CMS Open Data at an undergraduate college

J. HOGAN, K. JOHNSON, G. KNEFELKAMP, C. RODRIGUEZ, E. SCHARNICK, G. SCHINDLER, J. SISSON

BETHEL UNIVERSITY

6/8/2023

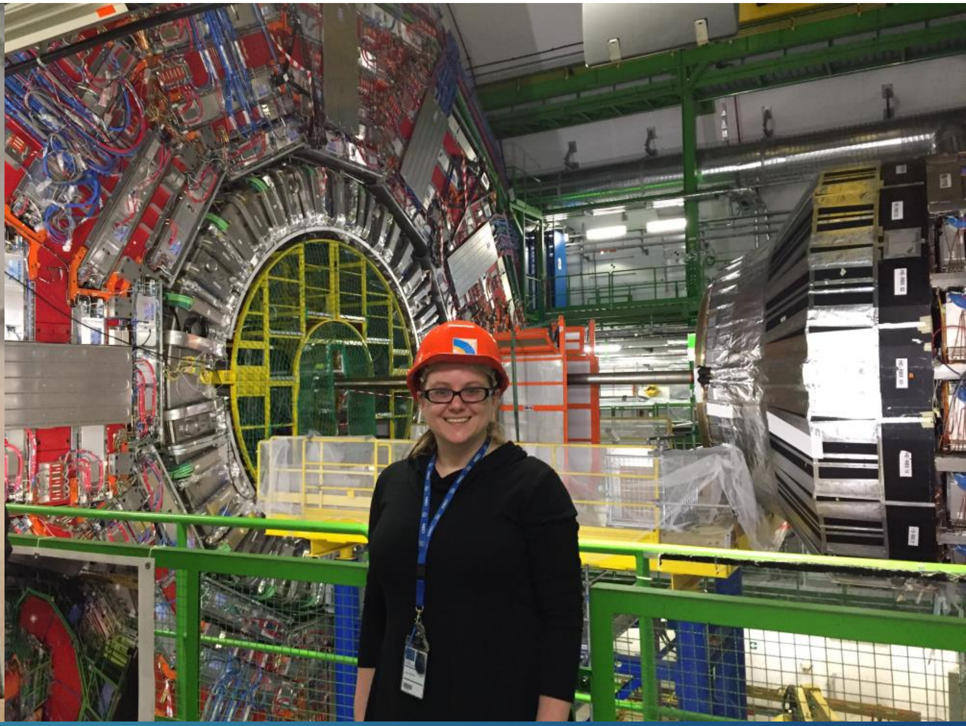
CMS Experiment

- Bethel's advanced labs integrate faculty research into the lab experience
  - Mine: Compact Muon Solenoid experiment at the CERN LHC

Latest technology → Earliest times

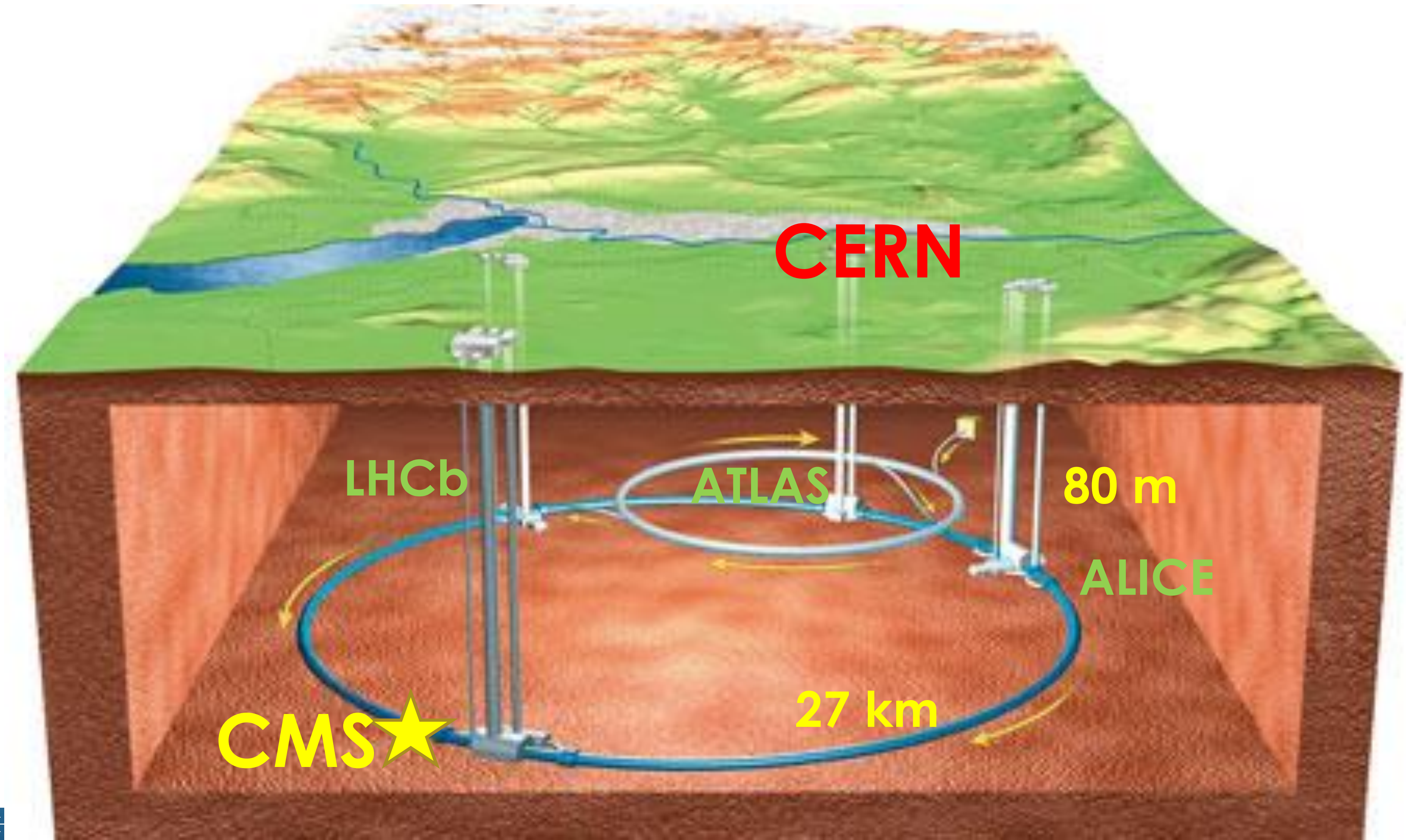
...to study the...

