High Throughput Grid Computing and High Energy Particle Physics

ASP2016

V. Of Rwanda, Kigali, Rwanda

August 18, 2016

Dr. Jaehoon Yu

Department of Physics

University of Texas at Arlington

Outline

- Introduction
- The problem
- A solution using the Computing Grid
- Performance of the Grid
- Conclusions

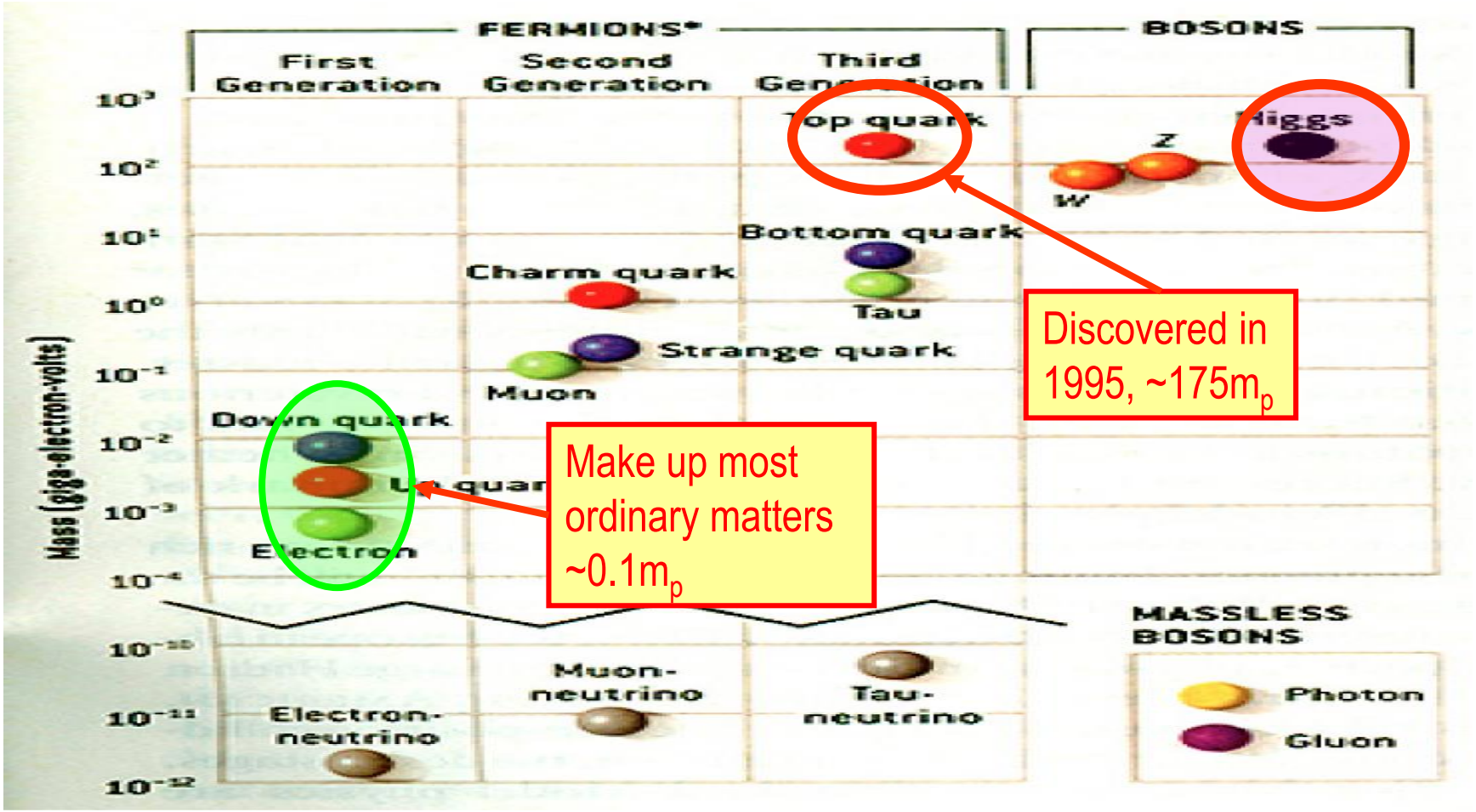
We always wonder…

- What makes up the universe?
- How does the universe work?
- What holds the universe together?
- How can we live in the universe well?
- Where do we all come from?

High Energy Physics

- Definition: A field of physics that pursues understanding the fundamental constituents of matter and basic principles of interactions between them.
- Known interactions (forces):
  - Gravitational Force
  - Electromagnetic Force
  - Weak Nuclear Force
  - Strong Nuclear Force
- Current theory: The Standard Model of Particle Physics (SU3xSU2XU1)

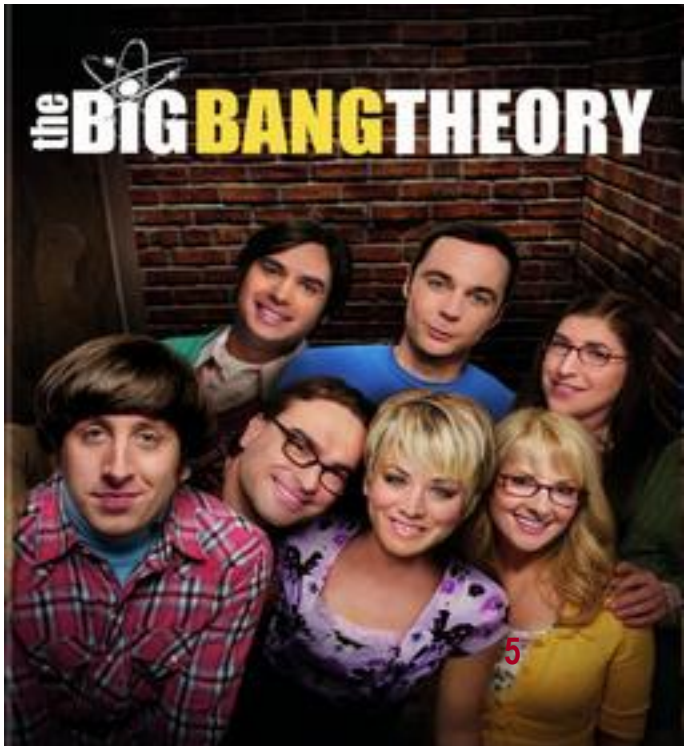
HEP and the Standard Model



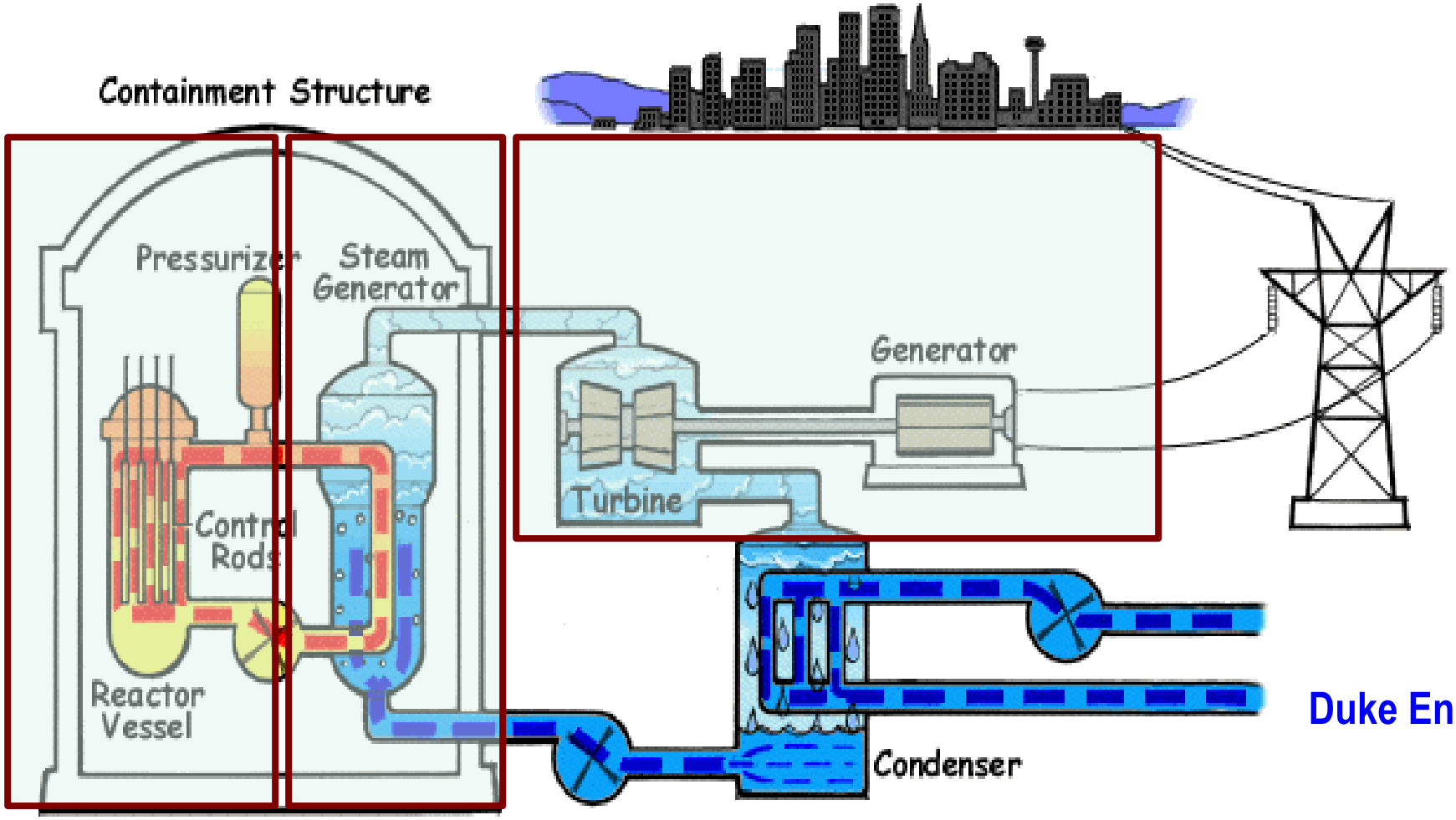
- Total of 16 particles (12+4 force mediators) make up all the visible matter in the universe! ➔ Simple and elegant!!!
- Tested to a precision of 1 part per million!

So what's the problem?

- Why is the mass range so large ($0.1m_p$ – $175\ m_p$)?
- Is the new particle we've discovered really the Higgs particle?
- Why is the matter in the universe made only of particles?
- Neutrinos have mass!! What are the mixing parameters, particle-anti particle asymmetry and mass ordering?
- Why are there only four apparent forces?
  - Were they all unified at the Big Bang?



How does a nuclear power plant work?



Duke Energy

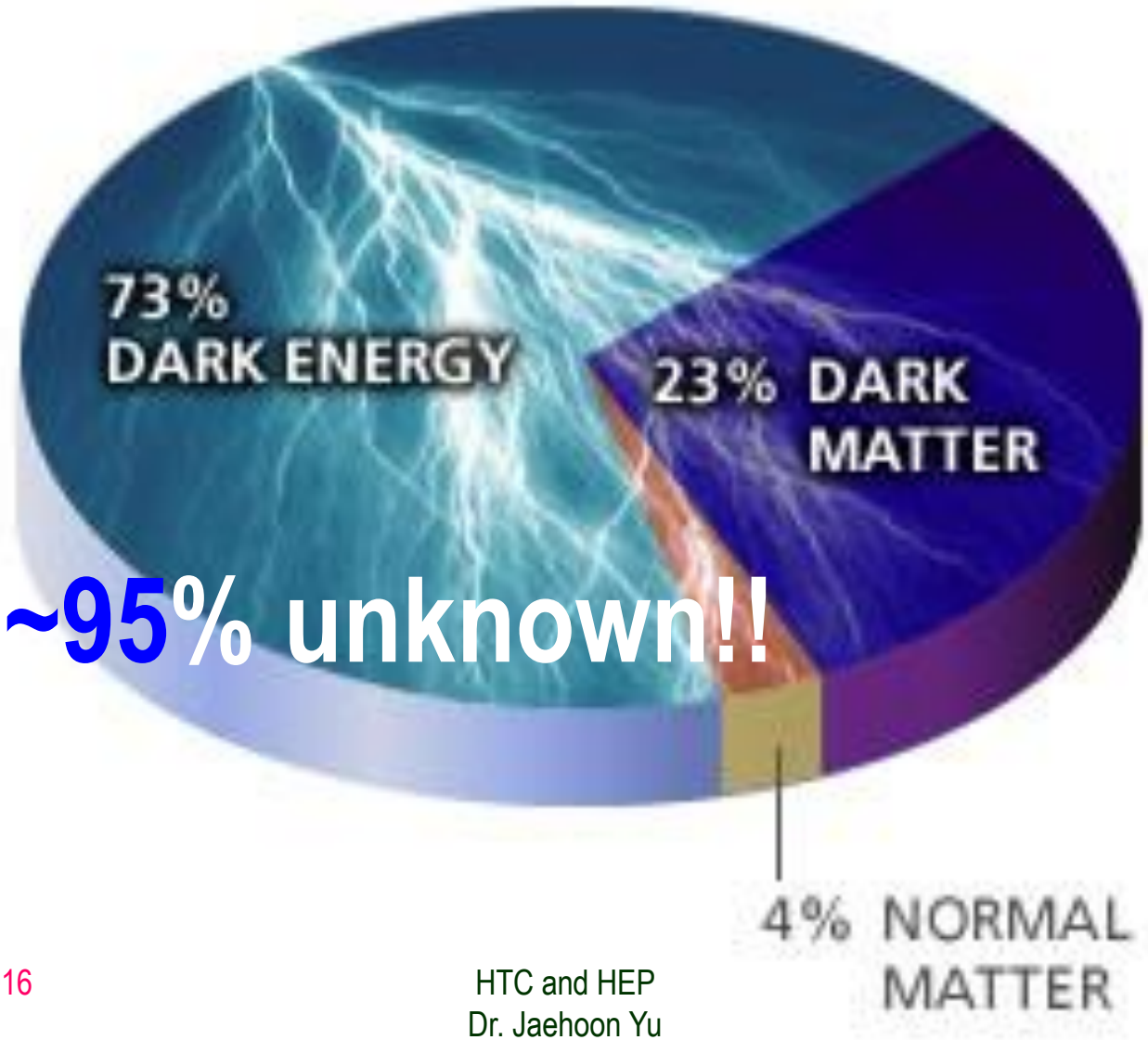
My 1000 year dream: Skip the whole thing!

Make electricity directly from nuclear forces!

So what's the problem?

- Why is the mass range so large ($0.1m_p$ – $175\ m_p$)?
- Is the particle we discovered really the Higgs particle?
- Why is the matter in the universe made only of particles?
- Neutrinos have mass!! What are the mixing parameters, particle-anti particle asymmetry and mass ordering?
- Why are there only four apparent forces?
  - Were they all unified at the Big Bang?
- Is the picture we present the real thing?

What makes up the universe?



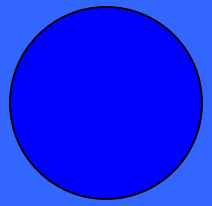
~95% unknown!!

So what's the problem?

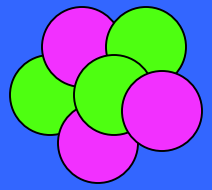
- Why is the mass range so large ($0.1m_p$ – $175\ m_p$)?
- Is the particle we discovered really the Higgs particle?
- Why is the matter in the universe made only of particles?
- Neutrinos have mass!! What are the mixing parameters, particle-anti particle asymmetry and mass ordering?
- Why are there only four apparent forces?
  - Were they all unified at the Big Bang?
- Is the picture we present the real thing?
  - What makes up the remaining ~95% of the universe?
- Are there any other particles we don't know of?
  - Big deal for the ongoing LHC Run and the future ones!
- How can we live well in the universe as an integral partner?

Accelerators are Powerful Microscopes.

They make high energy particle beams
that allow us to see small things.



seen by
low energy beam
(poorer resolution)



seen by
high energy beam
(better resolution)

Accelerators are also Time Machines.

