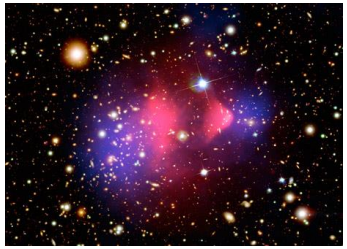
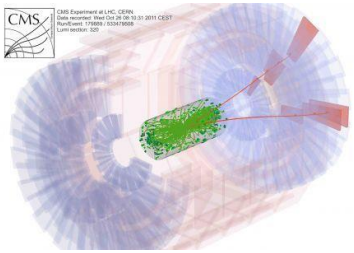
Project CODER

CMS Open Data Analysis EnviRonment

Sergei Gleyzer, University of Florida





USCMS Workshop on Open Data at UCF

June 15-17, 2016

Outline

- What is Project CODER
- Large Hadron Collider
- CMS Experiment at CERN
- CMS Open Data
- Plan for Today

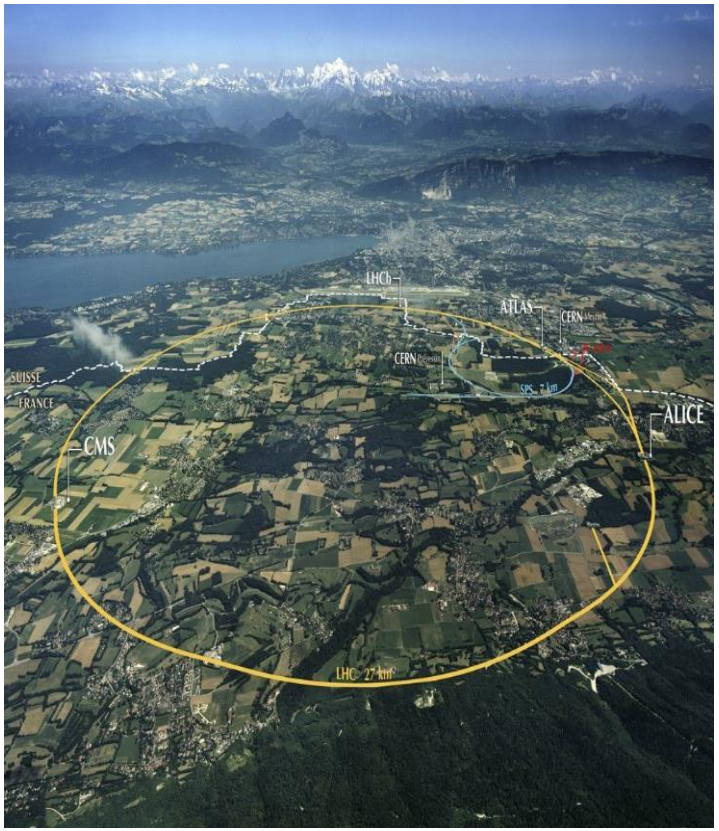
Project CODER

CMS Open Data Analysis EnviRonment

- **Interactive Analysis**
  - **Based on JUPYTER Notebooks**
- **Open High Energy Physics Data**
- **Compact Muon Solenoid (CMS) Experiment at CERN**

A Few Words About CERN

World Wide Web



CERN

Where the web was born

The Power of 3

Recurring theme of threes:

W: **What?**

W: **Where?**

W: **Why?**

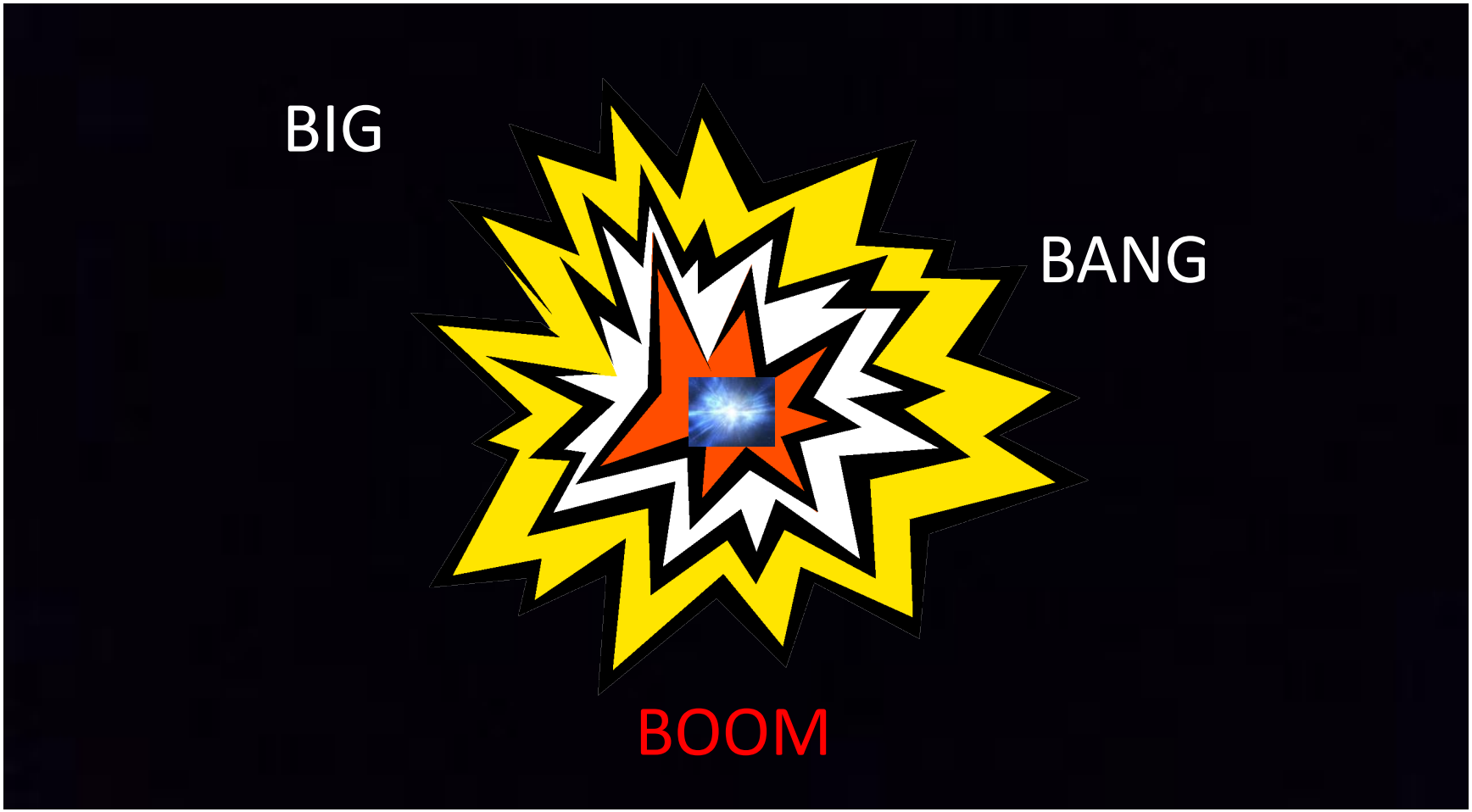
UF
UNIVERSITY of FLORIDA

What?

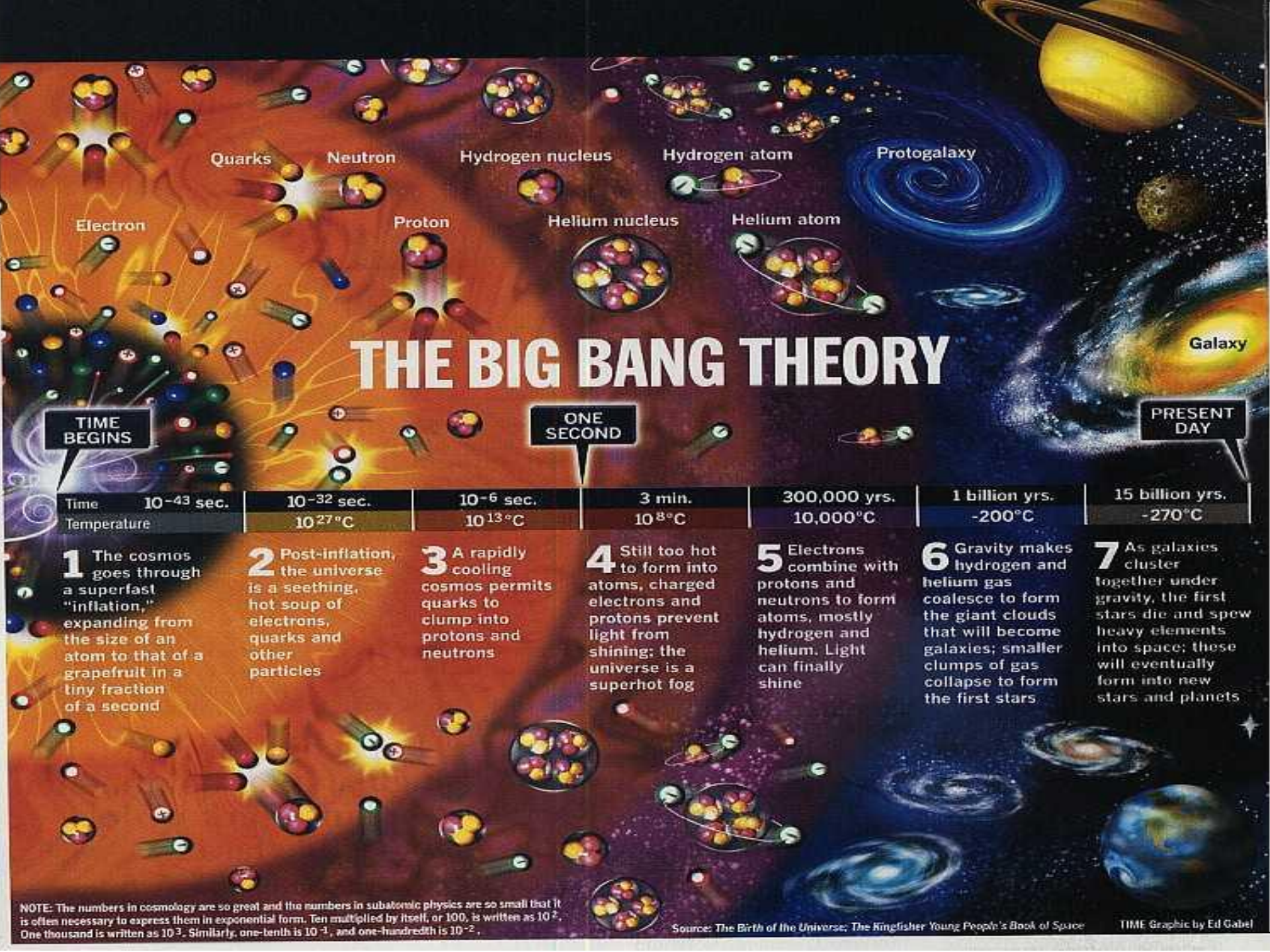
Where?

Why?