Largest machine → Smallest particles



The Large Hadron Collider

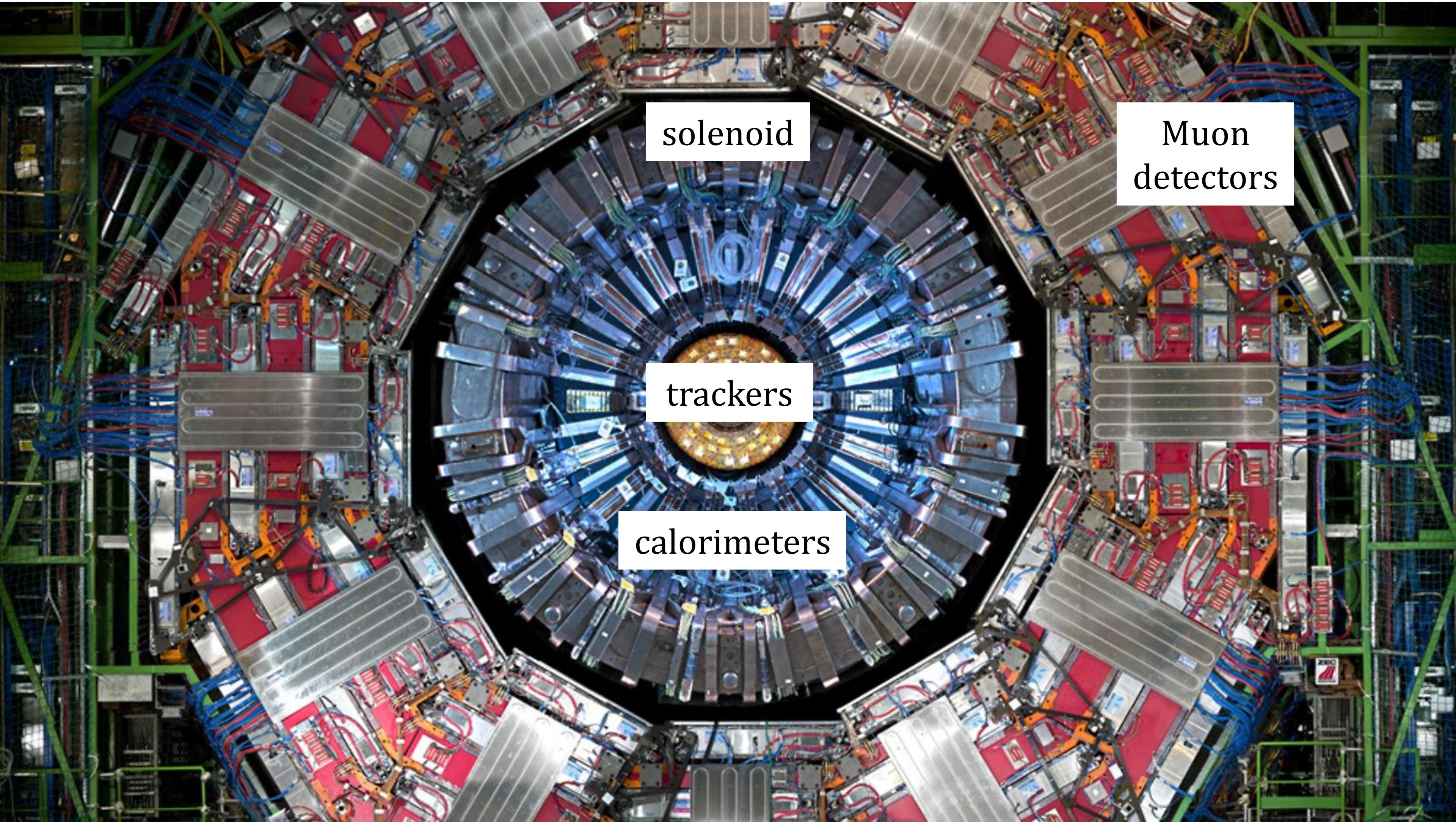
- Collides bunches of protons as well as ions



The CMS Detector

BETHEL UNIVERSITY

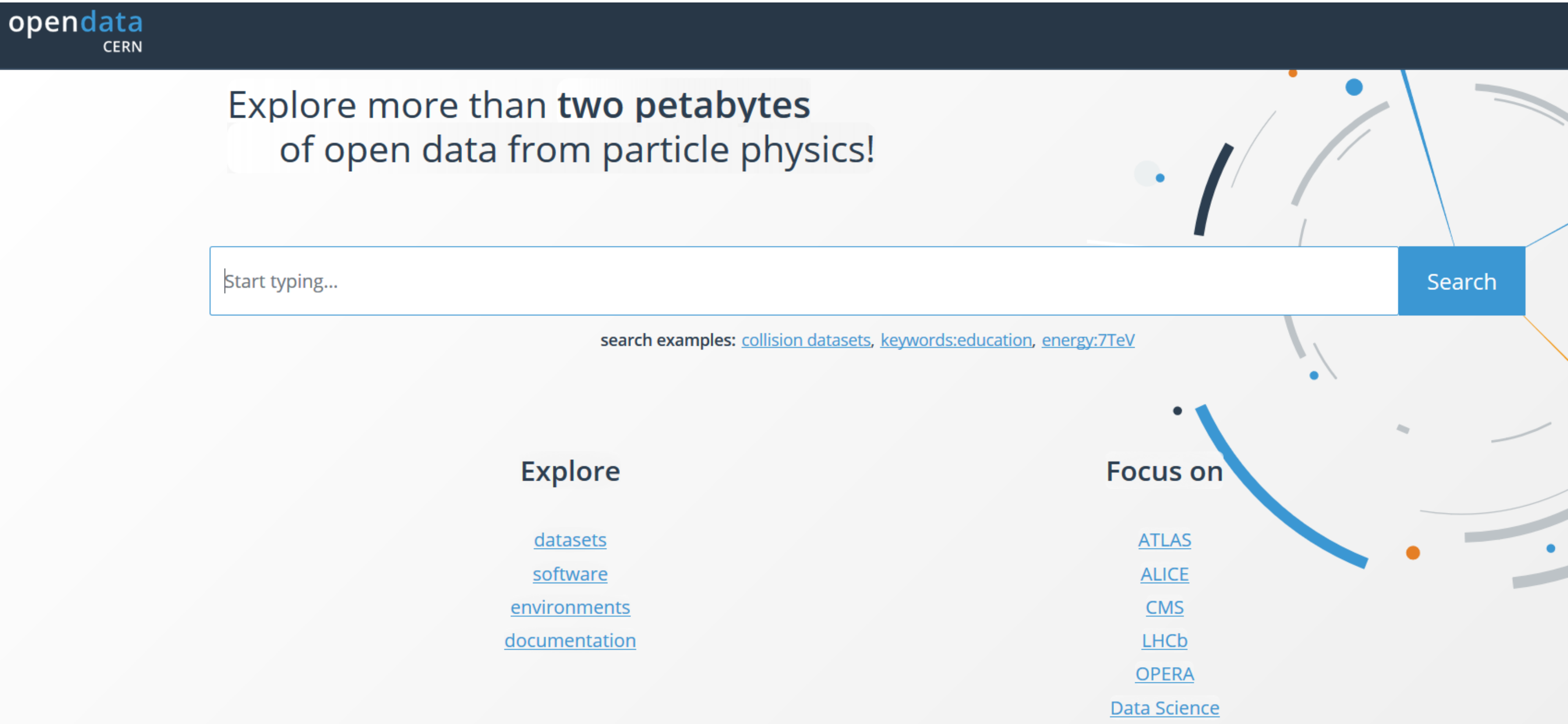
Lots of sub-sections optimized to detect different types of particles



