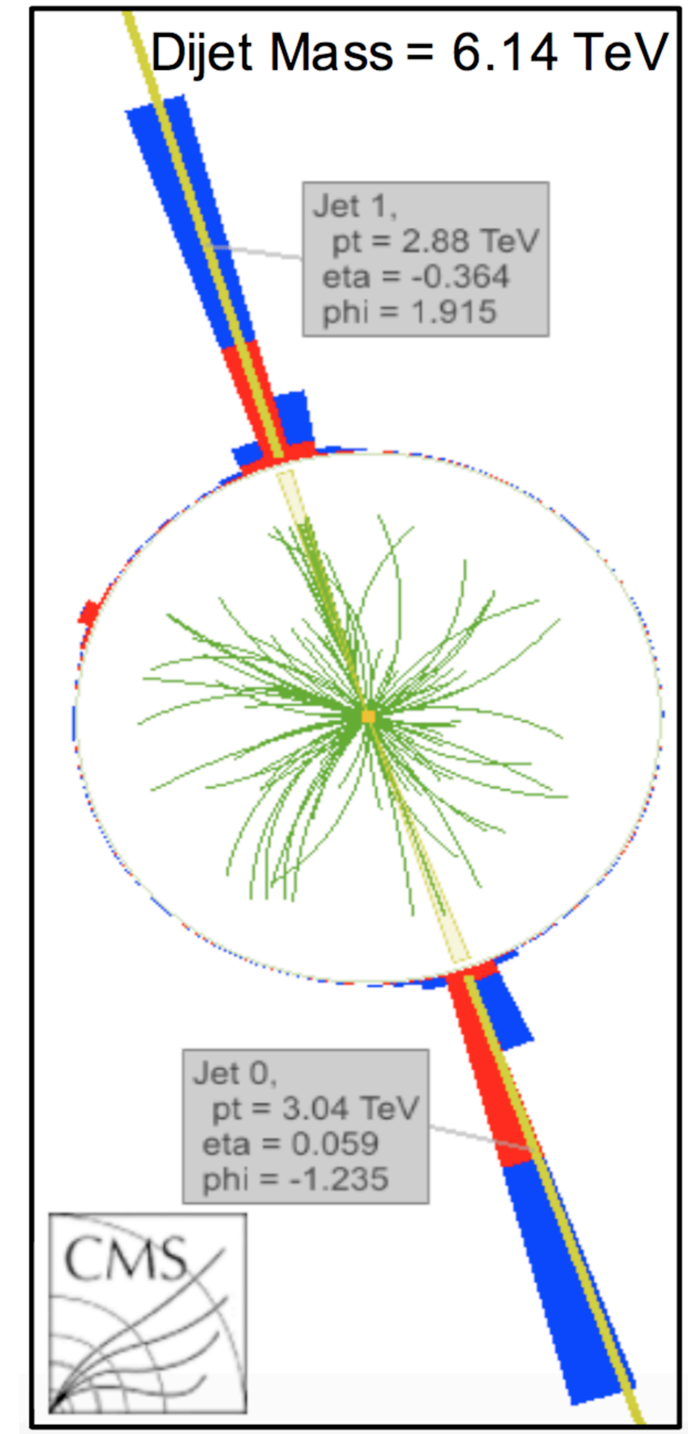




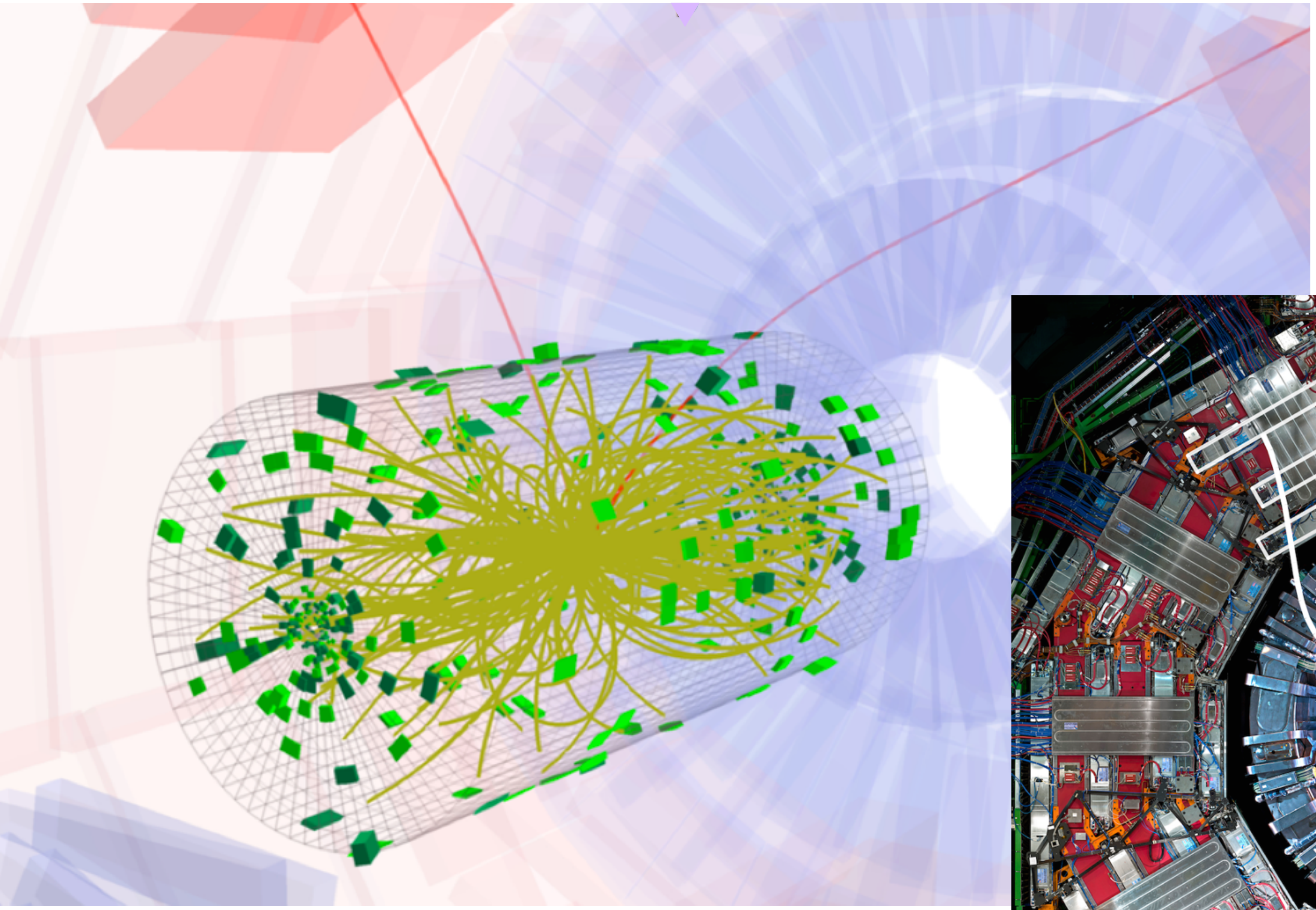
di-jets

$X \to j\,j$

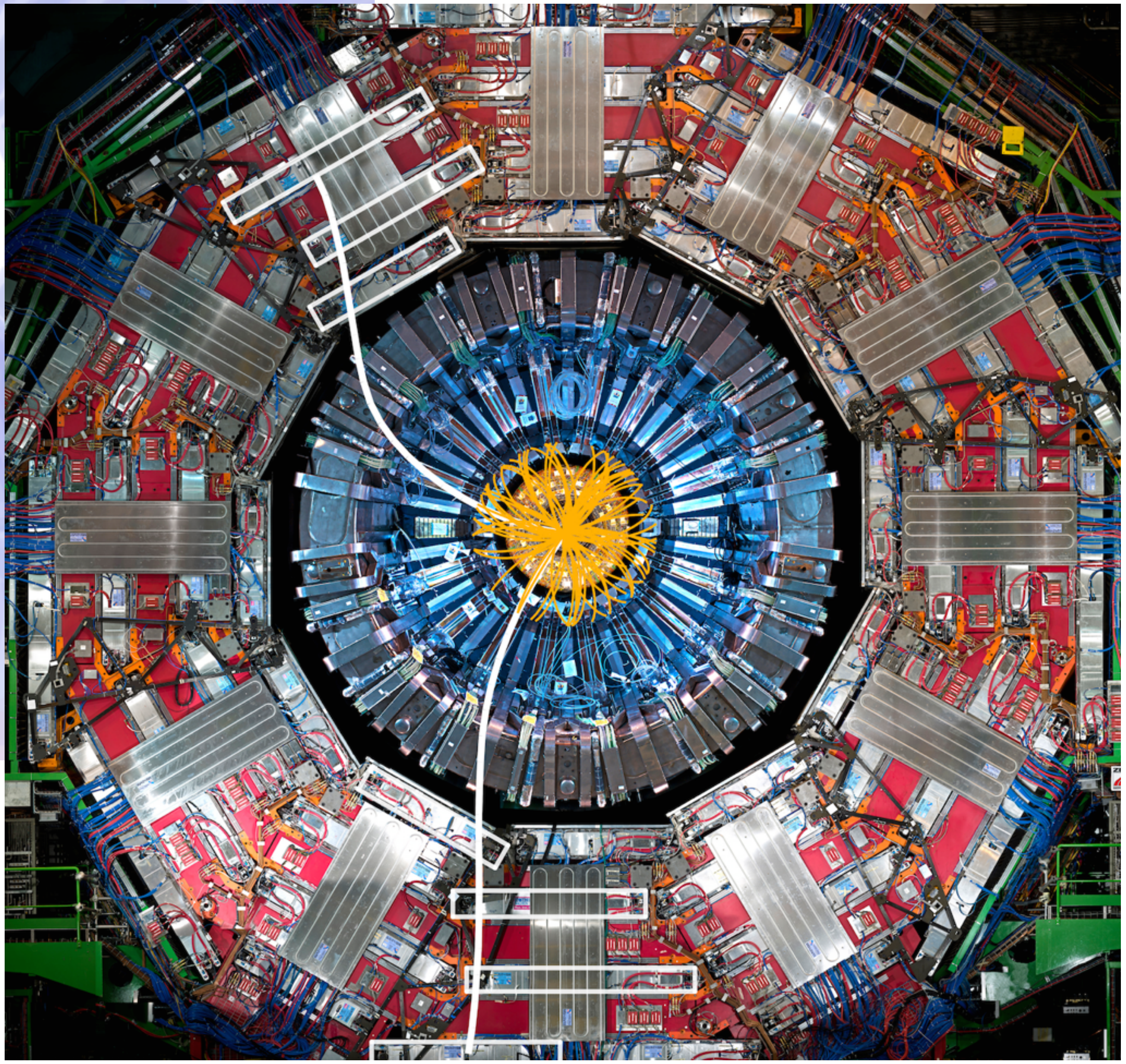




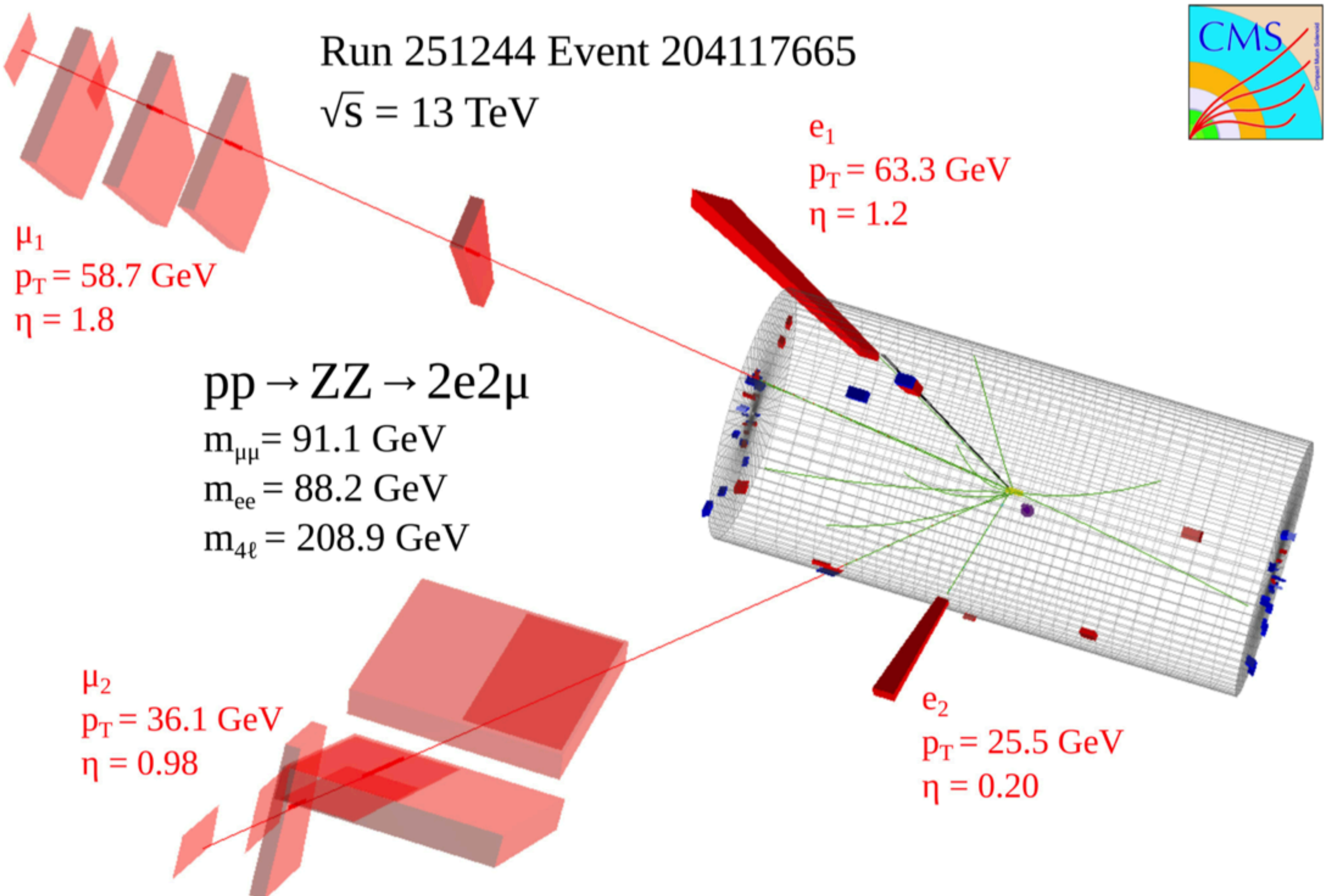
a di-muon event



X →μμ



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

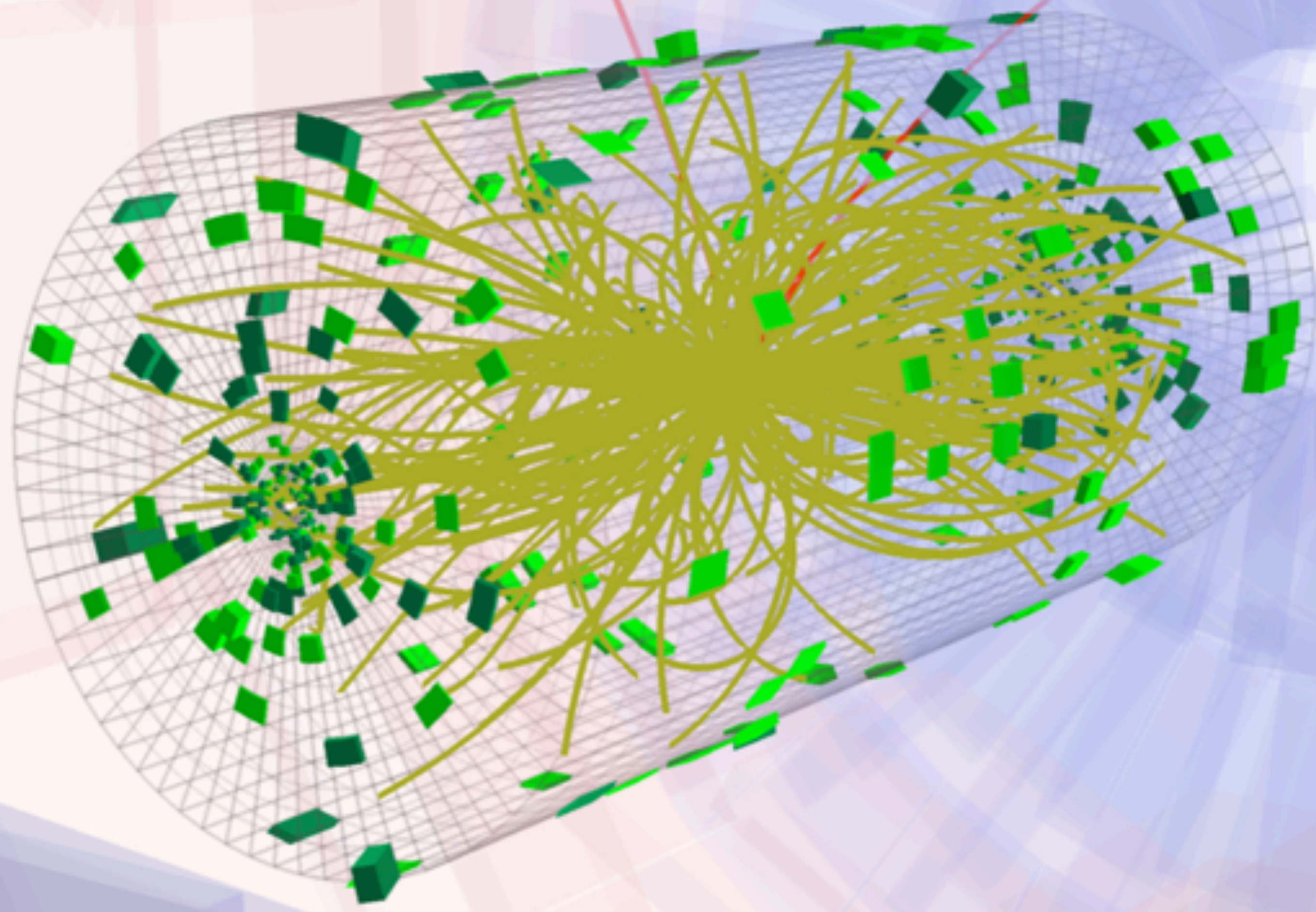


➡️ processes are explored with many (more) final state particles

ingredients of a physics measurement

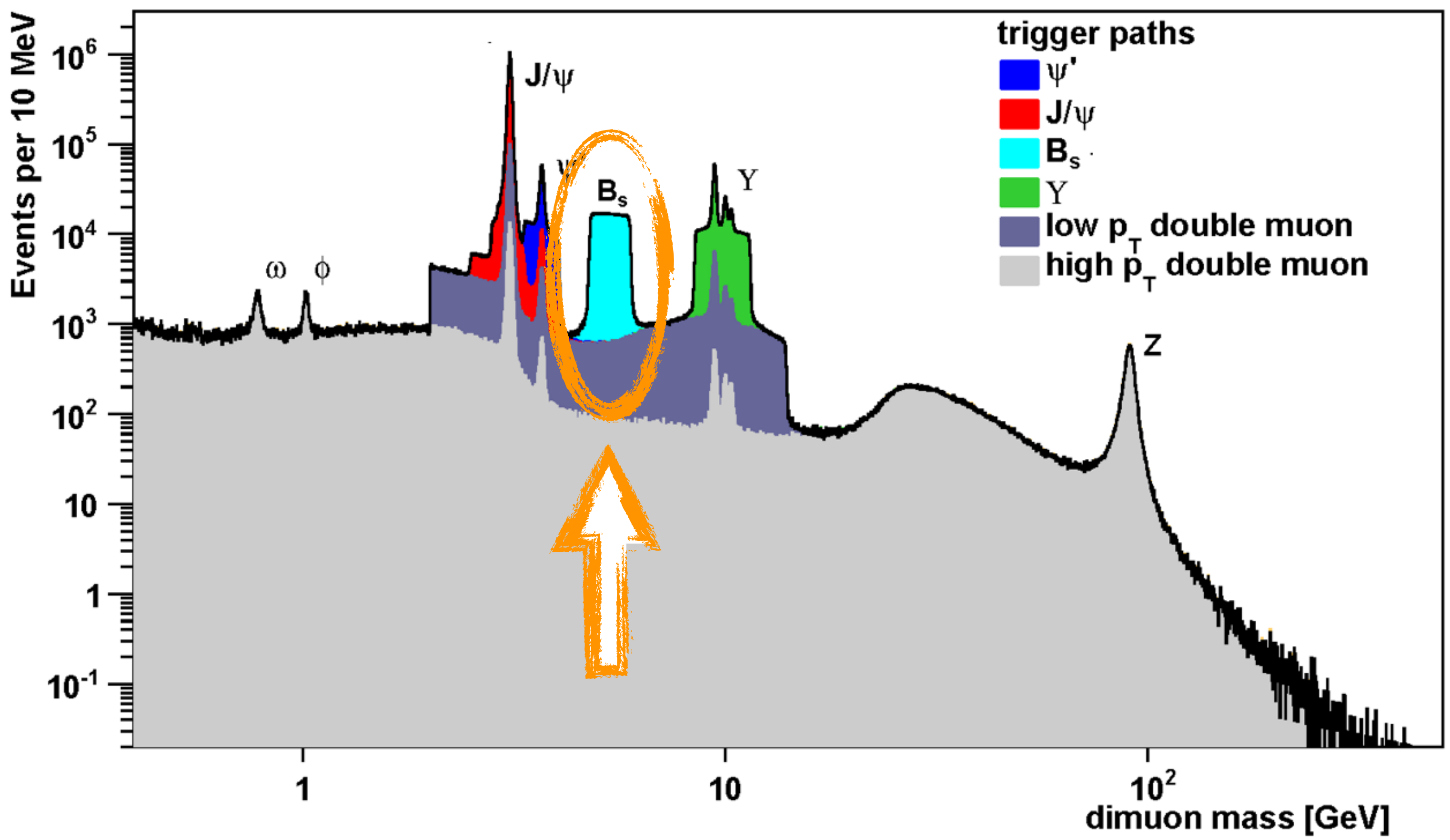
