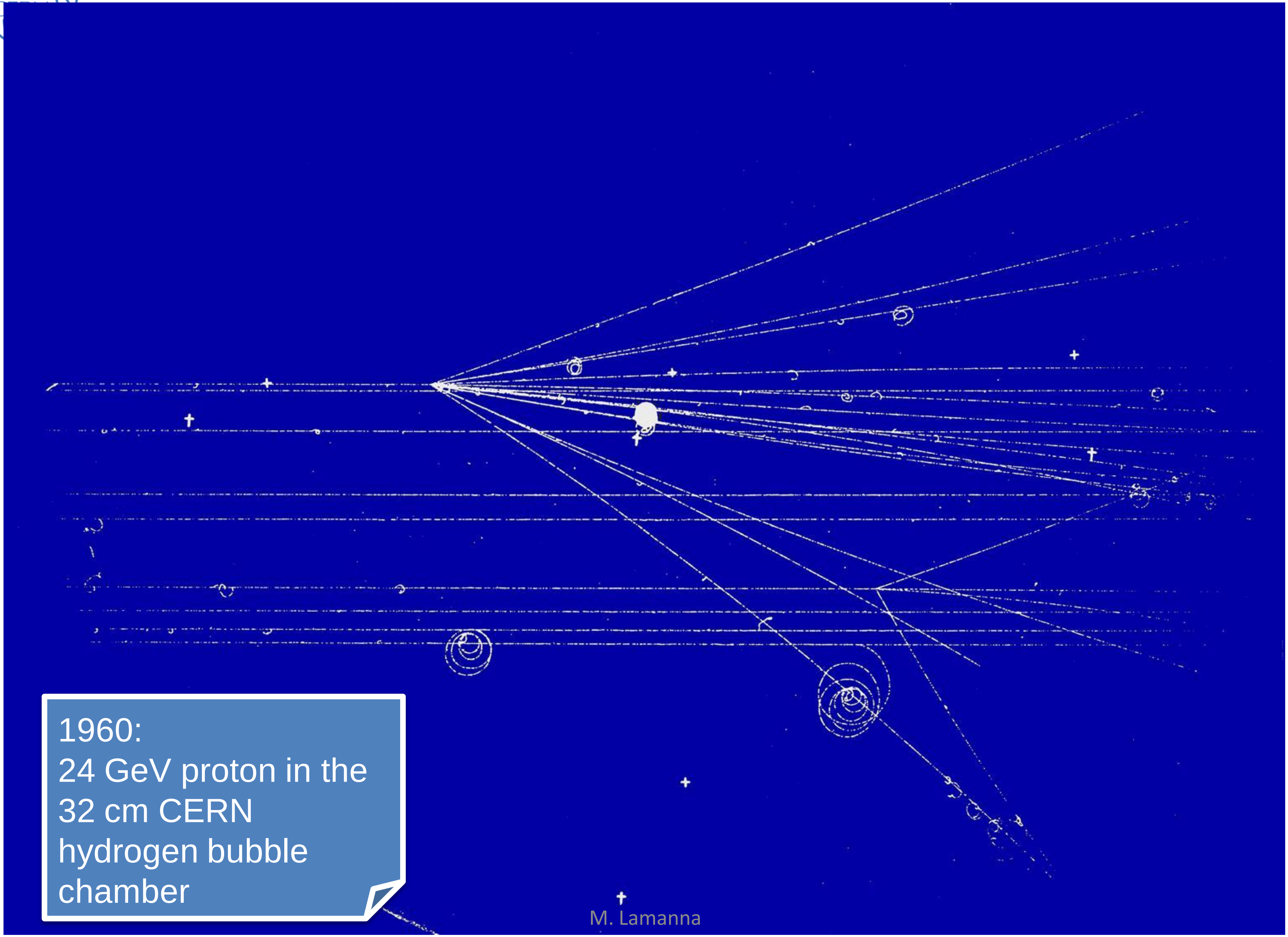
LHC Grid computing

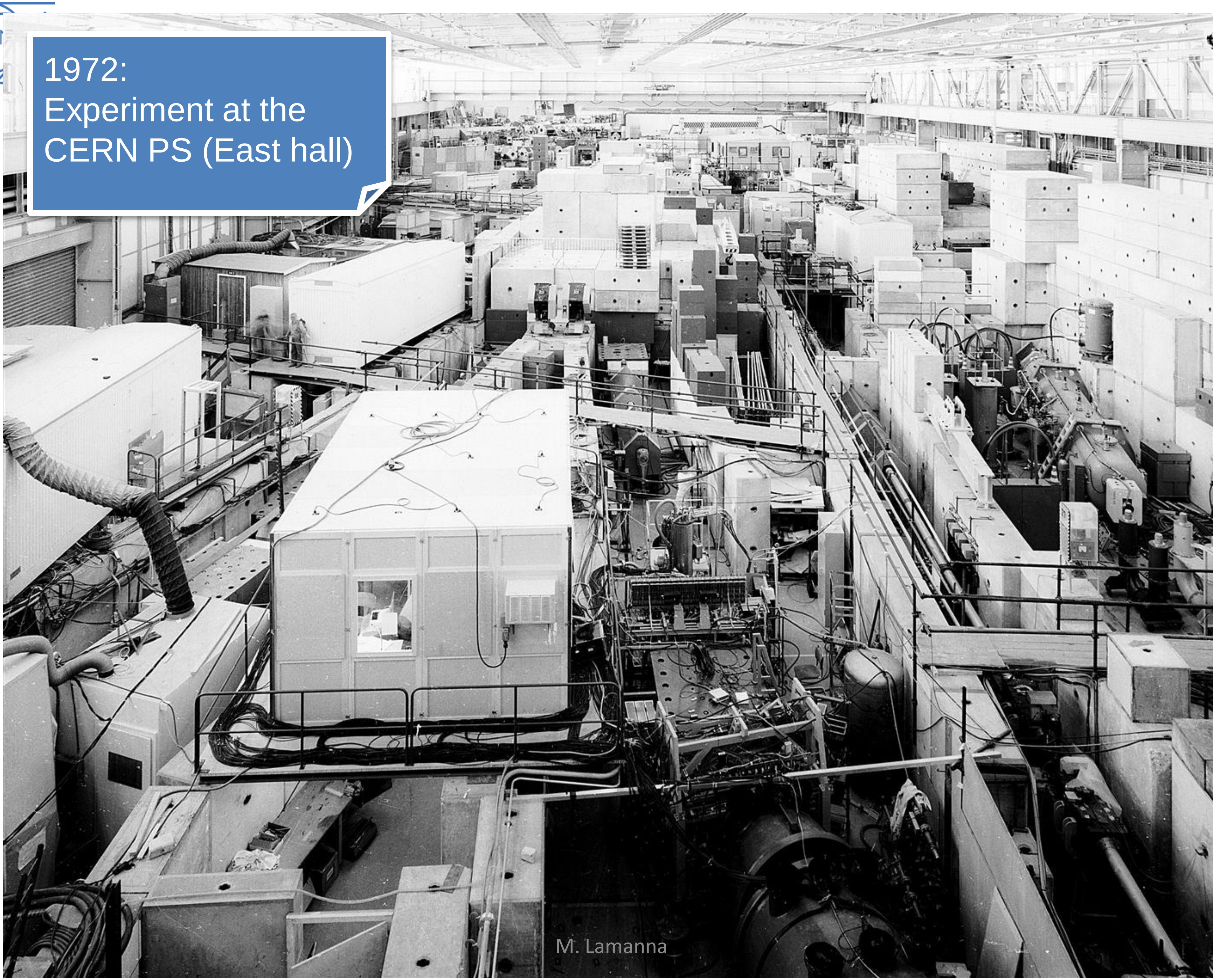
Massimo Lamanna

CERN
Information Technology Department
Grid Support Group



1960:
24 GeV proton in the 32 cm CERN hydrogen bubble chamber

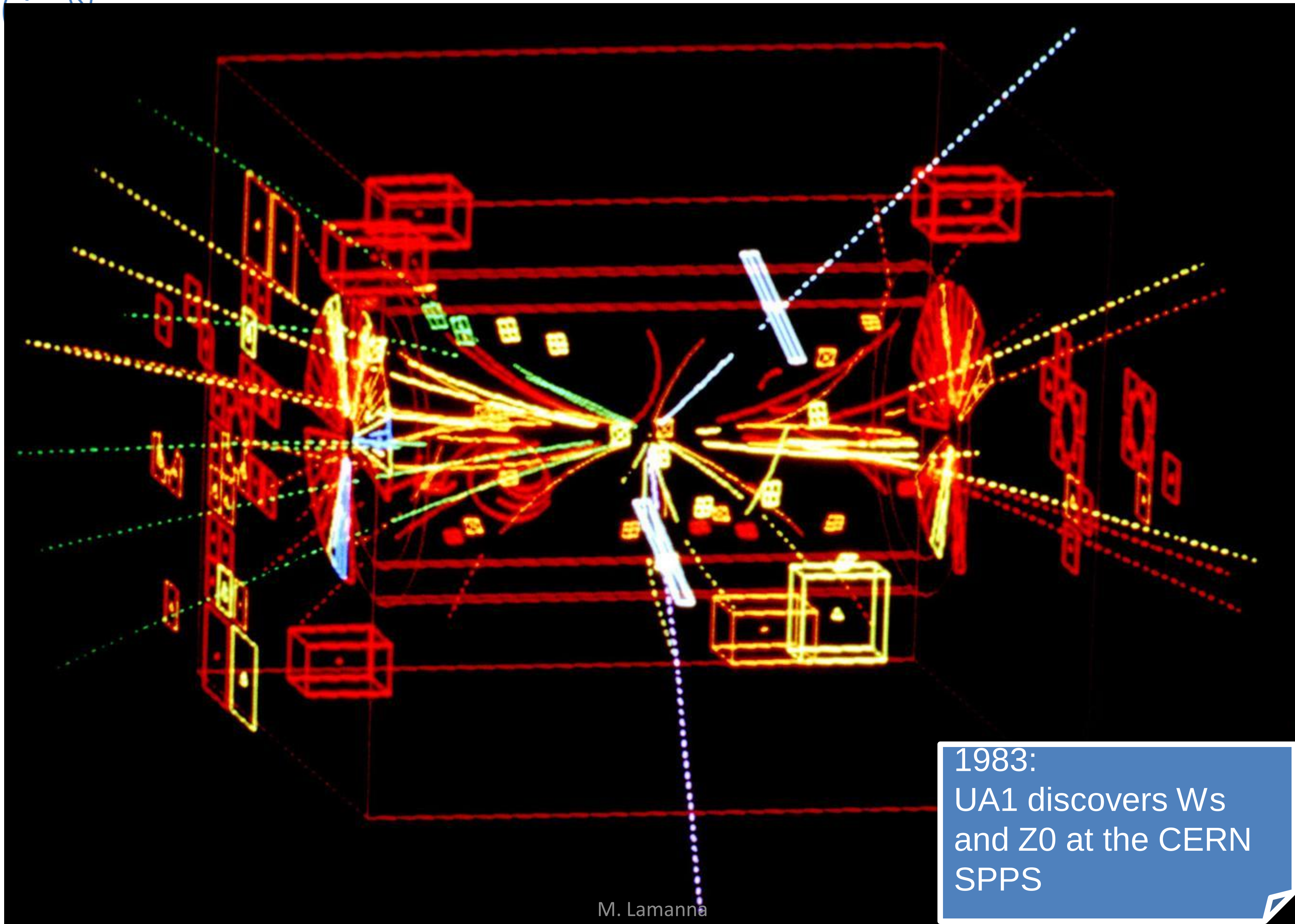
1960:
IBM 709 at Geneva airport



1972:
Experiment at the
CERN PS (East hall)

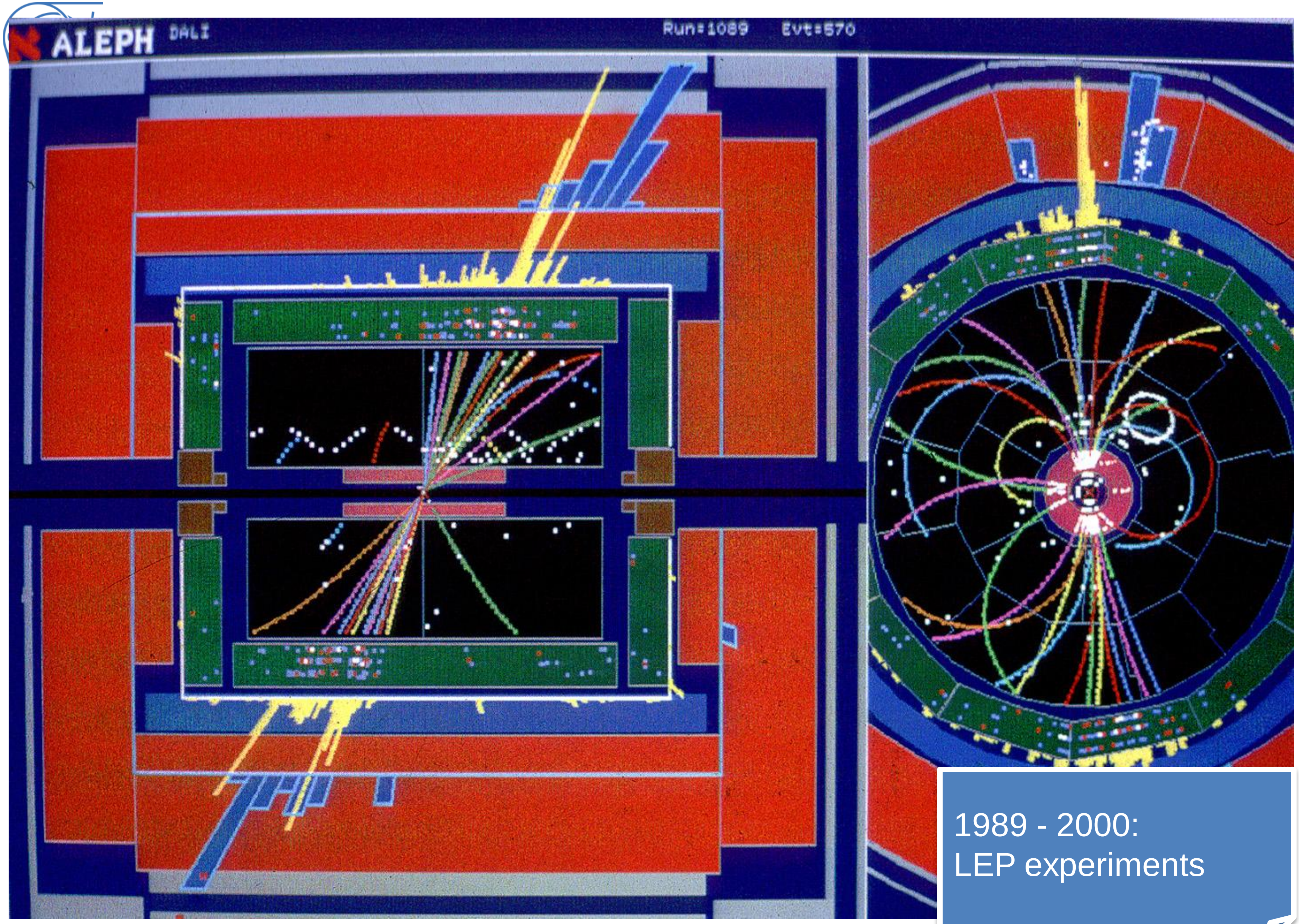
M. Lamanna

1972:
Installation of the
CDC 7600
(Bd 513)



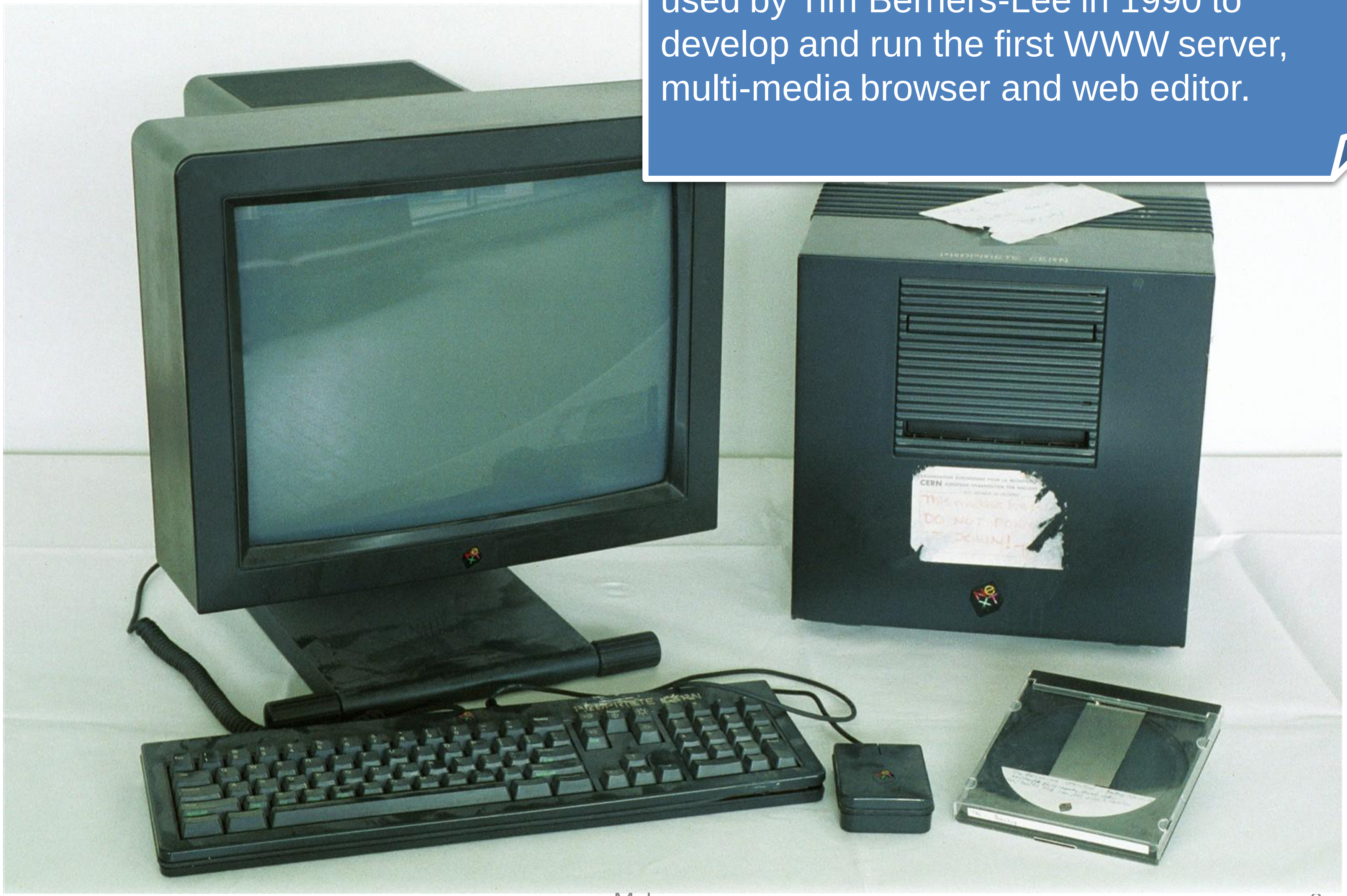
1983:
UA1 discovers Ws and Z0 at the CERN SPPS

1983:
Computer centre mainframes (IBM, Siemens, CDC)
(Bd 513)



1989 - 2000:
LEP experiments

1990:
The first web server: this machine was used by Tim Berners-Lee in 1990 to develop and run the first WWW server, multi-media browser and web editor.