CERN Open Data Portal

BETHEL UNIVERSITY

- CERN experiments release data to the public at intervals
- CMS has 2011/2012/2015 data and many simulated & derived datasets
- Software instructions and many educational exercises for lower levels



Open Data in the lab

BETHEL UNIVERSITY

- Modern Physics is our 4th-semester physics-sequence course
  - Required for Physics, Applied Physics, Electrical Eng, Mechanical Eng
  - Typical "baby quantum" sequence between relativity and nuclear/particle
  - Simultaneous with Computational Physics course for many students
- **Data analysis and statistical reasoning are critical learning objectives**
  - Open data lab hits many of our skill-building goals
  - Flexibility for more depth in many of these areas

Data analysis skills:

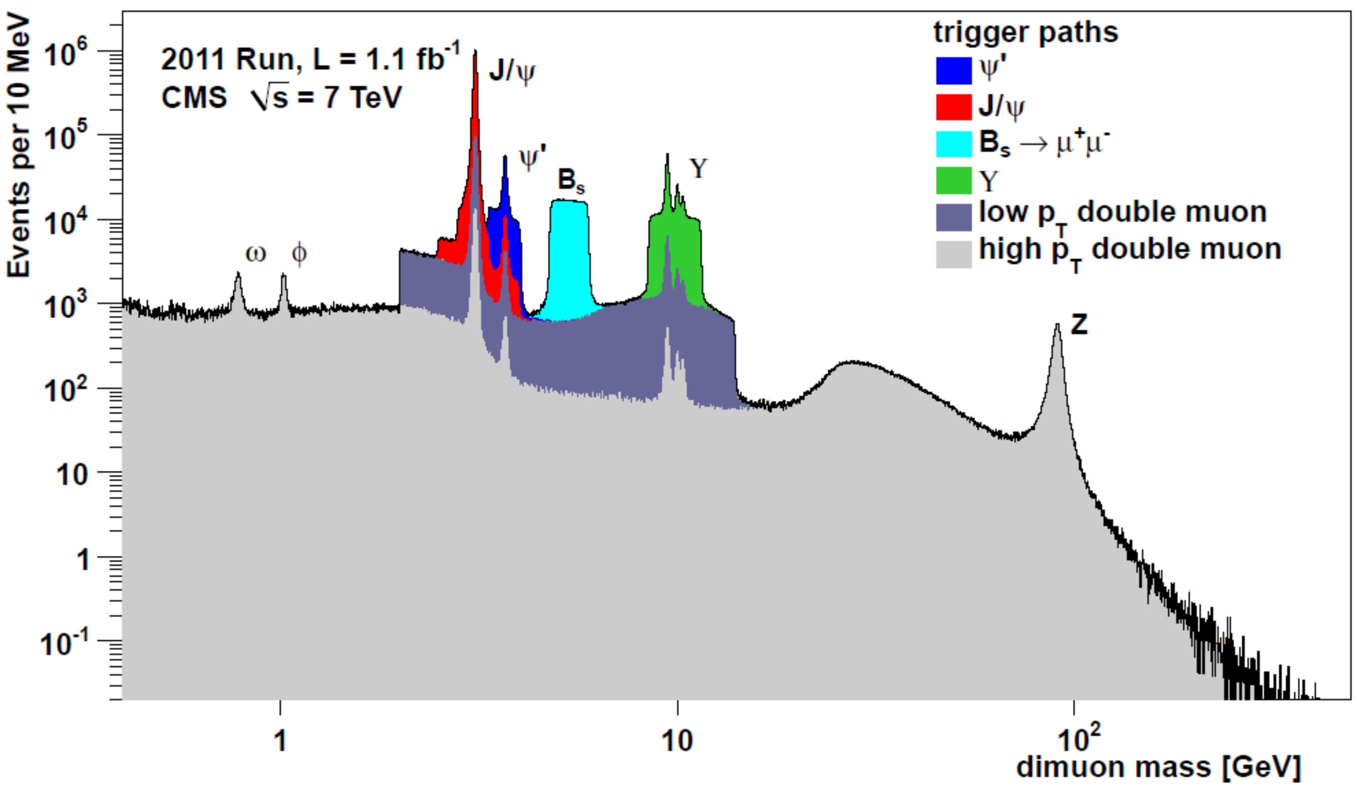
- Histograms
- Indistinguishable signal/background
- Fitting, goodness of fit metrics
- Propagation of uncertainties

Reinforce course content:

- Relativistic Dynamics
- Fundamental particles
- MATLAB programming (Comp Methods course)

Open Data in the lab

- First task for any new detector: confirmation of dimuon resonances
  - Common calibration tool
  - Consistency test of triggering mechanisms
  - **Simple and reliable final state accessible by students**



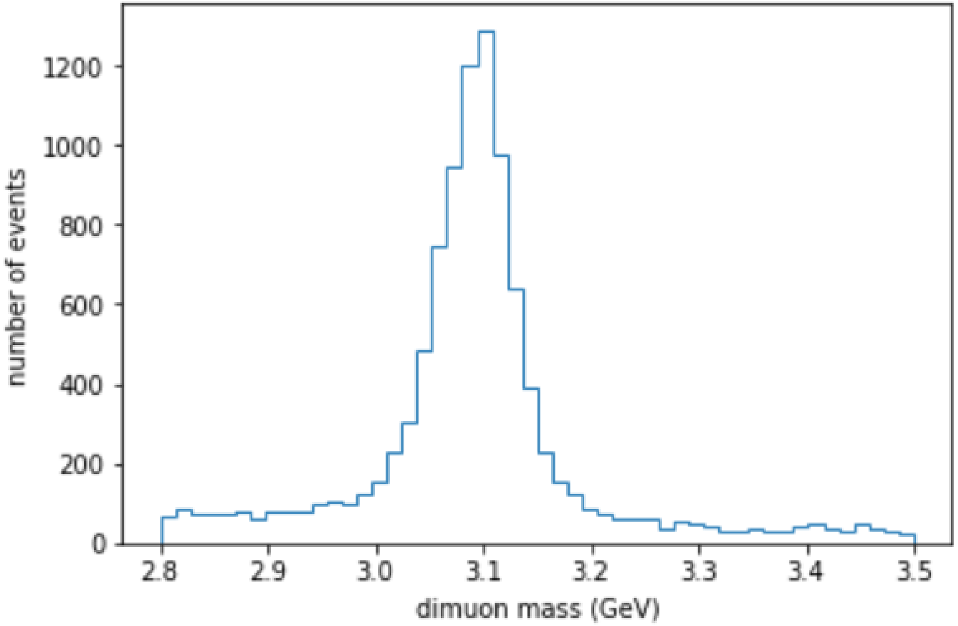
Reconstruction

Your first task is to load the CMS data file! Each data element has 8 pieces of information:

