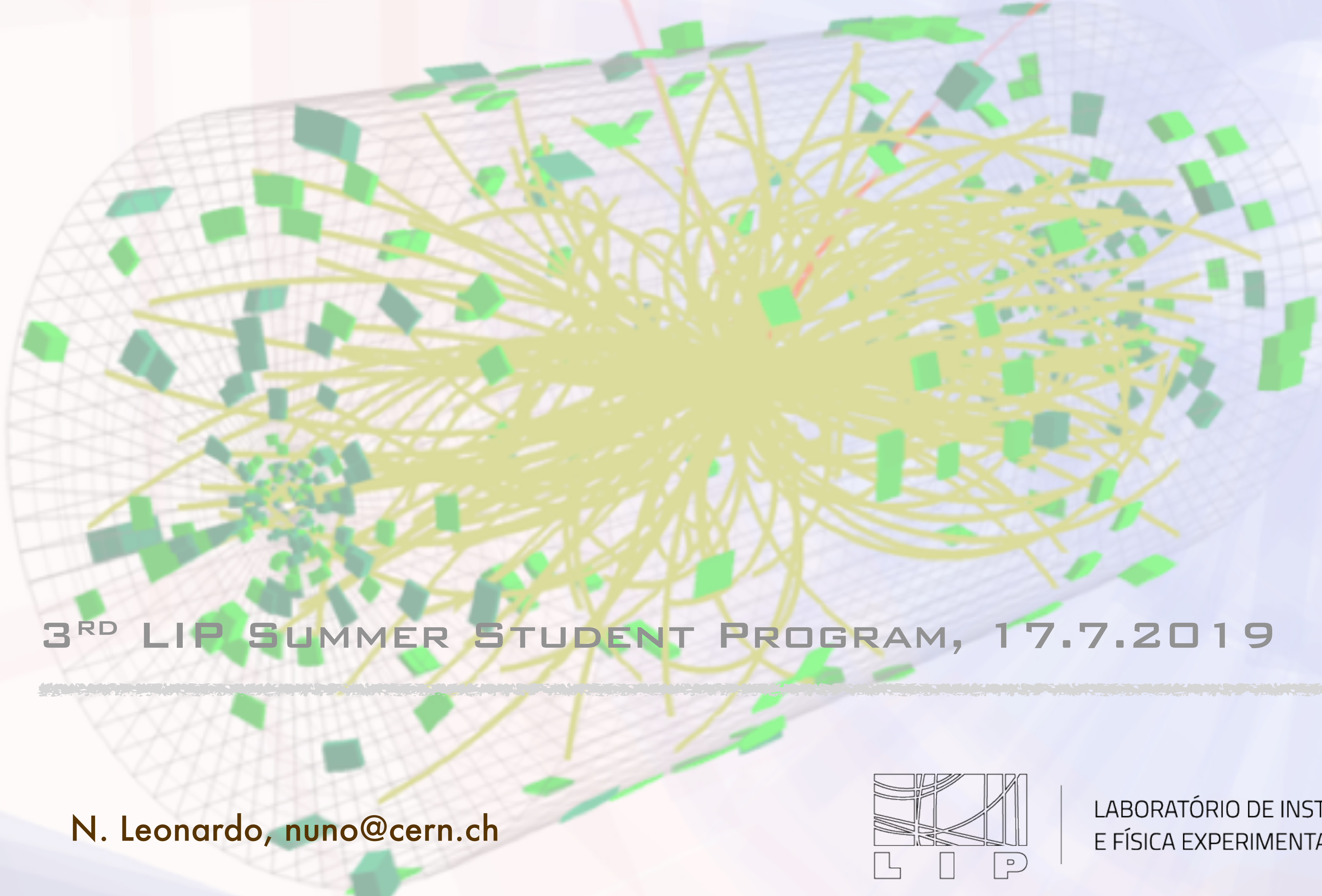
Data Analysis Tutorial



3RD LIP SUMMER STUDENT PROGRAM, 17.7.2019

N. Leonardo, nuno@cern.ch

LABORATÓRIO DE INSTRUMENTAÇÃO
E FÍSICA EXPERIMENTAL DE PARTÍCULAS

intro

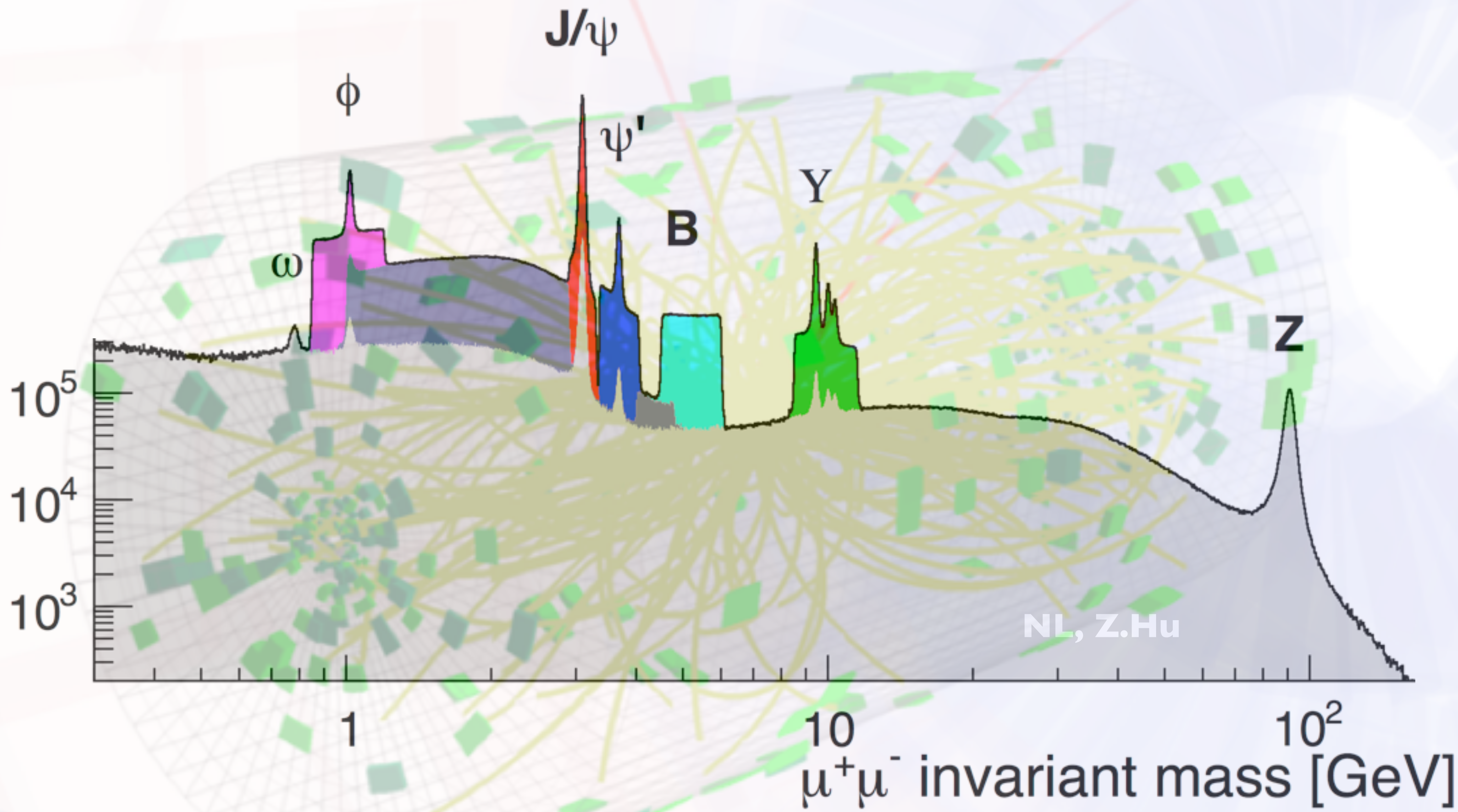
- this week you've seen already many nice physics results in the form of plenty of colourful looking plots
- this morning you've seen how particles are detected and their trajectories reconstructed
- on the last two afternoons you've had an introduction to HEP's baseline programming environment and tools, Linux, C++, ROOT
- just before you've heard about baseline concepts in statistics
- now we'll try to connect the dots by performing a data analysis