Start at the Beginning



The Big Bang Theory



THE BIG BANG THEORY

TIME BEGINS

ONE SECOND

PRESENT DAY

Time	10^{-43} sec.	10^{-32} sec.	10^{-6} sec.	3 min.	300,000 yrs.	1 billion yrs.	15 billion yrs.
Temperature		10^{27}°C	10^{13}°C	10^{8}°C	10,000°C	-200°C	-270°C

1 The cosmos goes through a superfast "inflation," expanding from the size of an atom to that of a grapefruit in a tiny fraction of a second

2 Post-inflation, the universe is a seething, hot soup of electrons, quarks and other particles

3 A rapidly cooling cosmos permits quarks to clump into protons and neutrons

4 Still too hot to form into atoms, charged electrons and protons prevent light from shining; the universe is a superhot fog

5 Electrons combine with protons and neutrons to form atoms, mostly hydrogen and helium. Light can finally shine

6 Gravity makes hydrogen and helium gas coalesce to form the giant clouds that will become galaxies; smaller clumps of gas collapse to form the first stars

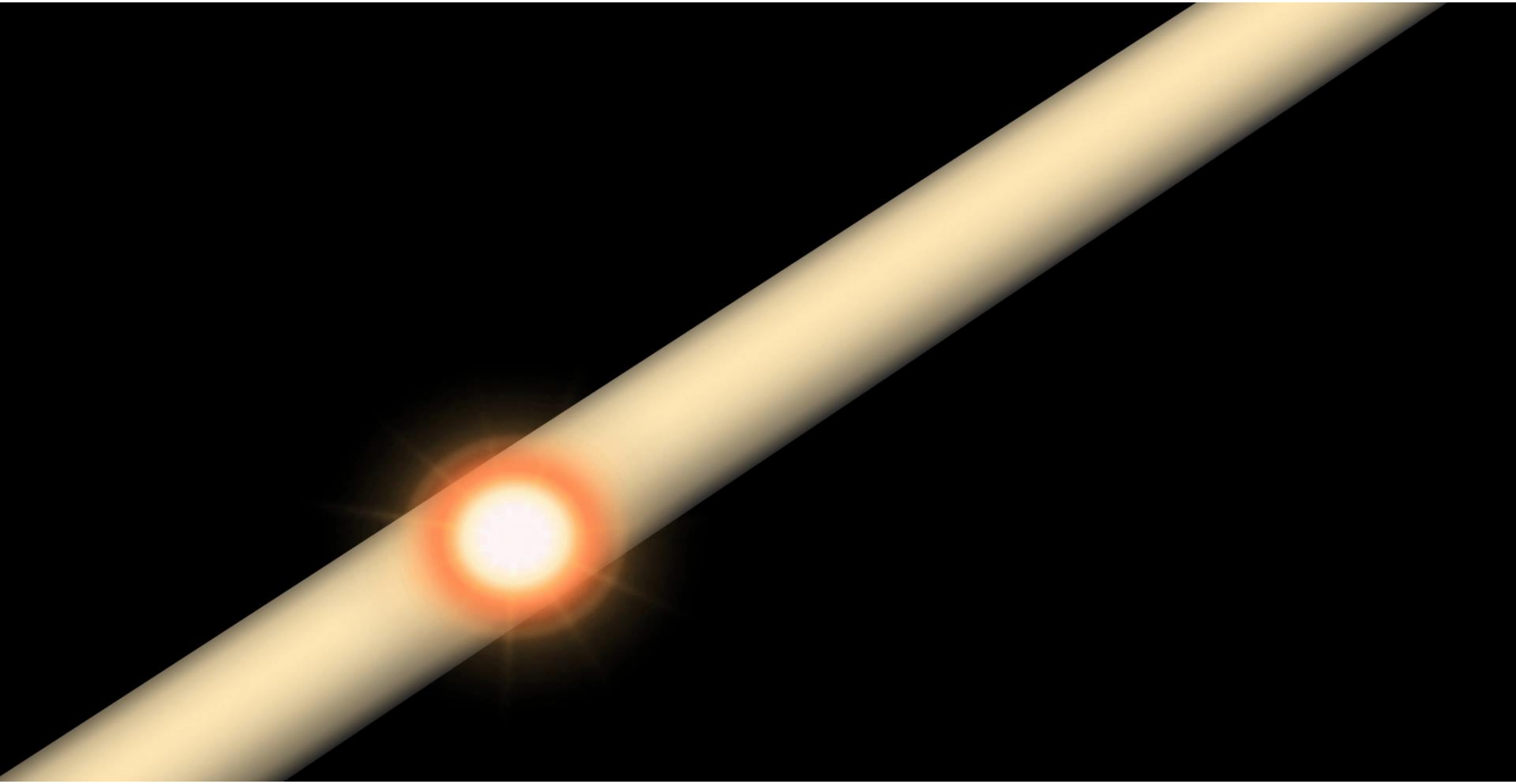
7 As galaxies cluster together under gravity, the first stars die and spew heavy elements into space; these will eventually form into new stars and planets

NOTE: The numbers in cosmology are so great and the numbers in subatomic physics are so small that it is often necessary to express them in exponential form. Ten multiplied by itself, or 100, is written as 10^2. One thousand is written as 10^3. Similarly, one-tenth is 10^{-1}, and one-hundredth is 10^{-2}.

Source: *The Birth of the Universe*; *The Kingfisher Young People's Book of Space*

TIME Graphic by Ed Gabel

Let's collide particles!



UF
UNIVERSITY of FLORIDA

What?

Where?

Why?

Large Hadron Collider

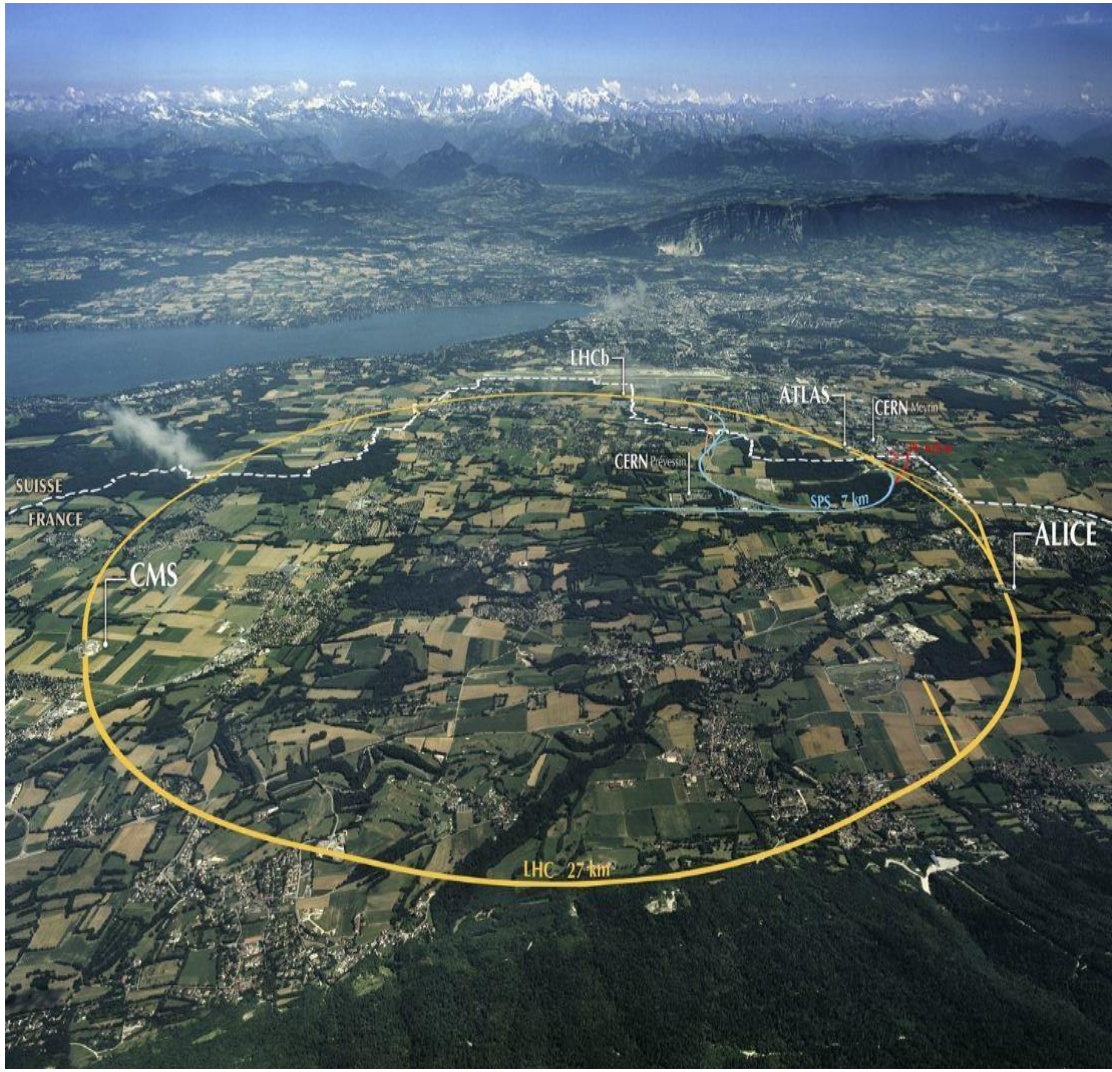
Crown jewel of modern science:

State of the art in science and engineering

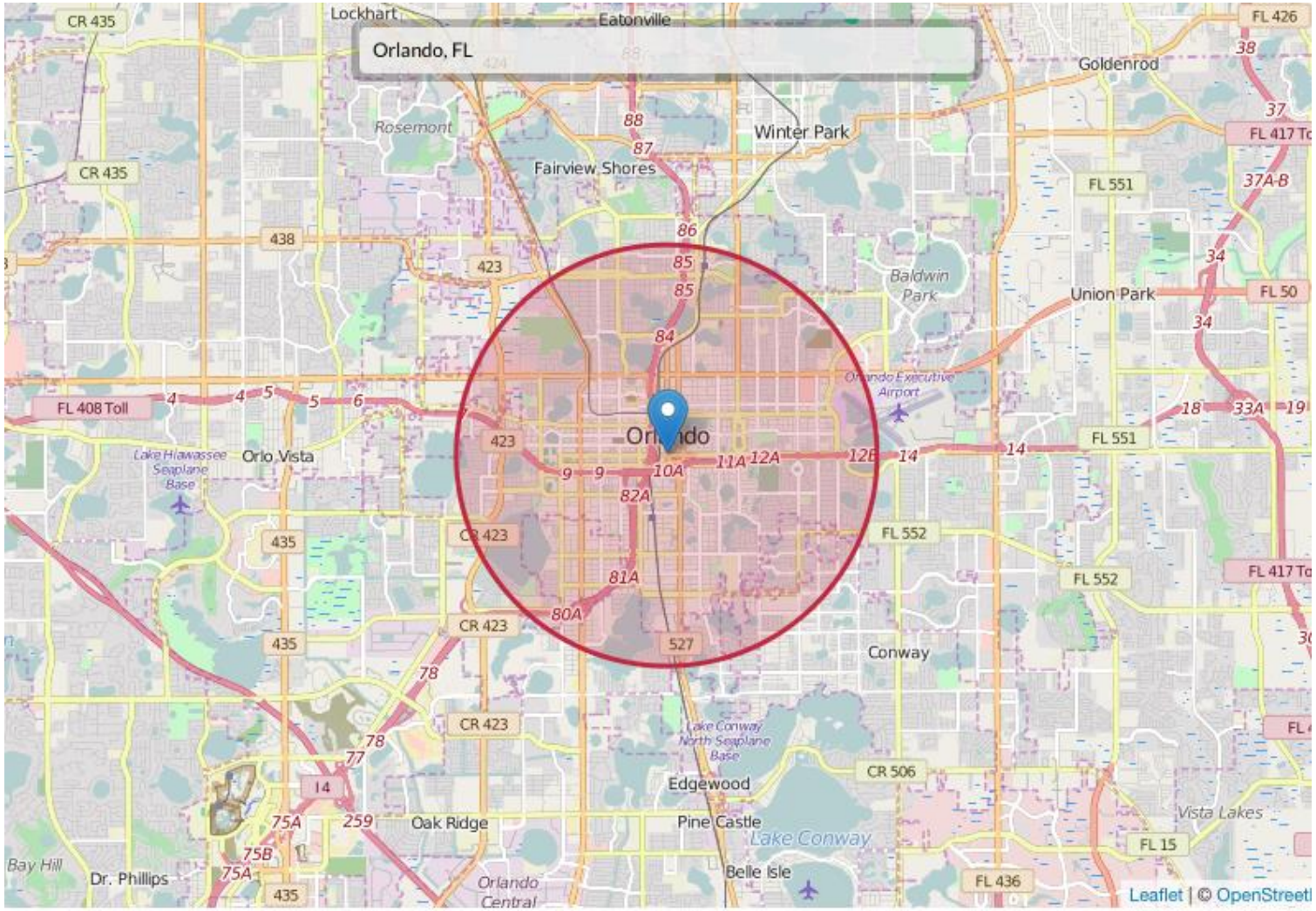
Operating since 2010

– **expected to run for next 20(+) years**

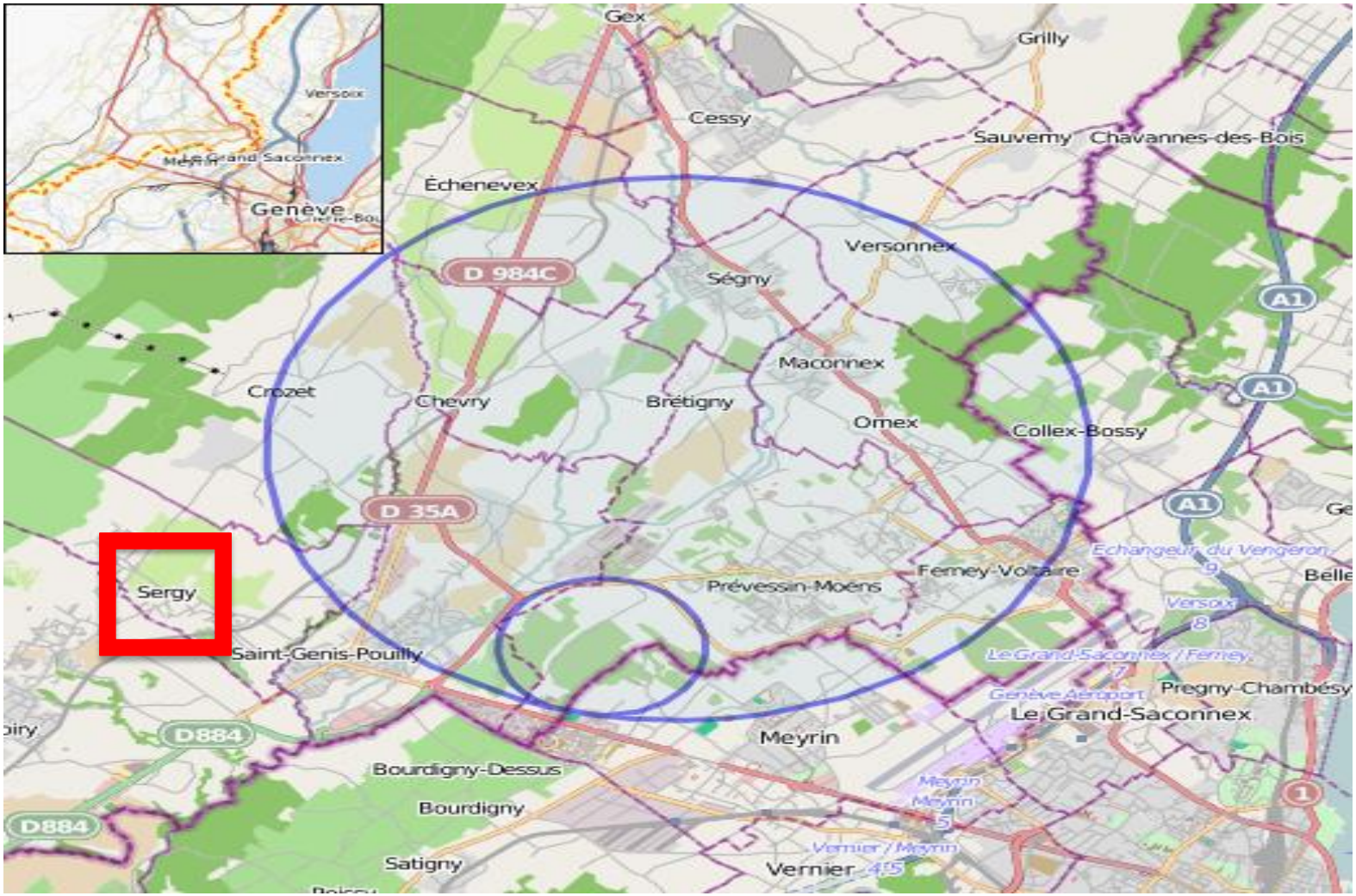
Amazing success of an international collaboration of thousands of scientists from across the world

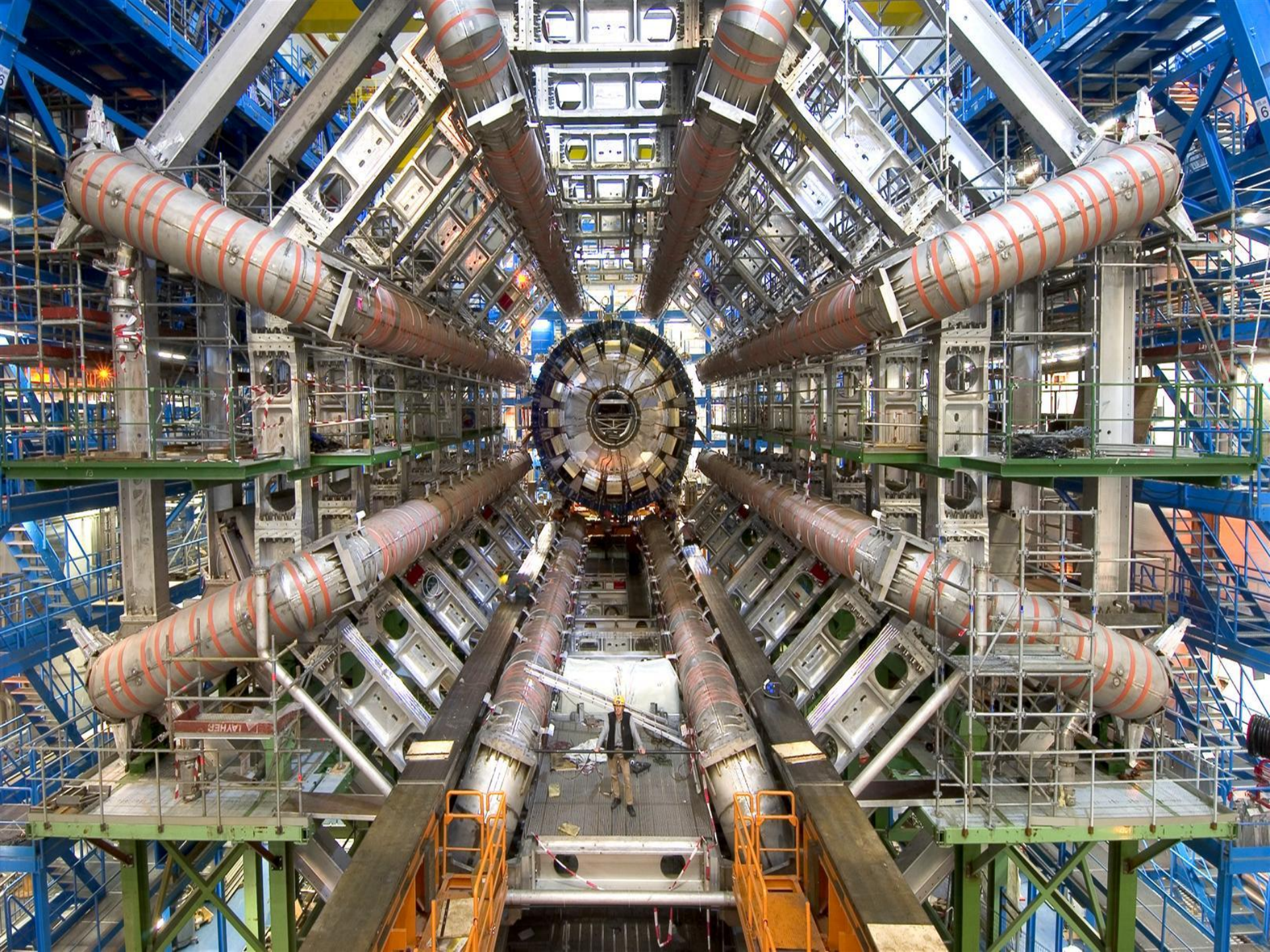


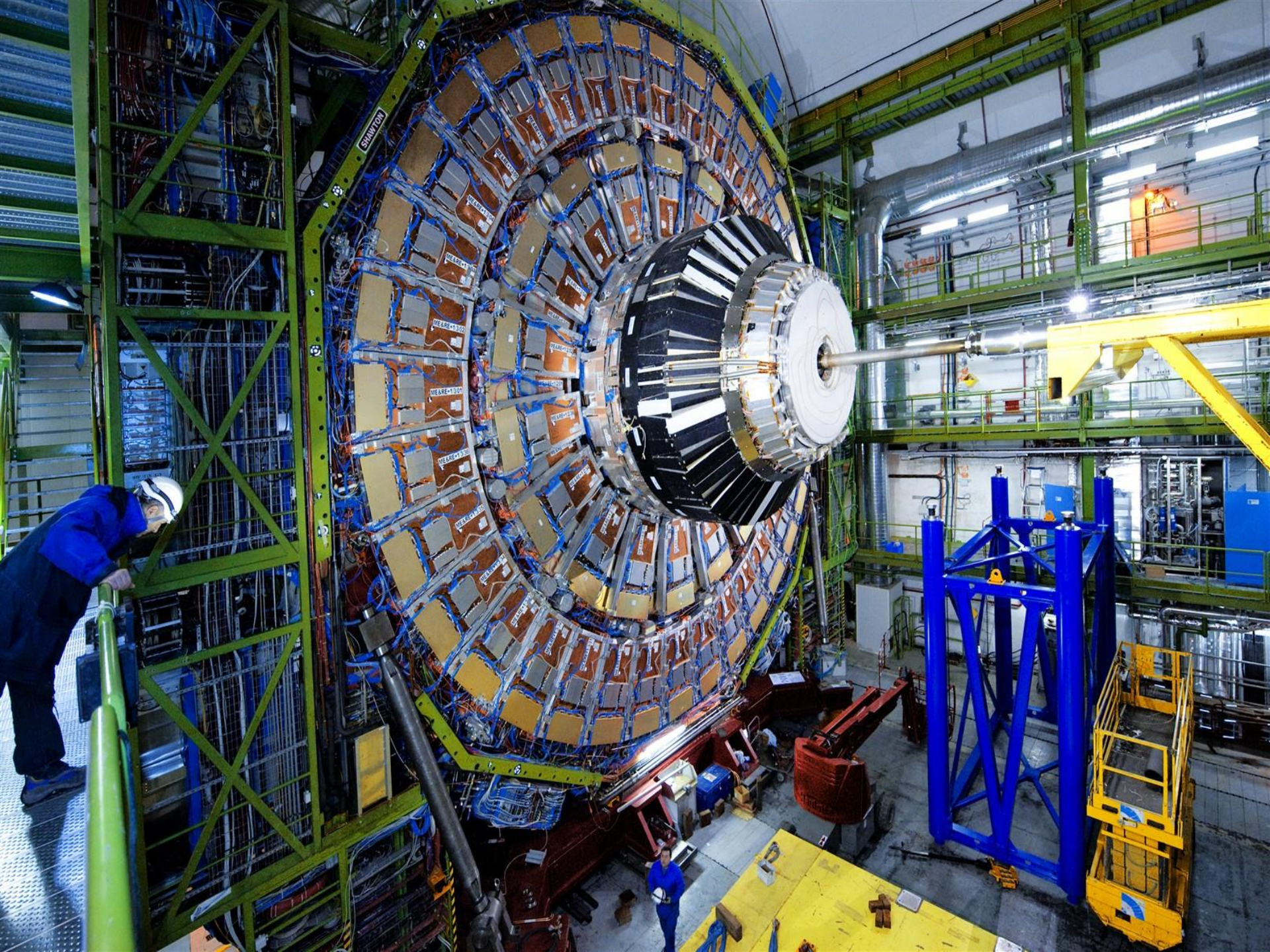
How Large is the LHC?