$B_s \to \mu\mu$

the ‘golden’ rare decay



searching for an *ultra-rare* decay: $B \to \mu\mu$

1. ONLINE SELECTION (TRIGGER)



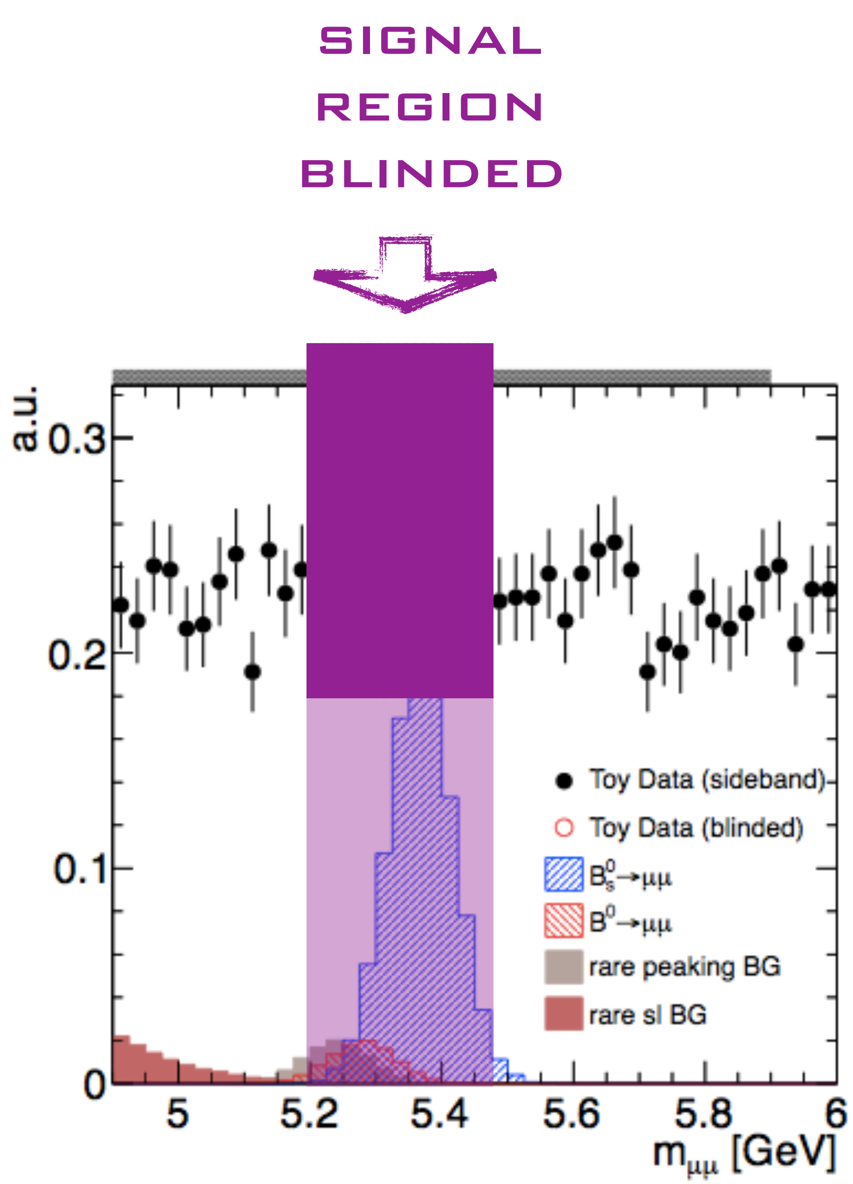
Dimuon Trigger

- L1 Hardware Trigger
  - p_T>3 GeV (few kHz)
- HLT Full tracking and vertexing
- HLT $B_s \to \mu\mu$
  - Leading and sub-leading μ p_T>3,4 (4,4) GeV $|\eta_{\mu\mu}|$<1.8 (1.8<$|\eta_{\mu\mu}|$<2.2)
  - p_T (μμ)>5 (4.8-6) GeV
  - 4.8 <m(μμ)< 6.0 GeV
  - $P(\chi^2/dof)$ >0.5%

