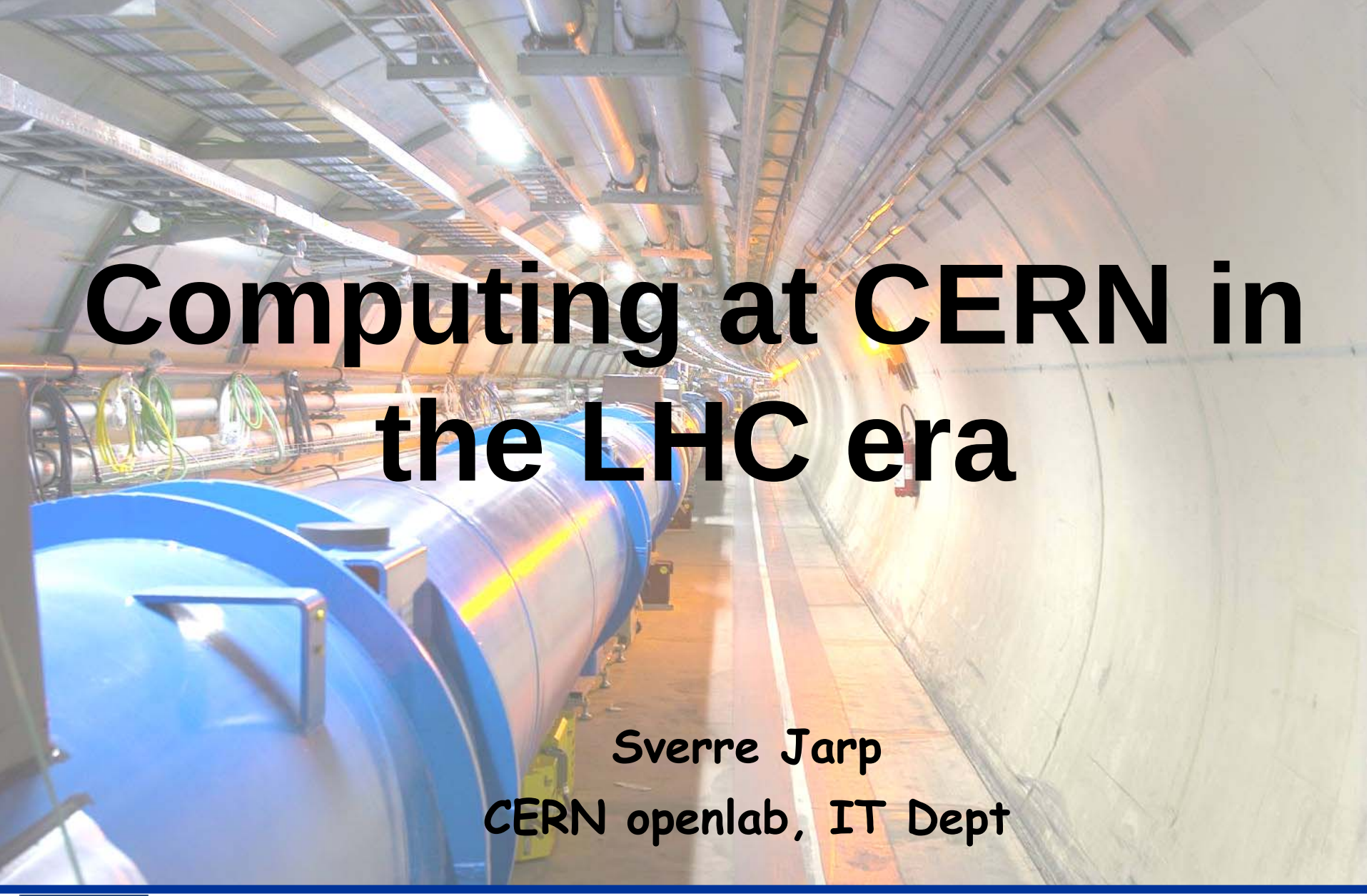


Computing at CERN in the LHC era

Sverre Jarp

CERN openlab, IT Dept

"where the Web was born"

Briefly about CERN

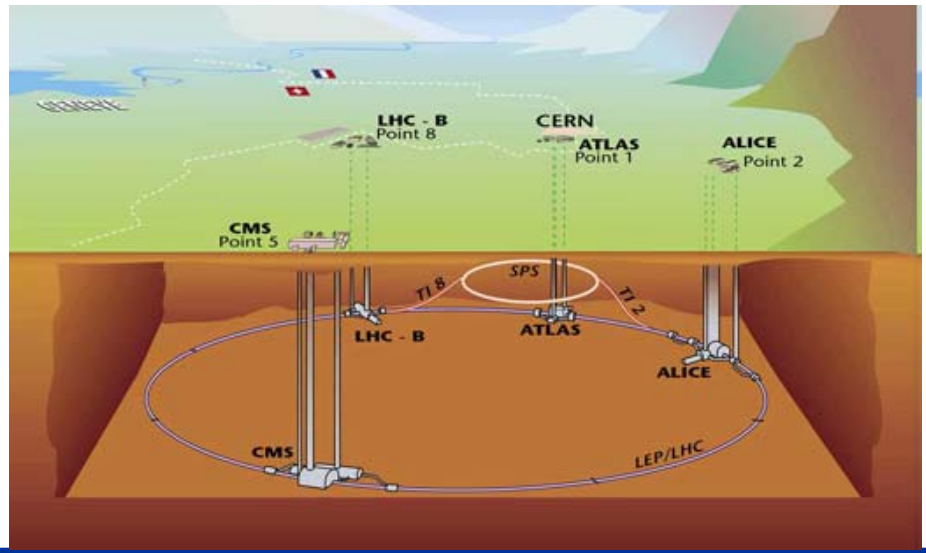
What is CERN?

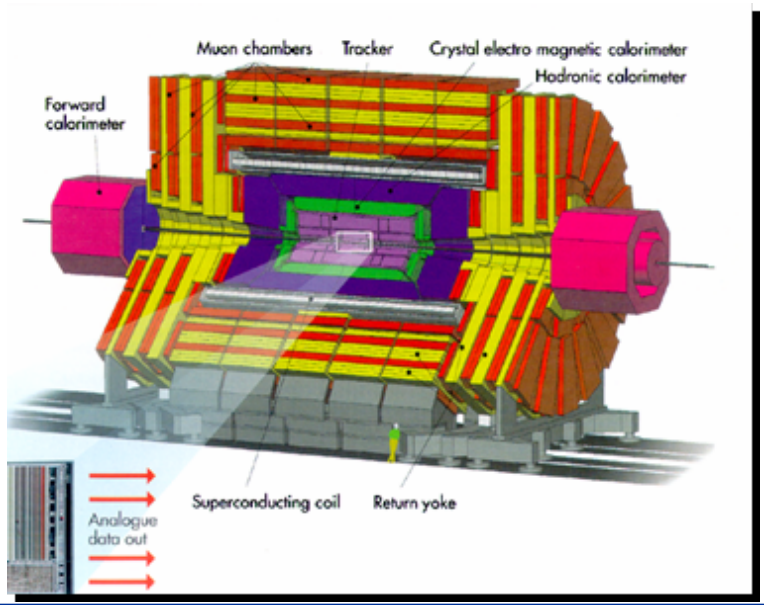
- CERN is the world's largest particle physics centre
- Particle physics is about:
  - elementary particles, the constituents from which all matter in the Universe is made
  - fundamental forces which hold matter together
- Particles physics requires:
  - special tools to create and study new particles