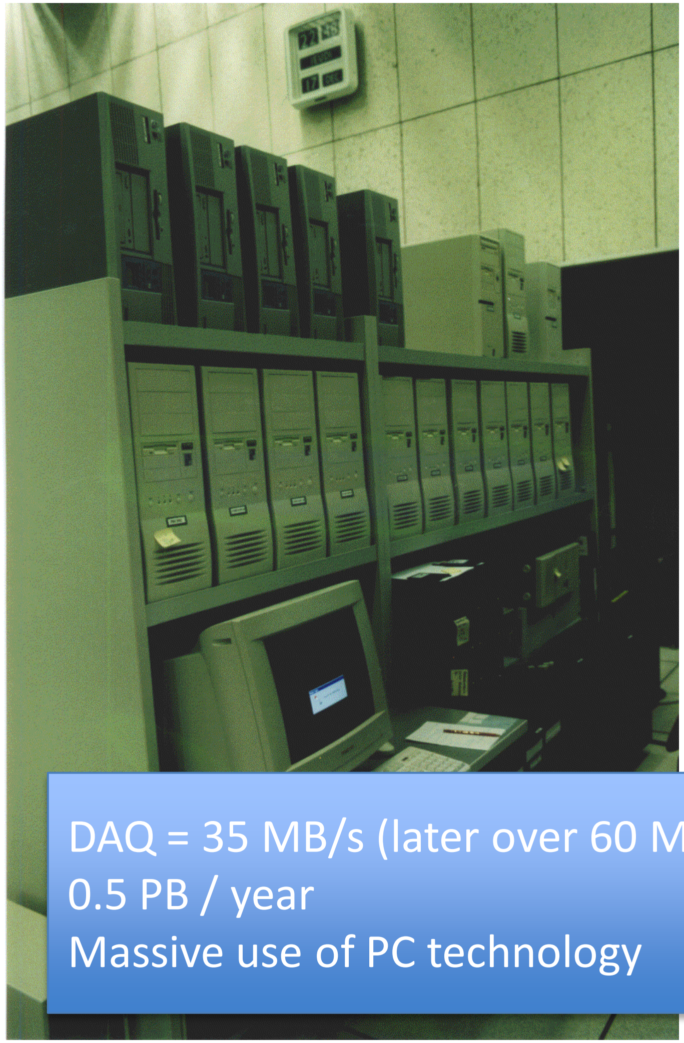


Early days of COMPASS (1998)

A new experiment being built…

The SM2 magnet…

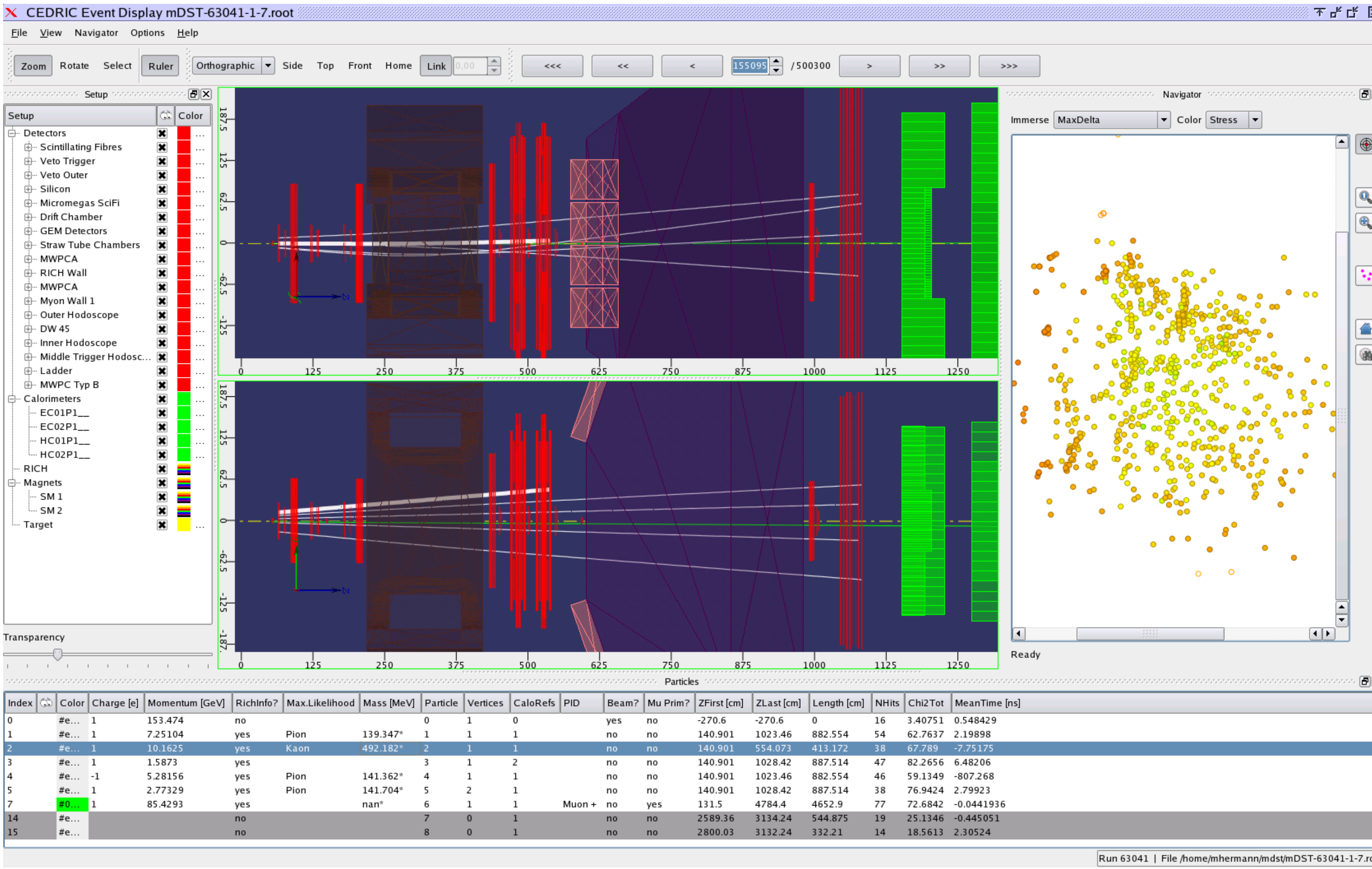
… with new computing challenges



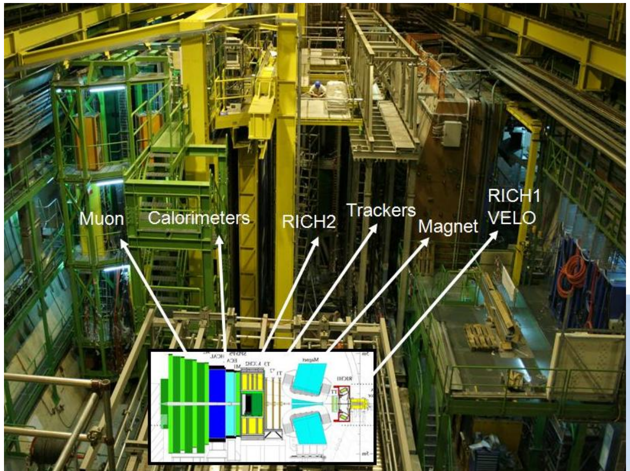
DAQ = 35 MB/s (later over 60 MB/s)
0.5 PB / year
Massive use of PC technology

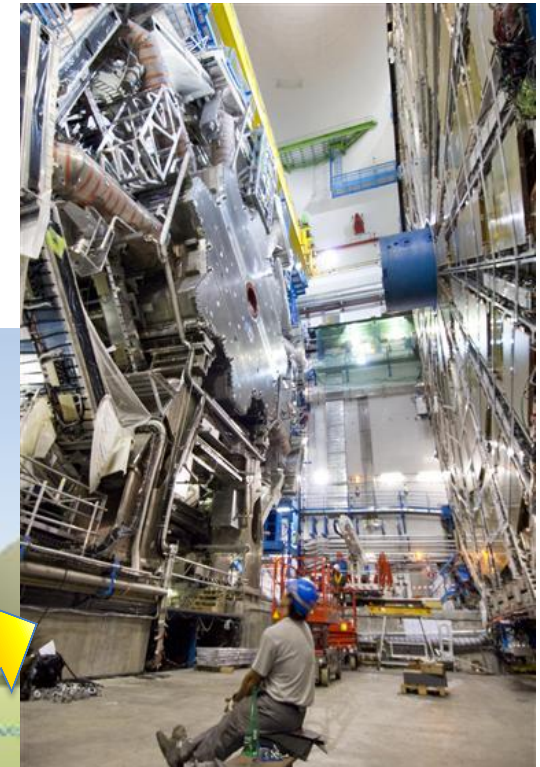
The COMPASS Computing Farm…

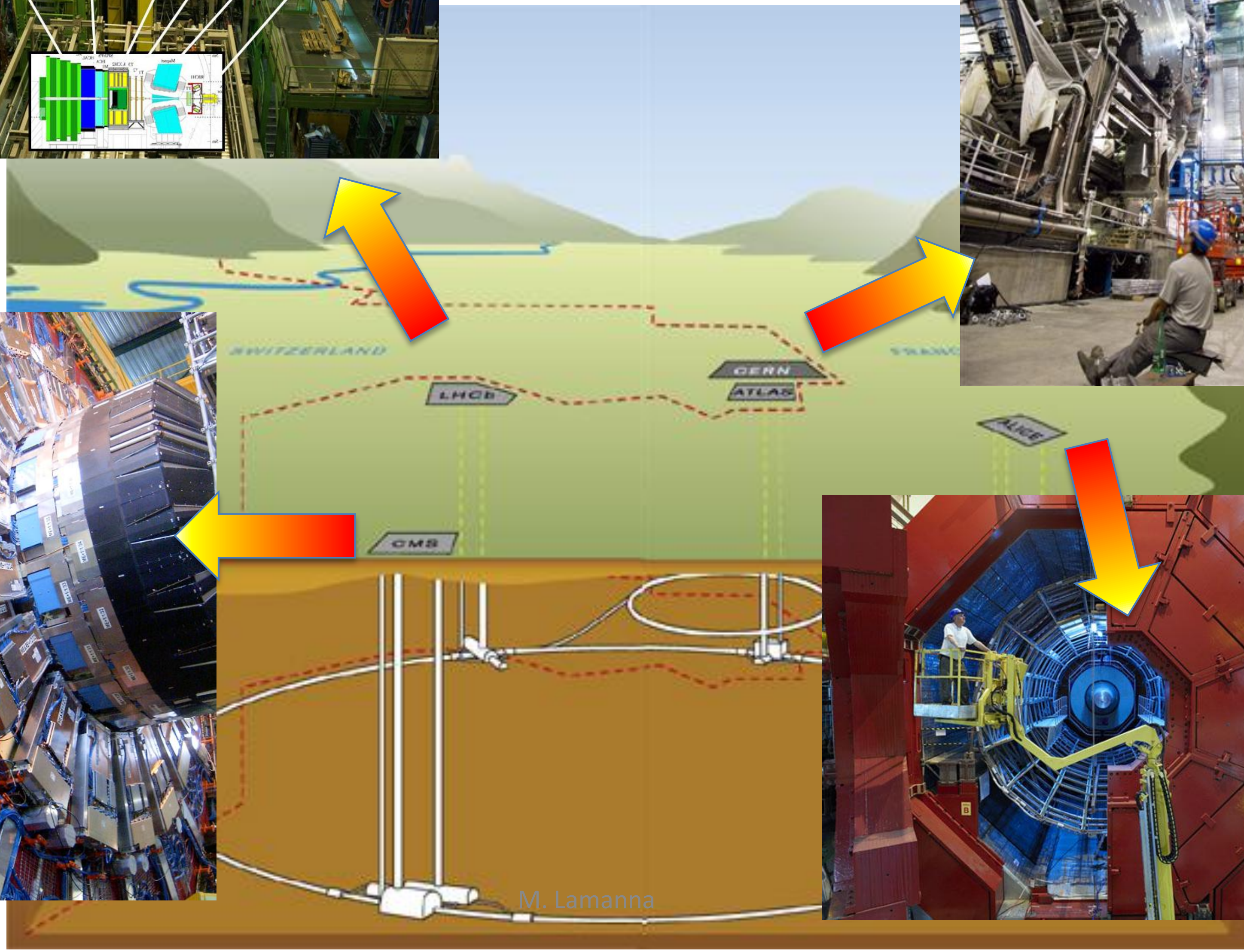
COMPASS Event Display

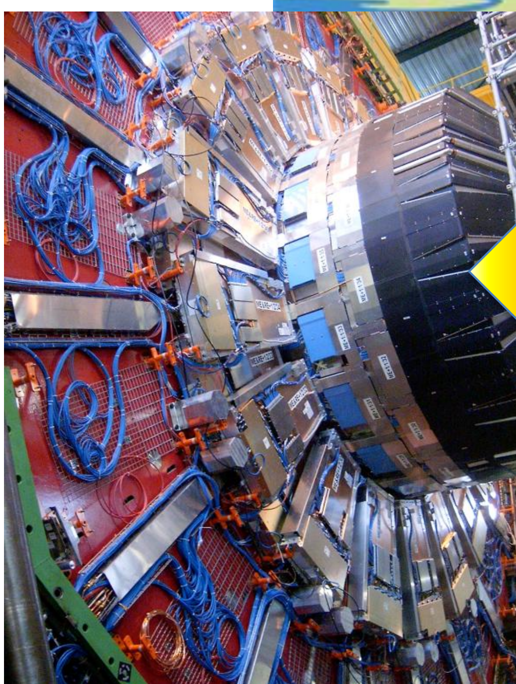


LHC









M. Lamanna

Trigger & Data Acquisition

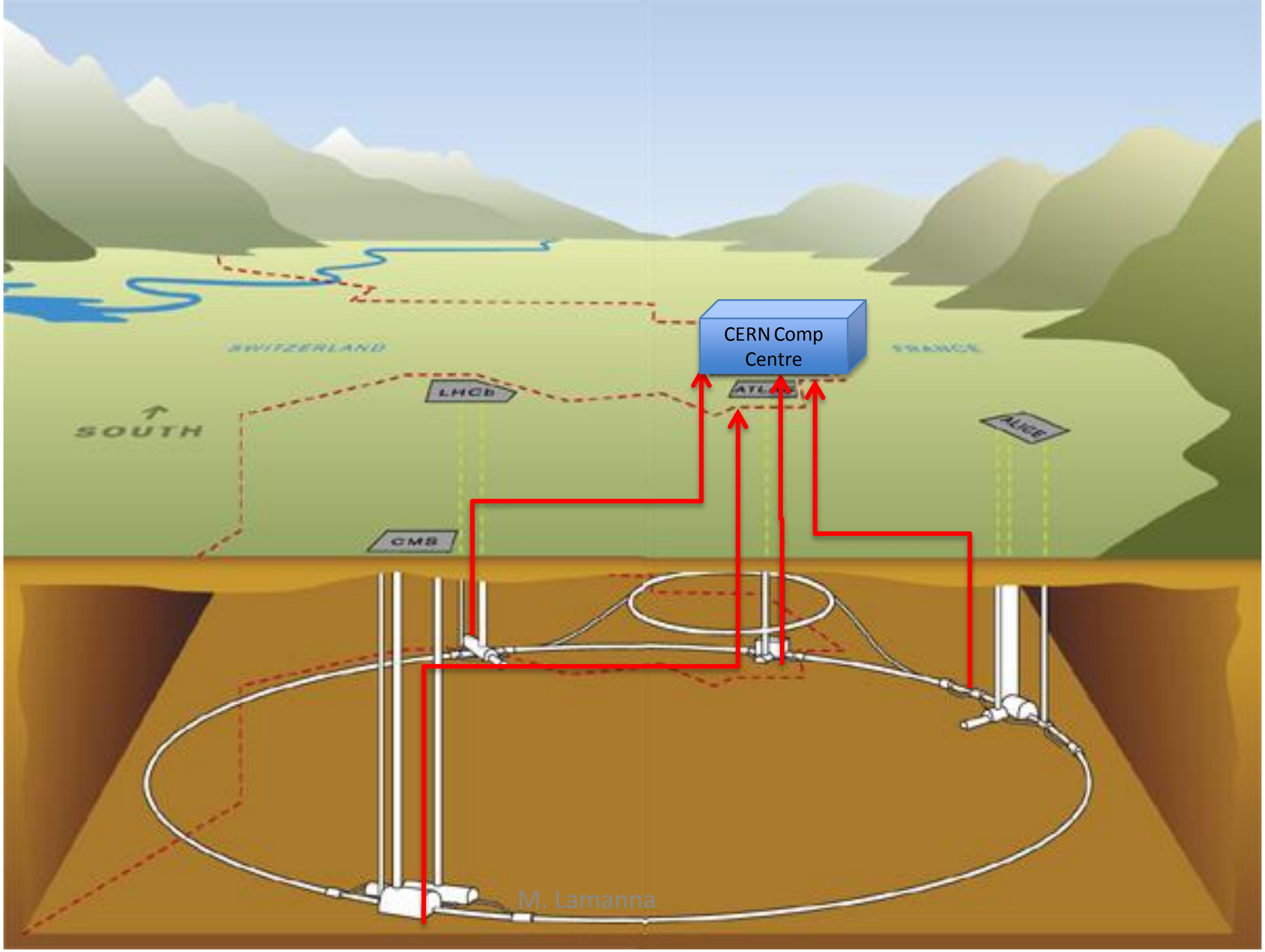
150,000,000 sensors embedded in the detector (ATLAS)

Beam interaction rate: 40 MHz – every 25 ns

20 events overlayed in the detector

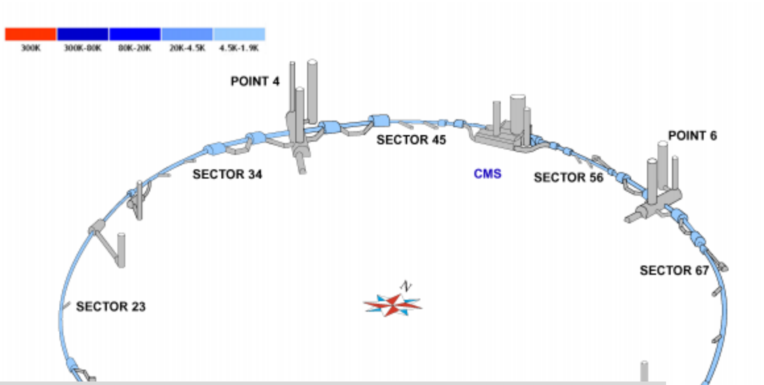
M. Lamanna

Data to the CERN COMPUTER CENTRE

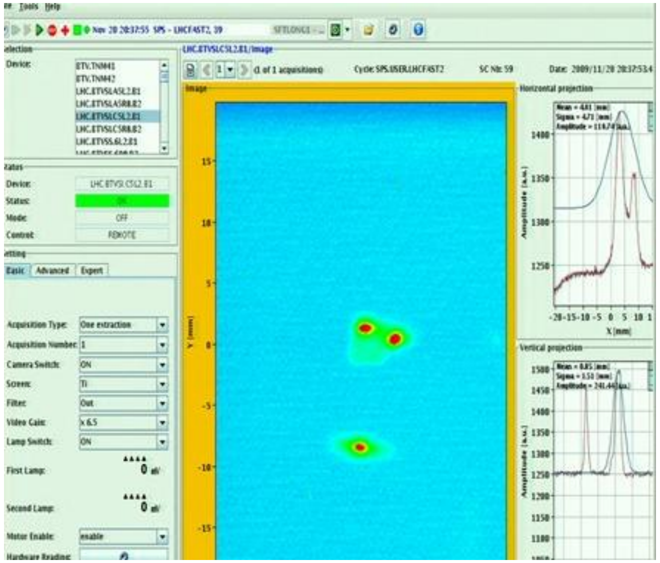


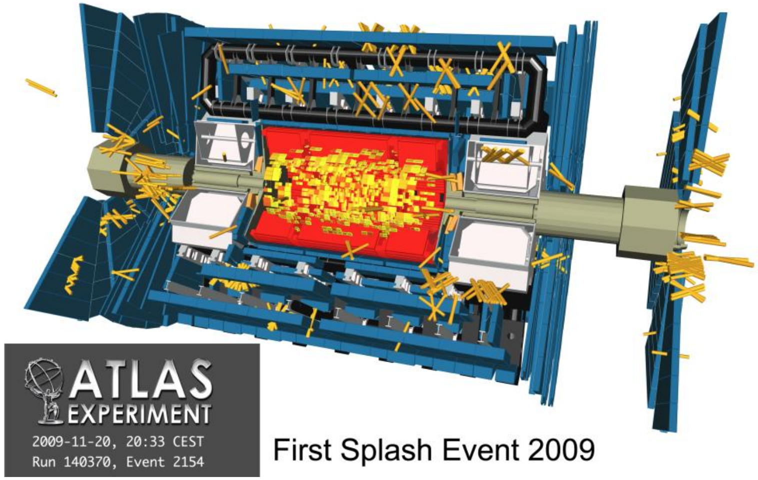
Data, data, Data!