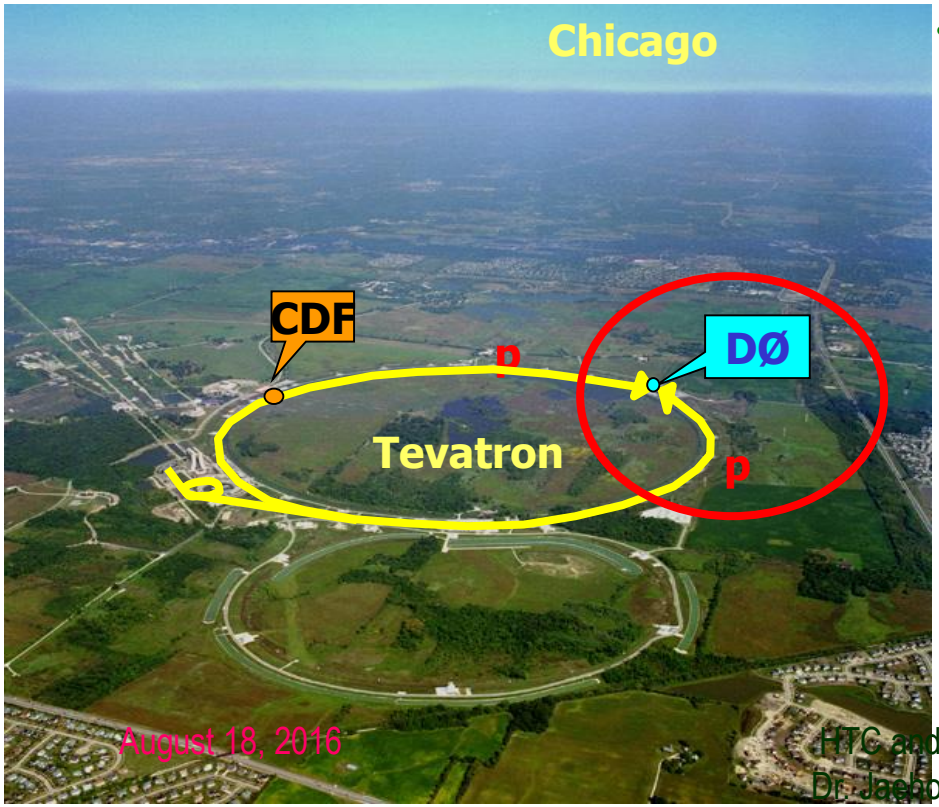
They make particles last seen
in the earliest moments of the universe.

Particle and anti-particle annihilate.

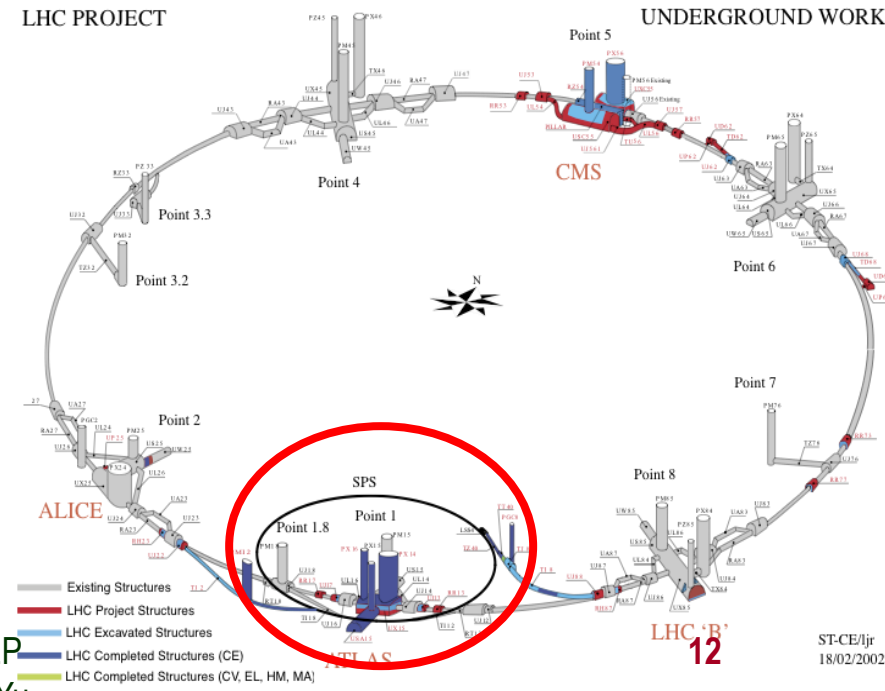
$$E = mc^2$$

Fermilab Tevatron and LHC at CERN

- World's Highest Energy proton-anti-proton collider
  - 4km (2.5mi) circumference
  - E_{cm}=1.96 TeV (=6.3x10^{-7}J/p➔ 13M Joules on the area smaller than 10^{-4}m^2)
  - Equivalent to the kinetic energy of a 20t truck at the speed 130km/hr
    - ~100,000 times the energy density at the ground 0 of the Hiroshima atom bomb
  - **Tevatron was shut down in 2011**
  - **Vibrant other programs running, including the search for dark matter with beams!!**

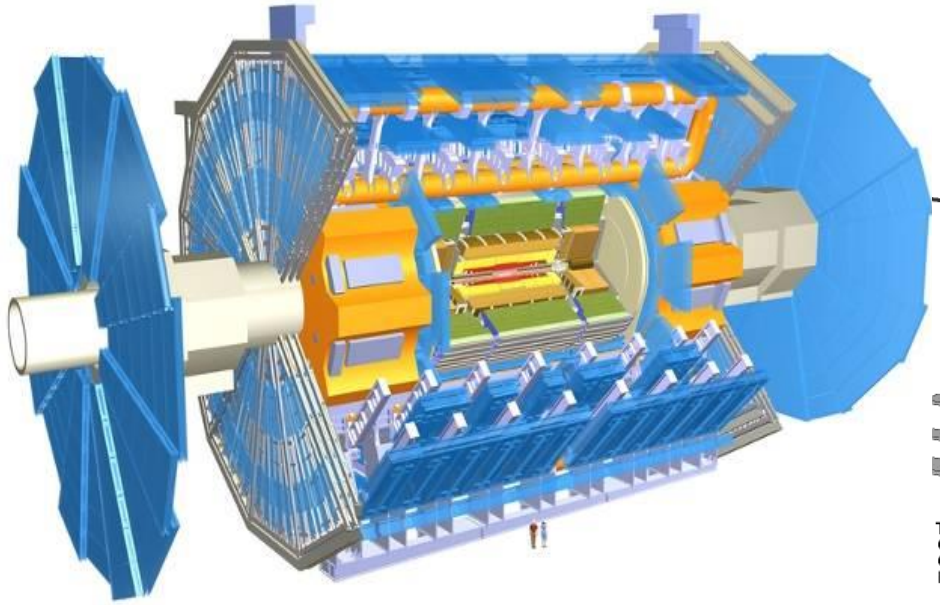


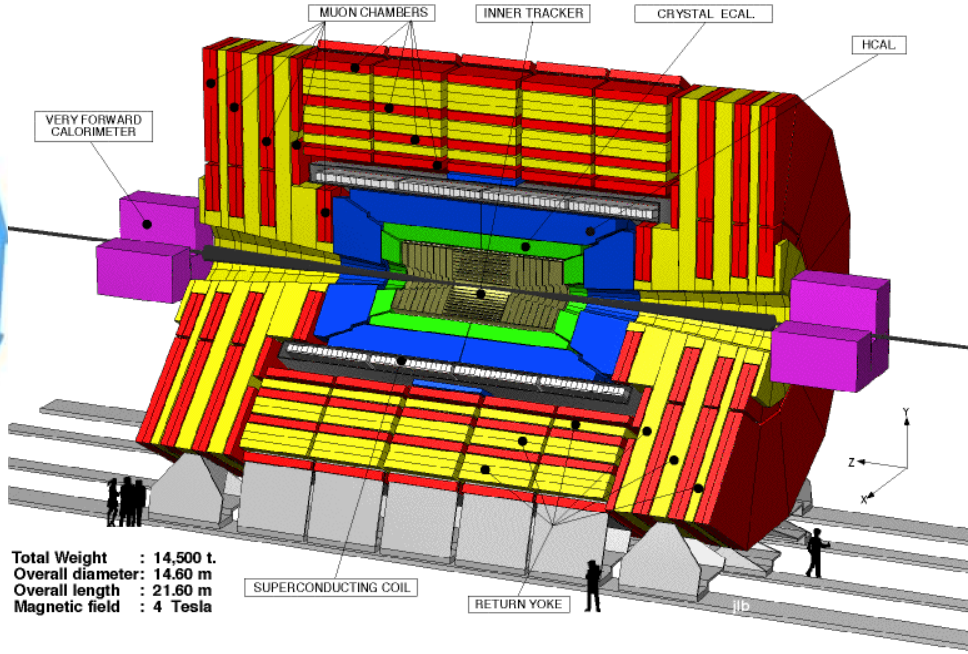
- World's Highest Energy p-p collider
  - 27km (17mi) circumference, 100m (300ft) underground
  - Design E_{cm}=14 TeV (=44x10^{-7}J/p➔ 360M Joules on the area smaller than 10^{-4}m^2)
  - ➢ Equivalent to the kinetic energy of a B727 (80tons) at the speed 310km/hr
    - ➢ ~3M times the energy density at the ground 0 of the Hiroshima atom bomb
- **Large amount of data accumulated in 2010 – 2013**
- **Beam returned after a 2 yr shutdown, data taking ongoing**



LHC @ CERN Aerial View

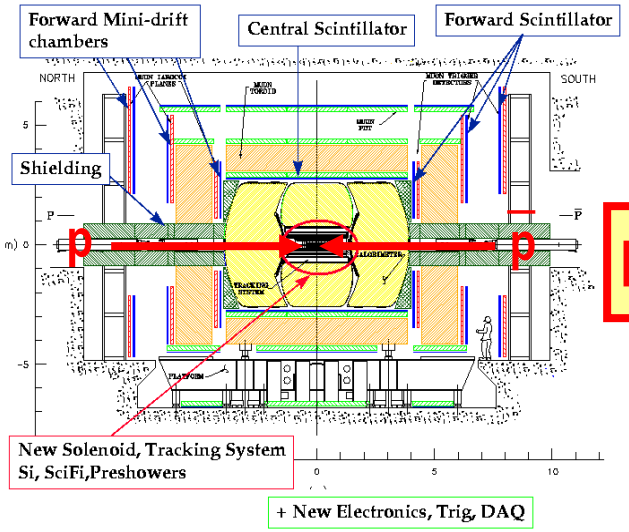
The ATLAS and CMS Detectors





- **Weighs 7000 tons and ~10 story tall**
- Records **200 – 400 collisions/second (out of 50million)**
- Records approximately **350** MB/second
- Records ~2 PB per year ➔ **200*Printed material of the US Lib. of Congress**

200x



Digital data

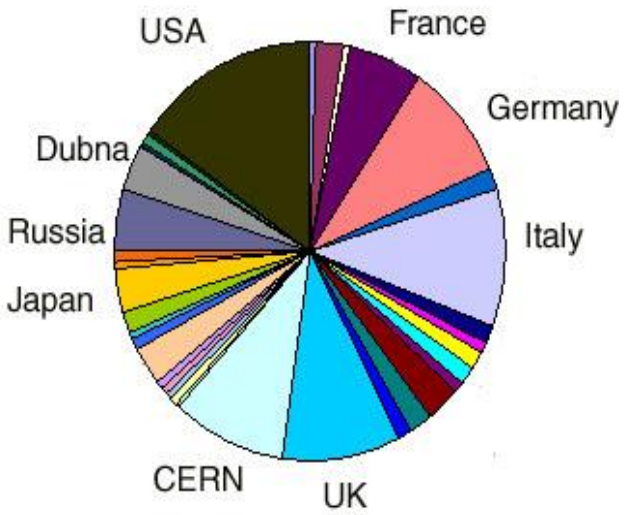
Data Reconstruction

The Problem

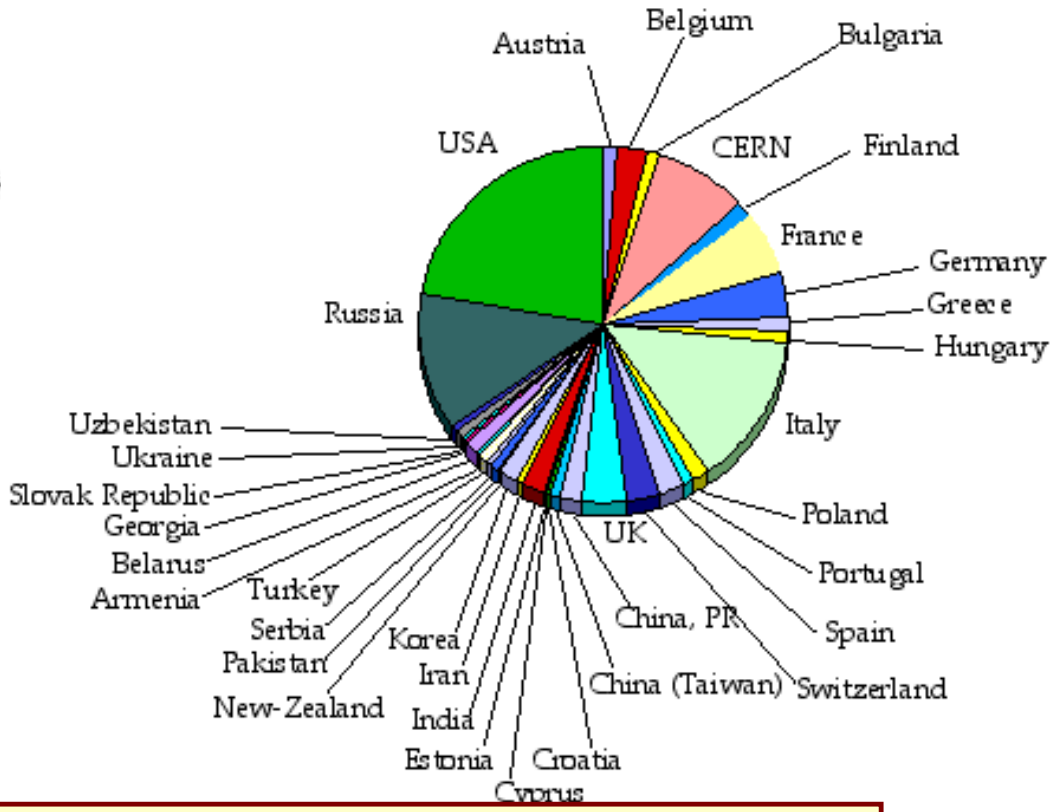
- Detectors are complicated and large ➔ Need large number of collaborators
  - They are scattered all over the world!

