Today's World-wide Computing Grid for the Large Hadron Collider (WLCG): A Petascale Facility - Moving to Exascale?

Sverre Jarp, CERN openlab CTO

18 May 2011

Agenda

- Quick overview of CERN and the Large Hadron Collider
- Computing by the LHC experiments
- CERN openlab and future R&D
- Conclusions

CERN and LHC

What is CERN?

- CERN is the world's largest particle physics centre
- Particle physics is about:
  - elementary particles, the constituents all matter in the Universe is made of
  - fundamental forces which hold matter together
- Particles physics requires:
  - special tools to create and study new particles
    - Accelerators
    - Particle Detectors
    - Powerful computer systems

CERN is also:

-2250 staff (physicists, engineers, technicians, …)

- Some 10'000 visiting scientists (most of the world's particle physicists)

They come from 500 universities representing 80 nationalities.

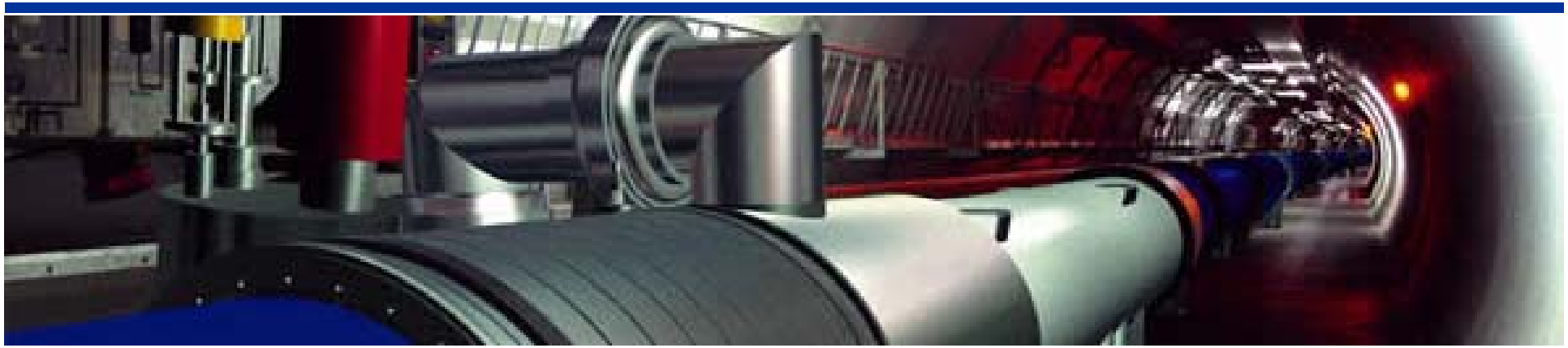
What is the LHC?

- The Large Hadron Collider can collide beams of protons at a design energy of 2 * 7 TeV
  - Inaugurated Sept. 2008; restart Nov. 2009
  - Reached 3.5 TeV (March 2010)
  - 2011/12: Two years at 3.5 TeV before upgrade

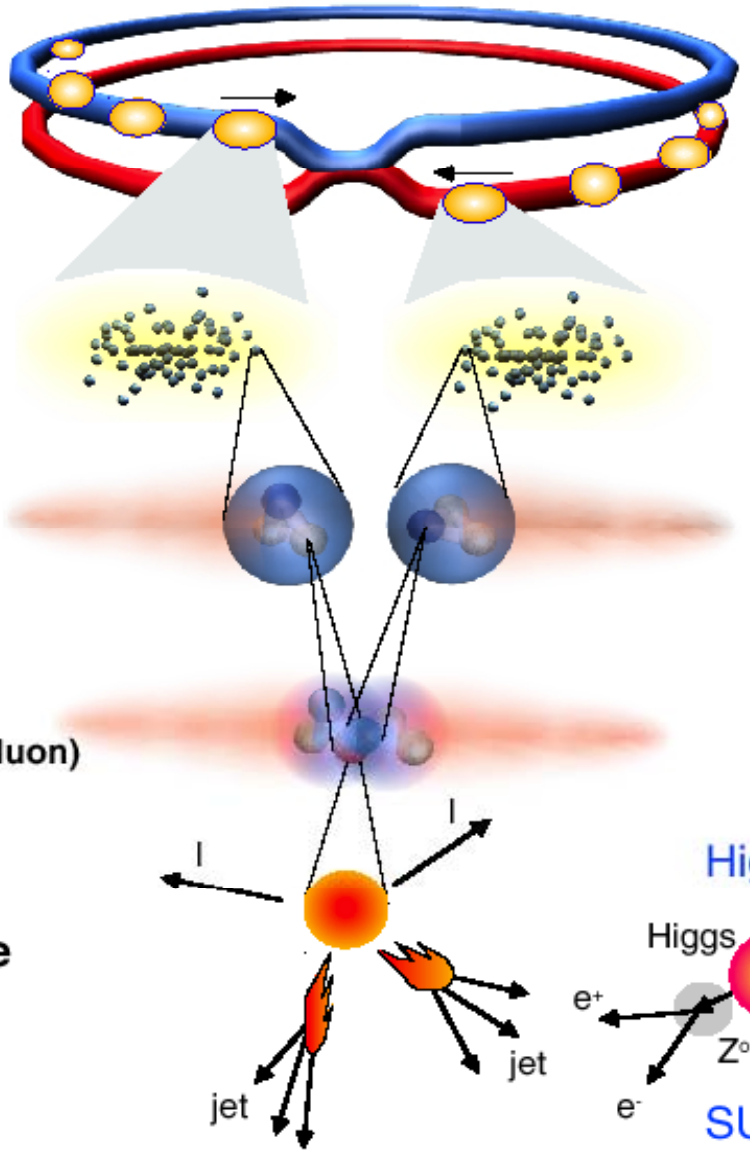
Four experiments, with detectors as 'big as cathedrals':
ALICE
ATLAS
CMS
LHCb

- Using the latest super-conducting technologies, it operates at about – 271˚C, just above the temperature of absolute zero. The coldest place in the Universe.

- With its 27 km circumference, the accelerator is the largest superconducting installation in the world.