CERN's tools

The special tools for particle physics are:

- **ACCELERATORS**, huge machines (inside a complex underground structure) - able to accelerate particles to very high energies before colliding them into other particles
- **DETECTORS**, massive instruments which register the particles produced when the accelerated particles collide
- **COMPUTING**, to reconstruct the collisions, to extract the physics data and to perform the analysis





CERN in Numbers

- 2500 Staff
- 6500 Users
- 500 Fellows and Associates
- 80 Nationalities
- 500 Universities
- Budget ~1200 MCHF/year
  (~730 M€/year)

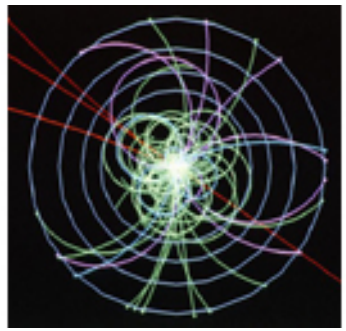
- **20 Member States:**
  Austria, Belgium, Bulgaria, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Italy, Netherlands, Norway, Poland, Portugal, Slovakia, Spain, Sweden, Switzerland and the United Kingdom.
- **8 Observers:**
  India, Israel, Japan, the Russian Federation, USA, Turkey, the European Commission and UNESCO

What is LHC?

LHC will be switched on in 2007

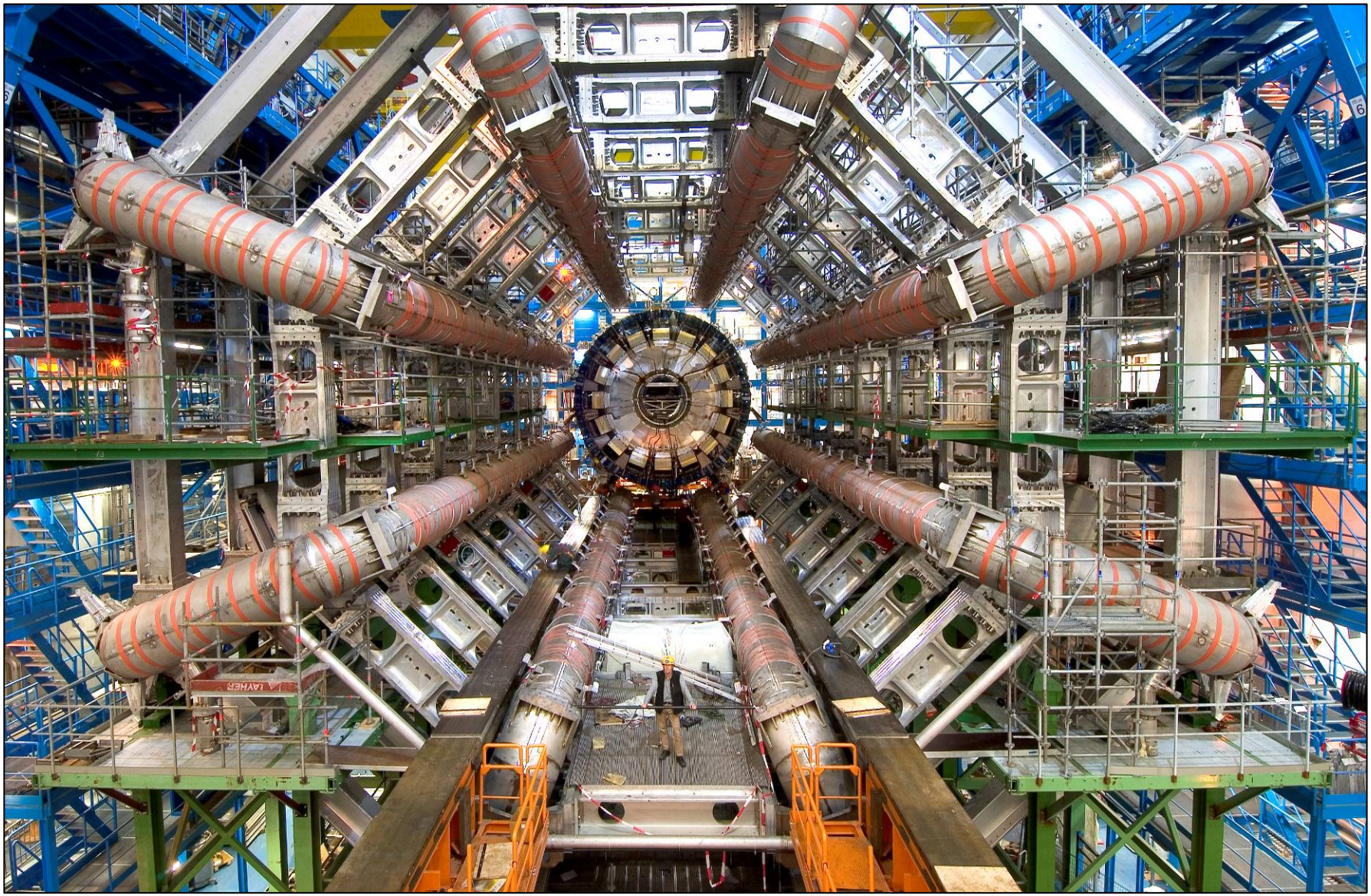
Four experiments, with detectors as 'big as cathedrals':
ALICE
ATLAS
CMS
LHCb

- It is a particle accelerator that will collide beams of protons at an energy of 14 TeV
- Using the latest super-conducting technologies, it will operate at about – 271°C, just above the absolute zero of temperature
- With its 27 km circumference, the accelerator will be the largest superconducting installation in the world.
- Its two proton beams will interact 40 million times per second (3000 bunches of 100 billion protons each)