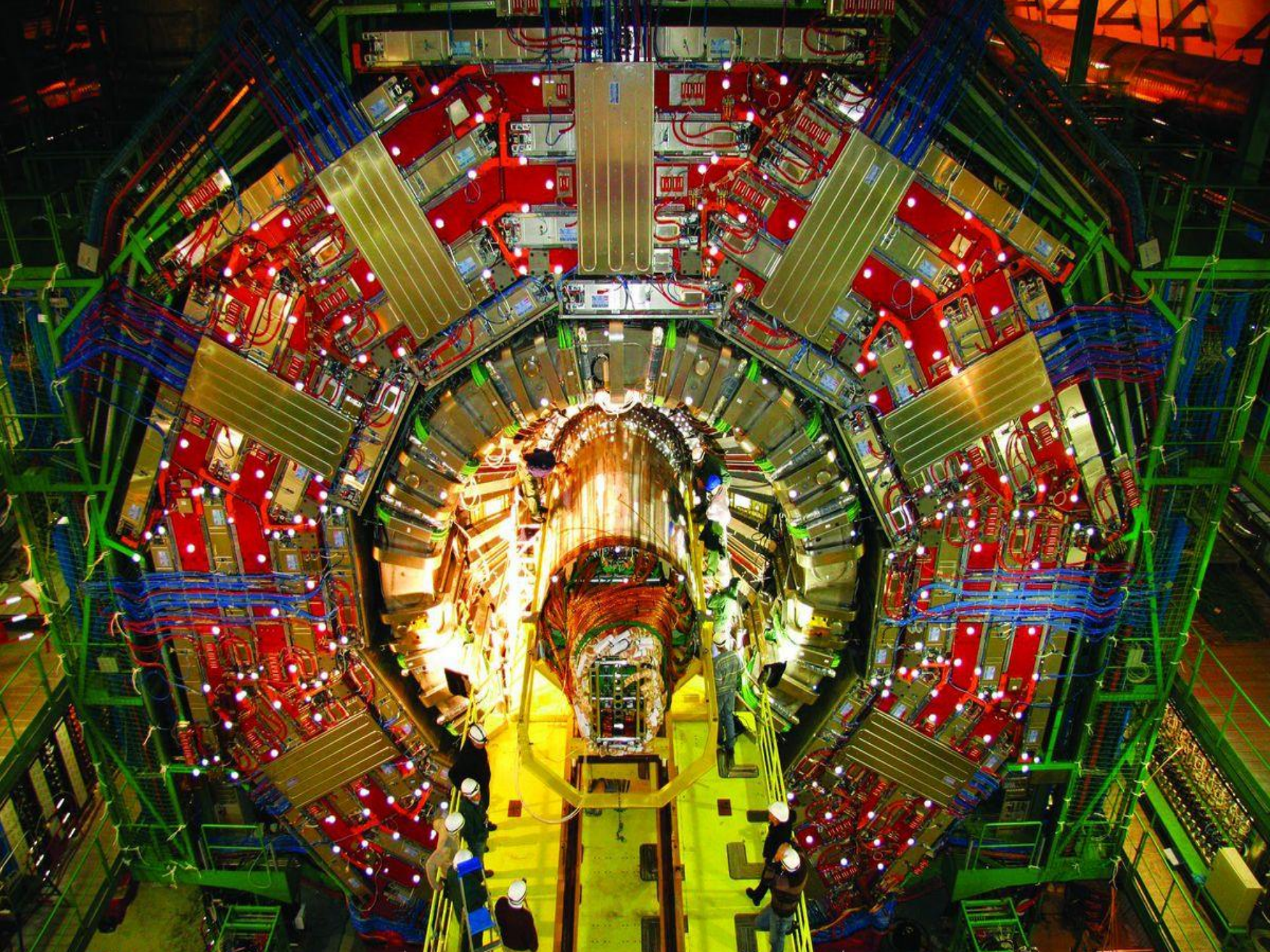


Large Hadron Collider






SHANTON



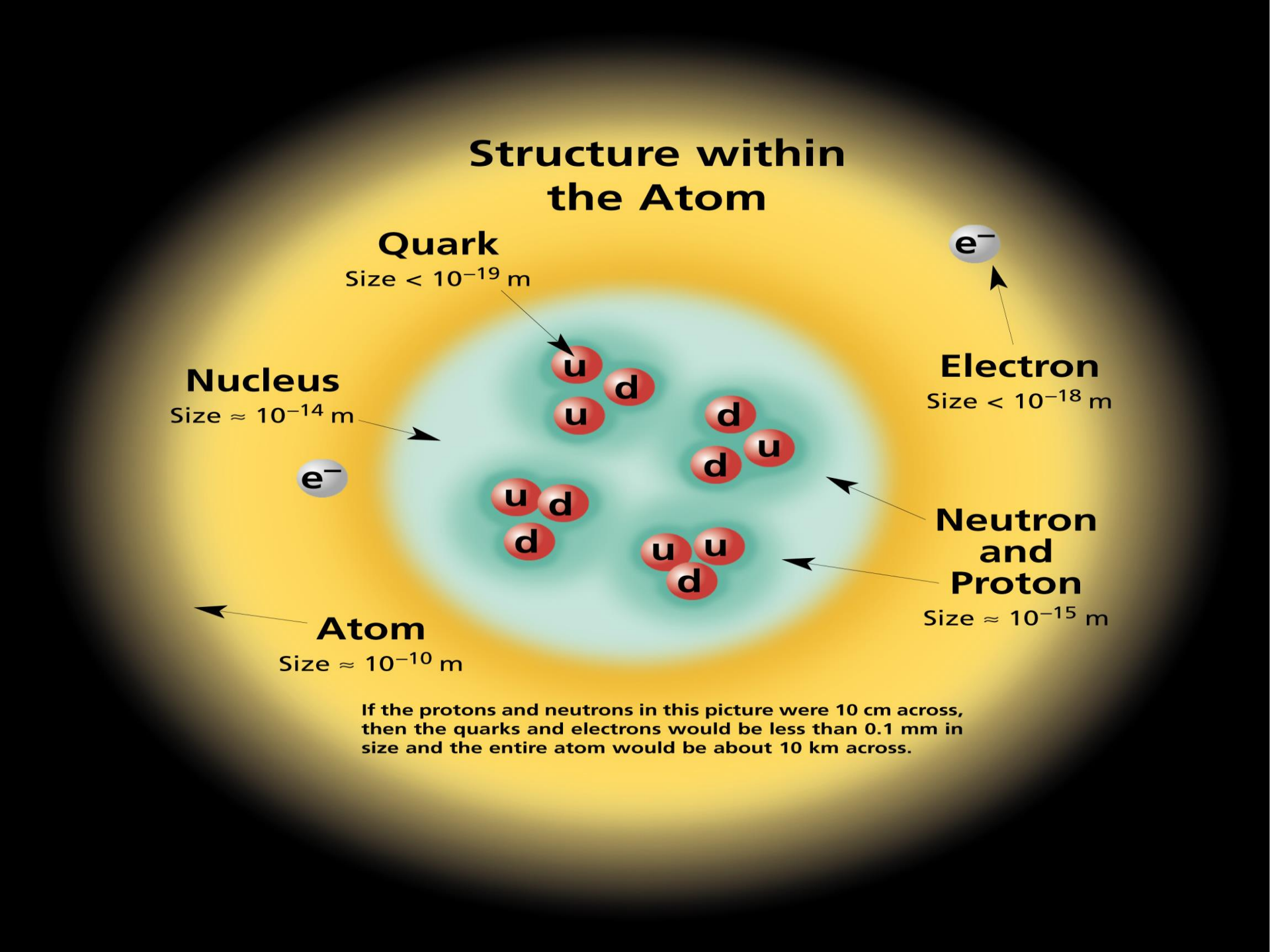
UF
UNIVERSITY of FLORIDA

CMS

What?

Where?

Why?


Structure within the Atom
Quark
Size < 10^{-19} m
Nucleus
Size ≈ 10^{-14} m
Atom
Size ≈ 10^{-10} m
Electron
Size < 10^{-18} m
Neutron and Proton
Size ≈ 10^{-15} m
e⁻
e⁻
u
d
u
d
d
u
d
u
d
u
u
d
If the protons and neutrons in this picture were 10 cm across, then the quarks and electrons would be less than 0.1 mm in size and the entire atom would be about 10 km across.