Collisions at LHC

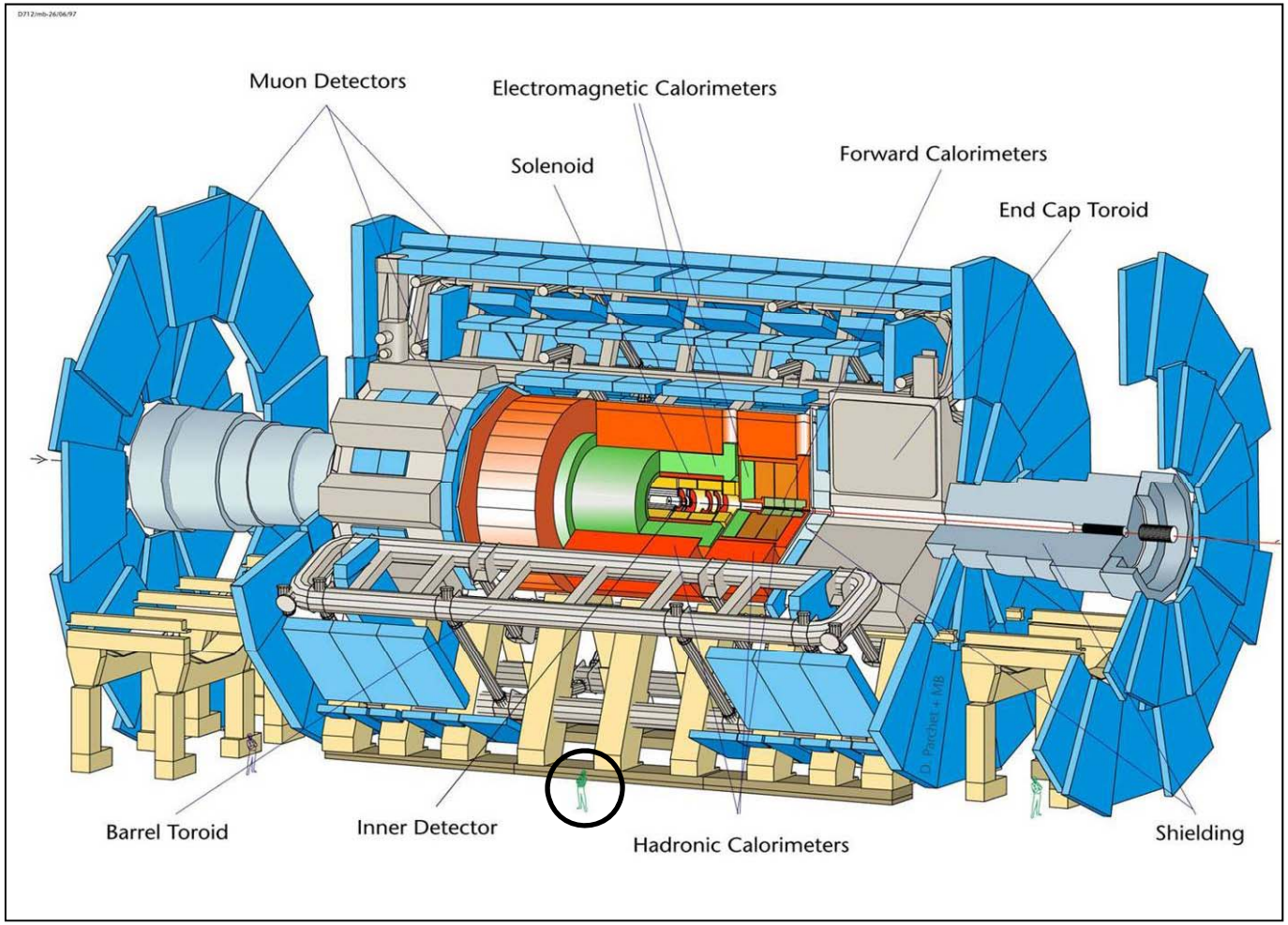


Proton-Proton	2835 bunch/beam
Protons/bunch	10^{11}
Beam energy	7 TeV (7×10^{12} eV)
Luminosity	10^{34} cm^{-2} s^{-1}
Crossing rate	40 MHz
Collisions ≈	10^{7} - 10^{9} Hz

Selection of 1 in
10,000,000,000,000

ATLAS

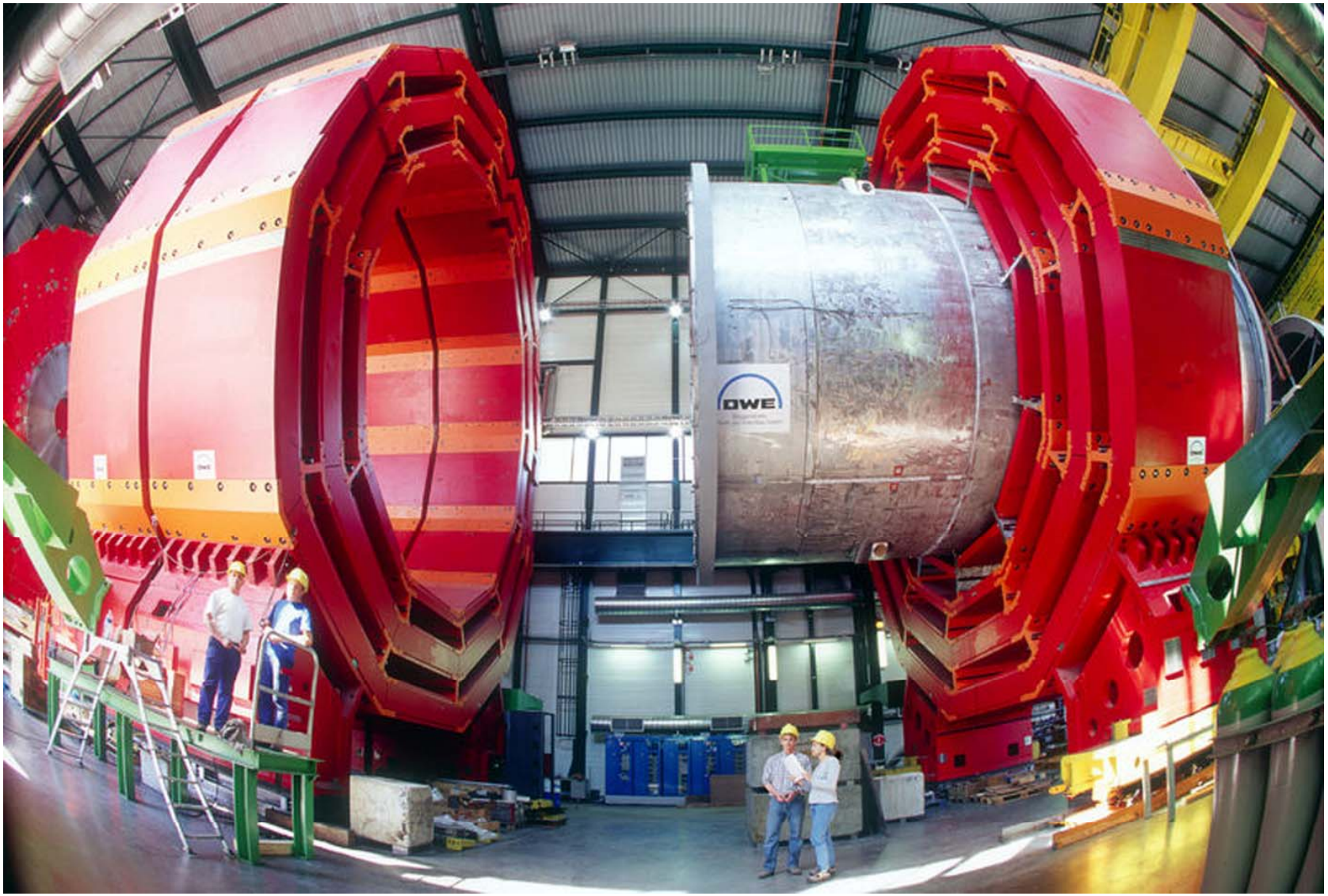
- General purpose LHC detector – 7000 tons



