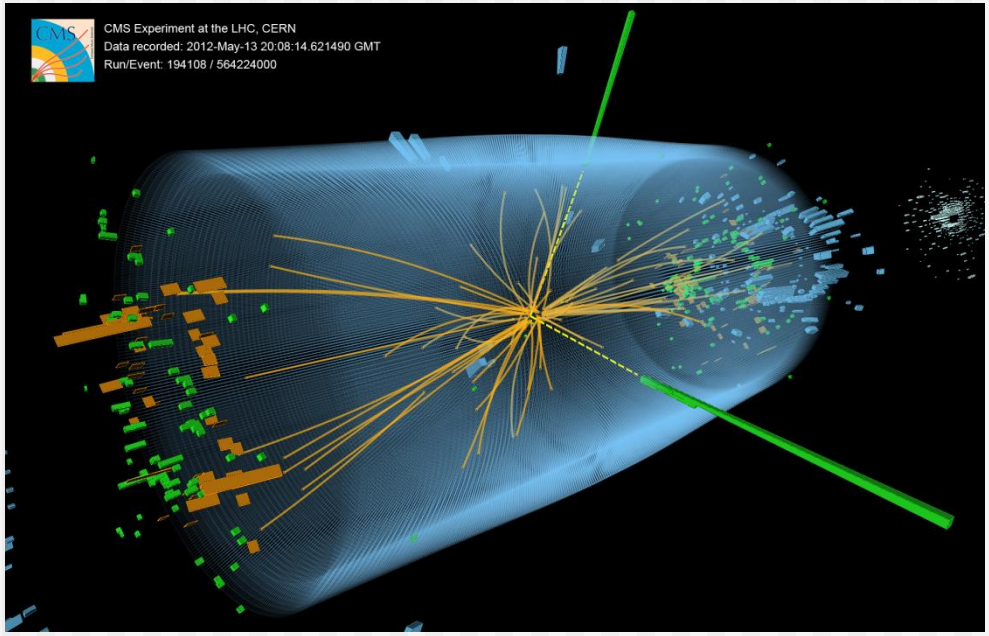
LHC Discoveries through Data Analysis



Christoph Grab (ETHZ)

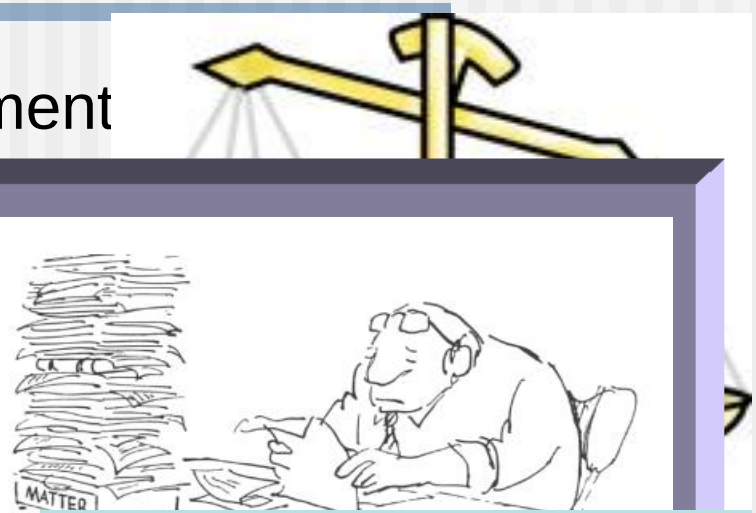
SOS17 Conference, Jekyll Island, March 26, 2013

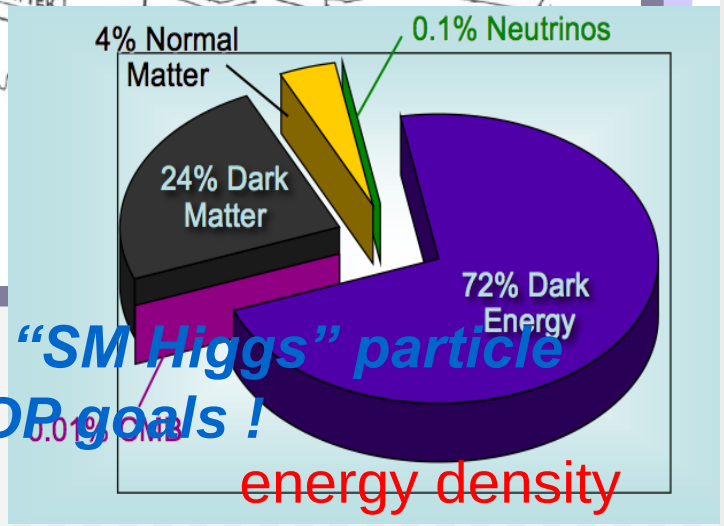
Contents

- What is the physicist's motivation?
- What are the challenges we face ?
- How did we solve it?
- What did we learn so far ?
- Outlook

Note: most examples are from CMS, but concepts are valid for all experiments ….

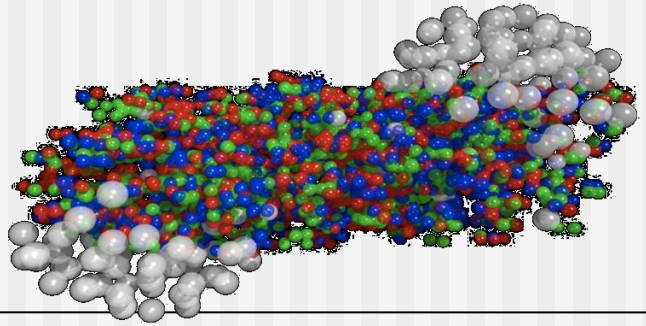
What are the Scientific Questions ?





Finding and measure the "SM Higgs" particle is one of the TOP goals !

- The Large Hadron Collider LHC at CERN is the machine, that allows to explore these uncovered territories, and hopefully provide some answers !

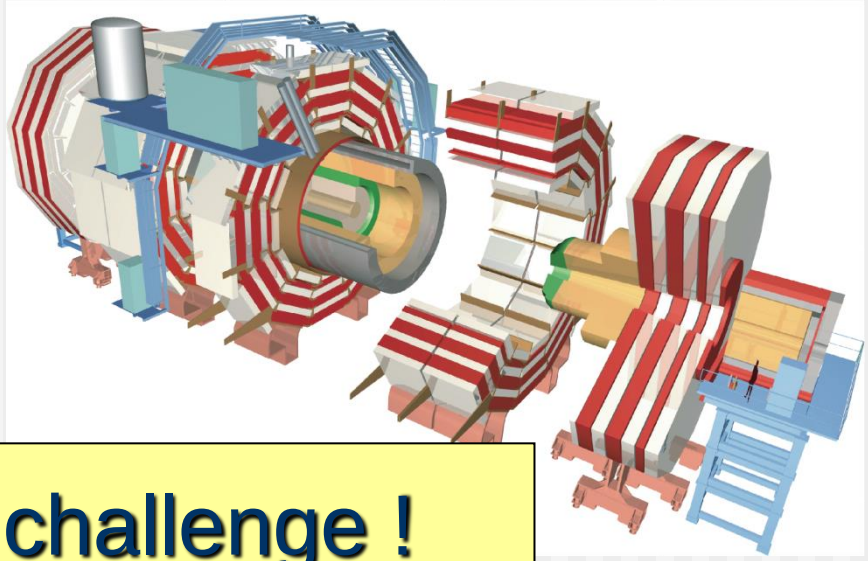


LHC – the tool for breakthroughs !

- Large Hadron Collider (LHC) collides protons at 3.5/4.0 TeV at typically 40/20 MHz
  ➔ a NEW regime

- Four experiments/detectors, with up to 3500 persons/experiment
  [Swiss participation in the three ATLAS, CMS and LHCb]

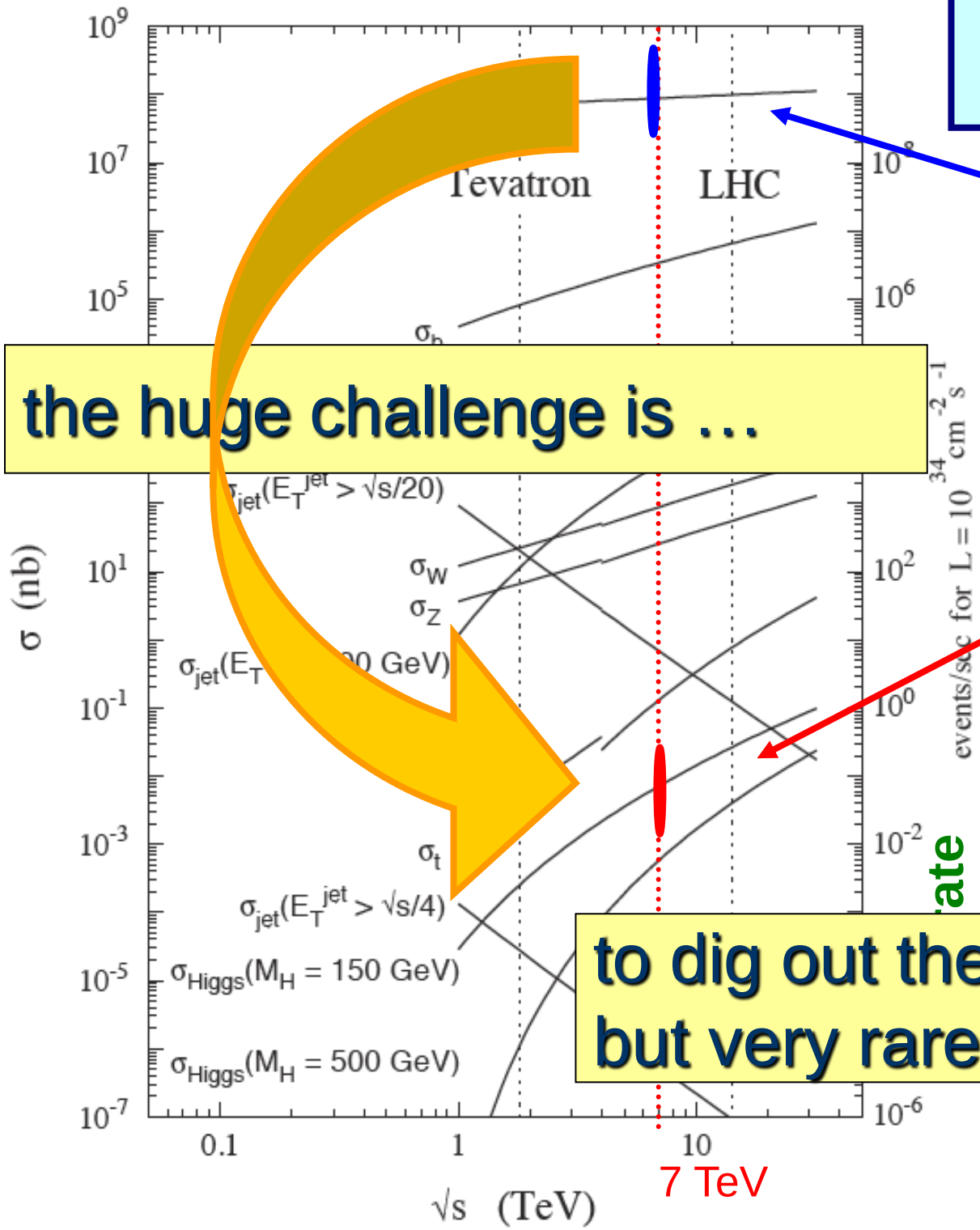
- Experiments record the particles produced in these collisions



- *Physicists around the world want to access and analyze these data as fast as possible to make discoveries !*

That's a challenge !

Cross sections ~ "Reaction probability"



the huge challenge is …

a large total reaction rate

σ_{tot} (pp) ~ 10^{11} pb

Searches for "new"

e.g. Higgs

very low rates

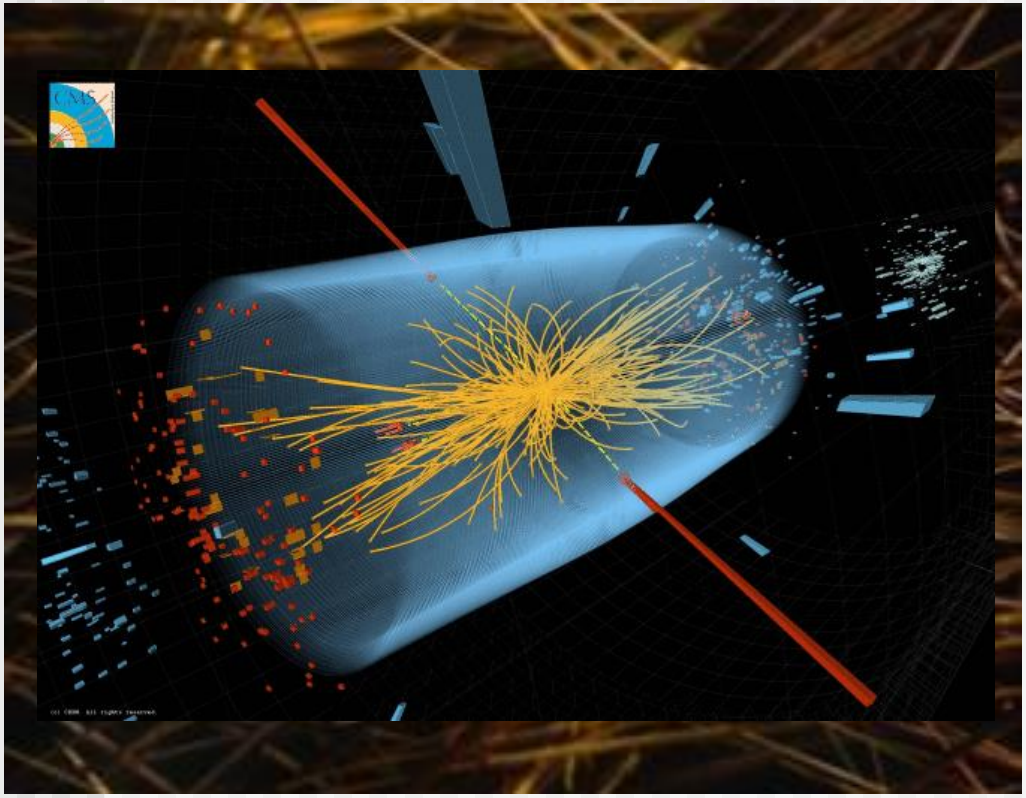
Only one out of a billion may be an exiting new event !

