Report from the
CHIPP Computing Board
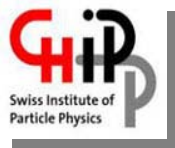
Gersau
Aug 2010

Christoph Grab

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Christoph Grab,

Issues
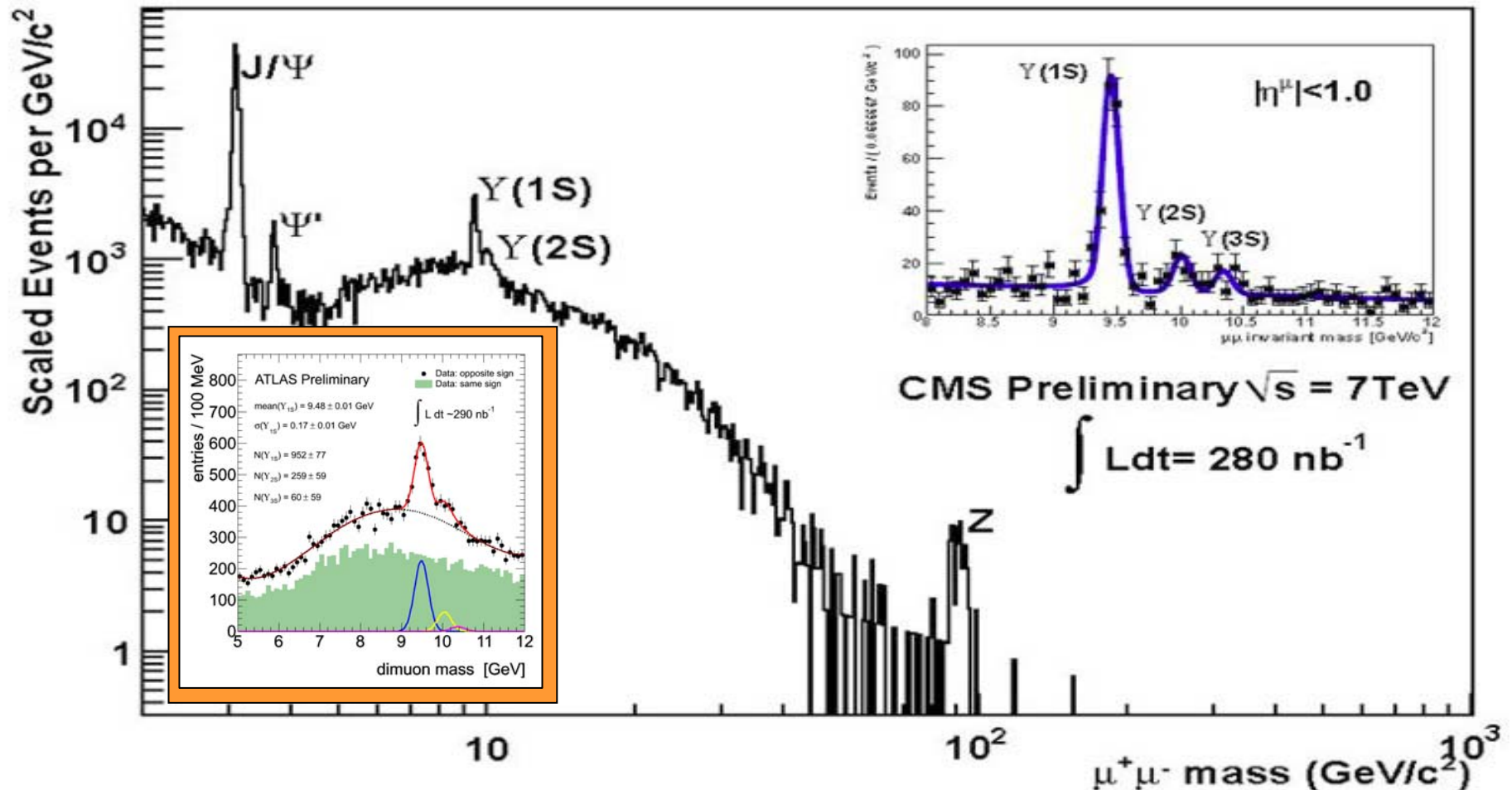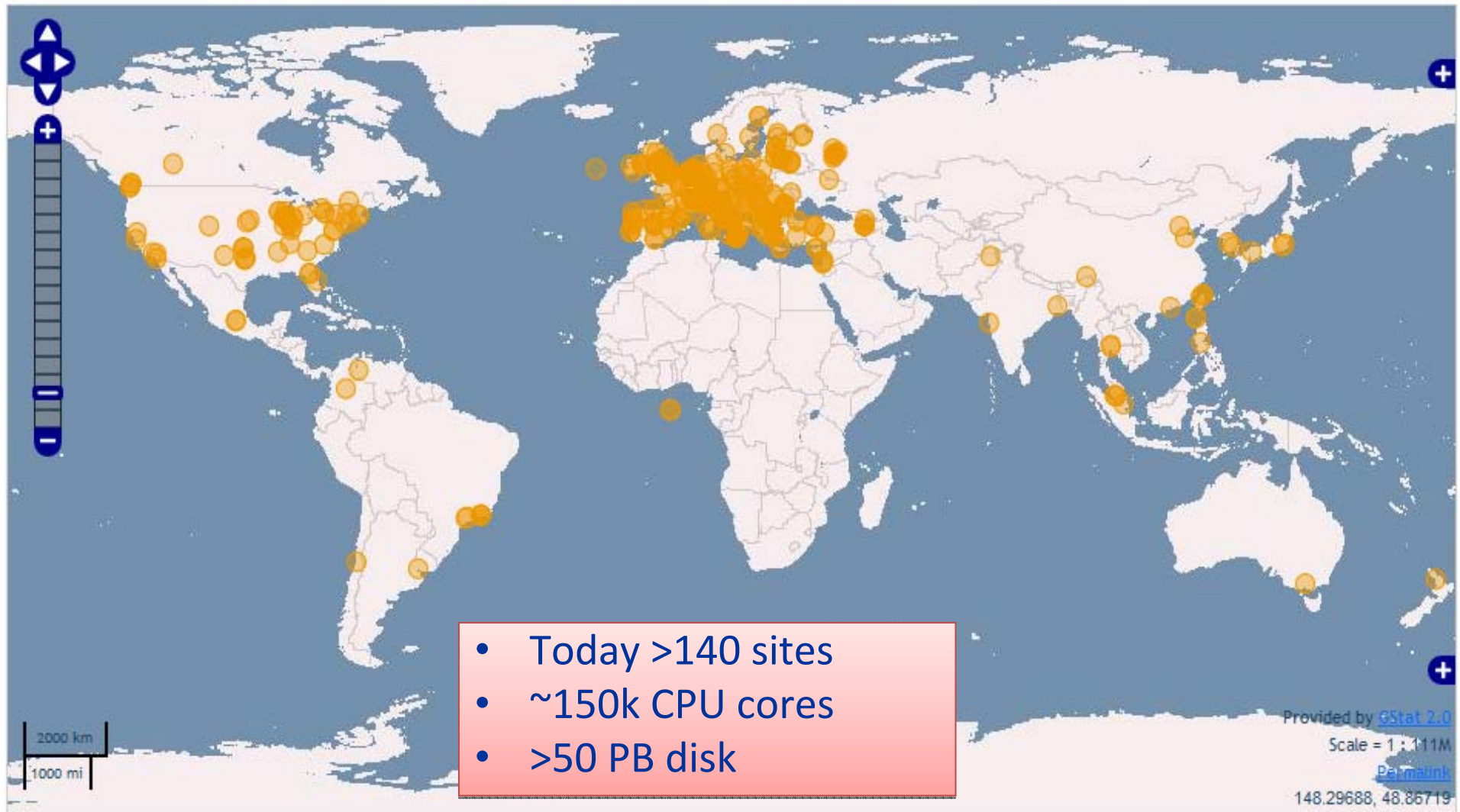
1. **Some general comments on Grid Operations**