searching for an ultra-rare decay: B→μμ

1. ONLINE SELECTION (TRIGGER)

2. BLIND THE DATA (AVOID BIAS)

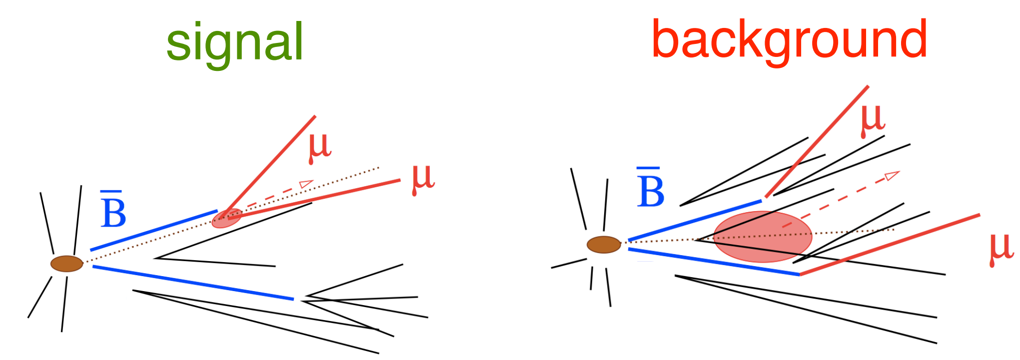


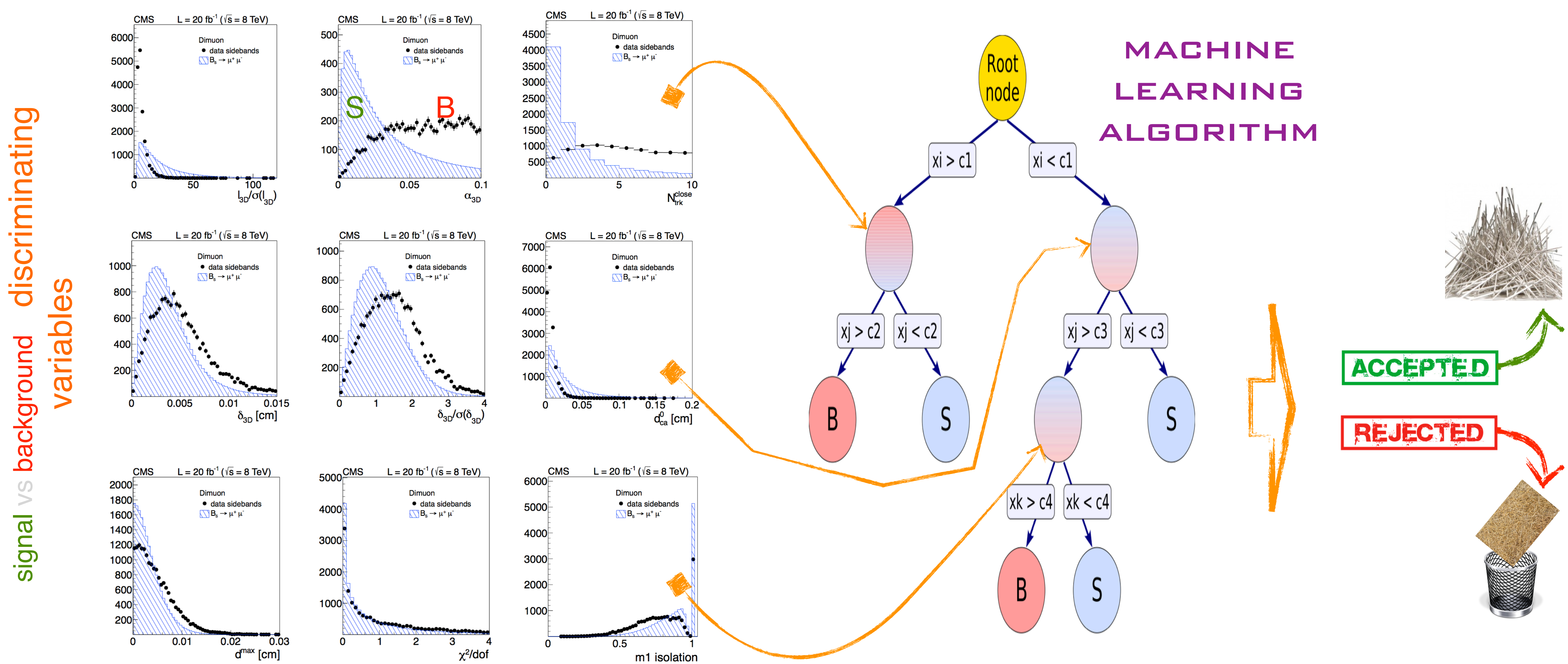
analysis procedure and event selection developed without inspecting the data in region where signal is expected

“box opening” only later, at final analysis stages

searching for an ultra-rare decay: B→μμ

1. ONLINE SELECTION (TRIGGER)
2. BLIND THE DATA (AVOID BIAS)
3. MULTIVARIATE SELECTION

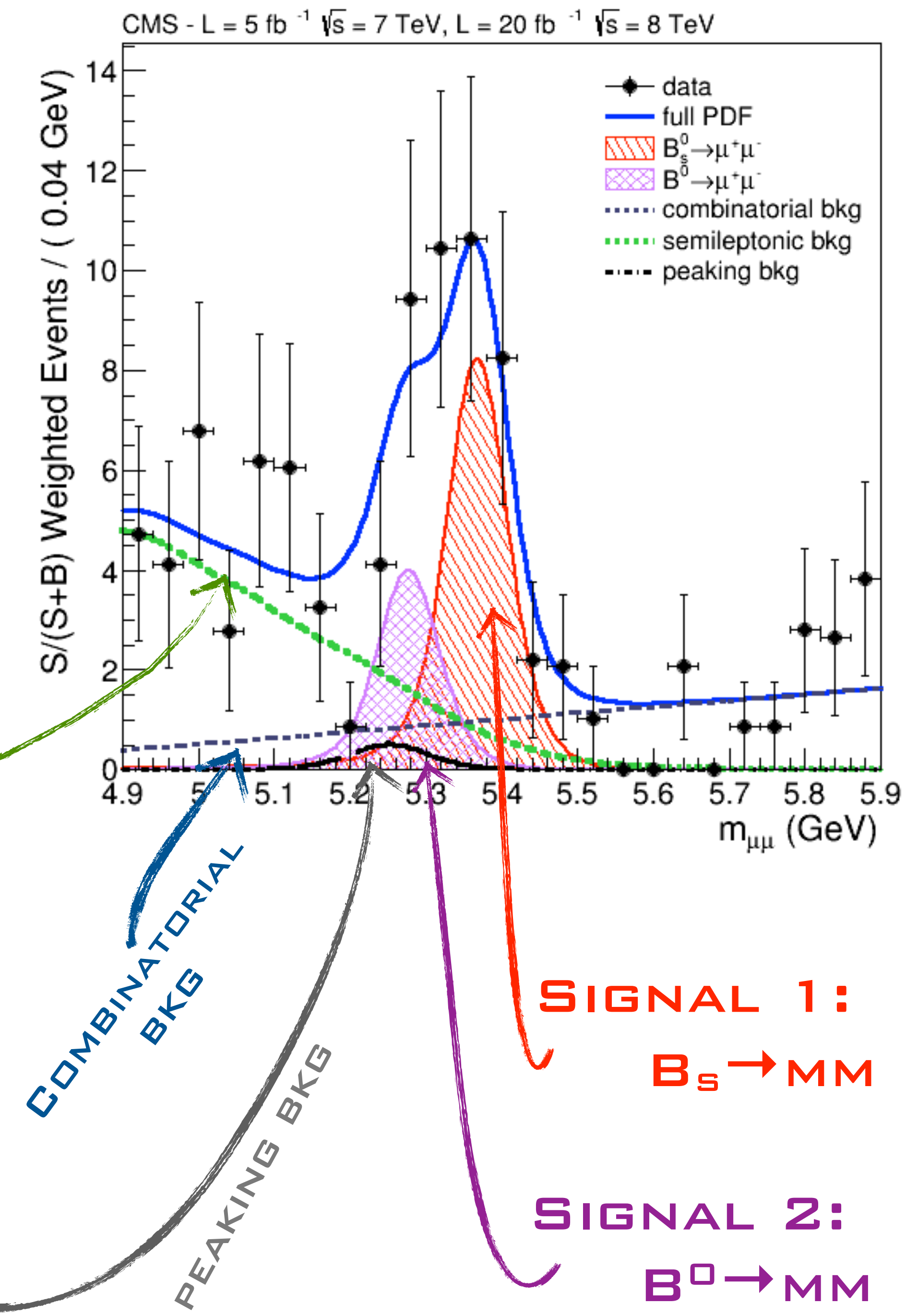


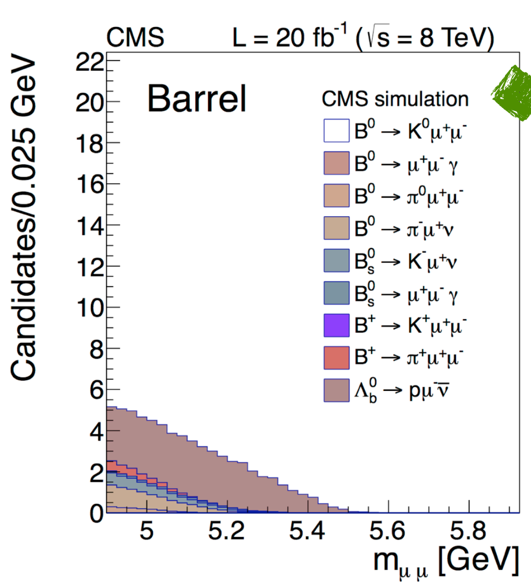


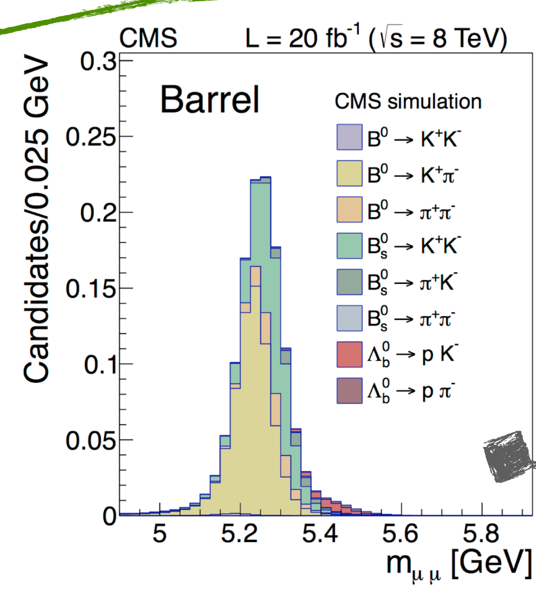
searching for an ***ultra-rare*** *decay:* **B→μμ**

1. ONLINE SELECTION (TRIGGER)
2. BLIND THE DATA (AVOID BIAS)
3. MULTIVARIATE SELECTION
4. **FIT THE DATA (LIKELIHOOD)**

Fit the data accounting for the various signal and background components

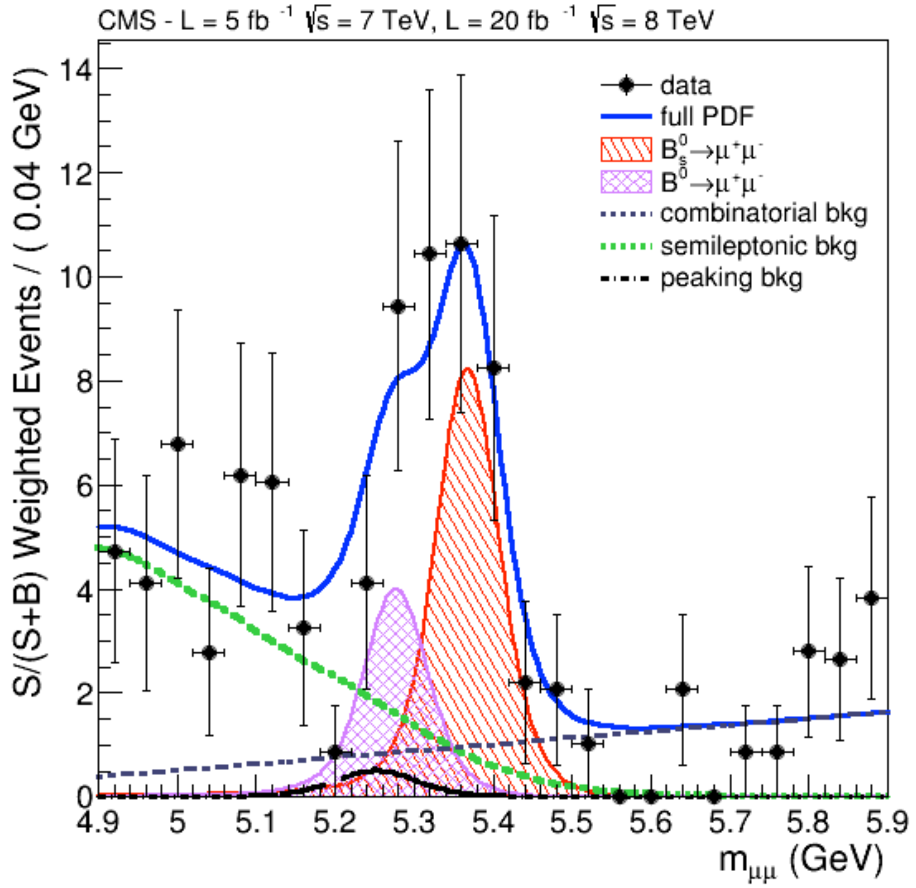




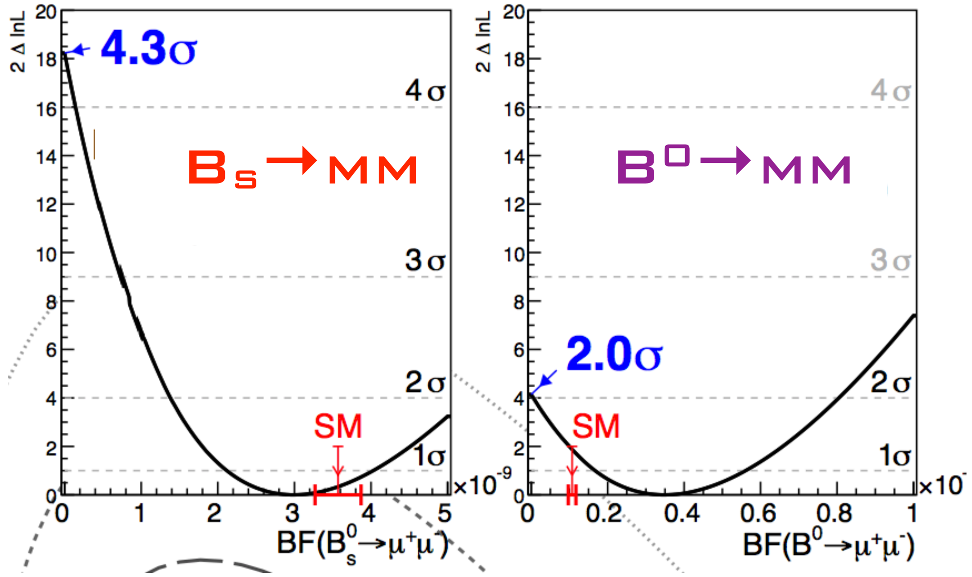


searching for an *ultra-rare* decay: B→μμ

1. ONLINE SELECTION (TRIGGER)
2. BLIND THE DATA (AVOID BIAS)
3. MULTIVARIATE SELECTION
4. FIT THE DATA (LIKELIHOOD)
5. STATISTICAL SIGNIFICANCE