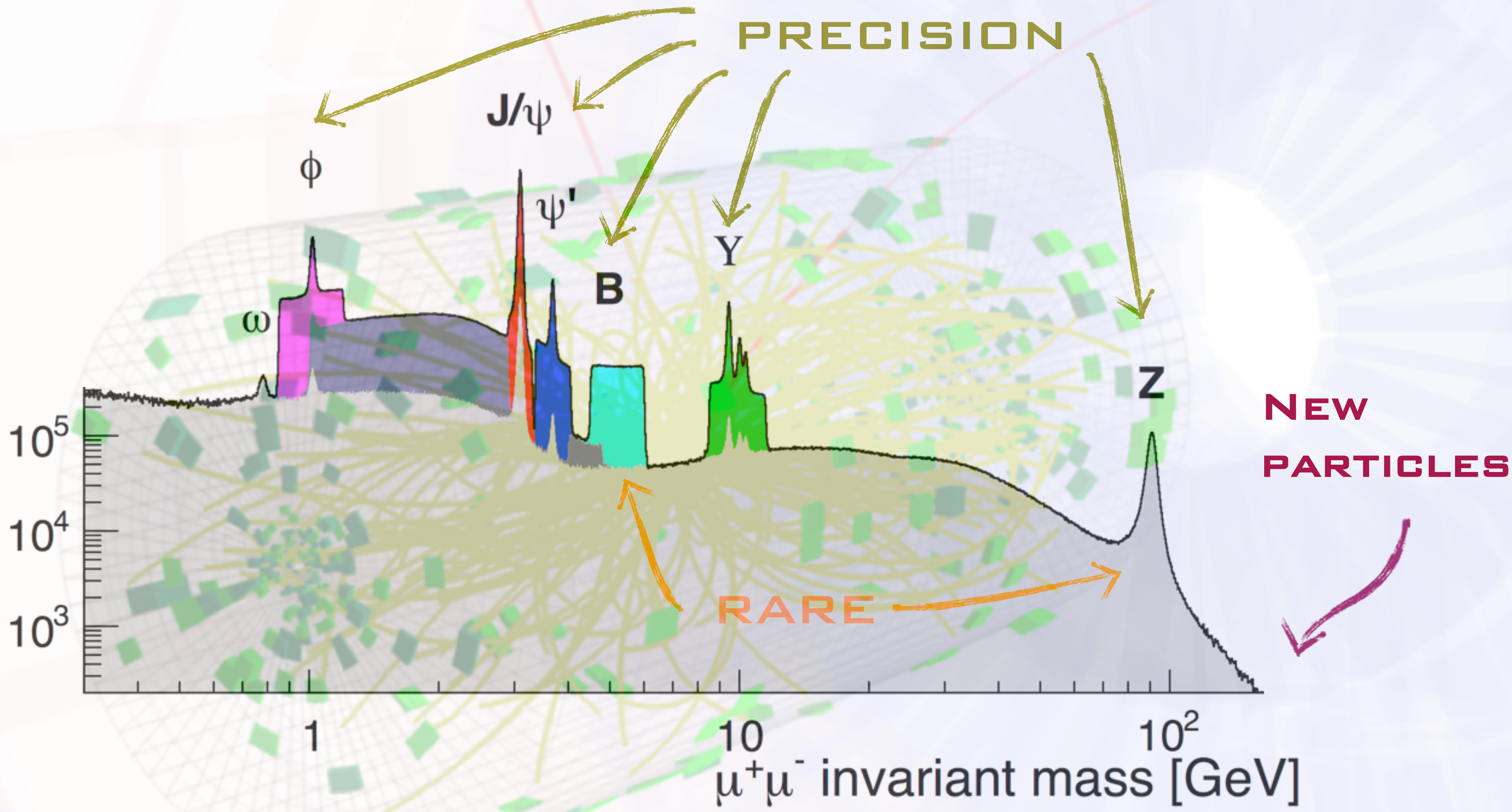
goals

- perform a simple data analysis
- manipulate data ntuples
- produce, process, and display data histograms
  - select different physics signals
  - plot kinematic distributions
  - selection criteria and efficiency
- extract physics parameters by performing a fit to the data
  - statistical errors
  - systematic errors

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

50 years of particle physics in one plot!