20^{th} of November 2009



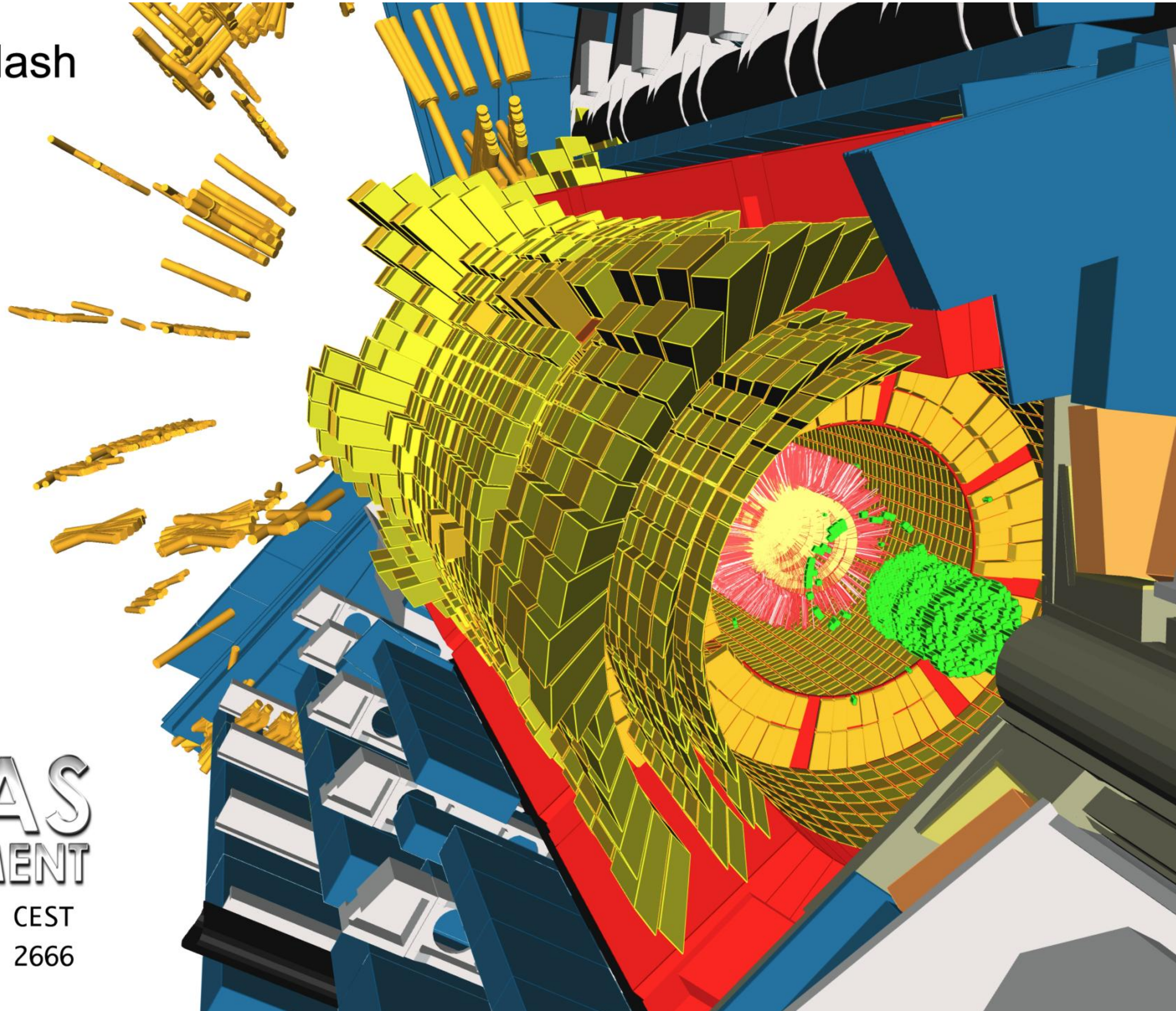
27 km tunnel (2000+ superconductive dipoles)
450 GeV injection energy
7 TeV max beam energy
3.5+3.5 TeV (March the 30^{th} world record)



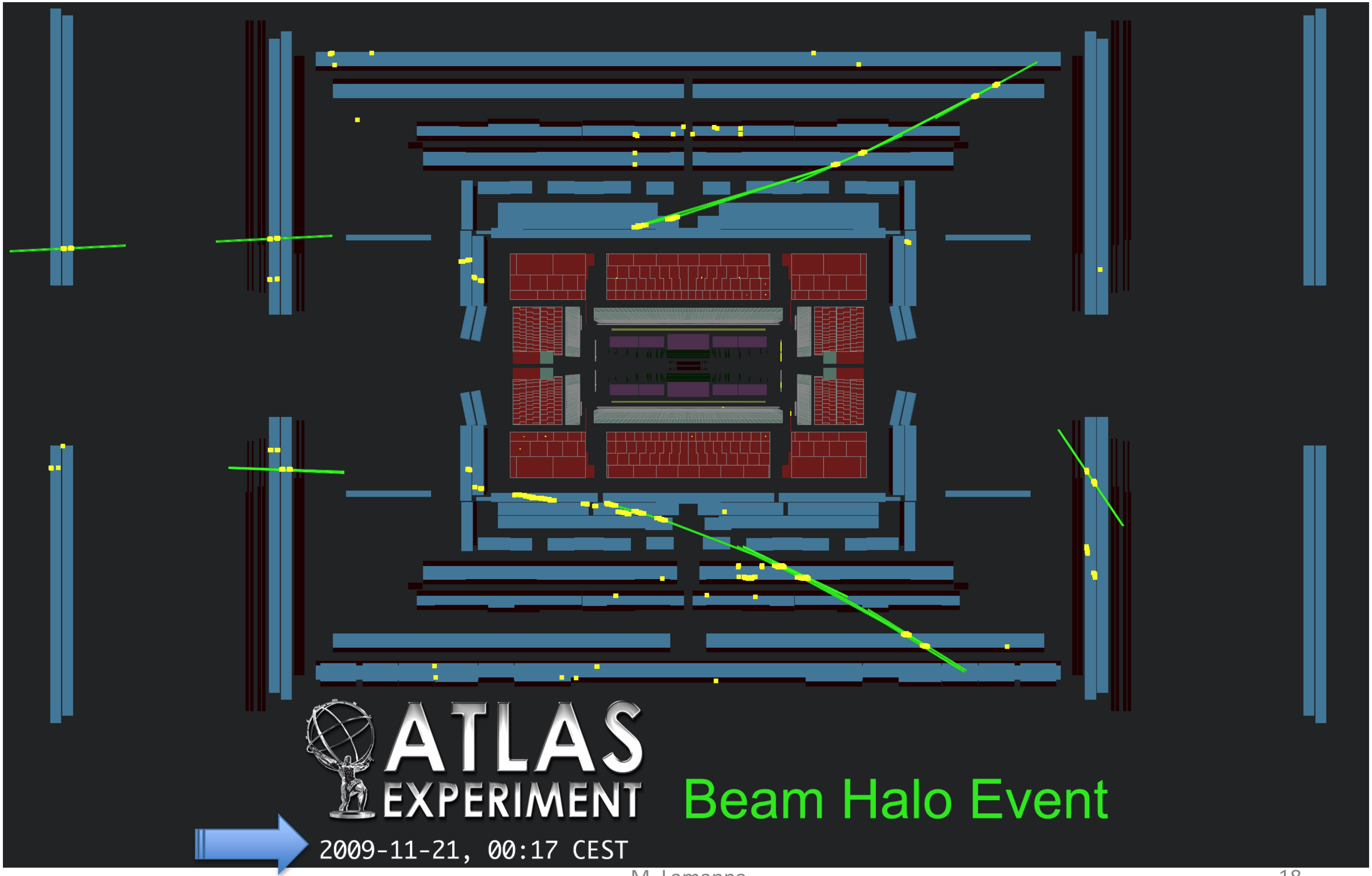


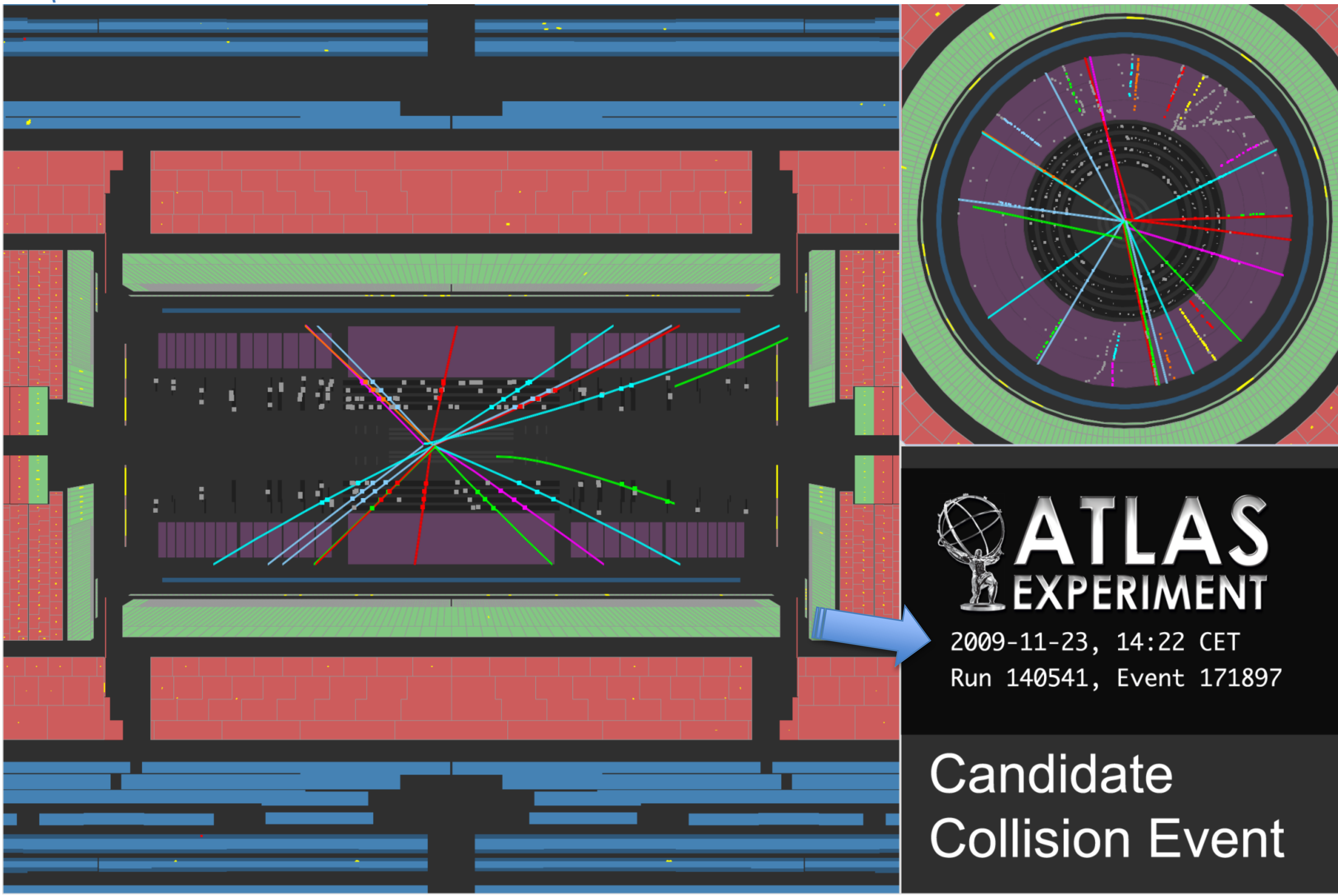
First Splash Event 2009

1st Beam Splash from Beam-2



2009-11-20, 23:32 CEST
Run 140370, Event 2666

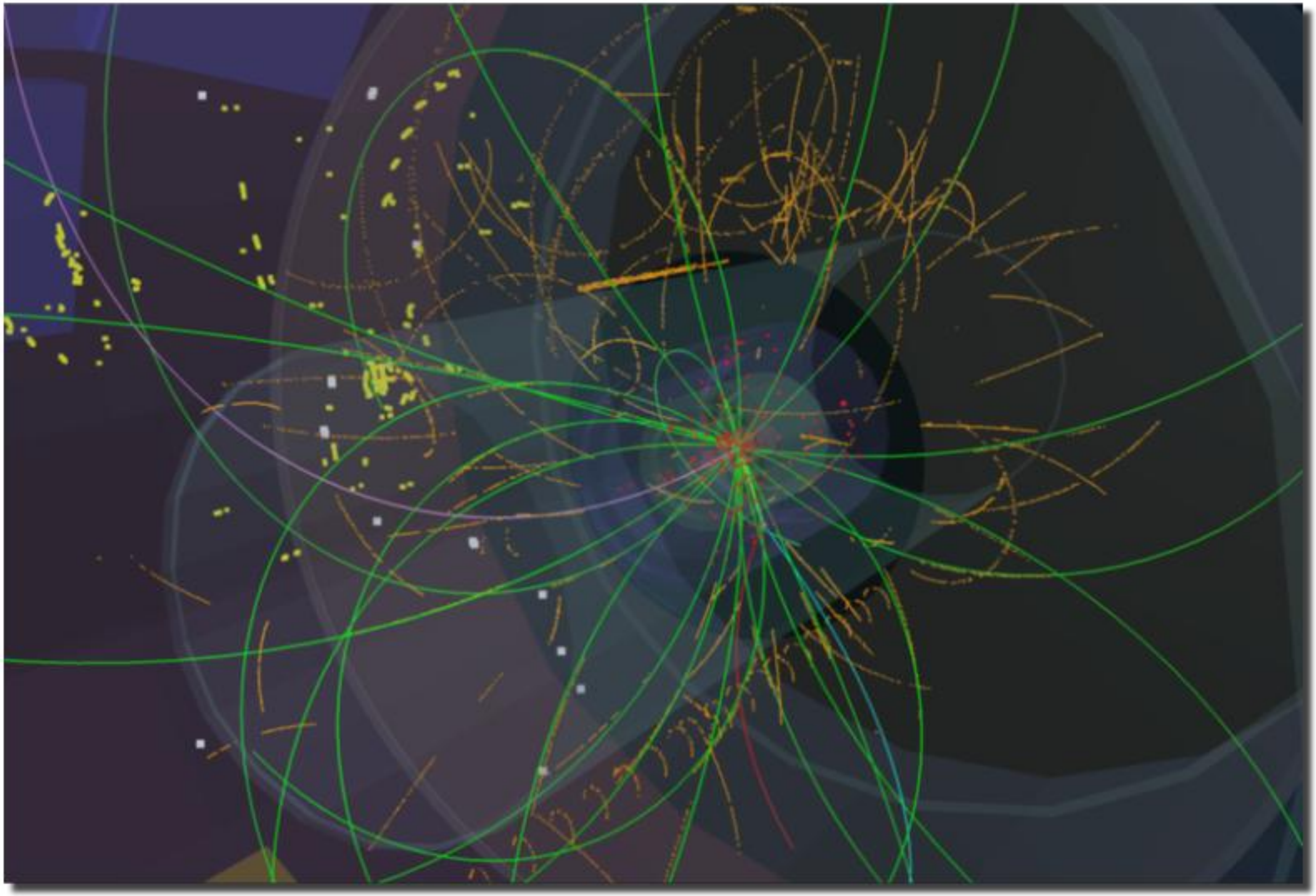

ATLAS
EXPERIMENT
Beam Halo Event
2009-11-21, 00:17 CEST



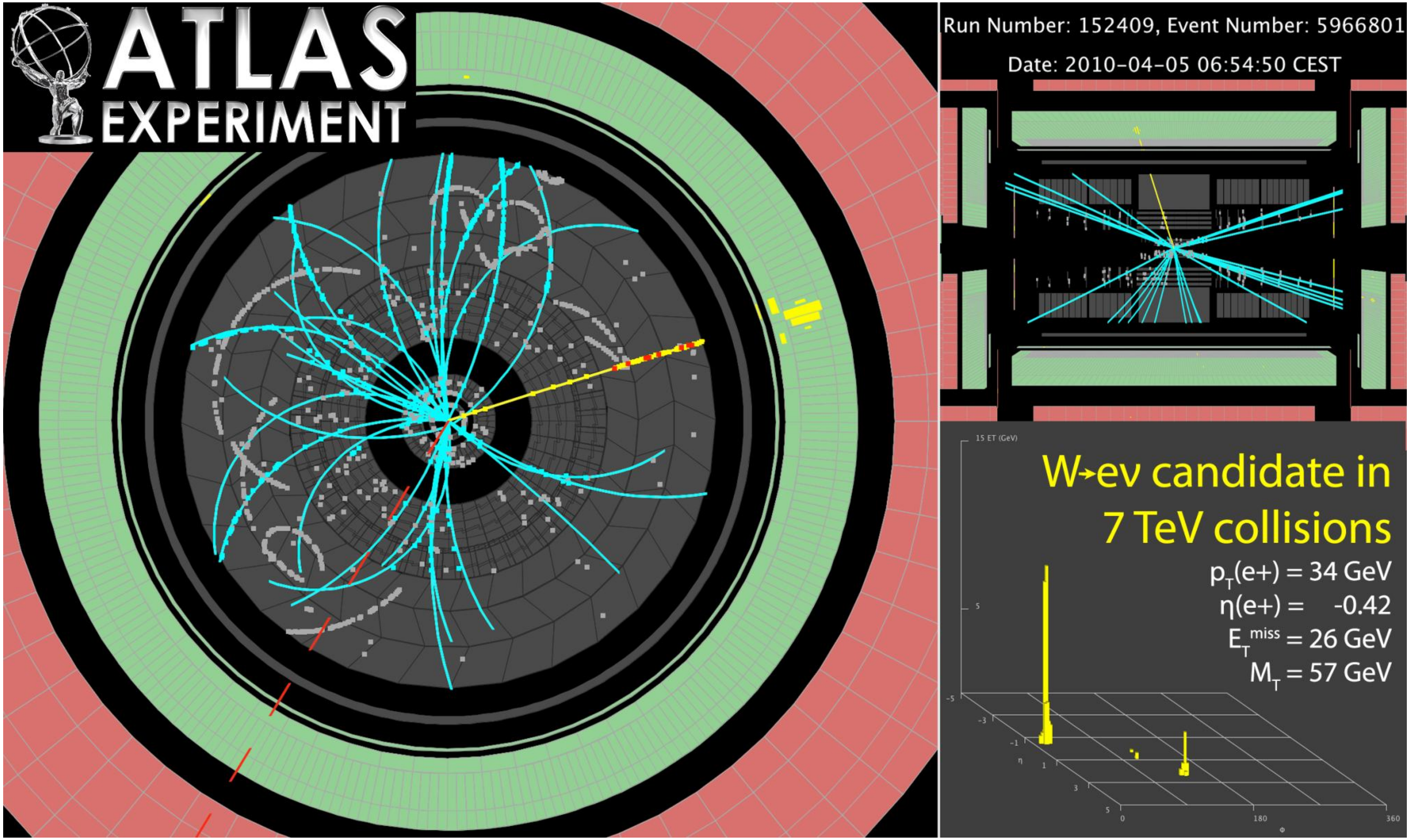
http://atlas.web.cern.ch/Atlas/public/EVTDISPLAY/events.html