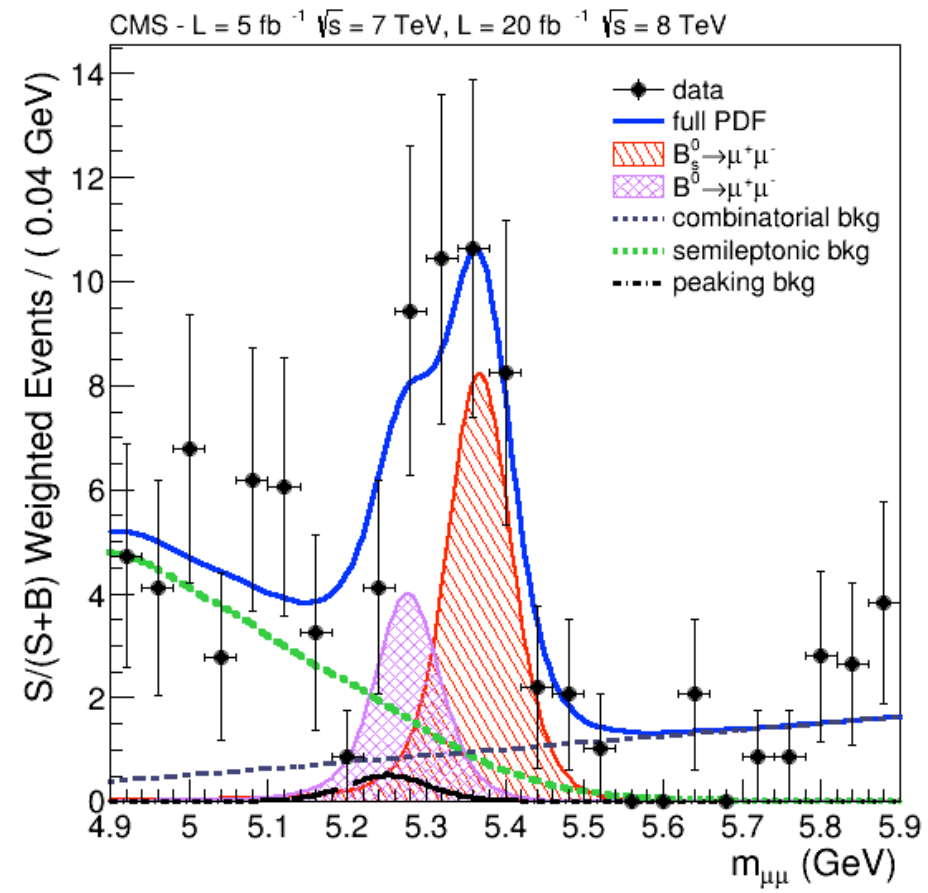


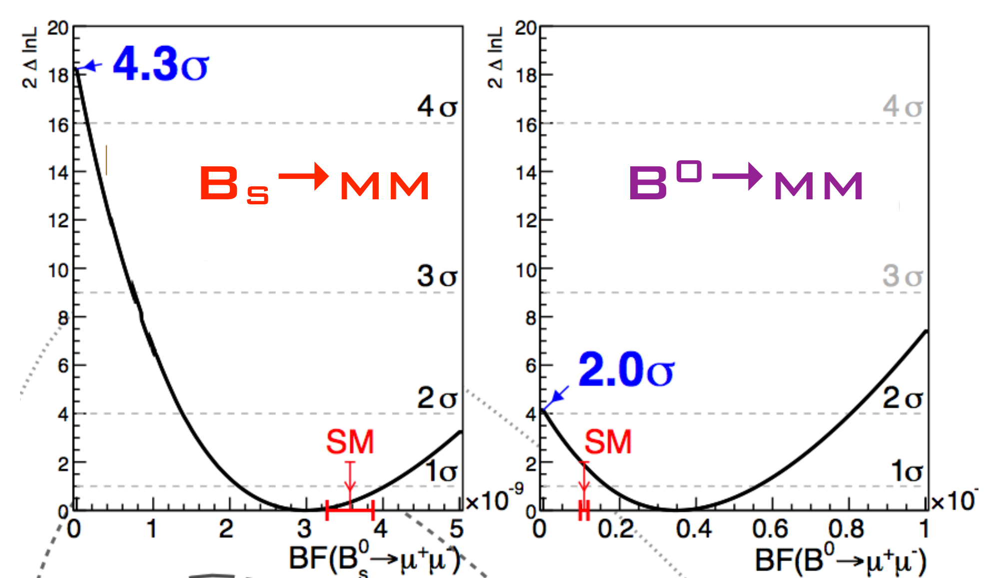
is the observed excess a genuine signal, or just a fluctuation of the background?



searching for an ultra-rare decay: B→μμ

1. ONLINE SELECTION (TRIGGER)
2. BLIND THE DATA (AVOID BIAS)
3. MULTIVARIATE SELECTION
4. FIT THE DATA (LIKELIHOOD)
5. STATISTICAL SIGNIFICANCE
6. **EXTRACT MEASUREMENT**

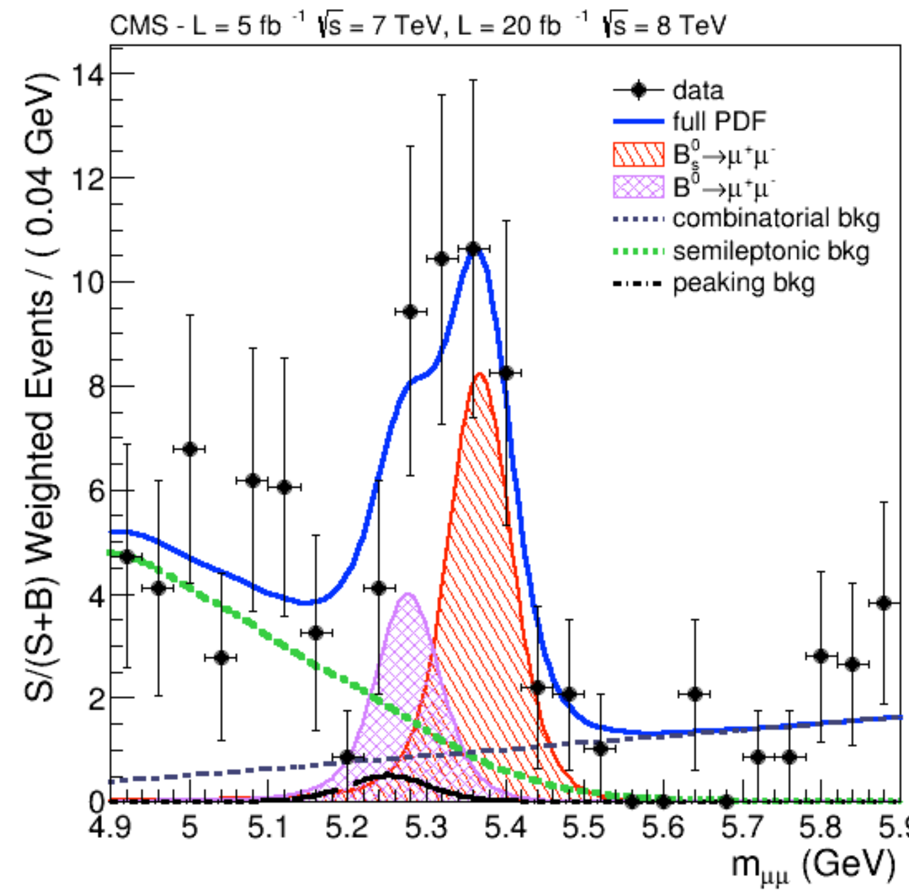




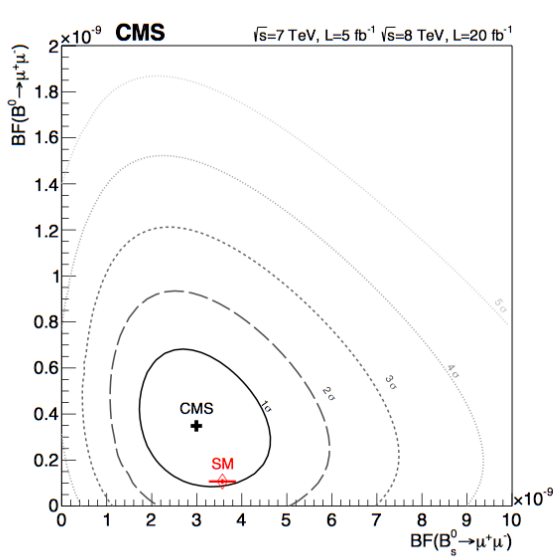
$$\mathrm{BR}(B_s \to \mu\mu) = \left(3.0^{+0.9}_{-0.8}\ (\mathrm{stat})^{+0.6}_{-0.4}\ (\mathrm{syst})\right) \times 10^{-9}$$

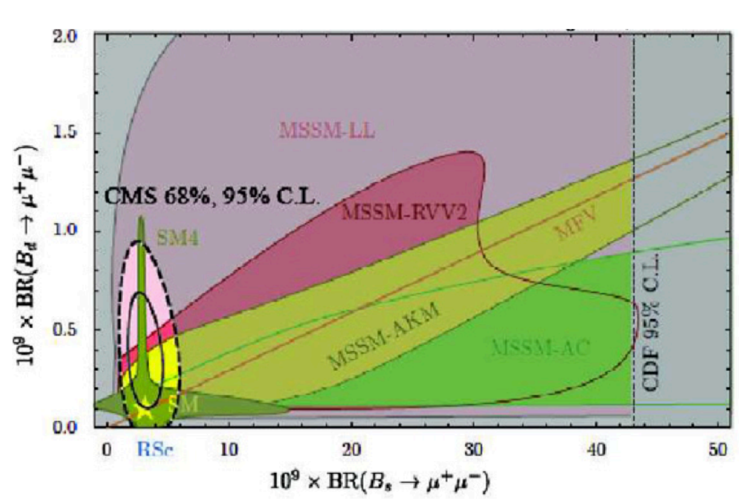
searching for an ultra-rare decay: B→μμ

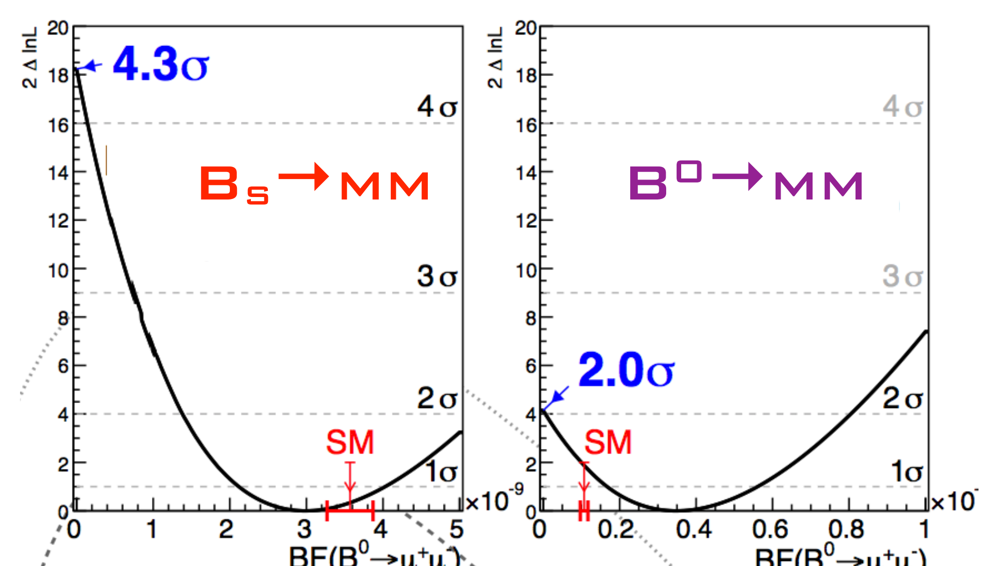
1. ONLINE SELECTION (TRIGGER)
2. BLIND THE DATA (AVOID BIAS)
3. MULTIVARIATE SELECTION
4. FIT THE DATA (LIKELIHOOD)
5. STATISTICAL SIGNIFICANCE
6. EXTRACT MEASUREMENT
7. **COMPARE TO THEORY**



