Tutorial on Data Analysis

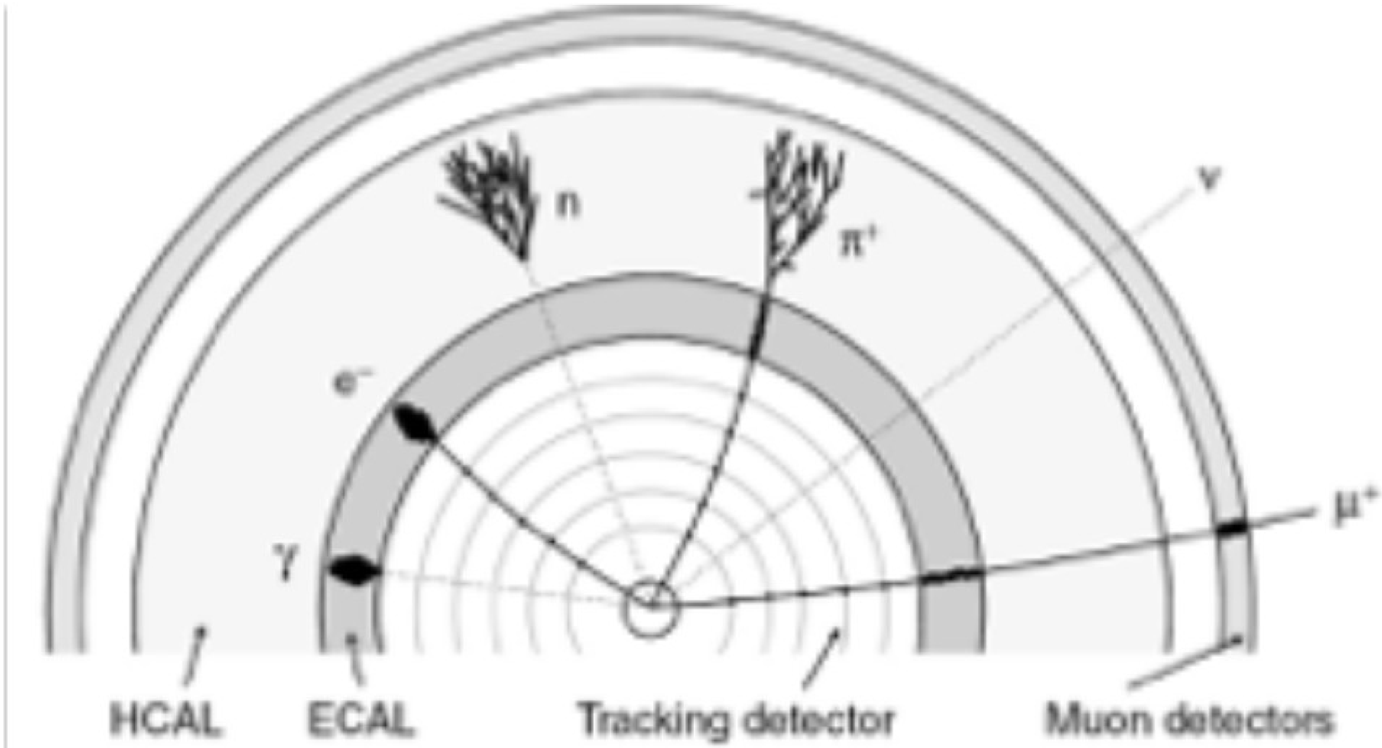
LIP internship program, 2025

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 fit
  - inspect statistical and systematic errors