HEP Collaborations

ATLAS



CMS

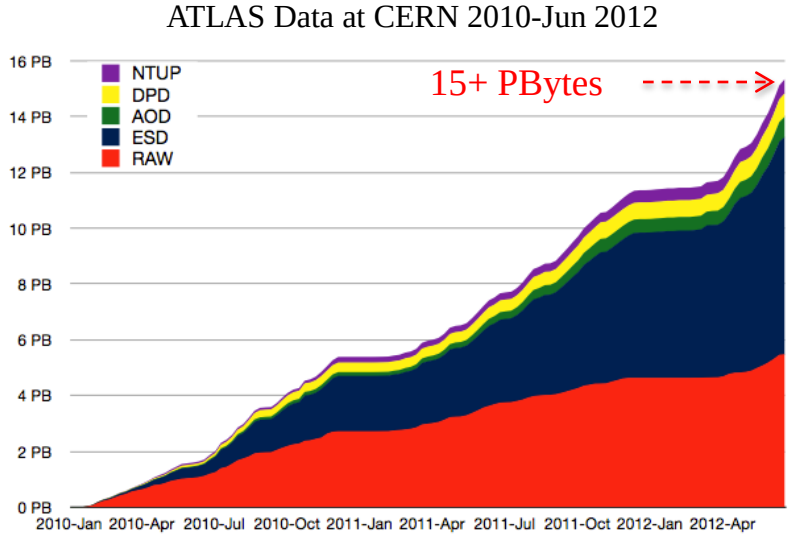


ATLAS+CMS 6000 Physicists and Engineers
Over 60 Countries, 250 Institutions

DUNE Collaboration at Fermilab: 800 collaborators, from 25 countries, 150 instituions

The Problem

- Detectors are complicated and large ➔ Need large number of collaborators
  - They are scattered all over the world!
  - How do we get them communicate quickly and efficiently?
  - How do we leverage collaborators' capabilities?
  - How do we utilize all the computing resources?
- Data size is large ~ 10 PB per year for raw data only
  - Entire data set 15+PB on disc
  - Where and how to store the la
  - How do we allow collaborators to access data in an efficient fa

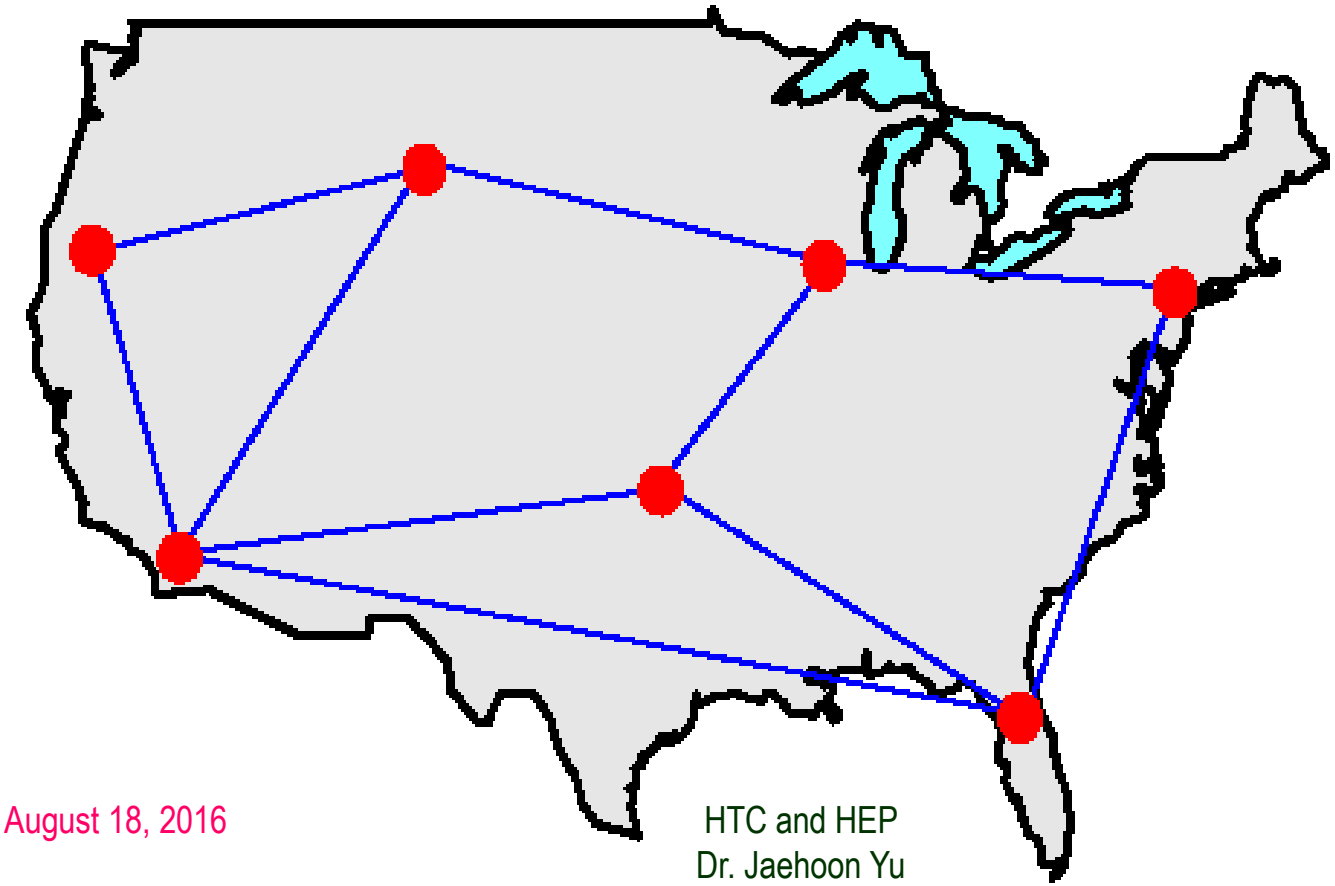


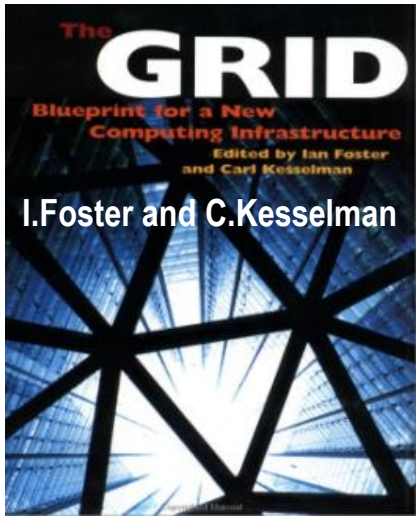
The Problem, cont'd

- How do we allow people's analysis jobs to access data and make progress rapidly and securely?
  - What is the most efficient way to get jobs' requirements matched with resources?
  - Should jobs go to data or data go to jobs?
  - What level of security should there be?
- How do we allow experiments to reconstruct data and generate the large amount of simulated events quickly?
  - How do we garner the necessary compute and storage resources effectively and efficiently?
  - What network capabilities do we need in the world?
- How do we get people to analyze at their desktops?

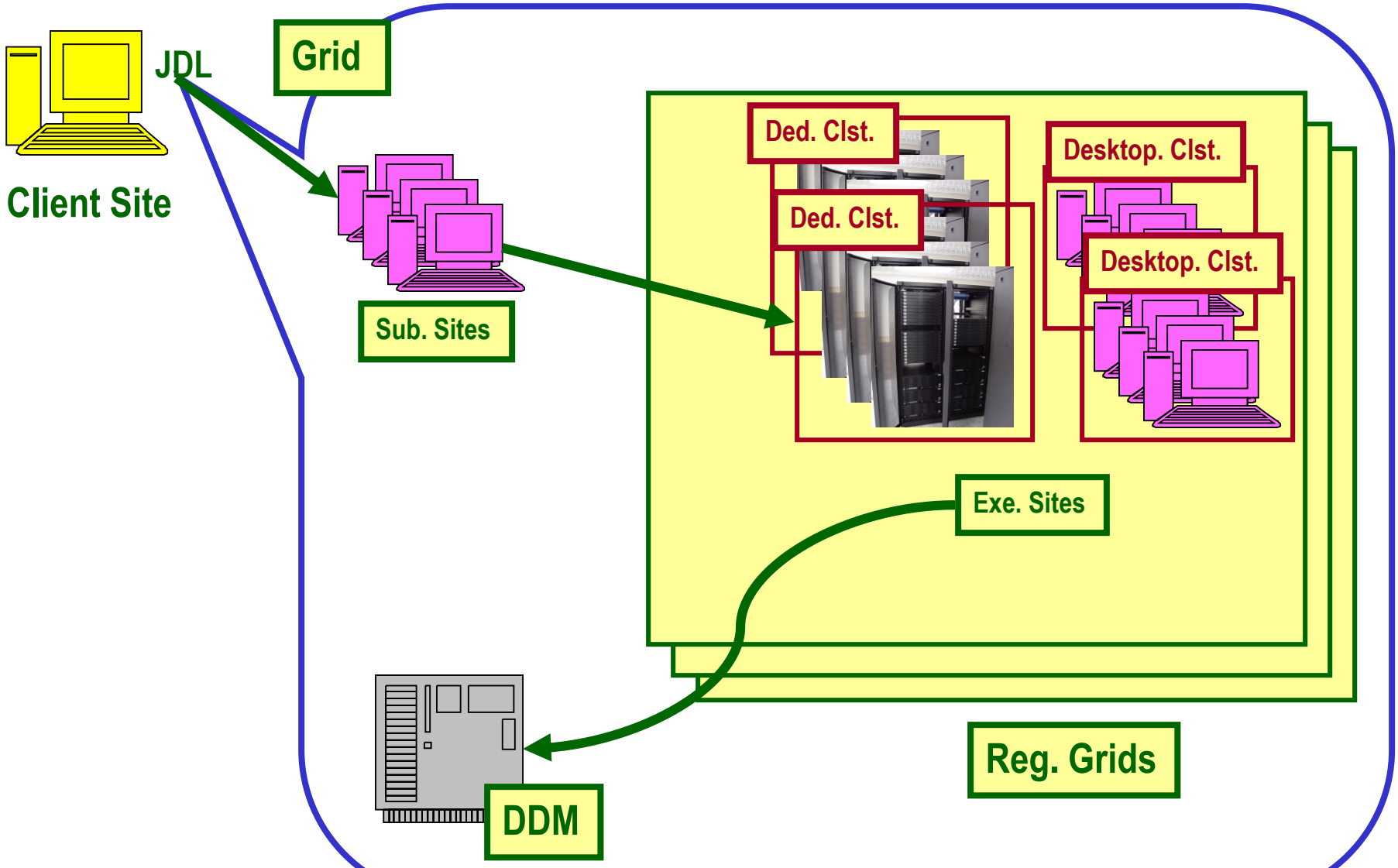
What is a Computing Grid?

- Grid, the definition: Geographically distributed computing resources configured for a coordinated use
- Physical resources & good network provide hardware capability
- The "Middleware" software ties them together

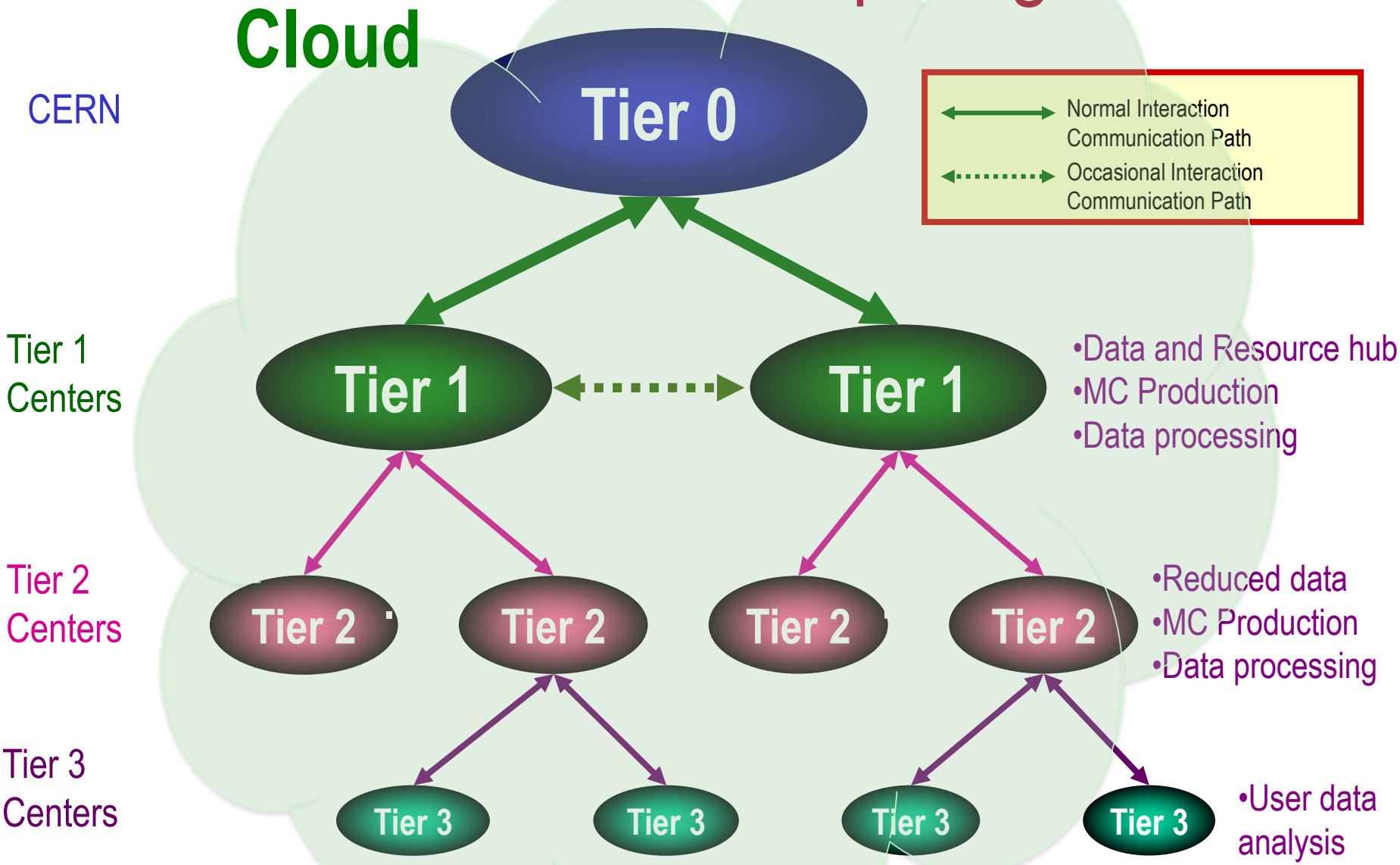




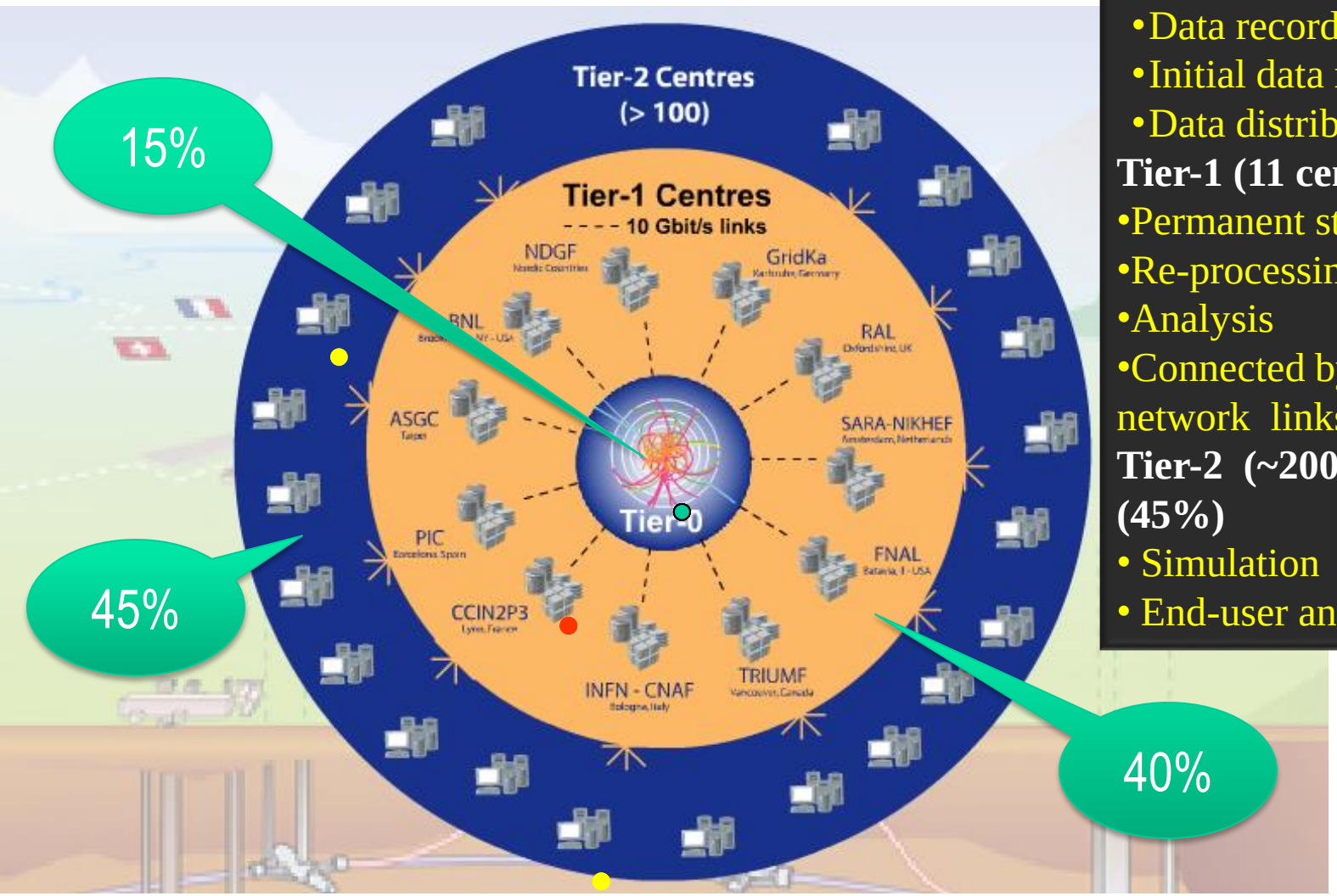
How does a computing Grid work?