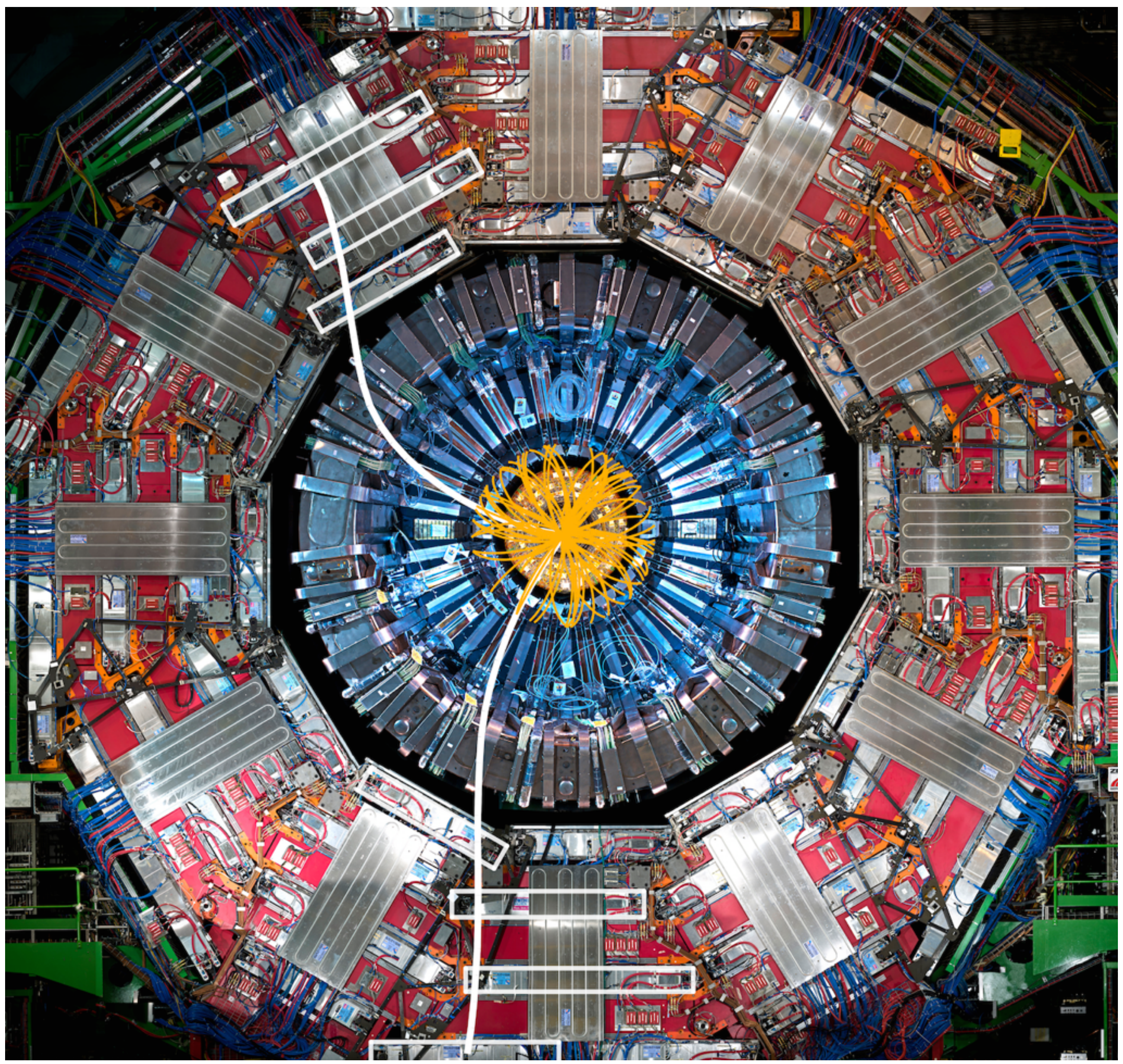


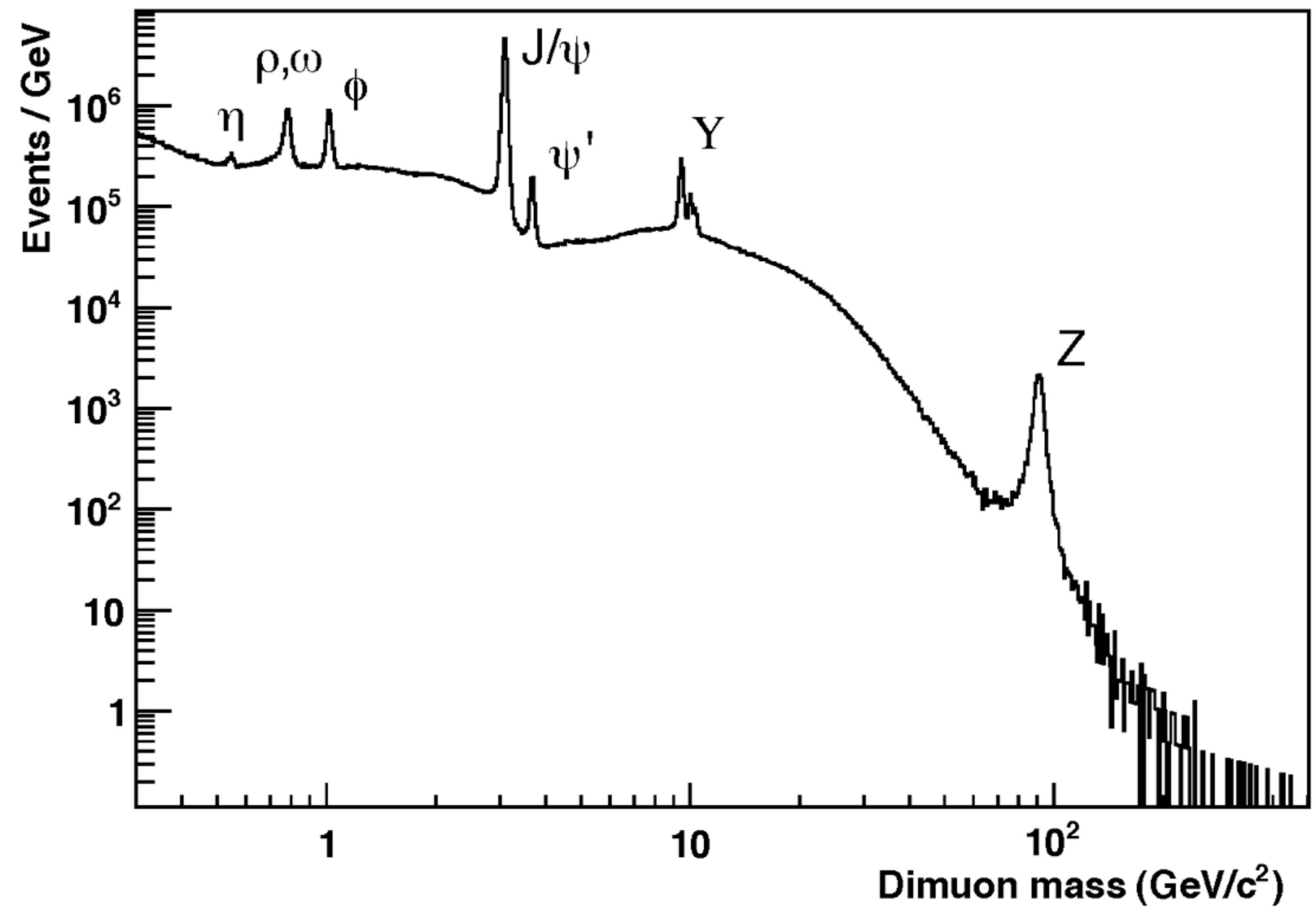


PRECISION
J/ψ
ϕ
ψ'
Υ
B
ω
Z
New
particles
RARE
10^5
10^4
10^3
1
10
10^2
μ+μ- invariant mass [GeV]

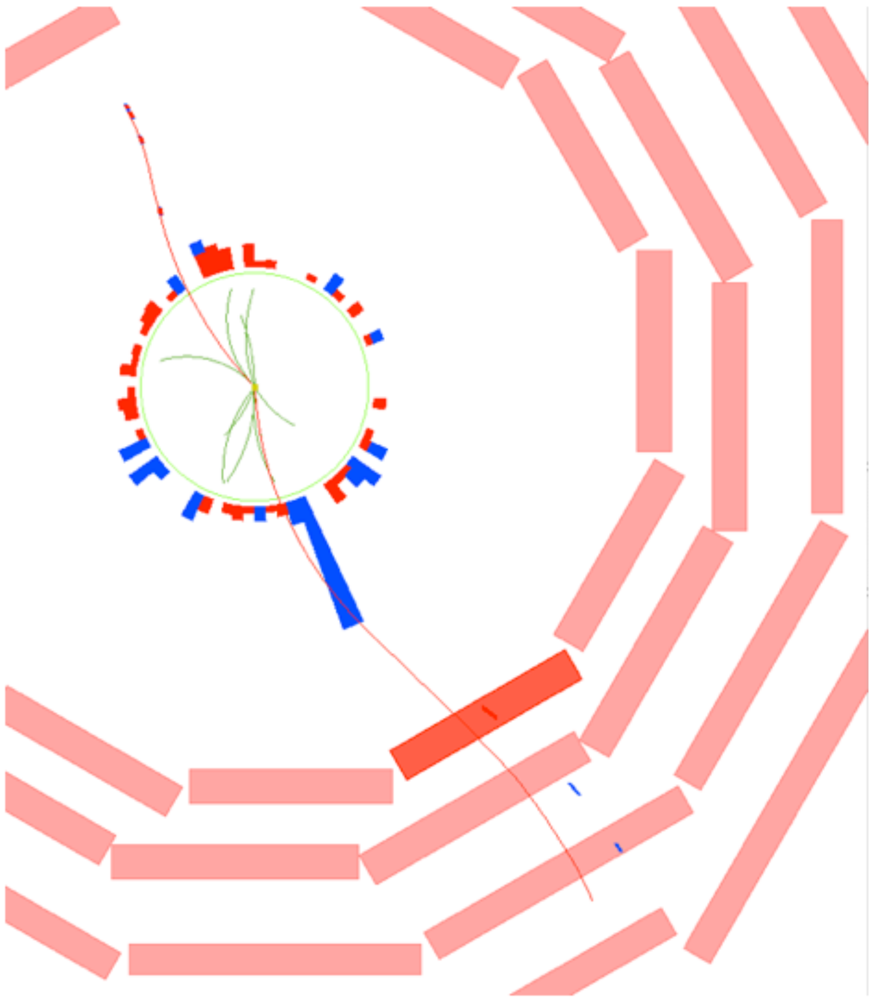
from detector to physics ...

?





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

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

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

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

setting up

- get the tutorial materials

```
wget http://cern.ch/nuno/datatutorial/tutorial.tgz
tar xvzf tutorial.tgz
cd  datatutorial
```

- start root

```
root -l

 root []
```

- check, load

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

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

inspecting the dataset

```
root [] TFile f("Skim4.root")
(TFile &) @0x1013ba520

root [] gDirectory→ls()
TFile**    Skim4.root
TFile*   Skim4.root
KEY: TTree oniaTree;5 Tree of Onia2MuMu
KEY: TTree oniaTree;4 Tree of Onia2MuMu

root [] oniaTree→Show()
  event              = 33412514
  dimuon_p4          = (TLorentzVector*)0x7fc205c82b10
  muonP_p4           = (TLorentzVector*)0x7fc205d41460
  muonN_p4           = (TLorentzVector*)0x7fc205d41b80
```

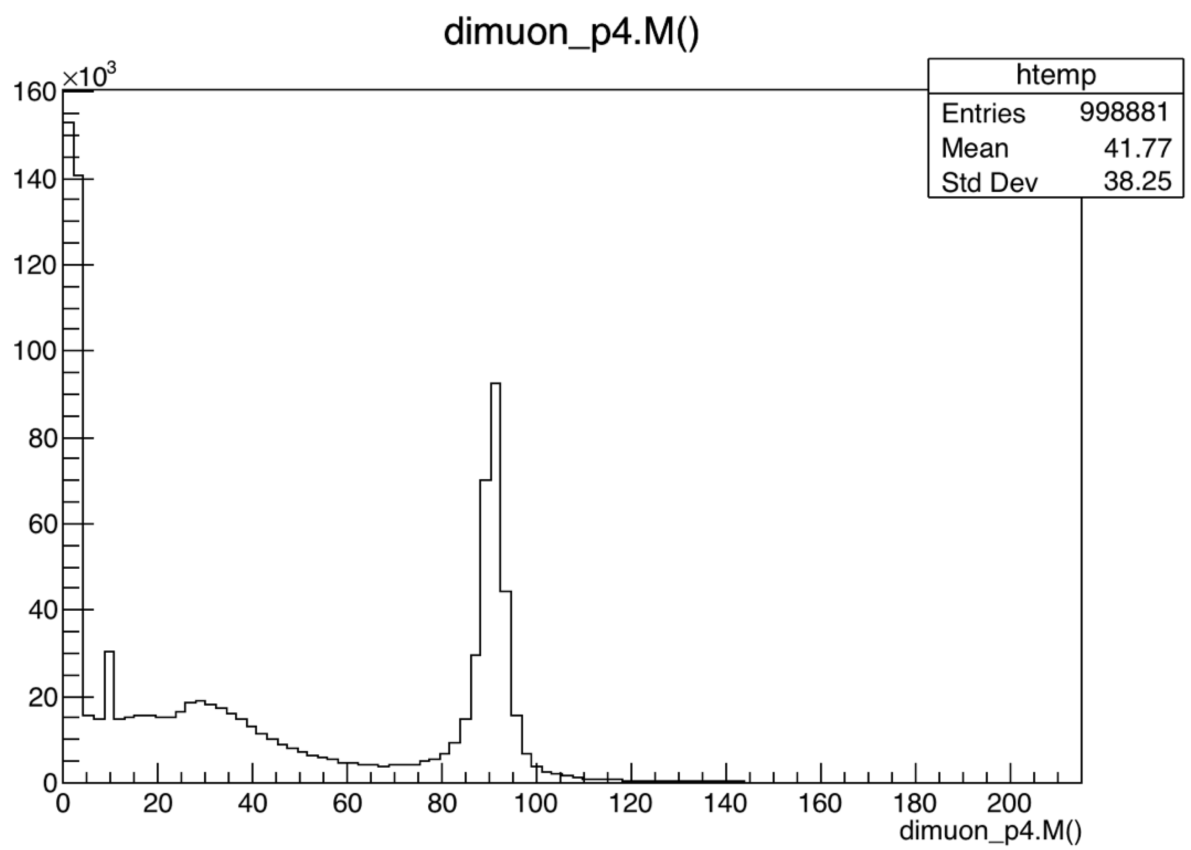
these are the particles' 4-momenta **P** https://root.cern.ch/doc/master/classTLorentzVector.html