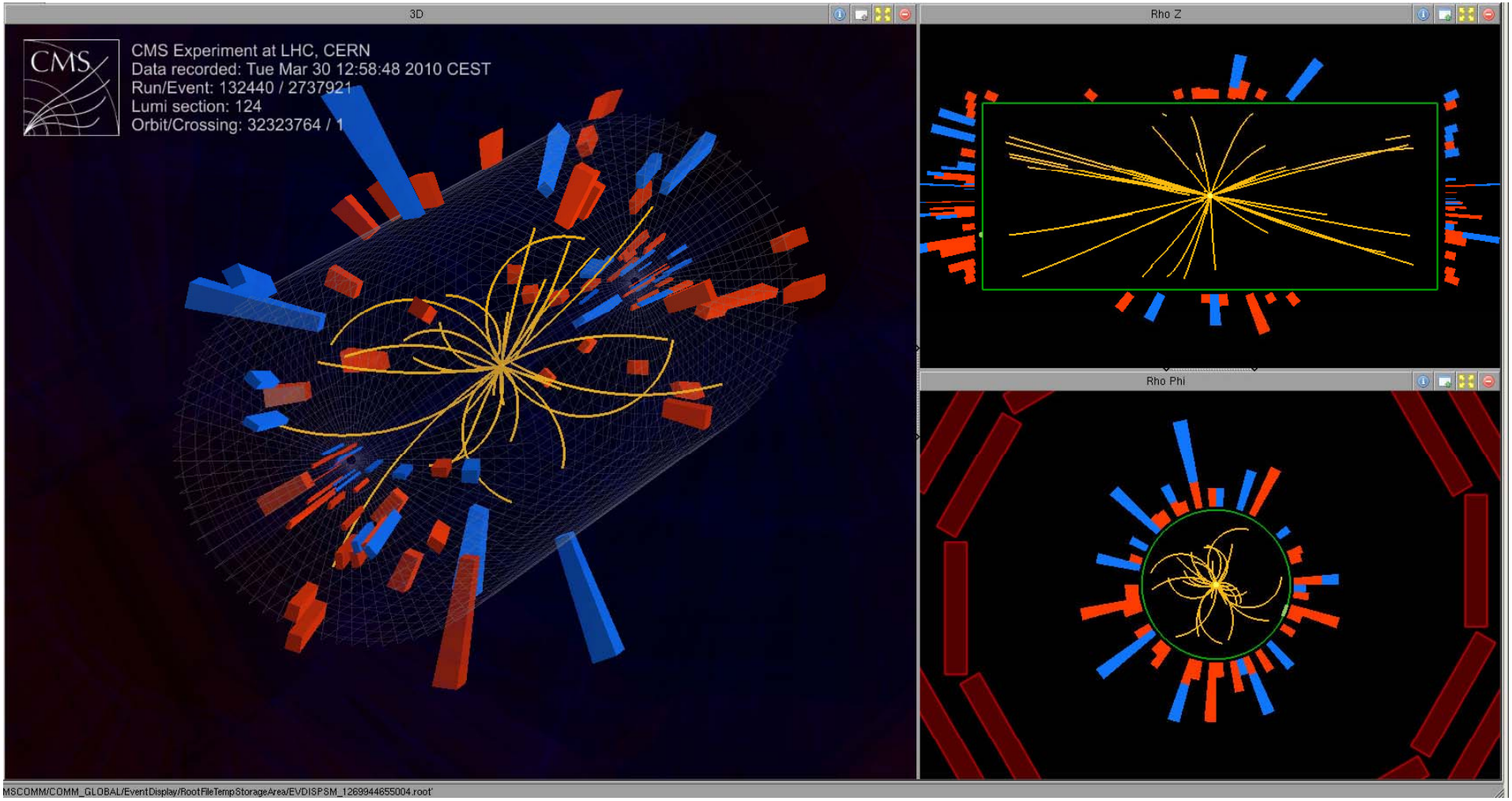
ATLAS under construction (2005)

Compact Muon Solenoid

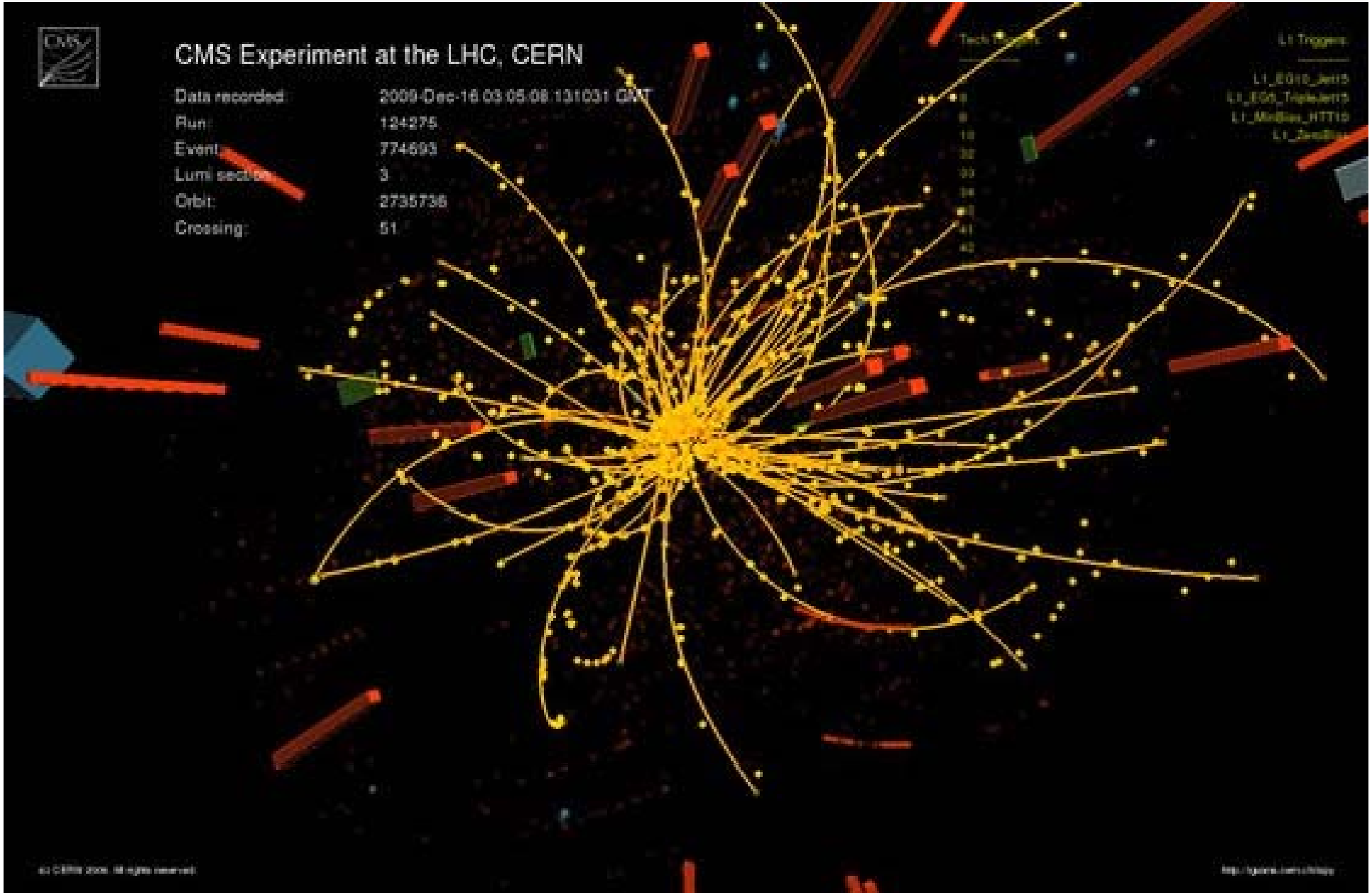
(CMS – 12500 tons)