Typical detector at particle colliders

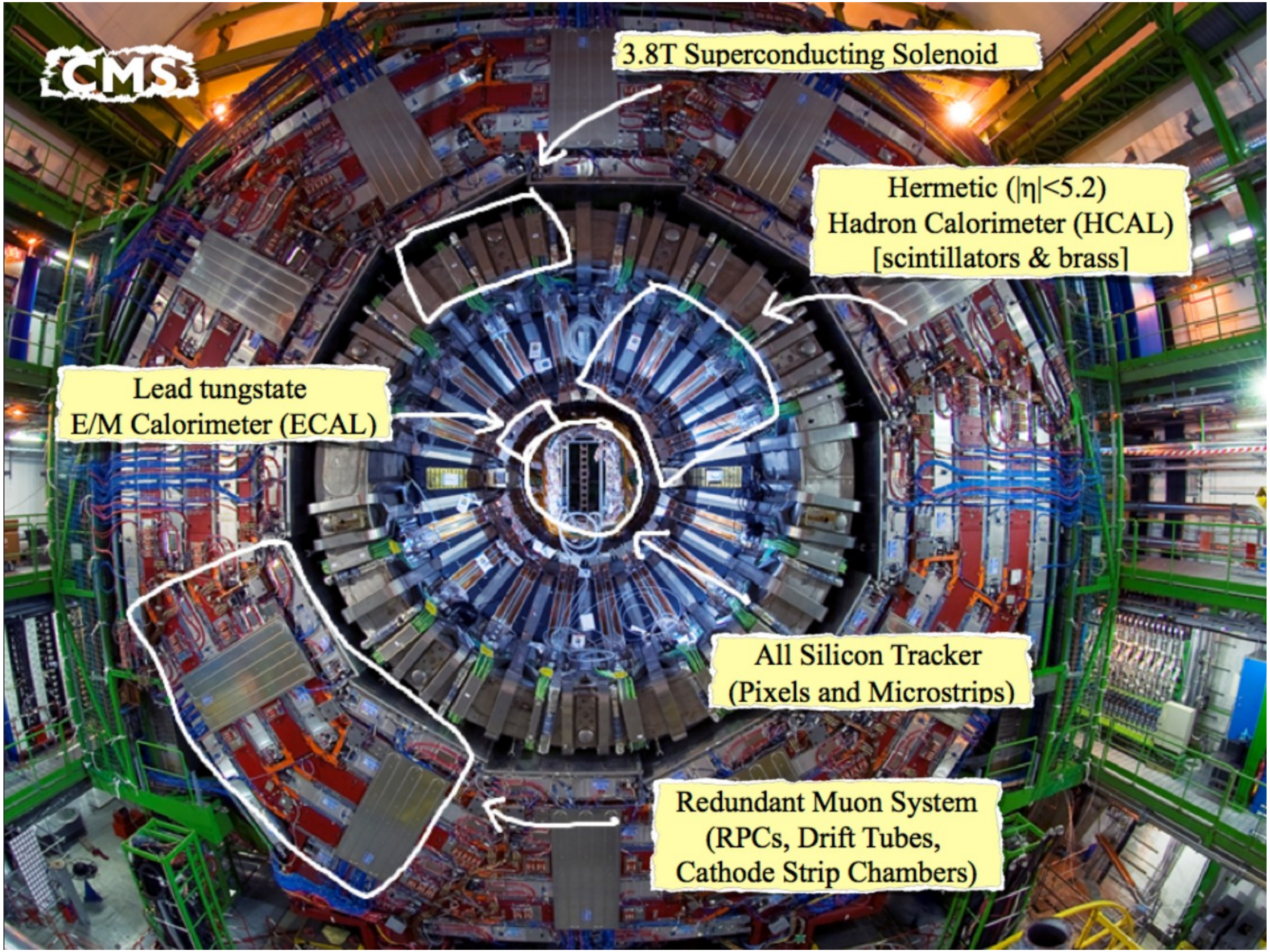


calorimeters:
measure particle's