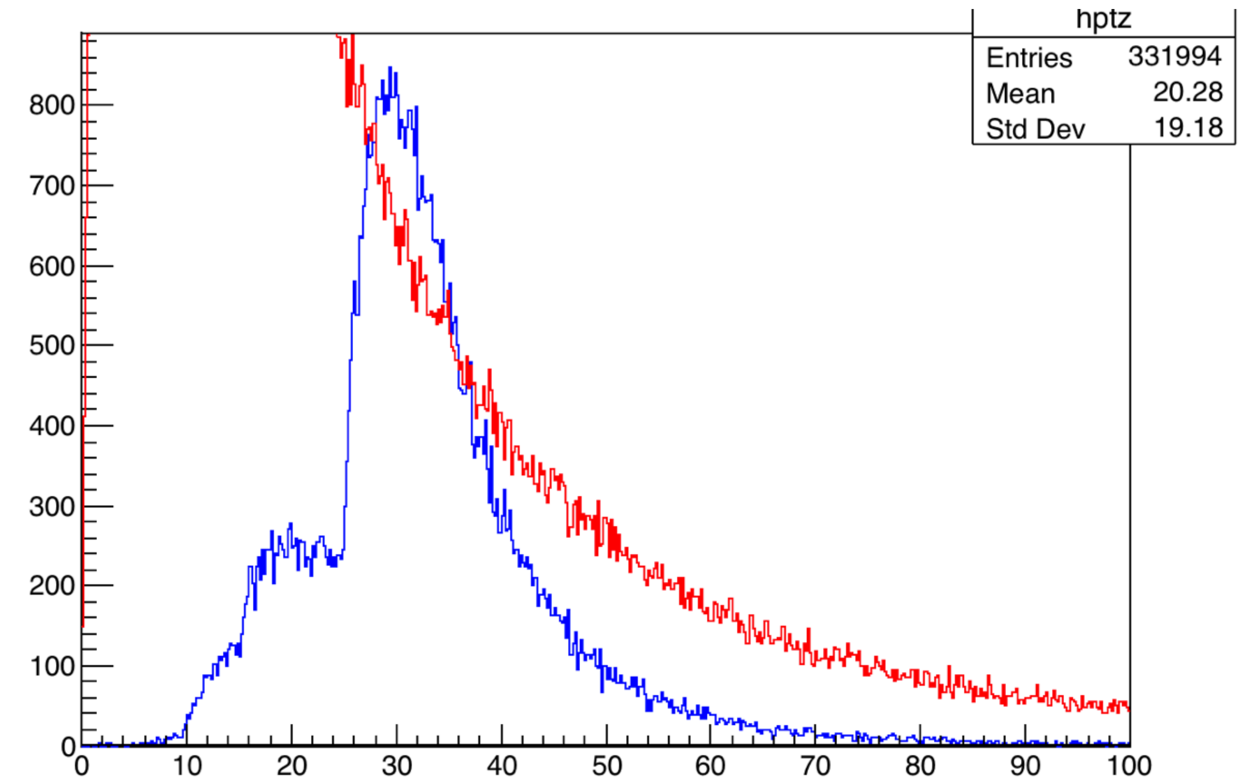
```
root [] oniaTree->Draw("dimuon_p4.M()")
Info in <TCanvas::MakeDefCanvas>:  created default TCanvas with name c1
```

invariant mass: `dimuon_p4.M()`

kinematic distributions

```
root [] oniaTree->Draw("dimuon_p4.Pt()")

root [] oniaTree-
>Draw("dimuon_p4.Pt()>>hptz(500,0,100)","dimuon_p4.M()>70")

root [] oniaTree-
>Draw("dimuon_p4.Pt()>>hptj(500,0,100)","dimuon_p4.M()>3.0&&dimuon_p
4.M()<3.2")

root [] hptz->SetLineColor(kRed)
root [] hptj->SetLineColor(kBlue)
root [] hptz->Draw("same")
root [] hptj->Draw("same")

root [] .q
```


dimuon_p4.M()
htemp
Entries 998881
Mean 41.77
Std Dev 38.25
×10³
160
140
120
100
80
60
40
20
0
0 20 40 60 80 100 120 140 160 180 200
dimuon_p4.M()

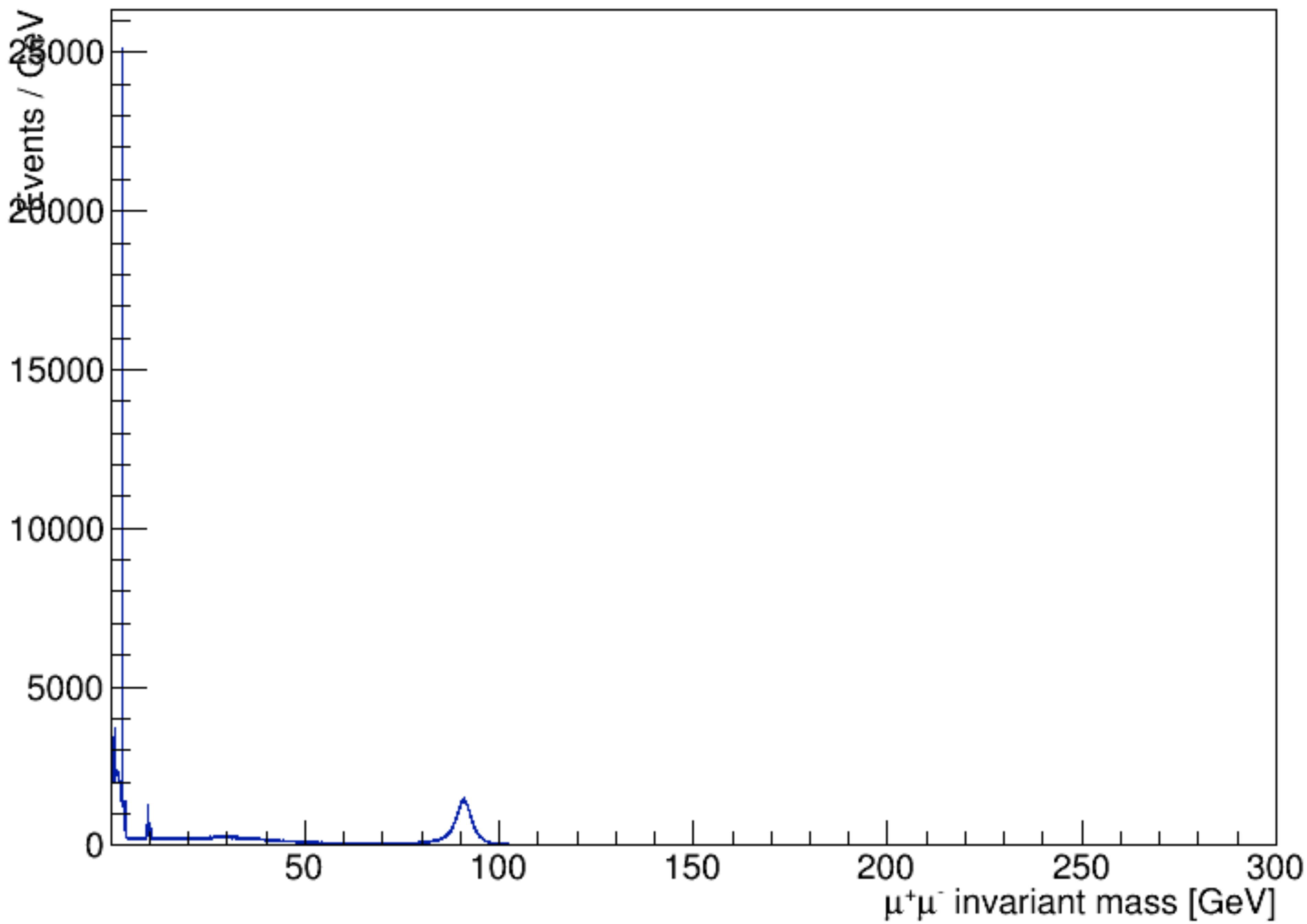

hptz
Entries 331994
Mean 20.28
Std Dev 19.18
800
700
600
500
400
300
200
100
0
0 10 20 30 40 50 60 70 80 90 100