`E1, E2, px1, px2, py1, py2, pz1, pz2`

Think:

- How will you use the 8 pieces of information to calculate the mass of X?
- How can you save only the events with a mass value inside your window?
- How can you calculate the relativistic kinetic energy of particle X?



Day 1: relativistic dynamics

1. Load the workspace
2. Conservation of energy and momentum
3. Apply $E^2 = p^2 + m^2$ (c = 1 is their "treat" for the exercise)
4. Assign each group a "mass window" surrounding a known meson that decays to muons (J/ψ, Upsilon, Z boson)
5. Select events with dimuon mass in the correct window
6. Store mass and kinetic energy of events in the window
7. Make histograms of both quantities with error bars
8. Save a workspace for next time

Want more depth?

- Have them explore a wide mass range and choose a window

Short on time?

- Remove kinetic energy elements
- Determine mass of the particle by eye from the histogram – Google search!
- End with "what did you discover?"

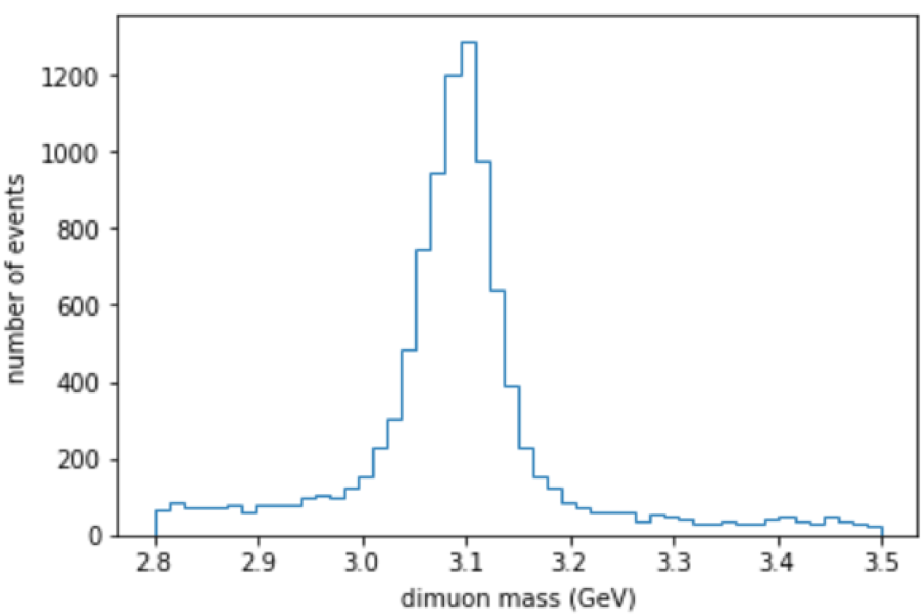
Background fitting

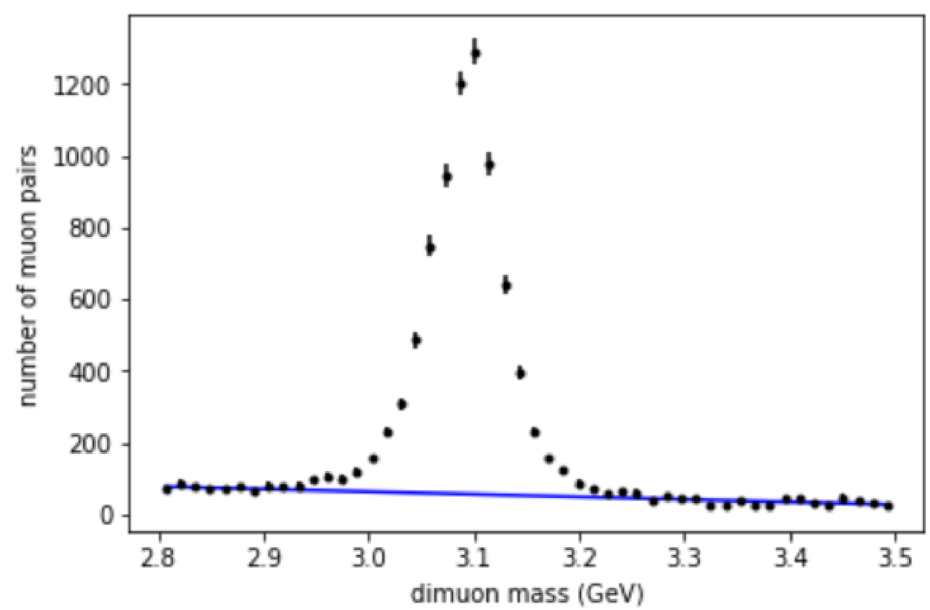
THINK:

Which type of curve do you expect will match your data best? Imagine a curve connecting the two sides under your peak.

EVALUATE your fit:

- Plotting: does the shape make any sense?
- Chi^2 is defined in Eq. 29. It describes the difference between the points and the fitted curve. points.
- OPTIMALLY, Chi^2 / (# points - # parameters) is around 1





Day 2: background subtraction

1. Determine “sideband” regions around the peak for fitting
2. Perform polynomial fits to just the background
3. Choose the best polynomial order by using a goodness-of-fit metric
4. Evaluate the polynomial throughout the whole mass range
5. Plot the function on top of the data
6. Subtract this background estimate from the data
7. Plot the new signal mass distribution, propagating error!

Want more depth?

- Explore other functions (using curve fitter from day 3)
- Edit the fitter to return the full covariance matrix and propagate error on background

Short on time?

- Remove goodness-of-fit evaluation
- Slim error propagation to suit your time needs – errors bars optional!