CMS event @ 3.5 TeV

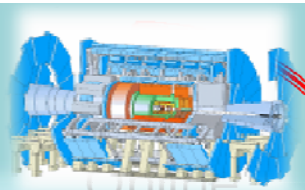


A CMS collision



LHC Computing

Data Handling and Computation for Physics Analysis



Online trigger and filtering

Selection & reconstruction

Offline Reconstruction

Event summary data

10%

Processed data

100%

Raw data

Event reprocessing

Batch physics analysis

1%

Event simulation

Offline Simulation w/GEANT4

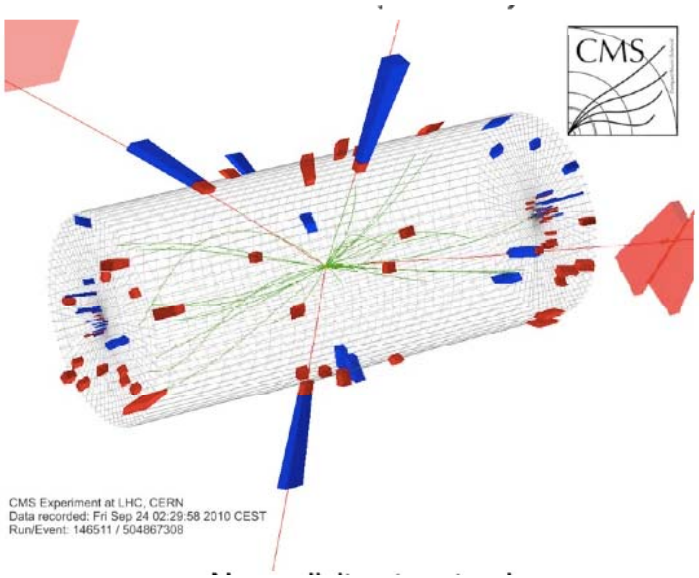
Offline Analysis w/ROOT

Analysis objects
(extracted by physics topic)

Interactive physics analysis

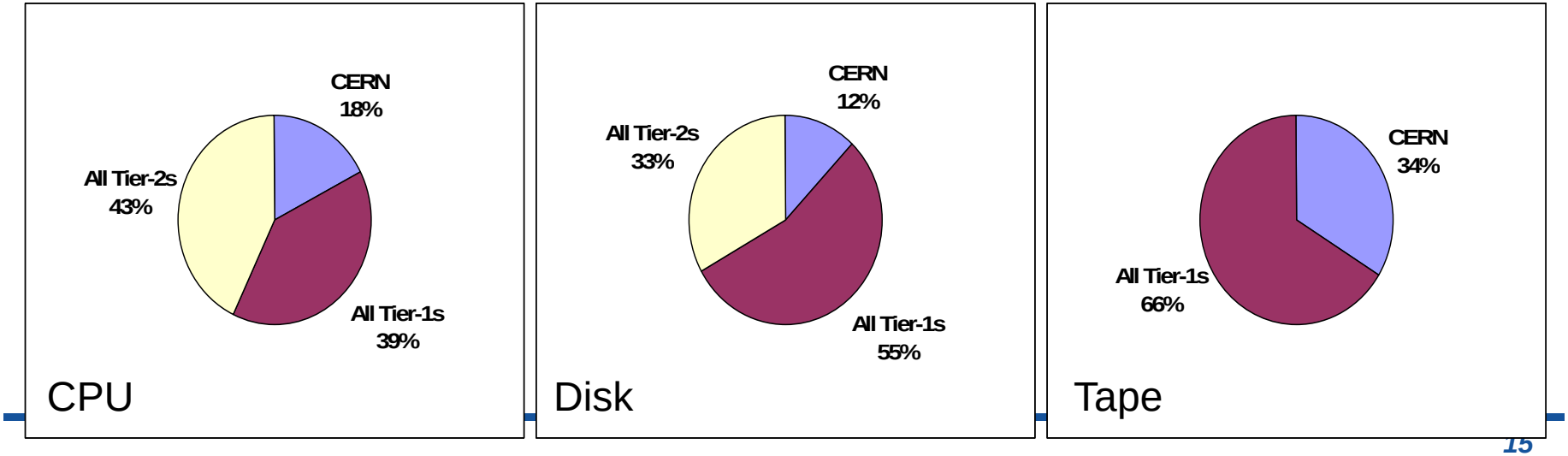
HEP programming paradigm

- All events are independent
- Trivial parallelism has been exploited by High Energy Physics for decades
  - Compute one event after the other in a single process
- Advantage:
  - Large jobs can be split into N efficient processes, each responsible for processing M events
    - Built-in scalability
- Disadvantage:
  - Memory needed by each process
    - With 2 – 4 GB per process
    - A dual-socket server with Octa-core processors
      - Needs 32 – 64GB



Rationale for Grids

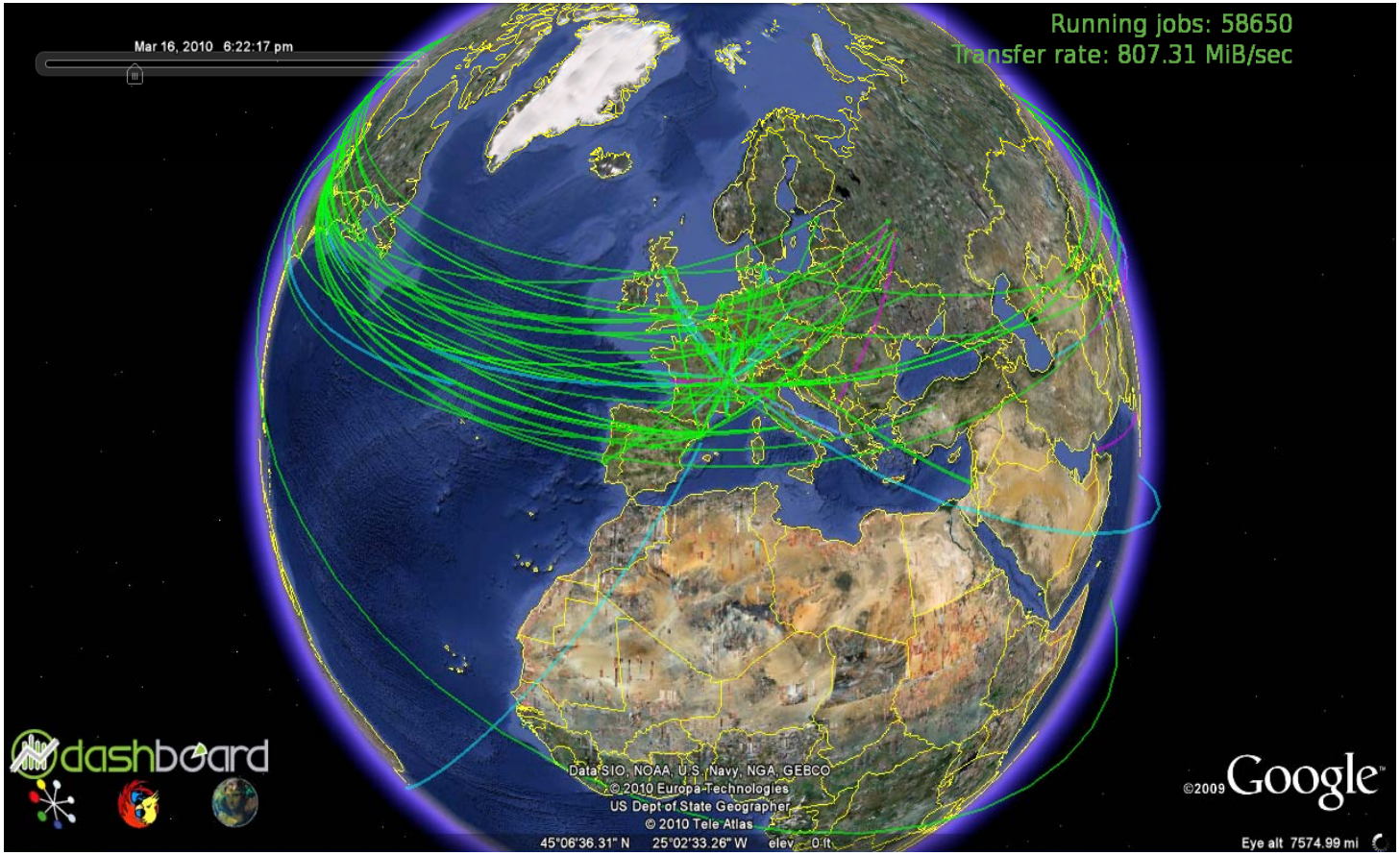
- **The LHC Computing requirements are simply too huge for a single site:**
  - Impractical to build such a huge facility in one place
  - Modern wide-area networks have made distances shrink
    - But, latency still has to be kept in mind
  - The users are not necessarily at CERN
  - Political resistance to funding everything at CERN
- **So, we are spreading the burden!**