Initial Idea of HEP Computing Model



Implemented ATLAS Grid Structure



Tier-0 (CERN): (15%)
- Data recording
- Initial data reconstruction
- Data distribution

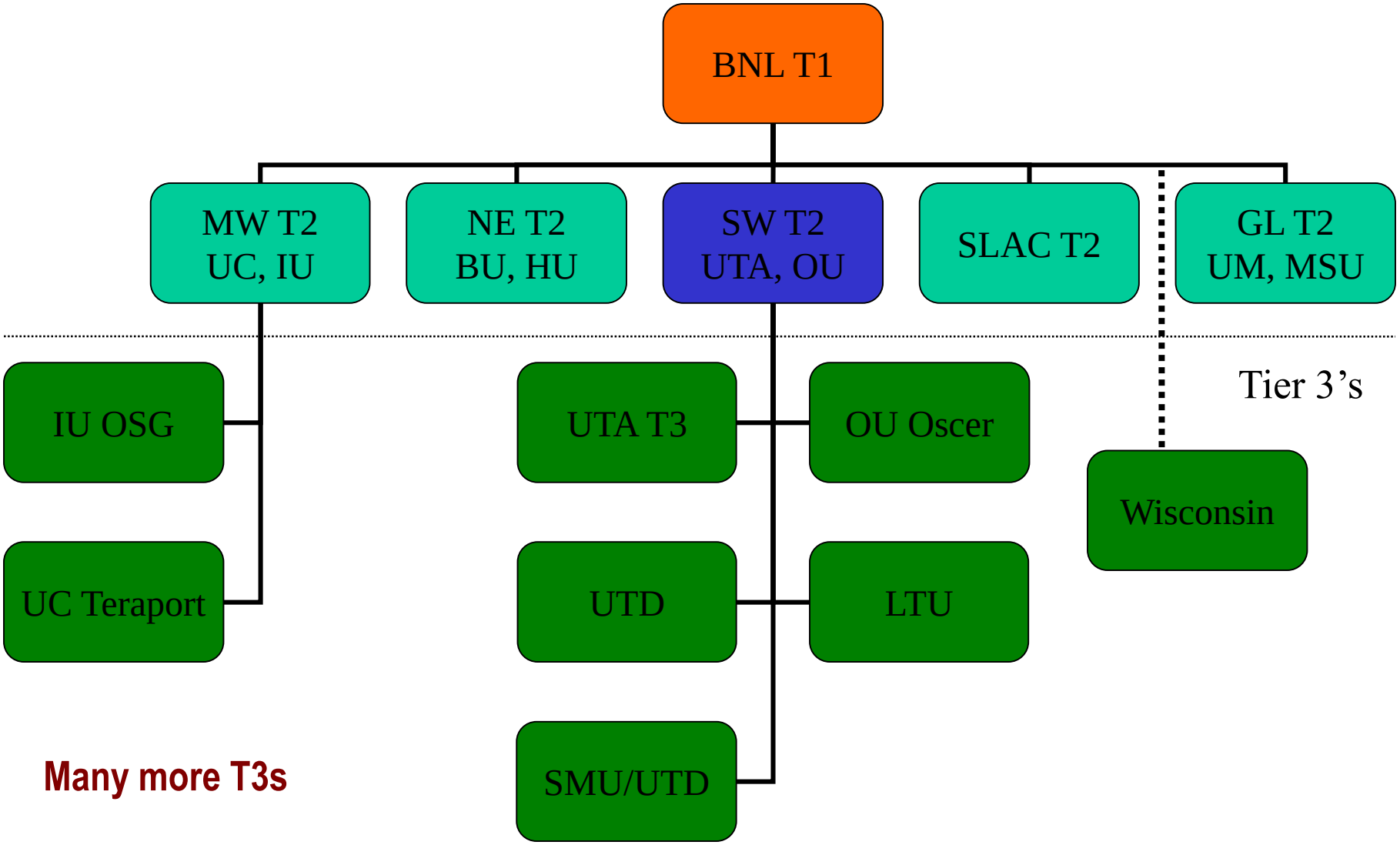
Tier-1 (11 centres): (40%)
- Permanent storage
- Re-processing
- Analysis
- Connected by direct 10 Gb/s network links

Tier-2 (~200 centres): (45%)
- Simulation
- End-user analysis

NORDUGRID
Grid Solution for Wide Area Computing and Data Handling

Open Science Grid

Tiered Example – US Cloud

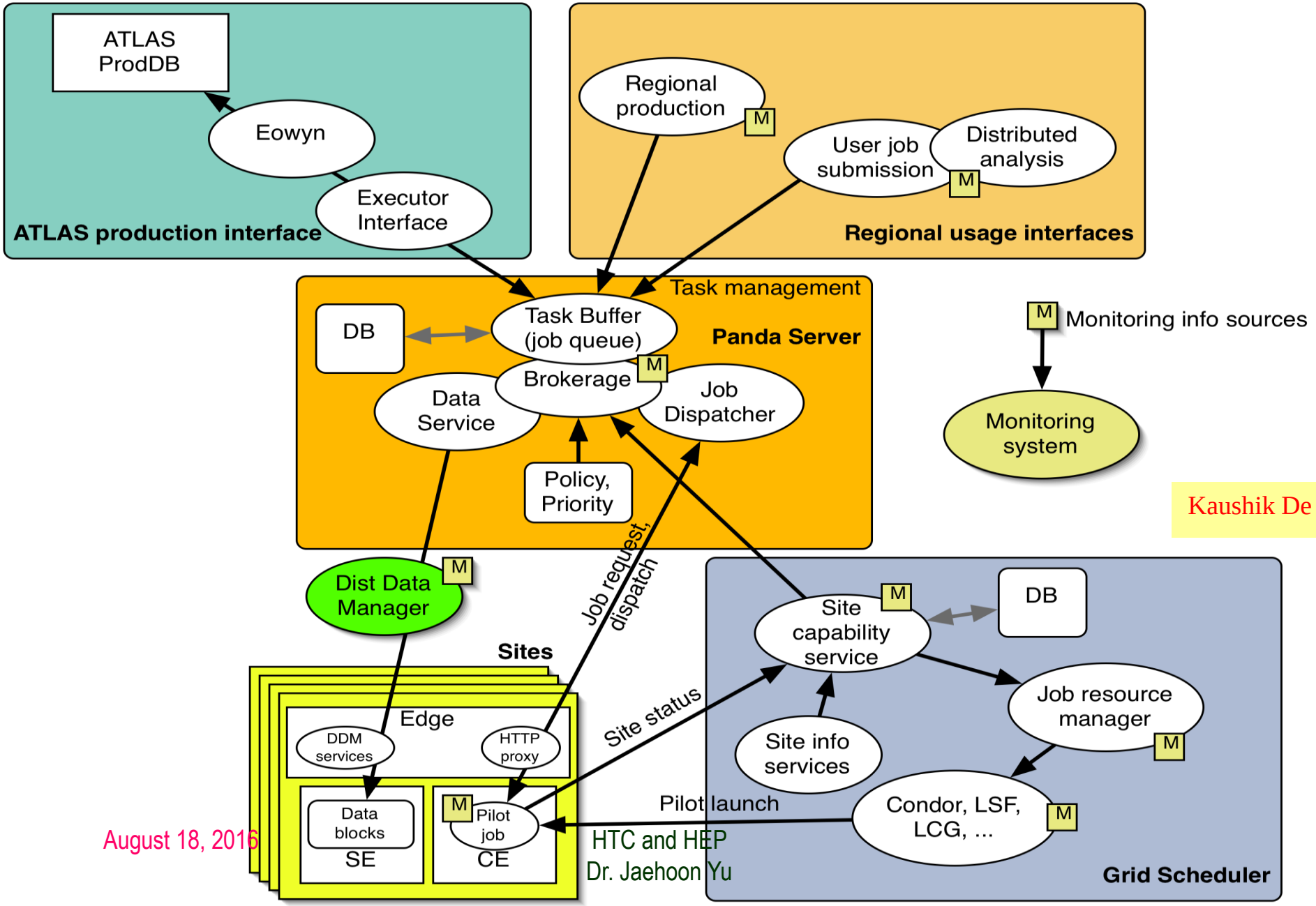


Many more T3s

ATLAS **P**roduction **an**d **D**istributed **A**nalysis System, Panda

- Designed for analysis as well as production
- Works with OSG, EGEE/LCG and others
- A single task queue and pilots
  - Apache-based Central Server
  - Pilots retrieve jobs from the server as soon as CPU is available low latency
- Highly automated with an integrated monitoring system
- Requires low operation manpower
- Integrated with ATLAS Distributed Data Management (DDM) system
- Not exclusively ATLAS and has spread throughout many different entities

ATLAS Panda Architecture

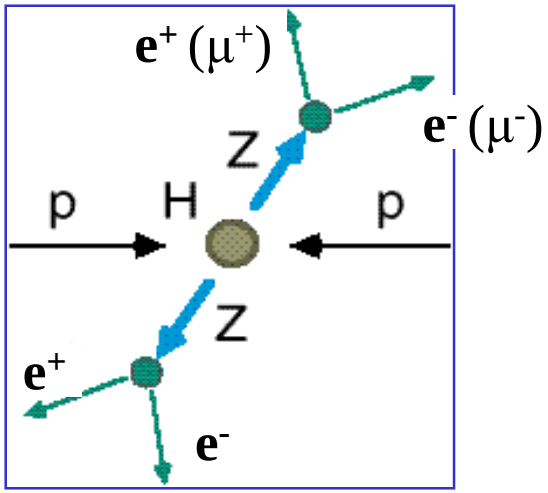


How do we look for the Higgs?

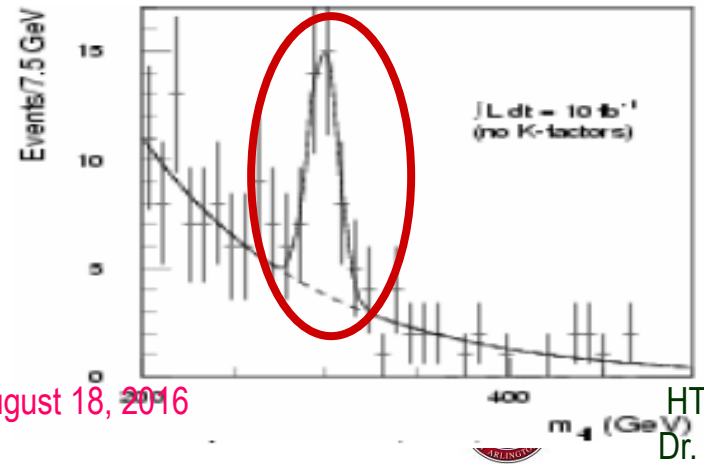
- Higgs particle is so heavy they decay into other lighter particles instantaneously
- When one searches for new particles, one looks for the easiest way to get at them
- Of many signatures of the Higgs, some are much easier to find, if it were the Standard Model Higgs
  - H→ γγ
  - H → ZZ* → 4e, 4⌠, 2e2⌠, 2e2⌡ and 2⌠2⌡
  - H→ WW*→2e2⌡ and 2⌠2⌡
  - And many more complicated signatures

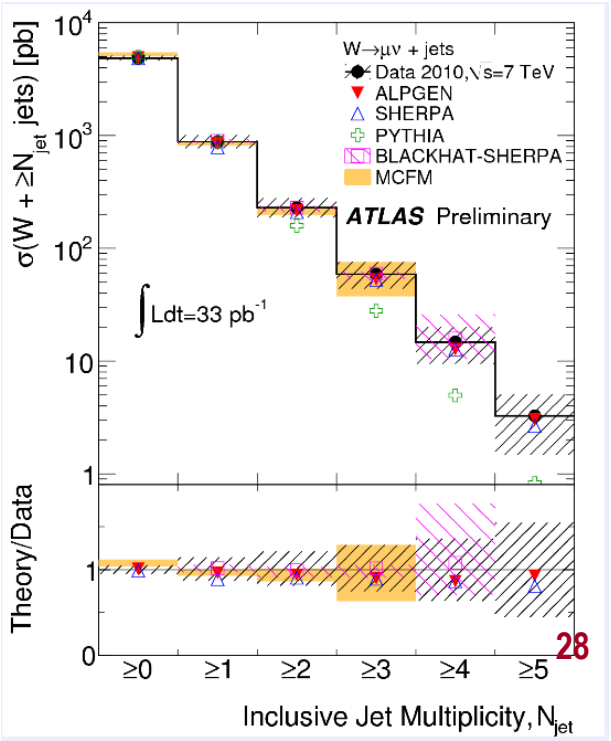
How do we look for the Higgs?

- Identify Higgs candidate events

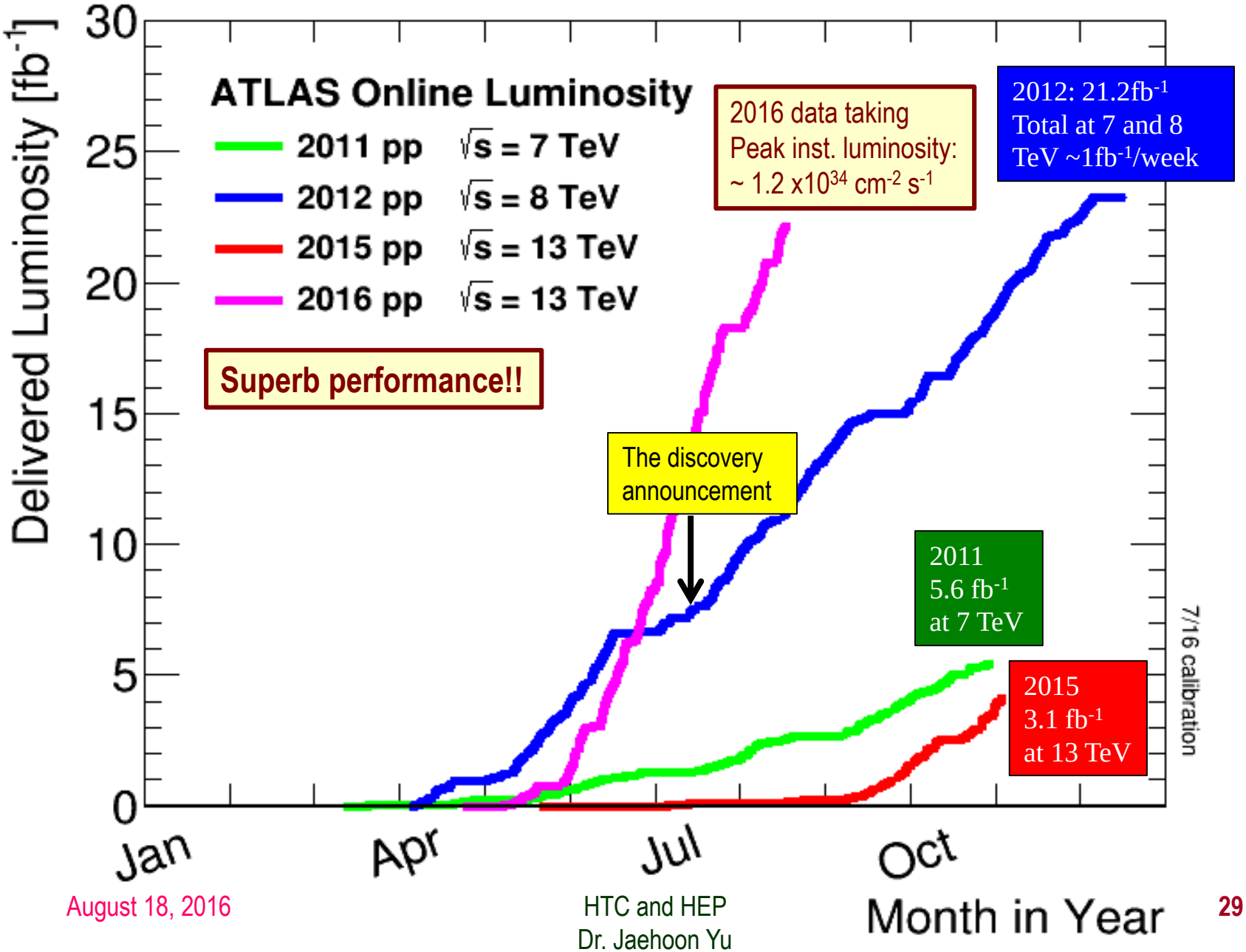


- Understand fakes (backgrounds)
- Look for a bump!!
  - Large amount of data absolutely critical