energy by absorbing it

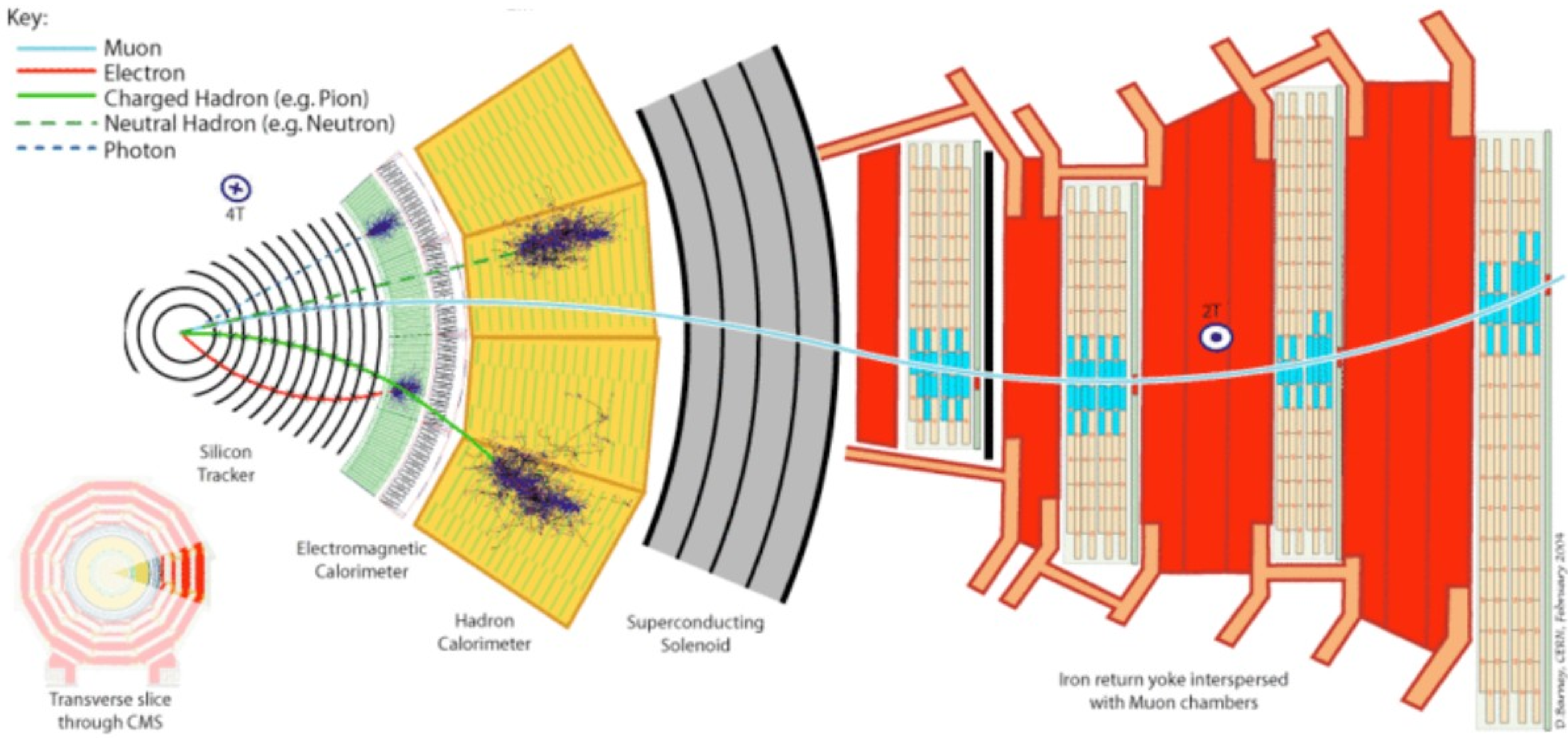
trackers:
detect trajectory
of charged particles