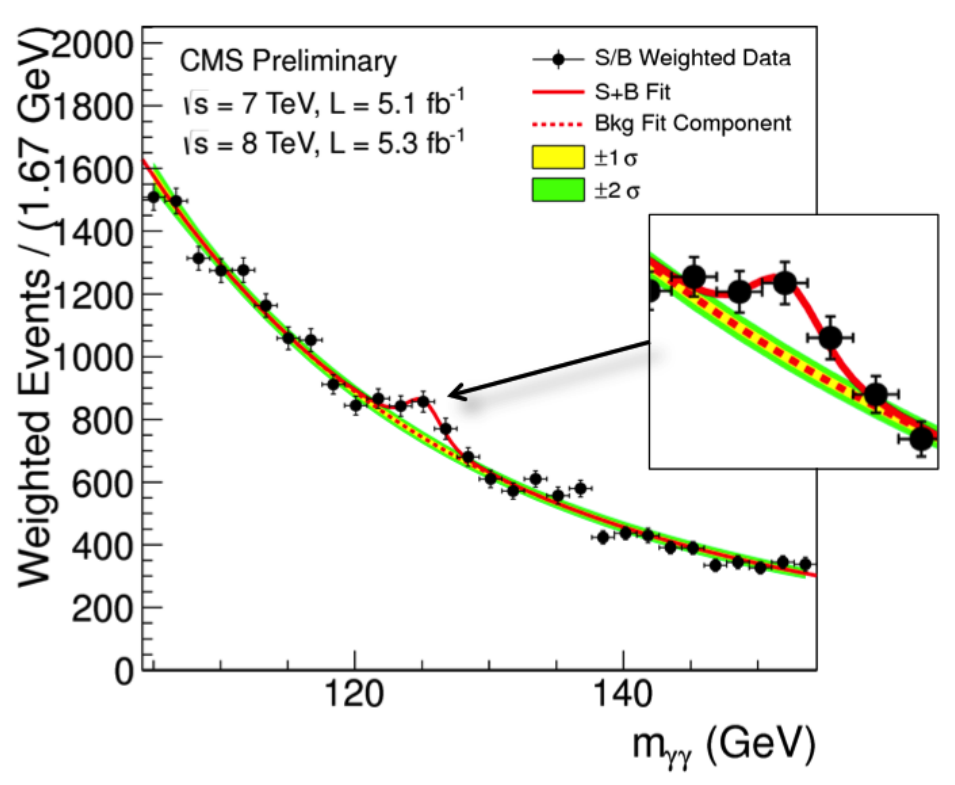
to dig out the very interesting, but very rare processes …

1 barn = 10^{-24} cm^2 ;

1 pb = 10^{-12} barn = 10^{-36} cm^2

Higgs: needle in the haystack…..



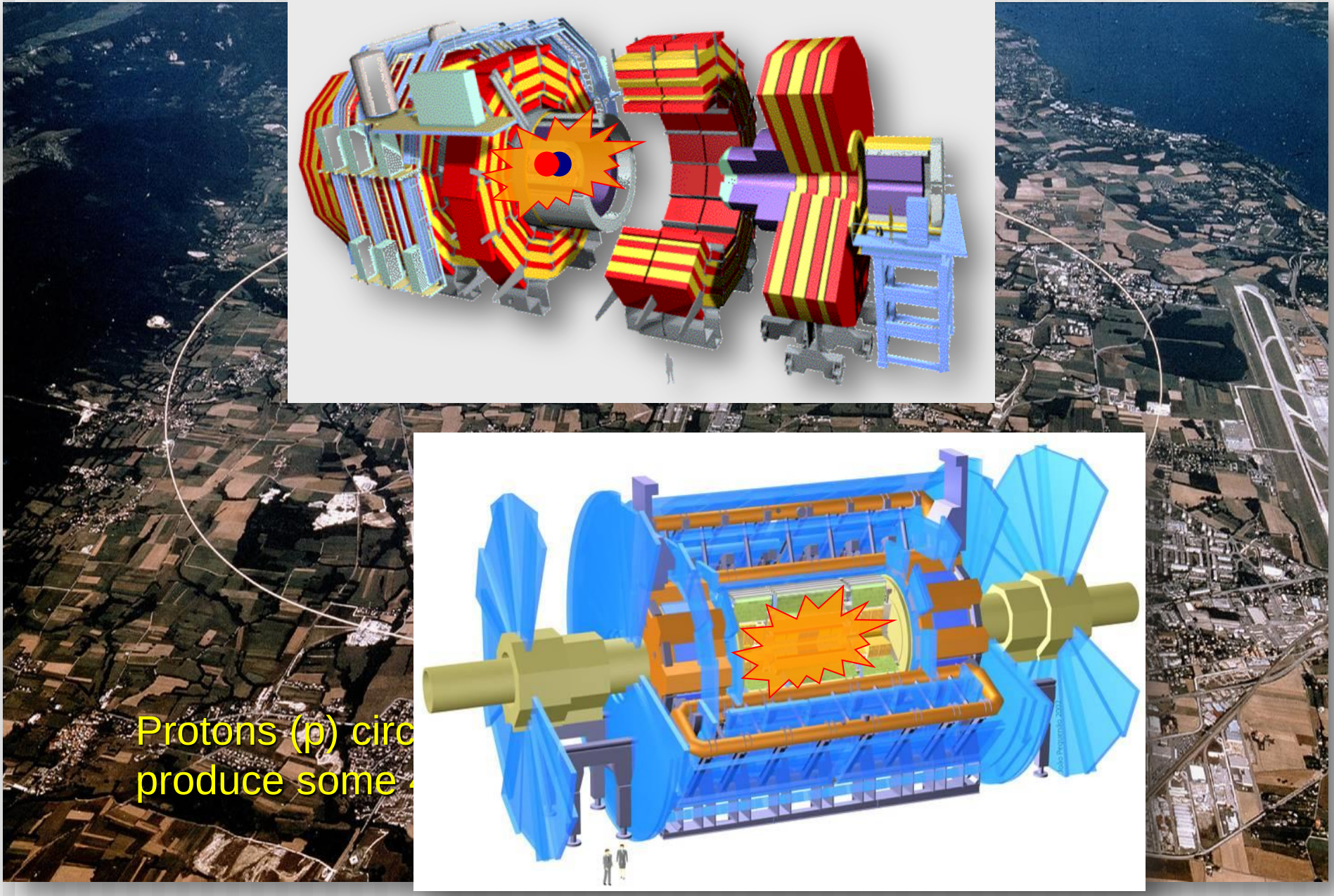


Every 50 nano-second new collisions….

Only about **ONE in a billion** is a really interesting reaction, e.g. a Higgs boson …

And may show up as a tiny signal on top of a large (highly reduced) background… not to be missed !

Large Hadron Collider (LHC)



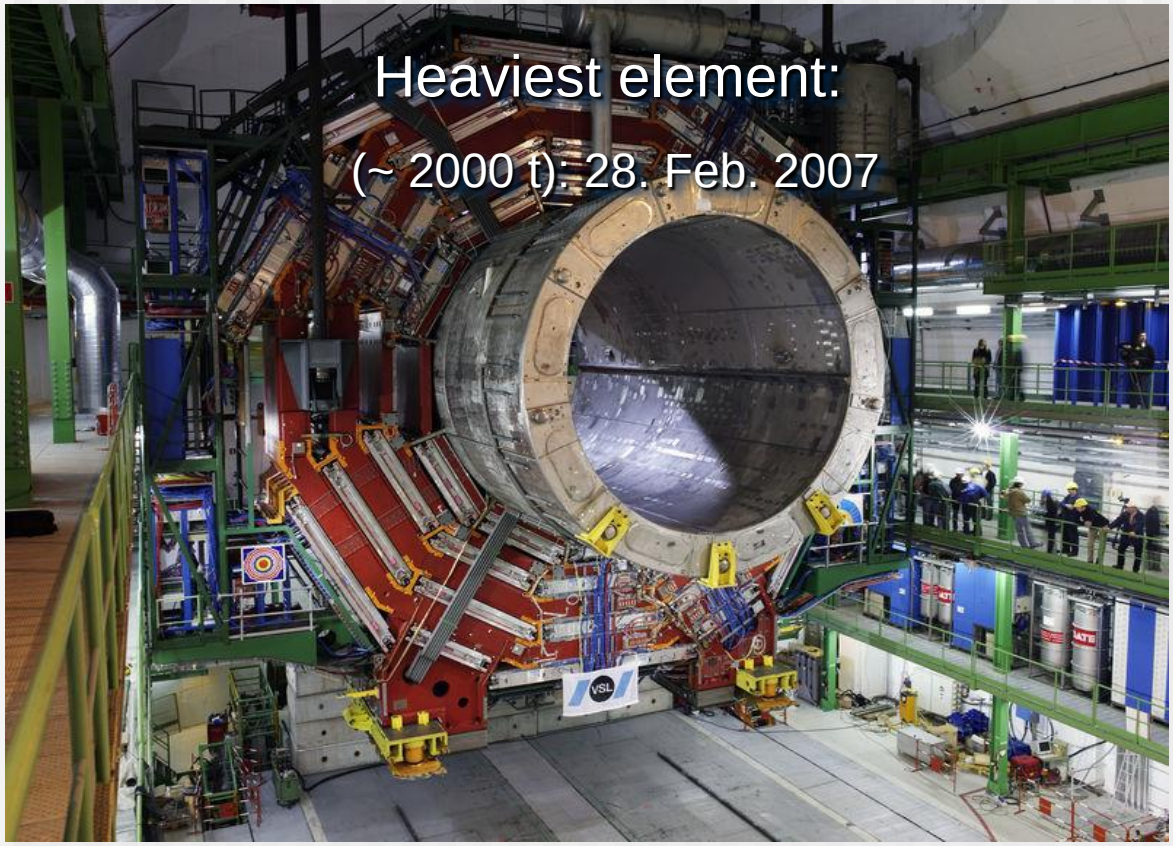
LHC Tunnel

CMS Detector in the Cavern

CMS collaboration: 41 Countries, 179 institutes

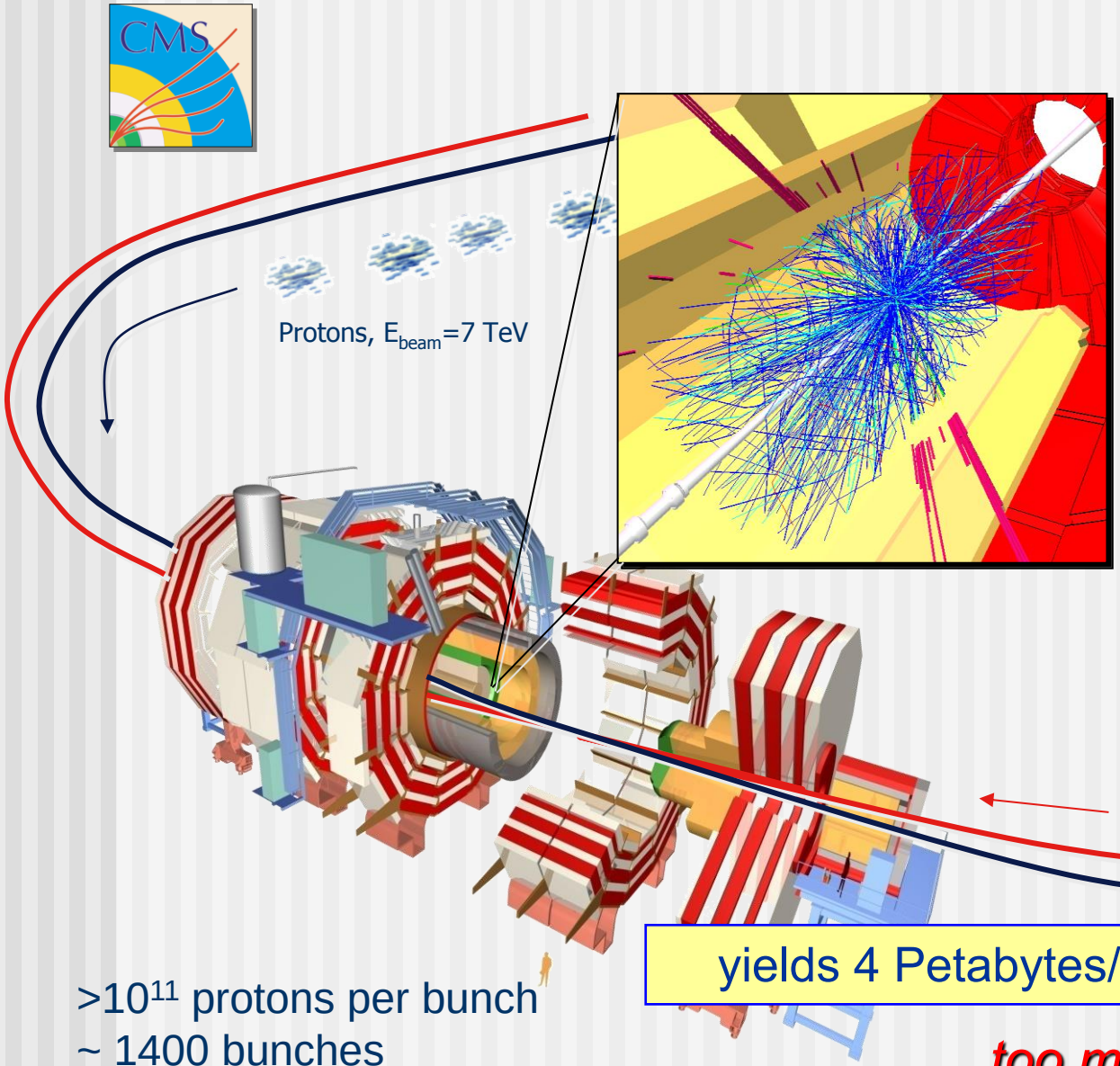
~3000 Authors including ~2200 PhD's and ~800 PhD students

CMS installation, 100 m below ground



Weight:	**12,500 t**
Diameter:	**15 m**
Length:	**21.6 m**
Field :	**4 Tesla**

Data from Proton-Proton Collisions (CMS)



At every p-p collision detector records all reaction products

Total ~ 100 million channels
~digital camera with 100 Mio Pixel, that records a picture 40 million times per second

Physicists analyze these events ("picture")

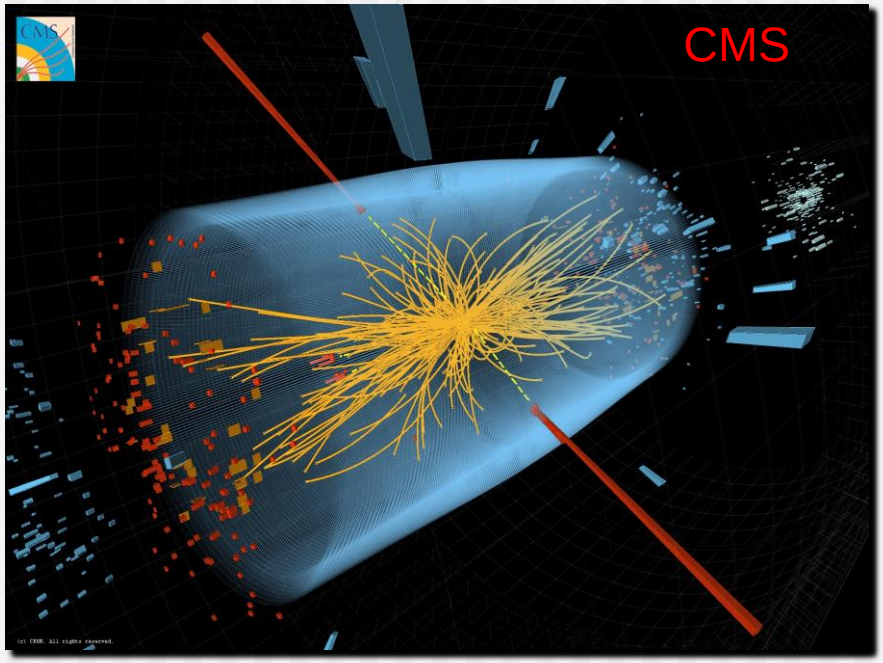
yields 4 Petabytes/sec = 4'000' 000'000' 000'000 B/s