3.5 TeV + 3.5 TeV collisions

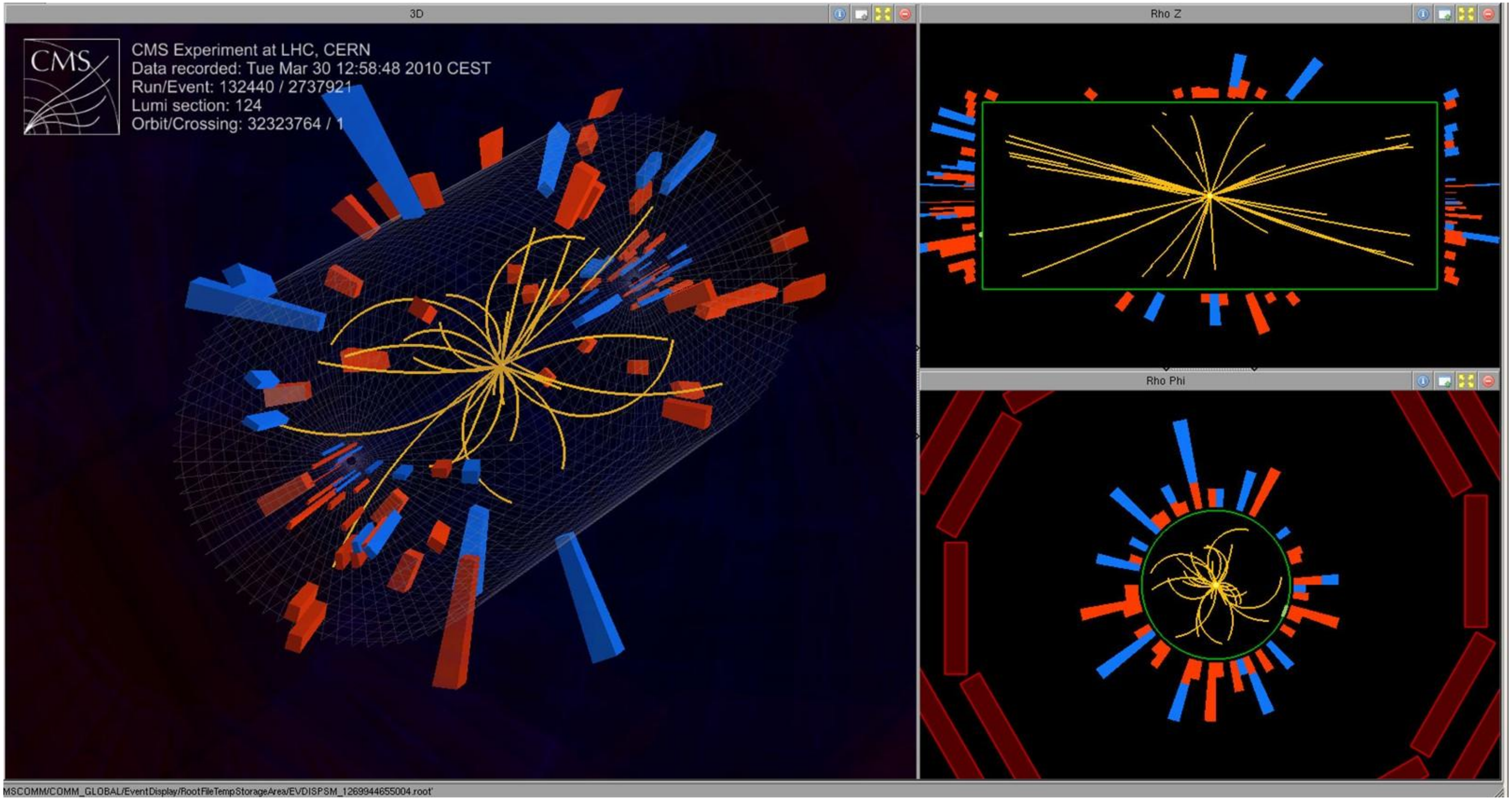
ALICE event



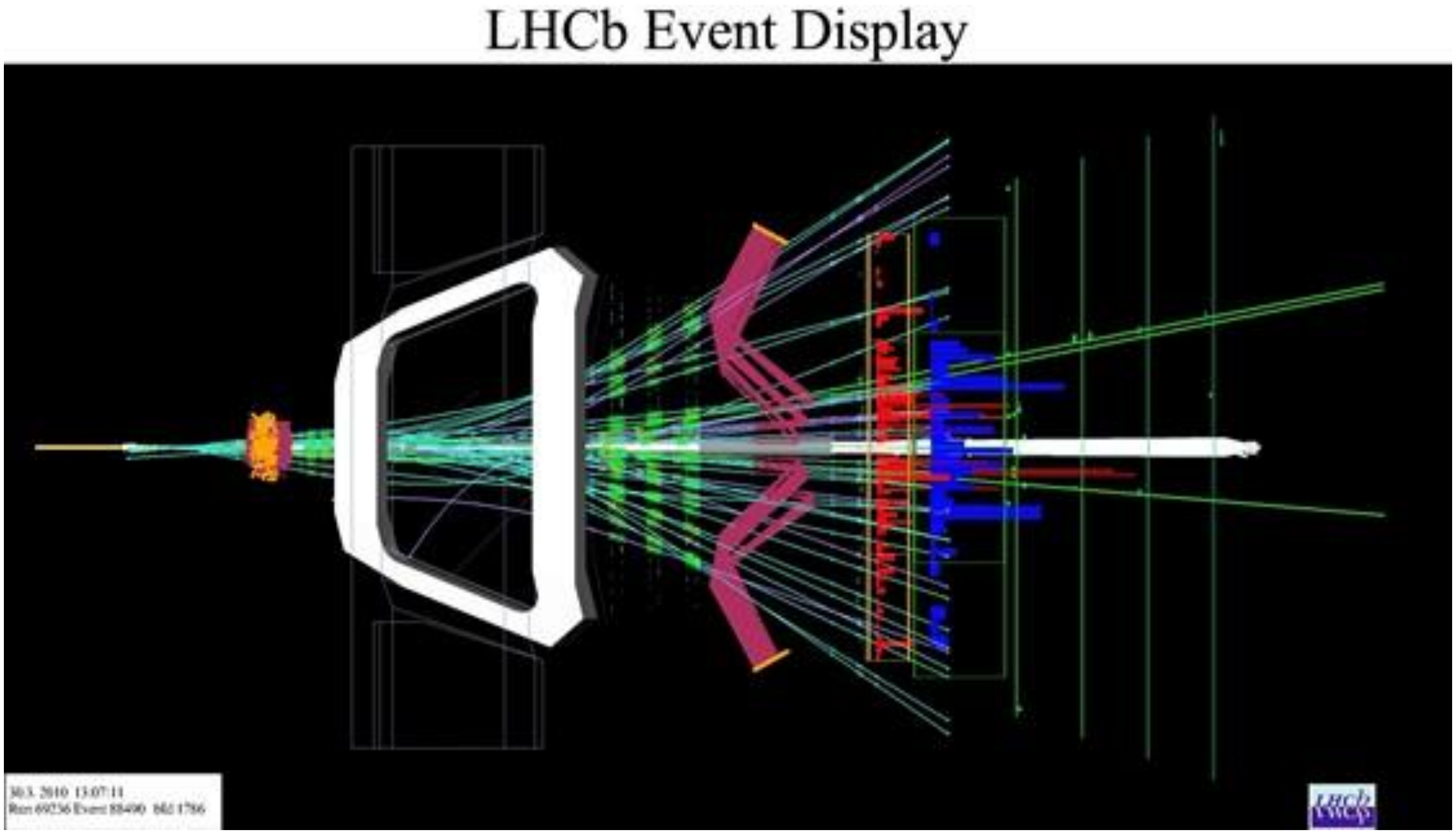
Very recent LHC pp collisions!



Collisions in CMS

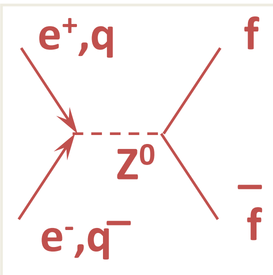


Collisions in LHCb

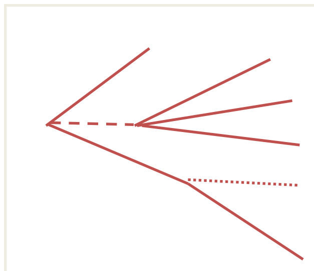


WHY COMPUTING?

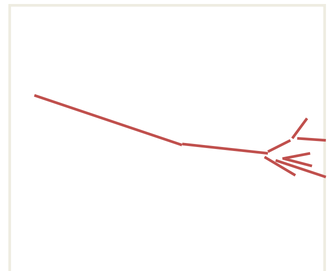
From Physics to Raw Data



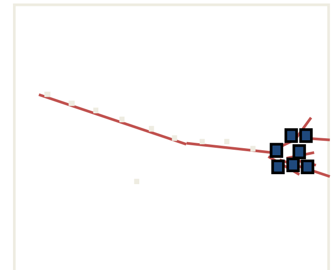
Basic physics



Fragmentation, Decay



Interaction with detector material
Multiple scattering, interactions



Detector response
Noise, pile-up, cross-talk, inefficiency, ambiguity, resolution, response function, alignment, temperature

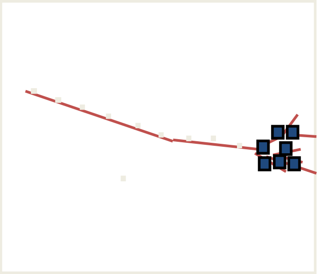
2037 2446 1733 1699
4003 3611 952 1328
2132 1870 2093 3271
4732 1102 2491 3216
2421 1211 2319 2133
3451 1942 1121 3429
3742 1288 2343 7142

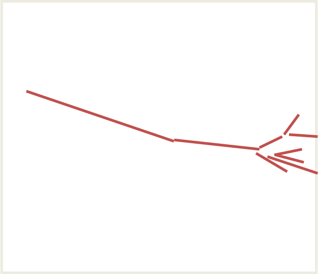
Raw data (Bytes)

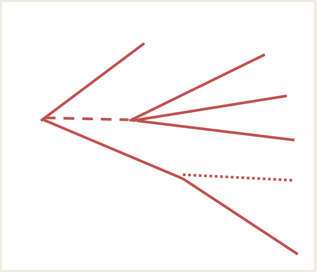
Read-out addresses, ADC, TDC values, Bit patterns

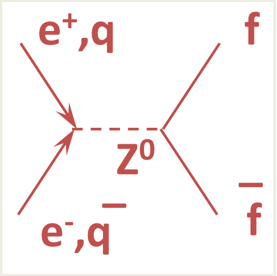
From Raw Data to Physics

2037 2446 1733 1699
4003 3611 952 1328
2132 1870 2093 3271
4732 1102 2491 3216
2421 1211 2319 2133
3451 1942 1121 3429
3742 1288 2343 7142









Raw data

Convert to physics quantities

Detector response
apply calibration, alignment,

Interaction with detector material
Pattern, recognition, Particle identification

Fragmentation, Decay
Physics analysis

Basic physics

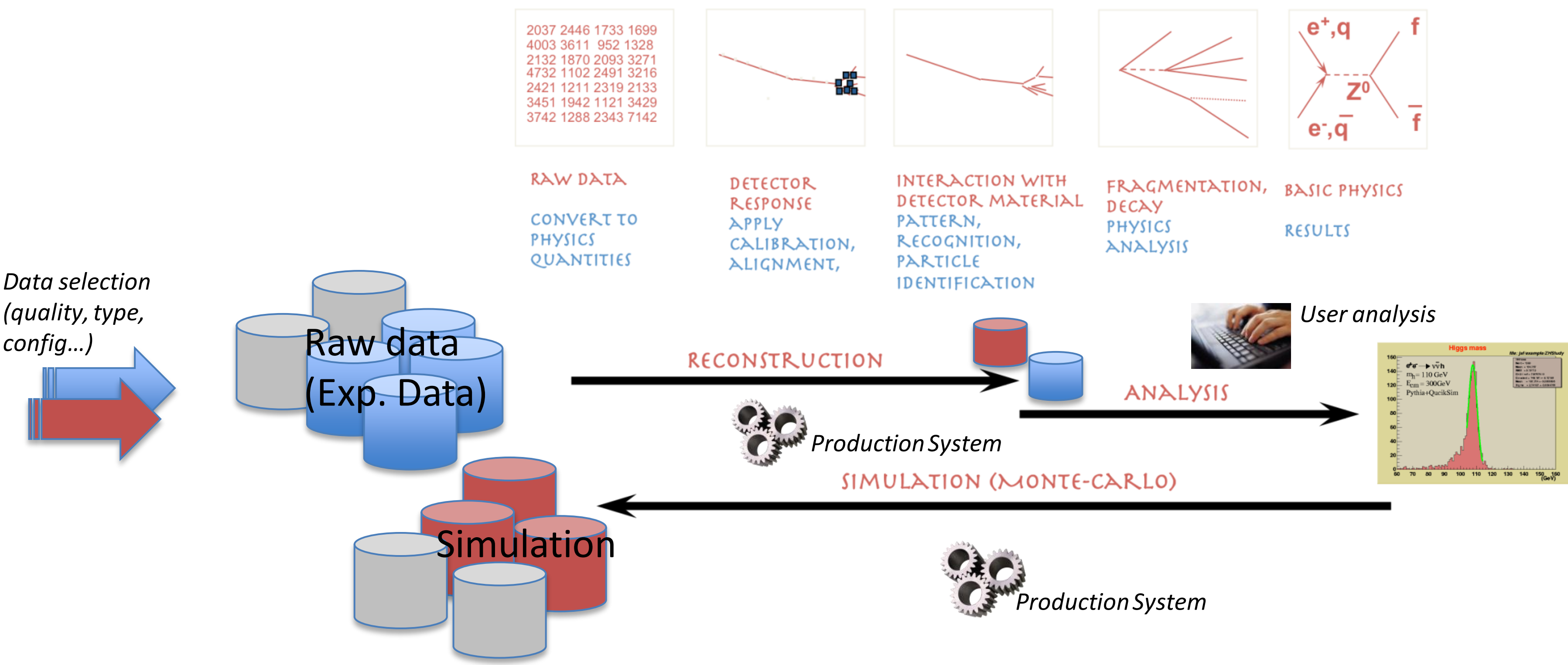
Results

Reconstruction

Analysis

Simulation (Monte-Carlo)

Analysis flow (user view)



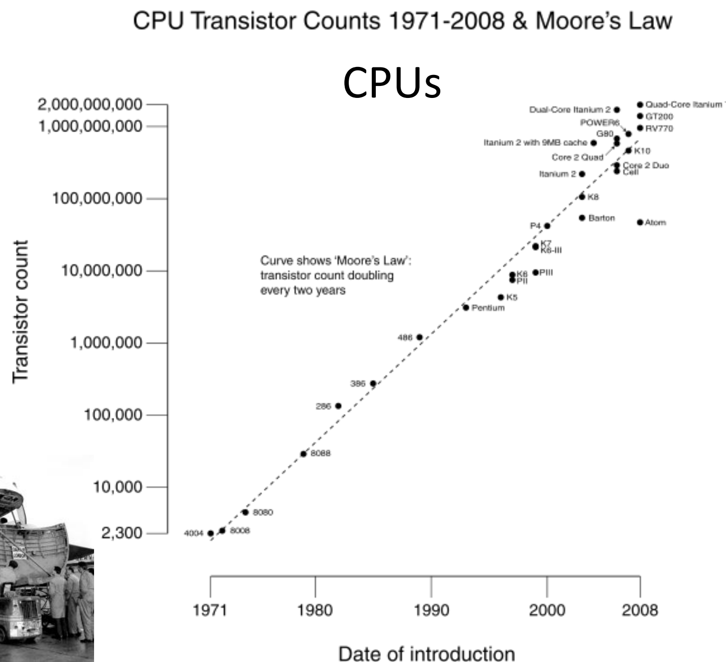
But how this is done *in practice*? Of course we need CPUs, disks, networks etc..
We cannot rush to the solution yet…

Dataset concept = collection of files. Only a small fraction of data in real DBs (e.g. calibrations)

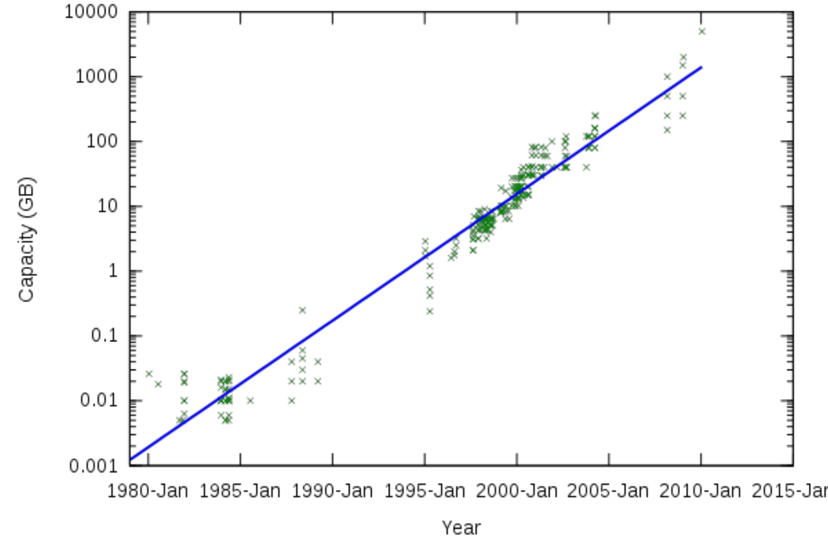
HOW COMPUTING?