the code

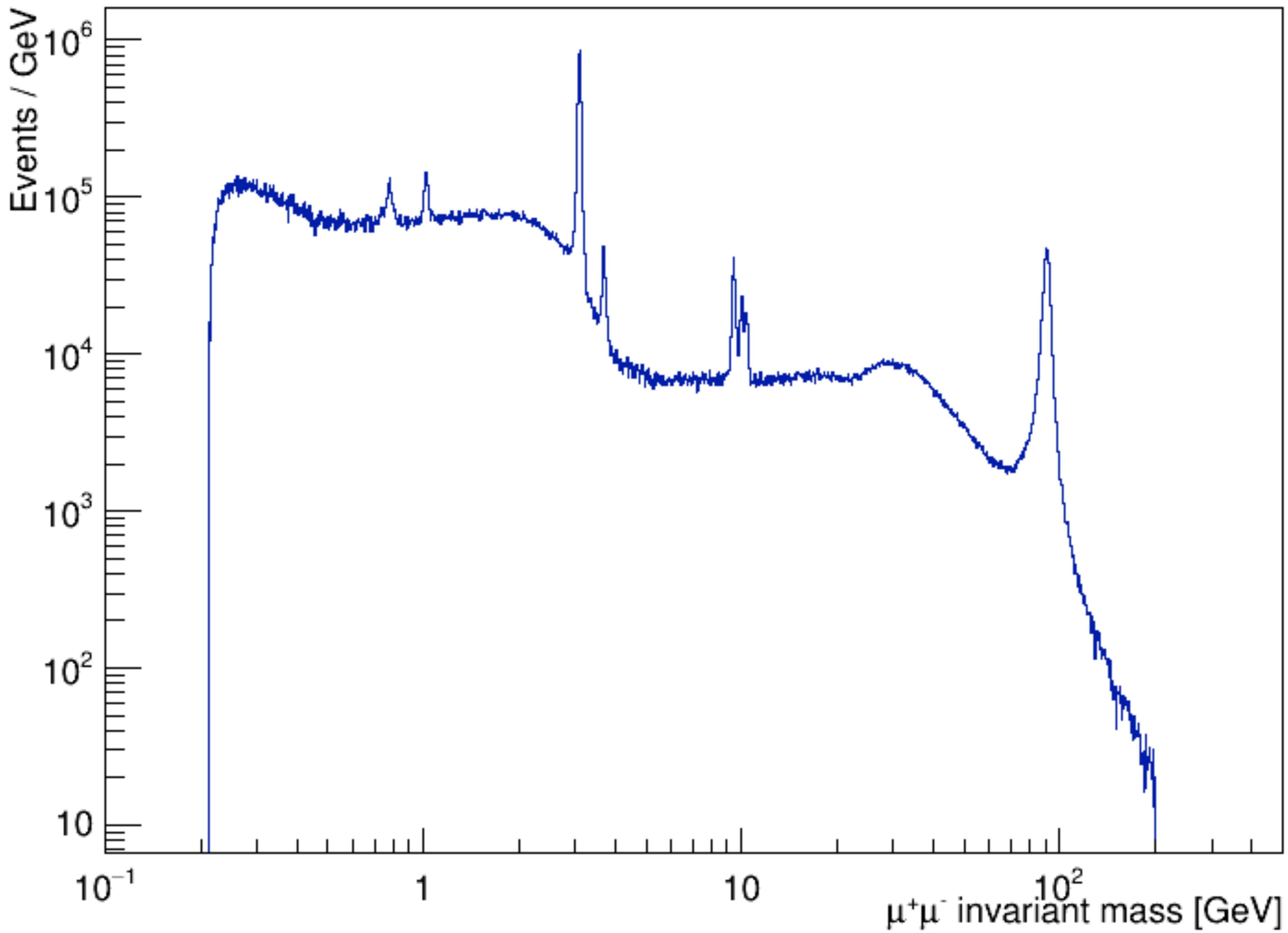
- main methods
  - GetSpectrum(): create the dimuon spectrum from the raw dataset
  - Cut(): allows to place selection cuts
  - SelectPeak(): allows to select one of the signals in the spectrum
  - FitPeak(): fits the data and extracts signal parameters

```
emacs dimuons.C &

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

ls plots
```

the ‘raw’ spectrum

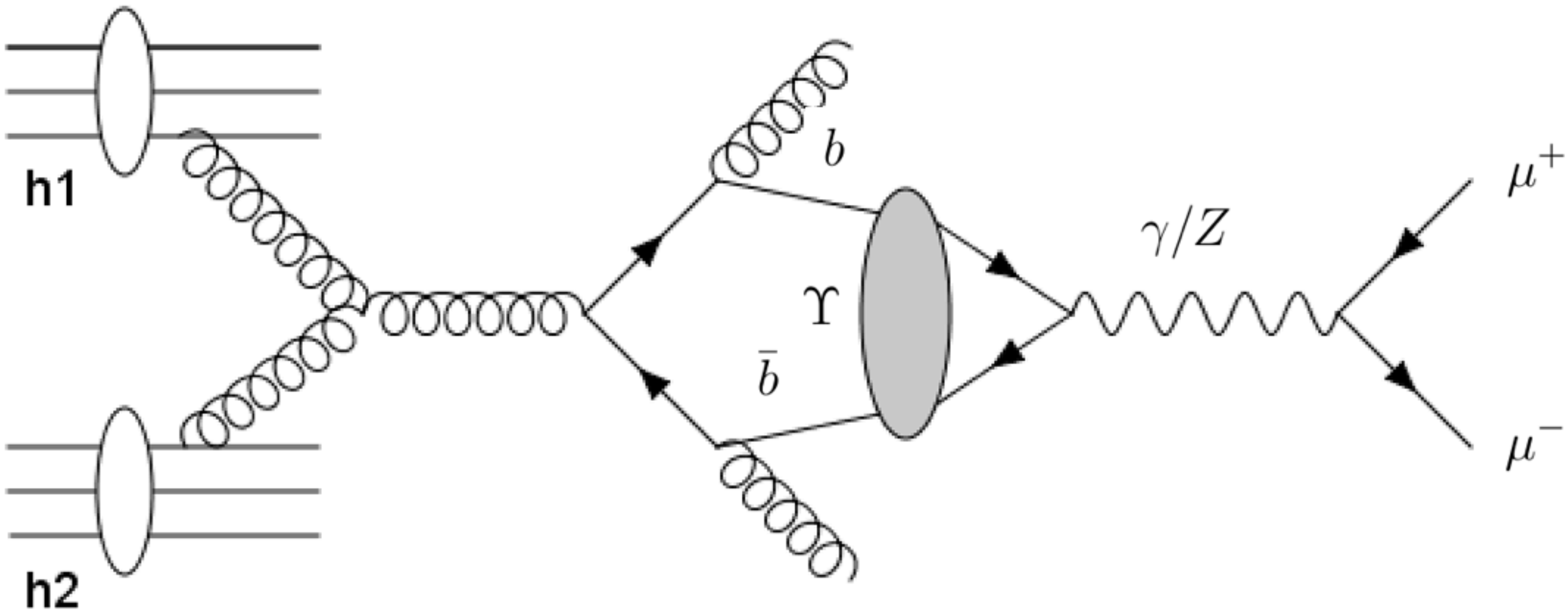


the 'right' spectrum

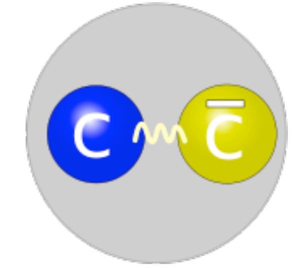


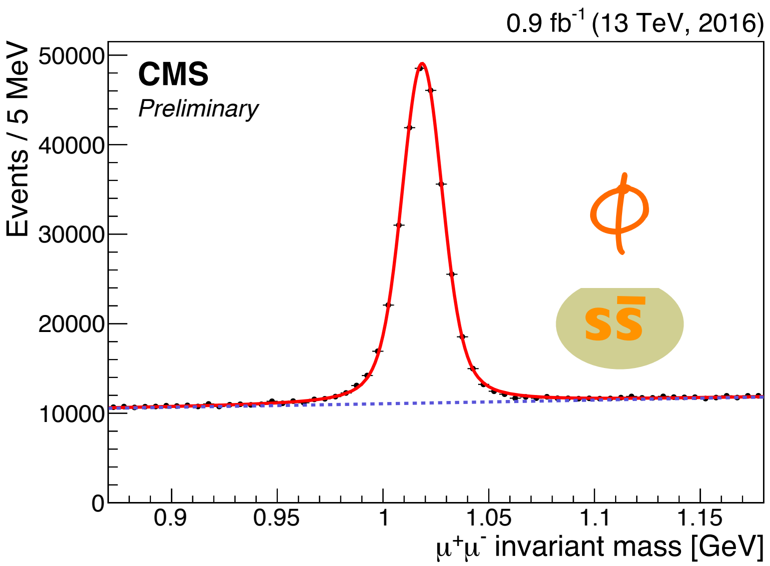
feature: variable bin widths, resolution-dependent, properly normalized, doubly-log scales