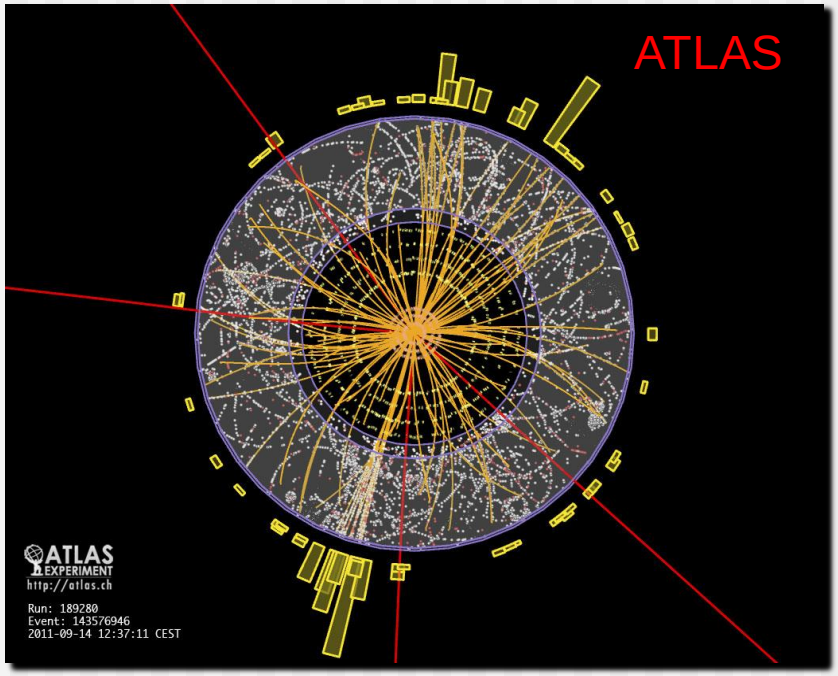
too much to be stored …

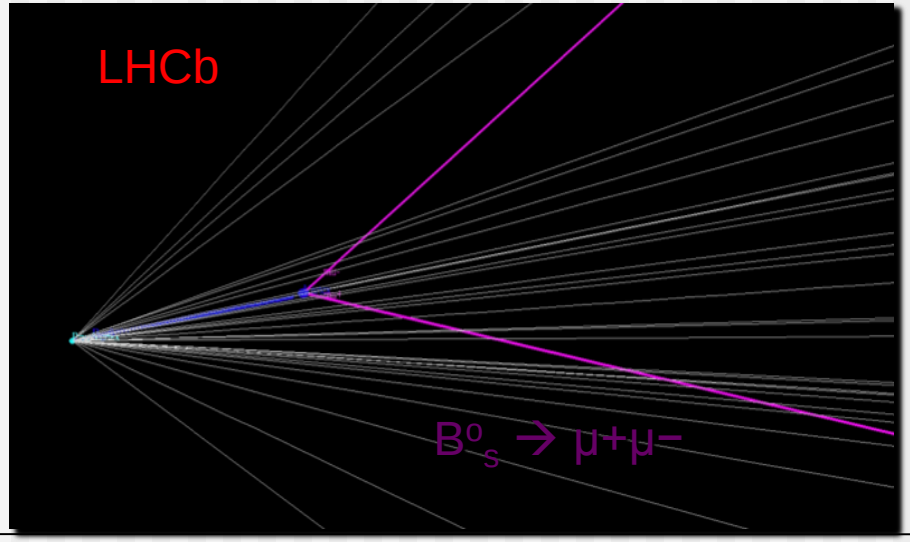
$>10^{11}$ protons per bunch
~ 1400 bunches
Collisions : $40 \cdot 10^6$ / sec

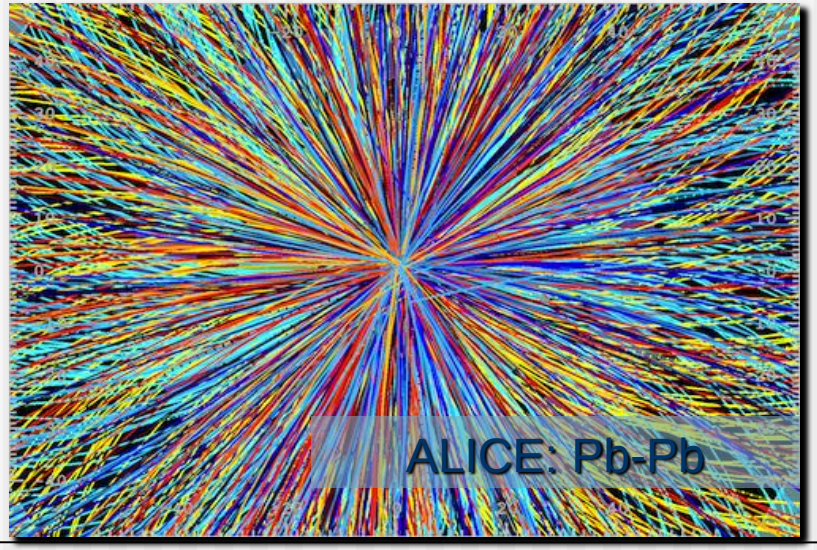
Data Harvest : high quality Event Pictures

Graphical representations of reaction products measured with the detectors

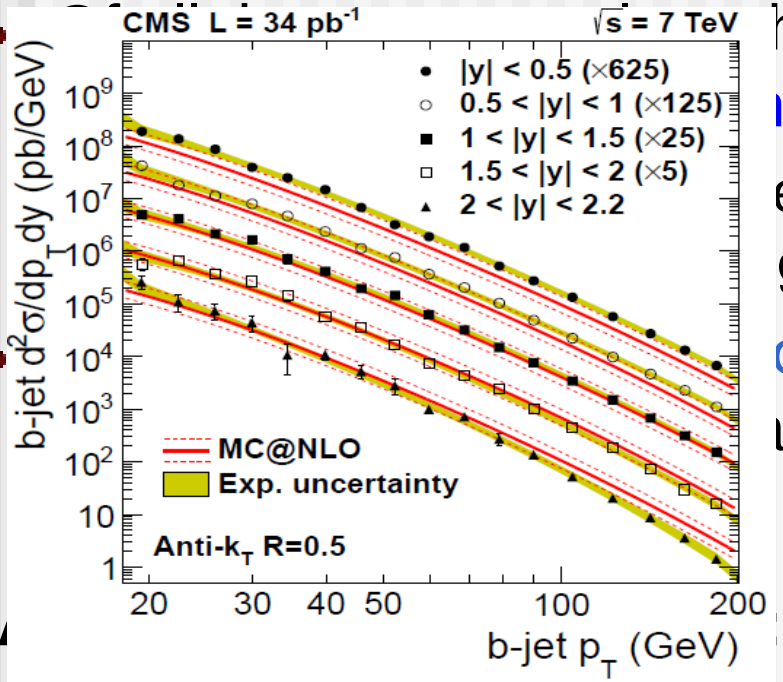




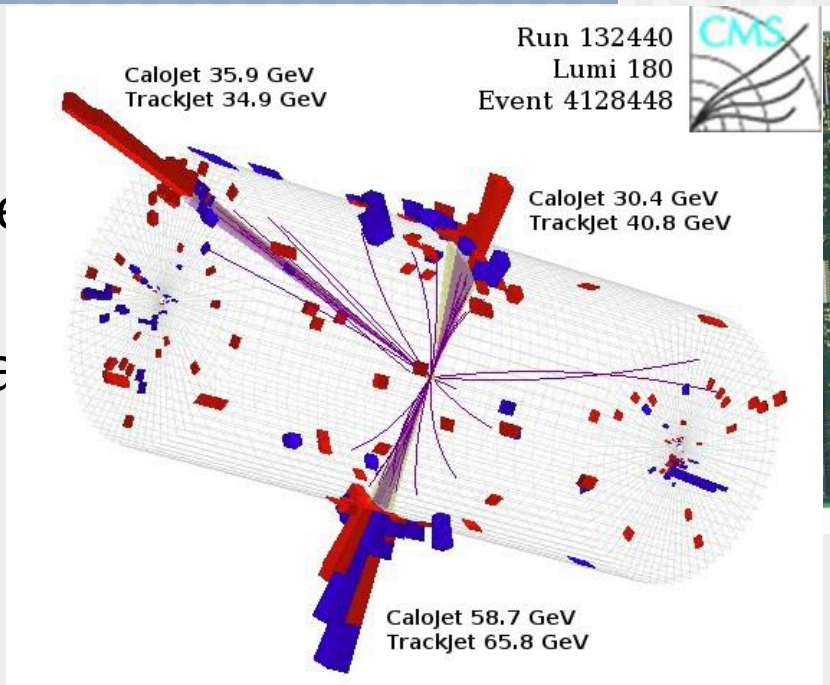




Overview of Data Reduction and Analysis



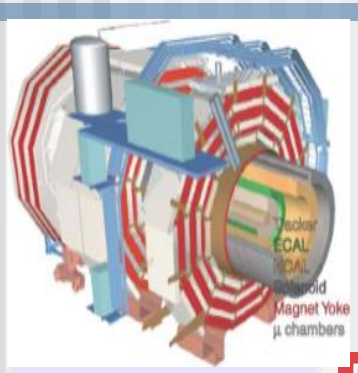
han ONE
ng" .
ed, intellige
gger"
ored on ta
analysis.



- reconstruct topologies → classify and "count" event types.
- Use self-developed optimized algorithms.
- Study statistical distributions of event classes, compare to theory.
- Need a good description of "known processes", and precise simulation and modeling of detector properties to get efficiencies and predictions.

- Every event is independent of the others → massively parallel process
- Distribute and analyze events worldwide.

Online Data Flow (CMS)



L1 Trigger
custom hardware, operates parallel

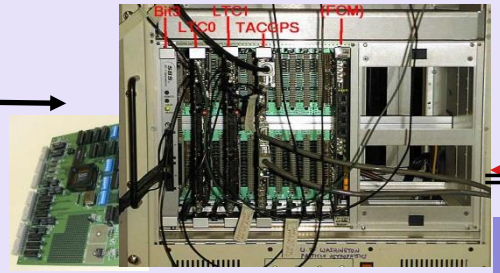
L2 + High Level Trigger
embedded proc. + commodity CPU farm

Offline data storage

Rate : 40 MHz
~1 PB/s

Rate: 75 kHz
75 GB/sec

Rate: 500 Hz
~500 MB/s



Combine detector info

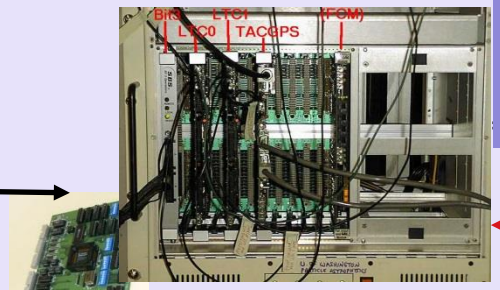
Write yes/no

Detector Front-End Electronics

Upper Level Readout 2

Analyze subdetector info

Keep event? yes/no



Reduction of: 10^5 in eventrate and 10^7 in datarate

Offline Analysis Chain (CMS)

data storage

reconstruction, calibration, alignment

Condense and store high-level data

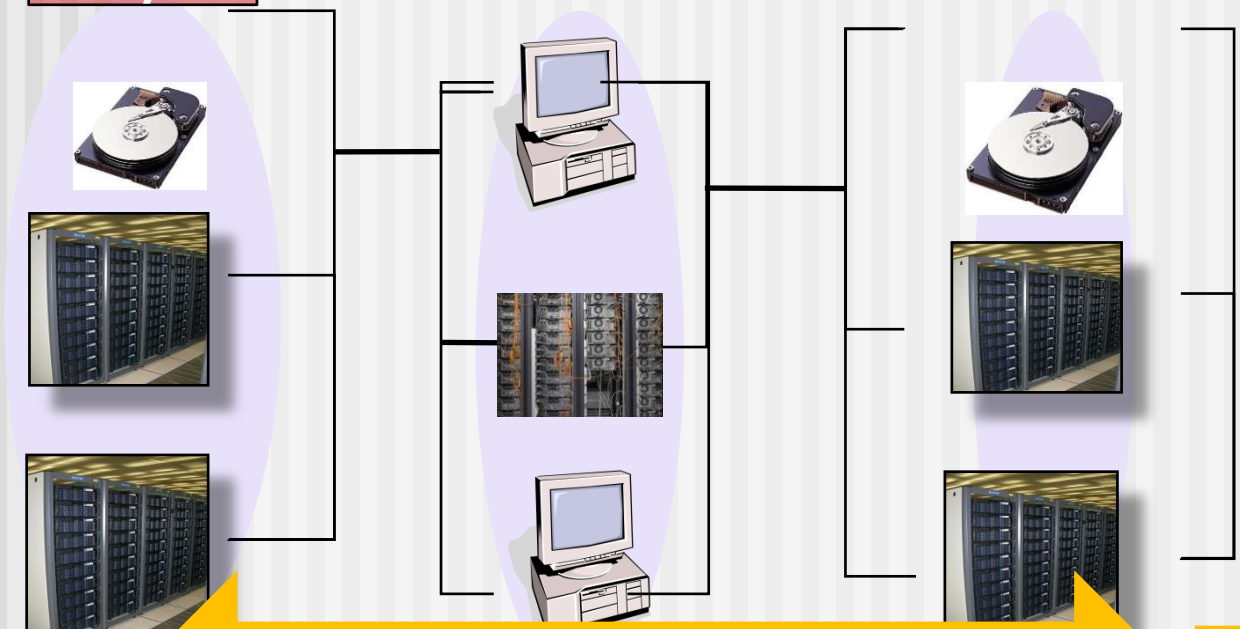
Individual physics analysis, algorithms, statistical methods

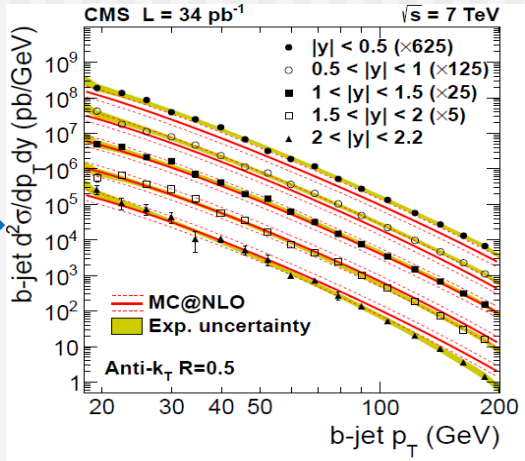
Data size: PByte

Tbyte

Gbyte

MByte





Common framework
centrally coordinated data access

Individual User/group Analysis
Chaotic data access

all the same steps for SIMULATION too ….