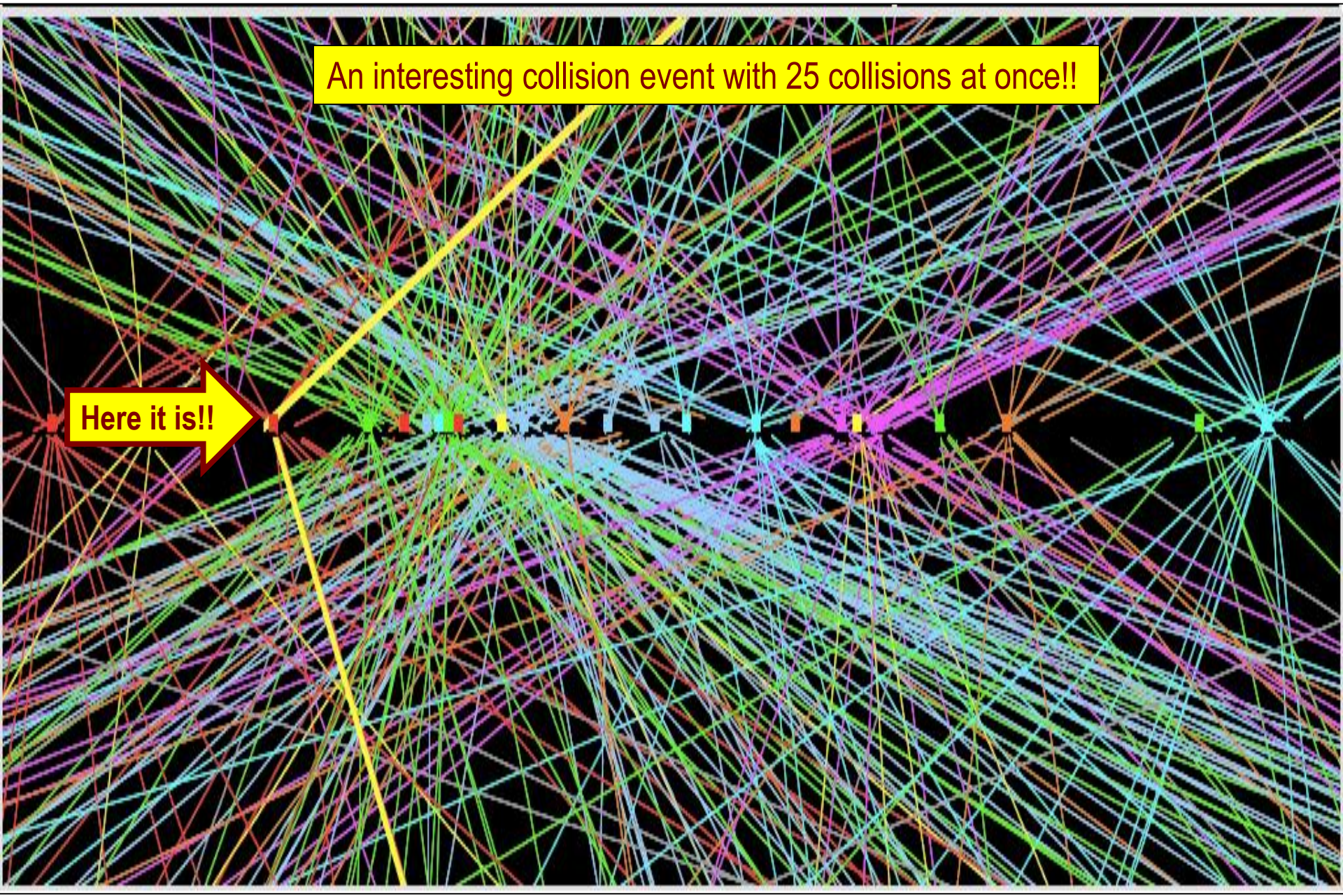




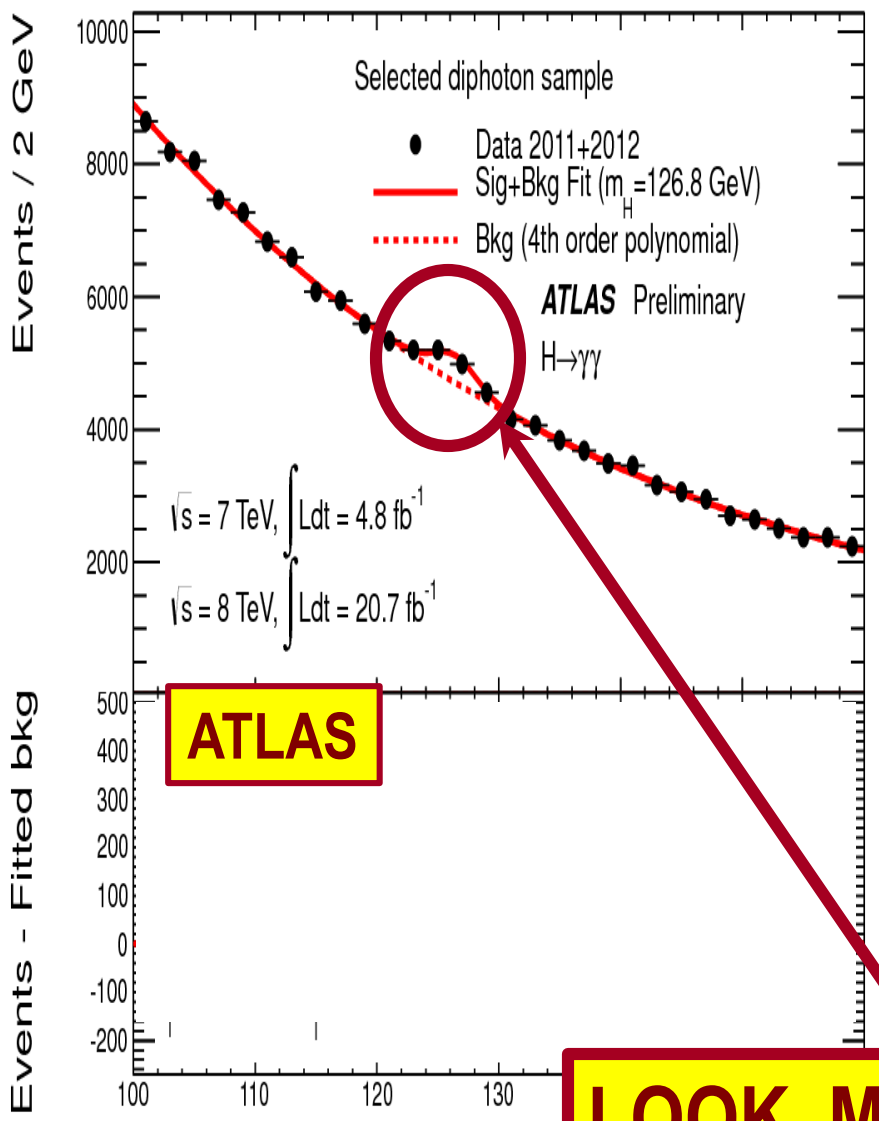
Amount of LHC Data

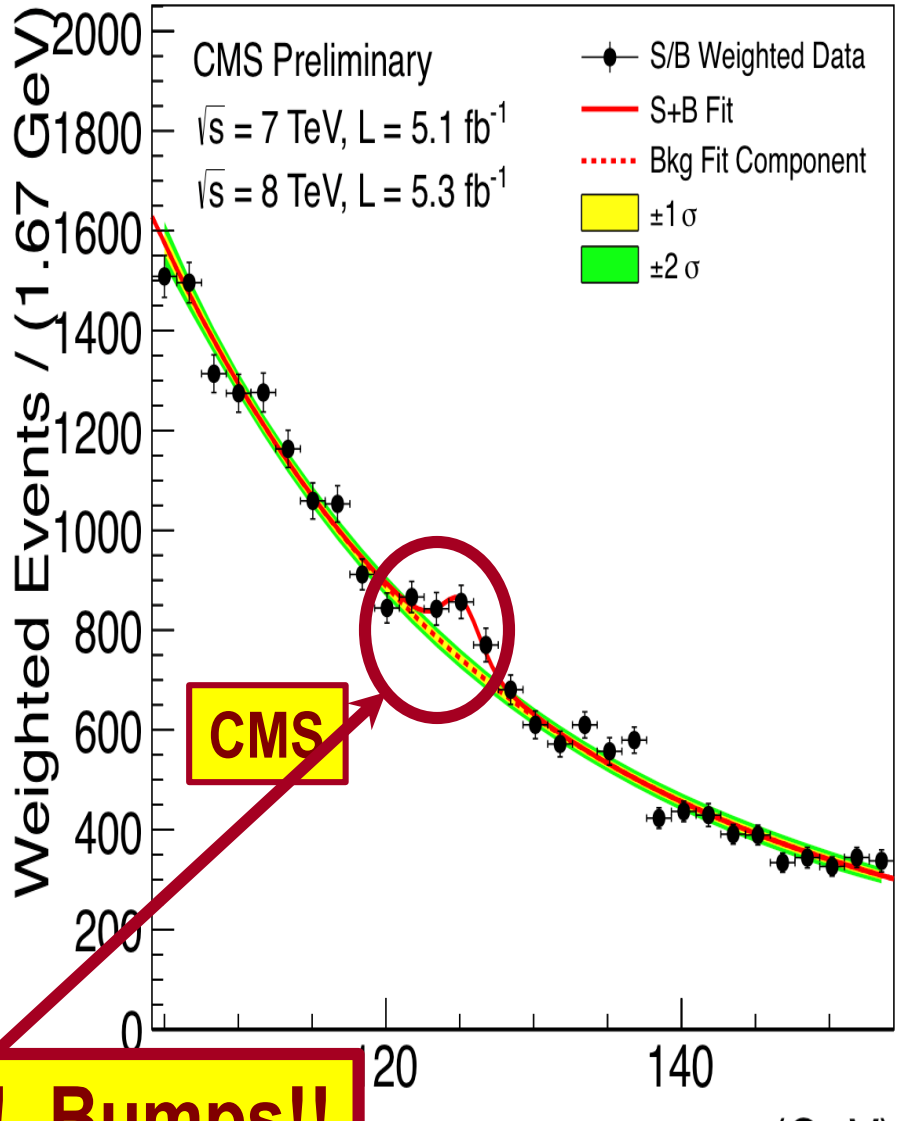


Challenges? No problem!



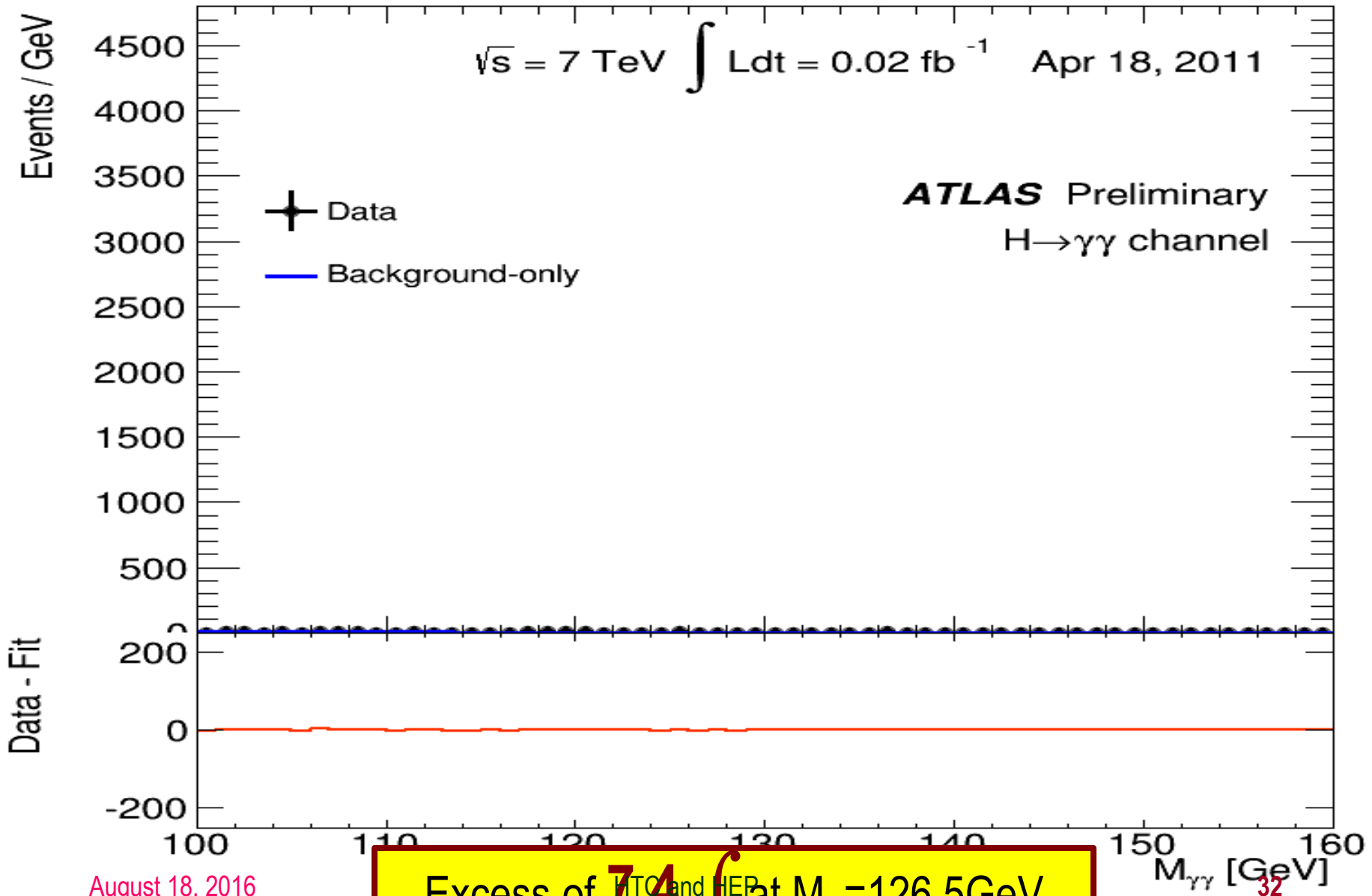
ATLAS and CMS Mass Bump Plots (H→©©)