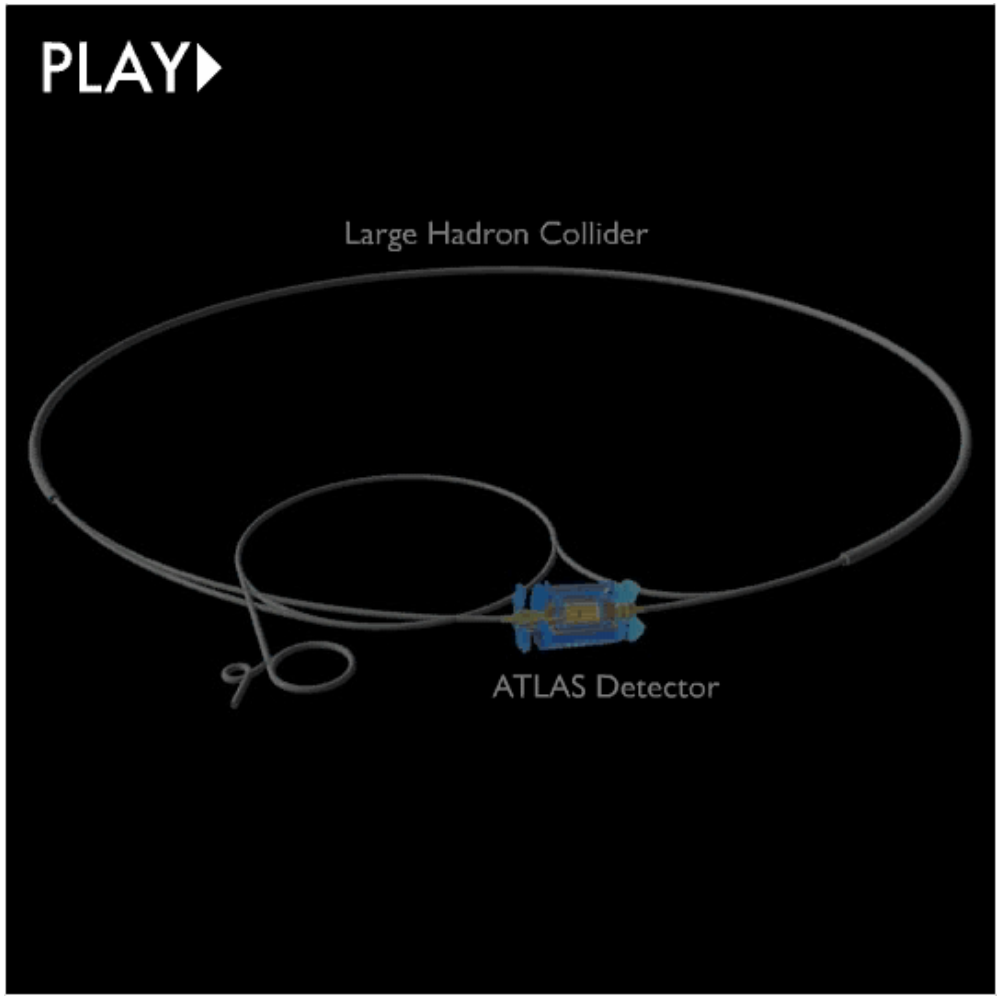


CERN openlab
for DataGrid applications

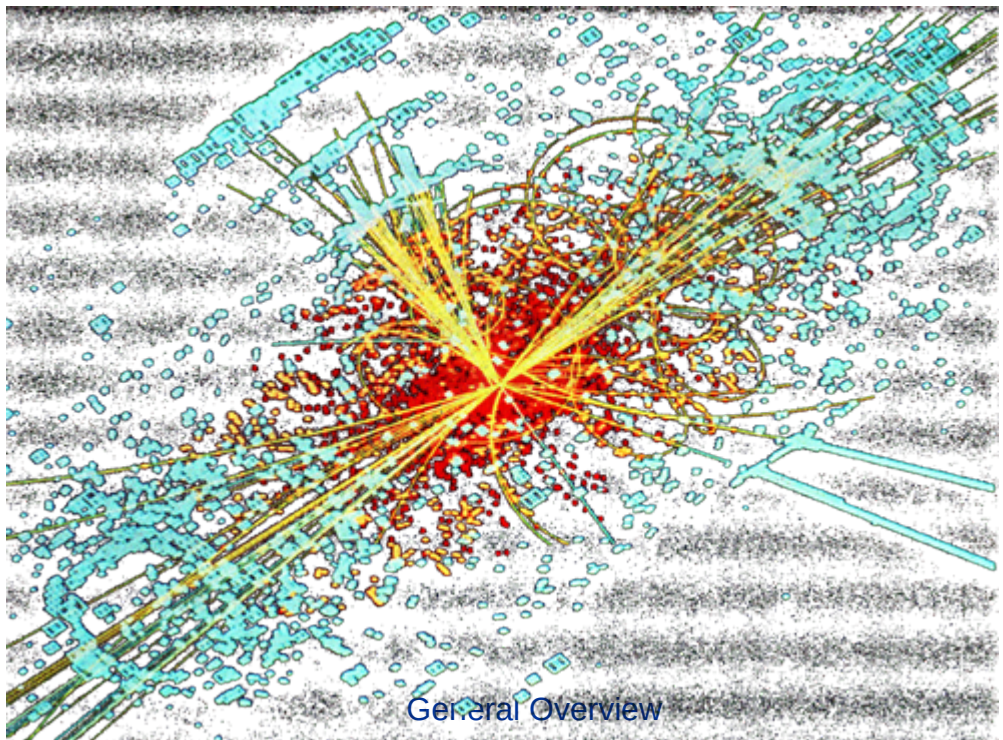
ATLAS construction



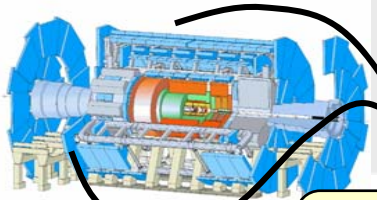
An ATLAS event



PHYSICS COMPUTING



Data Flow for Physics Analysis



detector

event filter (selection & reconstruction)

reconstruction

event data

processed data

100 %

raw data

10 %

event reprocessing

batch physics analysis

analysis

1 %

analysis objects
(extracted by physics topic)

event simulation

simulation

interactive physics analysis

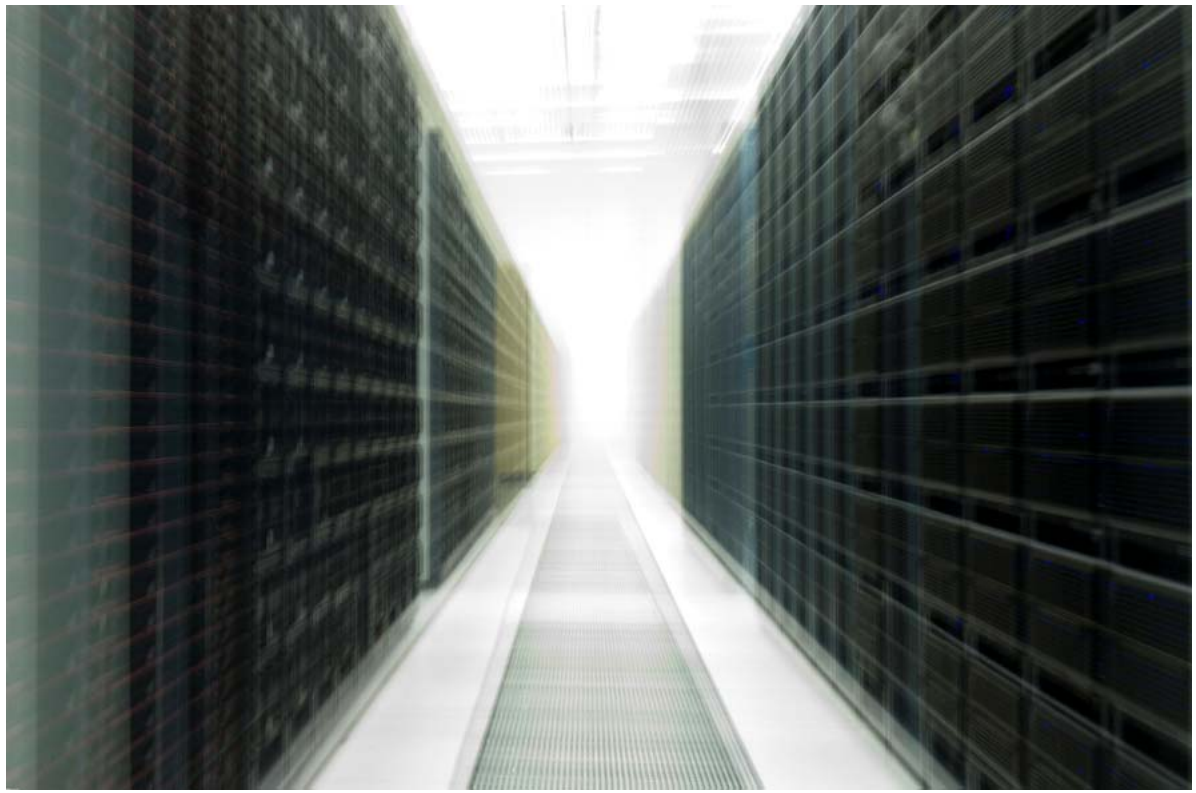
High Energy Physics Computing Characteristics