Analysis model mapped onto ..

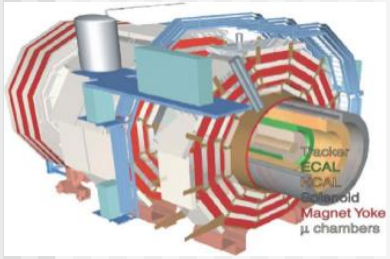
Hierarchical Structure -"WLCG"

Analysis model mapped onto

"WLCG =worldwide LHC Computing Grid"

Offline farm

Online system
DAQ, TRIGGER



Tier-0

CERN computer centre

T0: store raw data, reconstruct + distribute

Tier-1

National centres

FZK, USA, RAL, Italy, France

T1: store + reduce, re-reconstruct, distribute

Tier-2

Regional groups

CSCS, PISA, DESY, London

T2: simulations, group+user analysis

Tier-3

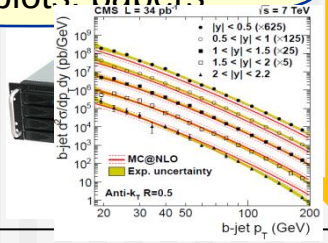
Institutes

ETHZ, PSI, UBE, UGE, EPFL

T3: end-user analysis

Servers+ laptops

T4: final plots, papers

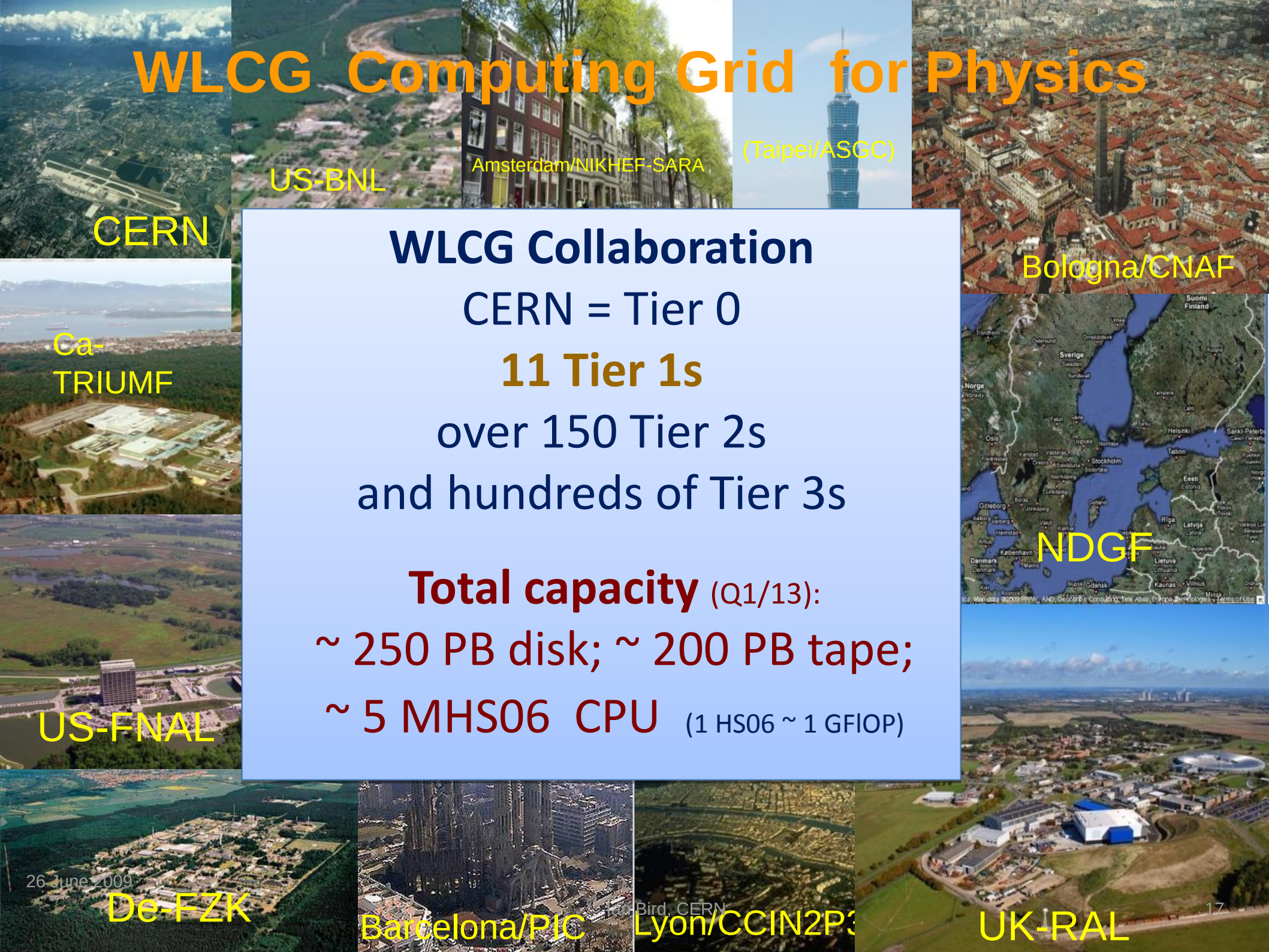


Common framework

Group + user framework

Useful model for Particle Physics but not necessary for others

WLCG Computing Grid for Physics



WLCG Collaboration

CERN = Tier 0

11 Tier 1s

over 150 Tier 2s

and hundreds of Tier 3s

Total capacity (Q1/13):

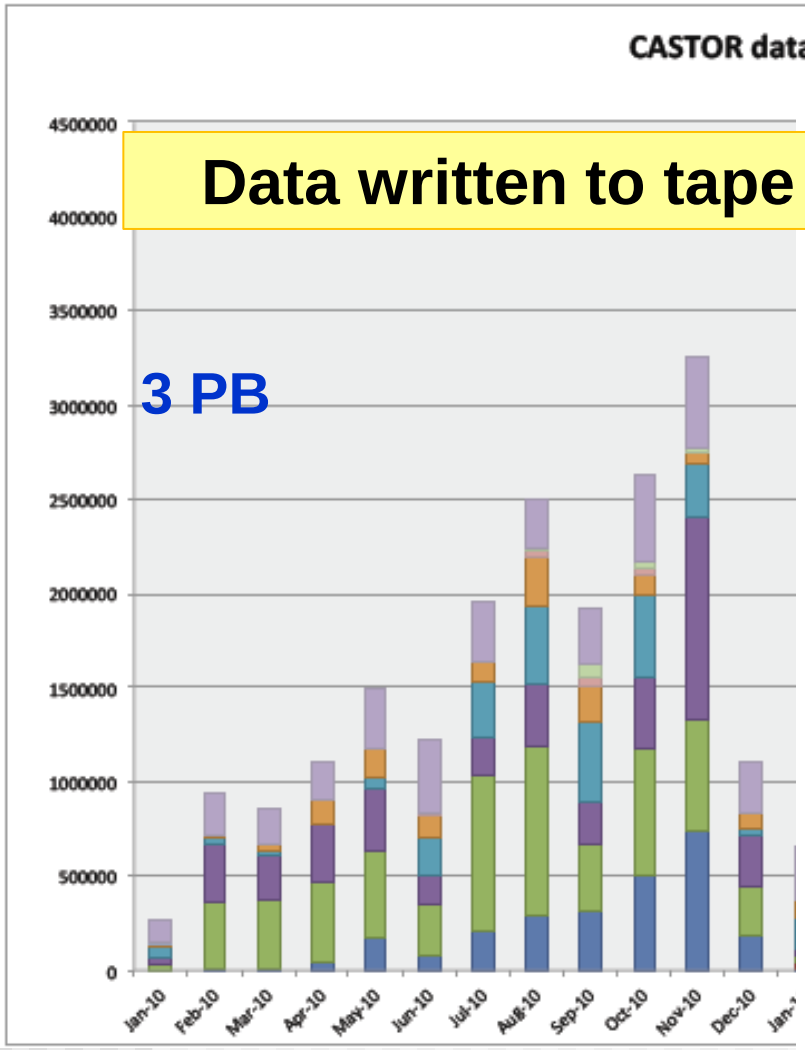
~ 250 PB disk; ~ 200 PB tape;

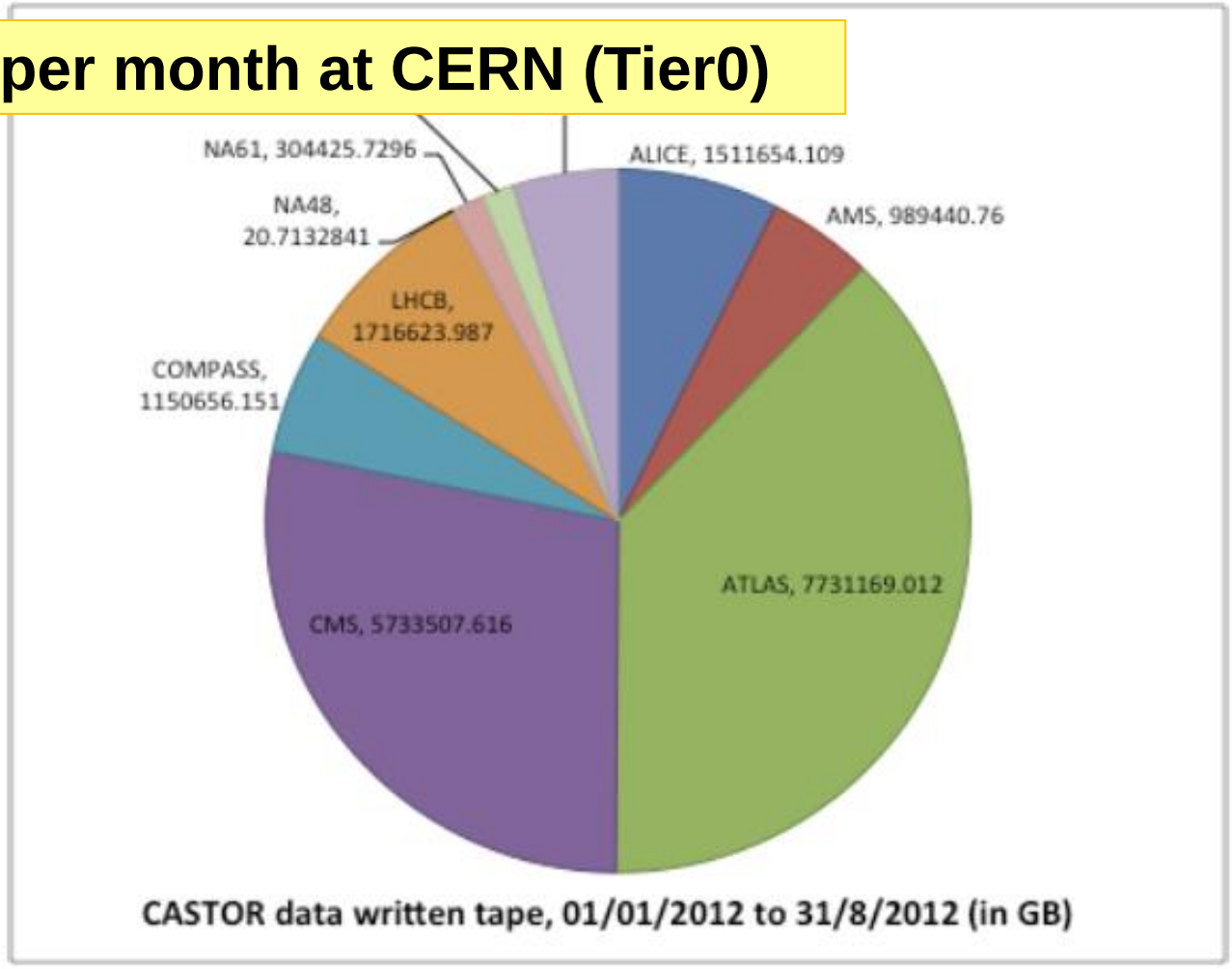
~ 5 MHS06 CPU (1 HS06 ~ 1 GFlOP)

A few real numbers …

Some Numbers - Storage Writing at CERN

- Data written at Tier-0 to tape ~ 1 PB per week
- Total data written in 8 months ~ 20 PB
- CERN passed 100 PB written on 3.2013