THE SM... AND BEYOND





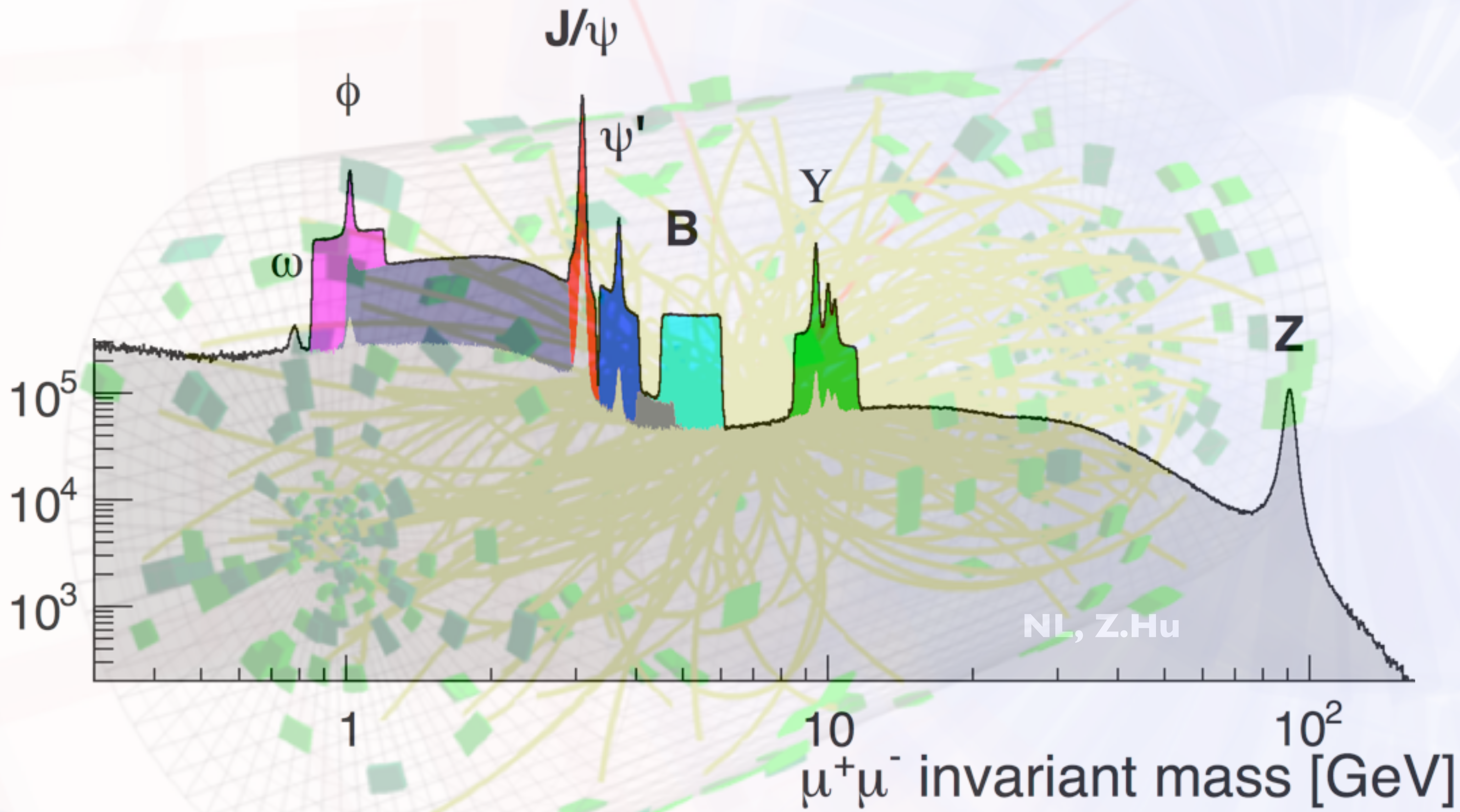


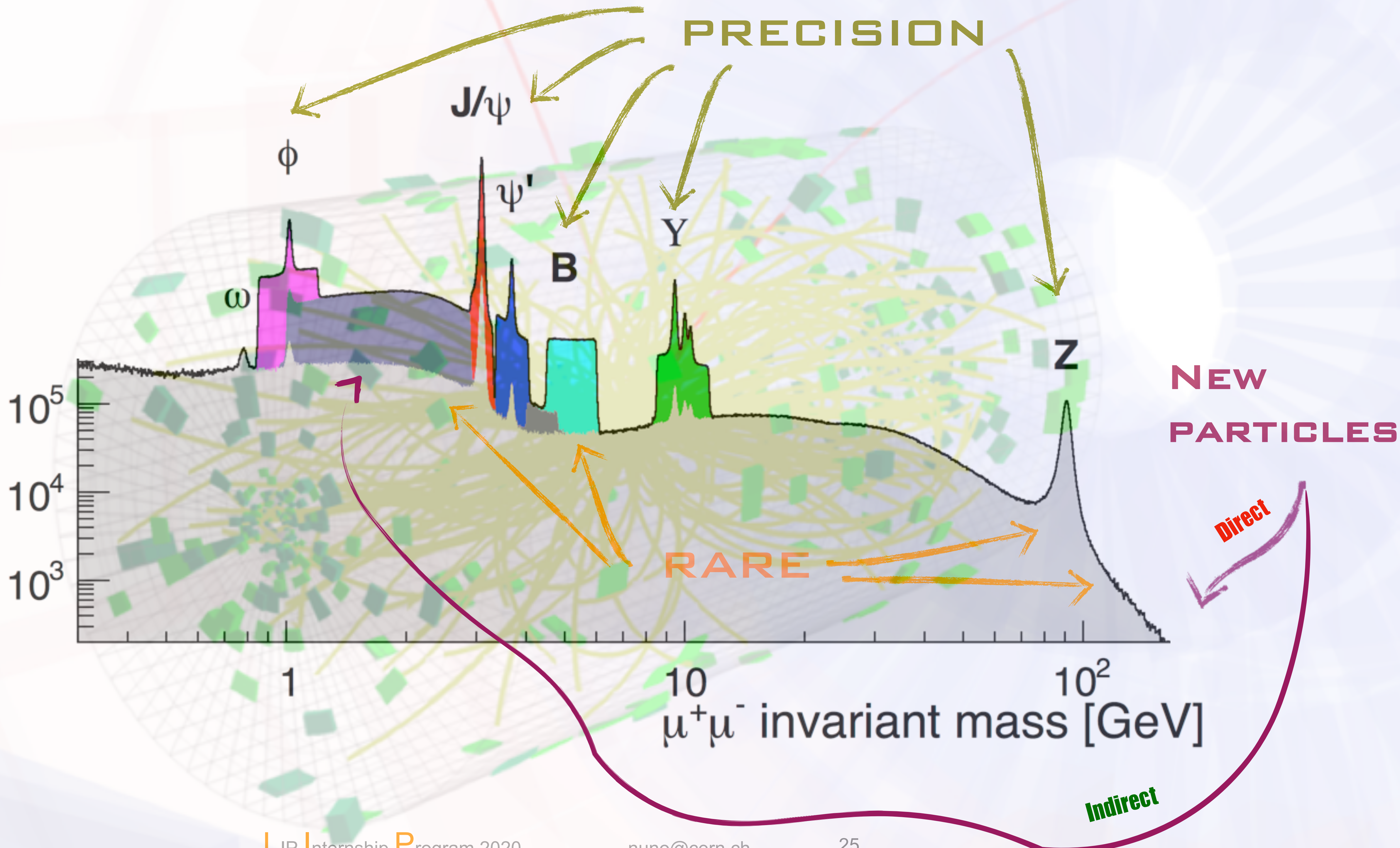
$$BR(B_S \to \mu\mu) = \left(3.0^{+0.9}_{-0.8}\ (stat)^{+0.6}_{-0.4}\ (syst)\right) \times 10^{-9}$$

the di-muon analysis

the di-muon spectrum ($X \rightarrow \mu\mu$)

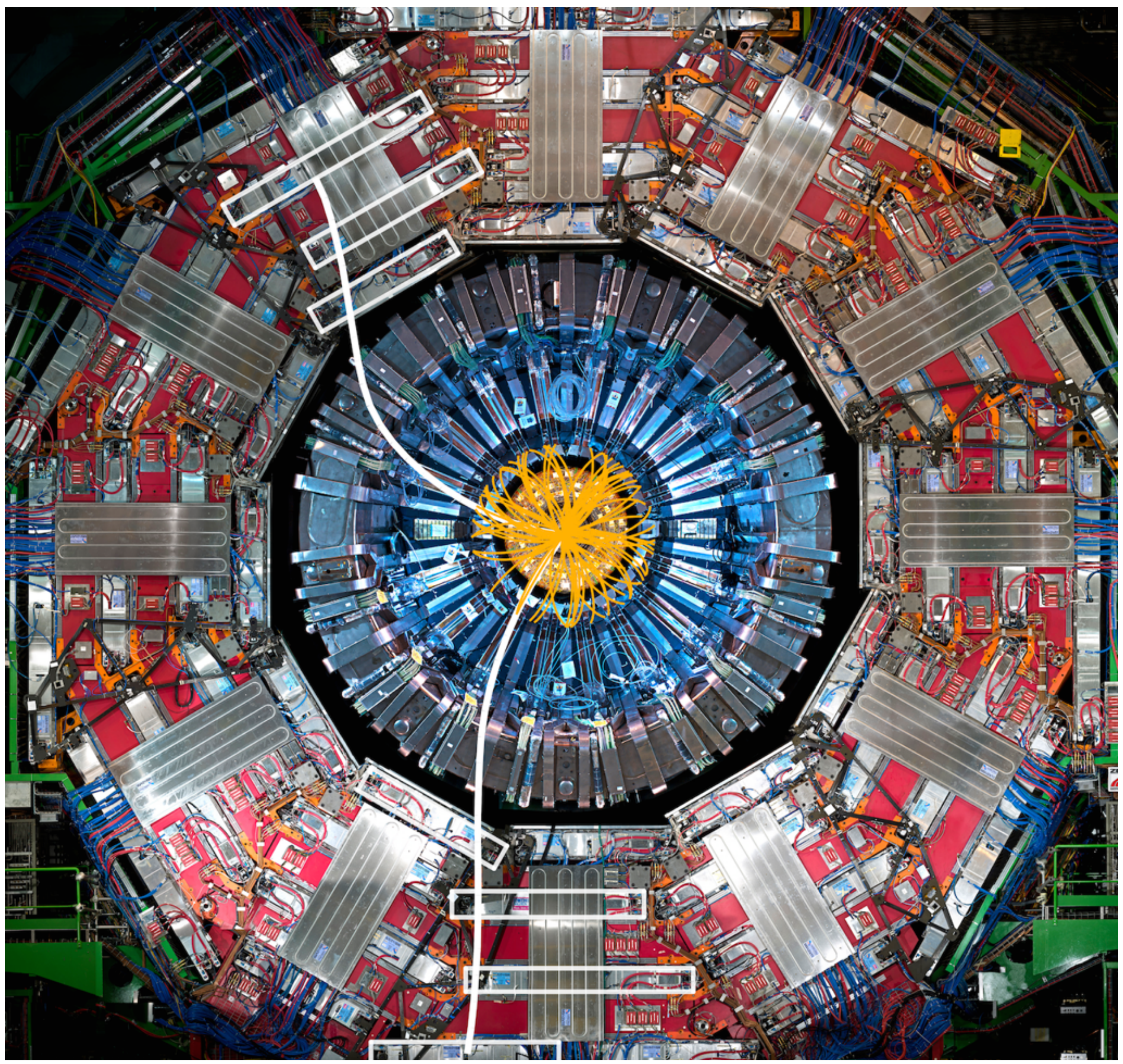
50 years of particle physics in one plot!

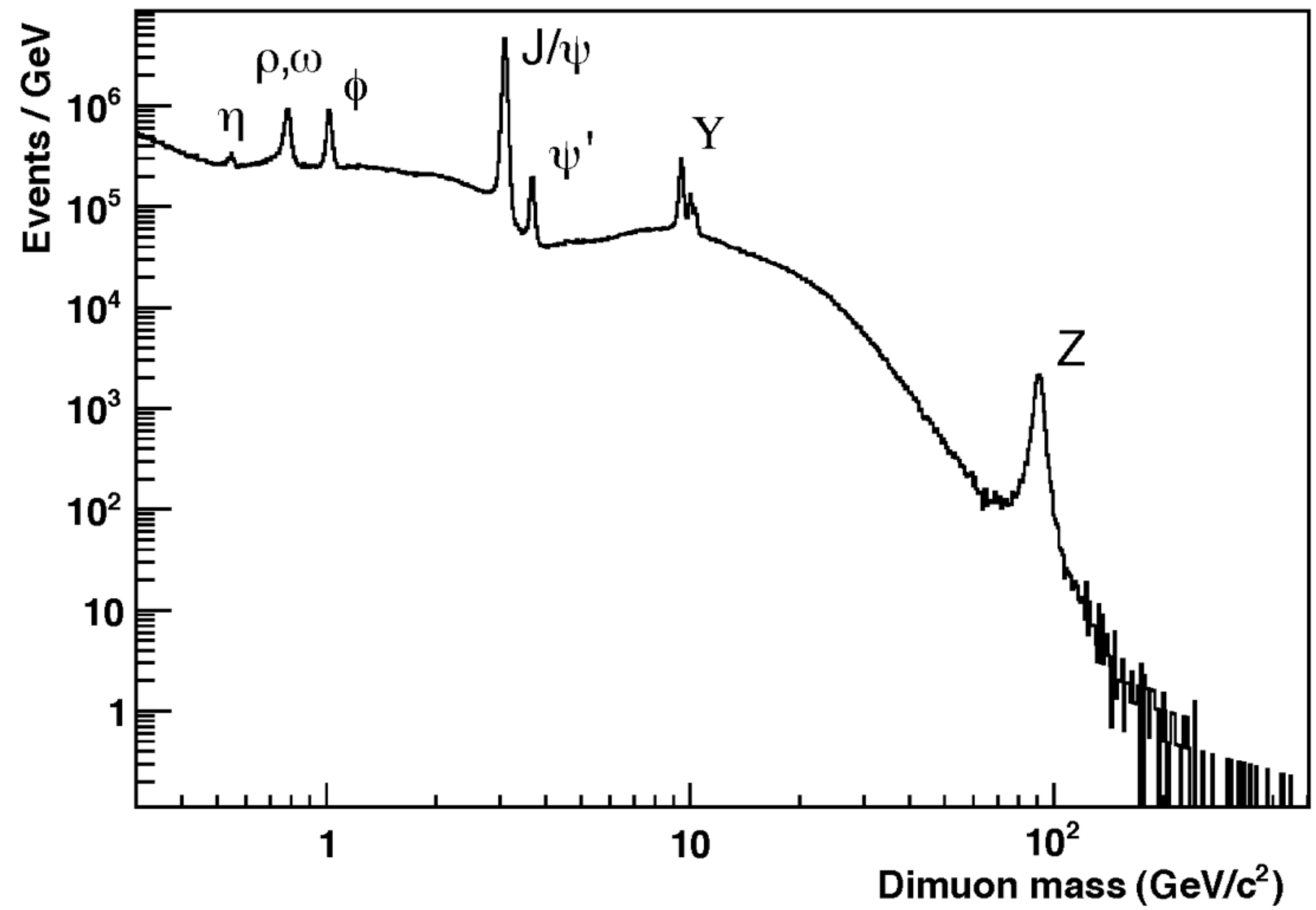



PRECISION
J/ψ
φ
ψ'
B
Υ
ω
Z
10^5
10^4
10^3
1
10
10^2
$\mu^+\mu^-$ invariant mass [GeV]
RARE
New particles
Direct
Indirect