Signal characterization

THINK:

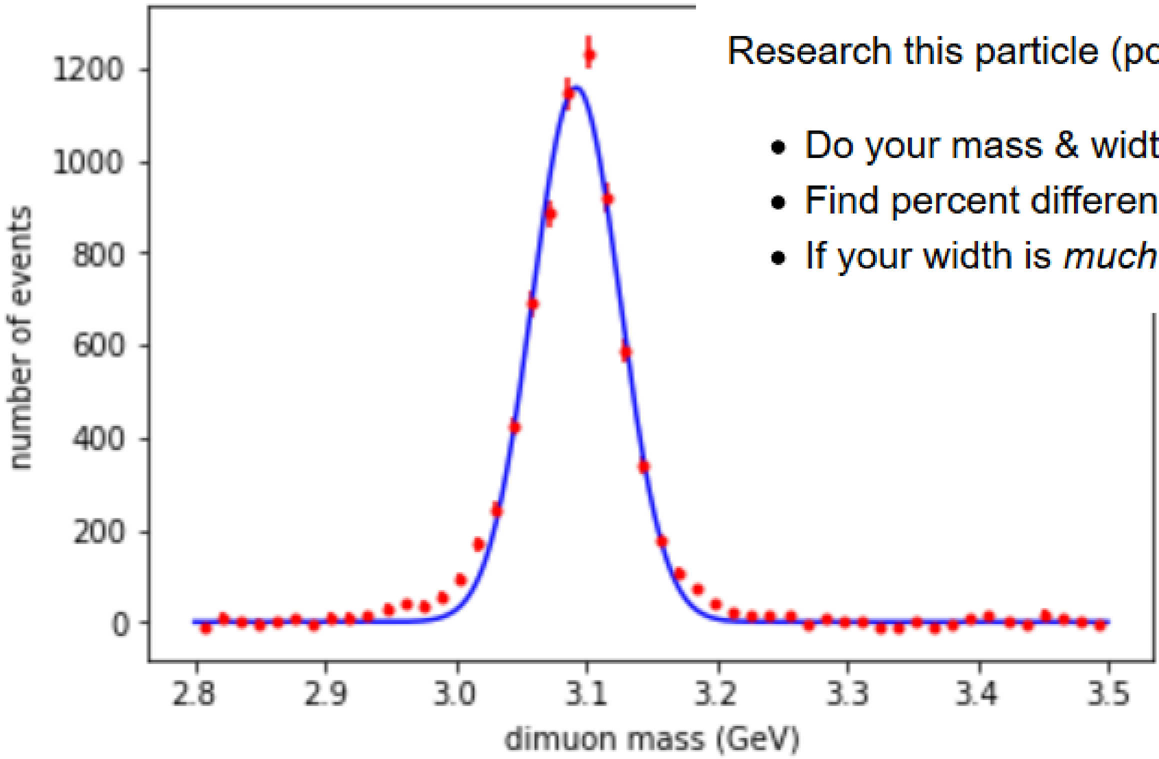
Which statistical distribution describes your signal peak?

THINK:

- Can you statistically distinguish signal from background?
- Can you find this particle with a web search for you mass?

Research this particle (pdg.gov), find its width (capital Gamma).

- Do your mass & width agree with the known values?
- Find percent differences and also discrepancy/significance.
- If your width is *much* larger than accepted, why might this be?



Day 3: characterization

1. Fit a Gaussian to the signal mass distribution
2. Determine mean, width, and their uncertainties
3. Associate mean and width to physical quantities (mass and lifetime)
4. Determine which particle this is!
   - Do the mass and width agree within uncertainties with known quantities?
   - For Z boson: yes. For light mesons: mass will agree and width will be several orders of magnitude too large because of the detector’s momentum resolution.
5. Write up their report

Short on time?

- Gaussian fit is only needed for width and uncertainty – can “eyeball” these
- Don’t ask about known width

Want more depth?

- Evaluate number of signal & background events by integration

Bring open data to your lab!

BETHEL UNIVERSITY

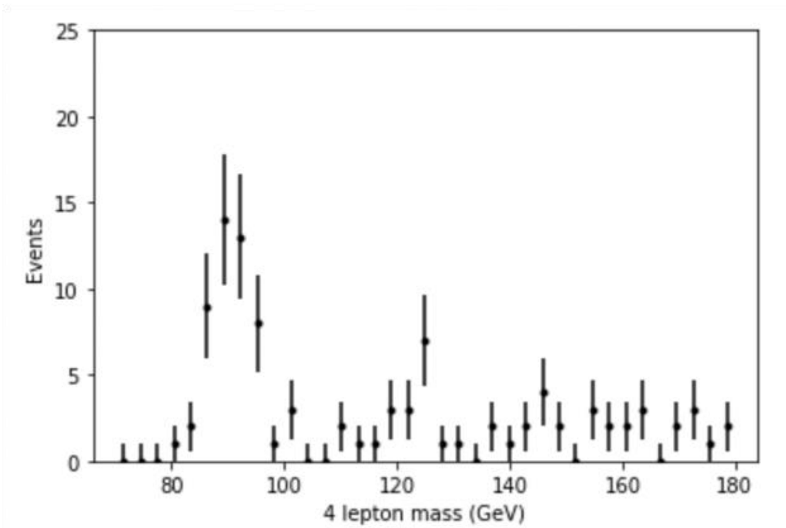
- Data is free for public use and all materials for this lab are on Github
- + expansions!