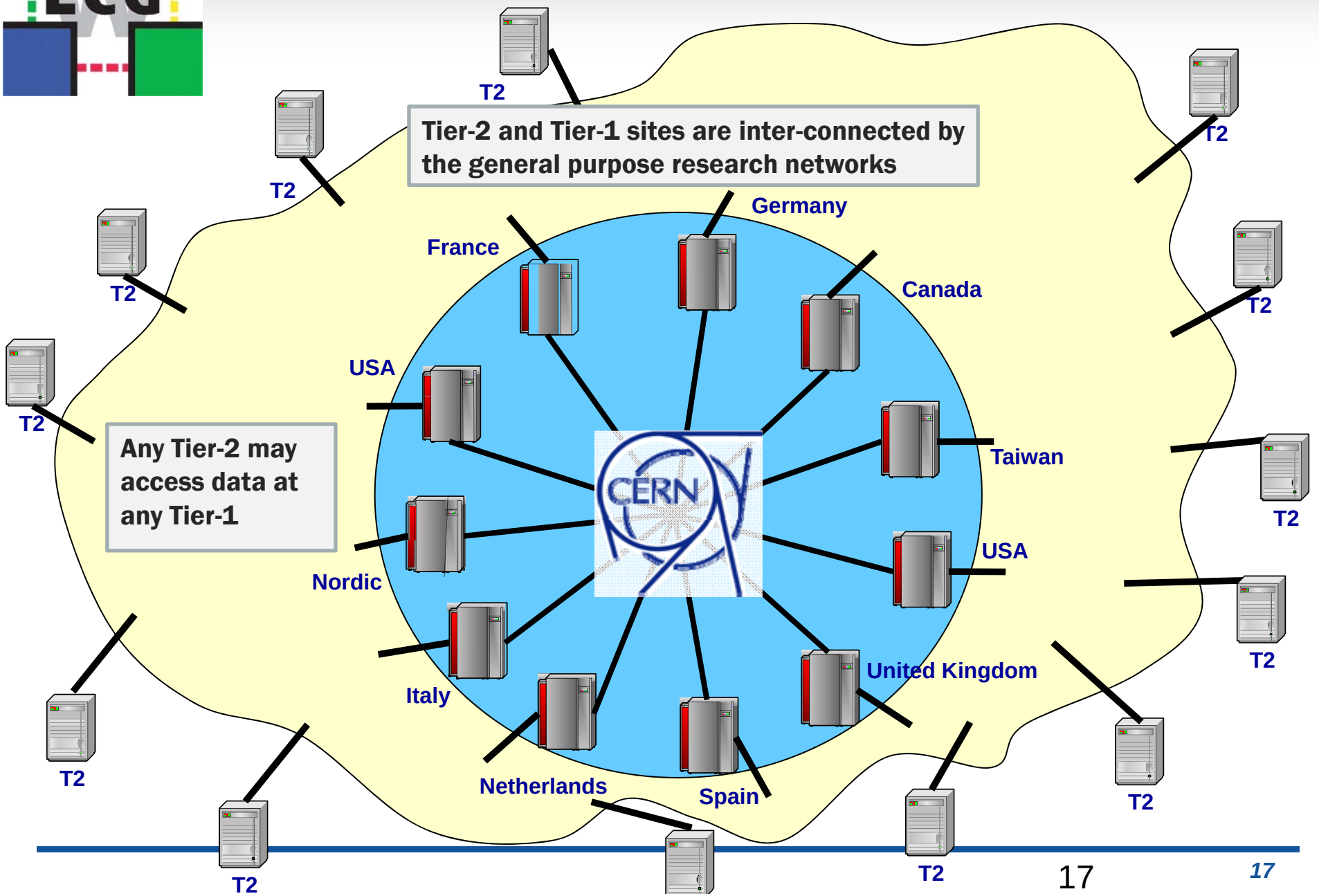
World-wide LHC Computing Grid

- **W-LCG: Largest Grid service in the world**

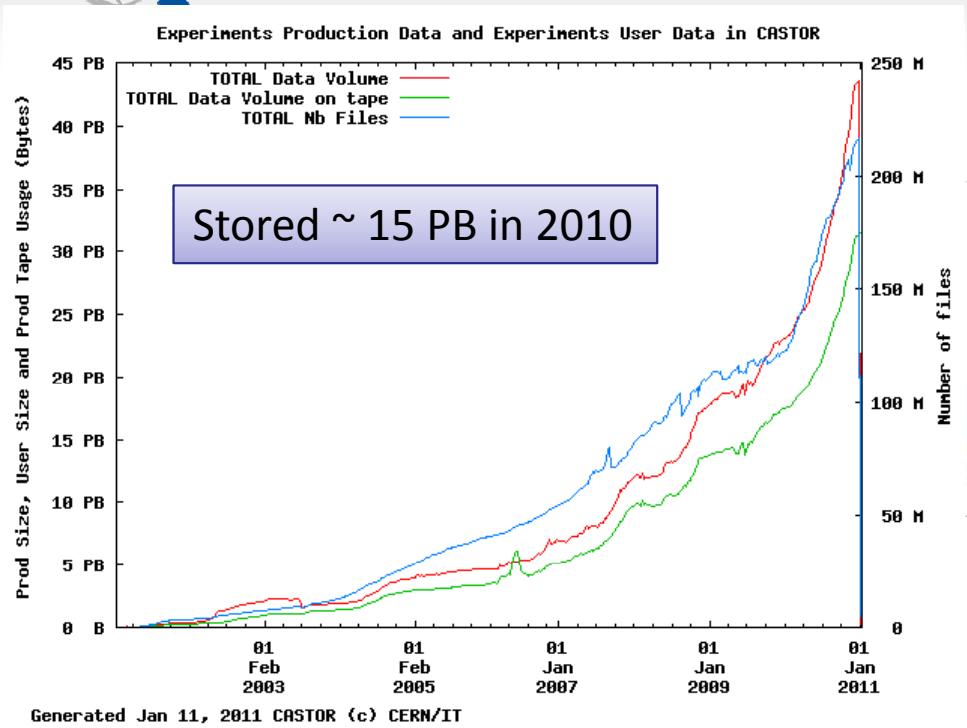
- **Built on top of EGEE and OSG**
- **Almost 160 sites in 34 countries**
- **More than 250'000 IA processor cores (w/Linux)**
- **One hundred petabytes of storage**