muons:
detected in outer
detector layers

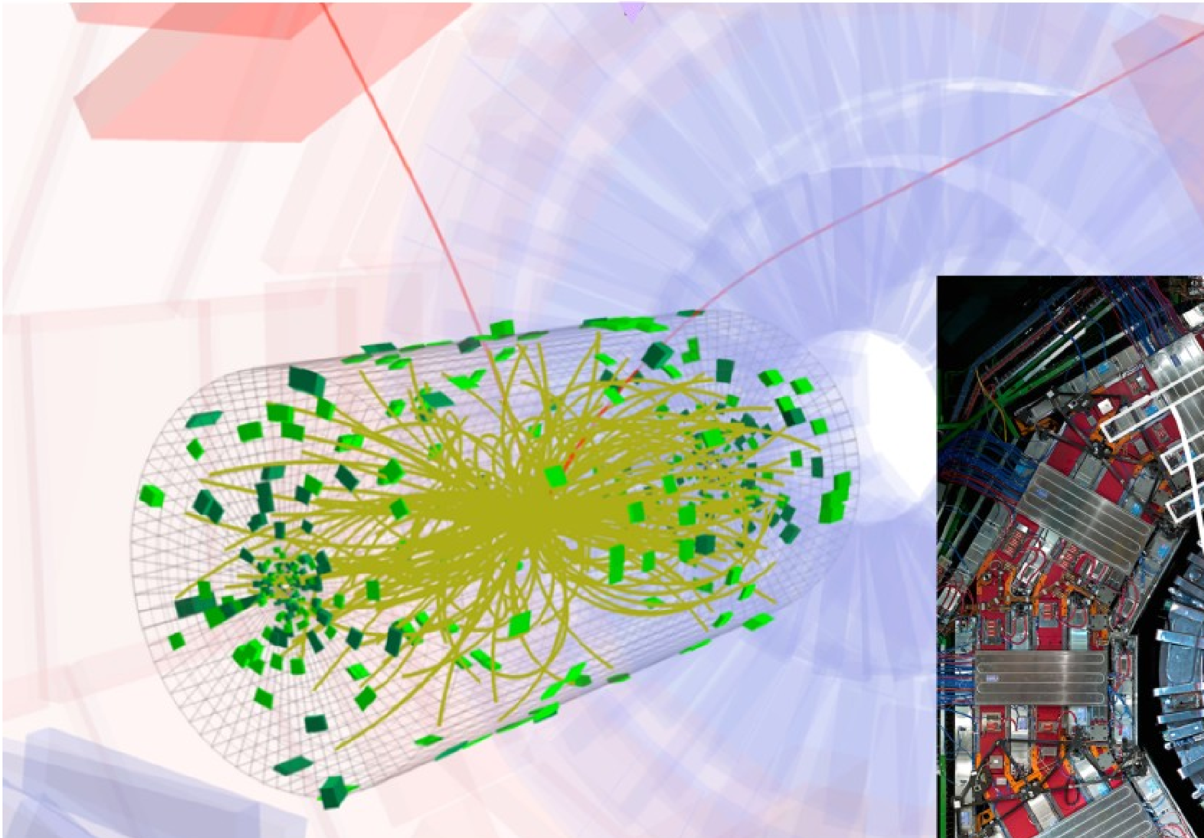
The CMS detector



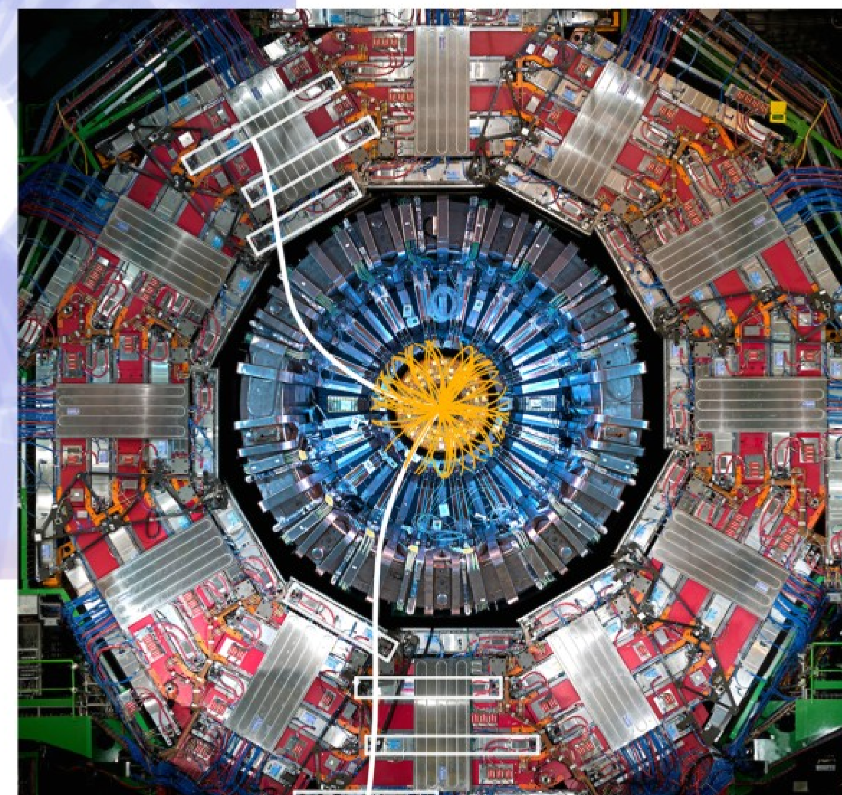
How do particles interact?



Two-muon events in CMS



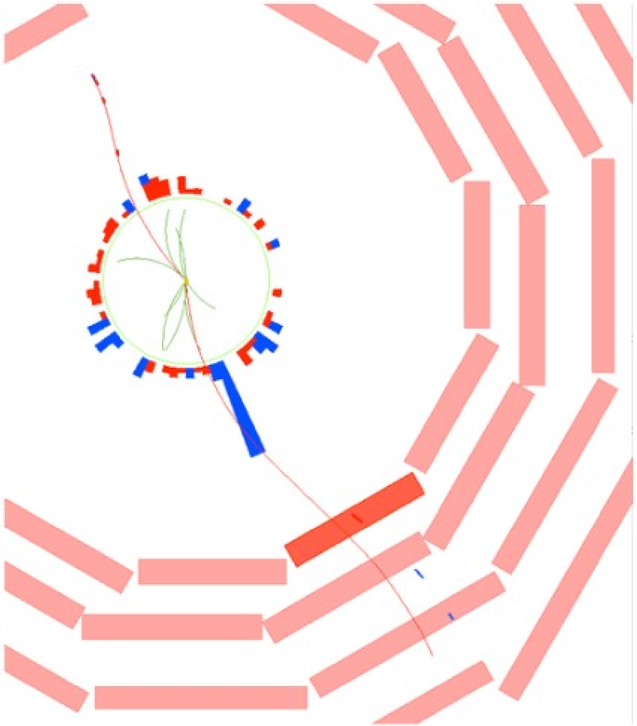
$X \to \mu\mu$



First look at the code and data

- Open the notebook in google colab
  - Download from indico agenda
  - If you never used google colab, follow instructions to set it up (simple login with your google account)
  - save a copy of the code in your area, so you can modify and run it
  - run the first blocks to set up root and open file with data
  - Let’s have a look at the content of the file!