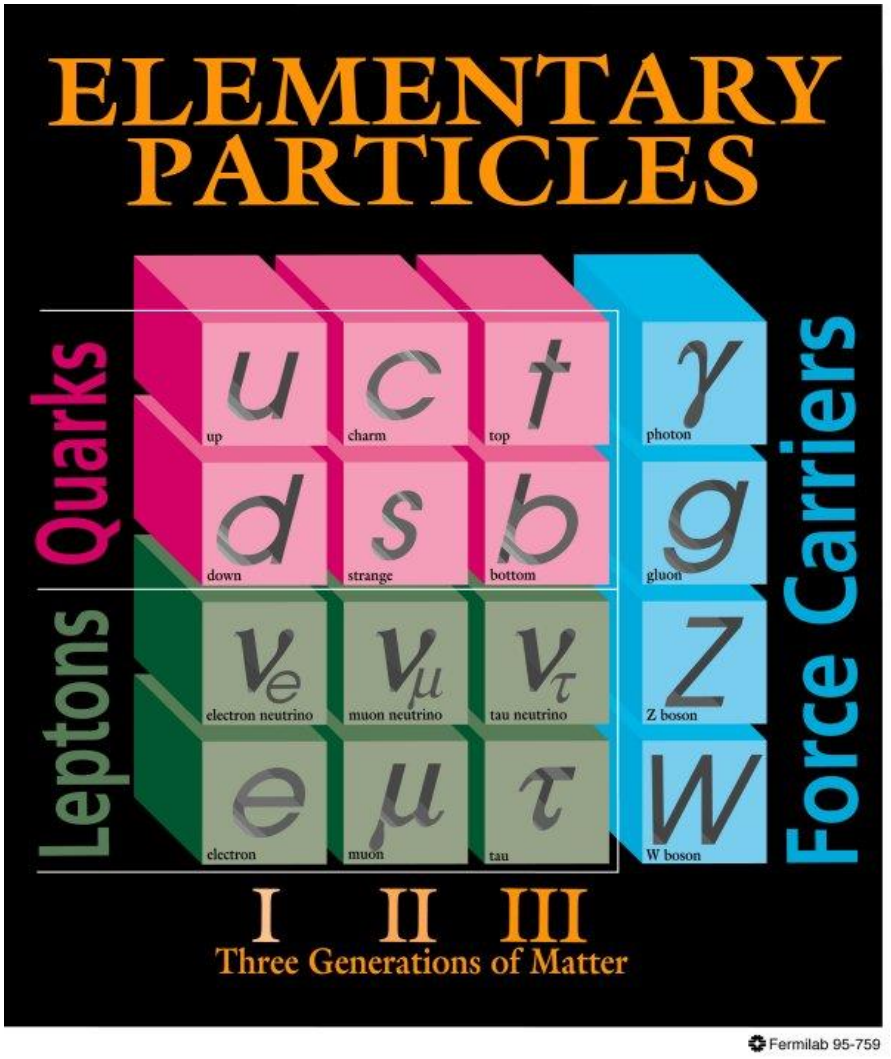
Standard Model

Extremely accurate theory of particles and interactions

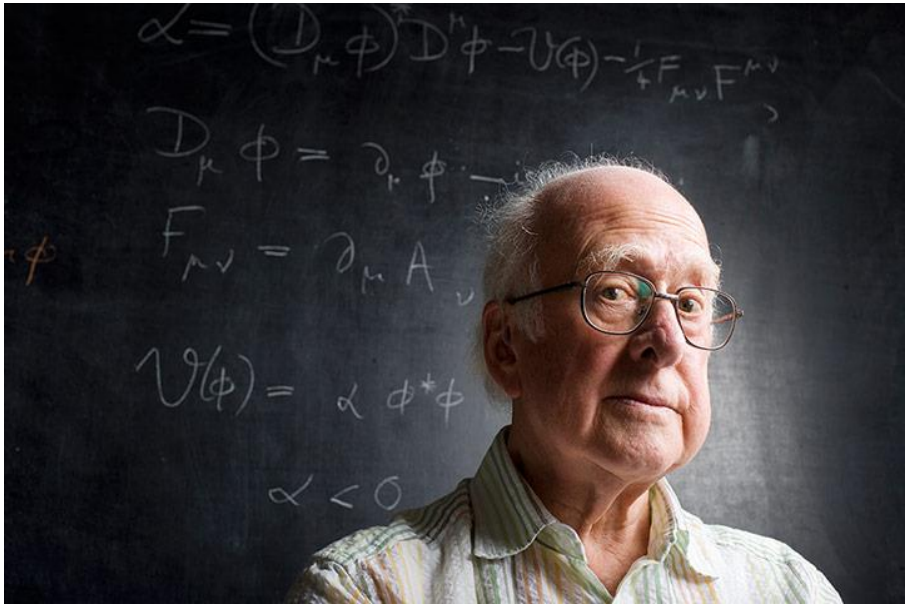
Staggering precision to many decimal points in all kinds of predicted observables

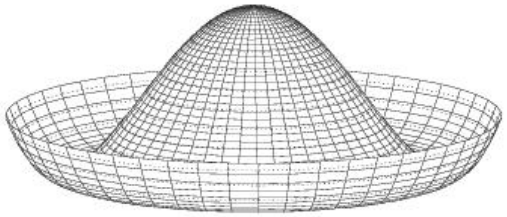
Around since 1970s

– **Many missing predicted particles found since then (t, c, g, Z, W)**



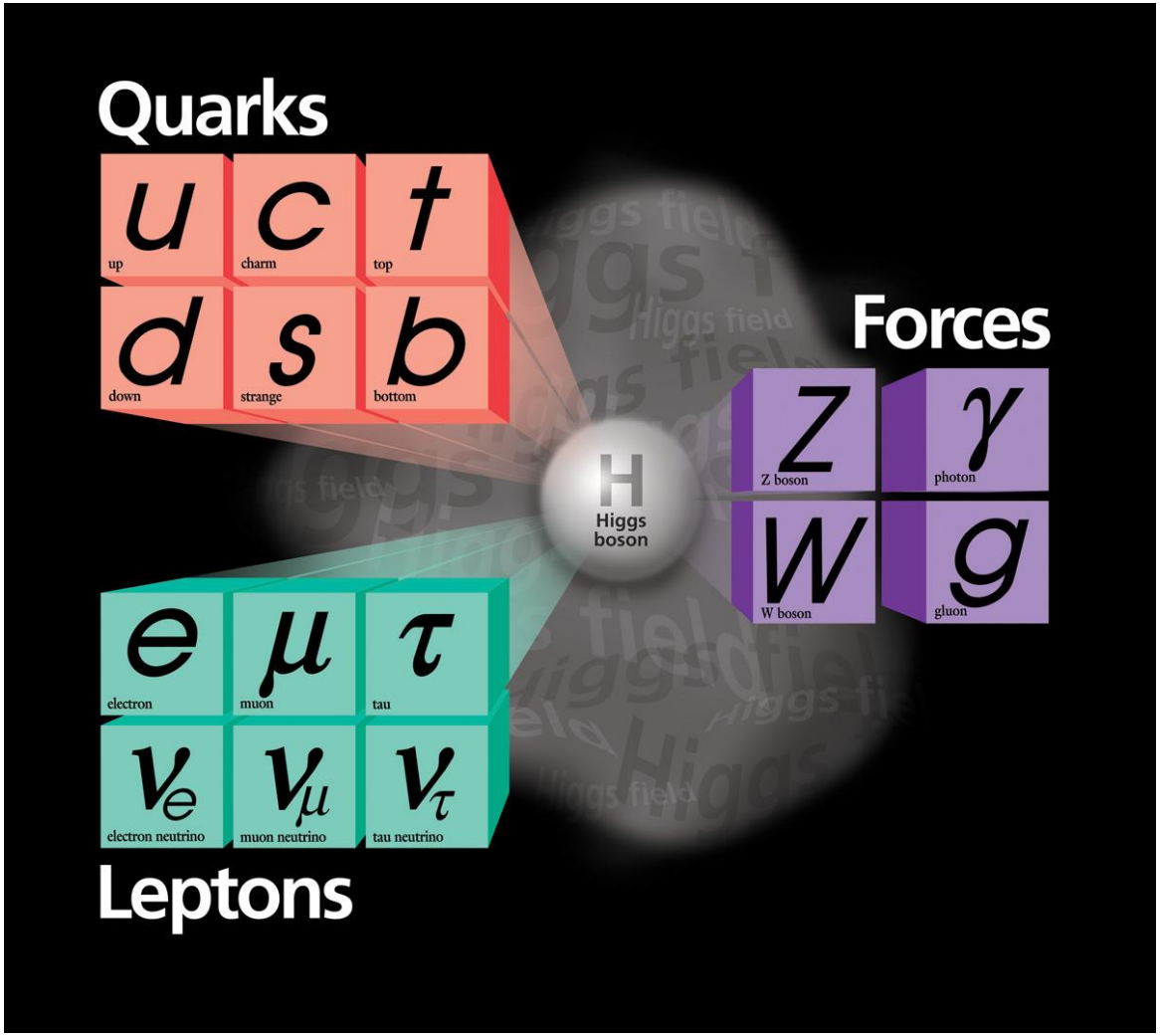
The Higgs Boson





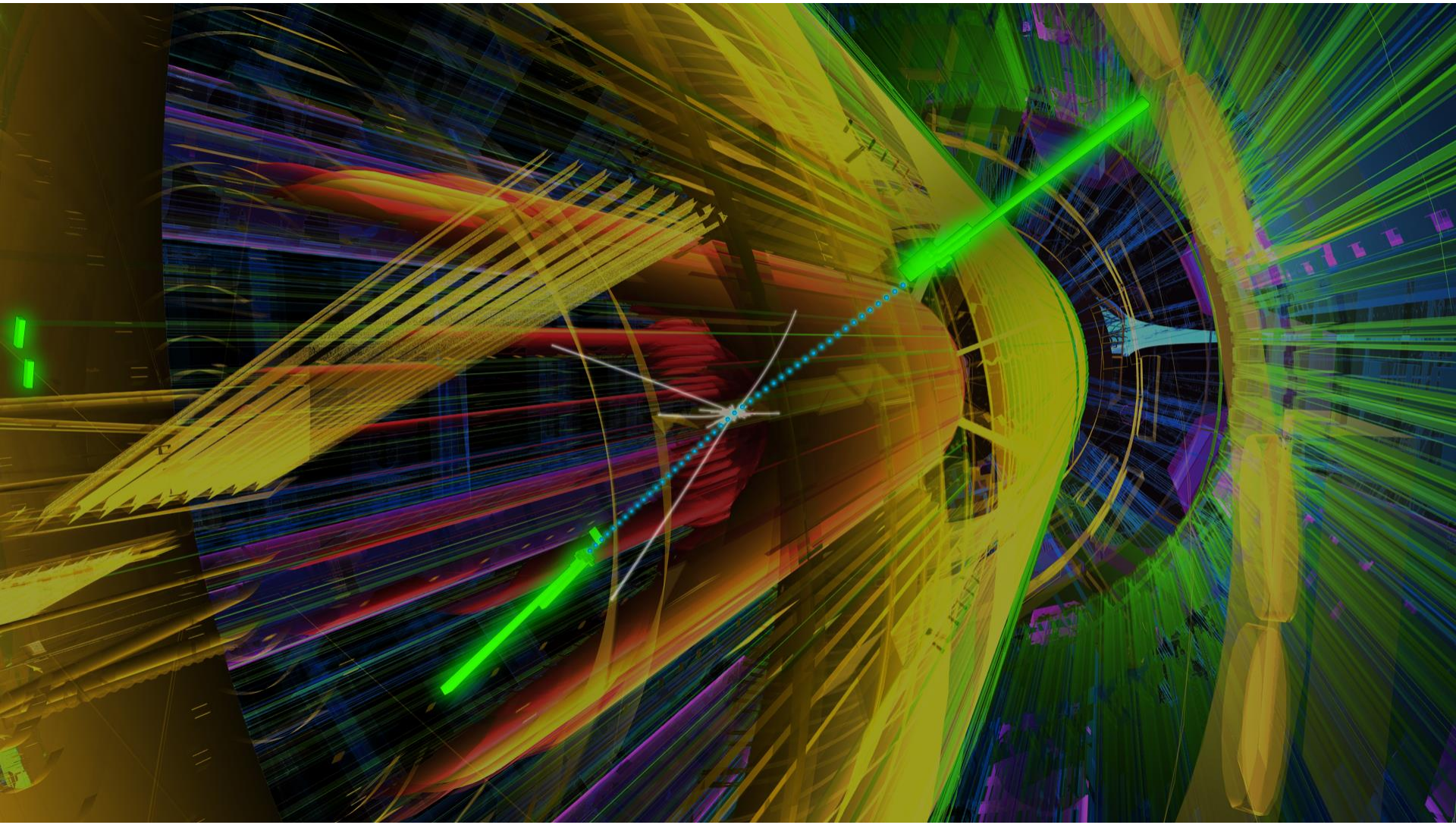
In 1964 Peter Higgs (and others) predicted the existence of a new particle (and field) that can explain how all other particles acquire mass