2. **Status Swiss Computing infrastructure for HEP analysis**
   ➔ tier-2 and tier-3s

3. **Relation CHIPP to Swiss NGI to European EGI**

The LHC computing grid "works successfully"

→ *proof is the physics output by all 4 experiments*

Worldwide resources

- Today >140 sites
- ~150k CPU cores
- >50 PB disk

CERN

US-BNL

Amsterdam/NIKHEF-SARA

Ca-TRIUMF

Bologna/CNAF

Taipei/ASGC

NDGF

US-FNAL

Barcelona/PIC

Lyon/CCIN2P3

UK-RAL

De-FZK

WLCG Collaboration Status

Tier 0; 11 Tier 1s; 64 Tier 2 federations

Today we have 49 MoU signatories, representing 34 countries:

Australia, Austria, Belgium, Brazil, Canada, China, Czech Rep, Denmark, Estonia, Finland, France, Germany, Hungary, Italy, India, Israel, Japan, Rep. Korea, Netherlands, Norway, Pakistan, Poland Portugal, Romania, Russia, Slovenia, Spain, Sweden, Switzerland, Taipei, Turkey, UK, Ukraine, USA.
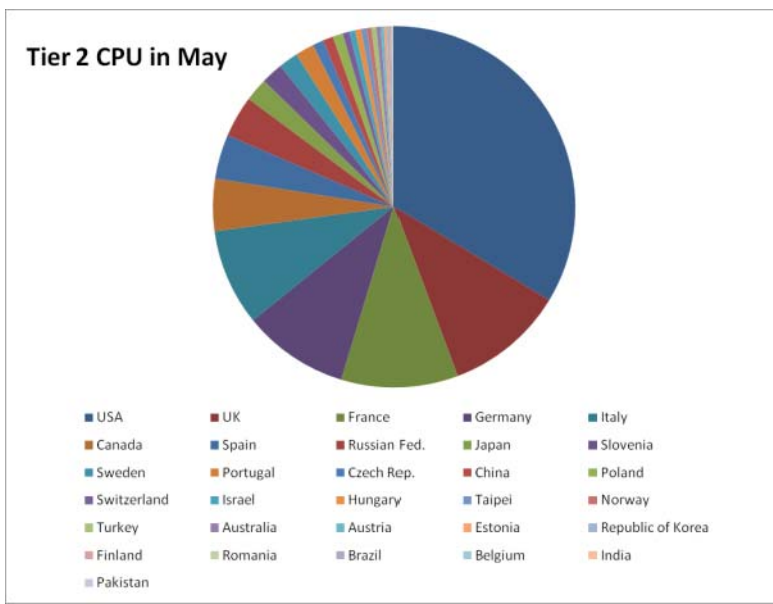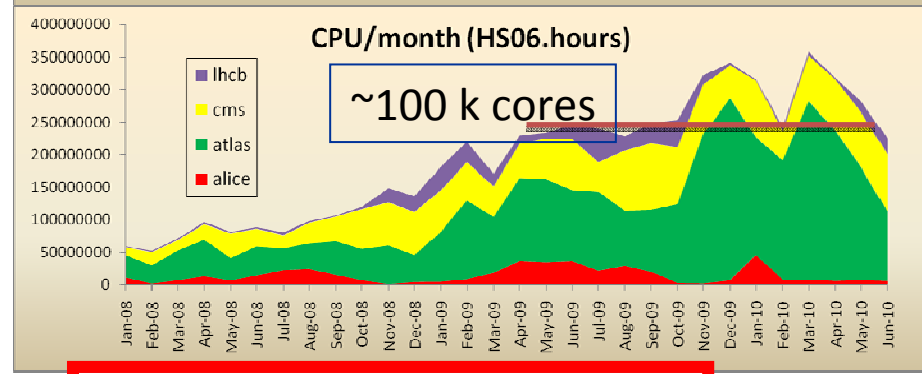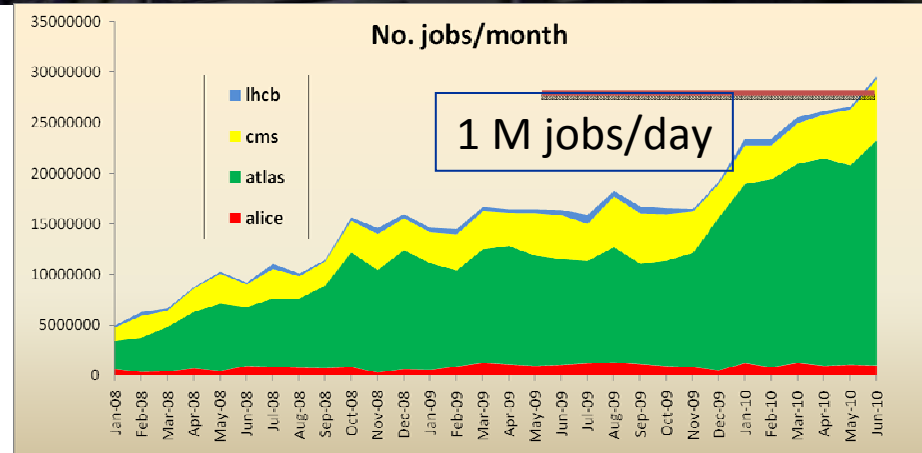
Use of CPU …

CPU Delivered in May

Tier2s 59%

8% CERN
6% kit
5% NL
4% in2p3
4% ral
3% bnl
3% fnal
3% cnaf
2% triumf
2% pic
1% ndgf
0% asgc
Tier2s

Tier 2 CPU in May

USA, UK, France, Germany, Italy, Canada, Spain, Russian Fed., Japan, Slovenia, Sweden, Portugal, Czech Rep., China, Poland, Switzerland, Israel, Hungary, Taipei, Norway, Turkey, Australia, Austria, Estonia, Republic of Korea, Finland, Romania, Brazil, Belgium, India, Pakistan