- **Independent events (collisions of particles)**
  - trivial (read: pleasant) parallel processing
- **Bulk of the data is read-only**
  - versions rather than updates
- **Meta-data in databases, but physics data in "flat" files**
- **Compute power measured in SPECint (rather than SPECfp)**
  - But good floating-point is important
- **Very large aggregate requirements:**
  - computation, data, input/output
- **Chaotic workload –**
  - research environment - physics extracted by iterative analysis, collaborating groups of physicists

  → Unpredictable → unlimited demand

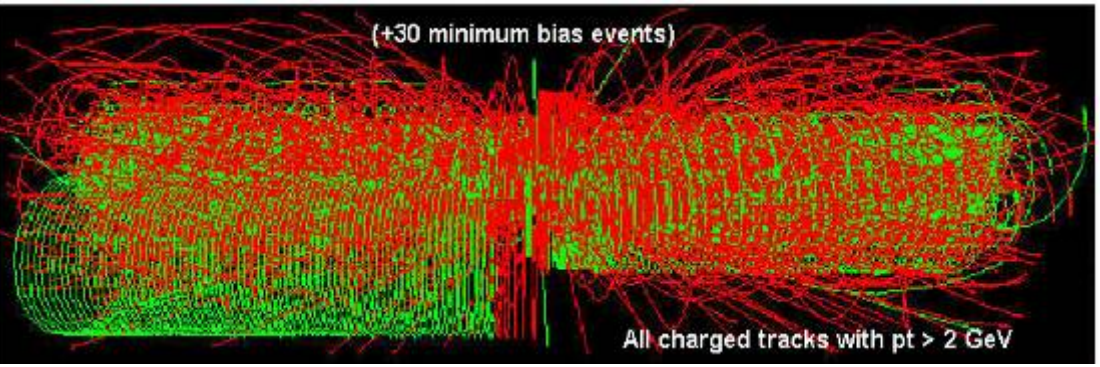
The Computing Environment

- High-throughput computing (based on reliable "commodity" technology)
  - Around 3000 (dual-socket Xeon) PCs with "Scientific Linux"
    - Now typically also "dual-core"
    - Quad-core expected for next acquisition

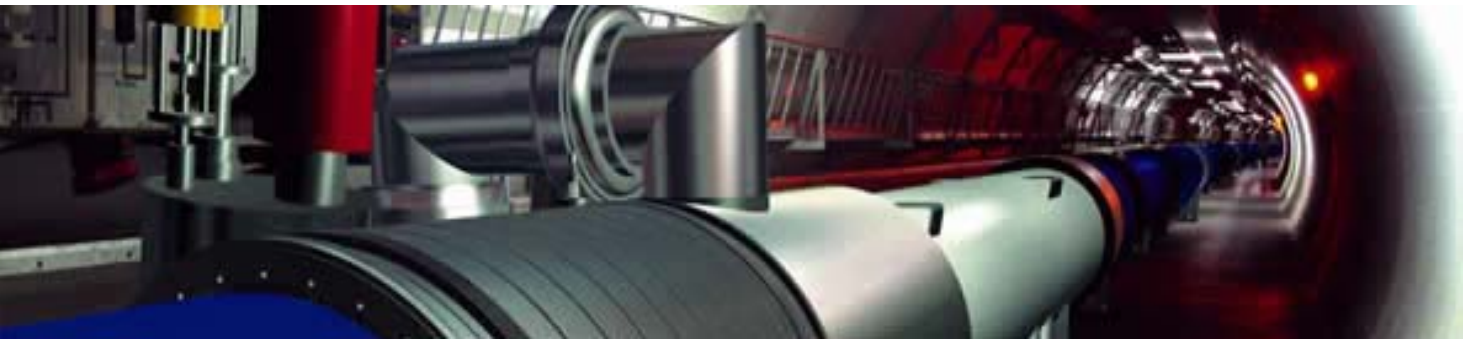


LHC DATA

Online computers filter out a few hundred "good" events per sec. Each event is ~1 MB.

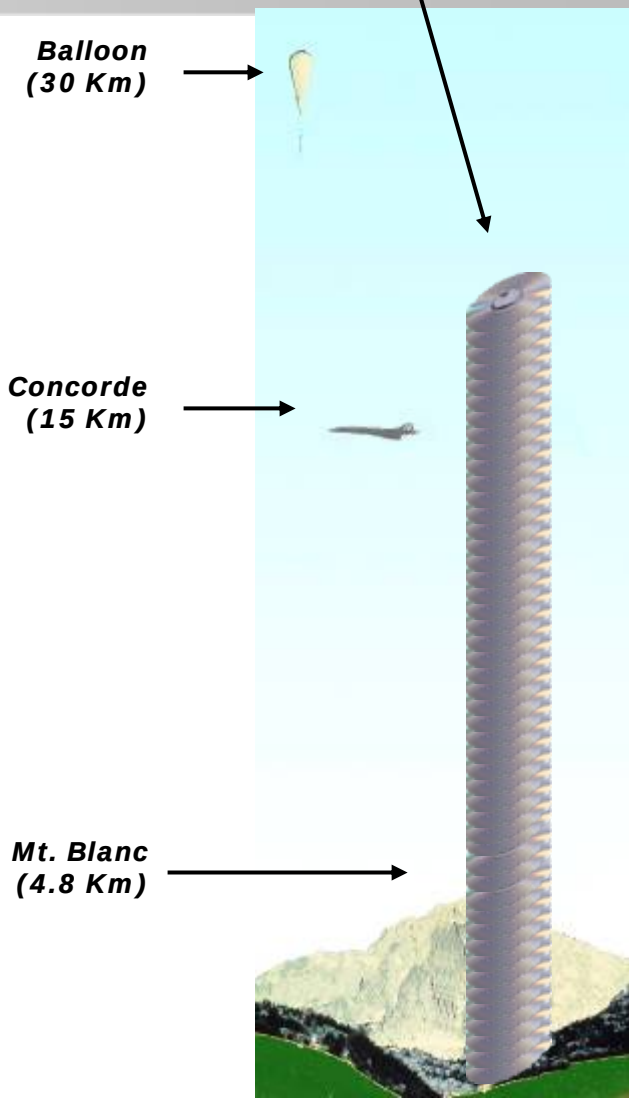


Which are recorded on disk and magnetic tape
at 100-1,000 Megabytes/sec → ~15 Petabytes per year
for all four experiments

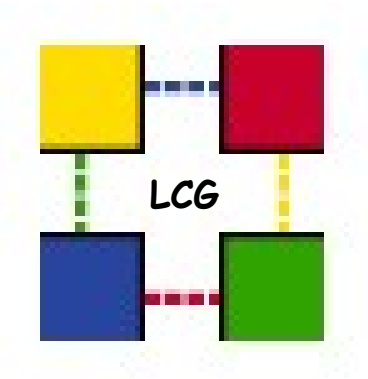


LHC data handling

- LHC data corresponds to about 20 million CDs each year
- Permanent storage → magnetic tape
- Transient storage → NAS servers
  - 20% of total