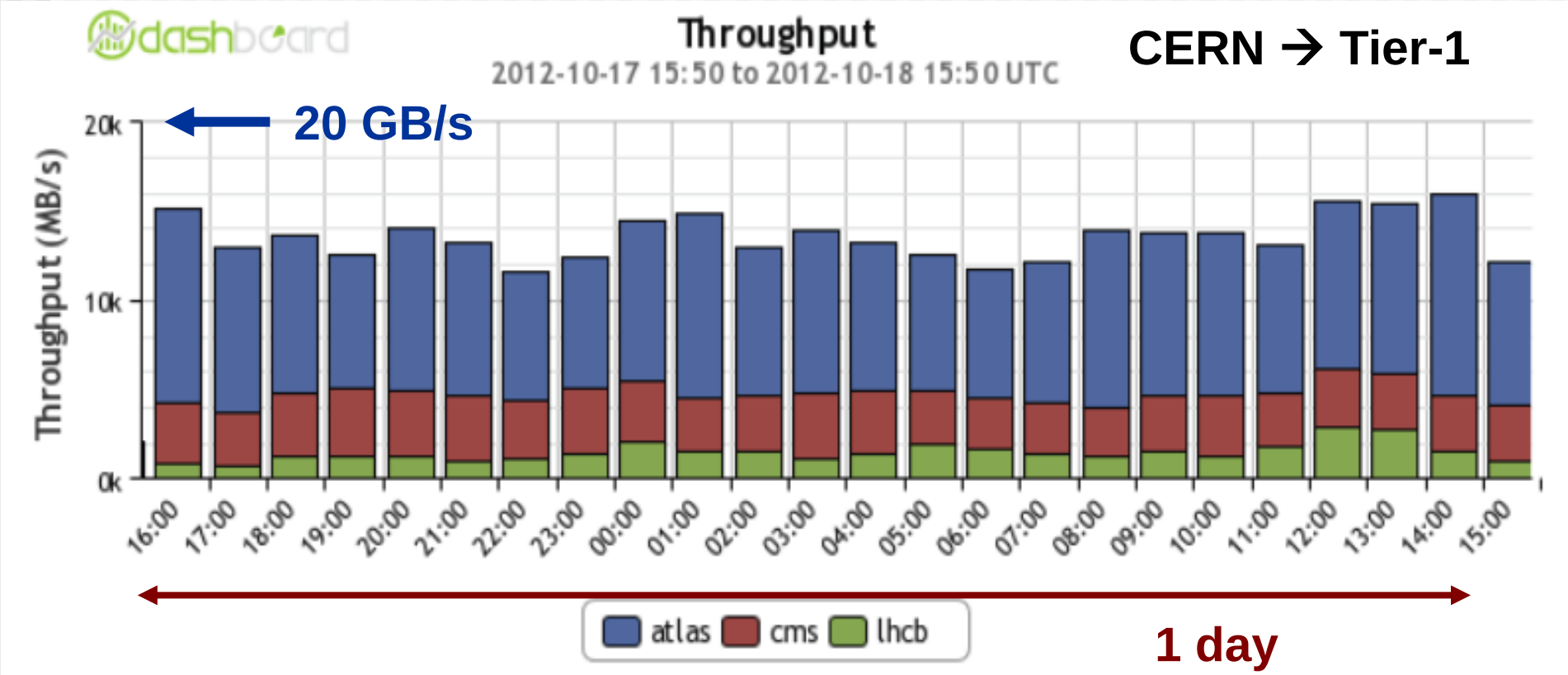




Some Numbers on Data Transfer

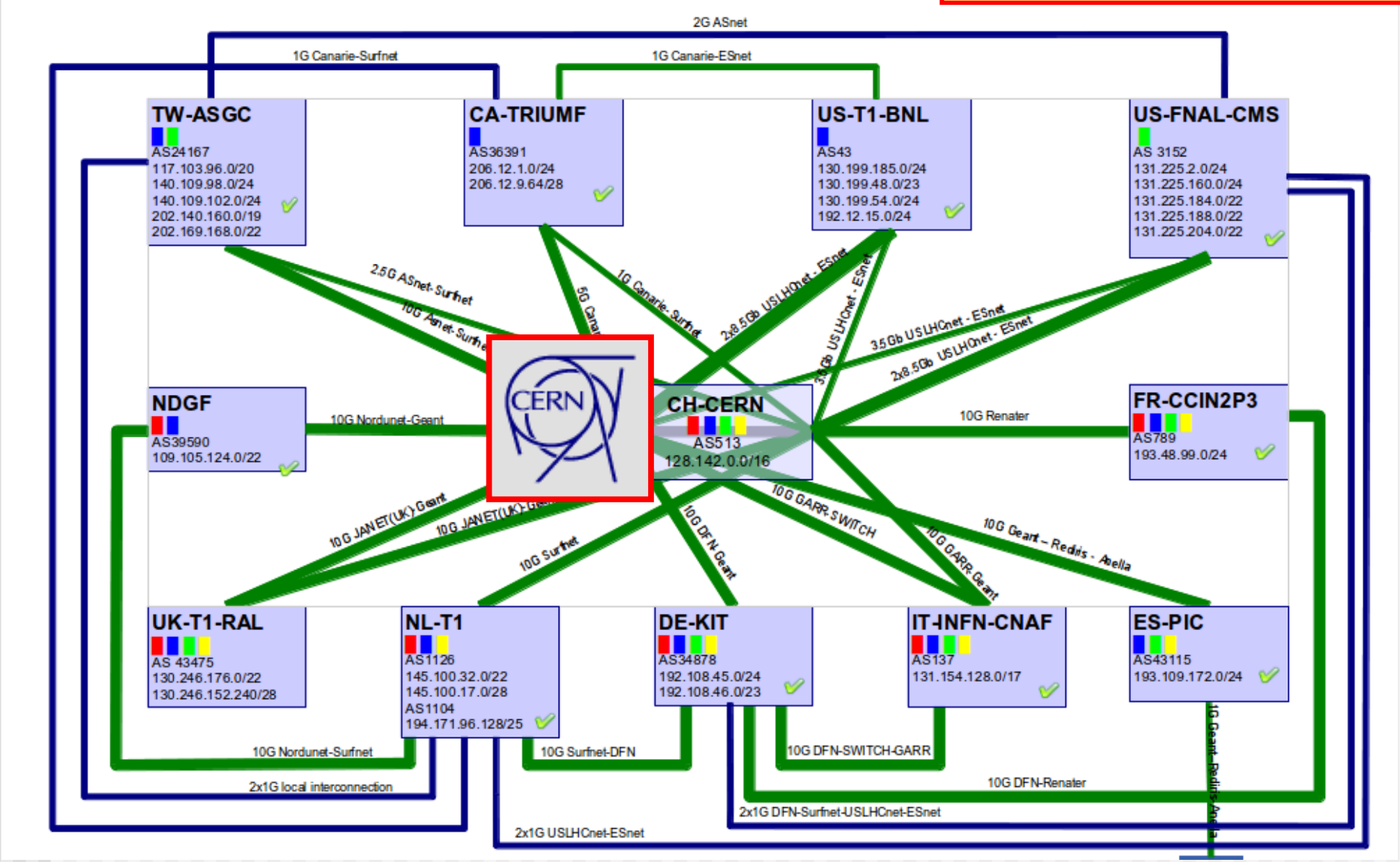
- **Data transferred**
  out of Tier-0 / CERN on average ~ 15 GB/s
  using dedicated optical private network "LHCOPN"

Average data throughput from CERN to Tier1s



LHC Optical Private Network

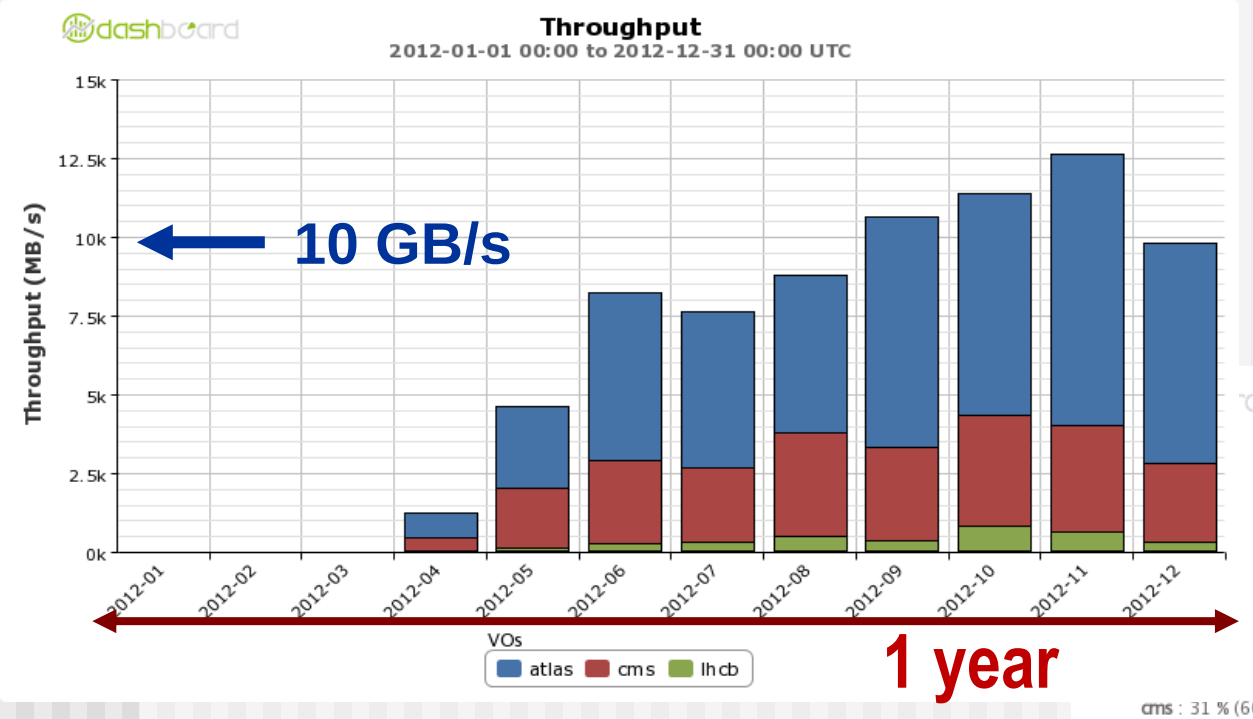
Tier0/1 → Tier1 data distribution



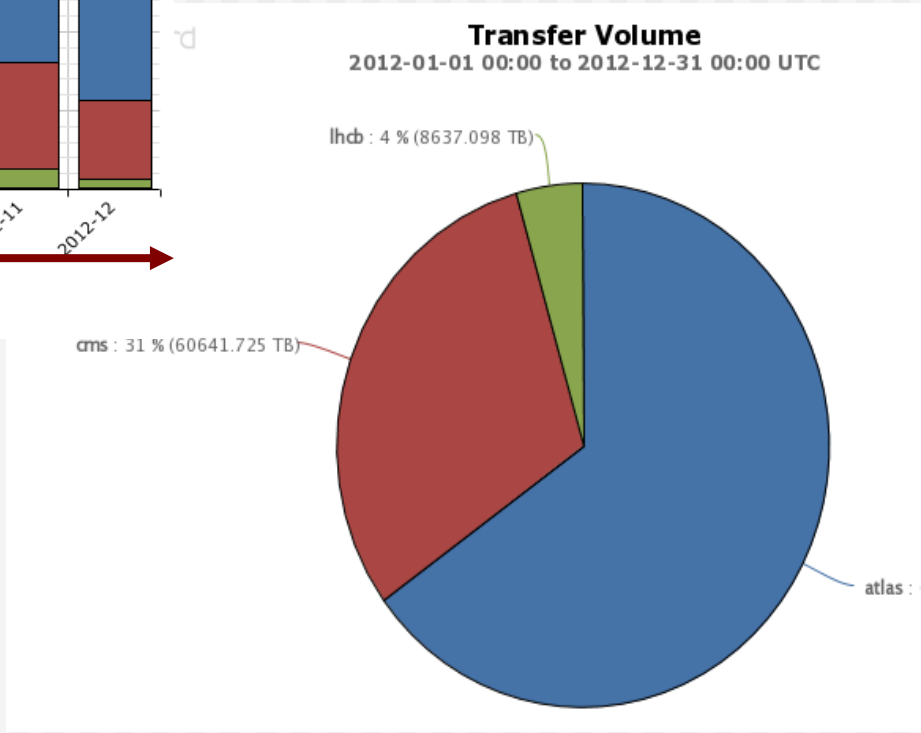
LHCOPN - Large Hadron Collider Optical Private Network : in production since May 2009 … dedicated link of the network to allow distribution of data from T0/T1 to T1s.

Global WLCG Data Transfers 2012

Average data distribution between Tier1s



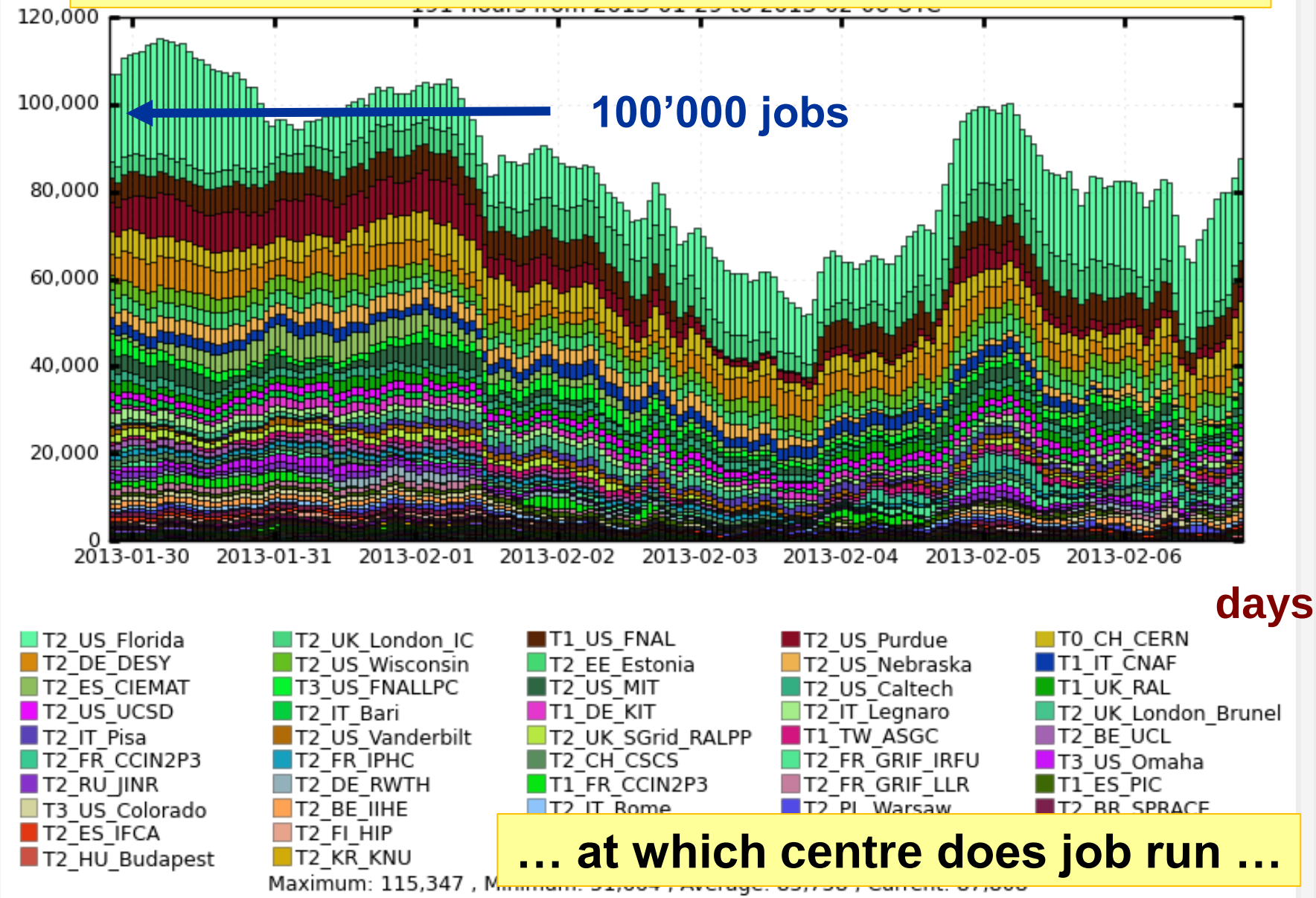
Constant stable throughputs of > 10 GB/s between centers