No. jobs/month

lhcb, cms, atlas, alice
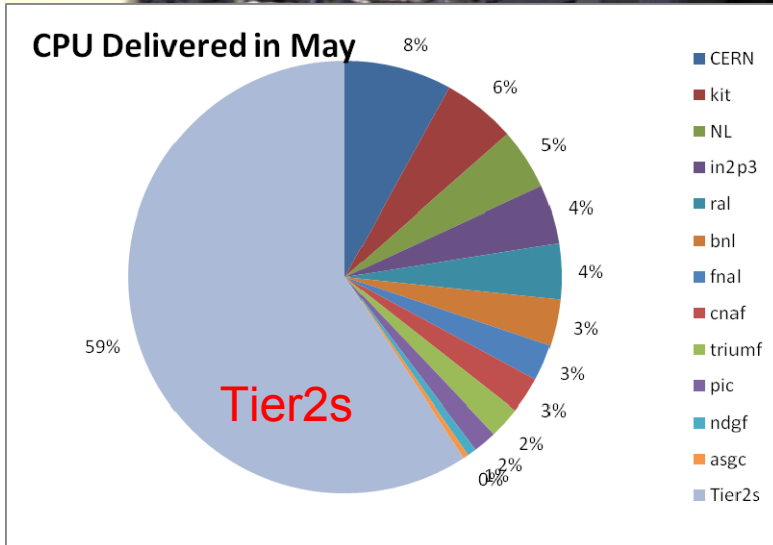
1 M jobs/day

CPU/month (HS06.hours)

lhcb, cms, atlas, alice

~100 k cores

- Peaks of 1M jobs/day now
- Use ~100k cores equivalent
- Tier 2s heavily used wrt Tier 1s

Data distribution for analysis

Data Distribution for Analysis

CMS

- Data transferred from Tier-1's

  - 49 Tier-2 sites received data

  - \> 5 PB transferred in last 120 days

  - average rate 562 MB/s

  - max rate 1407 MB/s

- Data transferred between Tier-2's

  - 41 Tier-2 sites received data

  - \> 2.5 PB transferred in last 120 days

  - average rate 254 MB/s

  - max rate 853 MB /s

  - full mesh approach

  - Data distribution re-balances itself

  - Datasets produced at Tier-2's can be distributed to others

Markus Klute, MIT   ICHEP - July 2010
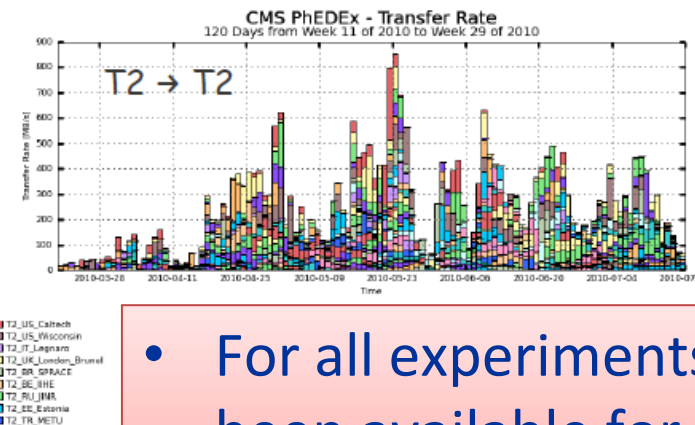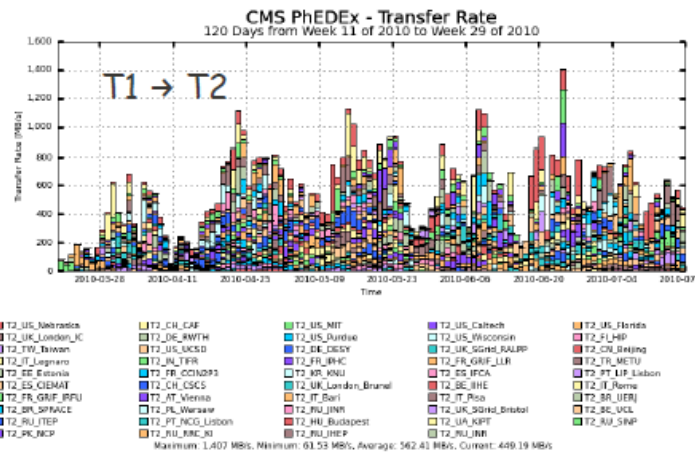
- For all experiments: early data has been available for analysis within hours of data taking

Status of the
Swiss Tier-2
Regional Centre

Cluster has been setup up in phases :

➔ has reached in Q1/2010 the "*originally planned LHC startup size*"

→ Technology choice so far: SUN blade centres + multicore CPUs

➔ *Swiss tier-2 is in routine operation ….* ☺

Since 2009

Implementations in Q4/09-Q1/10
- added a higher bandwidth backbone on Infiniband technology
- replaced local disk scratch system by a global Lustre system
- upgraded/added storage elements (Thors" with ZFS on disk)
- upgraded worker nodes (CPU) → faster, lower power, higher-density
  →and turned old nodes to ATLAS Switzerland for re-use in tier3s
- + maintain operation in parallel to upgrade with minimal interruption !