Two-muon invariant mass

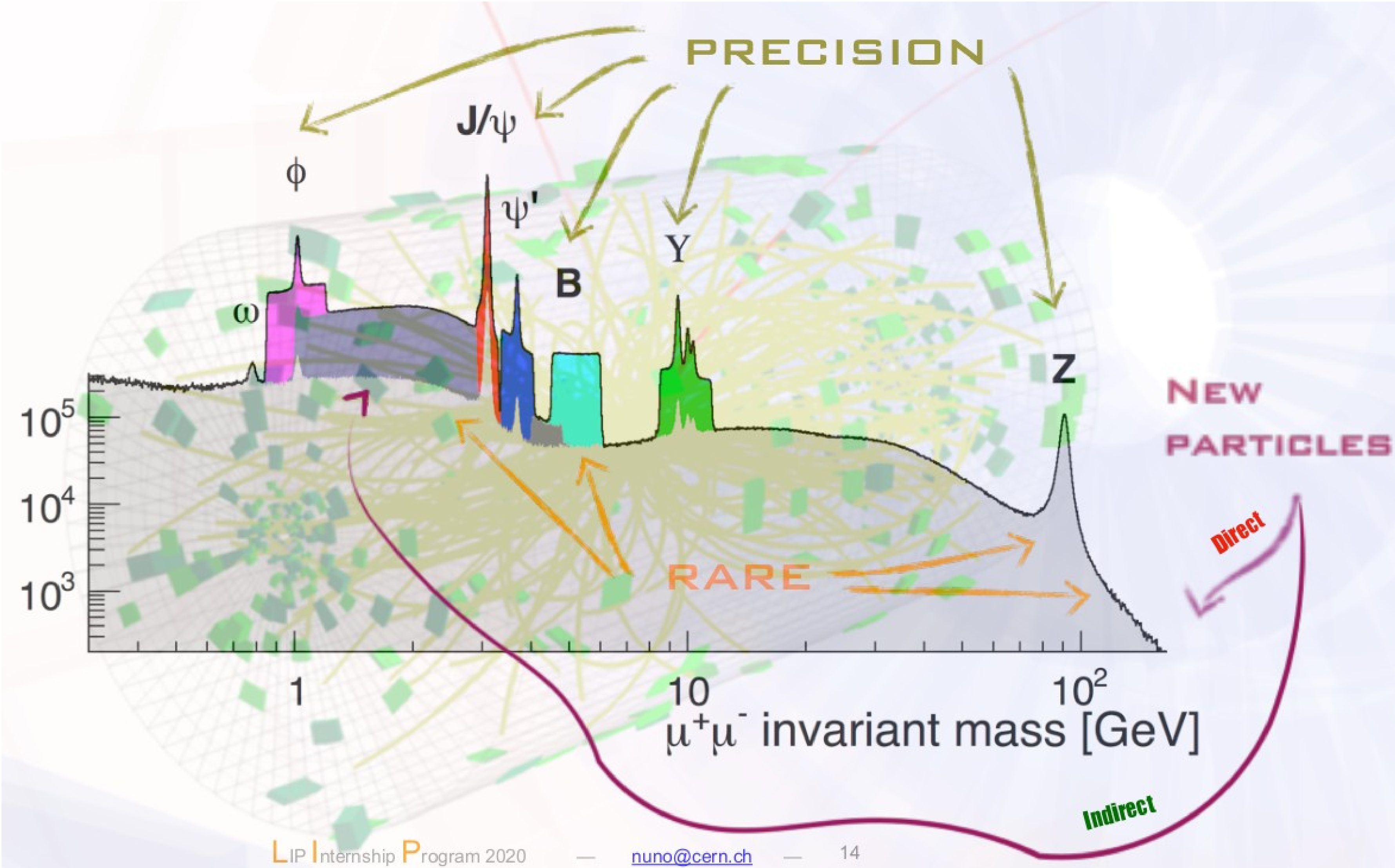


particle identification
- signal in muon chambers
→ it's a muon!
➠ $m = m(\mu) \sim 106 MeV/c^2$