Higgs Expansion

Sophomore Level Version

Output text files

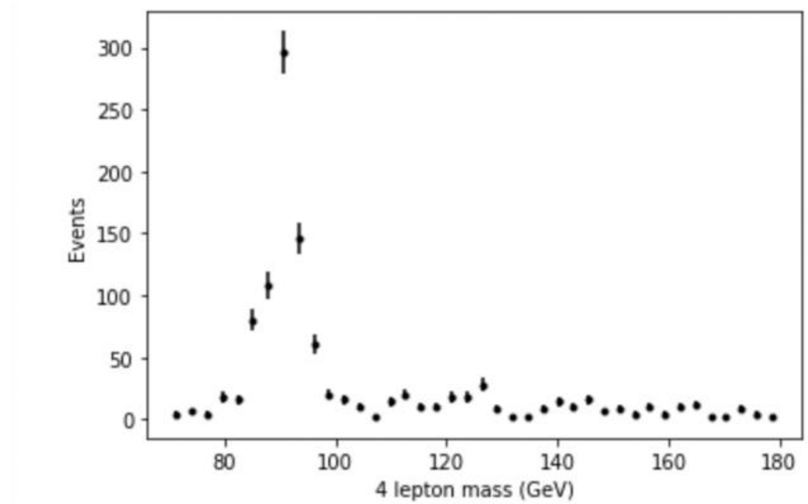
Used to make plots directly



Senior Level Version

Output Root files

Must be further analyzed to produce a significant Higgs peak



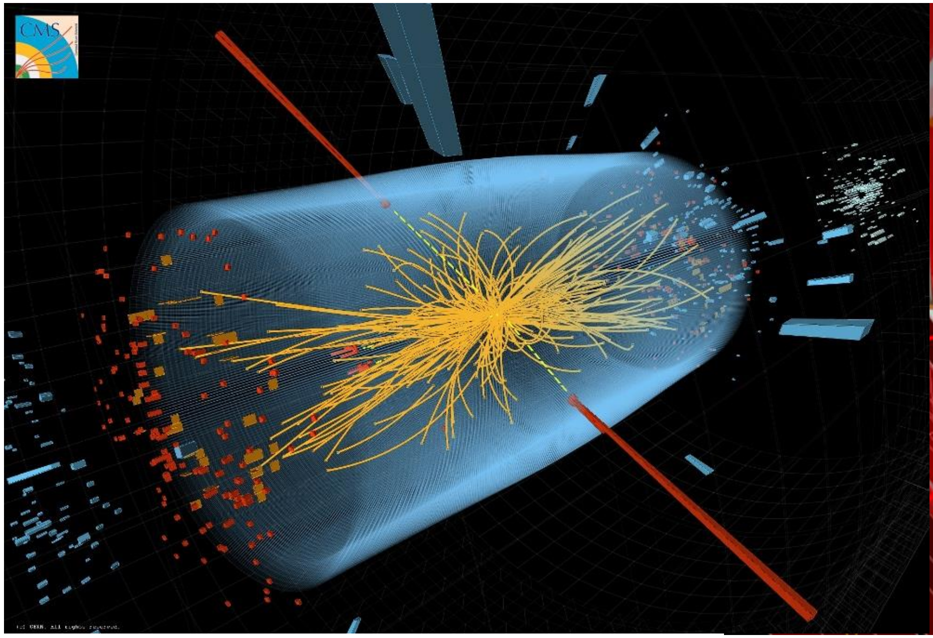
Assessment

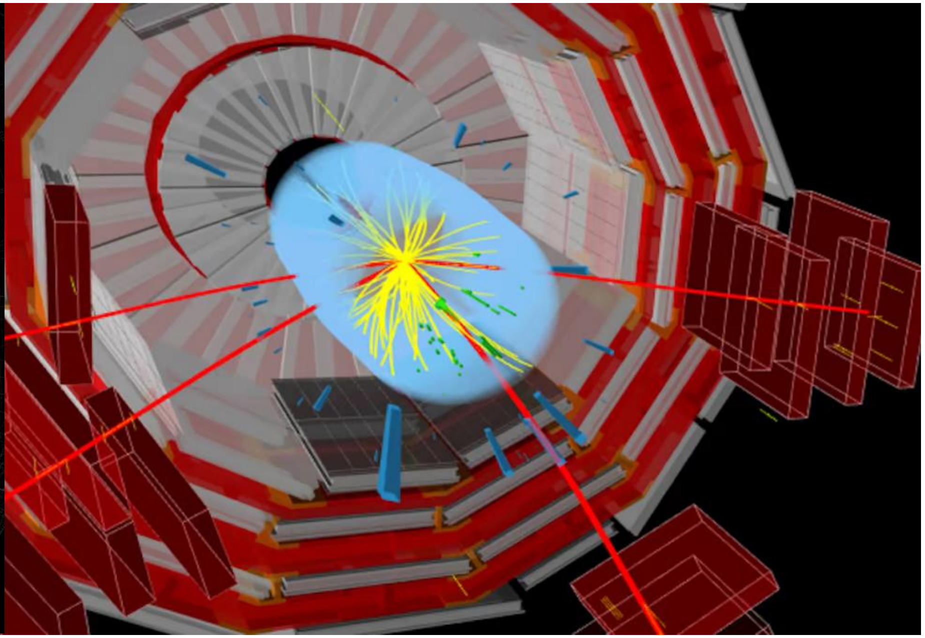
- Does this help them understand conservation? Particle decays?
- Does this help them understand histogram structure?
- N = 25 pre-survey (clickers) and N = 28 post-survey (final exam)

- Strong results the first time (80-90% correct):
  - understanding how histogram bins group data
  - understanding not ALL events in a peak are from the same physics process
  - understanding root cause of empty bins
- Growth after the lab:
  - Stating what the Poisson error bar on bin contents means (moderate growth)
  - Choosing the correct relativistic E-p conservation formula (moderate growth)
  - Seeing large peak as evidence for 1 particle whose mass is subject to the uncertainty principle (strong growth)

High school outreach with VR

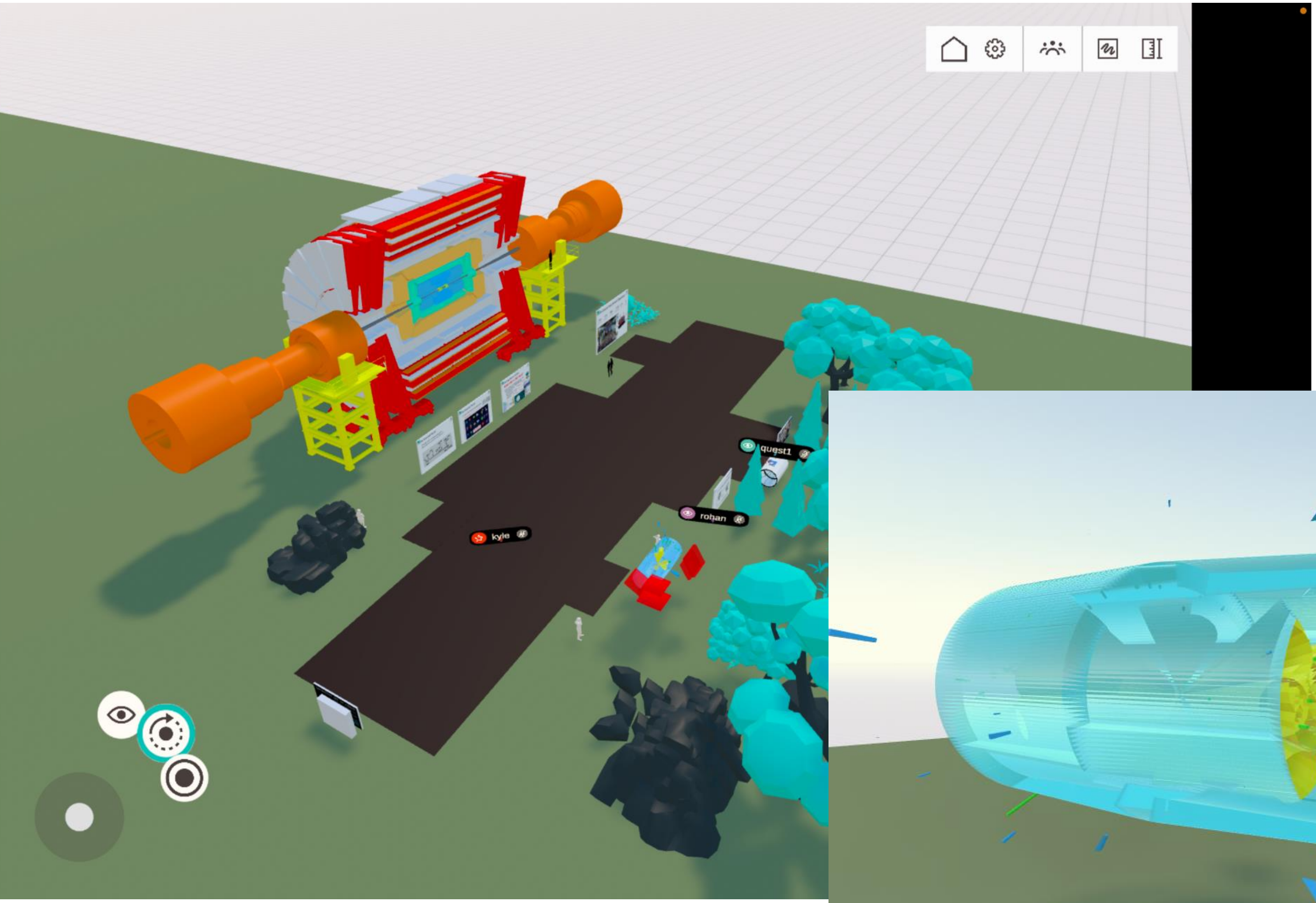
- Virtual reality "world" allows participants with any electronic device to explore the CMS detector!
- Arkio is a cross-platform 3D viewing software in which we've built two worlds using the model export capabilities of the iSpy event displays
- Check it out! https://www.arkio.is/download/ Room: odws Pass: 1234