200 PB shipped between WLCG centers in 2012

Monitoring Running Jobs of CMS

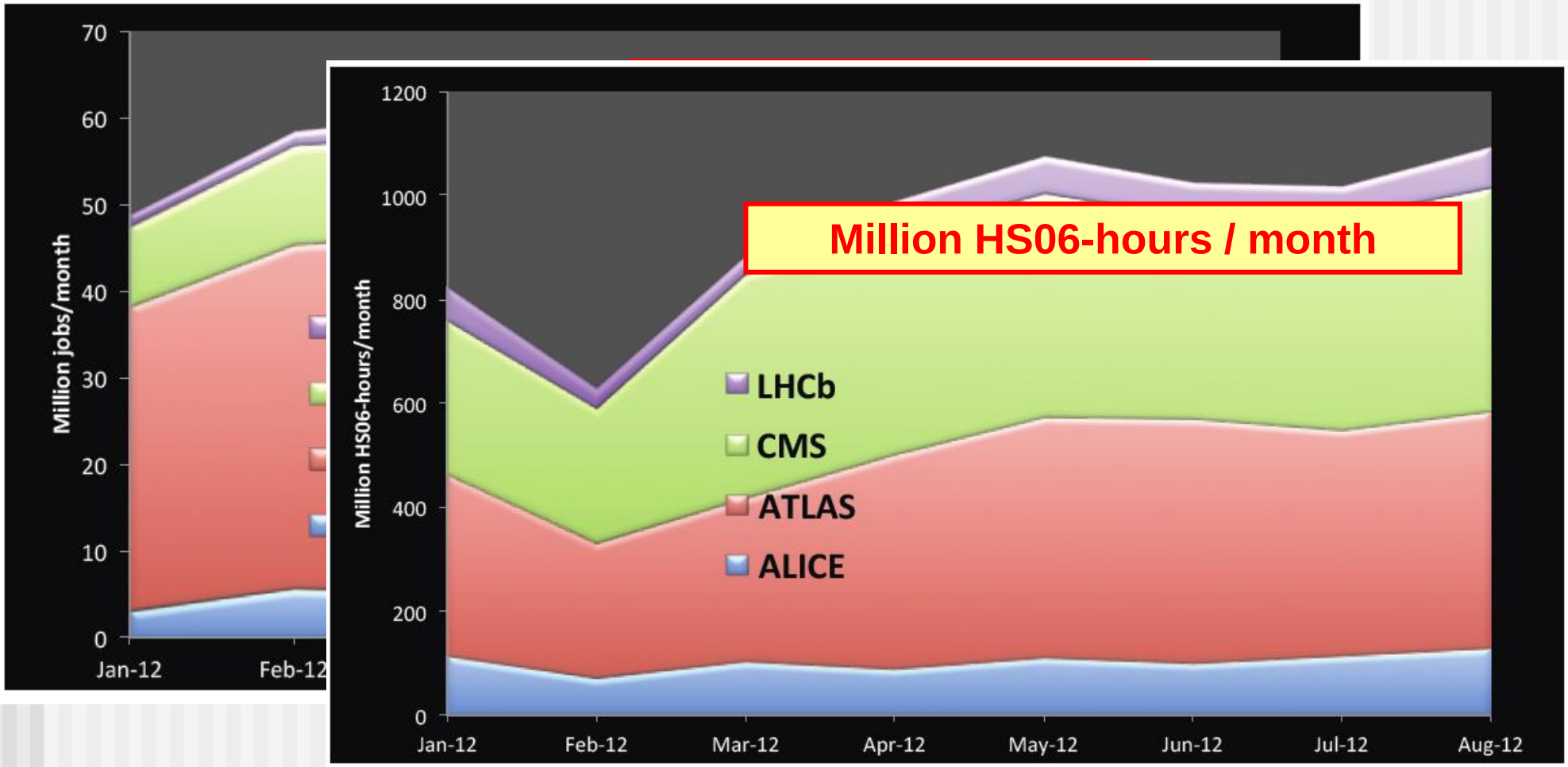
Number of simultanously running jobs of CMS only



… at which centre does job run …

Numbers of WLCG Jobs and CPU usage

- **Number of jobs of all experiments > 2 Million jobs/day**
  ➔ ~ 50 Million jobs/month
- **CPU usage : 10^9 HS06 *hours/month**
  (equivalent to 150 kCPU continues use; 1 HS06 ~ 1 GFLOP)



Why all these efforts ?

… ***PHYSICS*** is the motivation …

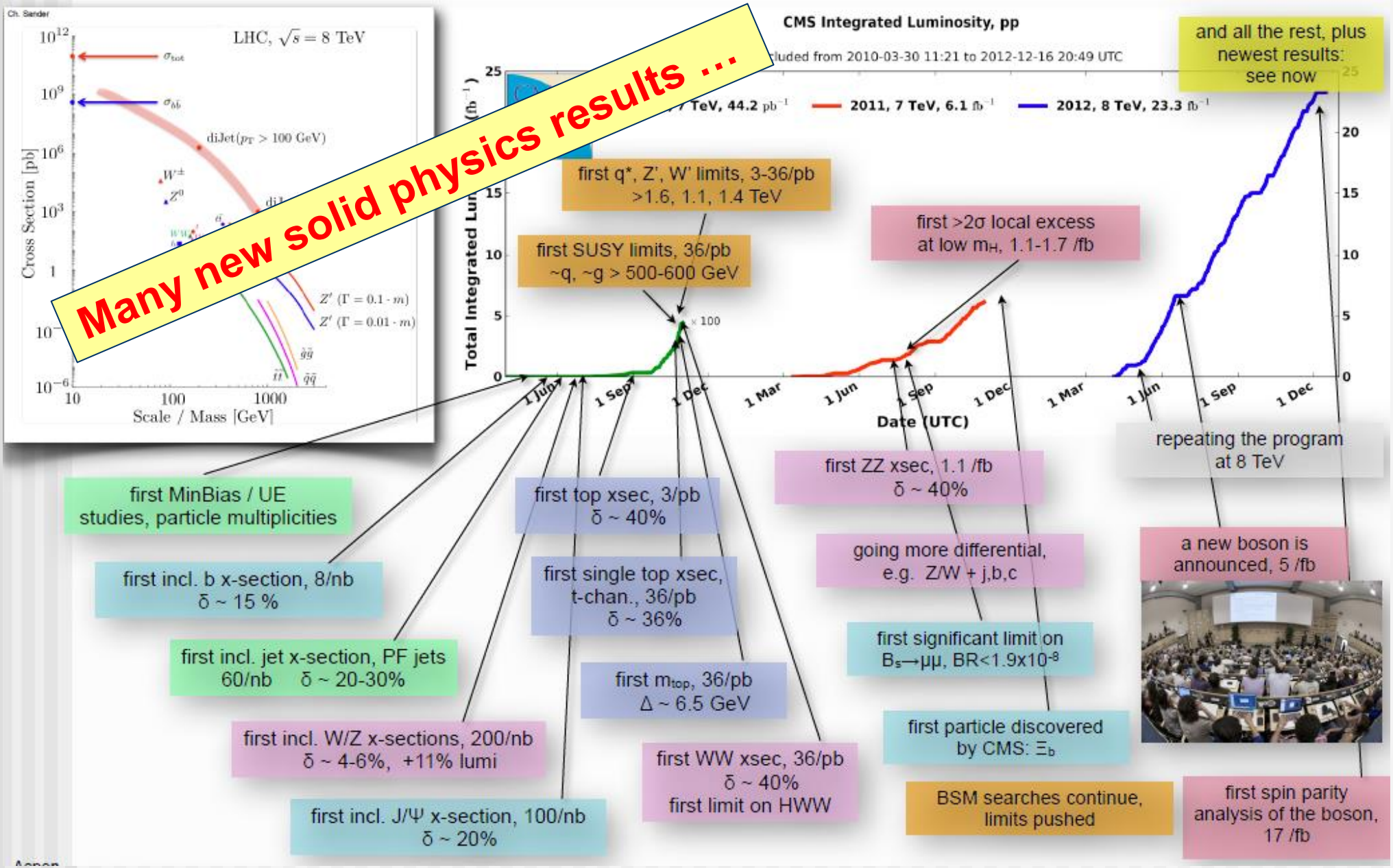
Reality
Experiment

Theory

$$S = \mathrm{i} \int \mathrm{d}^4x \, \mathcal{L}(x)$$

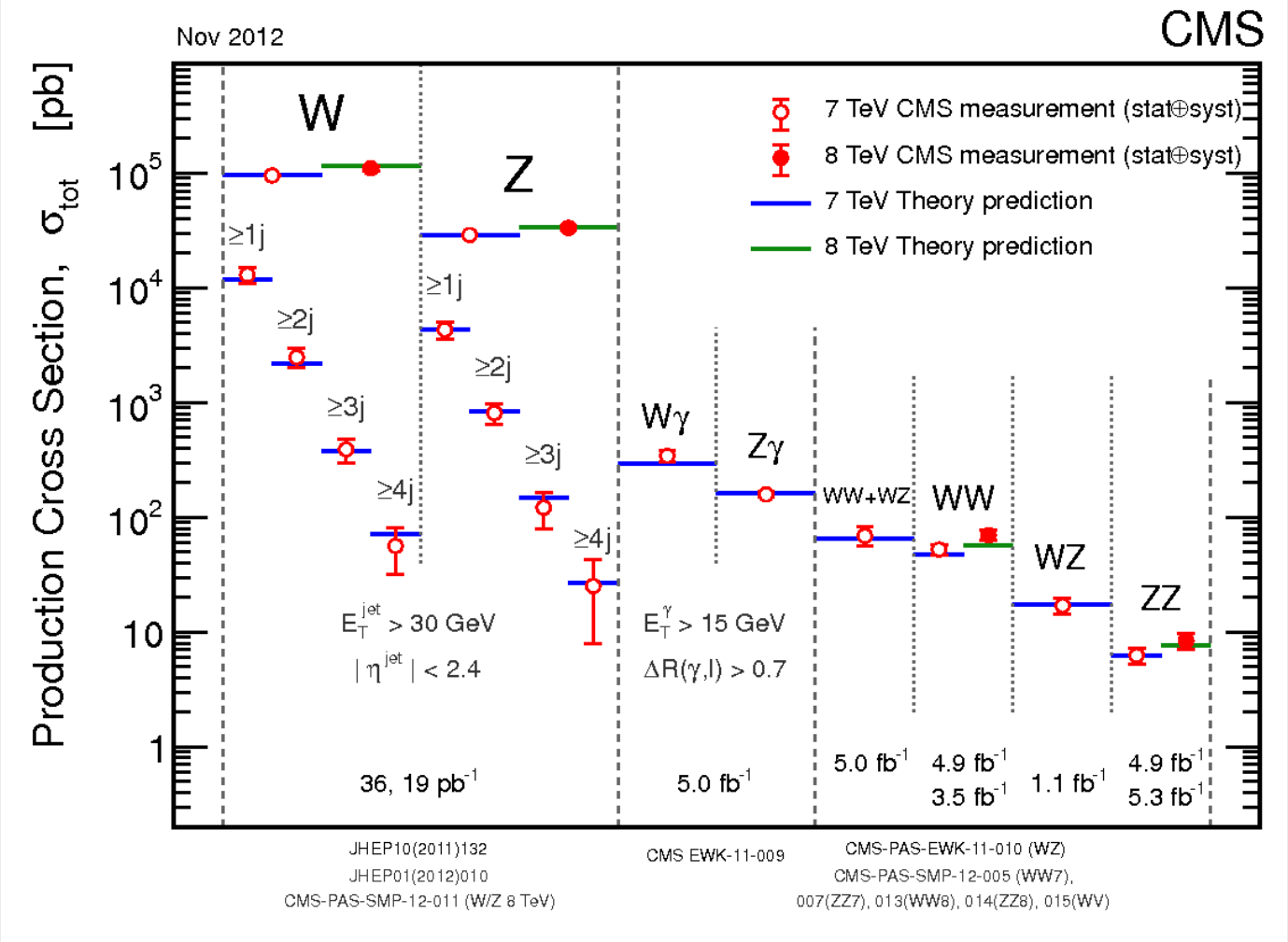
… just some highlights of results … !

Physics Results – Harvest of 3 Years



from G.Dissertori (Aspen 2013)