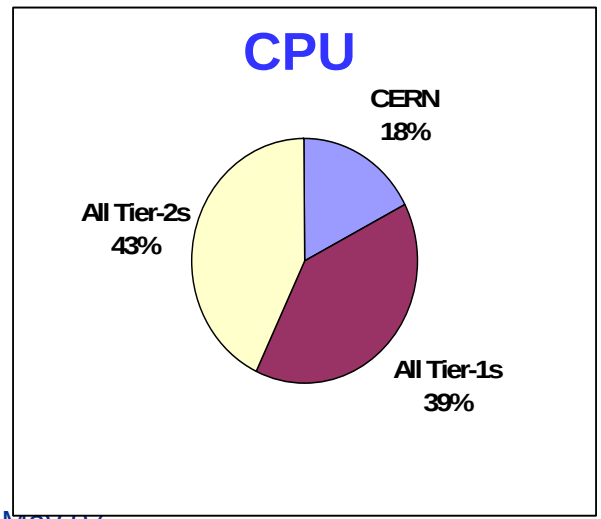


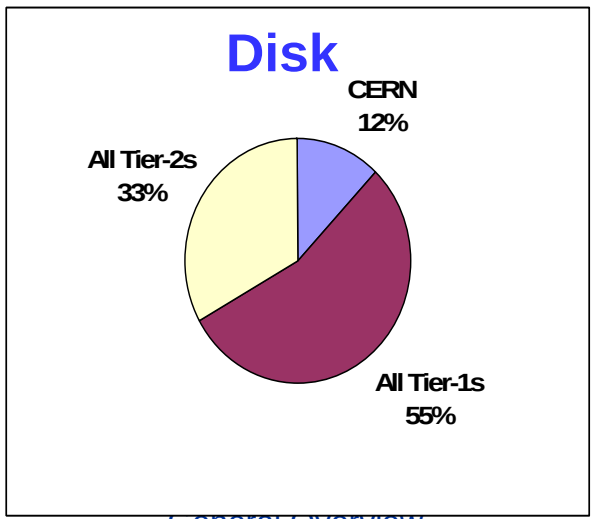
LCG (LHC Computing Grid)

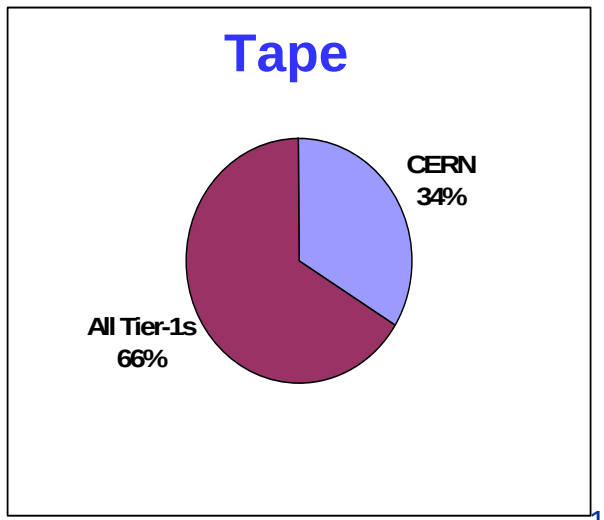


Why do we need a Grid?