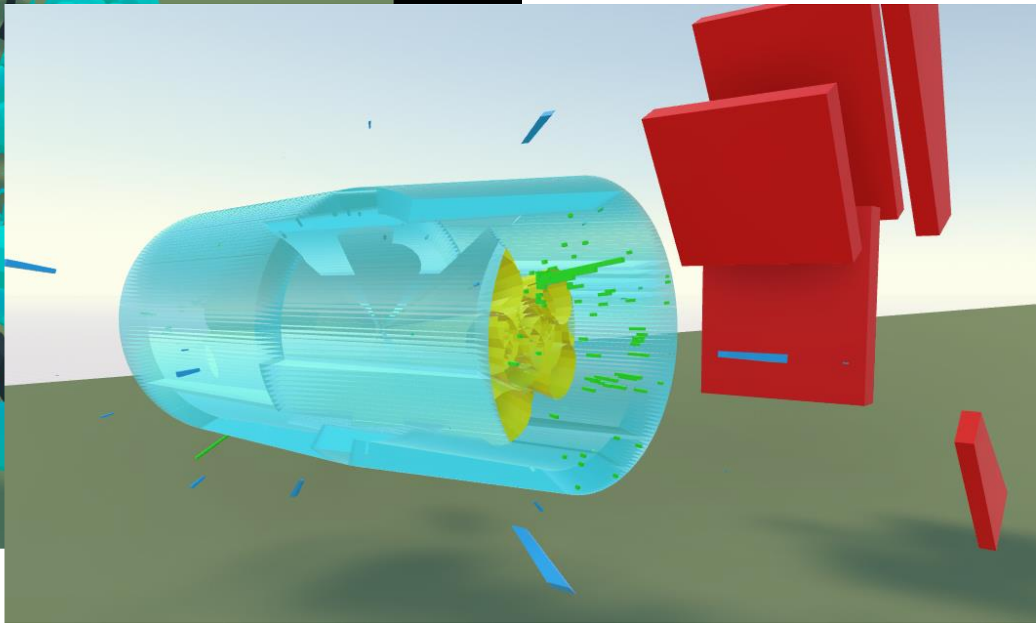




LEGO Particle Flow 2022

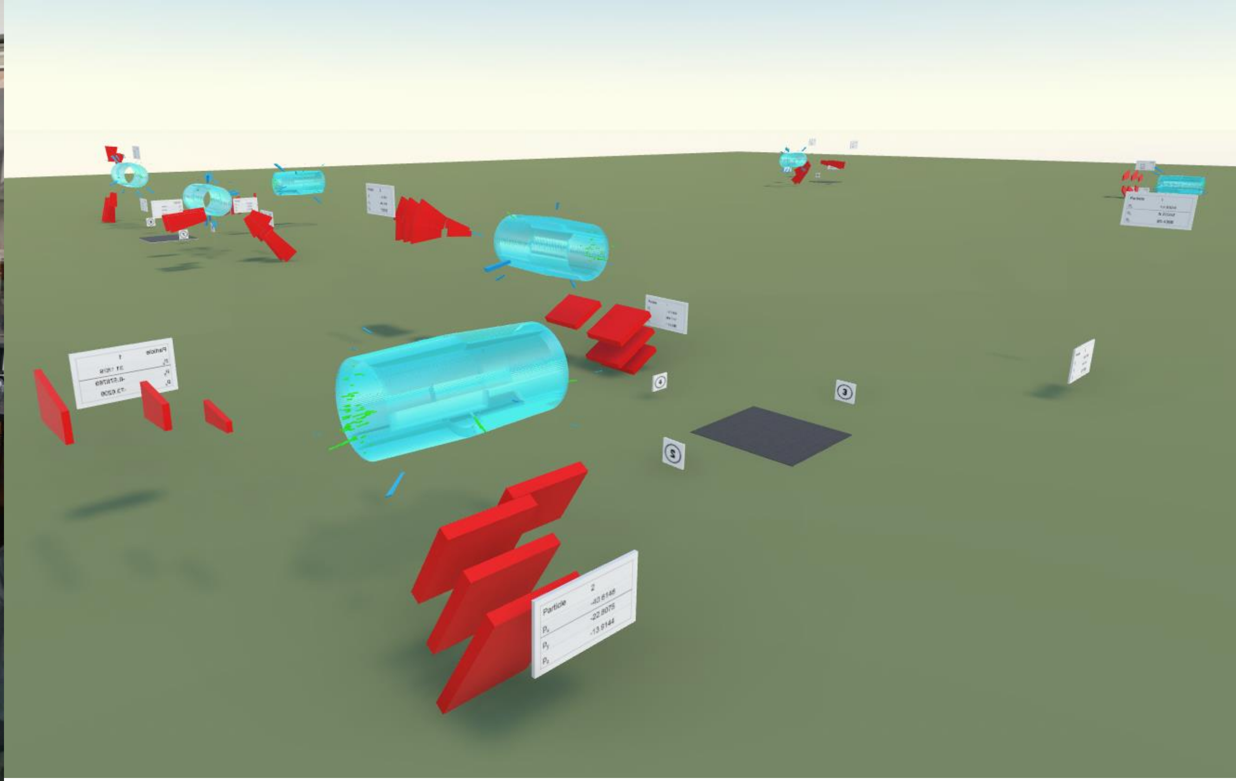
Aug 2022: we supplemented the LPC LEGO Particle Flow tutorial with a look inside the VR detector, and a peak at the event students would reconstruct