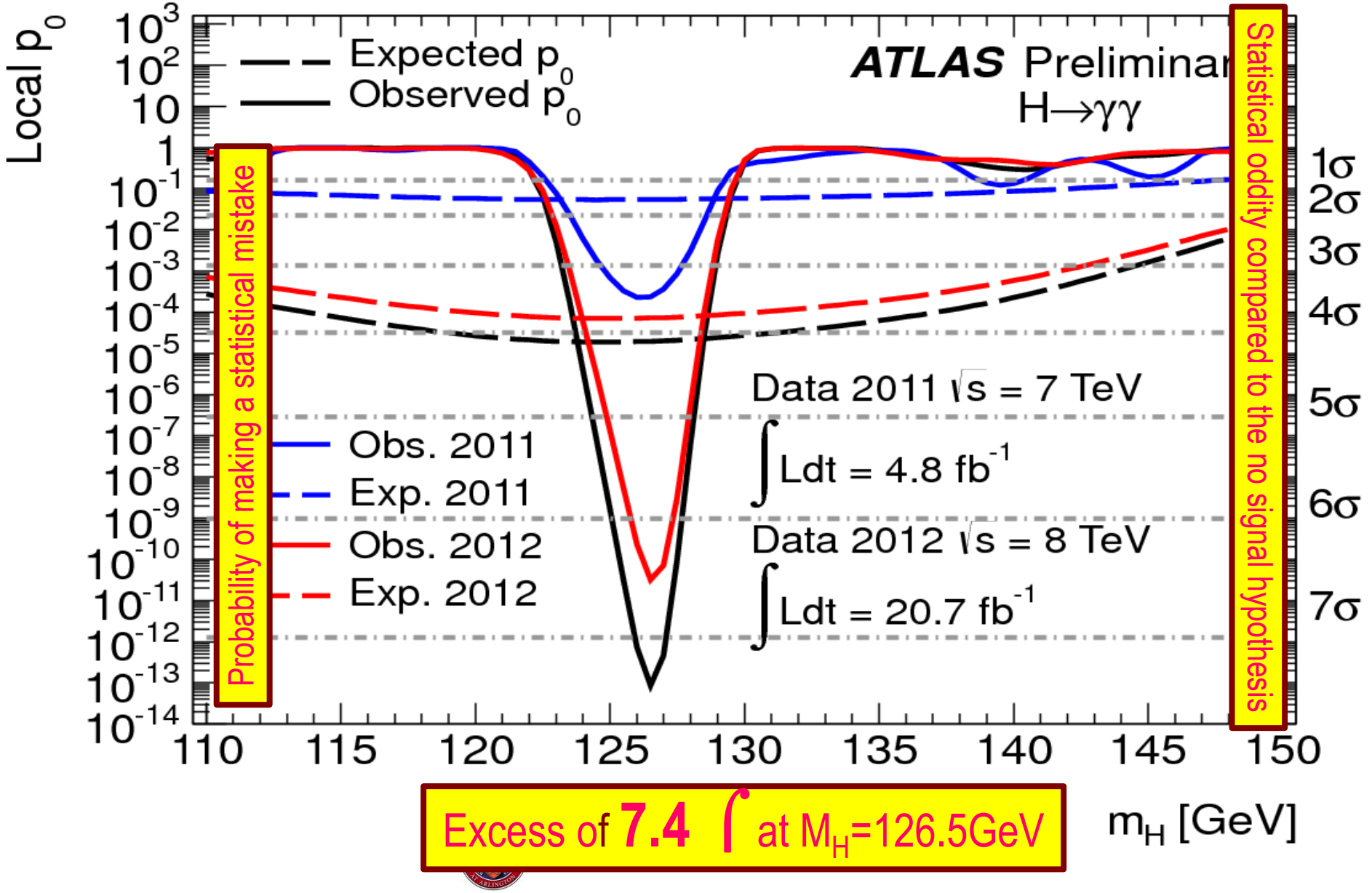
LOOK, Ma! Bumps!!

What did statistics do for Higgs→ γγ?

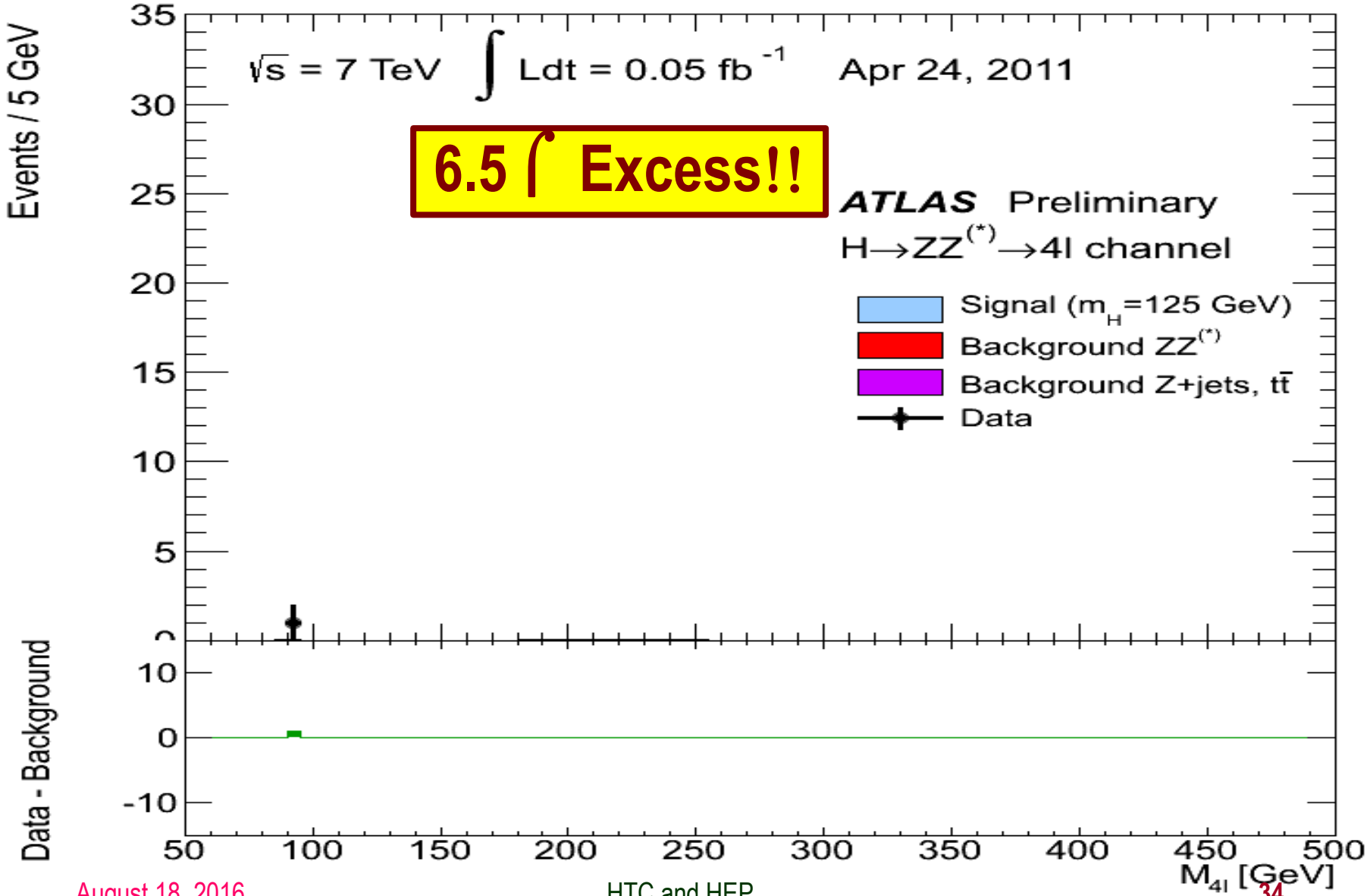


Excess of 7.4 σ at M_H=126.5GeV

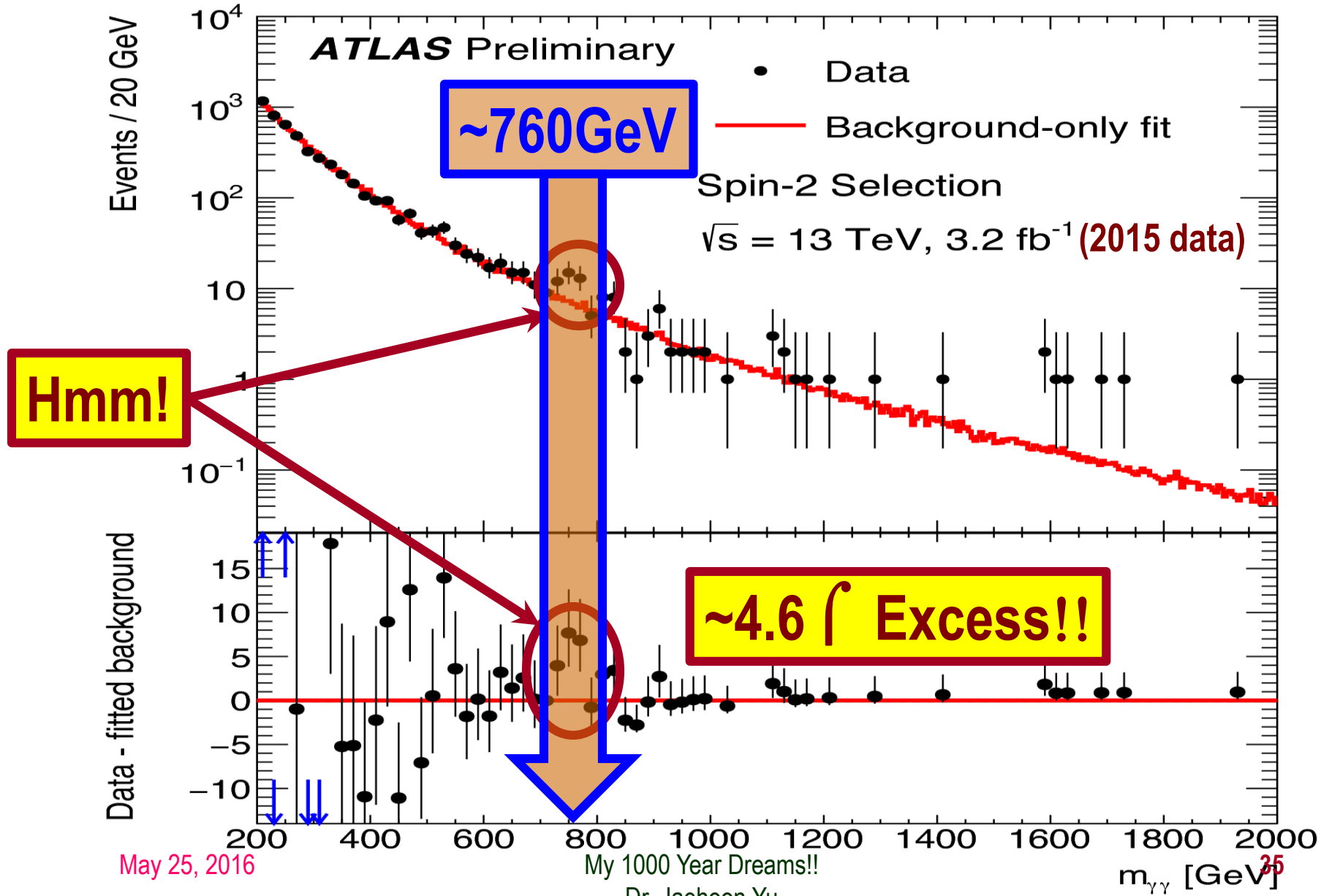
H→γγ Significance



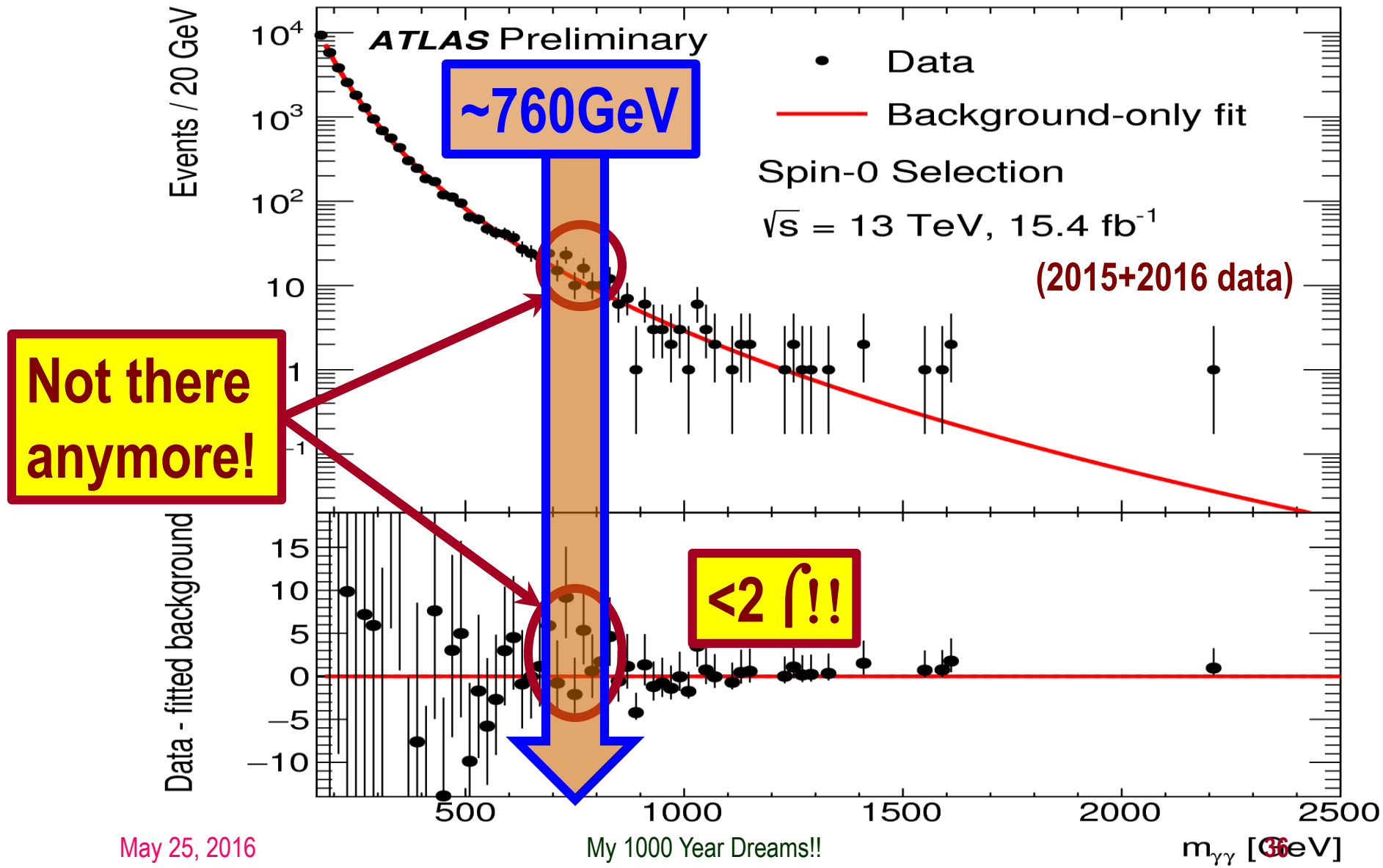
ATLAS Mass Bump Plot (H→4l)?



A hint of something new?