Moore law
(mass-market dynamics)

CPUs



Hard Disks

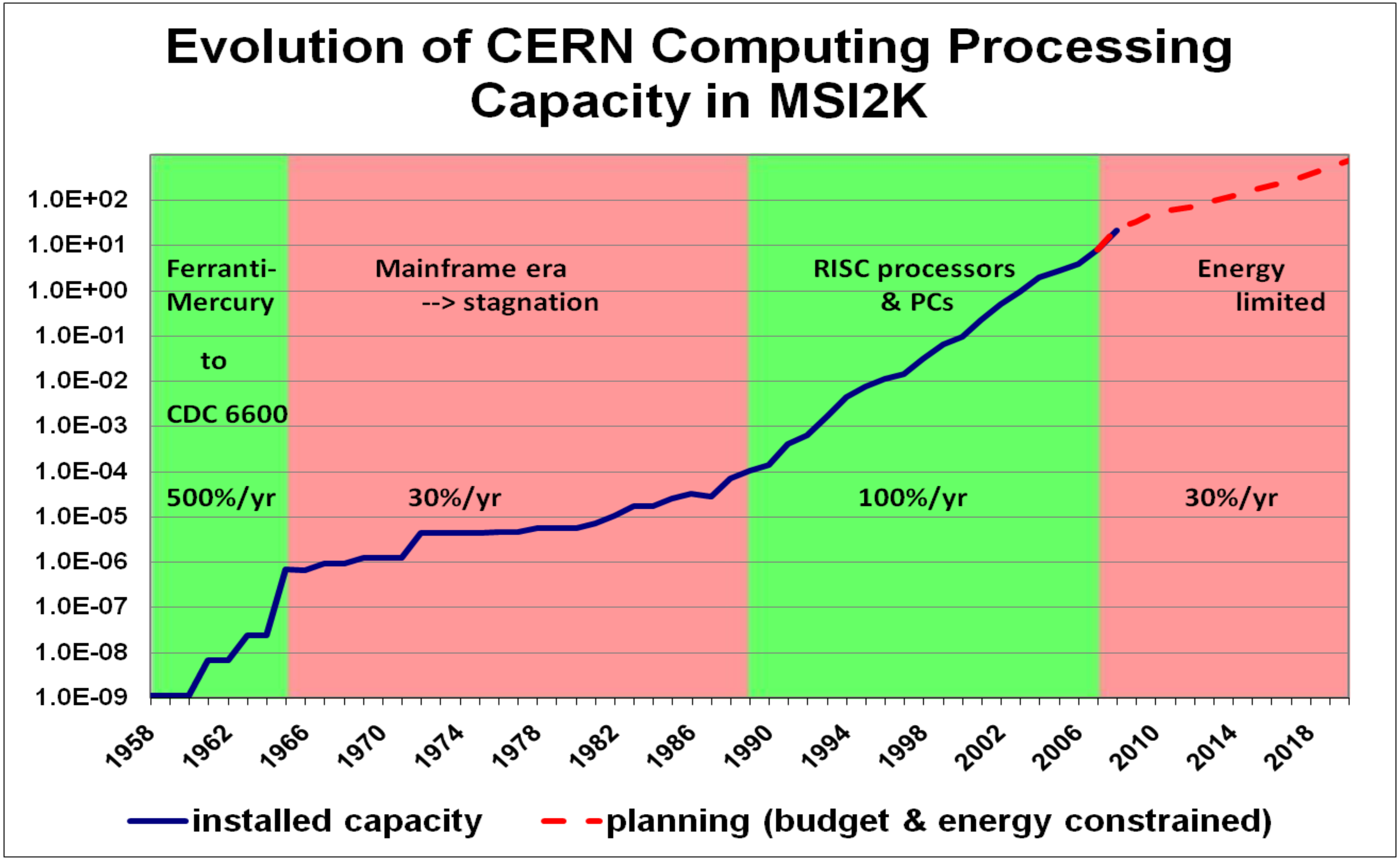


A nice way to express it is that in 2 years your next PC will cost the same but it will be twice as fast (and have twice as much disk space etc…)

http://en.wikipedia.org/wiki/Moore%27s_law

Commodity computing

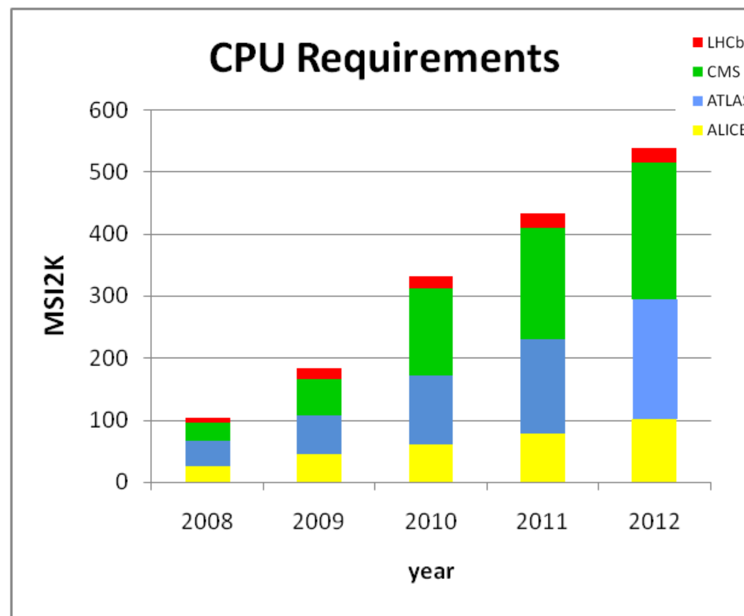
- Commodity hardware challenging the mainframe computing
  - Instead of an expensive supercomputer, buy lots of relatively inexpensive PCs
  - Total cost-of-ownership is an issue
  - On the other hand, your farm will be
    - Upgradable (buy more PCs if needed)
    - Evolutionary (change old out-of-guarantee PCs with new –more performing- ones
    - Cost-effective (buy the “cheapest” SPEC. No vendor lock!)
- The only way to cope with the CPU (and data storage!) request from new generation experiments
  - Late ’90: NA48 and COMPASS at CERN, several other experiments at FNAL, SLAC, DESY…
  - Nowadays: LHC experiments
- Limiting factors:
  - PC market evolution
    - Desktop vs laptops or other commodity devices
    - Multicore architectures
  - Power (includes cooling…) consumption



1.0E-03 MSI2K = 1,000 kSPEC INT 2000 ~ Pentium4 @ 3GHz (~3GFlops)

CPU

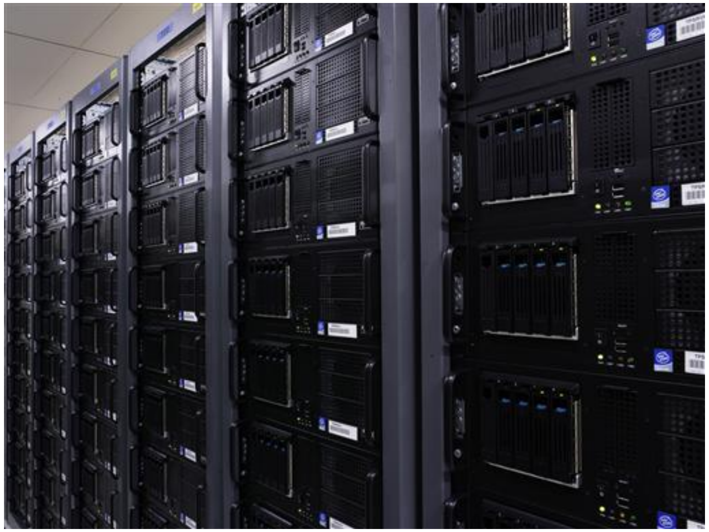
- "Number crunching" boxes
  - No resident scientific data
  - Shared facility for all CERN users (basically every physicists participating in an experiment at CERN)
- Faster PC means more SPECs per box
  - 1,000 SI2K ~ Pentium4 @ 3GHz (~3GFlops)
- Mass-market driven
  - MP3 encoding, Digital images processing, heavy office suites...

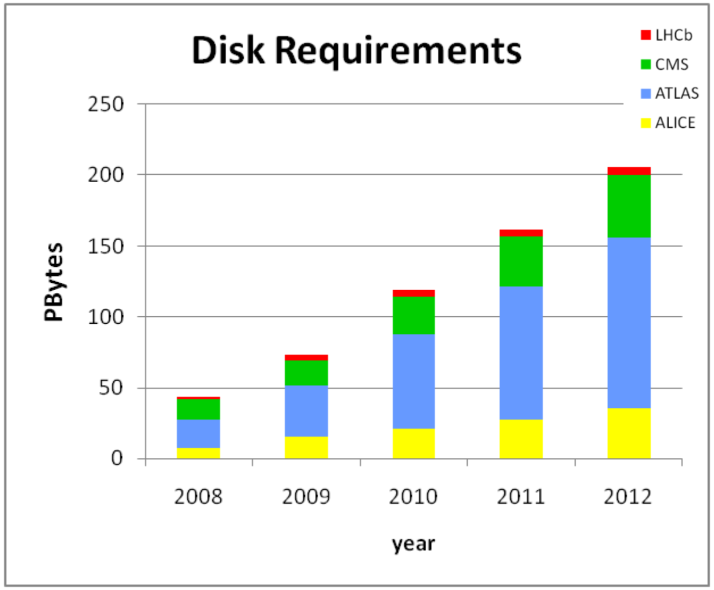


M. Lamanna

DISKS

- "Staging" area: they keep "hot" data
- Access to these disks is managed by dedicated PCs (serving/receiving data to the PC crunching numbers)
- Moore's law at work here!
  - Less expensive
    - GB/$ goes up
  - More compact
    - GB/cm3 goes up
  - Mass-market items!
    - Technology drive by you(storing digital-camera pics, MP3, etc...)





Tapes

- Data custodial
  - We build accelerator and experiments <u>to collect scientific data</u>
- Write-once / Read-many
  - At variance with backups (Write-once / Read-never)
- Evolving (a' la Moore's law) but more "gently" than PCs
  - No surprise, none of us has a tape library at home, I guess ☺
  - No mass-market as for CPU, RAMs and Disks
- Expect to store 40 PB (40,000 TB) of data per year
  - Scientific data, corresponding derived data (reconstruction, analysis), simulation data

Interesting facts (CERN Computer centre)

- Number of machines
  - About 4,500 batch (18,000 CPUs)
  - About 3,000 disk servers (50,000 hard drives)
  - Several hundred tape servers, console head-nodes, database and Grid servers etc.
- Storage Capacity
  - 5+ PB disk
  - 25+ PB tape (IBM and Sun/StorageTek)
  - There will be an additional 15 PB each year needed for the LHC data (3*10^6 DVDs!)
- Network Capacity
  - Connection at 10 gigabits/second to each Tier 1, plus backup, plus regular (firewalled etc) internet
  - Speed record: 1.1 TB in less than half hour (CERN-CalTech)
- Number of staff
  - CERN: ~2700 ; IT department (computing) ~250 and ~200 on shorter-term Grid projects

Early 2009 data

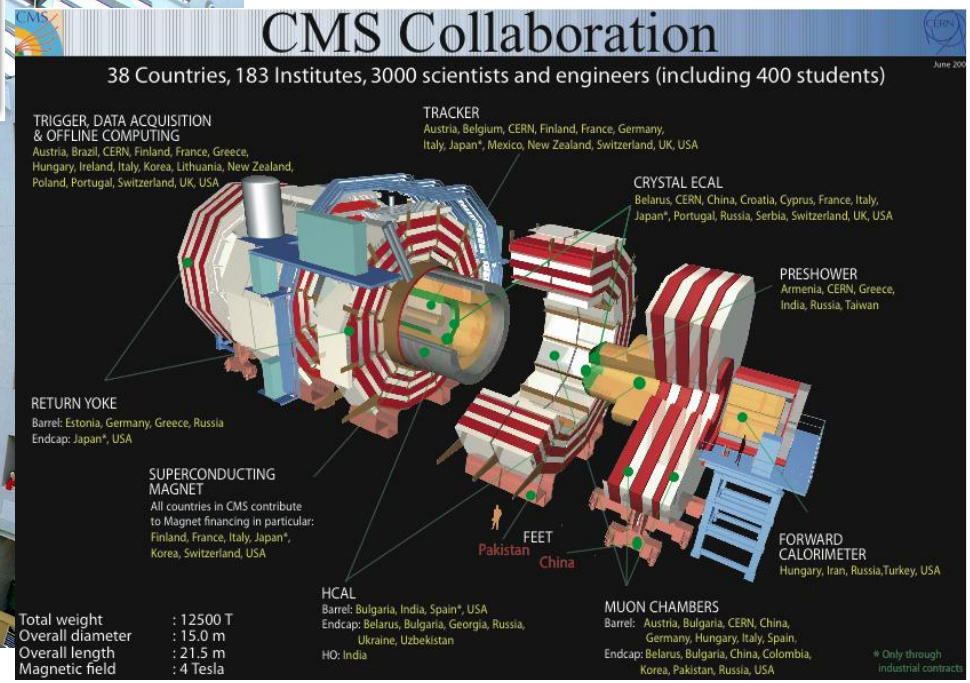
Fast ramp up going on for the LHC start

Service/Activity	Description	Data
Disk and Tape	Provide storage capabilities	1500 disk servers 5 PB disk space, 16 PB tape storage
Network Campus	Provide local CERN area network service	The Network core has a capacity of 4.8 Tbps The total number of network ports is: >1000 10 Gigabit ports >70000 Gigabit ports feeding 34000 hosts
External Networking	Provide connectivity and infrastructure with other 11 T1s	Out-coming Internet Back bone up to 60GB/s to 11 centers Speed record:1.1 TB transferred in less than half hour (Caltech-CERN)
NICE PC	PC cluster	5,500 active NICE PCs 1,500 Macs 1,100 CMF packages
Computing Facility	Provide local & Grid Computing Power	4500 nodes installed 16K CPUs available Up to 10K concurrent jobs
Computing Center cooling system	The system to cool the centre	more than 500,000 m3/h cold air to cool PCs three 1.5 MW chillers
CVS	Concurrent Version System	over 300 Software Projects and over 3000 users
Messaging services	Email, Ldap, Listbox, News, Fax, Antispam	17,015 mailboxes 7000 lists 2.8TB data 2.5 Million messages/day (~98% SPAM)
Web Services	Web Services , Search Engine, Verisign	- 8852 sites - including 930 sharepoint sites - 6 million hits per day - Bandwith 2 Tbyte per day
WLCG	Coordinating and operating WLCG grid activities	30,000 simultaneously 1 PB /month each ATLAS, CMS 1 GB/s (1.6 GB/s peak) to 12 sites 160 VOs
AFS	Andrew File System (distributed file system for CERN users)	15,000 users 27,000 volumes 25TB allocated

WHERE COMPUTING?

Another view of LHC experiments

ATLAS is more than 2,500 collaborators
From 169 universities from 37 countries
700 students!!!
CMS has a similar size. ALICE and LHCb collaborations are a bit smaller

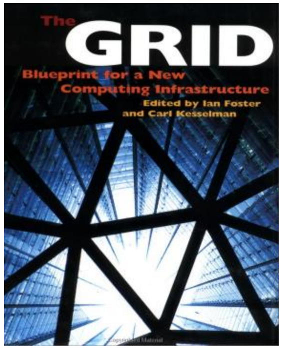


But do not be surprised if…

Computing goes worldwide…
…after all, we all got networks!
… and we invented the Web!!
… and now we play with Grids!!!

The Promise of Grid Technology

Ian Foster & Carl Kesselman - 1999



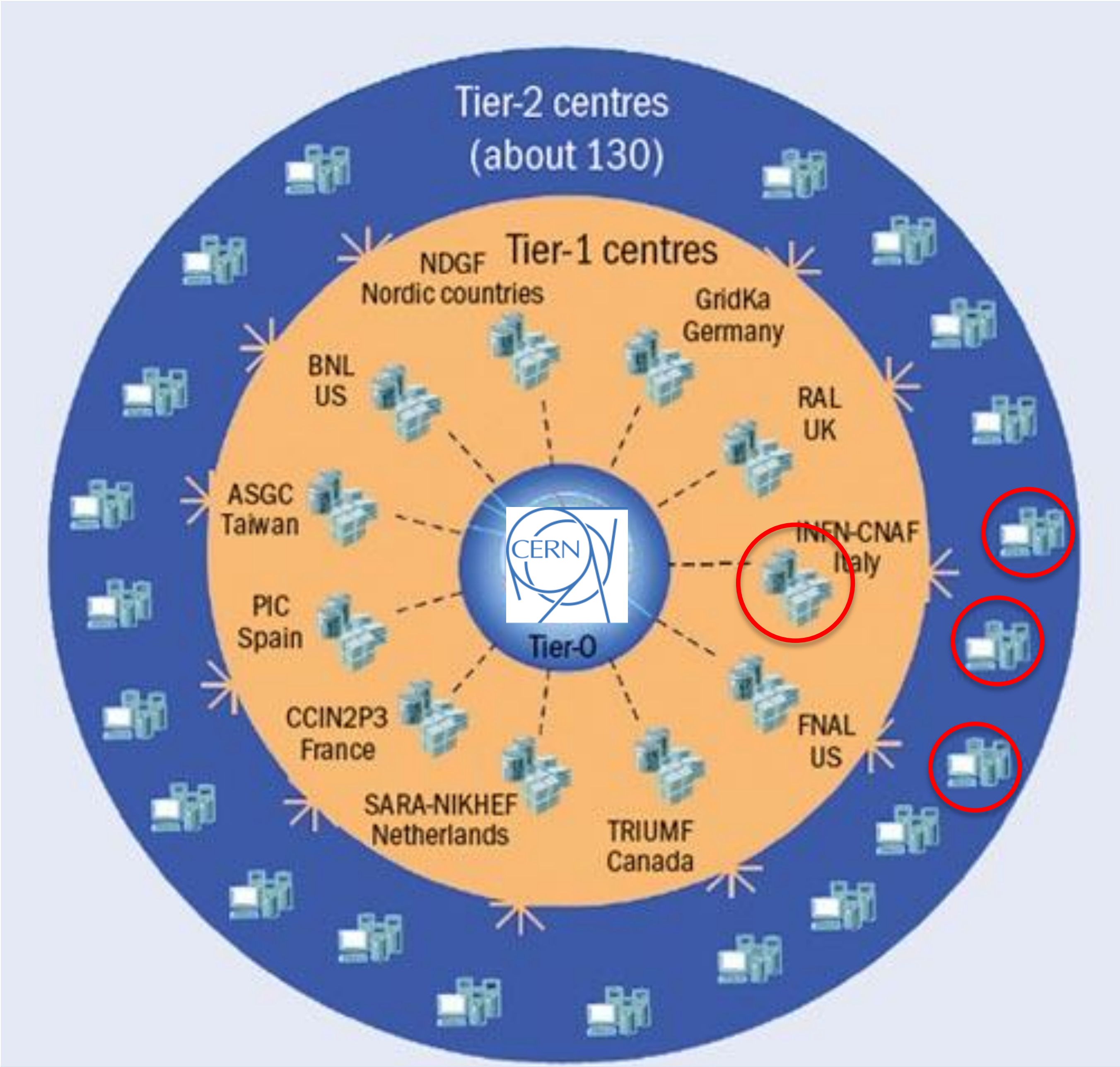
- The **Grid** – a virtual computing service **uniting the world wide computing resources** of particle physics
- The **Grid** provides the end-user with **seamless access** to computing power, data storage, specialised services
- The **Grid** provides the **computer service operation** with the tools to manage the resources, move the data around, monitor the behaviour of the services, alert operators to attend to potential problems

Sounds good!
Although very complicated ☺

Courtesy Les Robertson

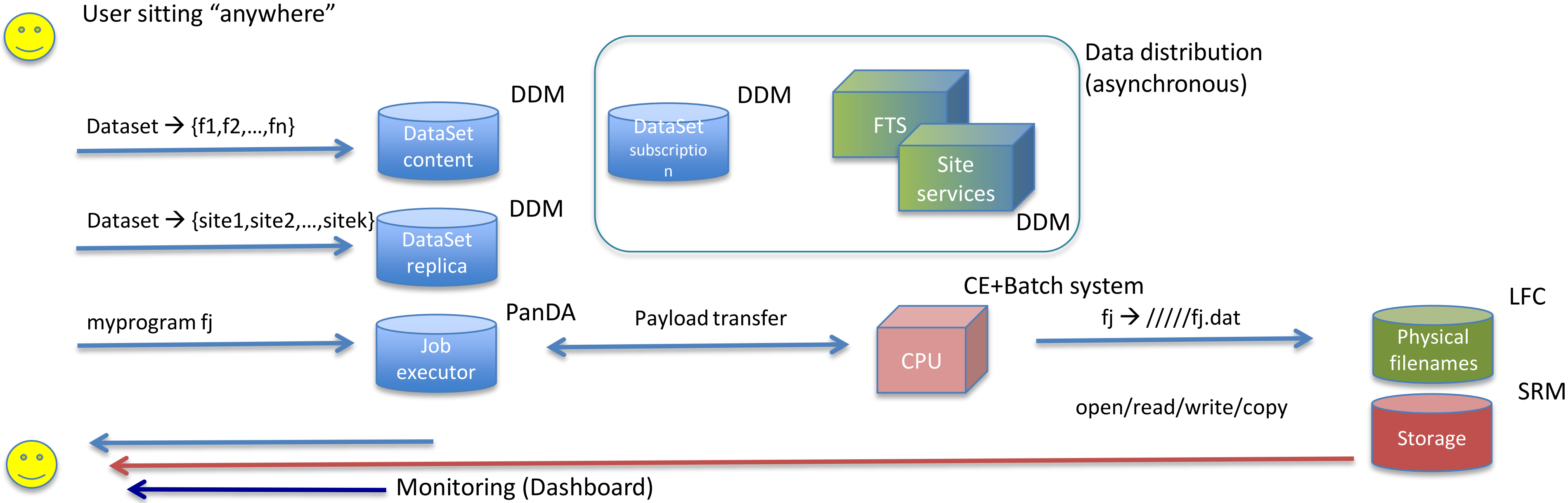
How does it work?