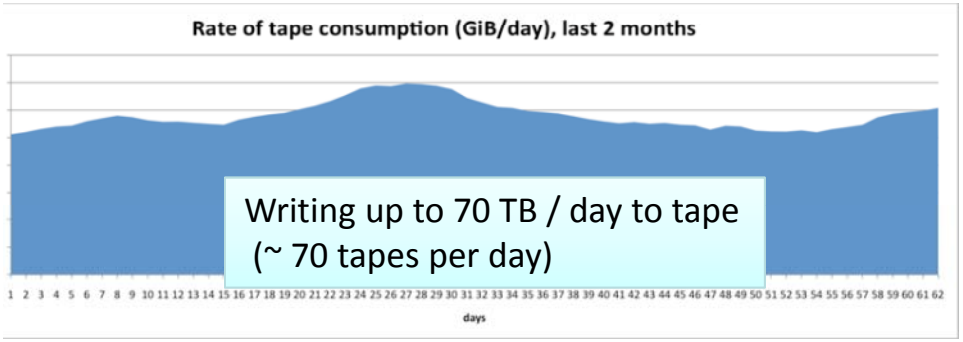
Excellent 10 Gb W-LCG connectivity



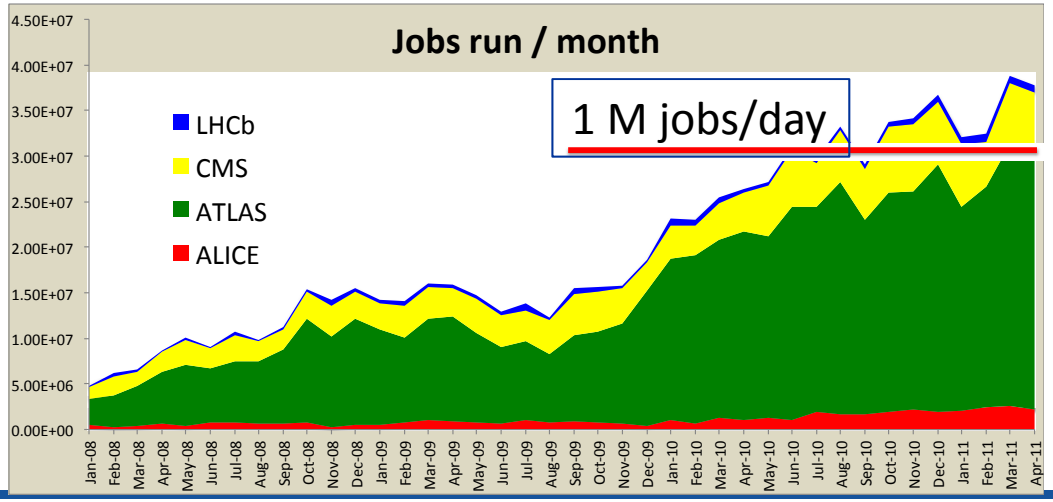
First year of LHC data (Tier0 and Grid)



- **Impressive numbers, we believe!**



Data written to tape (GB/day)

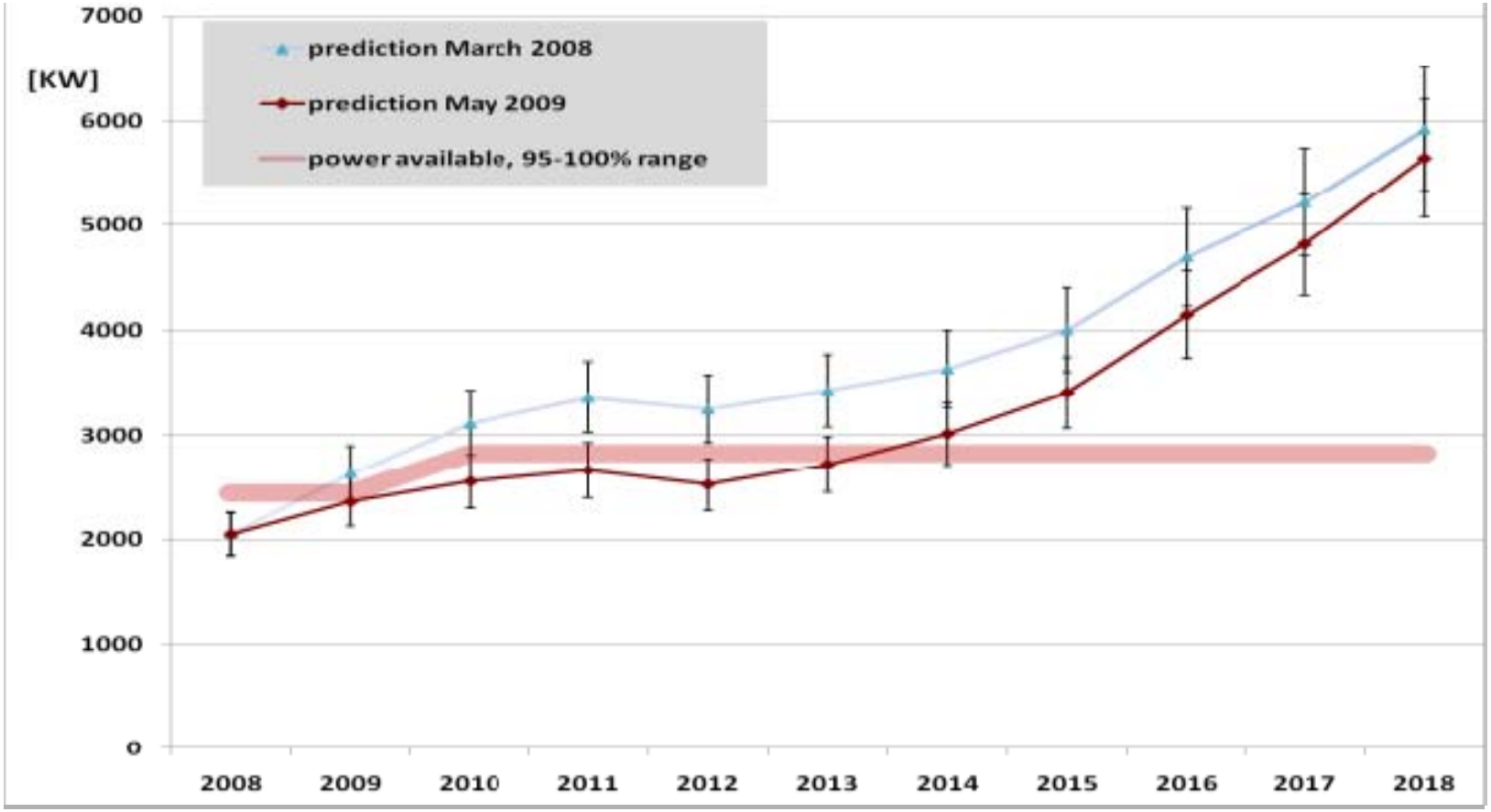


CERN's offline capacity

- High-throughput computing based on reliable "commodity" technology:
  - Scientific Linux
  - All inclusive: 7'800 dual-socket servers (64'000 cores)
  - Disk storage: 63'000 TB (usable) on 64'000 drives
  - Tape storage: 34'000 TB on 45'000 cartridges
    - 56'000 slots and 160 drives

Even CERN has a power problem

We are going to move from 2.9 MW to 3.5 MW.
Beyond this we will establish a remote Tier-0 in 2013!