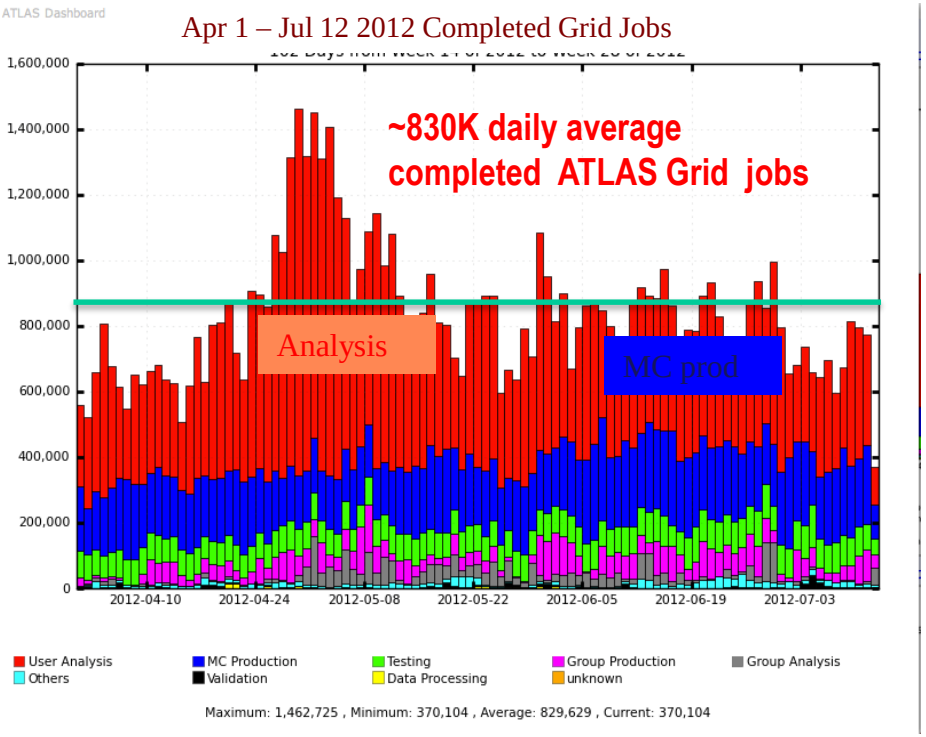


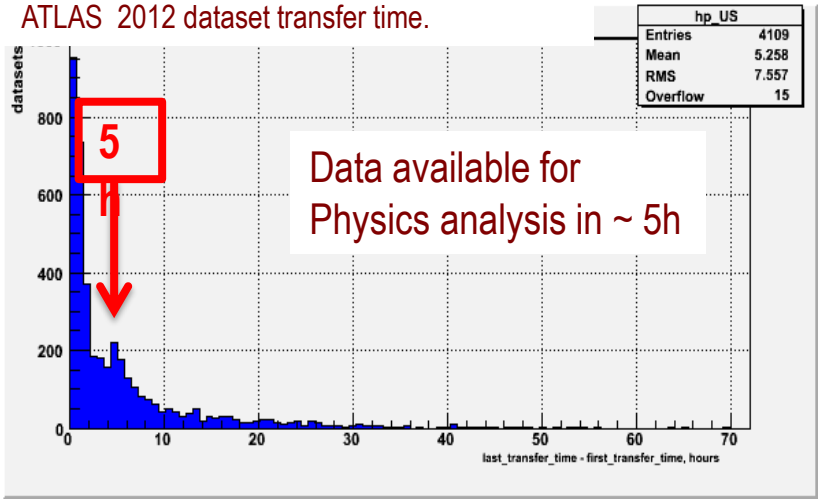
Disappeared after x4 data!!

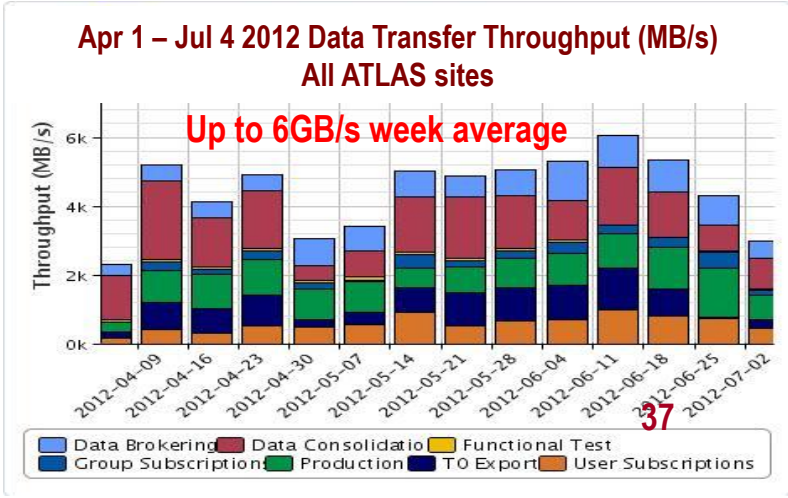


Performance of the Grid

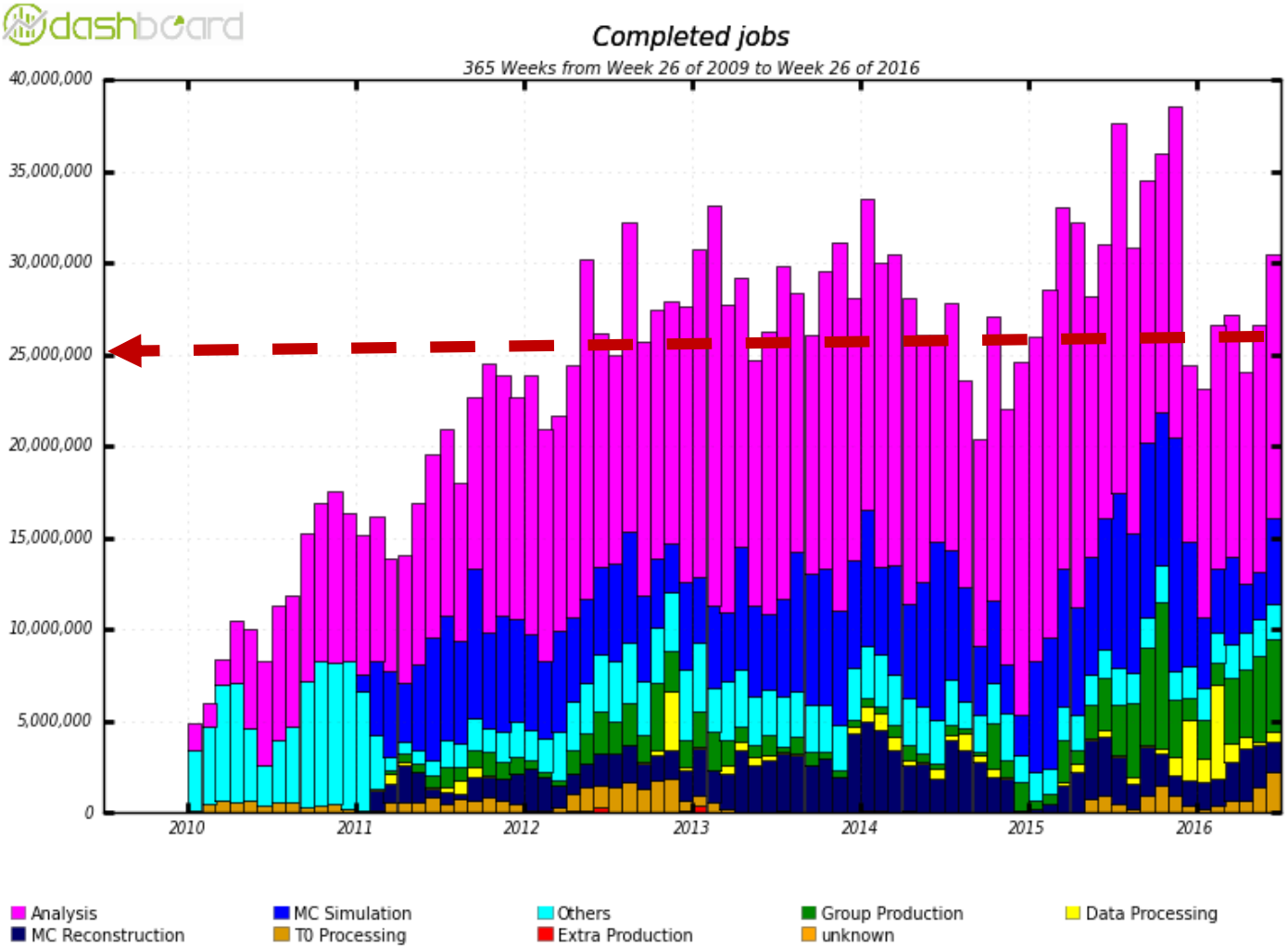
- ATLAS Distributed Computing on the Grid : 10 Tier-1s + CERN + ~70 Tier-2s +…(more than 80 Production sites)
- High volume, high throughput process through fast network!!







PanDA Performance



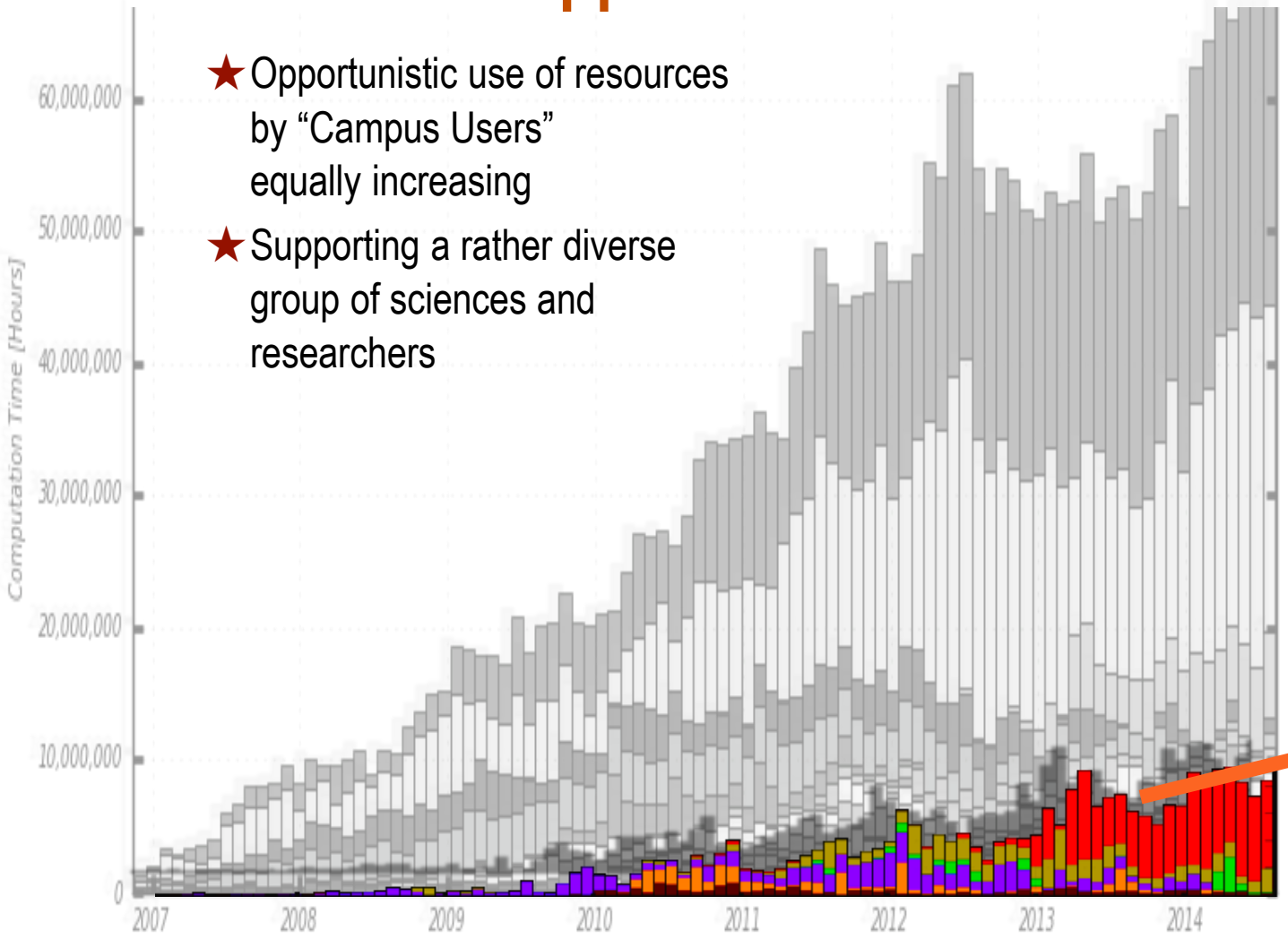
Current scale – 25M jobs completed every month at >hundred sites

First exascale system in HEP – 1.2 Exabytes processed in 2013

Kaushik De

Growing Use of "Owned" and of "Opportunistic" Resources

- ★ Opportunistic use of resources by "Campus Users" equally increasing
- ★ Supporting a rather diverse group of sciences and researchers



Opportunistic Use?

>90M CPU hours opportunistic use past 12 Months

Lotha

Resources Accessible via PanDA

Many Others

About 250,000 job slots used continuously 24x7x365

Now the commercial world picking up..

Early 90's

2004

1996

1998

2006

Many private entities fully utilized the internet communication we've developed to multi-trillion dollar venture!!

Now the concept of cloud being picked up, though not exactly the same idea behind it…

How sure are we with the discovery?

- The statistical significance of the finding is over 7 standard deviations

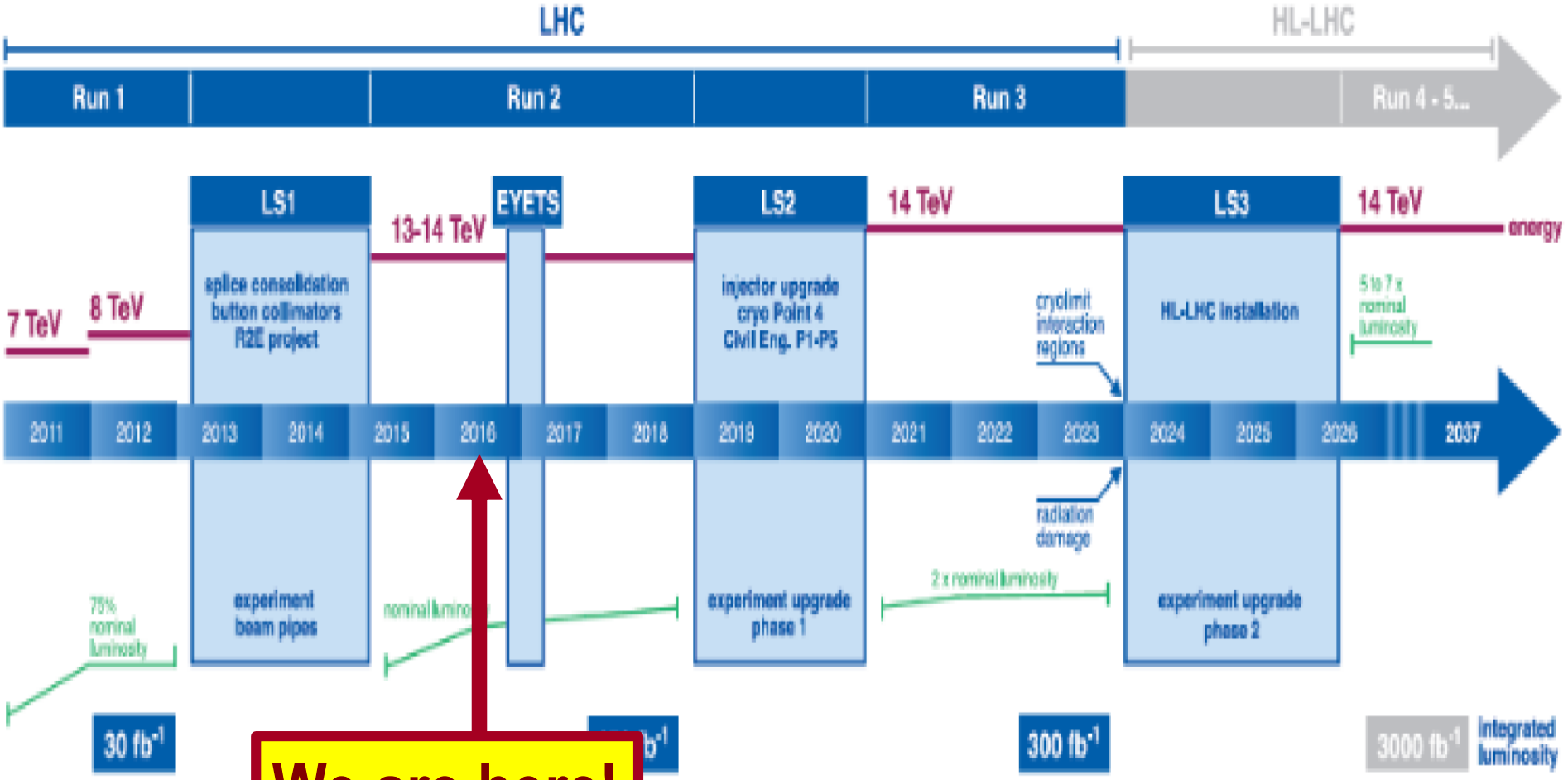
Statistical Significance Table

zσ	Percentage within CI	Percentage outside CI	Fraction outside CI
0.674 490σ	50%	50%	1 / 2
0.994 458σ	68%	32%	1 / 3.125
1σ	68.268 9492%	31.731 0508%	1 / 3.151 4872
1.281 552σ	80%	20%	1 / 5
1.644 854σ	90%	10%	1 / 10
1.959 964σ	95%	5%	1 / 20
2σ	95.449 9736%	4.550 0264%	1 / 21.977 895
2.575 829σ	99%	1%	1 / 100
3σ	99.730 0204%	0.269 9796%	1 / 370.398
3.290 527σ	99.9%	0.1%	1 / 1,000
3.890 592σ	99.99%	0.01%	1 / 10,000
4σ	99.993 666%	0.006 334%	1 / 15,787
4.417 173σ	99.999%	0.001%	1 / 100,000
4.891 638σ	99.9999%	0.0001%	1 / 1,000,000
5σ	99.999 942 6697%	0.000 057 3303%	1 / 1,744,278
5.326 724σ	99.999 99%	0.000 01%	1 / 10,000,000
5.730 729σ	99.999 999%	0.000 001%	1 / 100,000,000
6σ	99.999 999 8027%	0.000 000 1973%	1 / 506,797,346
6.109 410σ	99.999 9999%	0.000 0001%	1 / 1,000,000,000
6.466 951σ	99.999 999 99%	0.000 000 01%	1 / 10,000,000,000
6.806 502σ	99.999 999 999%	0.000 000 001%	1 / 100,000,000,000
7σ	99.999 999 999 7440%	0.000 000 000 256%	1 / 390,682,215,445

How sure are we with the discovery?