- ATLAS
  - Not substantially different for the other 3 LHC experiments
  - Heavily simplified…
- What do we want to achieve
  - The user wants to specify a subset of the data and run applications on it (chain of programs reading intermediate outputs)
  - Only at the end of the chain data sizes and computational complexity this can be (possibly) done on a laptop
    - 1000+ of physicists worldwide after the <u>same</u> data



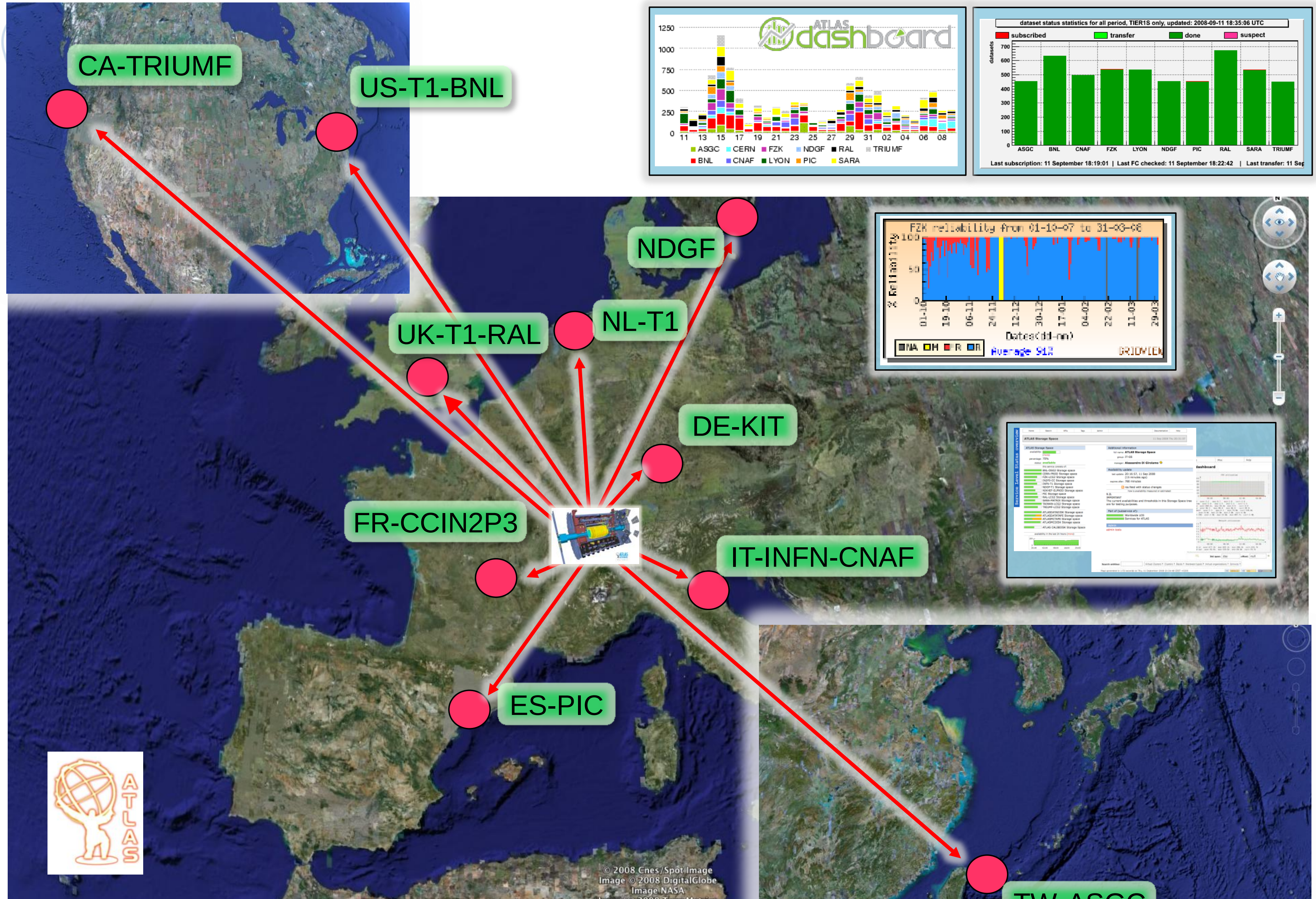
Cloud = Regional Structure

Behind the scenes…



Simplified

One job executed in one cloud



10 clouds
Only the Tier1 centre are shown
About 100 Tier2s in total

M. Lamanna

Powered by *middleware* !

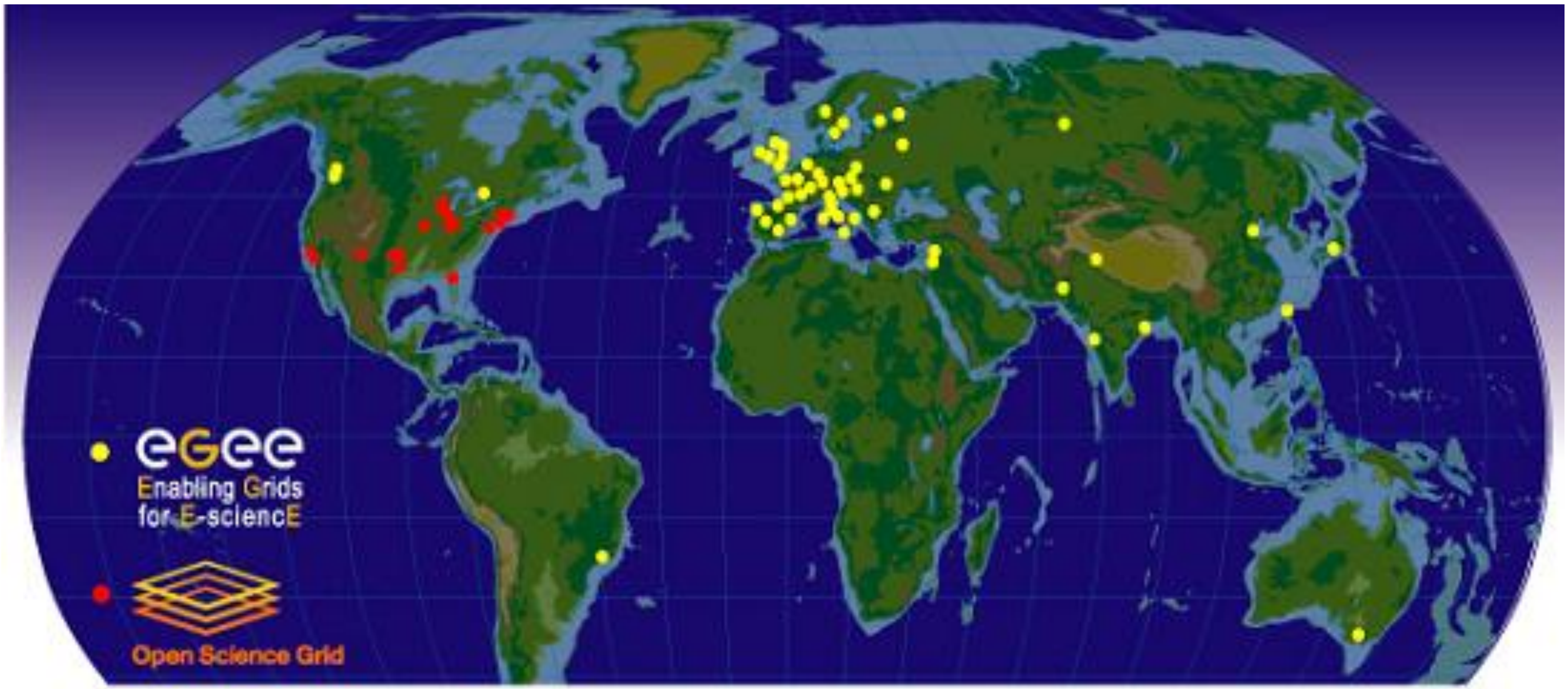
- Globus, Condor and EDG as starting point
  - Evolution lead by Europe (EDG → EGEE)
  - Other projects competing/collaborating
    - ARC, Unicore etc…
  - Large contribution from LHC community
    - And to a lesser extend from other users communities
    - “High-level” services
- Security
  - X509 infrastructure
  - Certificates → Proxies

Examples

- Computing Element (CE)
  - Gateway to CPU resources (typically a batch farm)
  - Examples: LCG-CE (LCG), CREAM (EGEE/INFN)
- Data transfer
  - Layer on top of gridFTP
  - Example: FTS (LCG/CERN)
- Workload Management System
  - CPU allocation for workload execution
  - Examples: gLite WMS (EGEE/INFN), glideinWMS (CMS), DIRAC (LHCb), PanDA (ATLAS), ALIen (ALICE), etc…
- Storage Element (SE)
  - Layer on top of storage solutions (Distr. FileSystem, Disk-Tape HSM-SAN, etc..)
  - SRM (Storage Resource Manager)
- Other services: information systems, catalogues, etc…

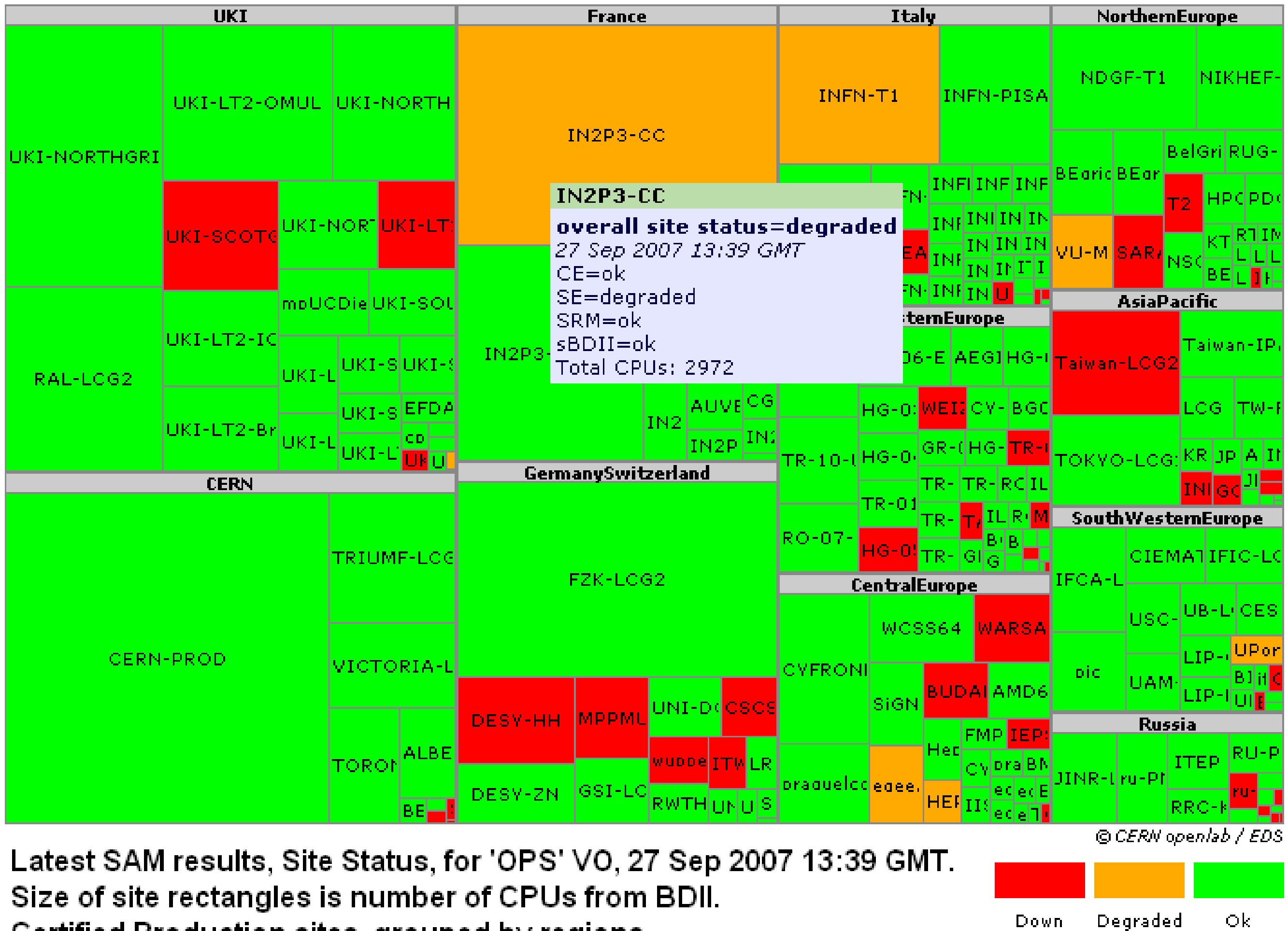
LHC Computing GRID

Worldwide infrastructure
(EGEE + OSG + NDGF)



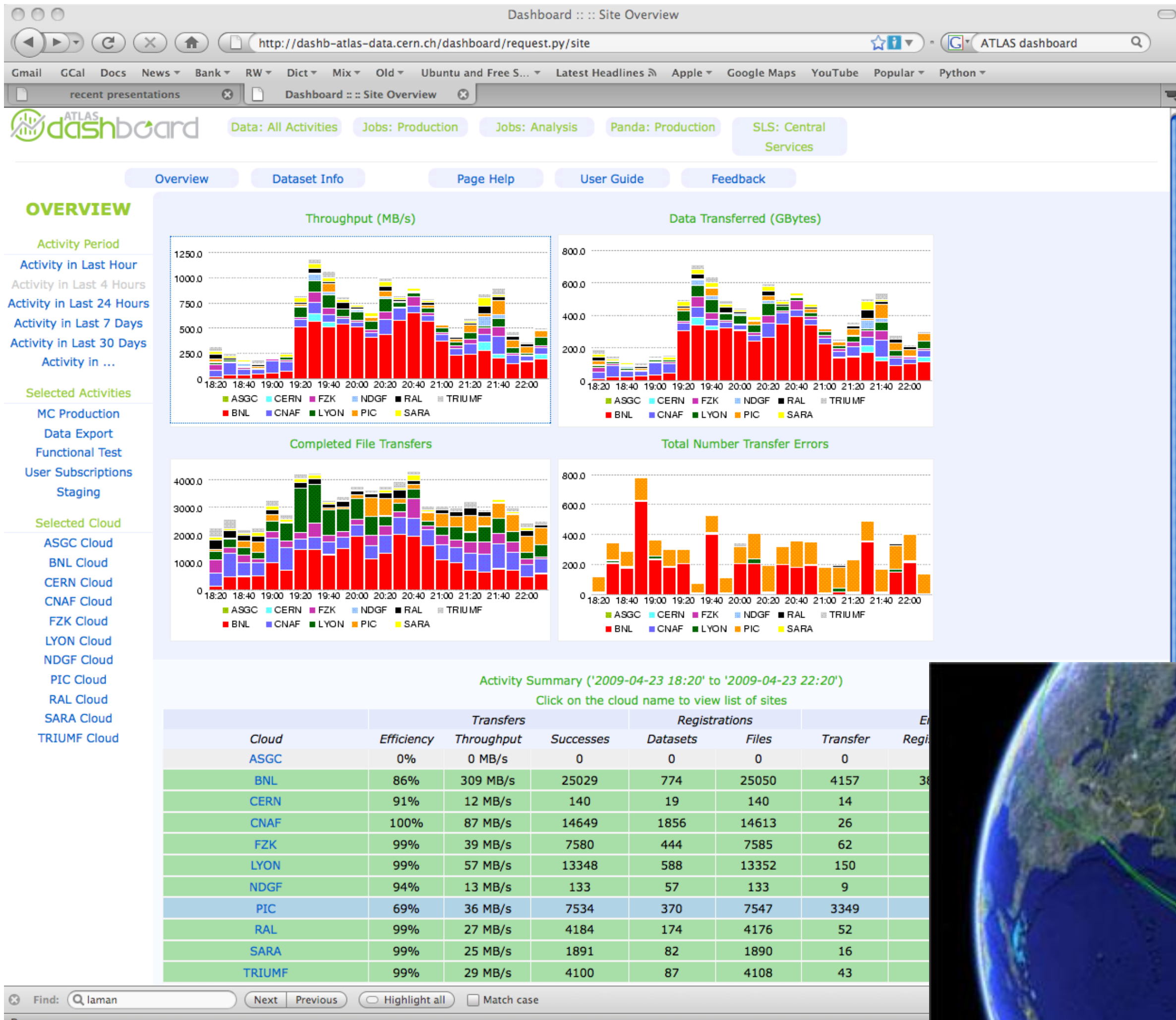
All together now…

- CPUs, disk and tapes from more than 200 sites
- Sharing hardware
  - Share CPUs
  - Replicate data across different sites (performance and data preservation)
- Collaborative effort
  - Complex operations
  - Share responsibilities
  - Solve problems together

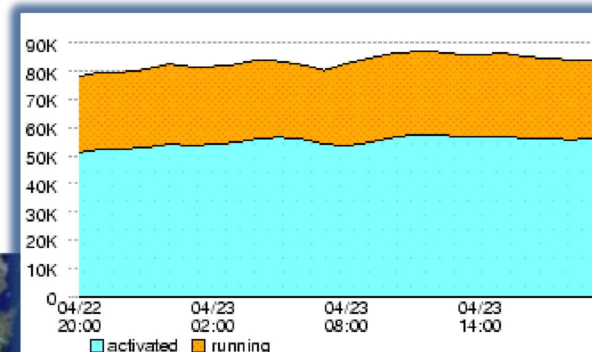


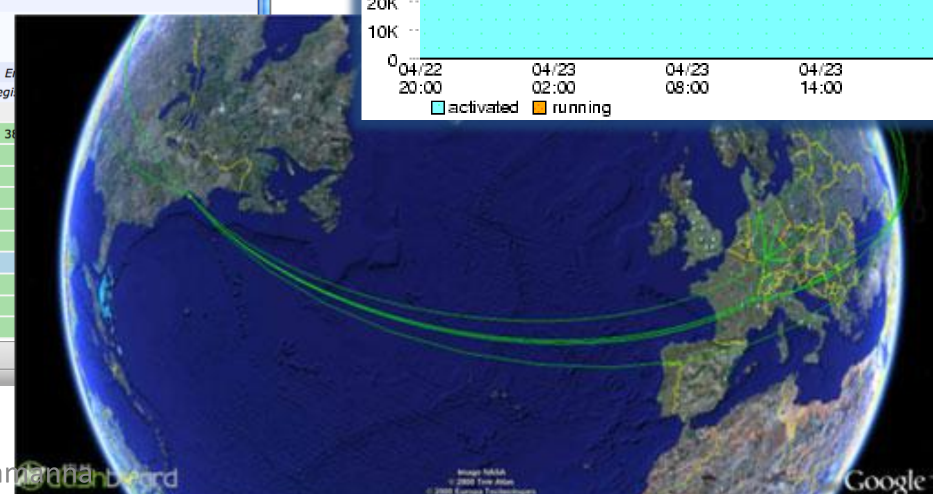
Latest SAM results, Site Status, for 'OPS' VO, 27 Sep 2007 13:39 GMT.
Size of site rectangles is number of CPUs from BDII.
Certified Production sites, grouped by regions.