W-LCG: A distributed supercomputer

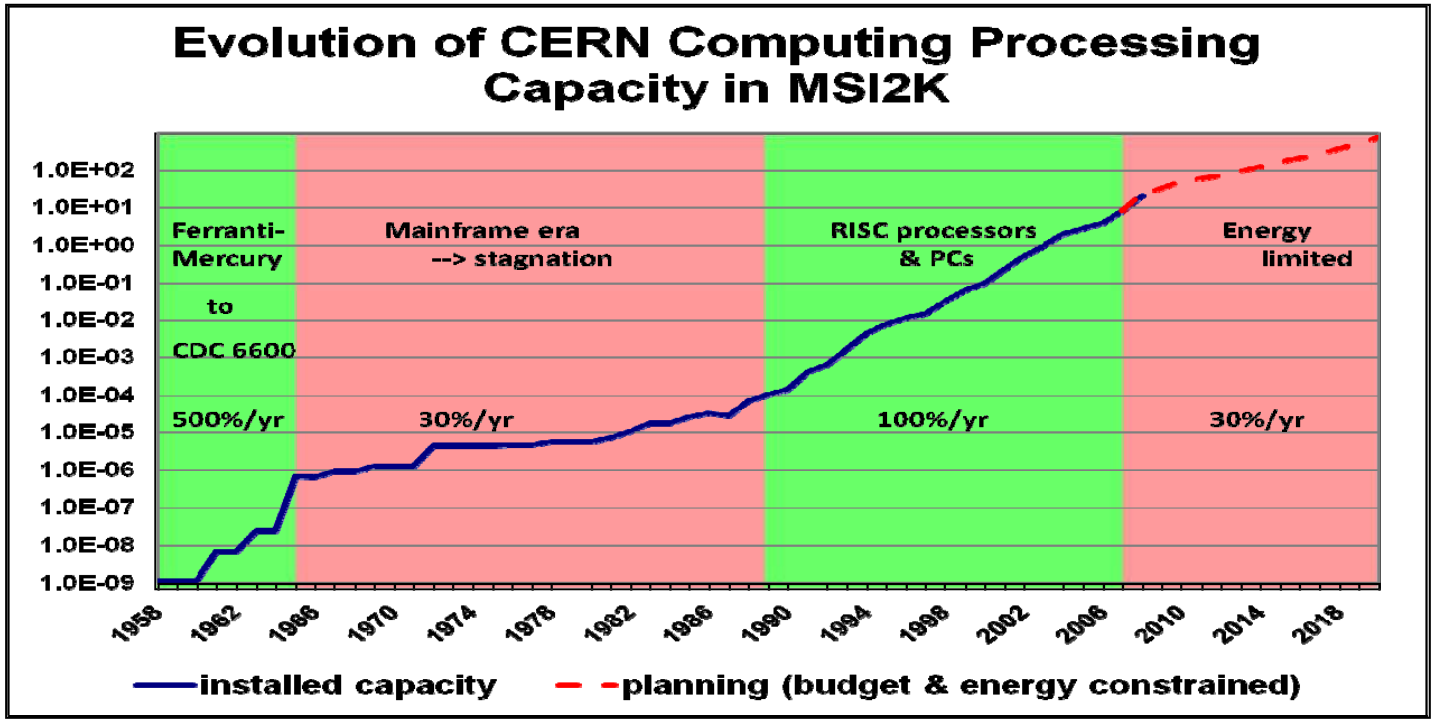
- Compared to TOP10 (Nov. 10)

Name/Location	Core count
Tianhe-1 (Tianjin)	186’368
Jaguar (Oak Ridge)	224’162
Nebulae – Dawning (NSCS)	120’640
Tsubame 2.0 (GSIC, Tokyo)	73’278
Hopper (DOE/NERSC)	153’408
Tera -100 – Bull (CEA)	138’368
Roadrunner (DOE/LANL)	122’400
Kraken XT5 (Tennessee)	98’928
Jugene (Jülich)	294’912
Cielo (DOE/SNL)	107’152

W-LCG
250’000
IA cores

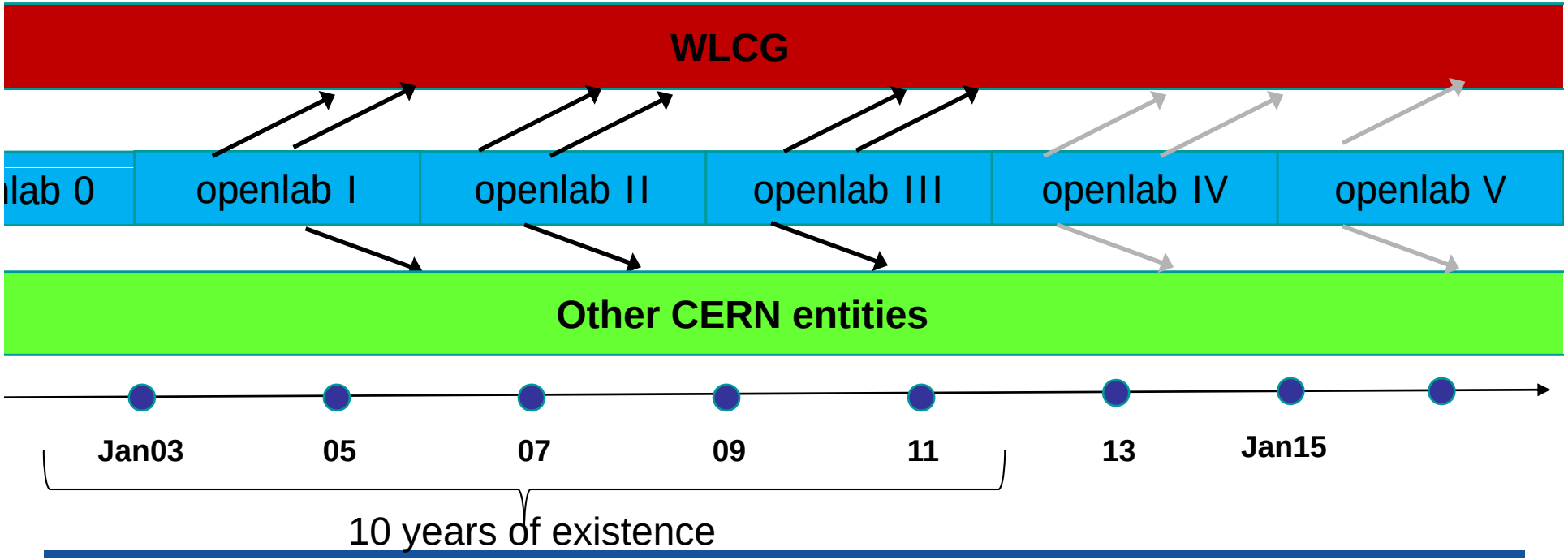
Insatiable appetite for computing

- During the era of the LEP accelerator (and beyond)
  - Compute power doubled every year
- We are desperately looking at all opportunities for this to continue