- The statistical significance of the finding is much bigger than seven standard deviations
  - Level of significance: much better than 99.999 999 999 7% (eleven 9s!!)
  - We could be wrong once if we do the same experiment 391,000,000,000 times (will take ~13,000 years even if each experiment takes only 1s!!)
- So what does this mean?
  - We have discovered the heaviest new boson we've seen thus far
  - It has many properties consistent with the Standard Model Higgs particle
    - It quacks like a duck and walks like a duck but…
  - We do not have enough data to precisely measure all the properties – mass, lifetime, the rate at which this particle decays to certain other particles, etc – to definitively determine its nature
- Precision measurements and searches in new channels ongoing

Long Term LHC Schedule

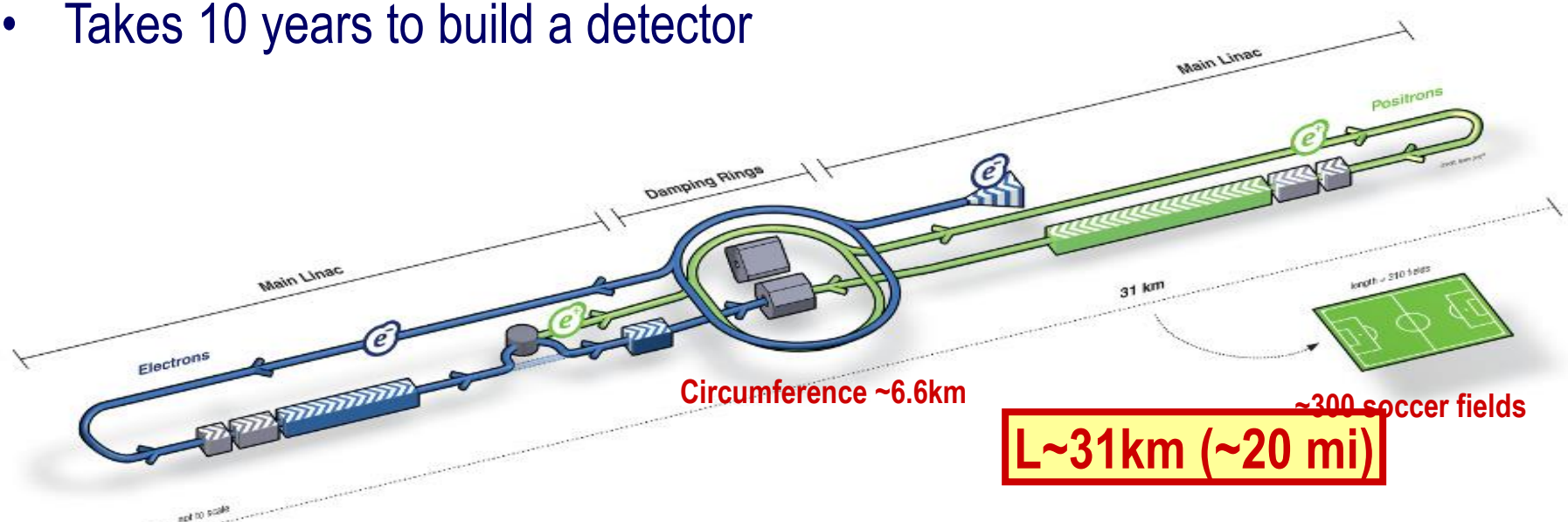


We are here!

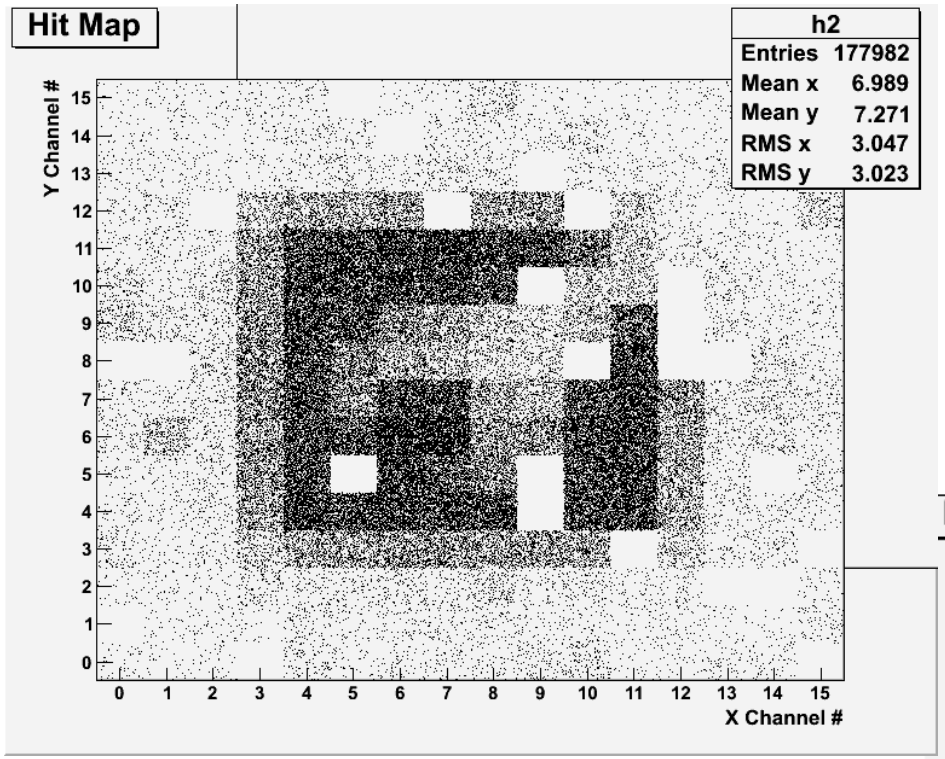
Collision data taking begun end of April, 2016!

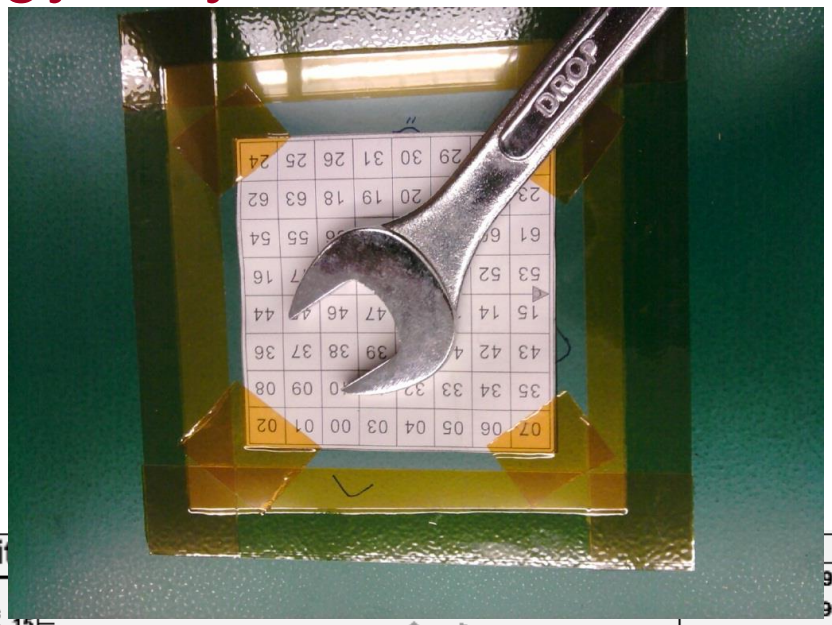
What next? Future Linear Collider

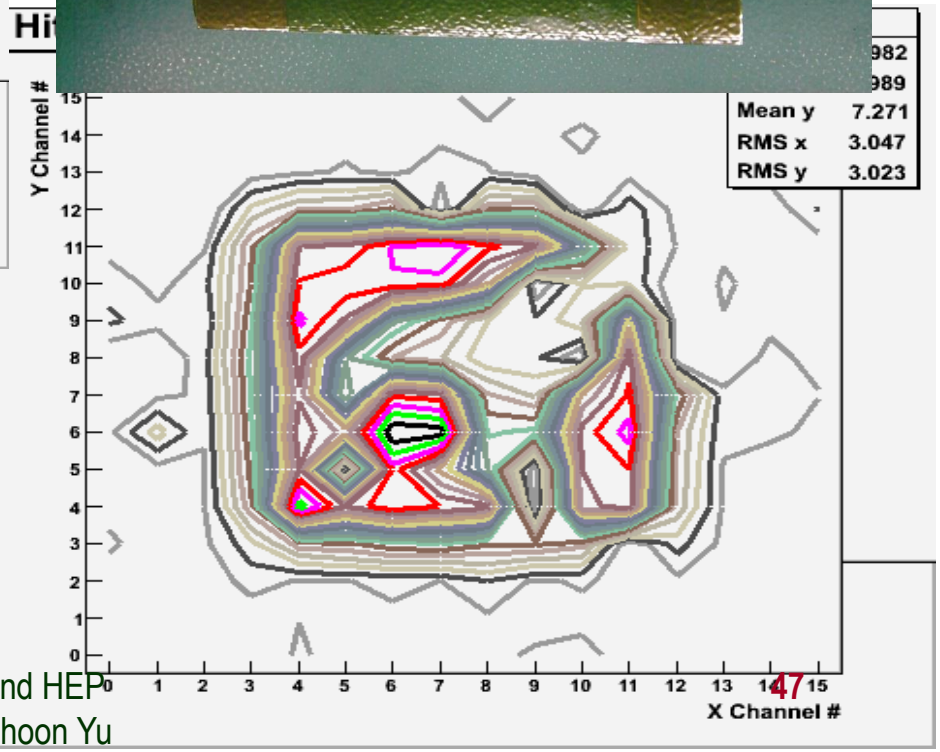
- Now that we have found a new boson, precision measurement of the particle's properties becomes important
- An electron-positron collider on a straight line for precision measurements
- 10~15 years from now (In Dec. 2011, Japanese PM announced that they would bid for a LC in Japan and reaffirmed by the new PM in 2013)
  - Our Japanese colleagues have declared that they will bid for building ILC
  - Japan just announced the selection of the site for the ILC the Week of Aug. 19!!
- Takes 10 years to build a detector



Bi-product of High Energy Physics Research



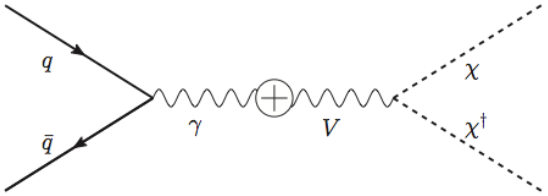




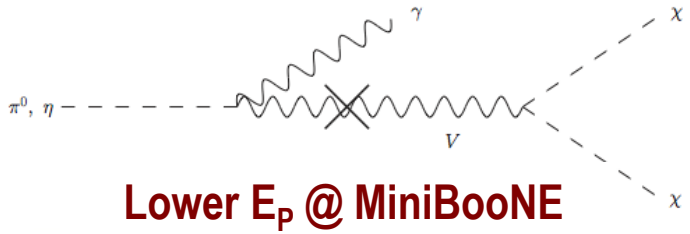
Can you see what the object is?
(GEM Detector X-ray Image)

Light DM Production at High Intensity Accelerator

- Now the Higgs particle, a part of only 5% of the universe, may've been seen
- It is time for us to look into the 95% of the universe using high intensity beams, like 1.2 – 2.3MW beams at Fermilab in the US!

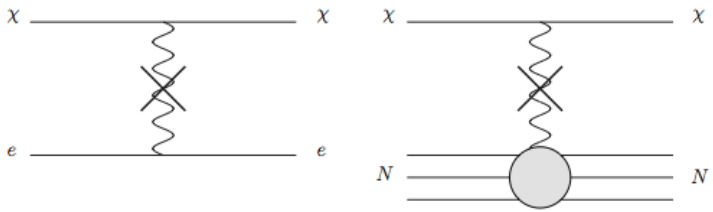


Higher E_P @ LBNE



Lower E_P @ MiniBooNE

- Detection of DM:



- How does a DM event look in an experiment?:

High Intensity Proton Beam

So are we done with the grid?

- LHC has performed extremely well!
- The data size will increase by 5 fold or more in run II
  - Computing will be under even more stress
- Grid computing infrastructure has served well thus far
  - 1500 ATLAS users process PBs of data & billions of jobs
- Identified limits in databases scalability, CPU resources, storage utilization, etc, are being addressed
- Time to plan for the future in HEP ➔ especially that in the high intensity neutrino physics
- And the use of the technology in everyday lives

Conclusions

- In the quest for the origin of the universe, High Energy Physics
  - Uses accelerators to look into extremely small distances
  - Uses large detectors to explore nature and unveil secrets of universe
  - Uses large number of computers to process data in a timely fashion
  - Large amount of data gets accumulated ➔ computing grid performed for expeditious data analyses
- HEP is an exciting endeavor in understanding nature
- Physics analyses at one's own desktop using computing grid sitting behind has happened!!
- Computing grid needed for other disciplines with large data sets
- Computing grid now going outside of HEP into everyday lives
- A true computing grid is revolutionizing everyday lives ….

What will you learn?

- From the underlying computing grid software infrastructure to the practical use of computing grid
- Today
  - Intro to Throughput Computing (HTC) – Kyle
  - Running jobs locally with CONDOR – Kyle, Julia, Pat
  - More about CONDOR – Kyle
- Tomorrow
  - More about HTC with CONDOR I&II – Horst
  - Introduction to Grids and Workflow Systems - Kyle
  - Running jobs remotely with DAGMan – Julia
  - The Grid and its use by the workflow systems – Kyle
  - Tutorial: Analysis example – Julia, Pat, Kyle, Horst & Jae
  - Strategies and technologies of handling large workflows– Kyle
  - Wrap up and further pointer – Jae

FFT: Number of beam particles per sec?

- What is the number of particles per second for an accelerator facility that can provide:
  - P MW of total beam power
  - of charged particles of energy E GeV?

$$N_p\left(\ /\text{sec};E\ GeV;P\ MW\right)=P/E\times 6.3\times 10^{15}\left(particles/\text{sec}\right)$$

  - What is the number of protons per second for 120GeV beams at 1.2MW?

$$N_p\left(\ /\text{sec};120\ GeV;1.2\ MW\right)=\frac{1.2}{120}\times 6.3\times 10^{15}\left(particles/\text{sec}\right)$$
$$=6.3\times 10^{13}\left(particles/\text{sec}\right)$$

  - What is the beam current? $I=N_p\times 1.6\times 10^{-19}$

$$=1.2\times 10^{-5}\left(C/\text{sec}\right)=12\,mA$$