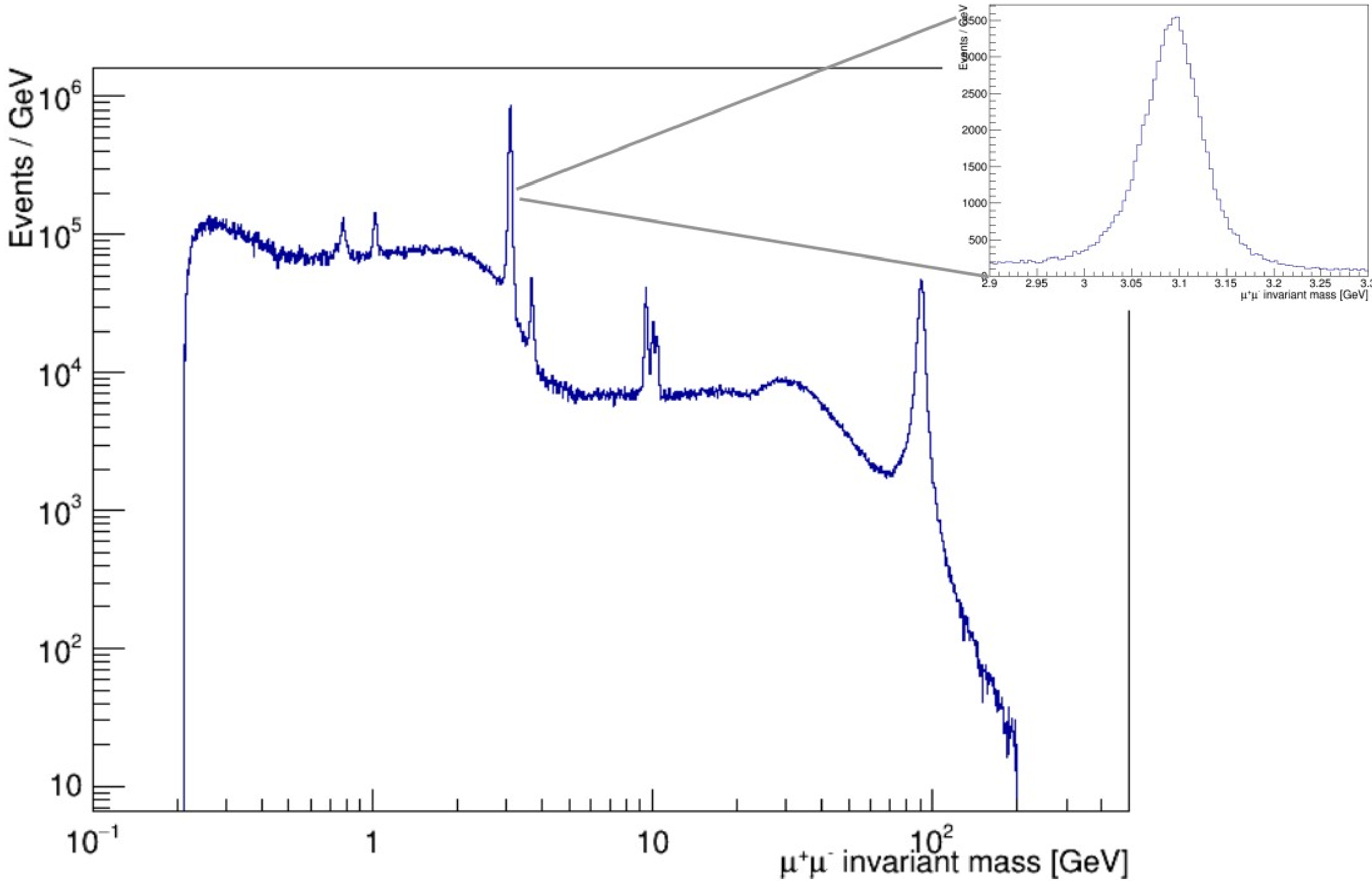
particle trajectory
- muon chambers but especially the silicon tracker
➠ linear momentum, $\underline{p} \equiv (p_x, p_y, p_z)$