➔ total resources after upgrade reaches "original design":
- WN: 768 cores a 3.7 k ~ 2500 kSI2k
- Storage: total of ~ 800 TB disk; no tapes
- Service nodes, VO-boxes, etc unchanged.

Growth corresponds to Swiss commitment in terms of compute resources supplied to the expt's according to the signed MoU with WLCG.

Reached the phase C ("LHC Startup Size") – operational since Q1/2010
➔ total of ~2500 kSI2k (768 cores) ; total ~800 TB storage

Swiss Tier-2 Size Evolution

CG_0710

Swiss Tier-2: Facts and Figures (2)
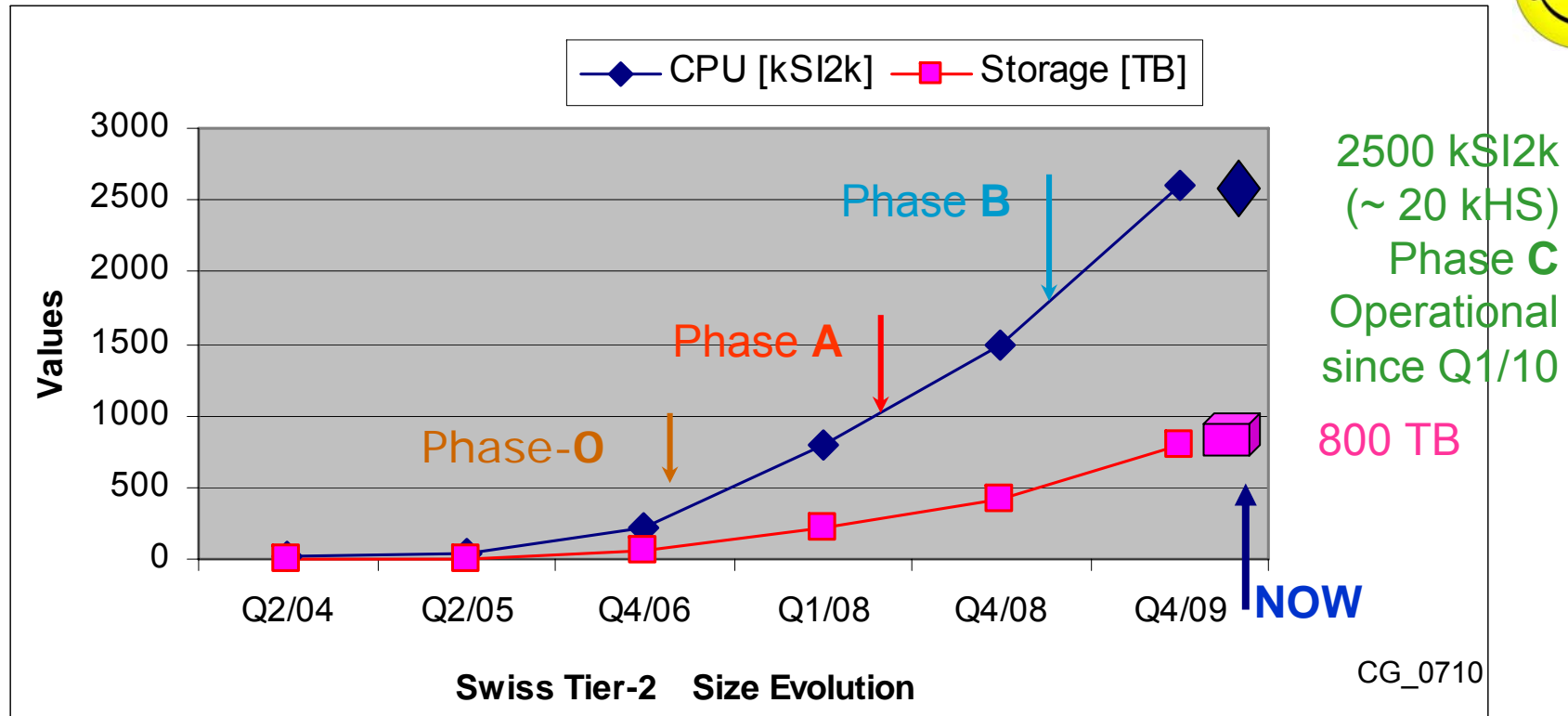
- **Operational Manpower at Tier-2:**
  - Operation at CSCS: ~ sums to 2.5 FTEs (IT experts, spread over 5 persons)
    (1.5 FTE by CSCS; 1 FTE by CHIPP since Q2/2010)
  - support of experiment specifics by scientists of experiments via
    one dedicated contact person per experiment (by CHIPP PostDocs)
    → in total an additional ~1.5 FTE   → *total for tier-2: ~ 4 FTE*