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

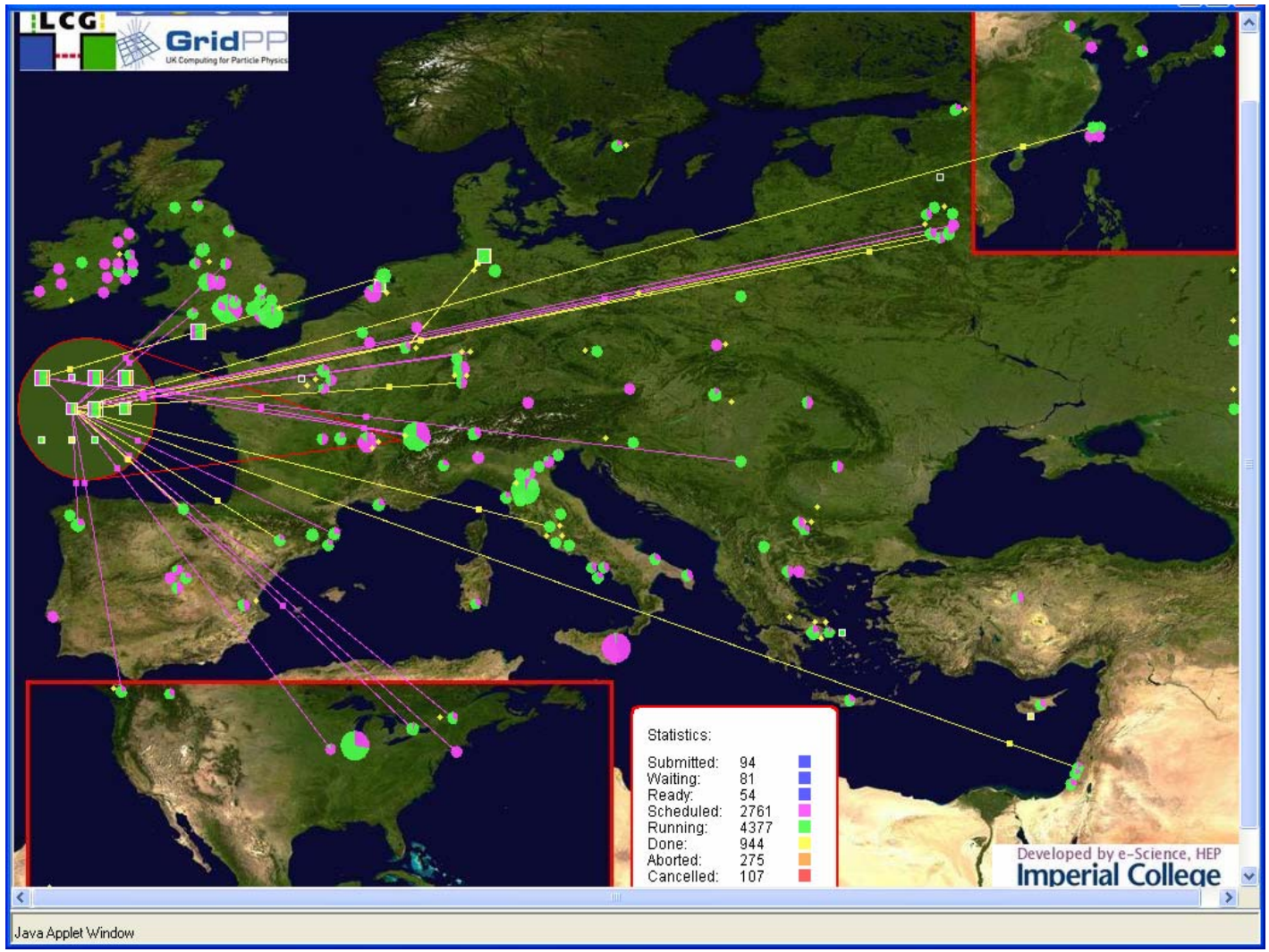




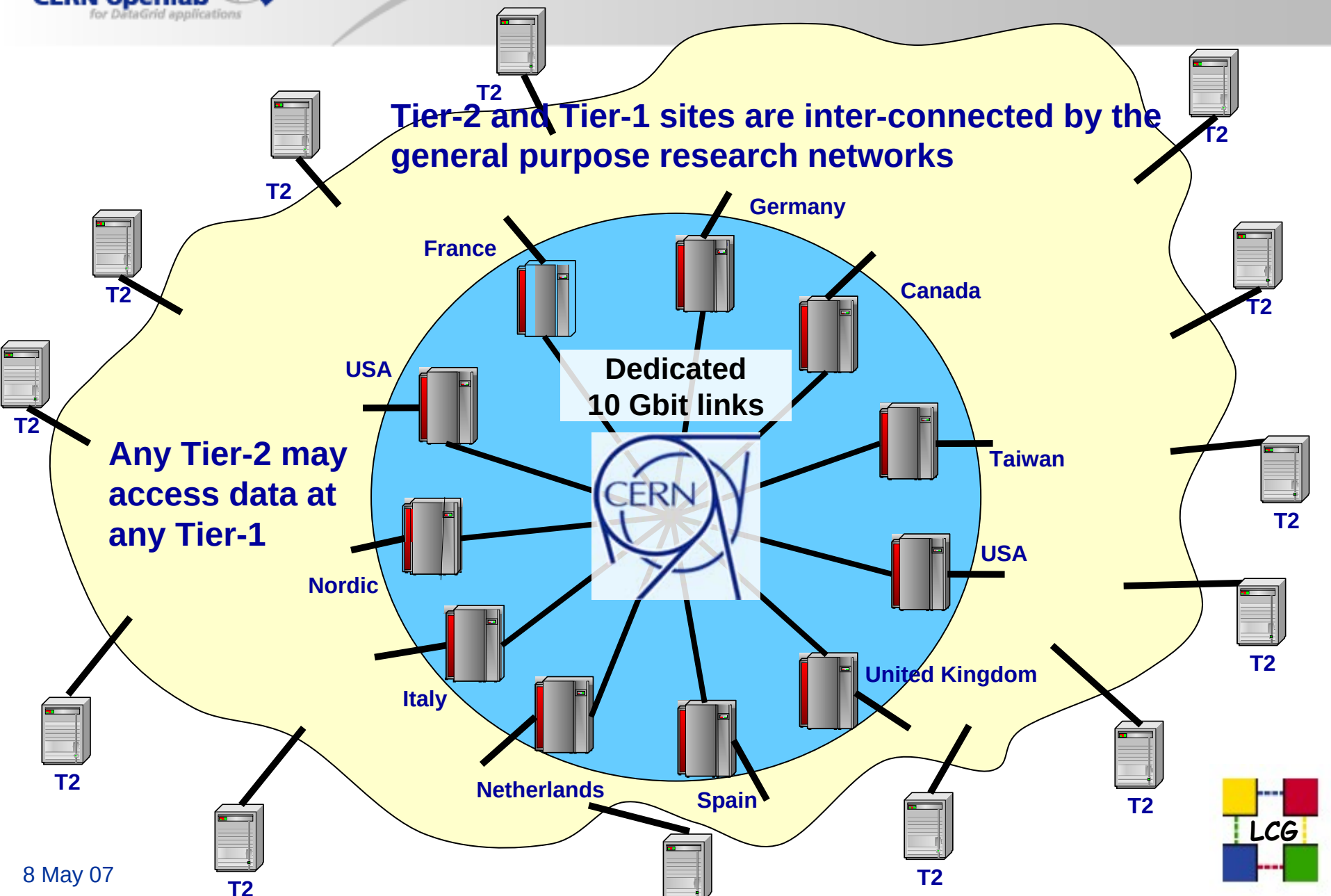


LCG-2

- Biggest Grid project in the world
- Almost 200 sites in 39 countries
- 30’000 IA-32 processors (w/Linux)
- Tens of petabytes of storage



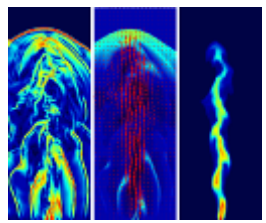
Powerful WAN as foundation



EGEE: Enabling Grids for E-sciencE

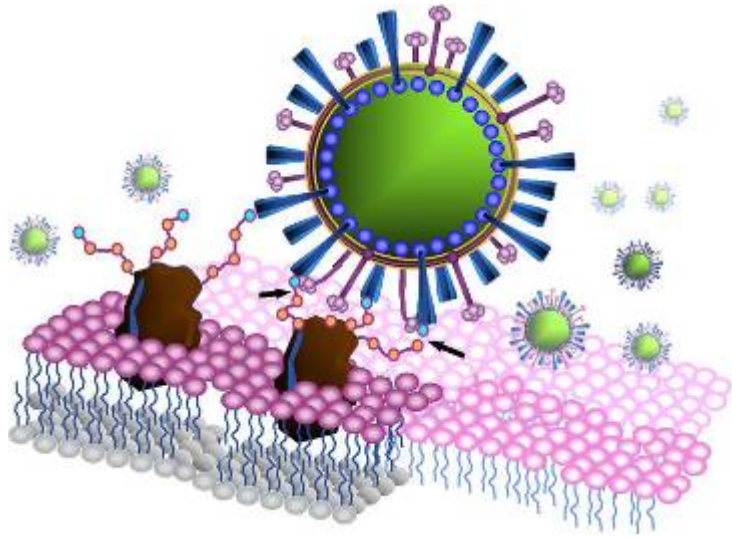
The EGEE project

- **EGEE**
  - 1 April 2004 – 31 March 2006
  - 71 partners in 27 countries, federated in regional Grids
- **EGEE-II**
  - 1 April 2006 – 31 March 2008
  - 91 partners in 32 countries
  - 13 Federations
- **Objectives**
  - Large-scale, production-quality infrastructure for e-Science
  - Attracting new resources and users from industry as well as science
  - Improving and maintaining “gLite” Grid middleware



Recent example: EGEE Attacks Avian Flu

- **EGEE used to analyse 300,000 possible potential drug compounds against bird flu virus, H5N1.**
- **2000 computers at 60 computer centres in Europe, Russia, Asia and Middle East ran during four weeks in April - the equivalent of 150 years on a single computer.**
- **Potential drug compounds now being identified and ranked.**



Neuraminidase, one of the two major surface proteins of influenza viruses, facilitating the release of virions from infected cells*. Image Courtesy Ying-Ta Wu, AcademiaSinica.*