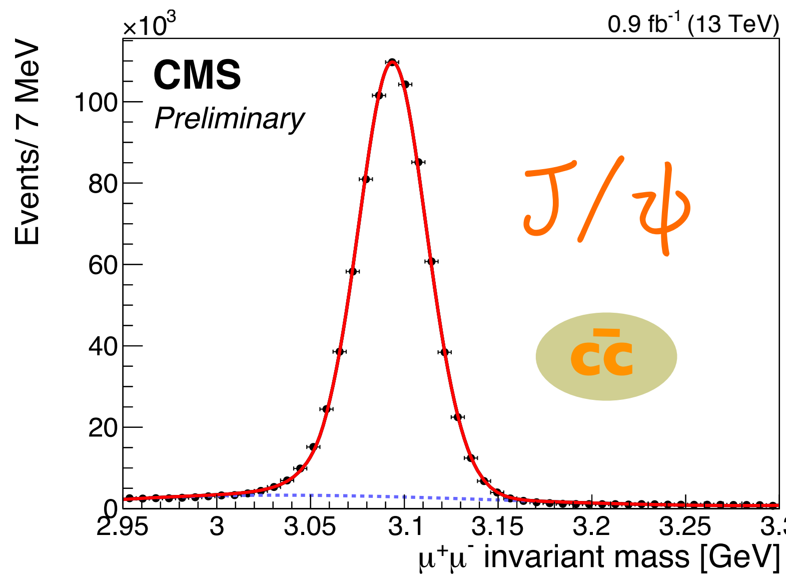
what are the peaks?

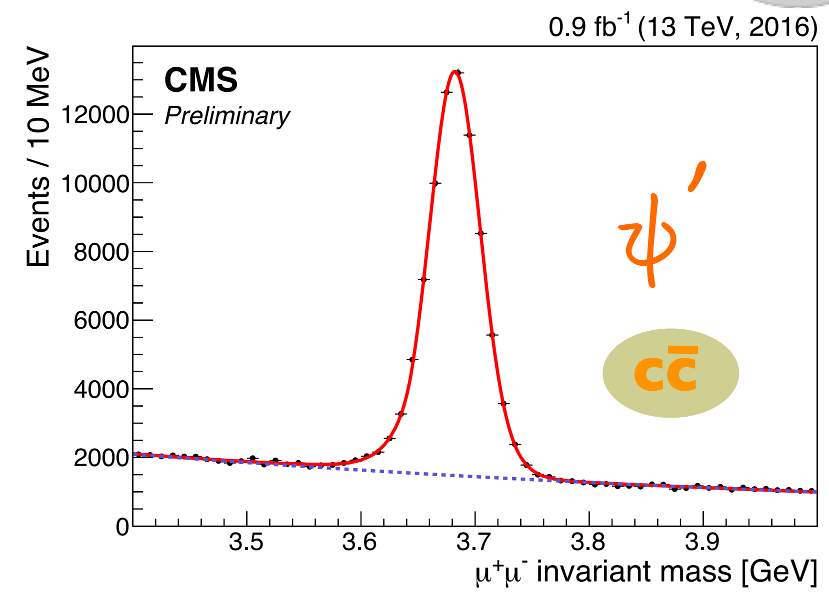


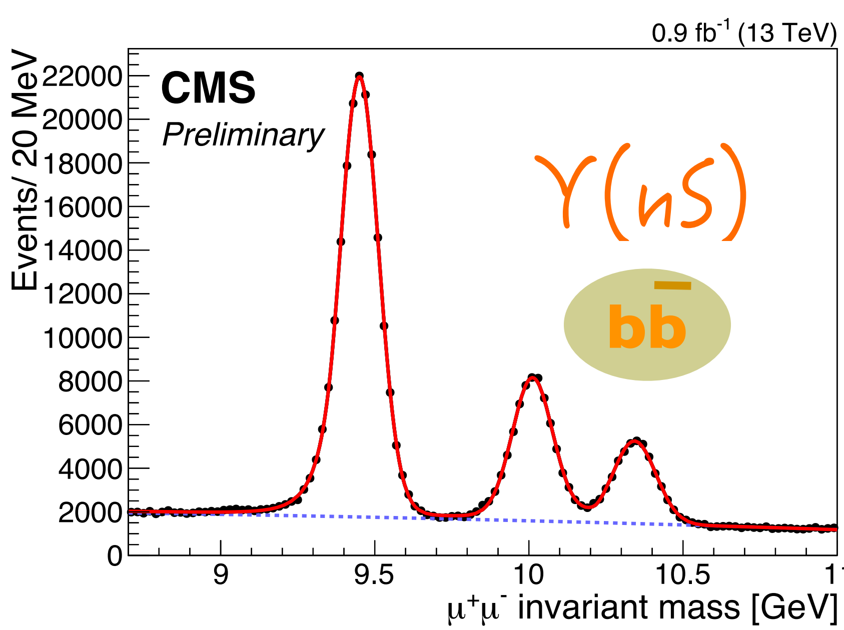
what are the peaks?