- **Support - strategy at CSCS:**
  - CHIPP cluster is integrated in new "Co-location Services";
  - CSCS is committed to continue support for CHIPP and CH-Tier-2 (as in MoU)

- **Financing** (HW and service) :
  - Financing of hardware mostly through SNF/FORCE (>95%),
    with some contributions by Universities + ETH + PSI

  - Infrastructure and operations provided by CSCS i.e. ETHZ
    (additional costs once at USI for infrastructure)
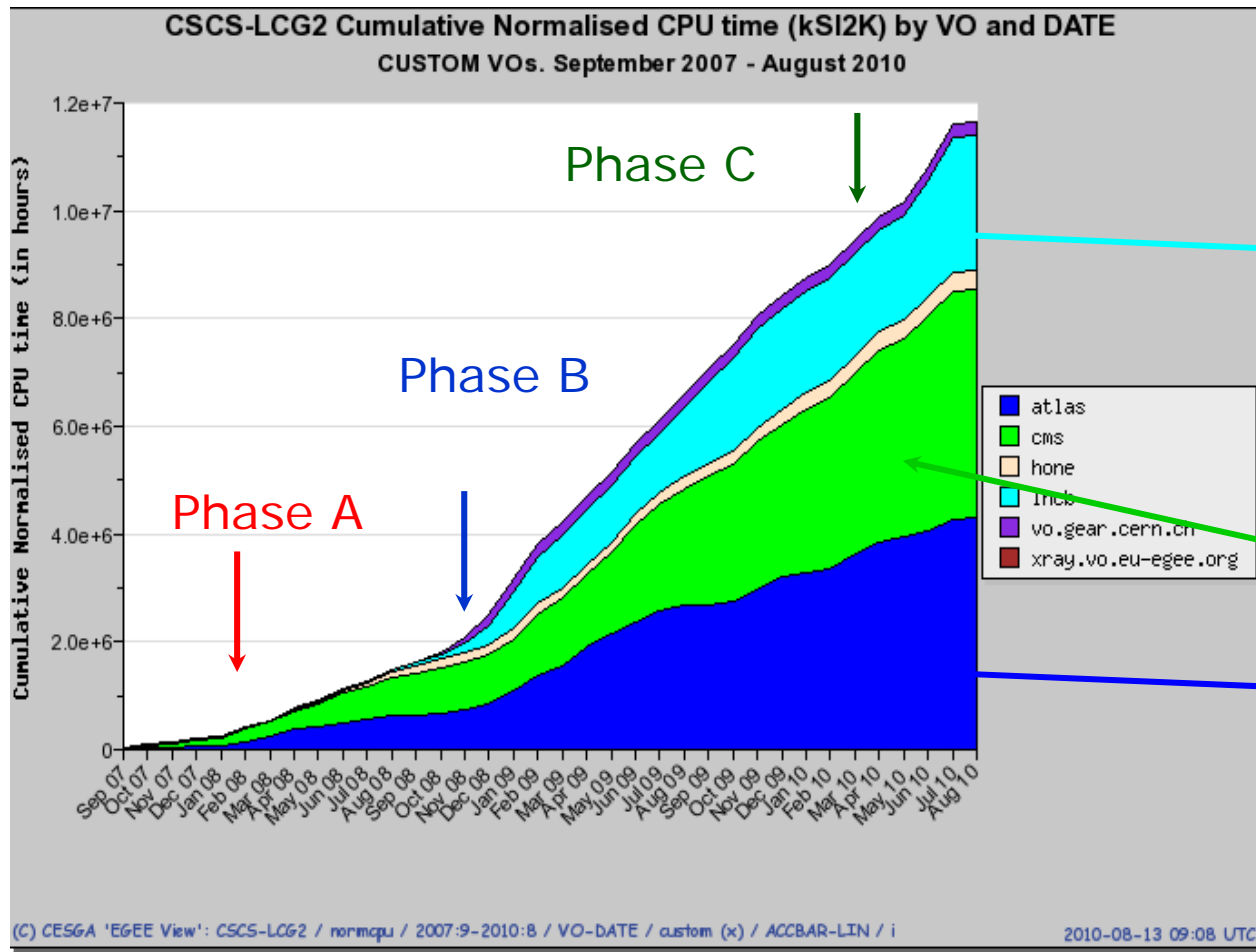
Swiss Tier-2: Planning

Note : present figures are still based on experiments numbers, estimated in the pre-data-taking era. ➔ experiments announced increased requirements …!

- Plans for 2010-2012 :
  - ➢ main goal is stability+reliability+availability; rolling replacements and "whatever additions required" to react to expt's needs
  - ➢ also ONE FTE person to support Tier-2

- **CSCS will move location to new campus at USI (Lugano);** foreseen in ~2012; also LHC shutdown in 2012…

- **Question: what about our cluster?**
  - ➢ Option 1: keep cluster running at Manno, setup new one in parallel at USI ➔ no interruption in operation; clean solution.
  - ➢ Option 2: move cluster : dismantle / transport /setup again. ➔ interruption of operation for months ???