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

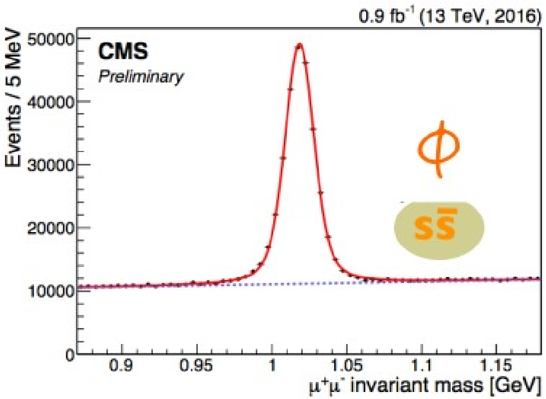
The dimuon spectrum

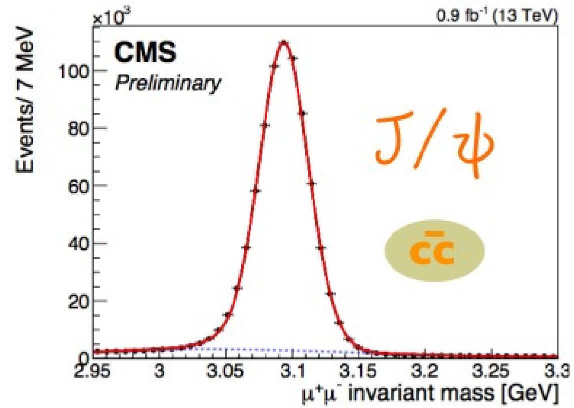


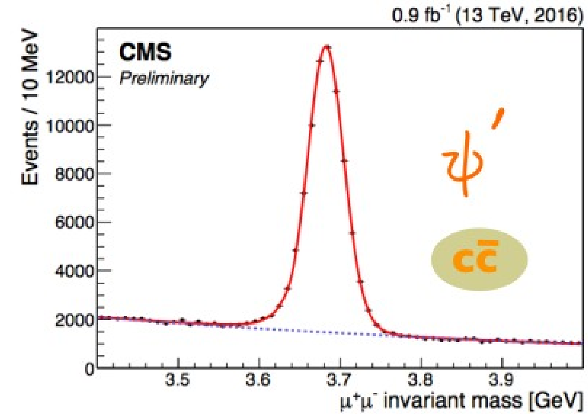
Back to the code: plot the dimuon invariant mass

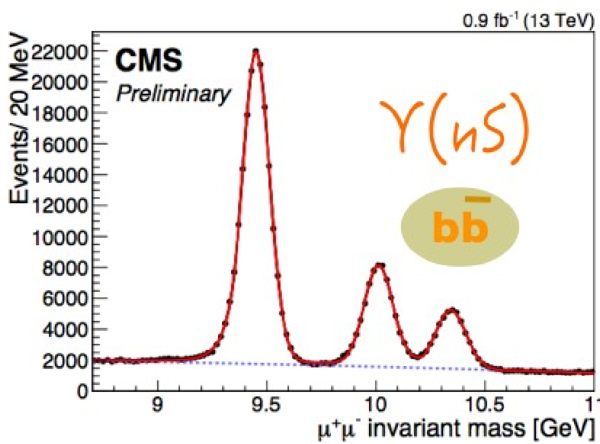


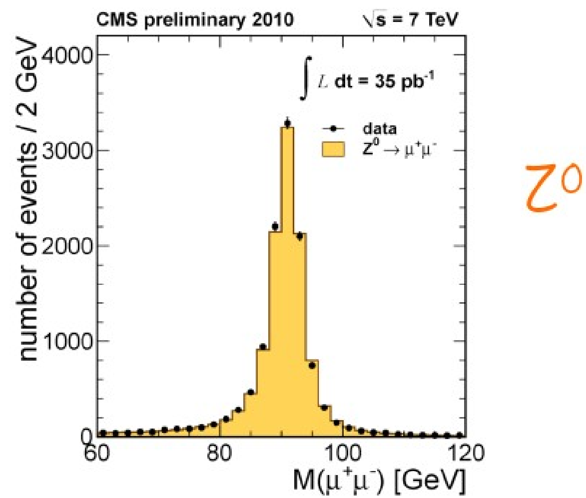
What are the peaks?