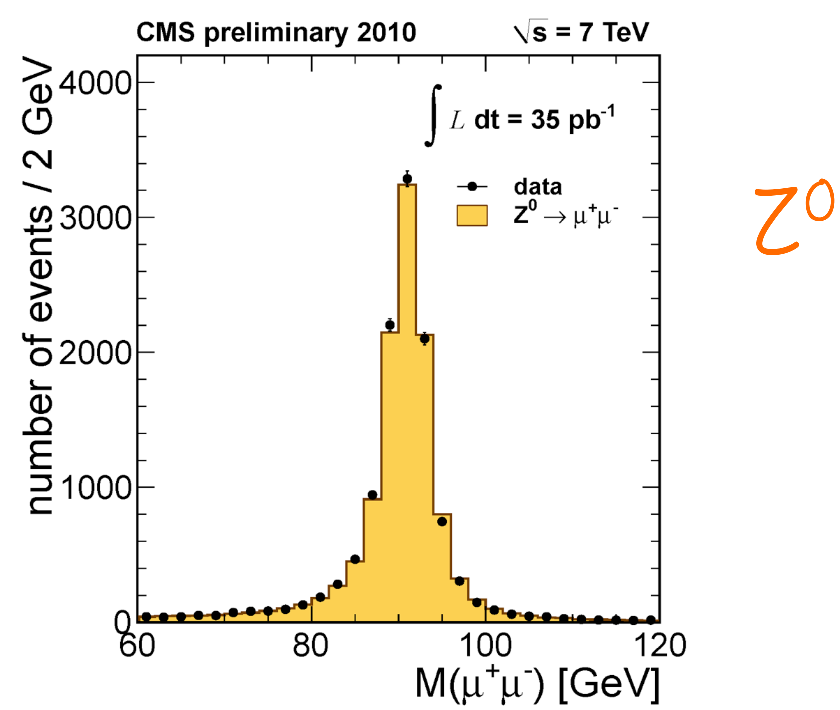






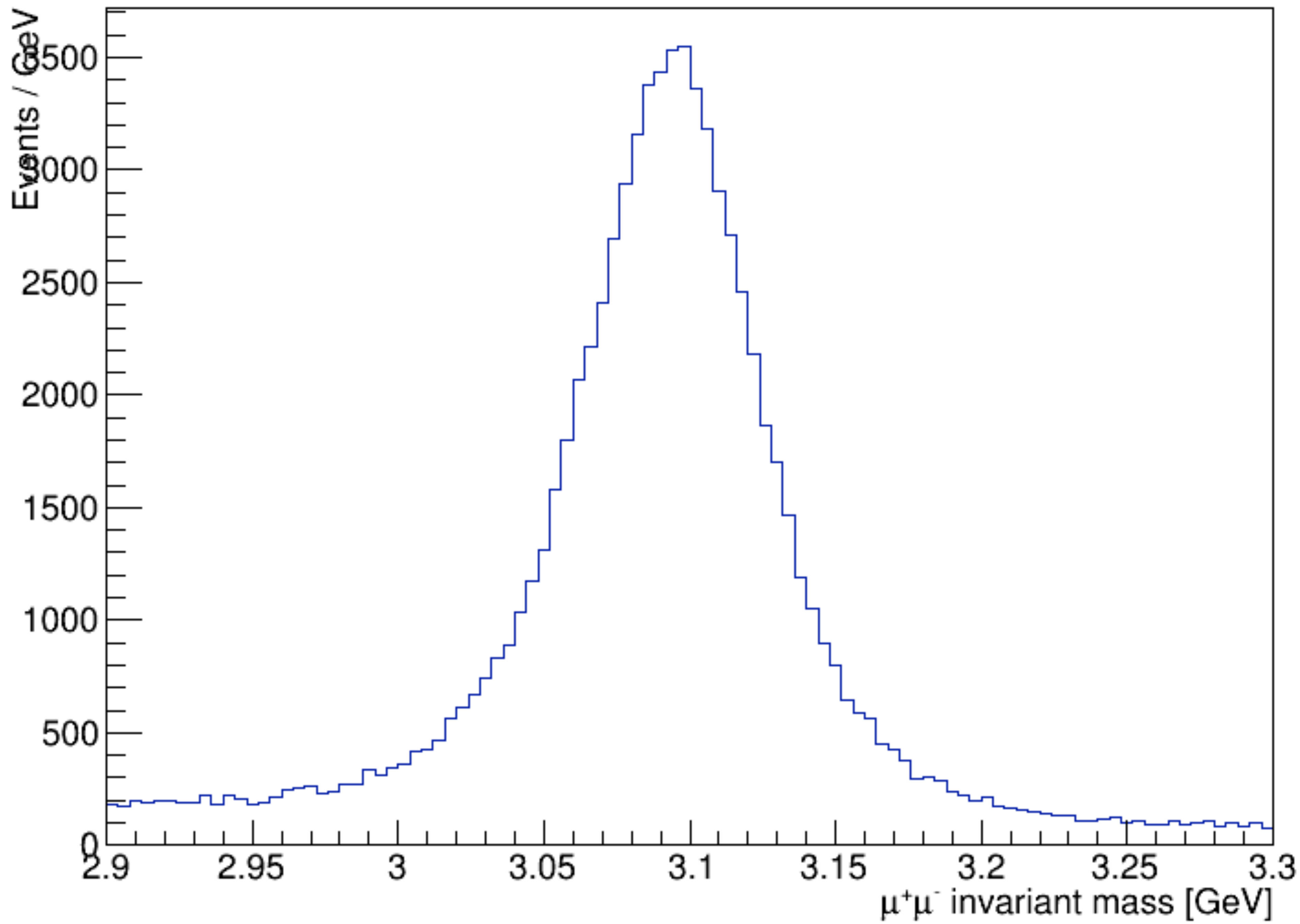






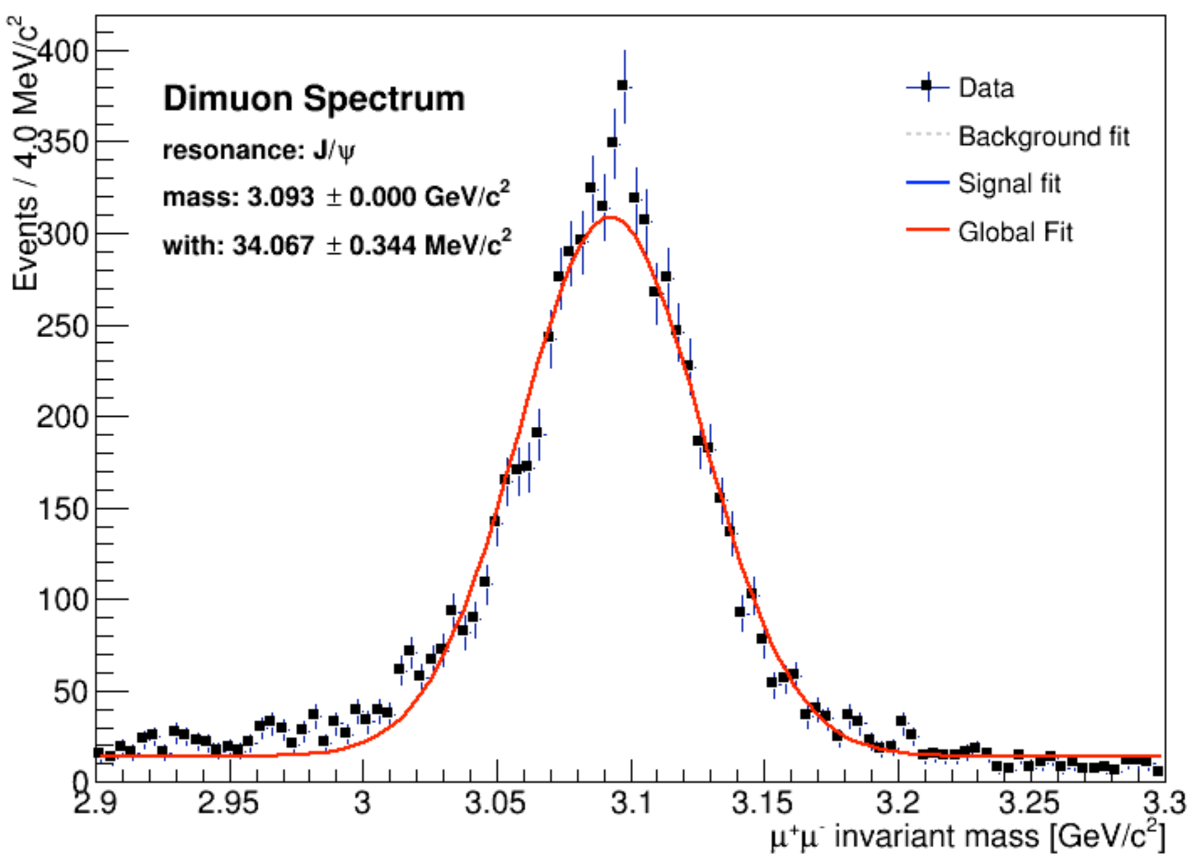
http://pdglive.lbl.gov

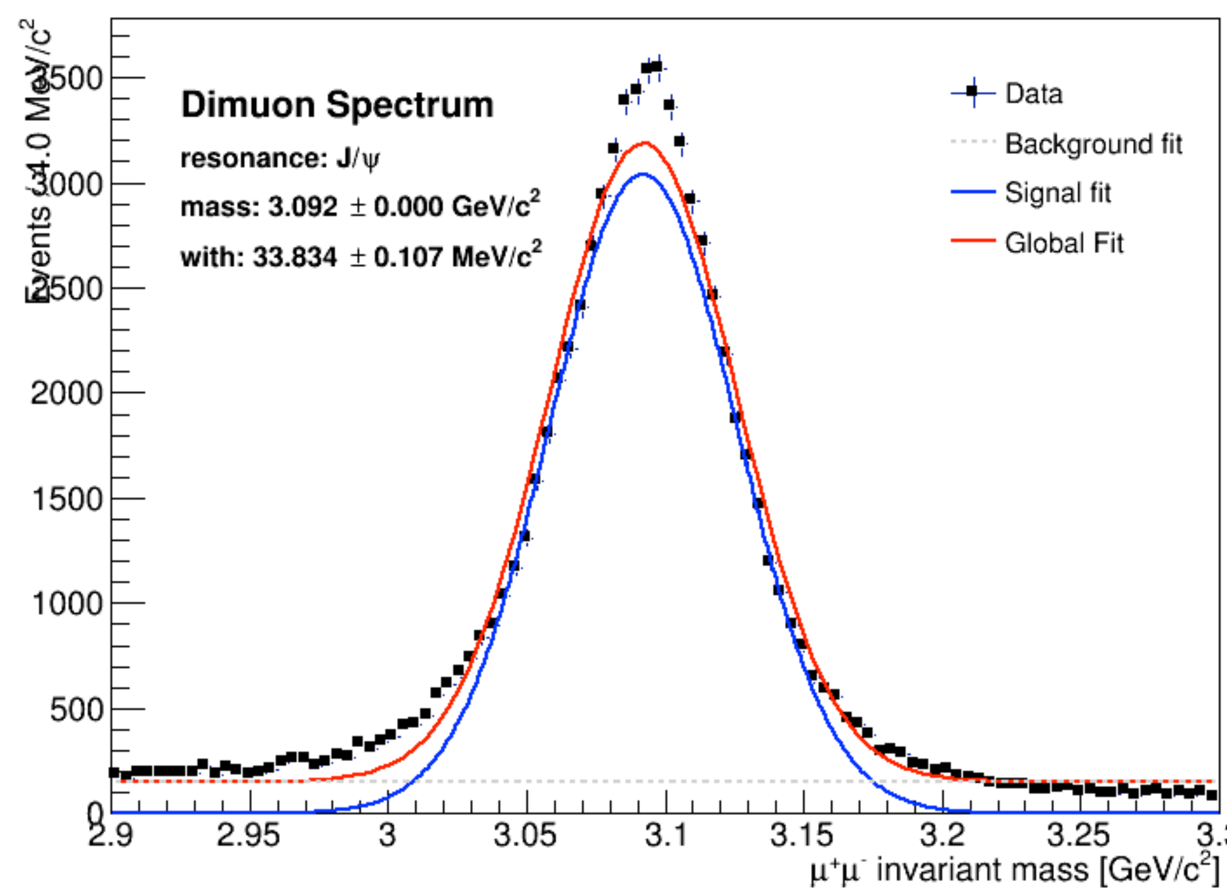
zoom in... and select a signal



fit the data

Simplified model: Gaussian (signal) + polynominal (bckr)





how can model be improved?
hint: final-state radiation

higher-yield distribution
not well described -- why?