Physics – Electroweak Processes

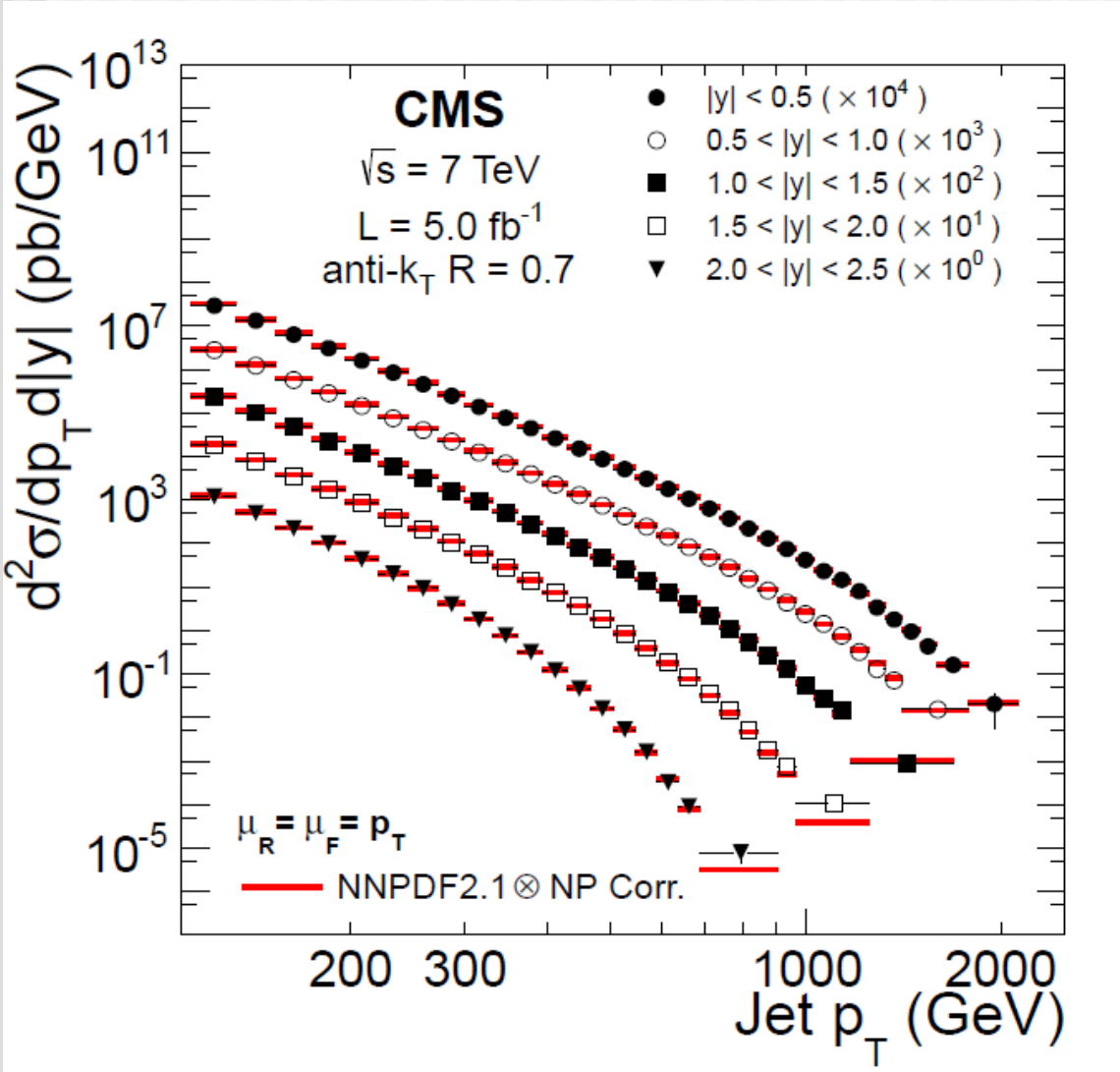


- **Theory (lines) describes data (points) remarkably well …**

No significant deviations found from theoretical predictions …

https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResults

Standard Model Physics



QCD Jet production cross section as example

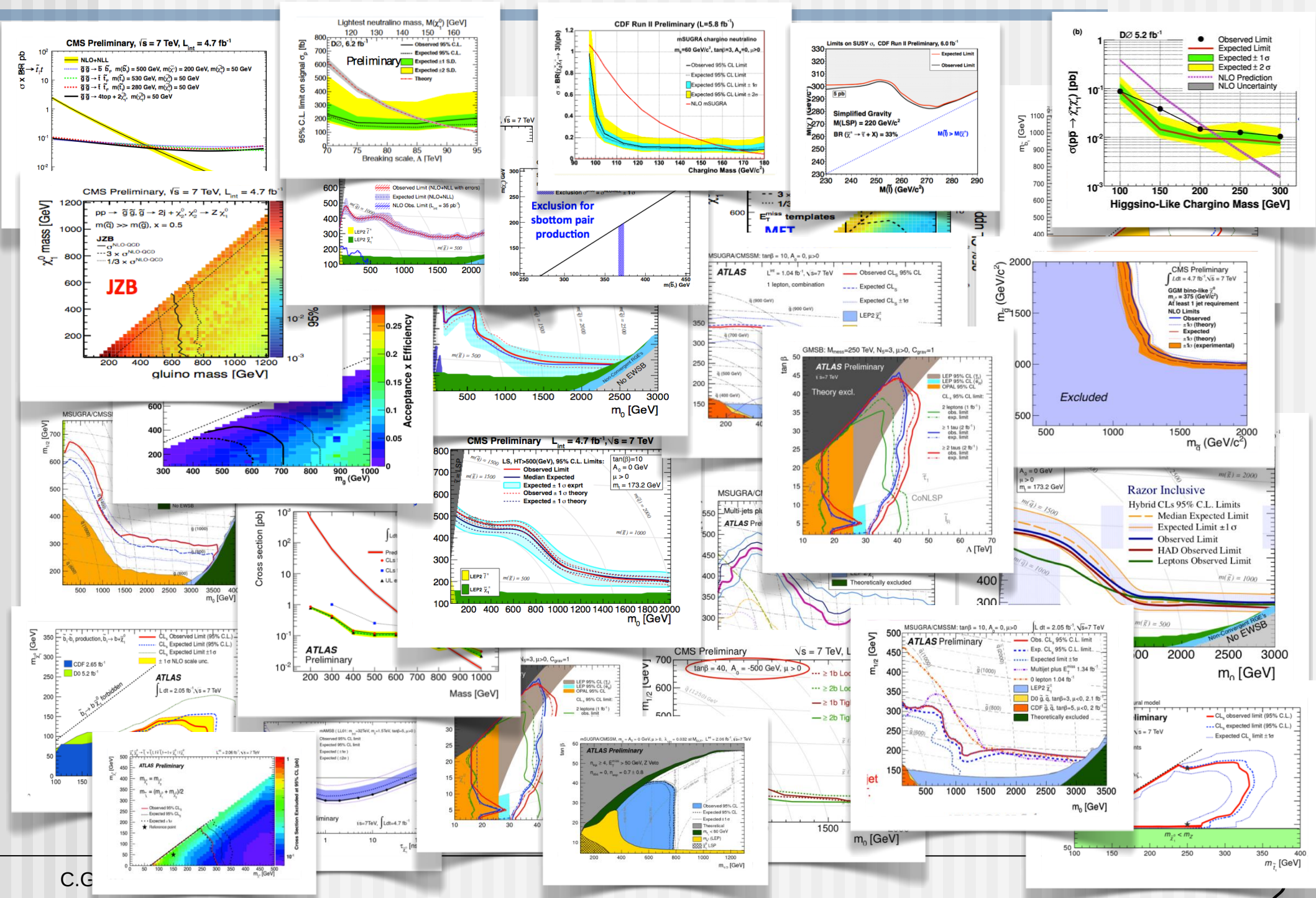
https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResults

Note the scale !

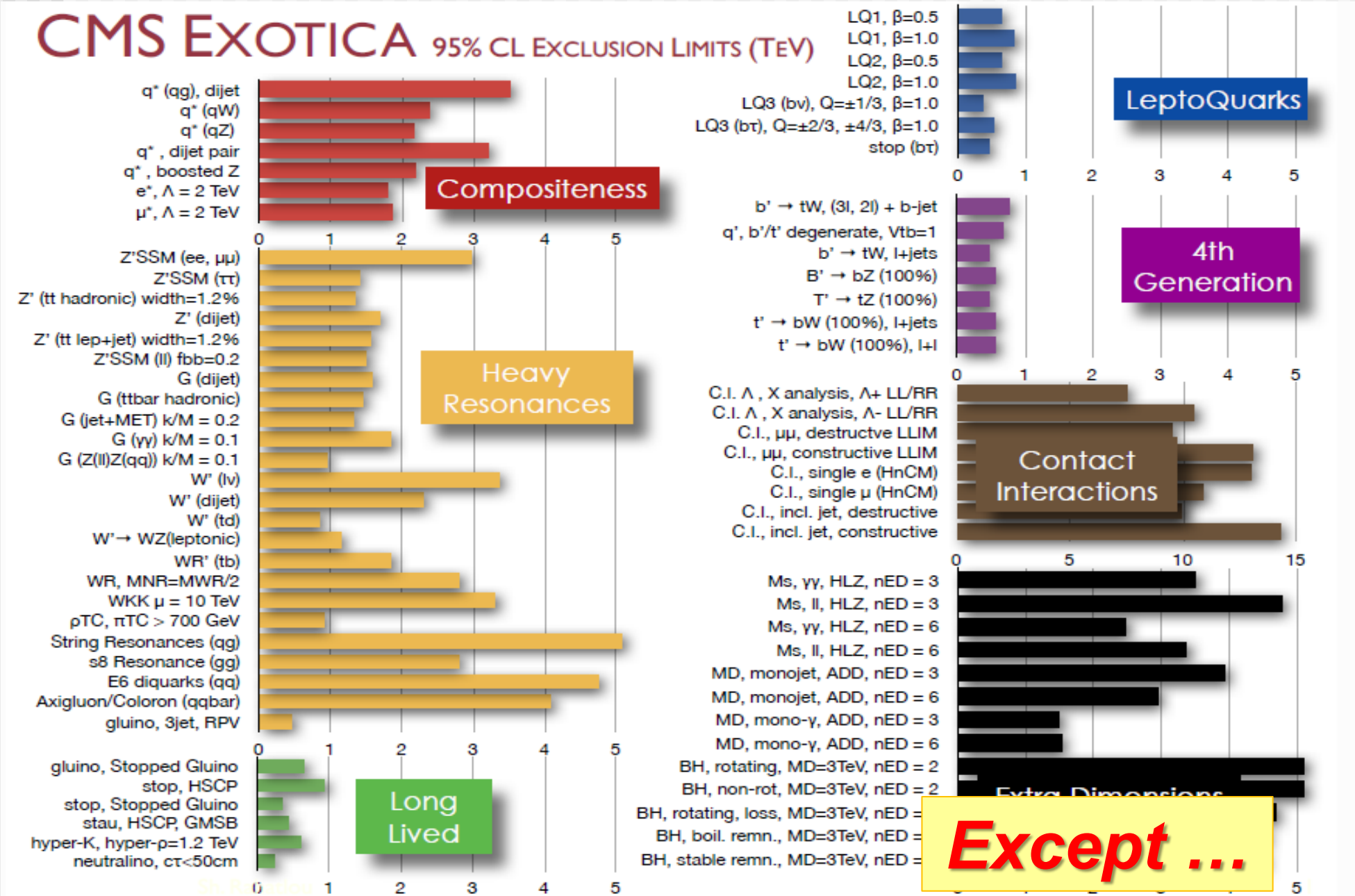
- **Theory describes data well over nearly 12 orders of magnitude!**

No significant deviations found from our best theoretical model, the "standard model" …

Many, many searches for new physics....

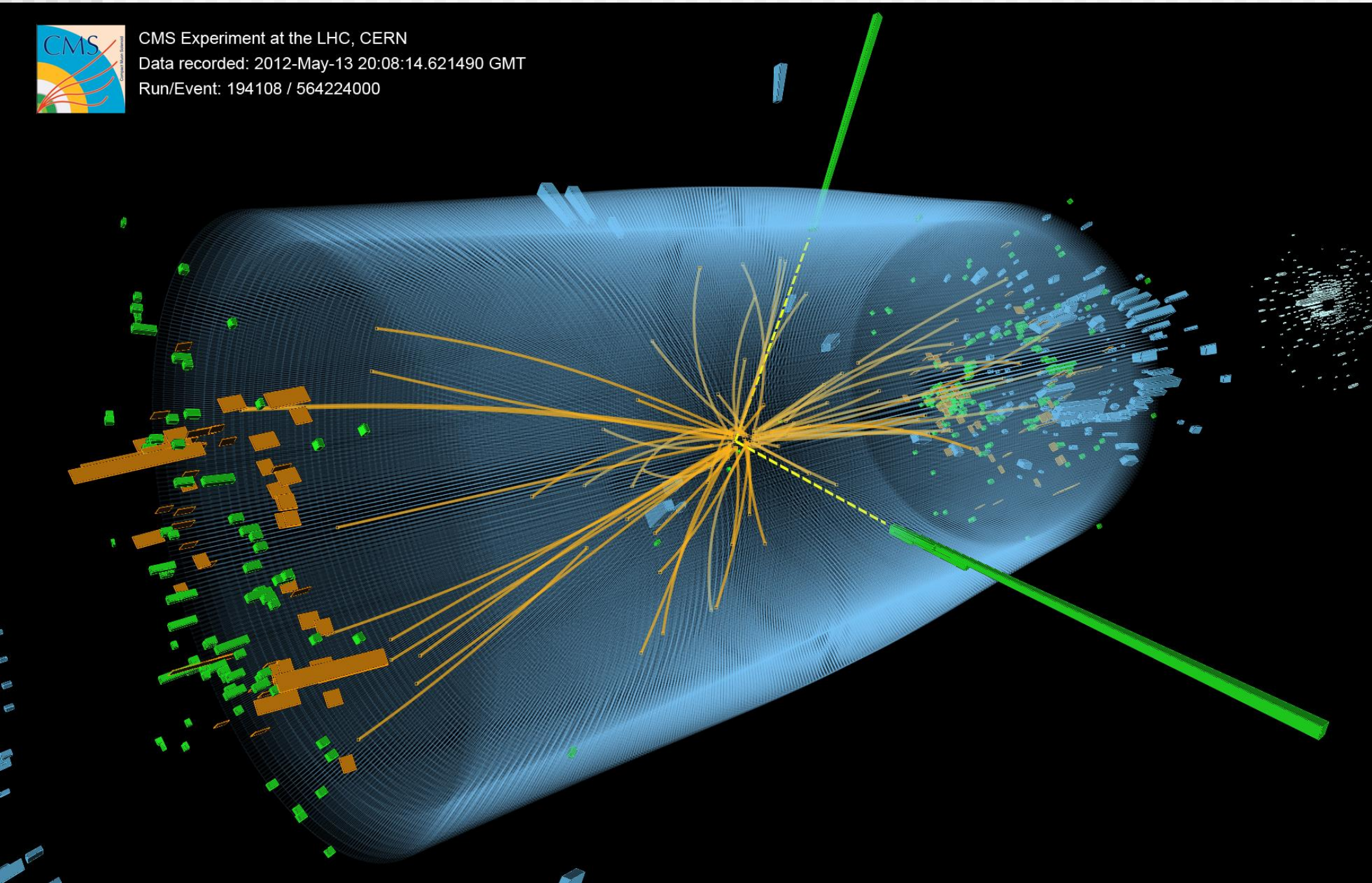


Nothing Exotic found … Limits on Mass/Scale



https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResults

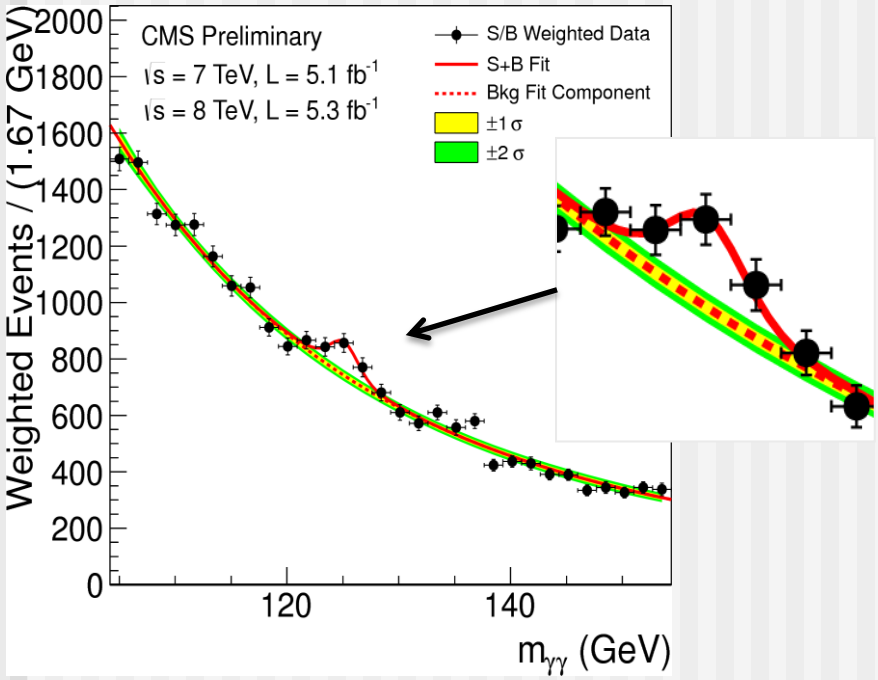
Event Candidate H → γγ



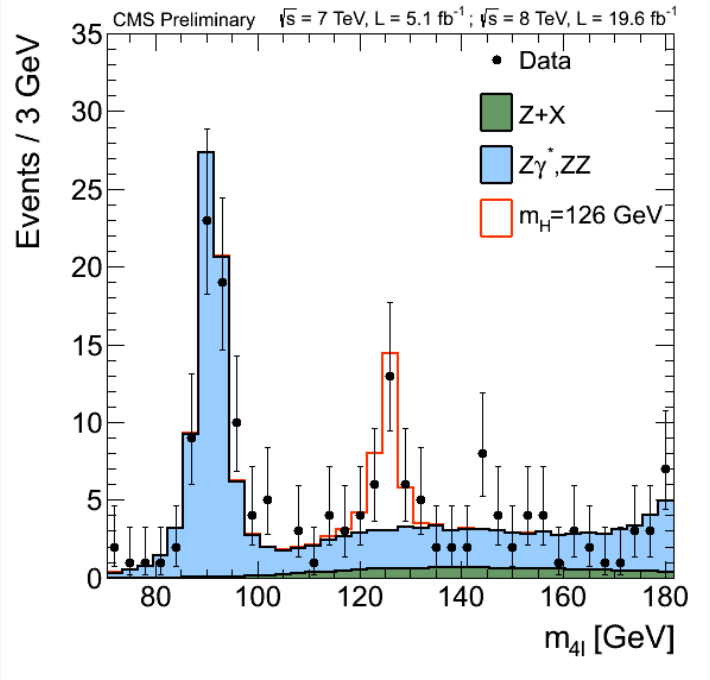
Search the SM Higgs Particle …

Example plots from CMS: more details on CMS and ATLAS websites.