The Higgs Boson

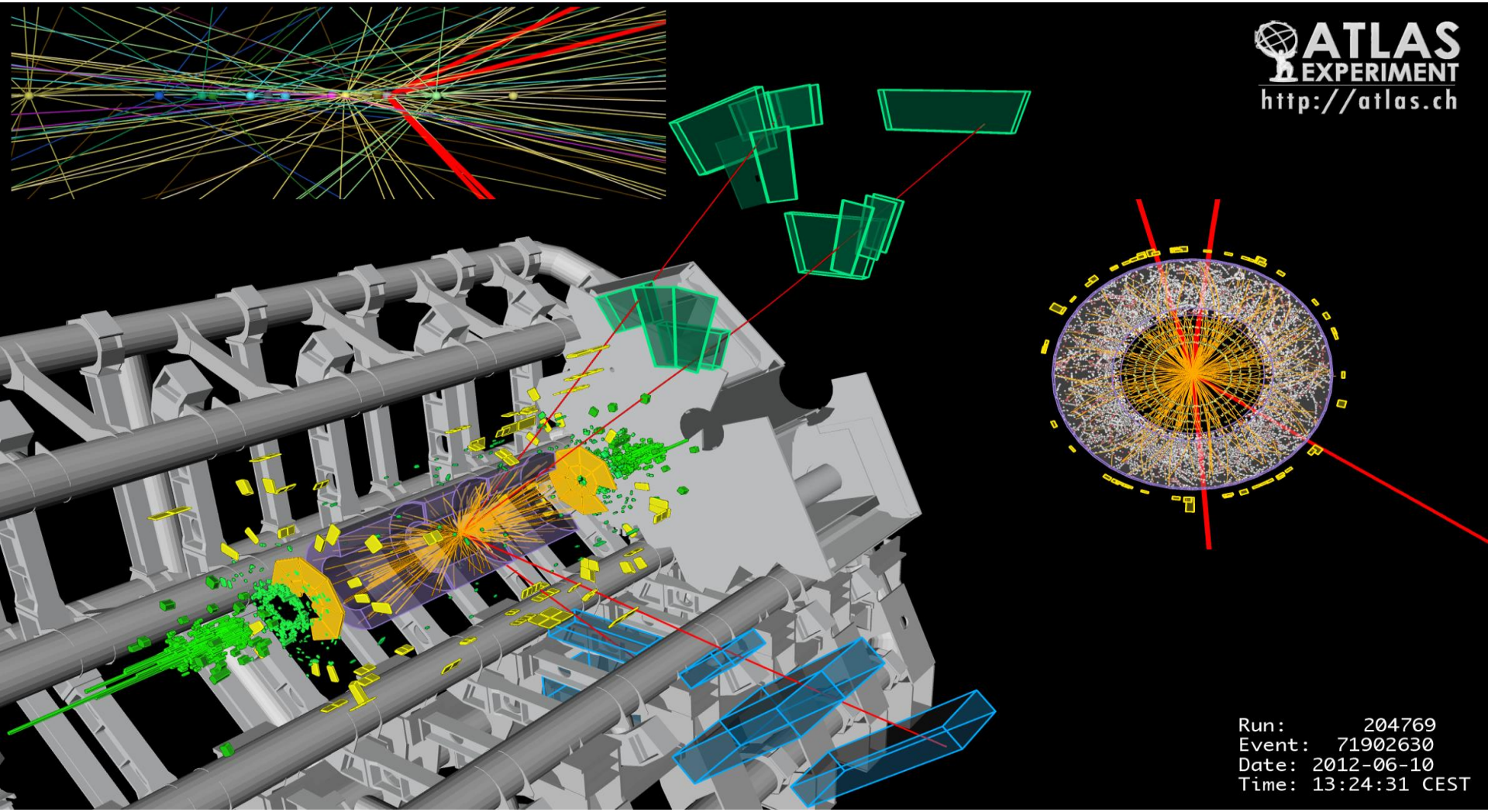


Higgs Boson Discovery

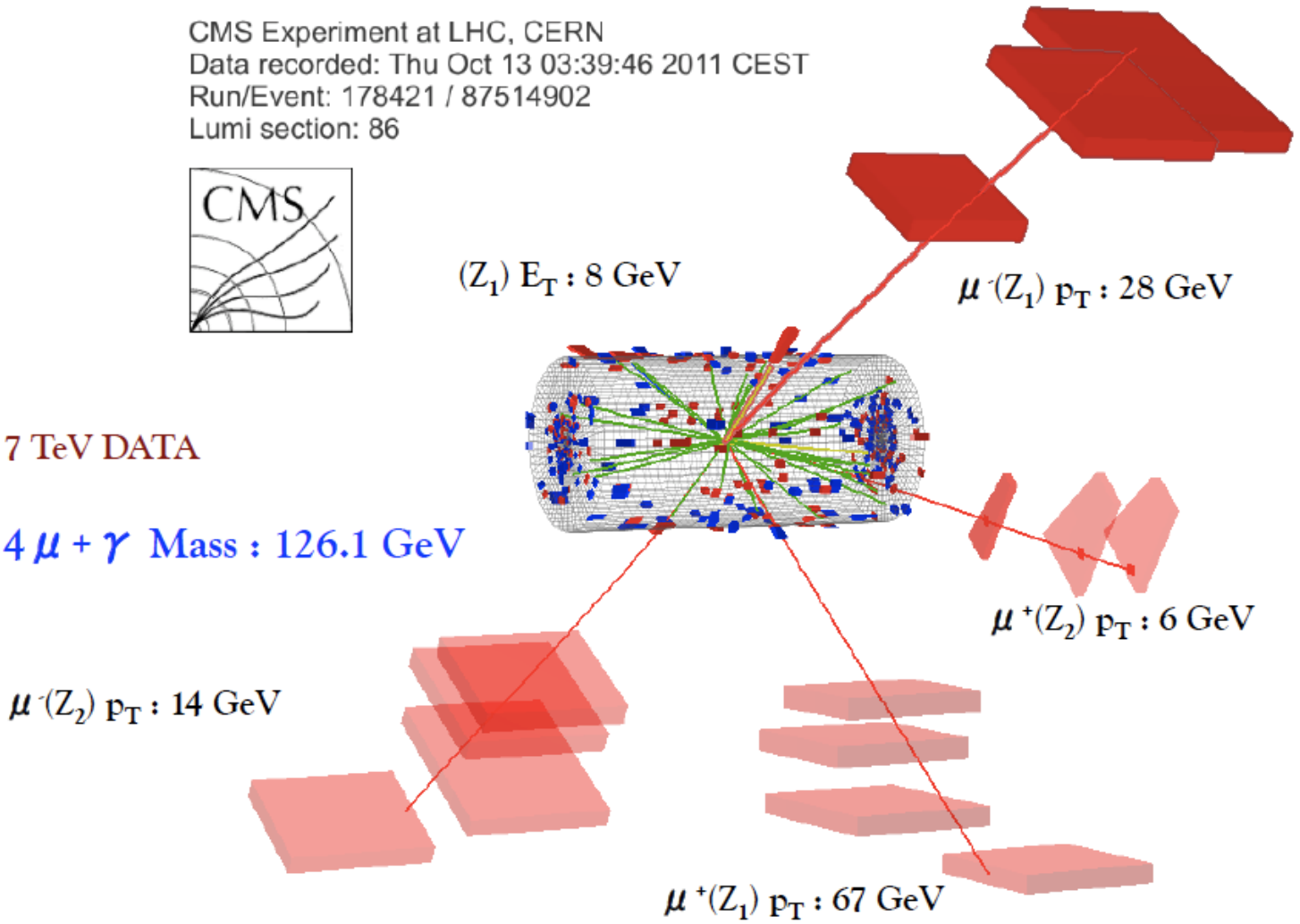
A Higgs Event



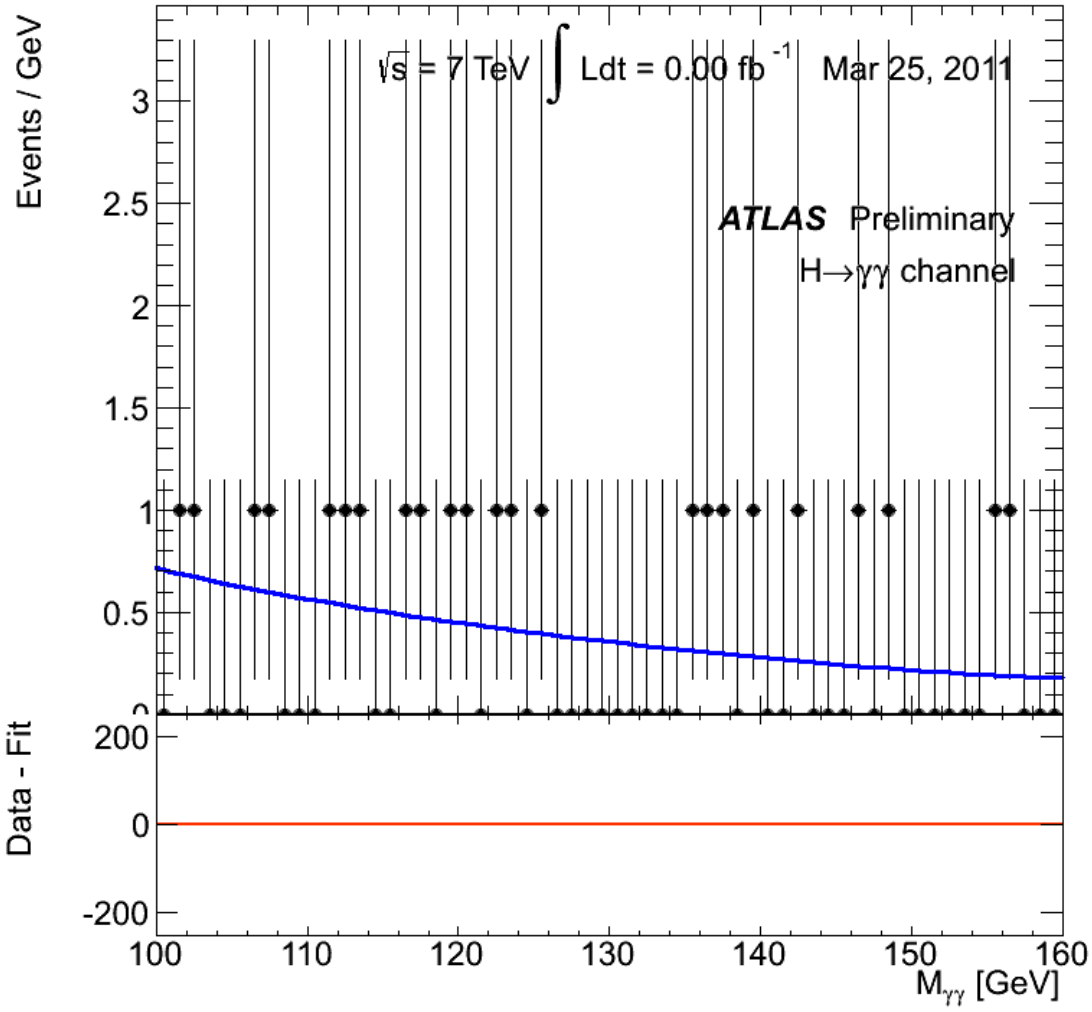
A Higgs Event



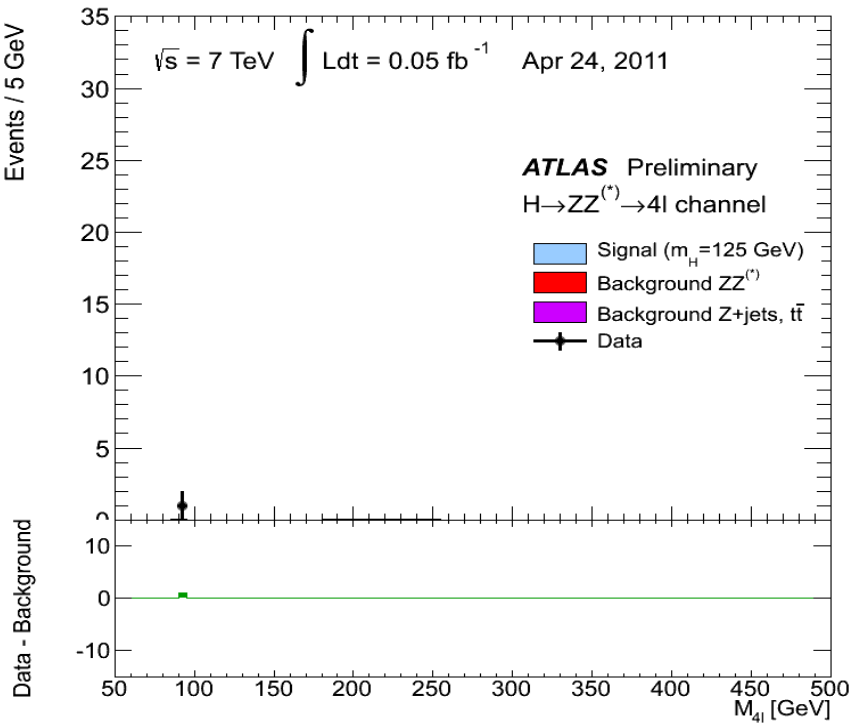
A Higgs Event



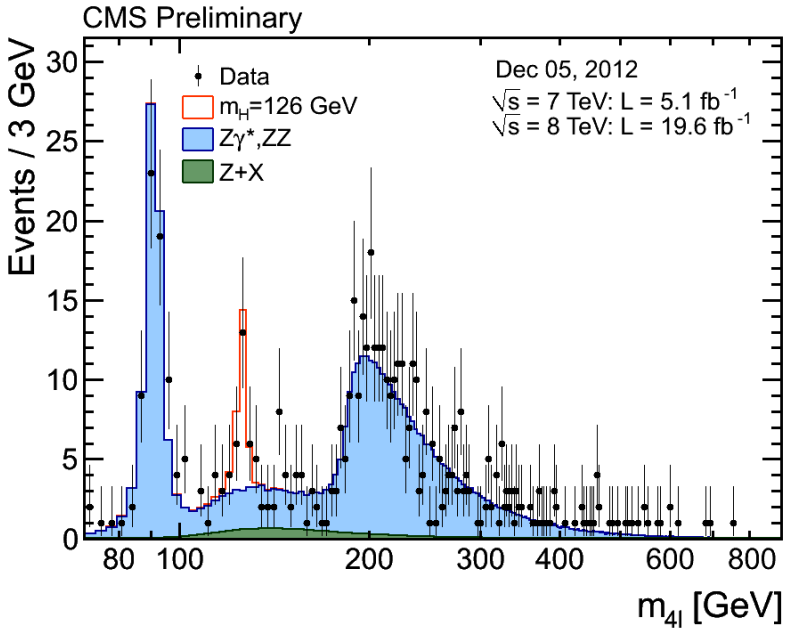
Higgs Boson


√s = 7 TeV ∫ Ldt = 0.00 fb -1 Mar 25, 2011
ATLAS Preliminary
H→γγ channel
Events / GeV