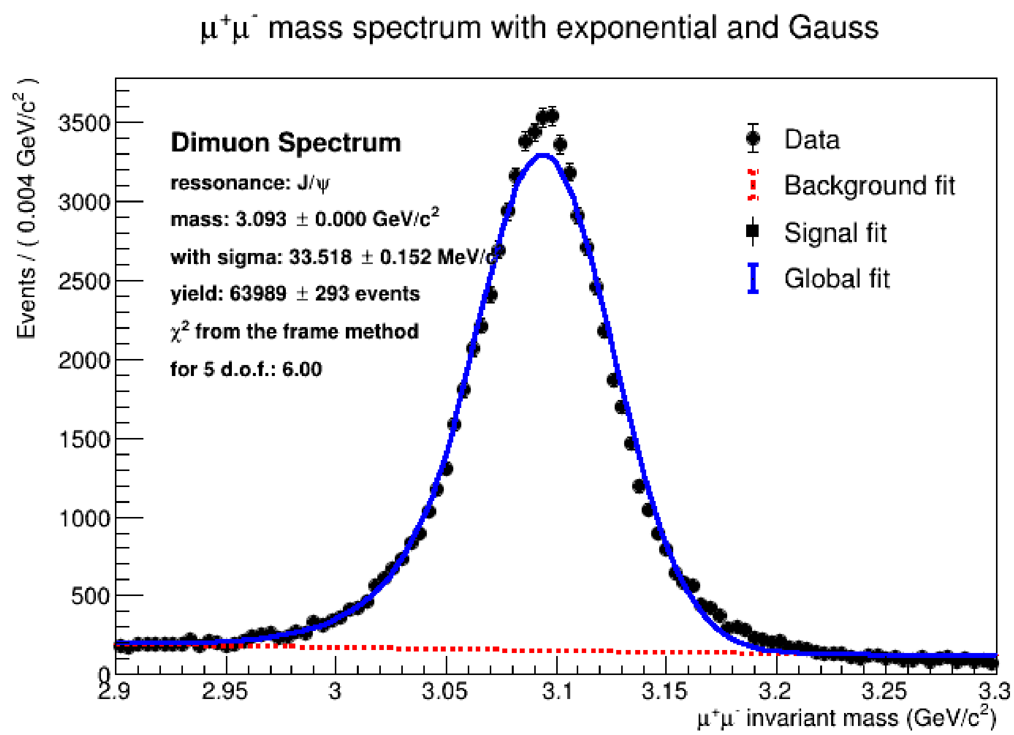




Check their measured properties at http://pdglive.lbl.gov

Fit the data!

- Choose your favourite peak (other than the J/ψ)
- Establish a fit model. Starting point:
  - signal: Gaussian function
  - background: exponential function
- Inspect quality of fit
  - can model be improved?
  - hint: final state radiation (μ → μɣ) may distort shape
- 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
  - $N \pm \sigma_{stat} \pm \sigma_{syst}$

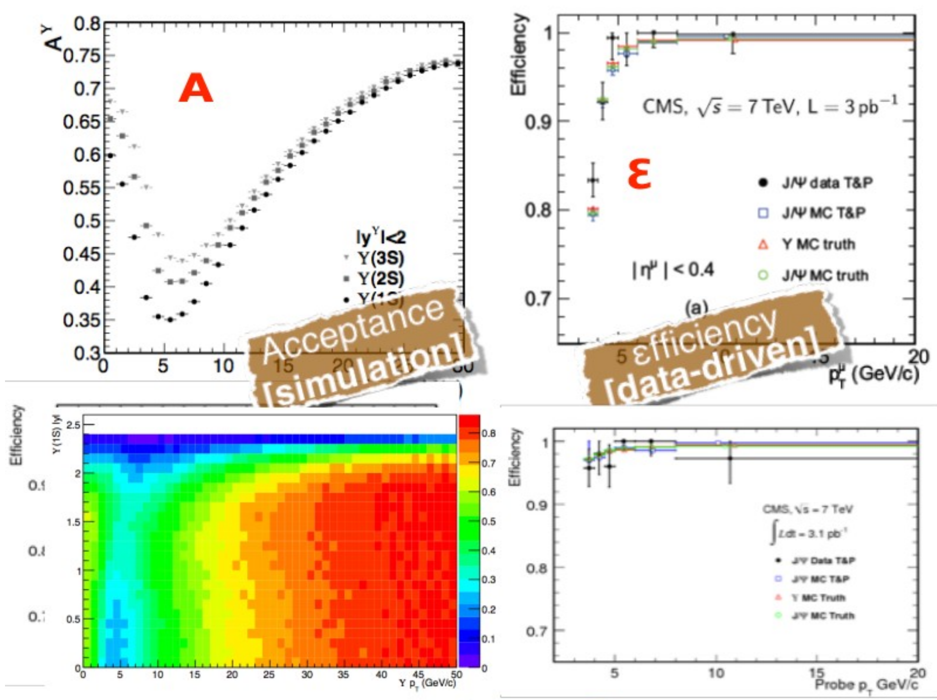
What do we learn from the yield?

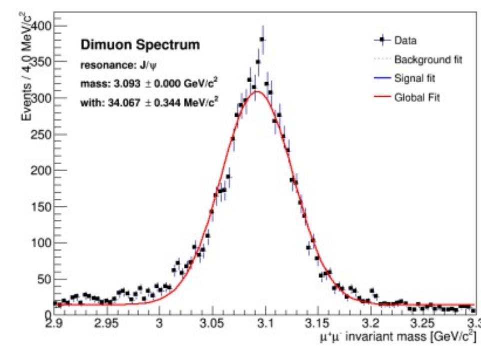
Cross section

“N=L.σ”

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

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





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