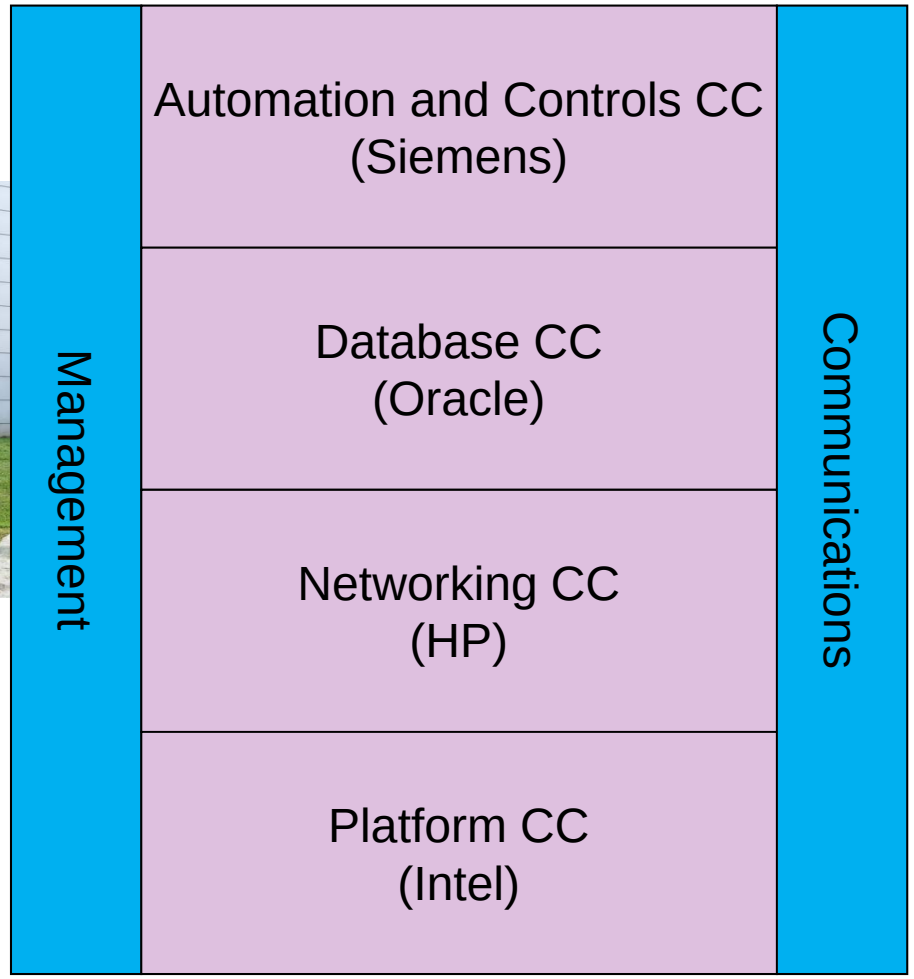
CERN openlab

- IT Department's main R&D focus
- Framework for collaboration with industry
- Evaluation, integration, validation
  - of cutting-edge technologies that can serve the LHC Computing Grid
- Sequence of 3-year agreements
  - 2003 – 2005: Phase I: the "opencluster" project
  - 2006 – 2011: Phase II & III: dedicated Competence Centres



CERN openlab structure

- **A solid set of Competence Centres**
  - With strong support from Management and Communications

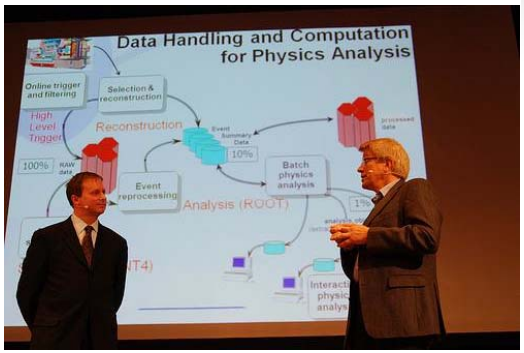


EXASCALE Capacity Computing R&D

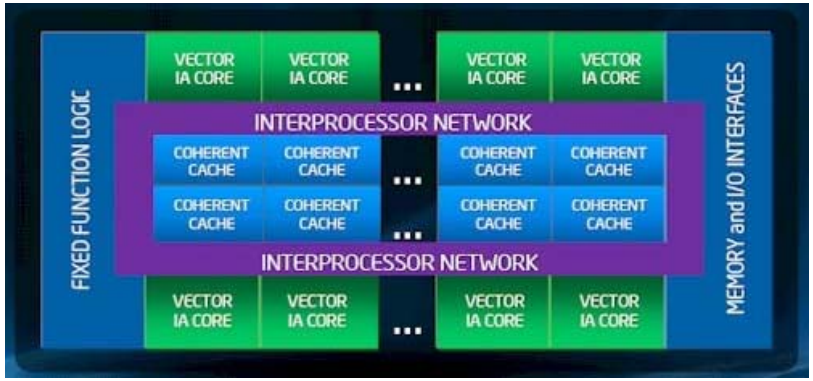
- **In openlab, we want to start an R&D project for Exascale**

- **Project goals:**
  - Identify constraints which might inhibit growth in CERN's Tier0 and in the W-LCG in the future.
  - Understand which software and hardware components must be moved towards the Exascale range.

Intel's "Many Integrated Core" Architecture



- **Announced at ISC10 (June 2010)**
  - S. Jarp on stage with K.Skaugen/Intel
- **Current version (codename "Knights Ferry SDP")**
  - Enhanced x86 instruction set with vector extensions
  - 32 cores + 4-way multithreaded + 512-bit vector units
- **Successful (easy) porting of our benchmark applications**
  - ALICE Trackfitter/Trackfinder
  - Multithreaded Geant4 prototype
  - Maximum Likelihood data analysis prototype



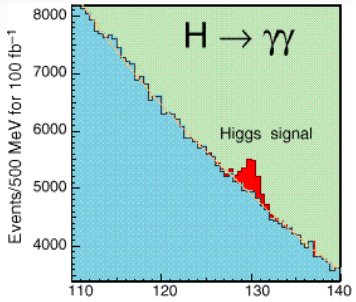
Graphics: INTEL

Conclusions

- **The Large Hadron Collider is foreseen to operate for the next 20 years!**
- **A Petascale Grid is currently in place for the computing tasks of the experiments**
- **We want to increase considerably the capacity of our Grid**
  - But, both power and cost are limiting factors
- **Planned and ongoing R&D activities should ease the move towards Exascale.**

BACKUP-2

High Energy Physics Computing Characteristics



- **Independent events (collisions of particles)**
  - trivial (read: pleasant) parallel processing
- **Millions of lines of in-house C++ code**
  - Most of the frameworks/toolkits are written by the physicists
- **Compute power scales with combination of SPECint and SPECfp**
  - Good double-precision floating-point (20% of total) is important!
  - Good math libraries needed
- **Current "HEPSPEC 2006" throughput benchmark for acquisitions (based on performance/W/CHF):**
  - 3 C++ jobs (INT) and 4 C++ jobs (FP)
- **Huge, but chaotic workload –**
  - research environment - physics extracted by iterative analysis

  → Unpredictable → unlimited demand