Data - Fit
M_γγ [GeV]

Higgs Boson

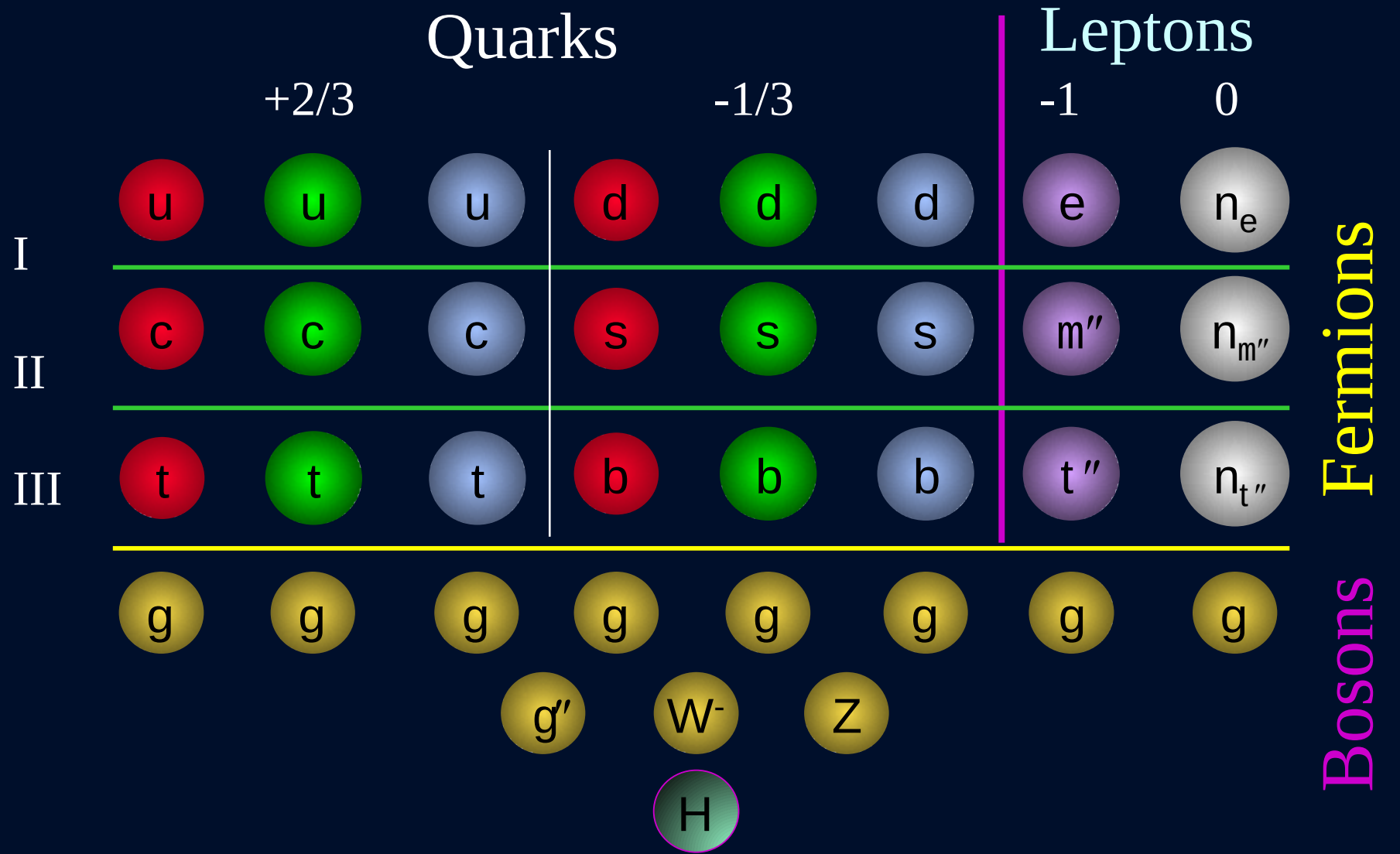


ATLAS

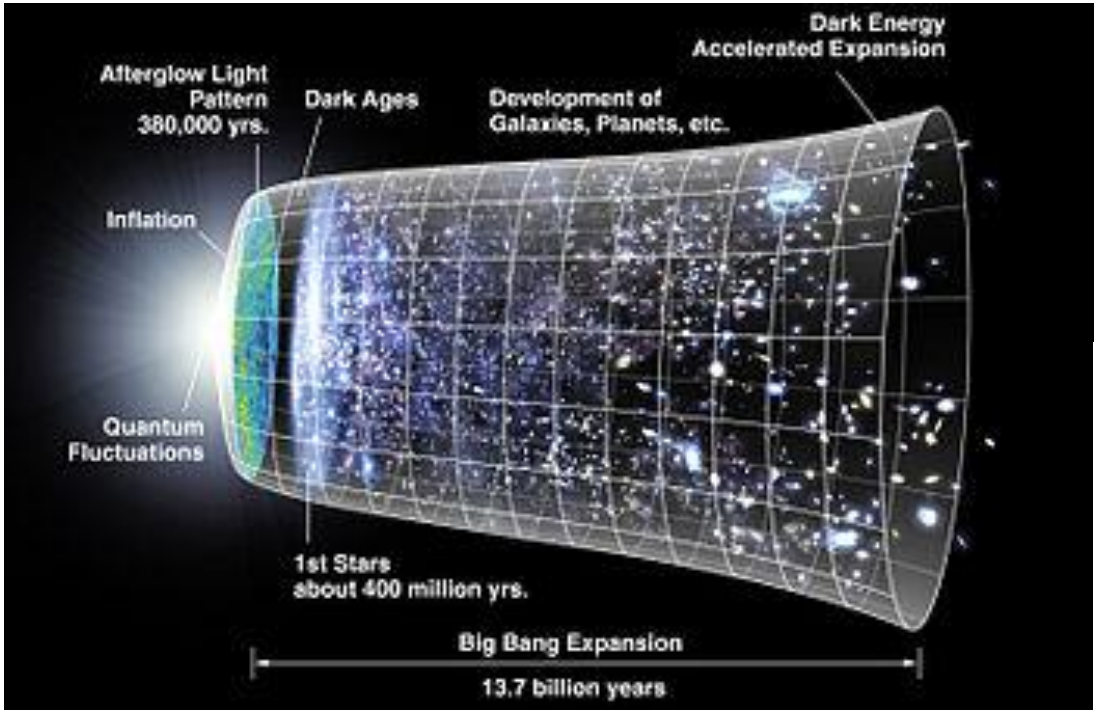


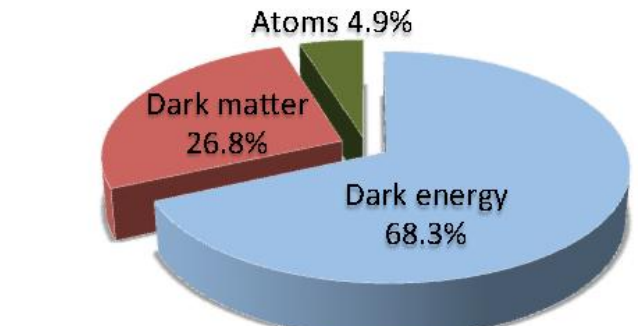
CMS

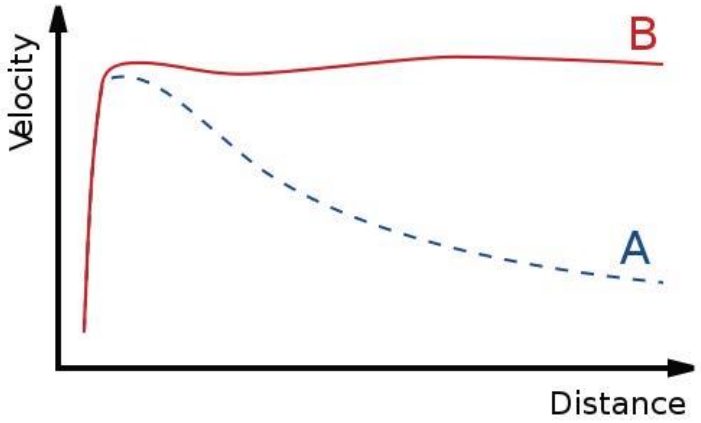
Standard Model (2016)



Energy Budget







Most of the stuff (95%) is currently missing from the Standard Model

Some Open Questions

What determines particle masses?