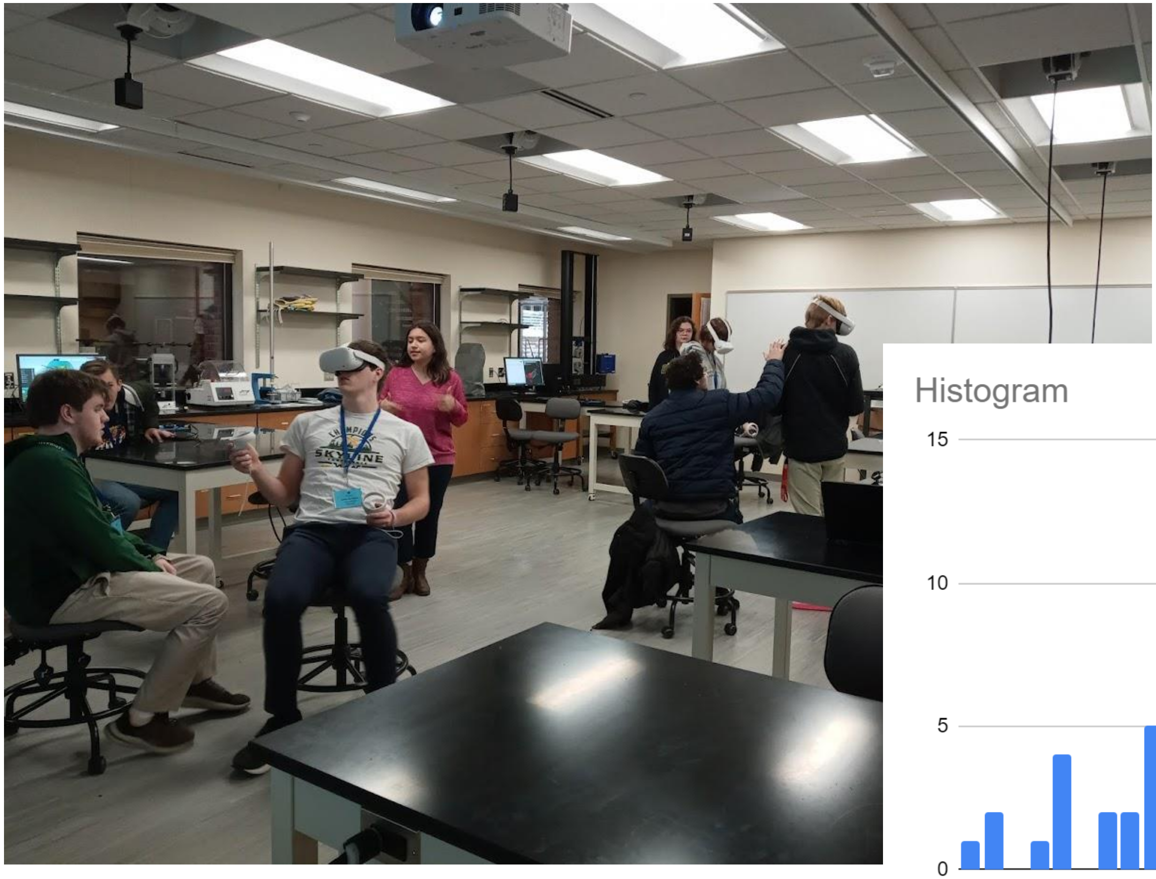
High school outreach with VR

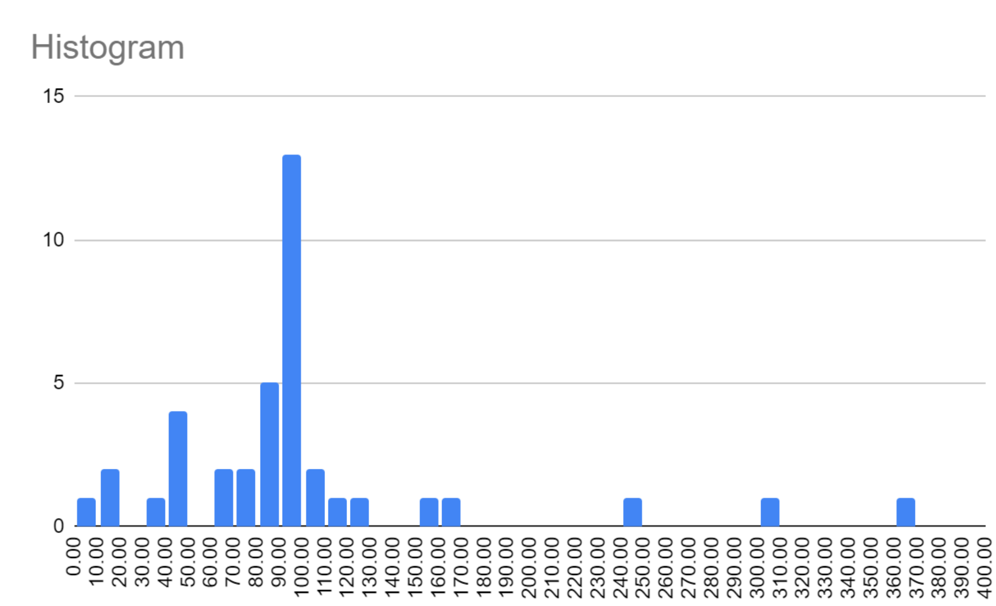
Jan 2023: we hosted high school physics students for a "CMS Masterclass" utilizing an expanded version of the Arkio exhibition. Several "pods" of dimuon Z boson events were created from which students could extract momentum and energy information.



High school outreach with VR

Following the Masterclass format, students also analyzed many events iSpy and all their computed invariant masses were collated into a histogram for discussion and discovery





Open Data Portal

- Education page: https://opendata.cern.ch/docs/cms-guide-for-education
- Displays: https://opendata.cern.ch/visualise/events/cms
  - http://cms-outreach.github.io/ispy/
- Higgs -> 4 leptons analysis (long!): https://opendata.cern.ch/record/5500
- Higgs -> tau tau analysis (outreach based): https://github.com/cms-opendata-analyses/HiggsTauTauNanoAODOutreachAnalysis
- Machine learning tutorial: https://github.com/FNALLPC/machine-learning-das
  - Uses OpenData! Could run notebooks with mybinder.org
  - Both fully-connected and convolutional network tutorials
  - “Jet image” optional notebook for “computer vision” type processing