Mass of 2 Photons



Mass of ZZ → 4 leptons

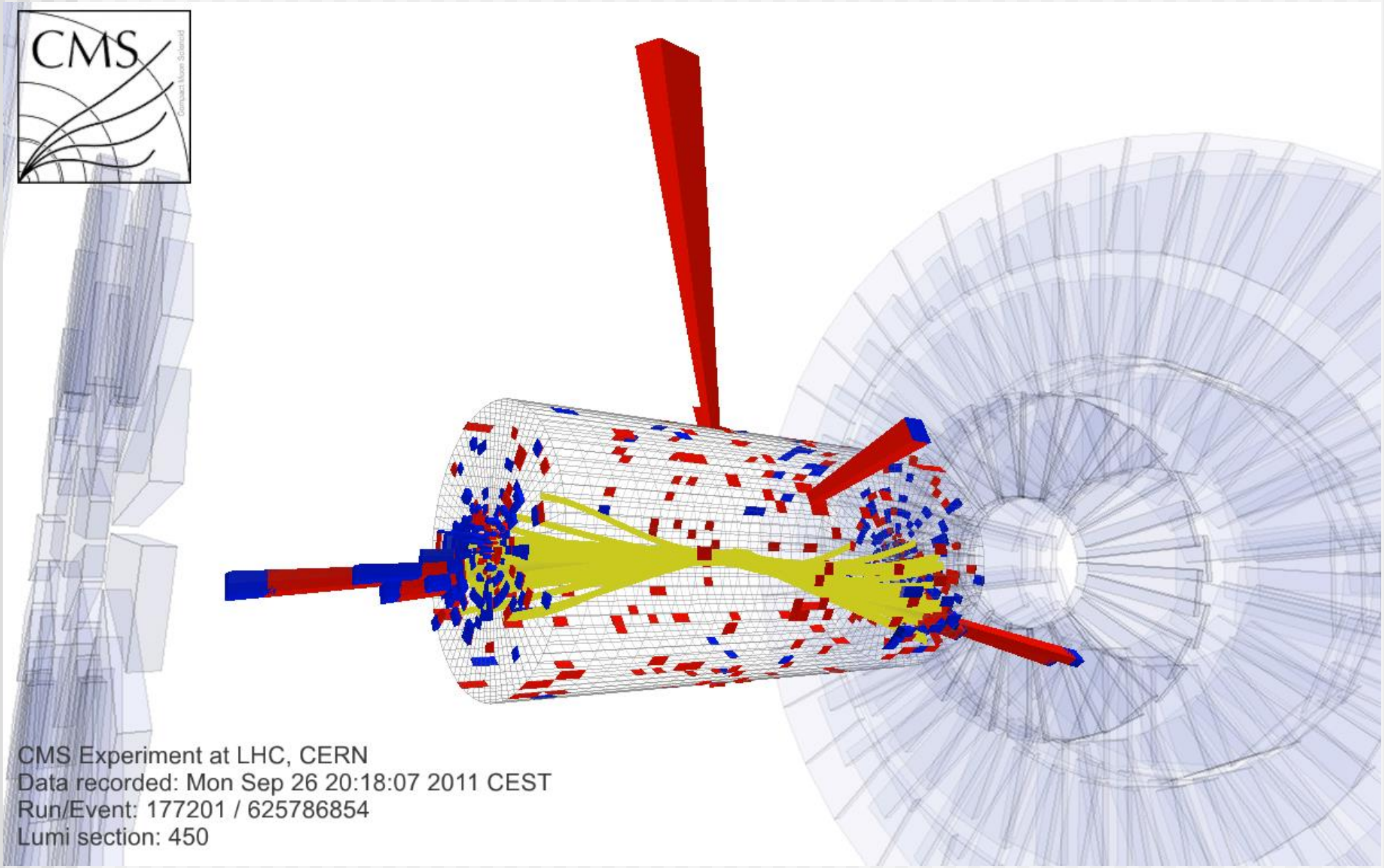


https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResults

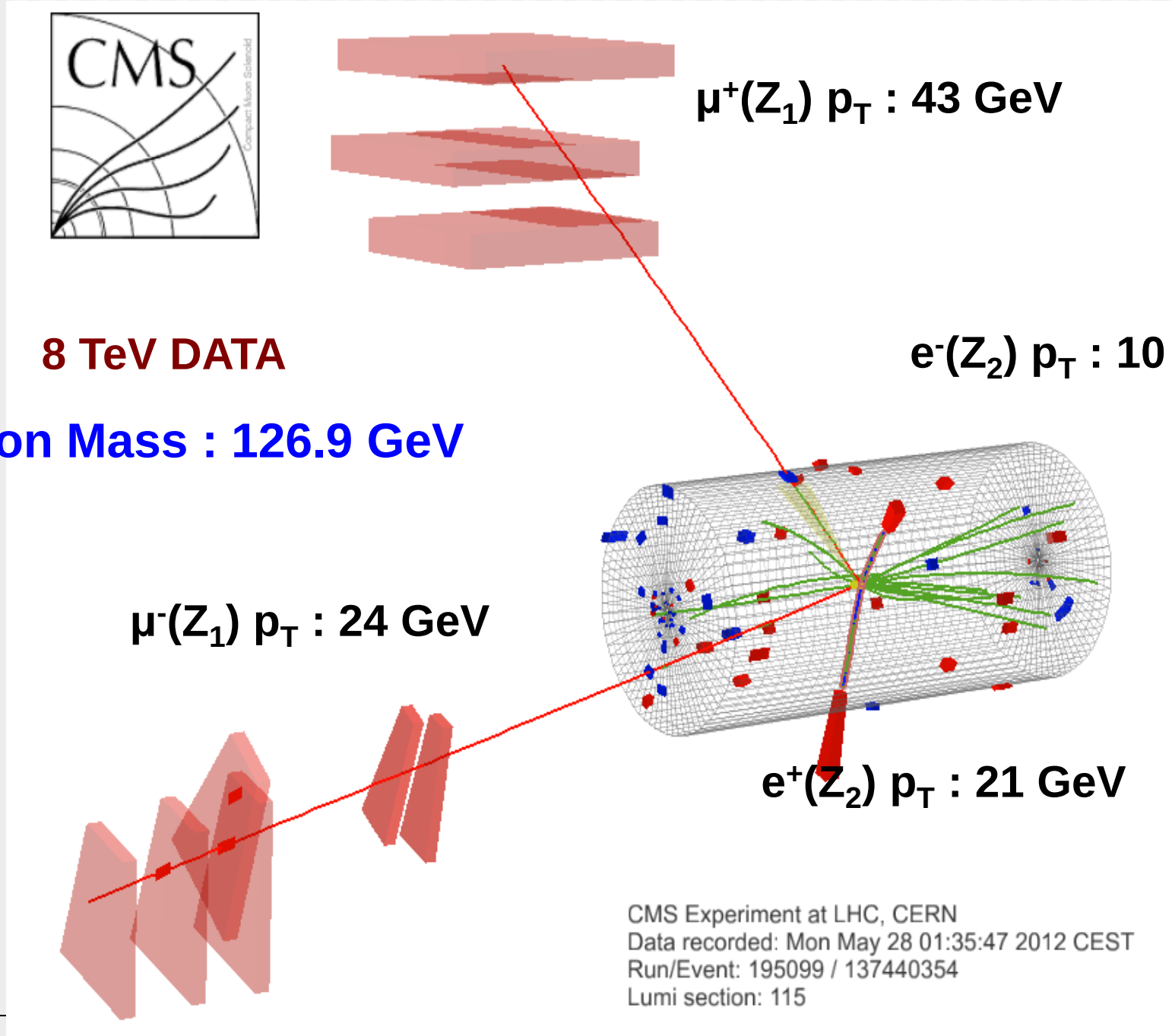
- Measurements show existence of new particle „H“ at 125 GeV!
- Both ATLAS and CMS see it independently (significance close to 7σ)

- All findings are consistent so far with the **standard model Higgs** !
- More data and further analyses are needed to pin down its properties

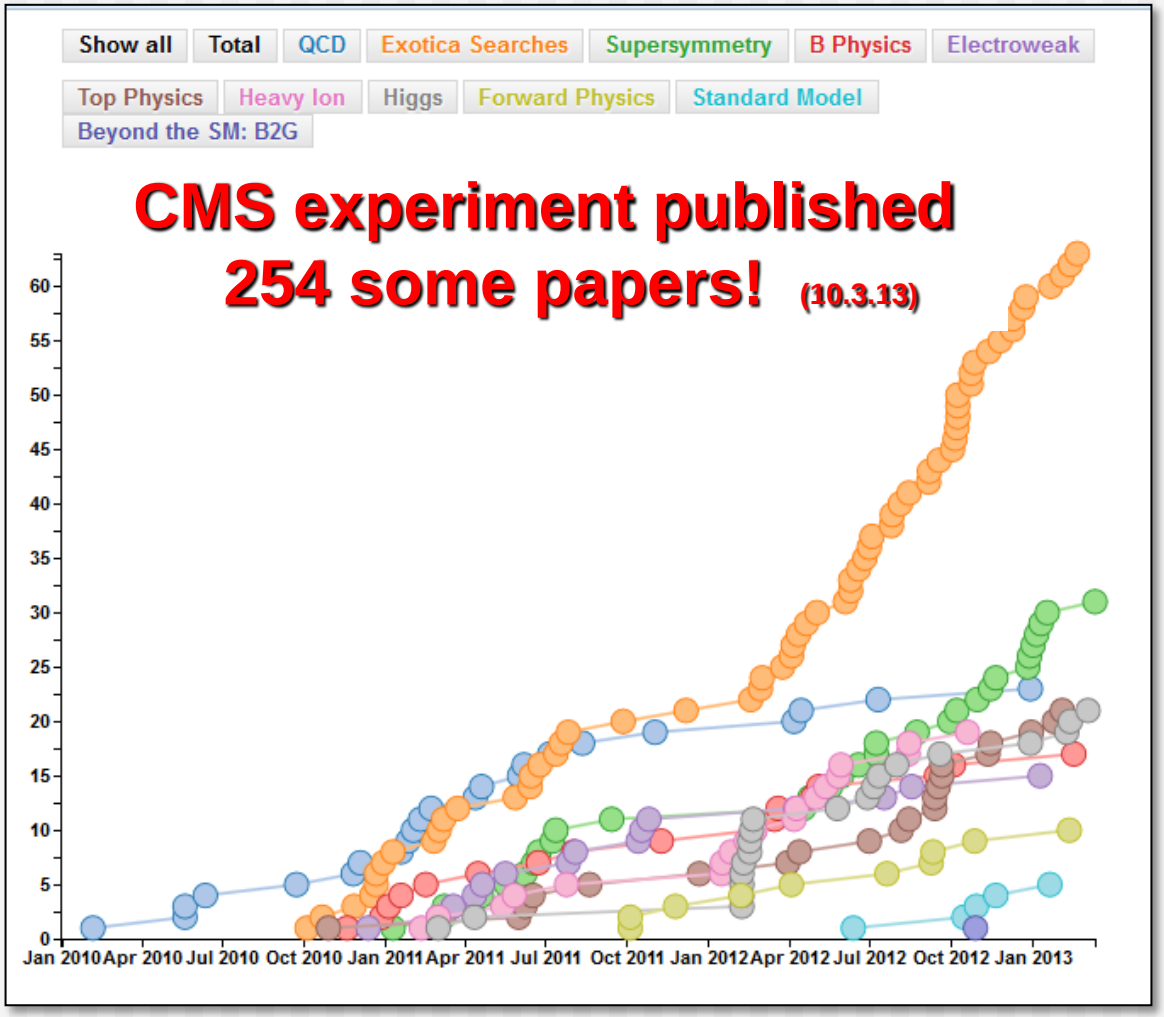
Event Candidate $\gamma\gamma$ +2 jets



Event Candidate H → ZZ → 4 leptons



Harvest of three Years of Physics …



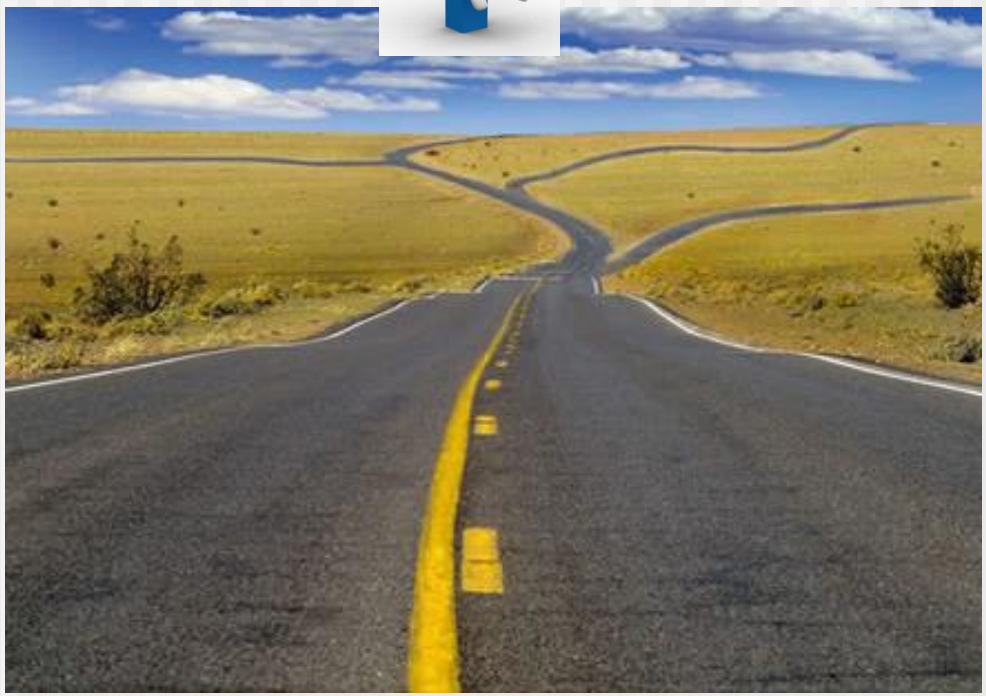
https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResults

This is CMS only, but the other experiments performed similar ….

The Future …

- Machine LHC + Experiments ran well for three years
- Now 18 months shutdown → increase beam energy
- Increase in rate → NEED for increased compute resources by about factor 2!

- Planning goes up to ~2023
- Surprises ahead ???



The Large Hadron Collider (LHC) opened a new chapter of Particle Physics in the TeV regime

Many new results achieved –

Discovered a new particle, "H" (?!),
next we need to pin down its properties

Our search for answers to other most fundamental questions of modern physics continues, as e.g.

- New forms of matter
- Unification of forces …….