Are there new particles?

Where do observed symmetries come from?

Why is gravity so weak?

What is Dark Matter/Dark Energy?

Near Future

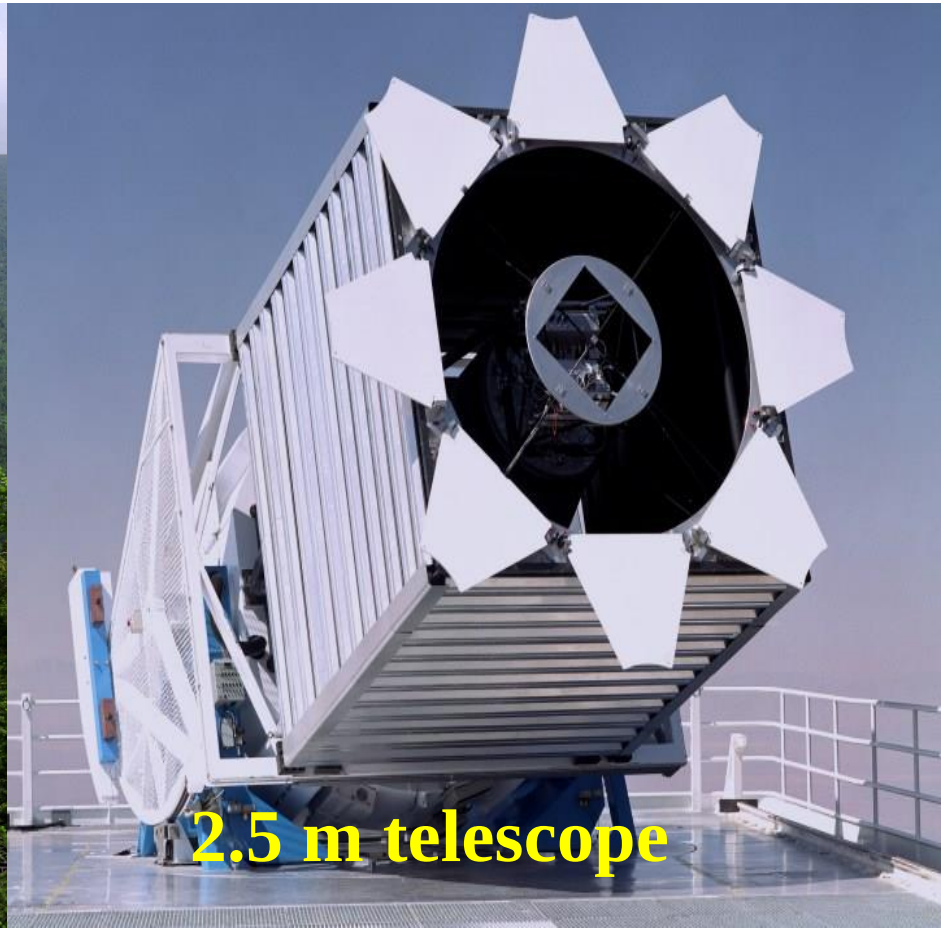
The Large Hadron Collider is again collecting data this summer after a short shutdown

Particle collision energy is 13 TeV (tera electron volts)

Higher collision energy means higher probability to create interesting (rarer) events. More data will hopefully bring more discoveries.

How Big is Big Data?

Sloan Digital Sky Survey



Collected more data in the first two weeks

than was collected in the history of astronomy

Large Synoptic Survey Telescope

3200 Megapixel camera

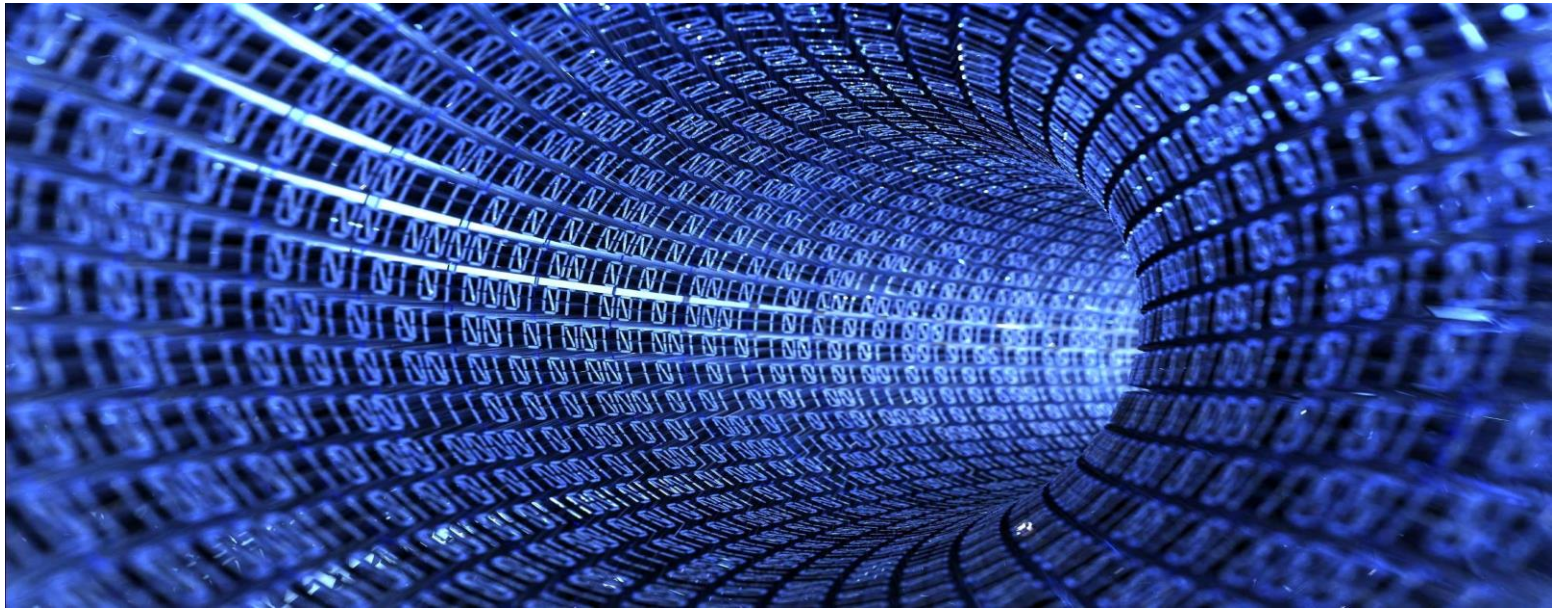
Will create a movie of the sky in different frequencies for ~10 years

Data-taking expected to begin in 2022

LSST
Large Synoptic Survey Telescope

Big Data

Project	Expected Data	Period
SDSS	100 Tb	2000 - 2015
LSST	100 000 Tb	2022 - 2032
LHC	15 000 000 Tb	2010 - 2035



Open Data

Office of Science and Technology Policy

About OSTP | Pressroom | OSTP Blog | Divisions | Initiatives | R&D Budgets | Reso

Expanding Public Access to the Results of Federally Funded Research

Subsc

Posted by **Michael Stebbins** on February 22, 2013 at 12:04 PM EDT

E-Mail Tweet Share +

The Obama Administration is committed to the proposition that citizens deserve easy access to the results of scientific research their tax dollars have paid for. That's why, in a policy memorandum released today, OSTP Director John Holdren has directed Federal agencies with more than $100M in R&D expenditures to develop plan to make the published results of federally funded research freely available to the public within one year of publication and requiring researchers to better account for and manage the digital data resulting from federally

LHC Open Data

CERN Accelerating science

Sign in Directory

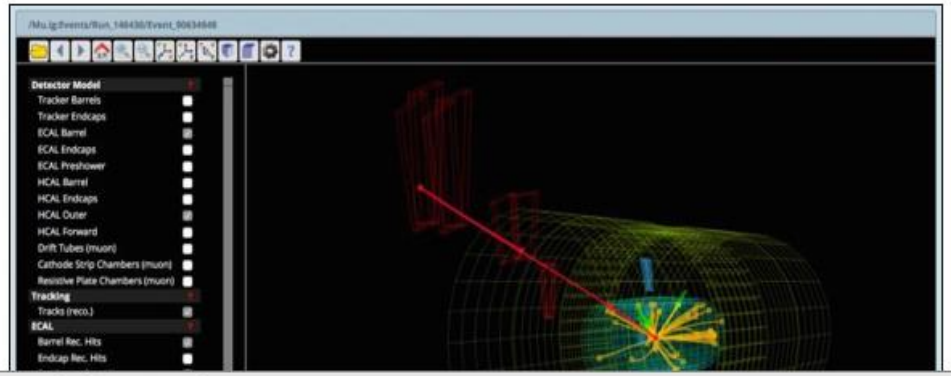
About CERN | Students & Educators | Scientists | CERN people

Accelerators | Experiments | Physics | Computing | Engineering | Updates

CERN makes public first data of LHC experiments

Posted by Cian O'Luanaigh on 20 Nov 2014. Last updated 20 Nov 2014, 16.59.

Voir en français



ABOUT CERN

About CERN

Computing

Engineering

Experiments

How a detector works

CERN UPDATES

LHC filling with liquid helium at 4 kelvin

17 Dec 2014

CERN's Large Hadron Collider gears up for run 2

LHC Open Data

Since 2014 CMS has released all its collected collision data from 2010 and 2011 runs

opendata
CERN

http://opendata.cern.ch

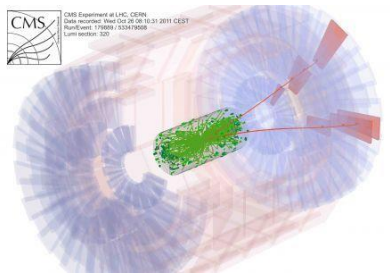
- **Same format as used by physics analyses with tools and basic instructions**
- **also some simulated (Monte Carlo) data**

JUPYTER Notebooks

- Open source web-based application blends code with elements such as text, figures, links
  - Excellent integration of instructions and executable code
    - Great for interactive analysis and teaching demonstrations of anything that involves code
  - Can be run locally, on a server, laptop or smartphone
    - All you need is a browser

Today

- **Start with Python programming using Jupyter notebooks**
  - **A short primer on ROOT**
- **Useful Particle Physics Concepts**
- **LHC Event Displays**



- **Today, Tomorrow:**
  - **Basic Examples of CMS Open Data Analysis**