CERN openlab

www.cern.ch/openlab

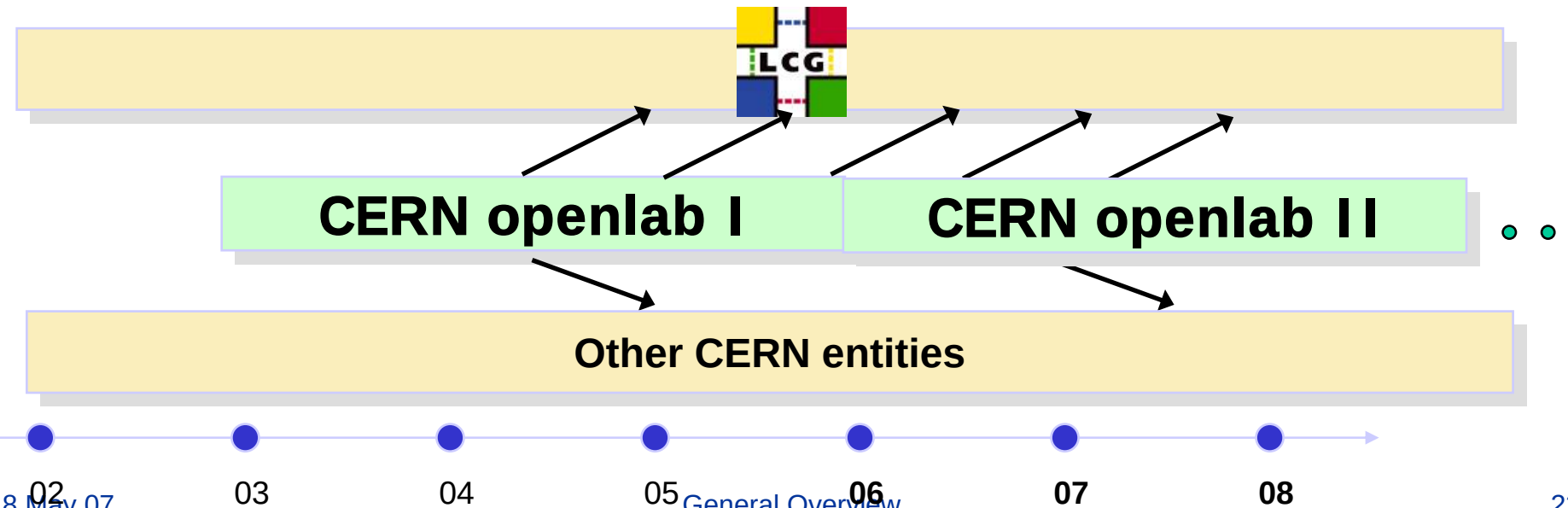
PARTNERS

CONTRIBUTORS

EDS

CERN openlab

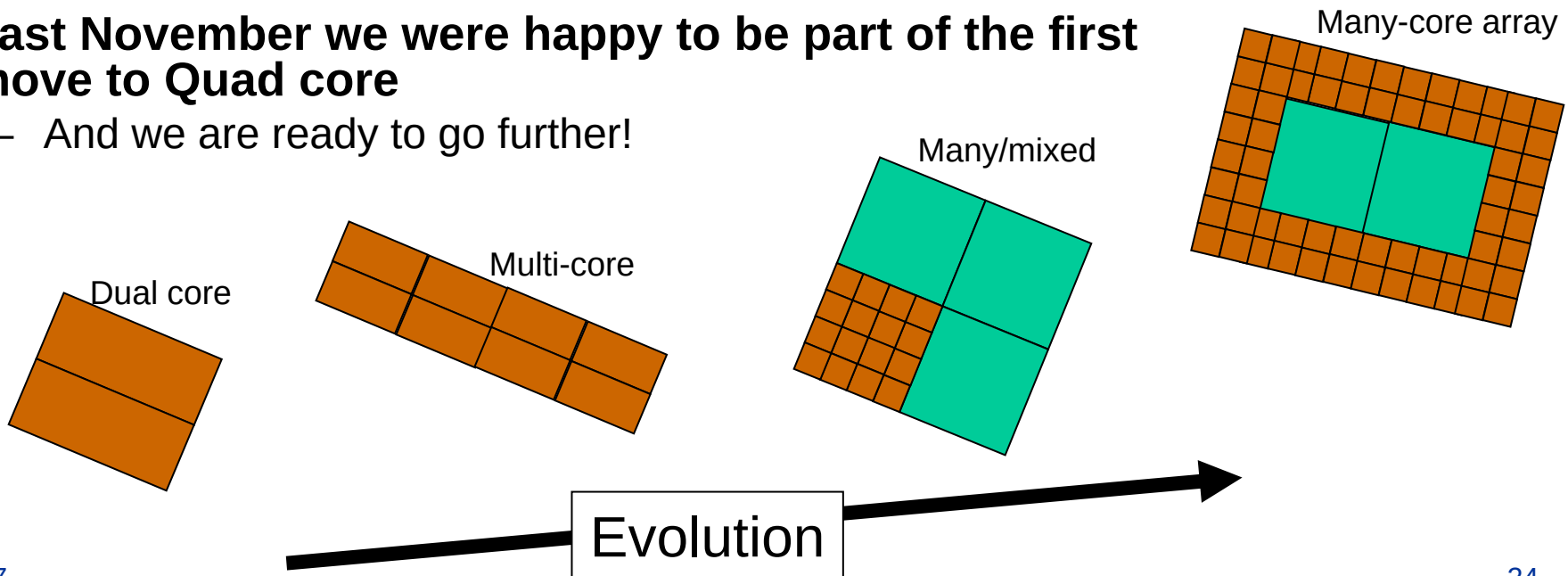
- CERN-IT department's main R&D focus
- Framework for collaboration with industry
- Evaluation, integration, validation
  - of cutting-edge technologies that can serve the LHC Computing Grid (LCG)
- Sequence of 3-year agreements
  - 2003 – 2005: the "opencluster" project
  - 2006 – 2008: openlab Phase II with new projects:
    - Platform, Grid, databases, Network/Security



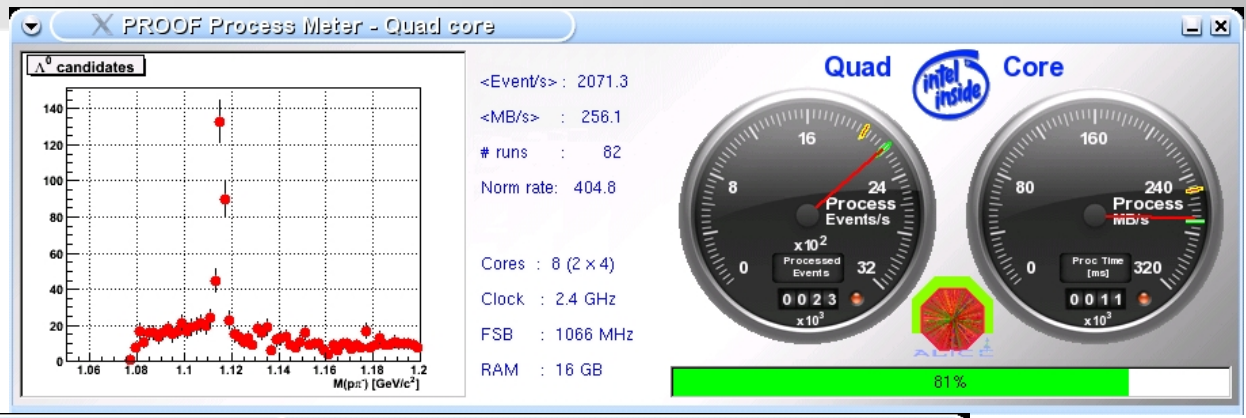
Multicore/Manycore

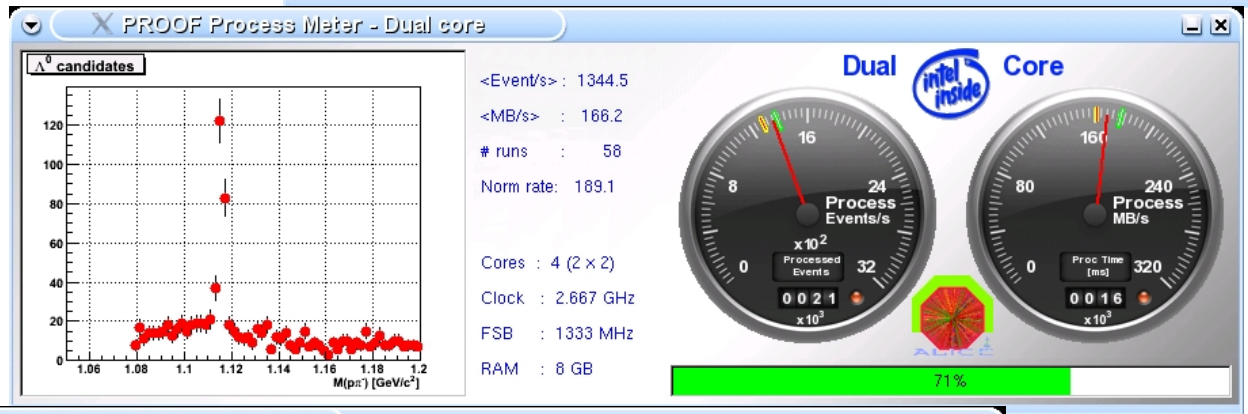
- **Our “high throughput” computing model is maybe ideally suited:**
  - Independent processes can run on each core, provided that:
    - Main memory is added
    - Bandwidth to main memory remains reasonable
  - Testing, so far, has been very convincing
    - Initially on Dual Core systems (Dempsey, Woodcrest, Montecito, etc.)