  => Evaluations of costs in progress; option 1 preferred.
 *Feedback welcome !*

Some Performance Plots
of Tier-2 Operations

CSCS-LCG2 Cumulative Normalised CPU time (kSI2K) by VO and DATE
CUSTOM VOs. September 2007 - August 2010

Phase A
Phase B
Phase C

Incremental CPU-usage over last years

LHCb
CMS
ATLAS

- **Tier-2** is up and has been in stable operation for years ! continuous contributions of resources to experiments.

- Spare cycles given to other VO (eg. H1, theory (CERN) …)

CSCS-LCG2 Normalised CPU time (kSI2K) per VO
CUSTOM VOs. September 2007 - August 2010
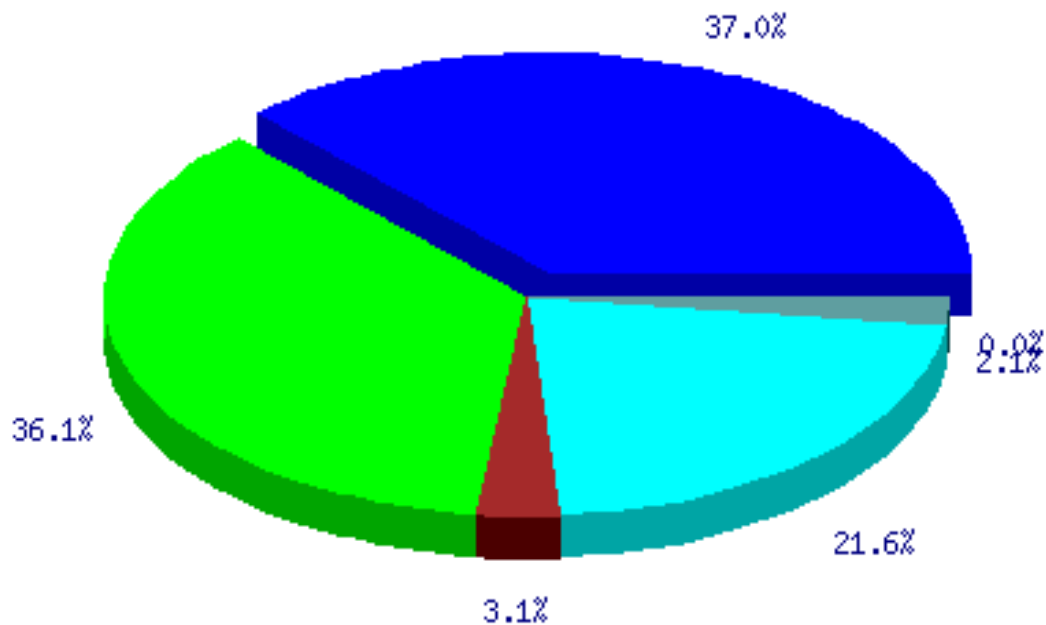
Swiss Tier-2

atlas
cms
hone
lhcb
vo.gear.cern.ch
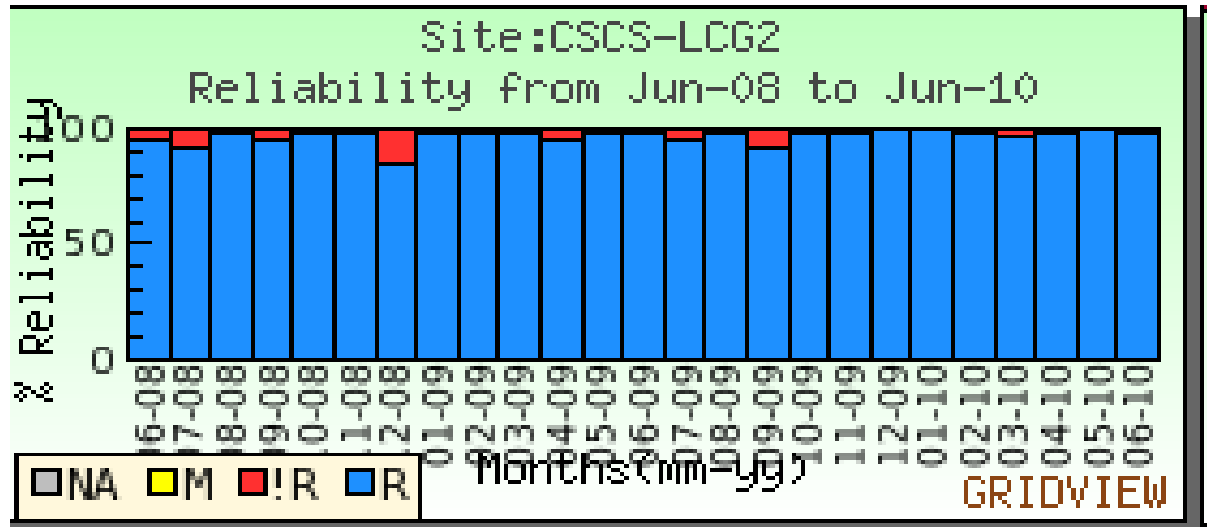Others