ATLAS ON THE GRID

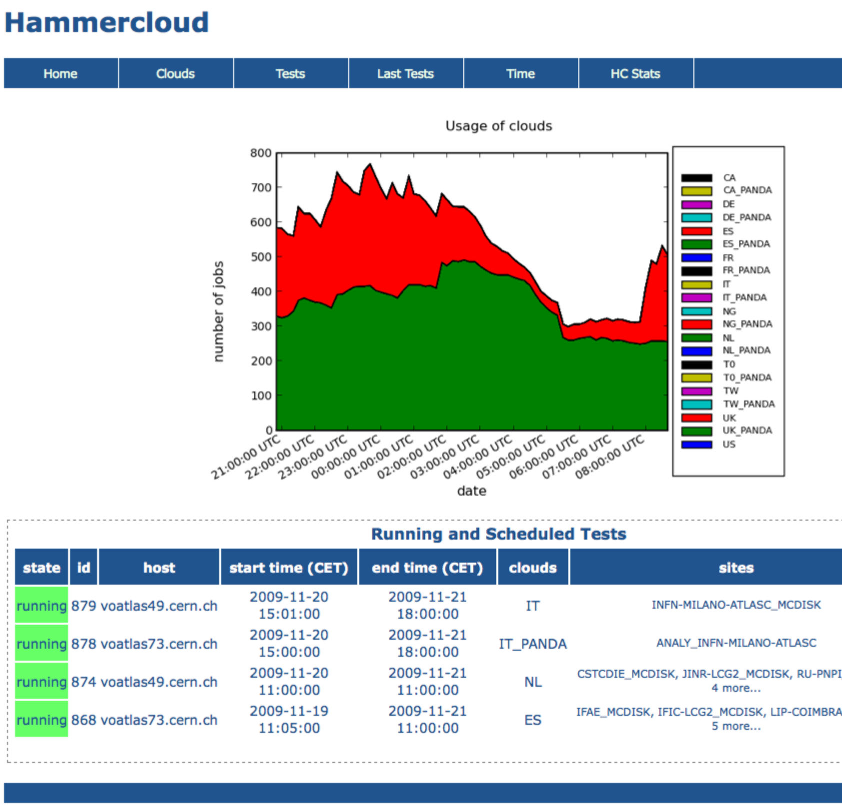


- Data distribution (left)
- Job distribution (bottom)





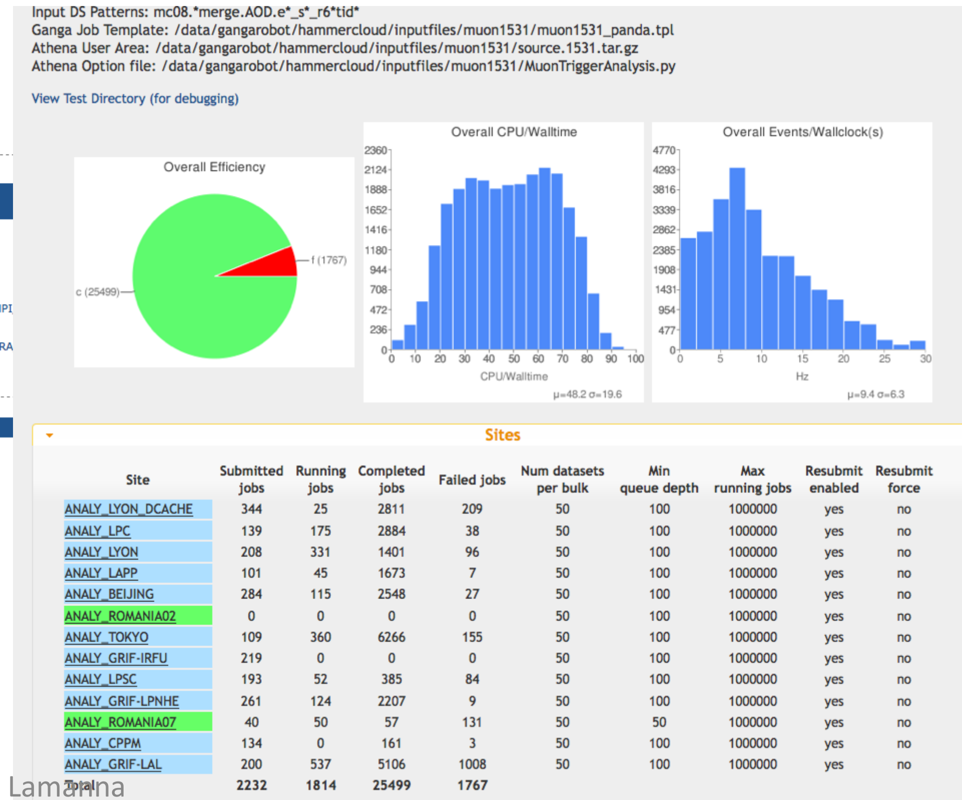
HammerCloud



Running and Scheduled Tests

state	id	host	start time (CET)	end time (CET)	clouds	sites
running	879	voatlas49.cern.ch	2009-11-20 15:01:00	2009-11-21 18:00:00	IT	INFN-MILANO-ATLASC_MCDISK
running	878	voatlas73.cern.ch	2009-11-20 15:00:00	2009-11-21 18:00:00	IT_PANDA	ANALY_INFN-MILANO-ATLASC
running	874	voatlas49.cern.ch	2009-11-20 11:00:00	2009-11-21 11:00:00	NL	CSTCDIE_MCDISK, JINR-LCG2_MCDISK, RU-PNPI 4 more...
running	868	voatlas73.cern.ch	2009-11-19 11:05:00	2009-11-21 11:00:00	ES	IFAE_MCDISK, IFIC-LCG2_MCDISK, LIP-COIMBRA 5 more...

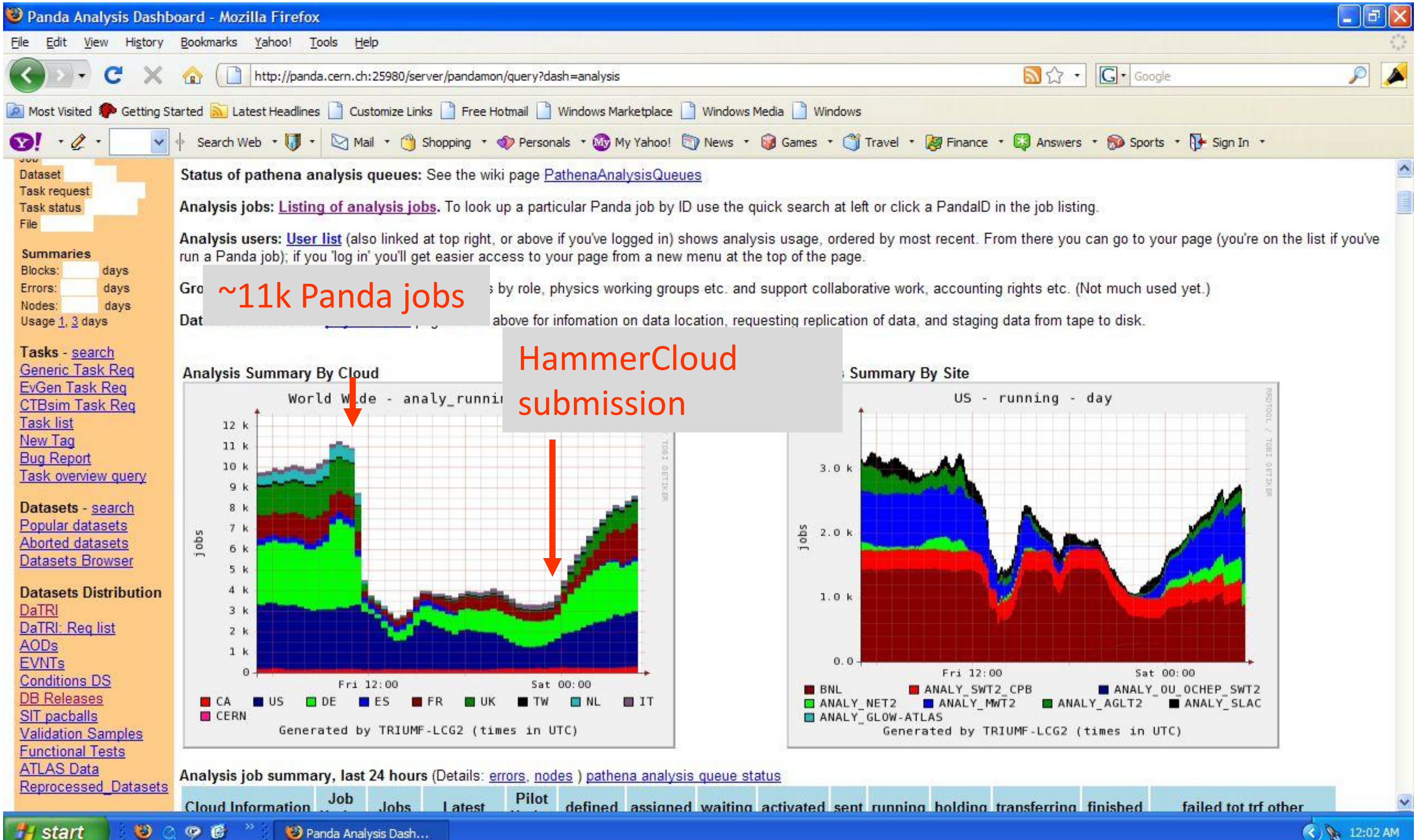
M. Palladin Università Udine – CERN OpenLab



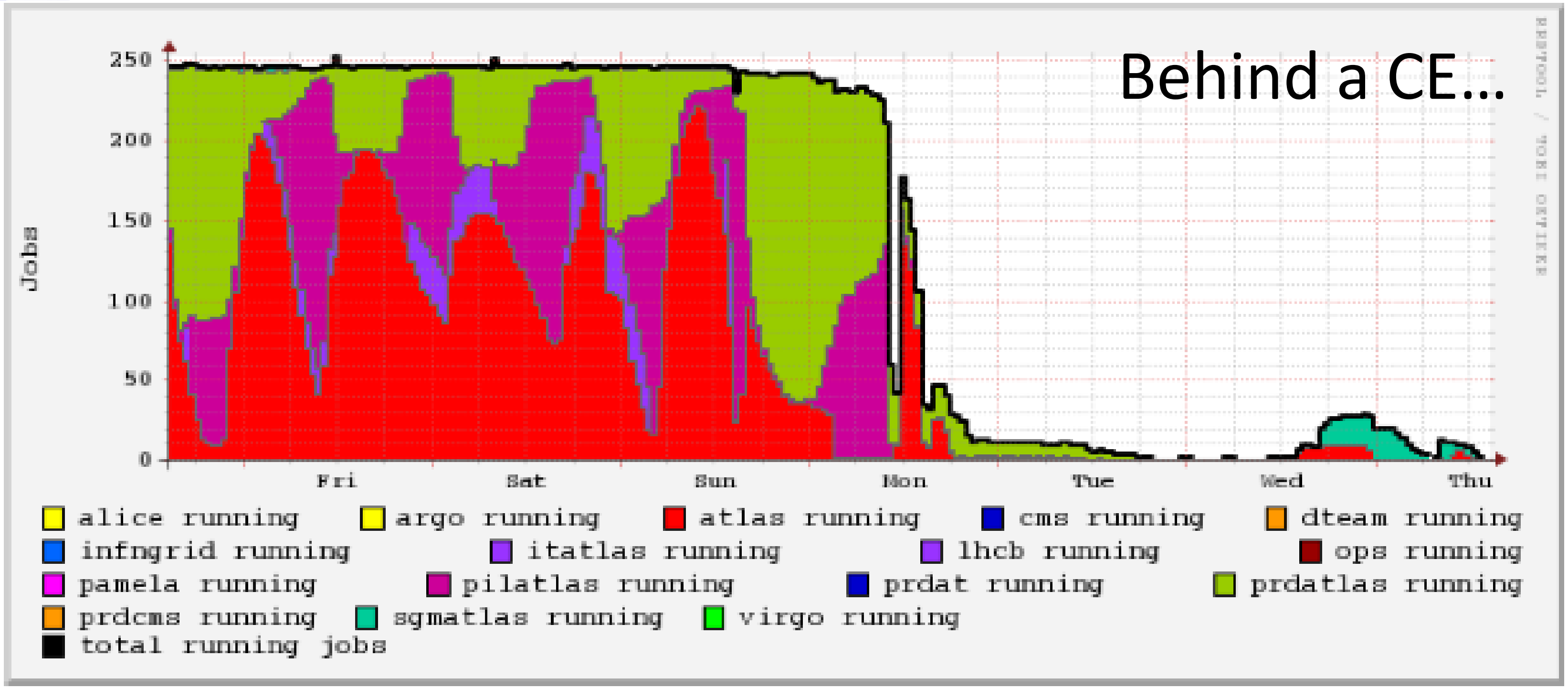
Sites

Site	Submitted jobs	Running jobs	Completed jobs	Failed jobs	Num datasets per bulk	Min queue depth	Max running jobs	Resubmit enabled	Resubmit force
ANALY_LYON_DCACHE	344	25	2811	209	50	100	1000000	yes	no
ANALY_LPC	139	175	2884	38	50	100	1000000	yes	no
ANALY_LYON	208	331	1401	96	50	100	1000000	yes	no
ANALY_LAPP	101	45	1673	7	50	100	1000000	yes	no
ANALY_BEIJING	284	115	2548	27	50	100	1000000	yes	no
ANALY_ROMANIA02	0	0	0	0	50	100	1000000	yes	no
ANALY_TOKYO	109	360	6266	155	50	100	1000000	yes	no
ANALY_GRIF-IRFU	219	0	0	0	50	100	1000000	yes	no
ANALY_LPSC	193	52	385	84	50	100	1000000	yes	no
ANALY_GRIF-LPNHE	261	124	2207	9	50	100	1000000	yes	no
ANALY_ROMANIA07	40	50	57	131	50	50	1000000	yes	no
ANALY_CPPM	134	0	161	3	50	100	1000000	yes	no
ANALY_GRIF-LAL	200	537	5106	1008	50	100	1000000	yes	no
Total	2232	1814	25499	1767					

Oct. 30 @ midnight



Balancing within activities



INFN-NAPOLI

- Green = production (PanDA production)
- Red = WMS analysis
- Purple = Pilot (PanDA analysis)
- Blue = WMS local share

Grid is about cooperation

- "Obviously" all computing centres cooperate!
- An interesting dimension is cooperation (sharing, exchange of information,…) across different scientific and technical communities
  - Because of **experience** *(long history of high performance innovative computing)*, **present situation** *(we have to cope with LHC physics requirements)* and **culture** *(large international collaborations)* HEP can help and collaborate with other sciences!

- This is very actually concrete!
  - Grid (projects) is an ideal playground for this
    - CERN is the lead partner of the largest Grid infrastructure (EGEE – EU funded)
    - EGEE is a multi-science grid project (It contributes to LCG but supports several other communities, like biomedical, earth sciences, nuclear fusion, astrophysics etc…

Enabling Grids for E-sciencE

ITU conference (2006)

The problem:
Assign frequencies for digital radio and television (international treaty)

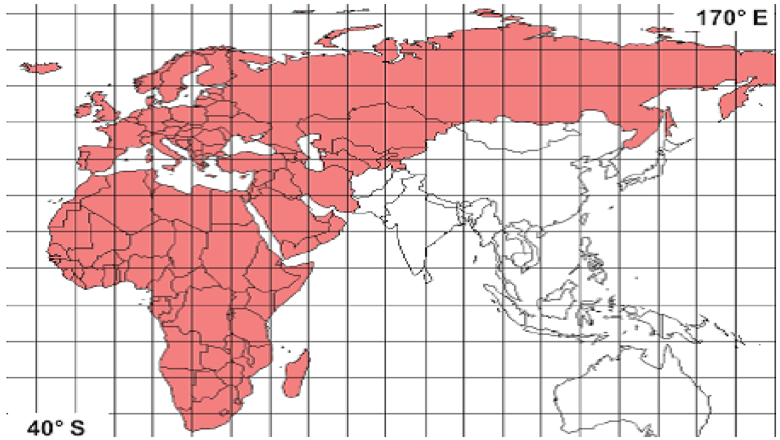
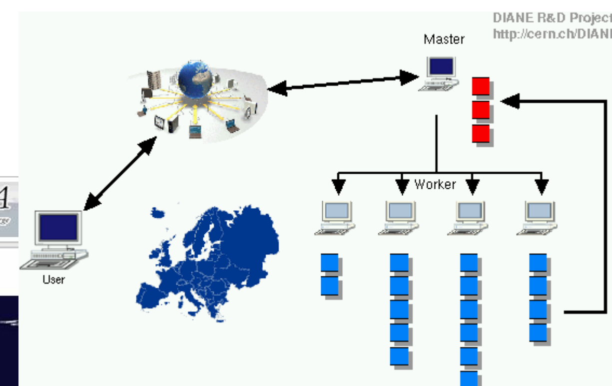
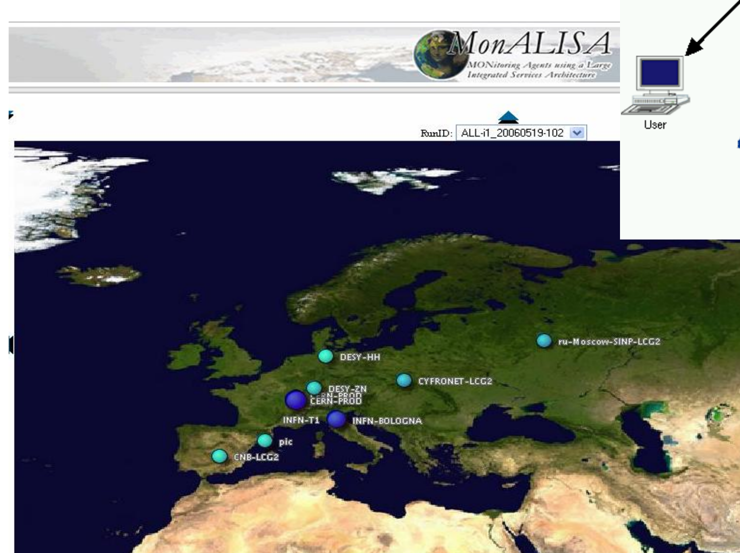