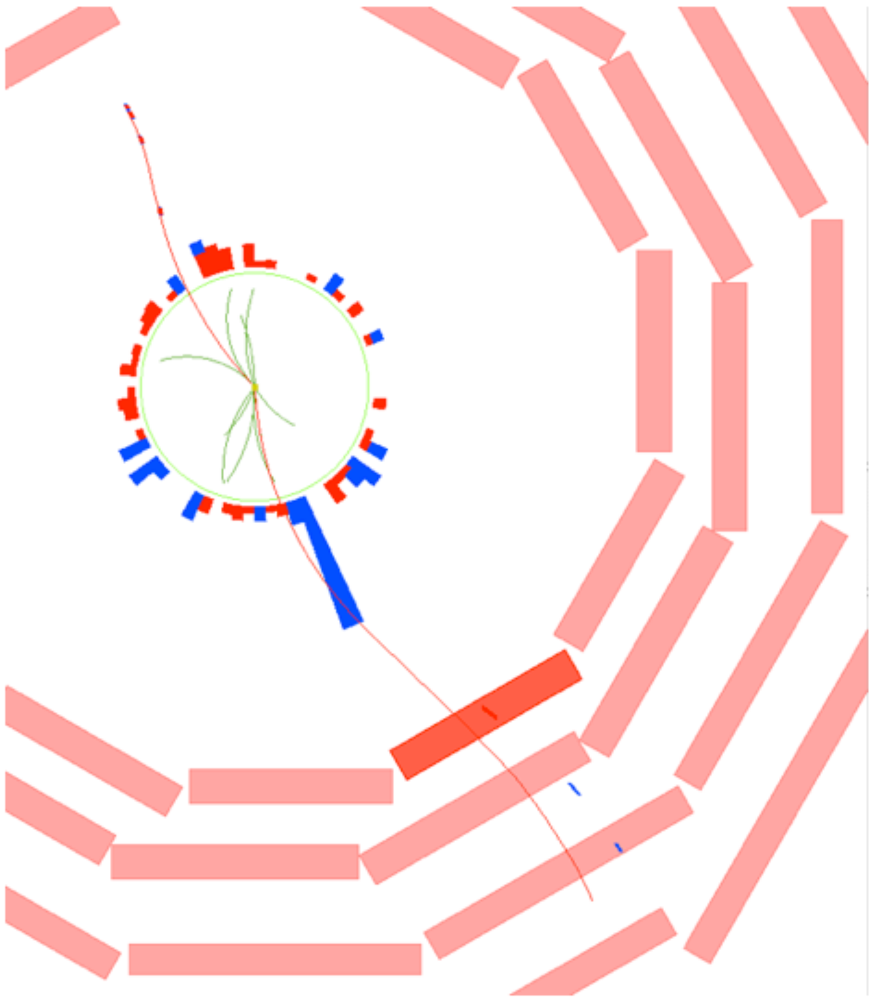
from detector to physics ...

?





di-muon ‘invariant mass’ ?

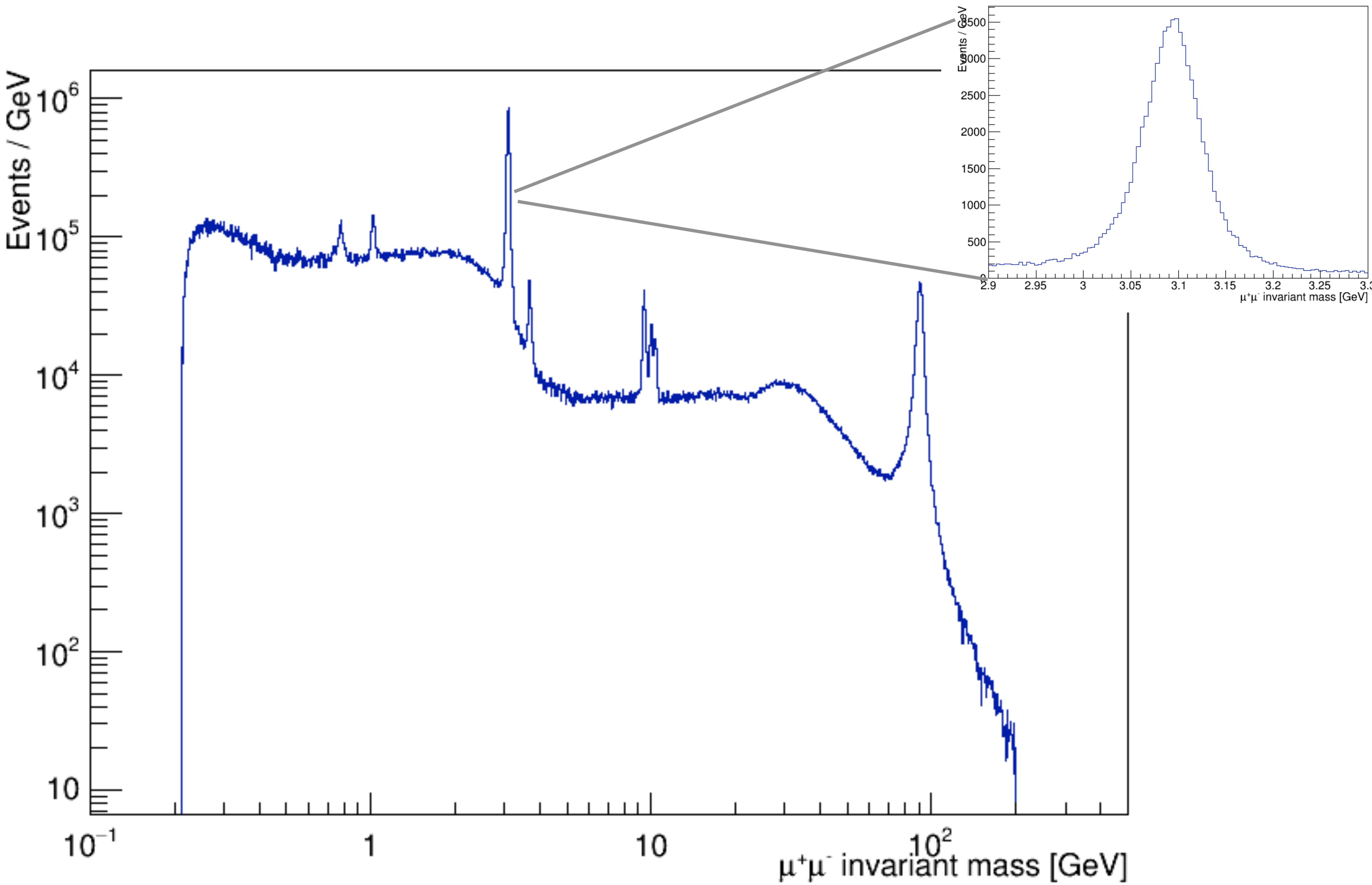


particle identification
- signal in muon chambers
→ it’s a muon!
➠ $m = m(\mu) \sim 106 MeV/c^2$

particle trajectory
- muon chambers but especially the silicon tracker
➠ linear momentum, $\underline{p} \equiv (p_x, p_y, p_z)$

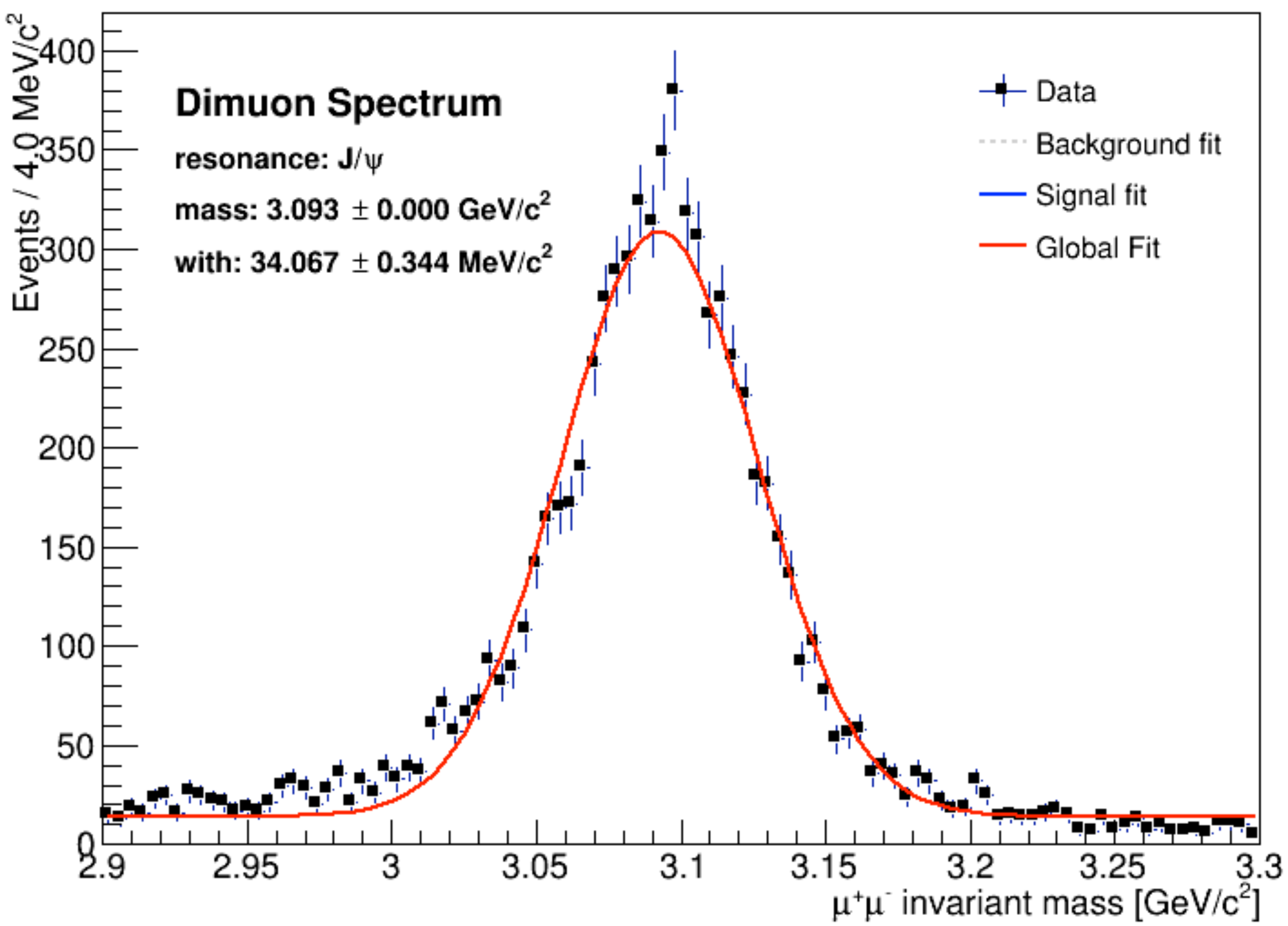
➠ form 4-momentum of each muon: $\mathbf{P}_\mu \equiv (E, p_x, p_y, p_z)$
➠ that of the di-muon pair $\mathbf{P}_{\mu\mu} = \mathbf{P}_{\mu 1} + \mathbf{P}_{\mu 2} = \mathbf{P}_{\mathbf{x} \to \mu\mu}$
➠ invariant mass $\mathbf{P}_{\mu\mu} \cdot \mathbf{P}_{\mu\mu} = \mathbf{M}_{\mu\mu}^2 = (\mathbf{M_x})^2$

the reconstructed di-muon spectrum



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

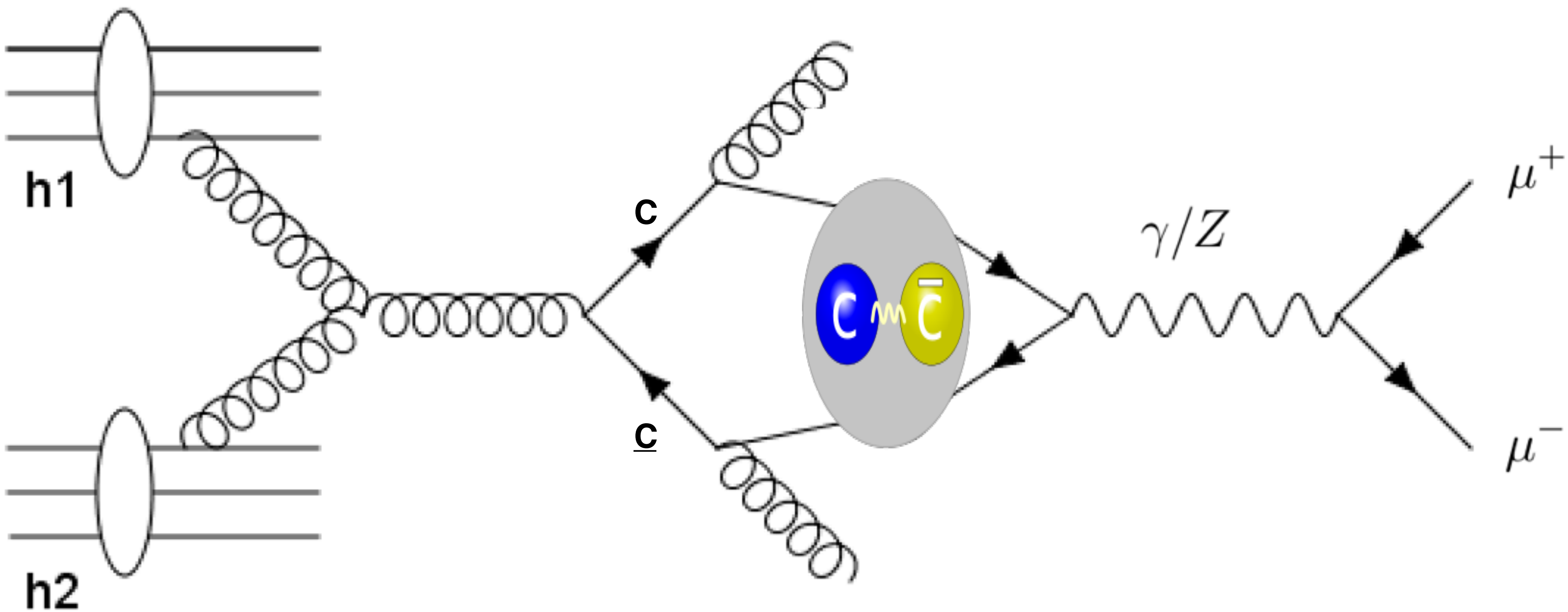
fit the data



- inspect quality of fit
  - can model be improved?
  - hint: final state radiation ($\mu \rightarrow \mu\gamma$) may distort shape

- 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**
  - $N \pm \sigma_{stat} \pm \sigma_{syst}$

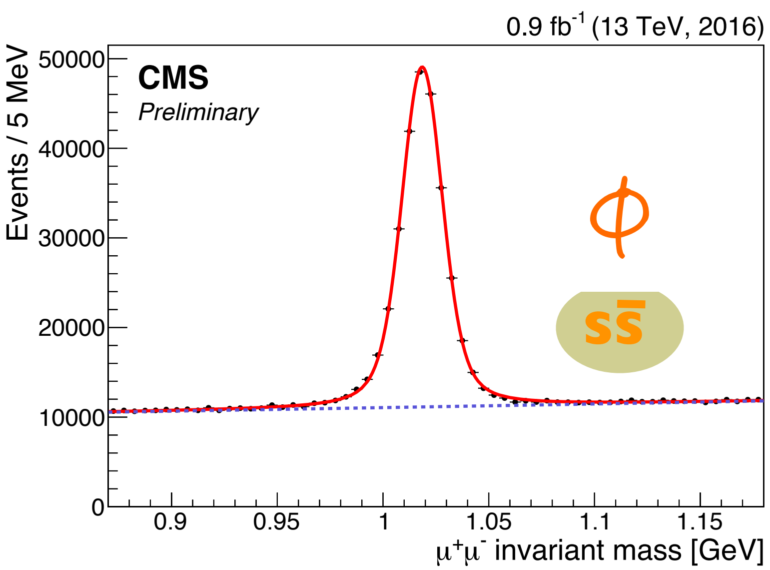
what's the physics process ?