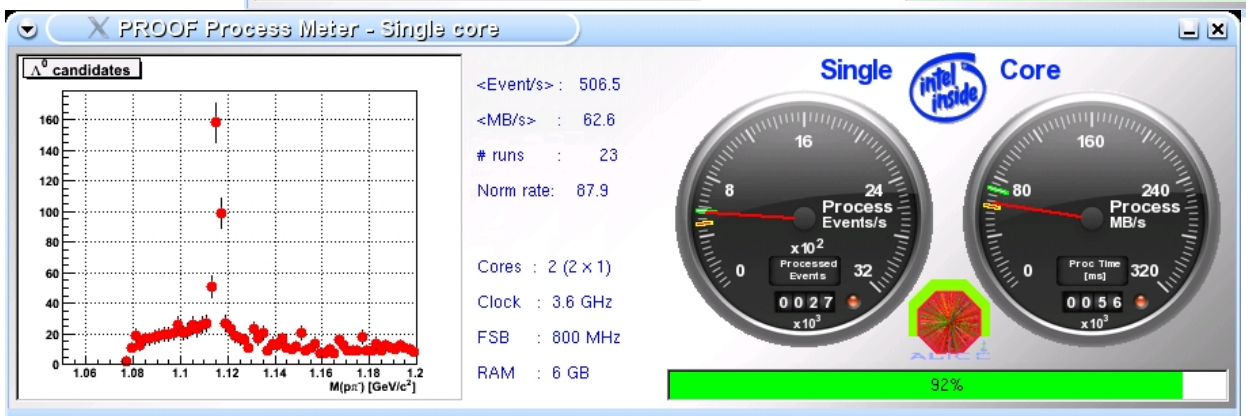
- **Last November we were happy to be part of the first move to Quad core**
  - And we are ready to go further!



Multicore comparisons



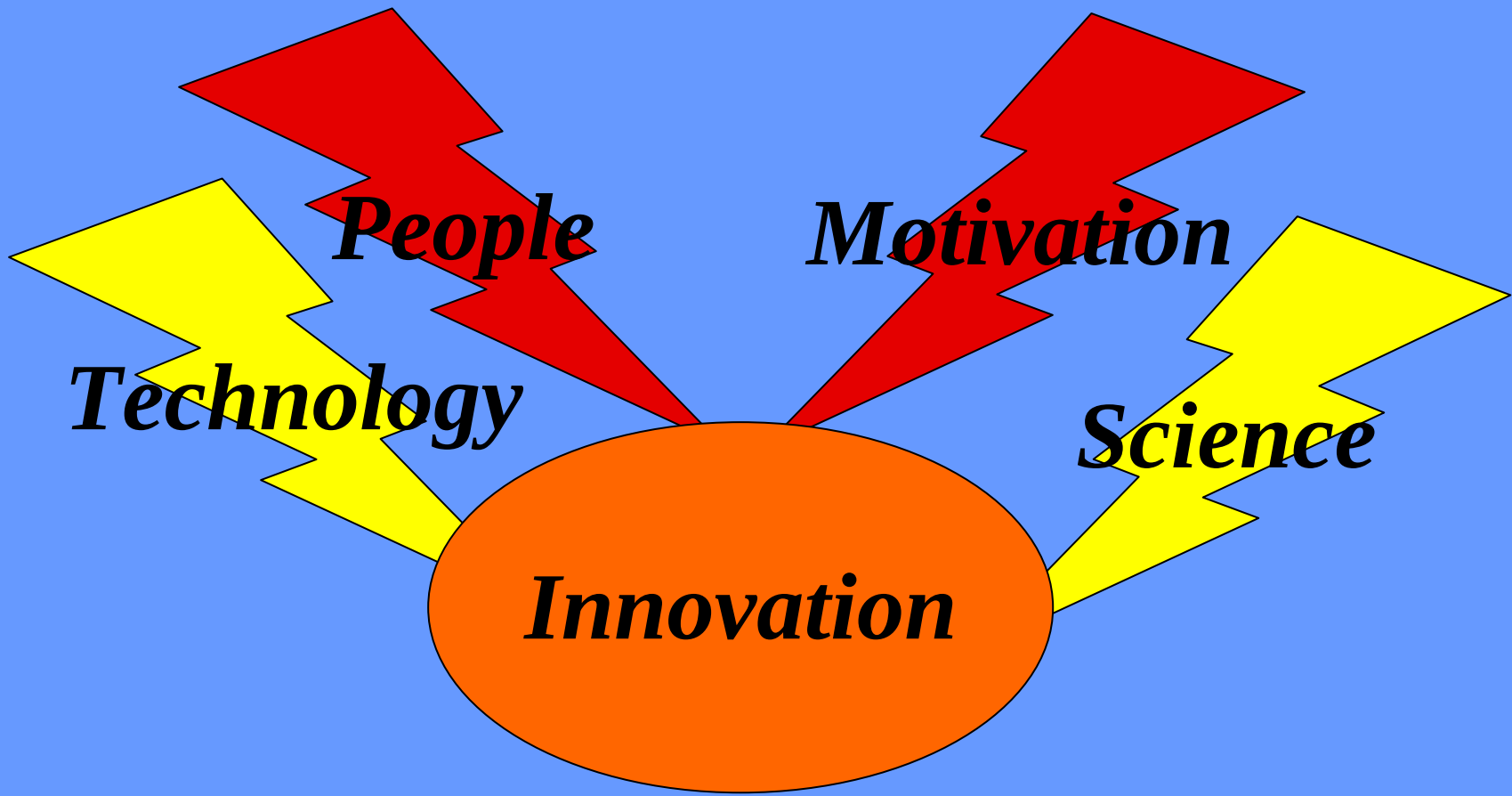




Conclusions

- **CERN is busily preparing for the arrival of LHC data in one year's time!**
  - New and exciting technologies will be used to cope with the data
    - 10 Gb networking
    - Terabyte disk and tape technology
    - 64-bit processors with multicore and virtualization capabilities
  - Our Grid offers seamless integration, all around the globe
    - Together with our partners (EU, industrial partners, other Physics Labs, other sciences) we expect to continue to come up with interesting proofs-of-concept and technological spin-off !
- **High Throughput Computing is "on the move" !**

LHC Computing



Grids serving science

- Physics/Astronomy ***(data from different kinds of research instruments)***
- Medical/Healthcare ***(imaging, diagnosis and treatment)***
- Bioinformatics ***(study of the human genome and proteome to understand genetic diseases)***
- Nanotechnology ***(design of new materials from the molecular scale)***
- Engineering ***(design optimization, simulation, failure analysis and remote Instrument access and control)***
- Natural Resources and the Environment ***(weather forecasting, earth observation, modeling and prediction of complex systems: river floods and earthquake simulation)***