Figure 1
The extent of the planning area for the RRC-06

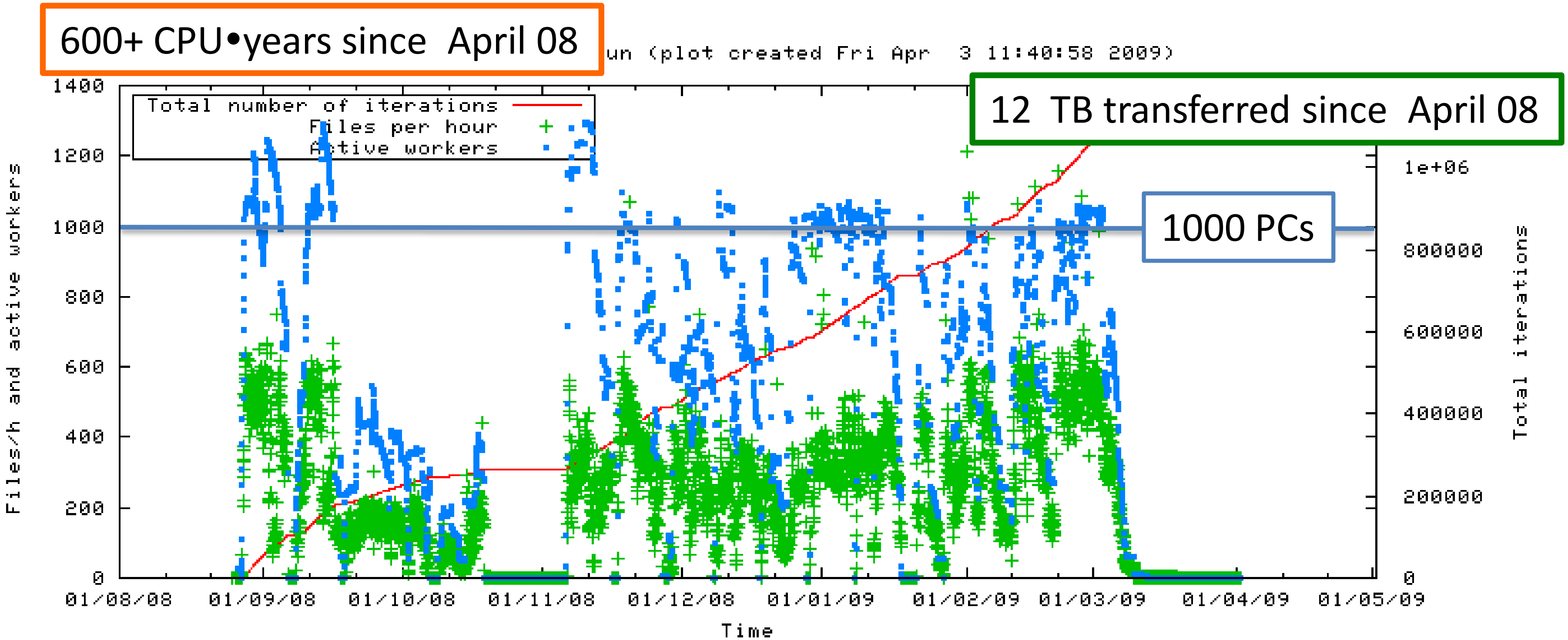
Critical point:
Need on dependability: verify (iteratively) the compatibility between radio stations





Solution:
Use the EGEE grid + a system used in ATLAS and LHCb to increase the reliability of the Grid

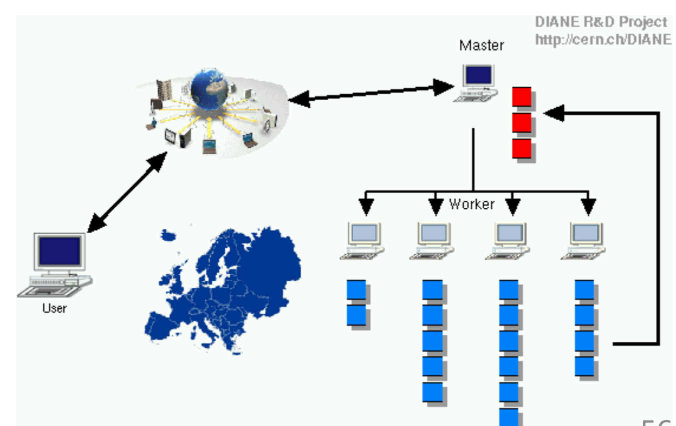
QCD on the Grid



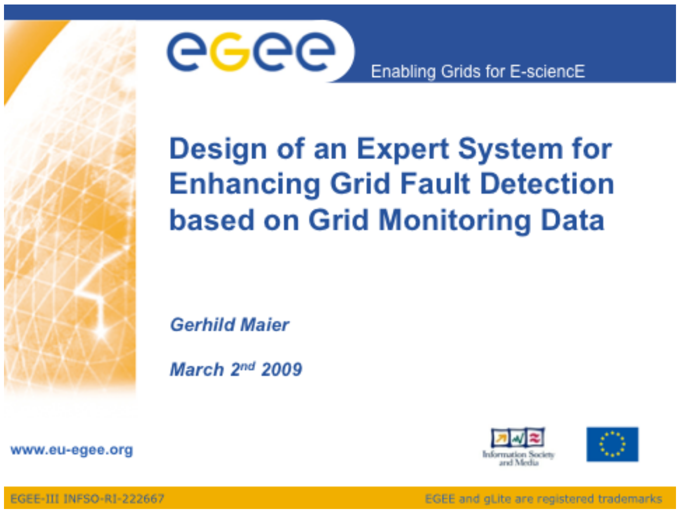
- Several day in 2007 (first campaign)

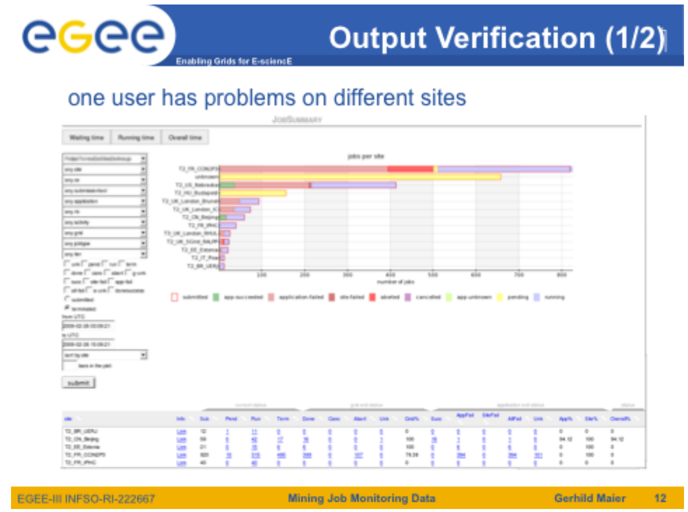
- 12 months of running in 2008/9
  - second campaign, several periods, graph Sep08-Mar09

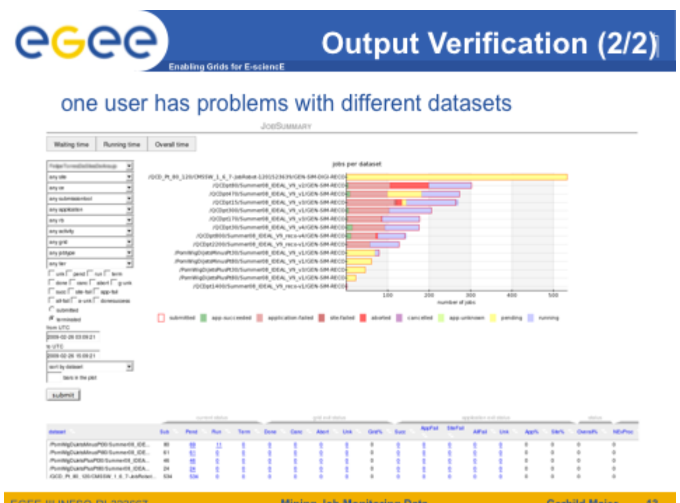
- Results regularly presented to leading conferences:
  - Lattice Conference -- Ph. De Forcrand (ETH and CERN)

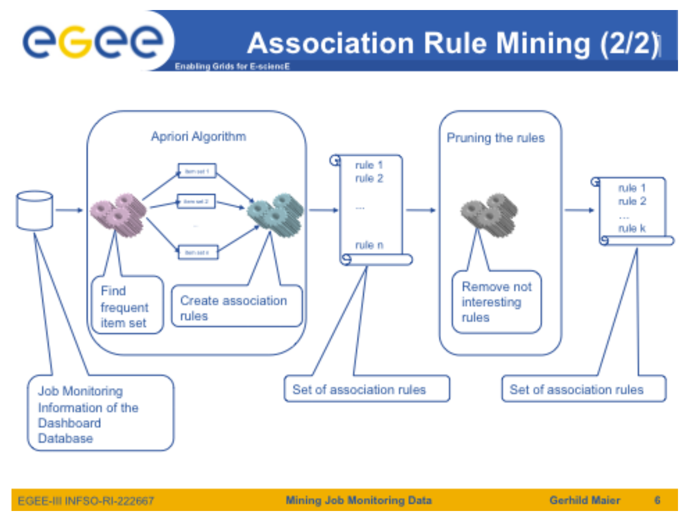


“Interactions” with computer science (PhD students)



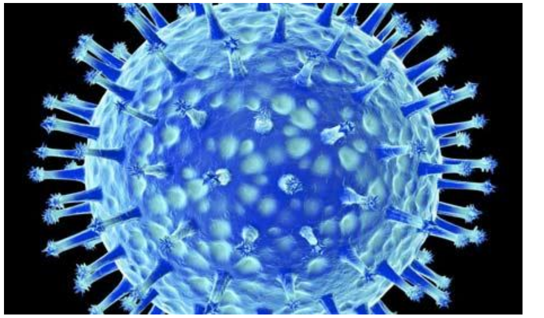






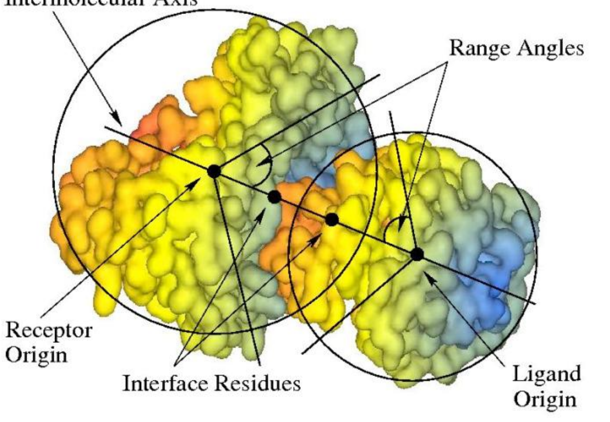
Gerhild Maier (CERN and Linz University)

Bird Flu



- Basic idea:
  - Compute how a given chemical interacts with a protein (e.g. belonging to a virus)
  - High affinity means the chemical is a potential drug against the virus
- In silico (i.e. use your PC):
  - Scan millions of chemicals (~10^3 s per chemical-protein pair)
    - With 1,000 PCs, 1 docking per second
  - Good candidate given to biologist (verification longer –and more expensive- than in silico docking)
    - In practice, you enrich the initial sample saving time (and money)
    - Essential to fight to pandemic (H5N1) or neglected diseases (like Malaria)
- WISDOM collaboration
  - Malaria
  - H5N1 (Bird Flu)

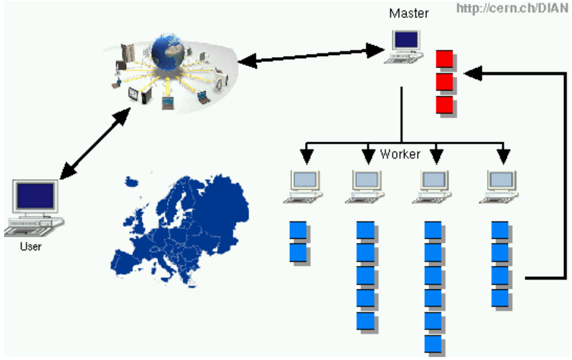
Molecular docking



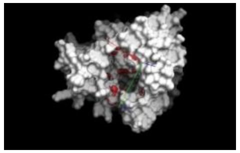
...More communities sharing tools designed for LHC and being useful in other domains!

Always reuse the same pattern:

1. Divide the problem in small task that can be distributed independently
2. Distribute these tasks to "dumb" workers (Grid processes)
3. Recollect the results



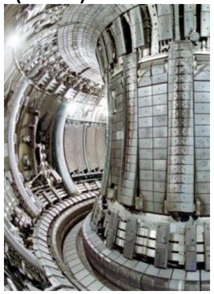
http://cern.ch/DIANE



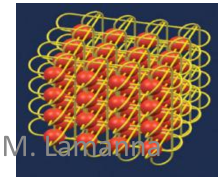
Academia Sinica Genomics Research Center

HARP

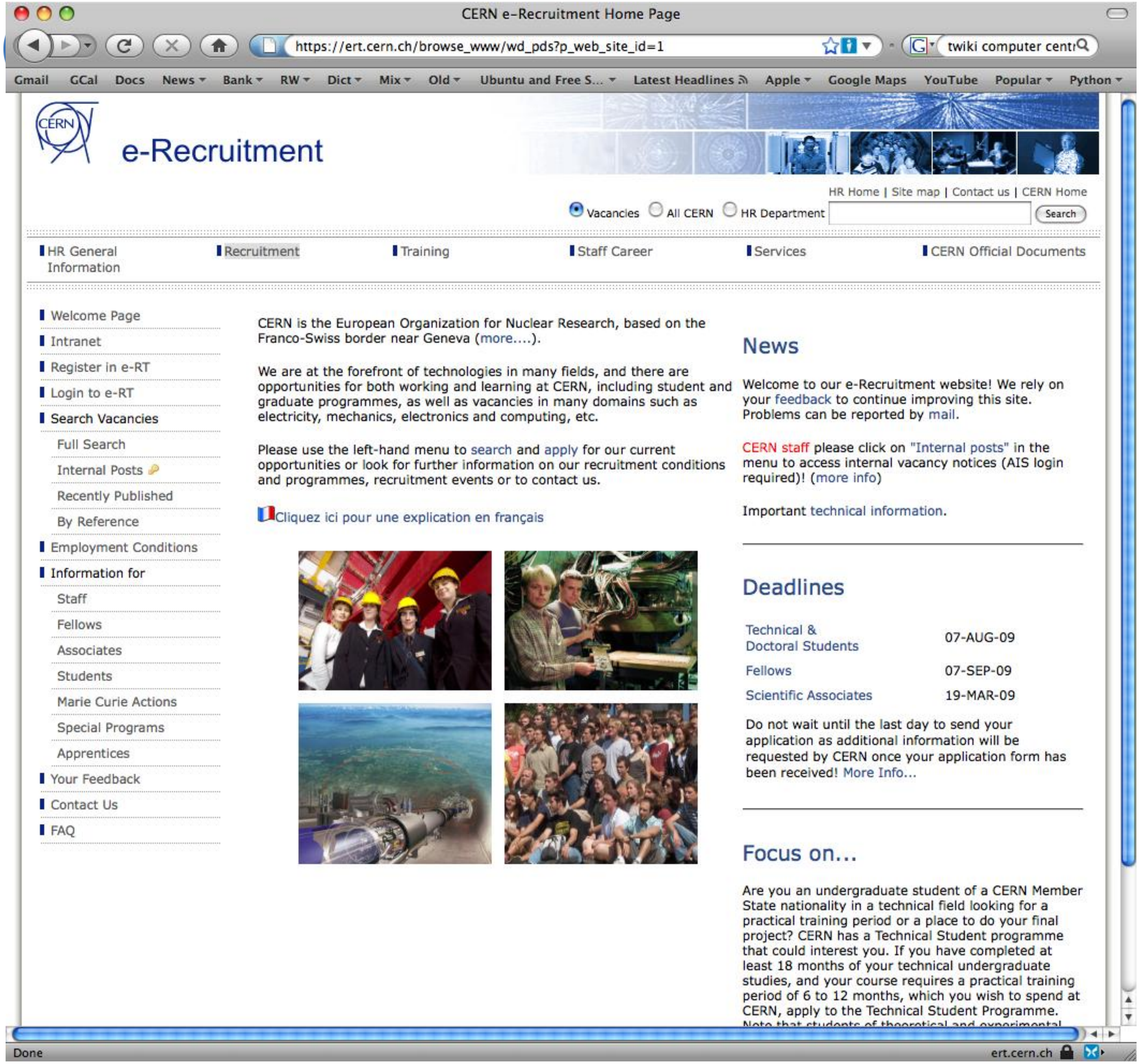
Nuclear fusion (ITER)



Lattice QCD



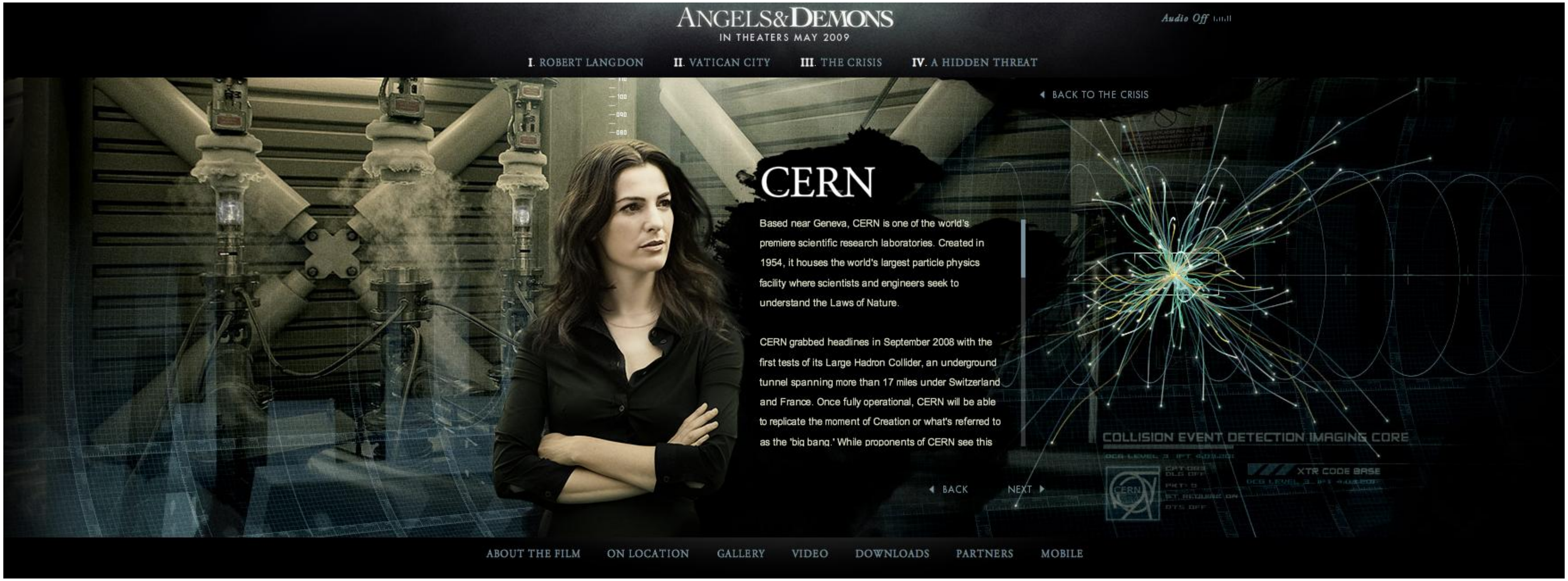
http://cern.ch/ganga



CERN options for students

- University level (BS/Master)
  - Summer student
  - OpenLab summer students
- Master thesis
  - Technical student (non physicist)
- PhD students
  - Doctoral students
- Young scientists/engineers
  - Fellowship
  - Other programmes

Questions?


ANGELS&DEMONS
IN THEATERS MAY 2009
Audio Off
I. ROBERT LANGDON
II. VATICAN CITY
III. THE CRISIS
IV. A HIDDEN THREAT
BACK TO THE CRISIS
CERN
Based near Geneva, CERN is one of the world's premiere scientific research laboratories. Created in 1954, it houses the world's largest particle physics facility where scientists and engineers seek to understand the Laws of Nature.
CERN grabbed headlines in September 2008 with the first tests of its Large Hadron Collider, an underground tunnel spanning more than 17 miles under Switzerland and France. Once fully operational, CERN will be able to replicate the moment of Creation or what's referred to as the 'big bang.' While proponents of CERN see this
BACK
NEXT
COLLISION EVENT DETECTION IMAGING CORE
XTR CODE BASE
ABOUT THE FILM
ON LOCATION
GALLERY
VIDEO
DOWNLOADS
PARTNERS
MOBILE