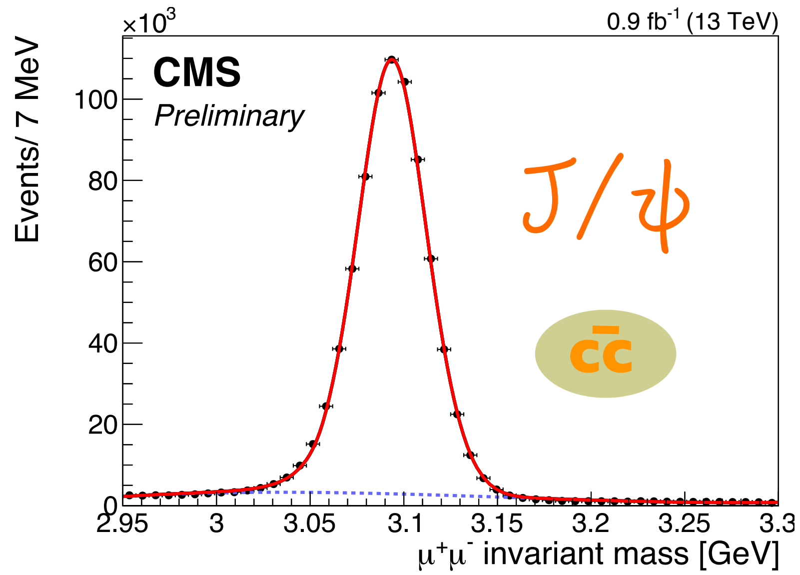


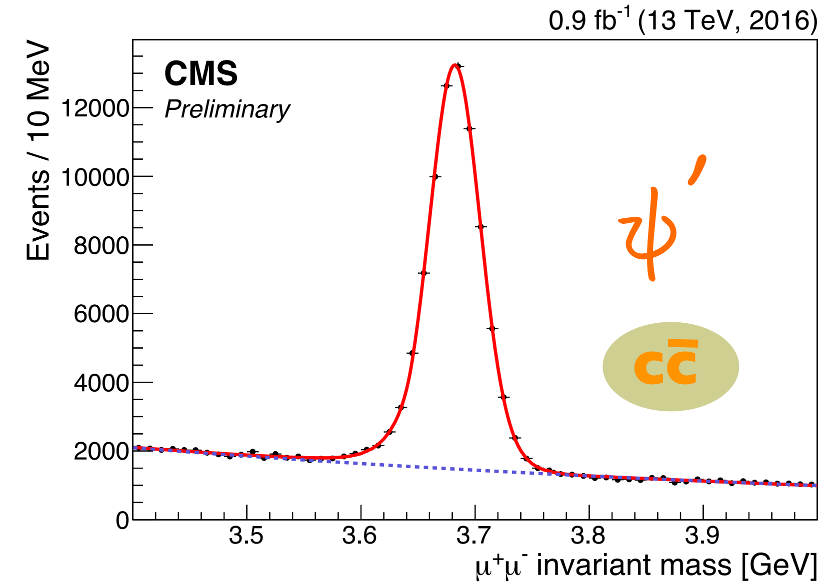
production: strong force

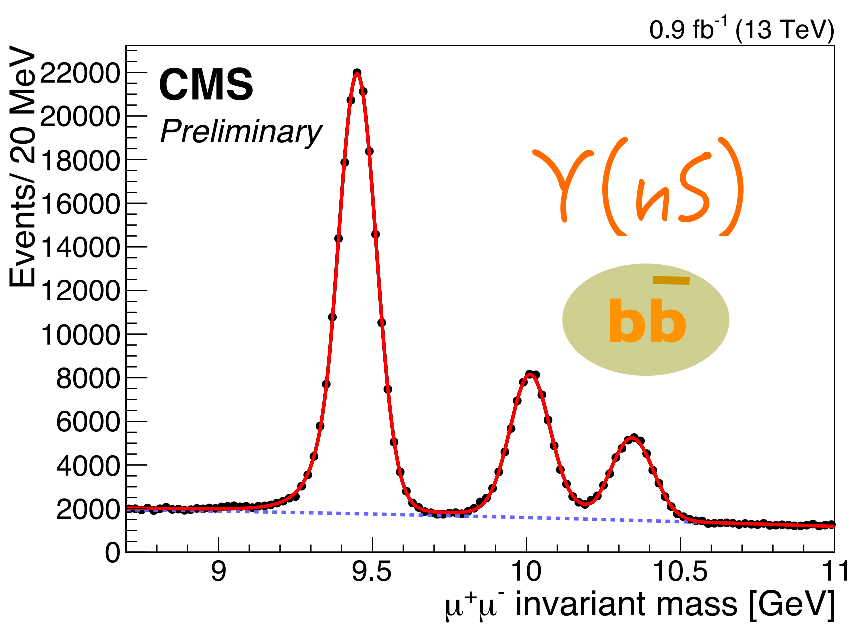
decay: electroweak force

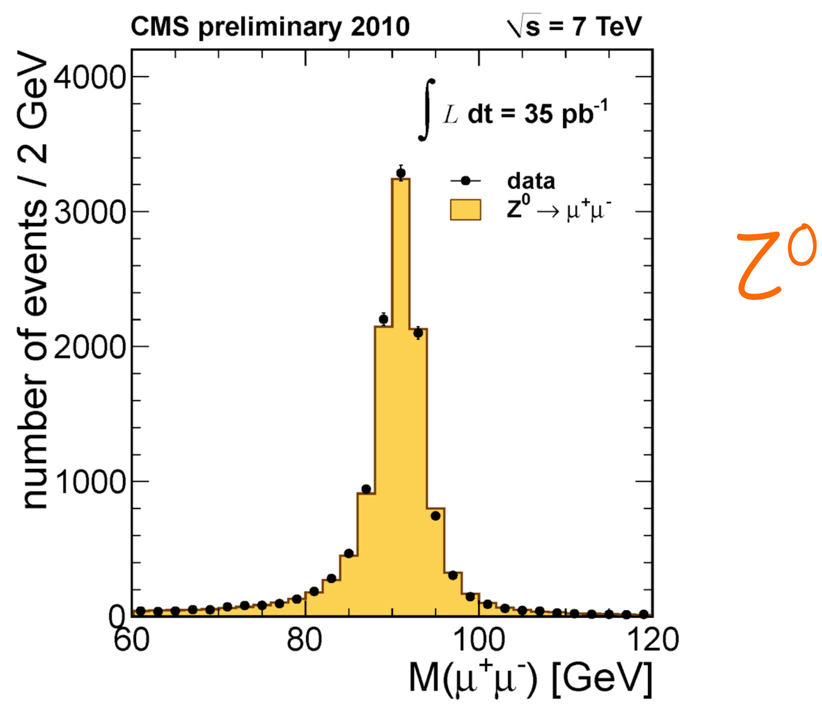
what are the peaks?











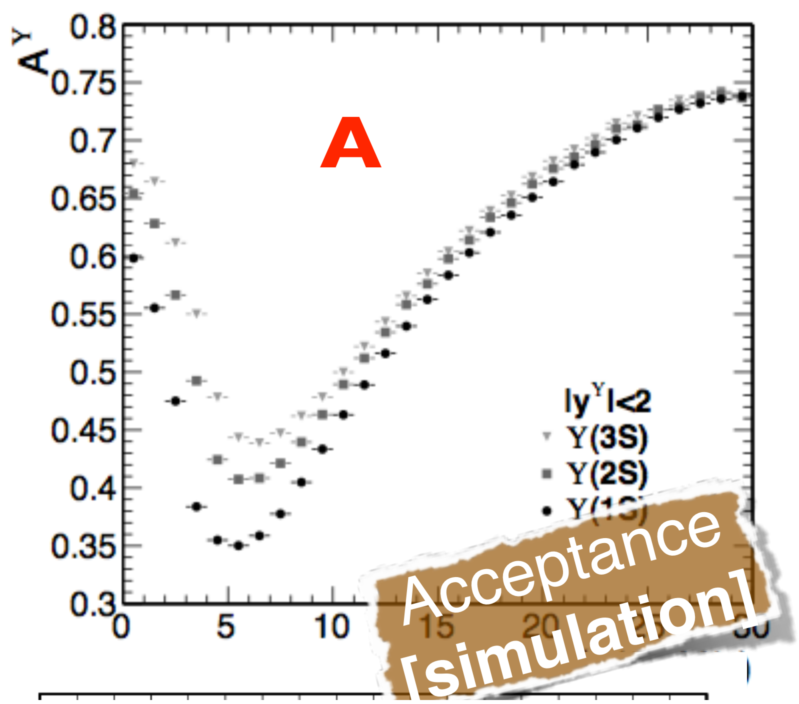
Check their measured properties from: http://pdglive.lbl.gov

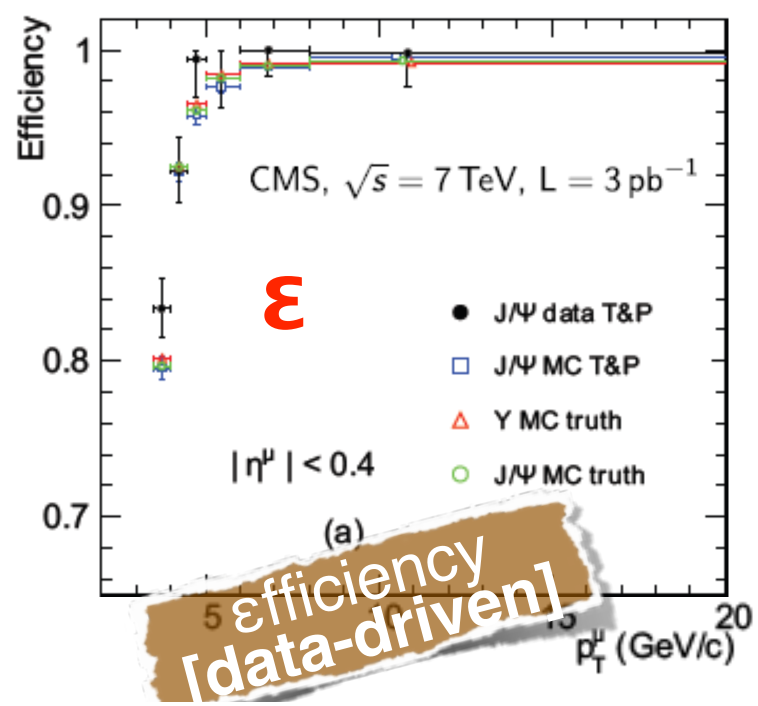
production cross section

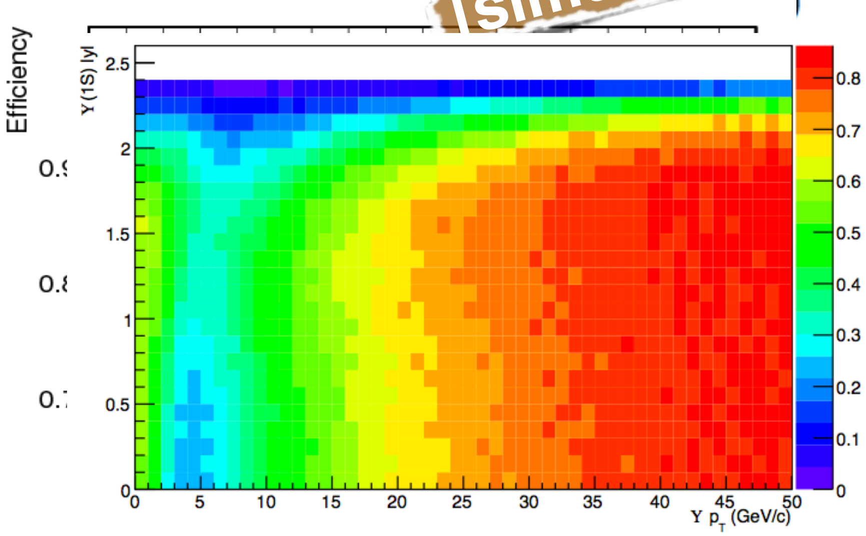
"N=L.σ"