exercise

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

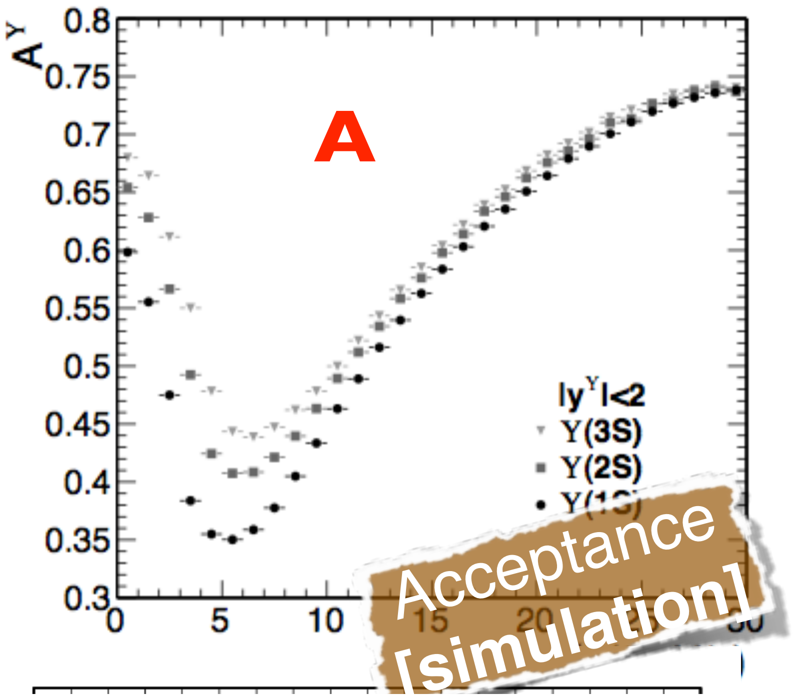
☑ congratulations: you've grasped the ingredients of a physics measurement, the production cross-section of your chosen particle!

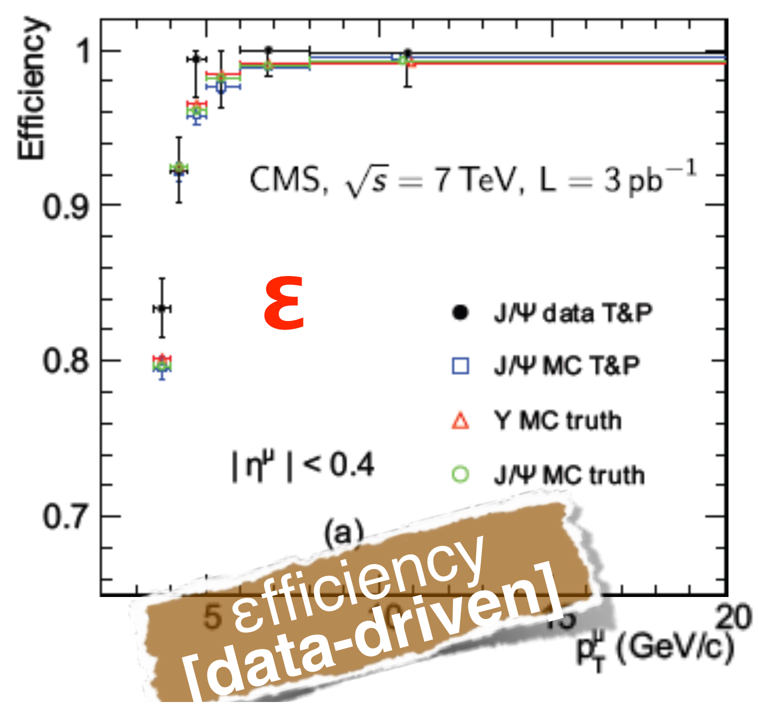
cross section

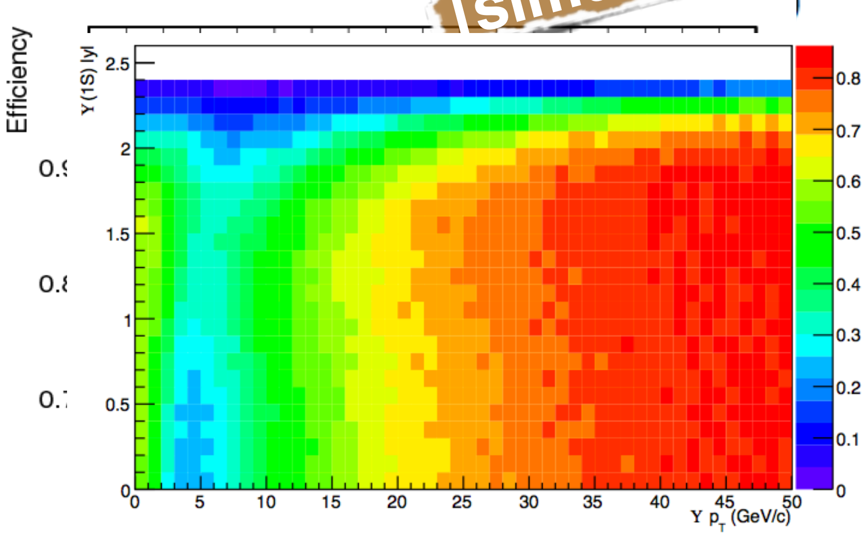
an effective area of interaction
unit: barn, 1b = 10^{-28} m^2 = 100fm^2

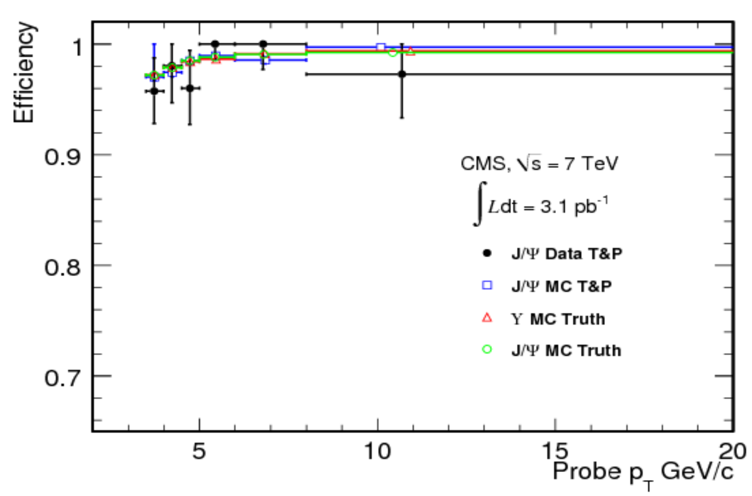
"N=L.σ"

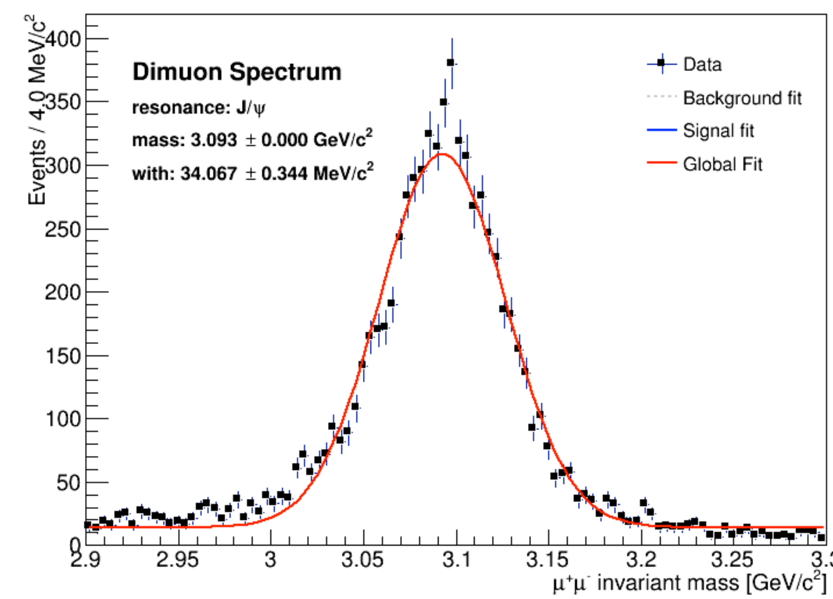
$$\frac{d^2\sigma(Q\overline{Q})}{dp_T\,dy}\mathcal{B}\left(Q\overline{Q}\rightarrow\mu^+\mu^-\right)=\frac{N_{fit}(Q\overline{Q})}{\mathcal{L}\cdot\mathcal{A}\cdot\epsilon\cdot\Delta p_T\cdot\Delta y}$$











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