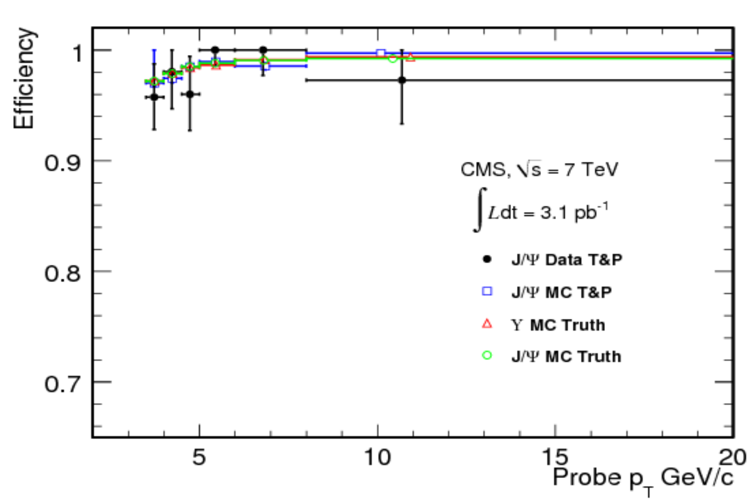
$$\frac{d^2\sigma(Q\overline{Q})}{dp_T\,dy}\mathcal{B}\left(Q\overline{Q}\to\mu^+\mu^-\right)=\frac{N_{fit}(Q\overline{Q})}{\mathcal{L}\cdot\mathcal{A}\cdot\epsilon\cdot\Delta p_T\cdot\Delta y}$$

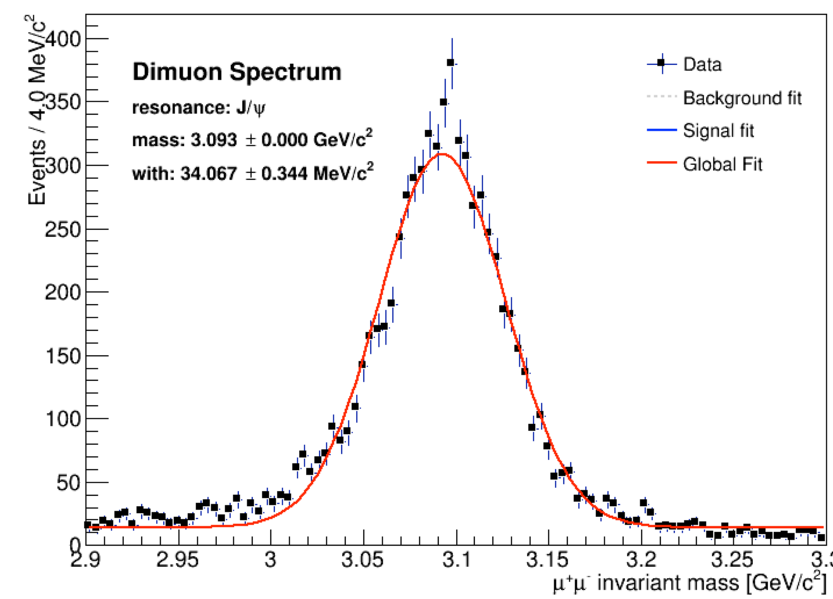
an effective area of interaction
unit: barn, 1b = 10^{-28} m^2 = 100fm^2











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

towards the exercise

setting up

- get the tutorial materials

  ...

- start root

```
root -l
```

```
root []
```

- check, load

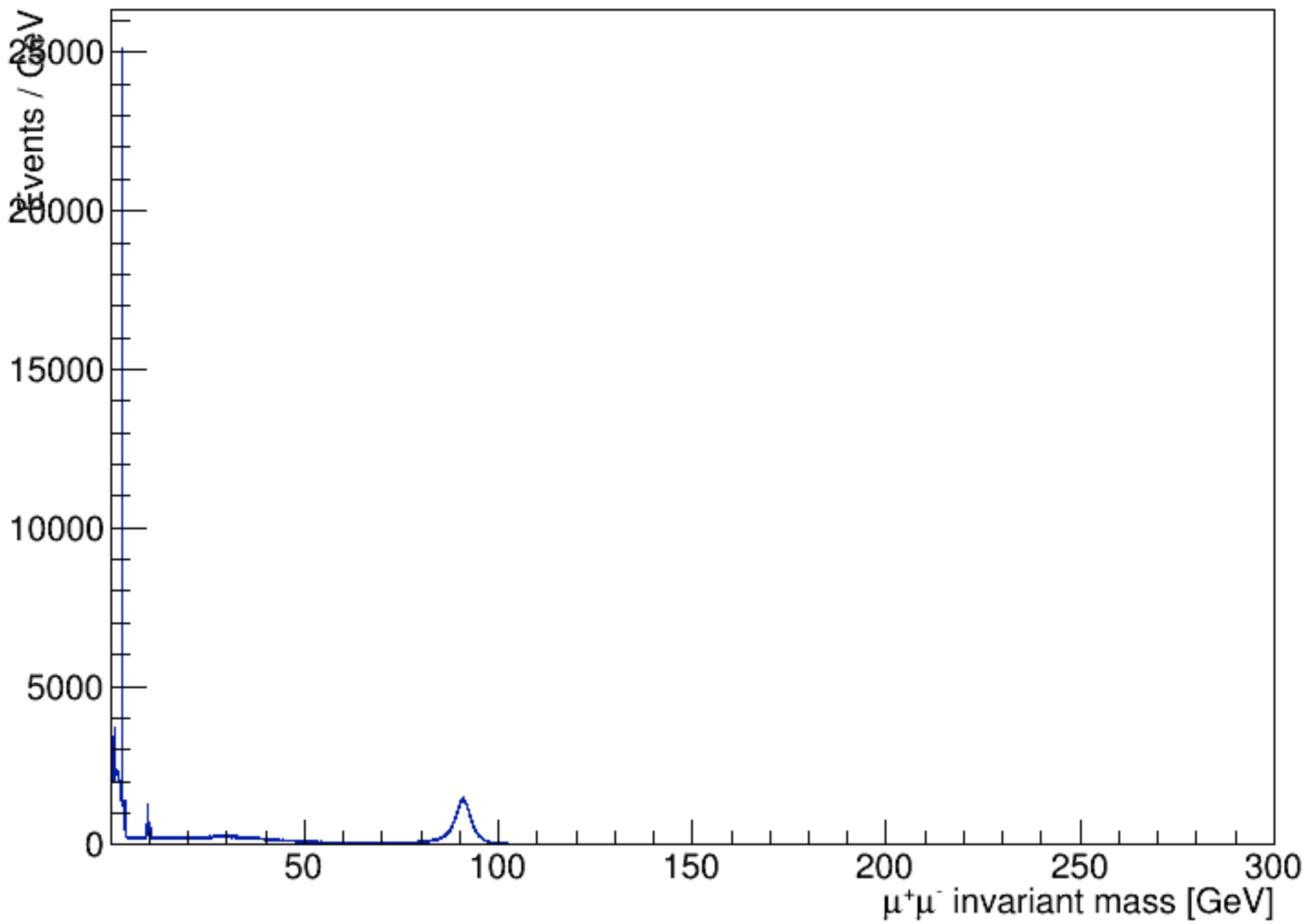
```
root [4] .!pwd
/Users/nuno/datatutorial
root [5] .!ls
Skim4.root dimuon.h  dimuons.C

root [6] .!mkdir plots
root [7] .!ls
Skim4.root dimuon.h   dimuons.C  plots
```

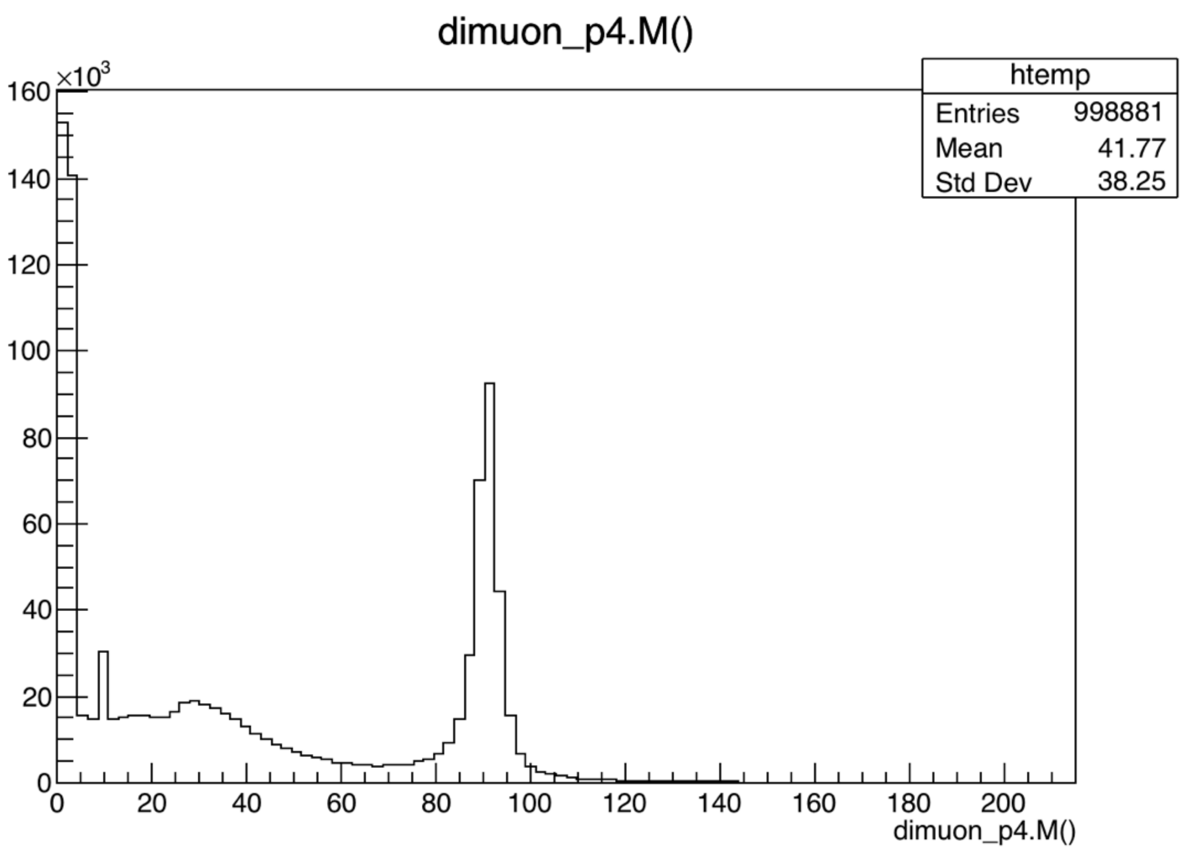
the code

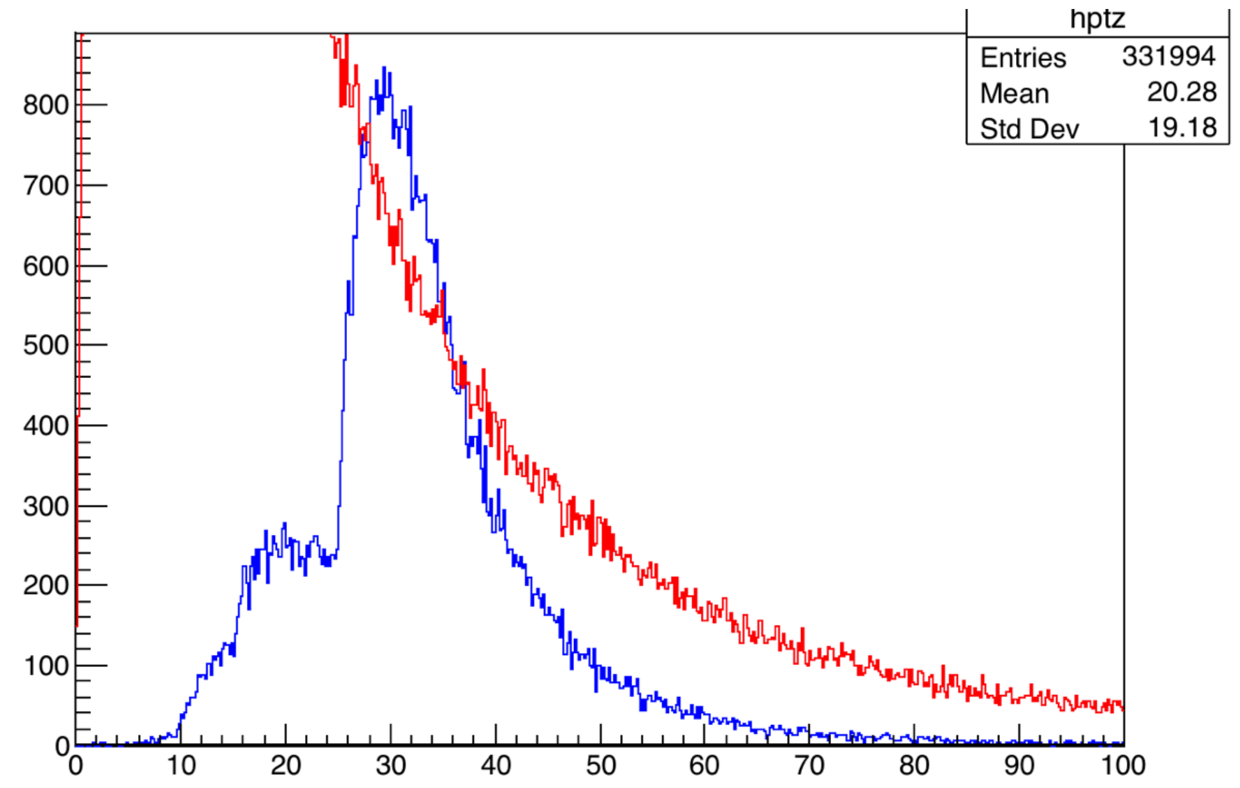
- main methods
  - GetSpectrum(): create the dimuon spectrum from the raw dataset
  - Cut(): allows to place selection cuts
  - SelectPeak(): allows to select one of the signals in the spectrum
  - FitPeak(): fits the data and extracts signal parameters

```
emacs dimuons.C &

root -l -b -q dimuons.C++

ls plots
```

the ‘raw’ spectrum




dimuon_p4.M()
htemp
Entries 998881
Mean 41.77
Std Dev 38.25
×10^3
160
140
120
100
80
60
40
20
0
0 20 40 60 80 100 120 140 160 180 200
dimuon_p4.M()


hptz
Entries 331994
Mean 20.28
Std Dev 19.18
800
700
600
500
400
300
200
100
0
0 10 20 30 40 50 60 70 80 90 100

Check the exercise instructions

Tutorial on data analysis and fitting with ROOT for the LIP Internship Program

https://github.com/aboletti/LIP-analysis-tutorial