37.0%

36.1%

3.1%

0.0%

21.6%

ATLAS
CMS
LHCb
H1
others

- Shares between VOs overall pretty well balanced

- **CSCS-LCG site**
  reliability
  for last 2 years
  06/08 – 07/10

- **CSCS-LCG Site**
  availability
  for last 2 years
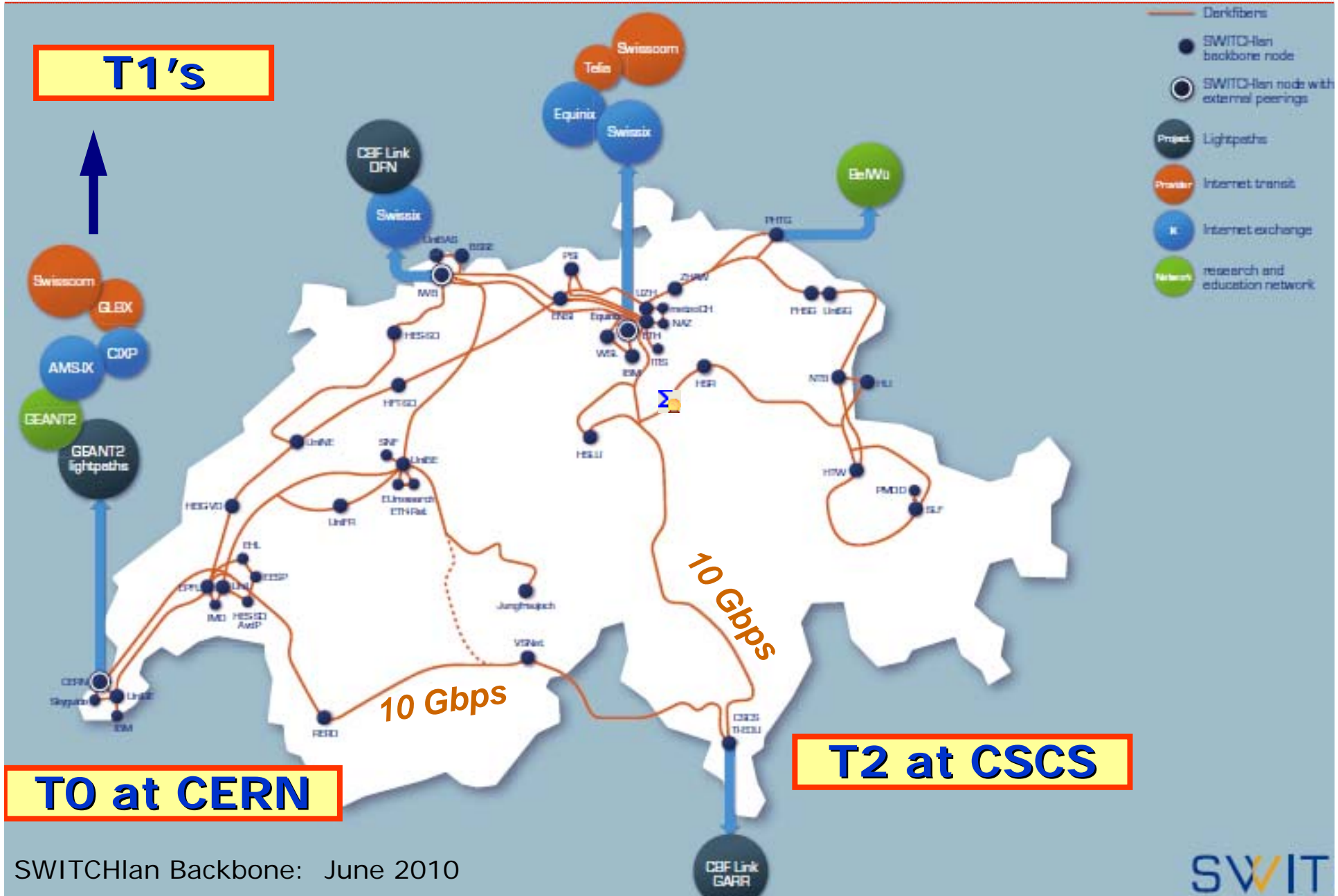 **06/08 – 07/10**
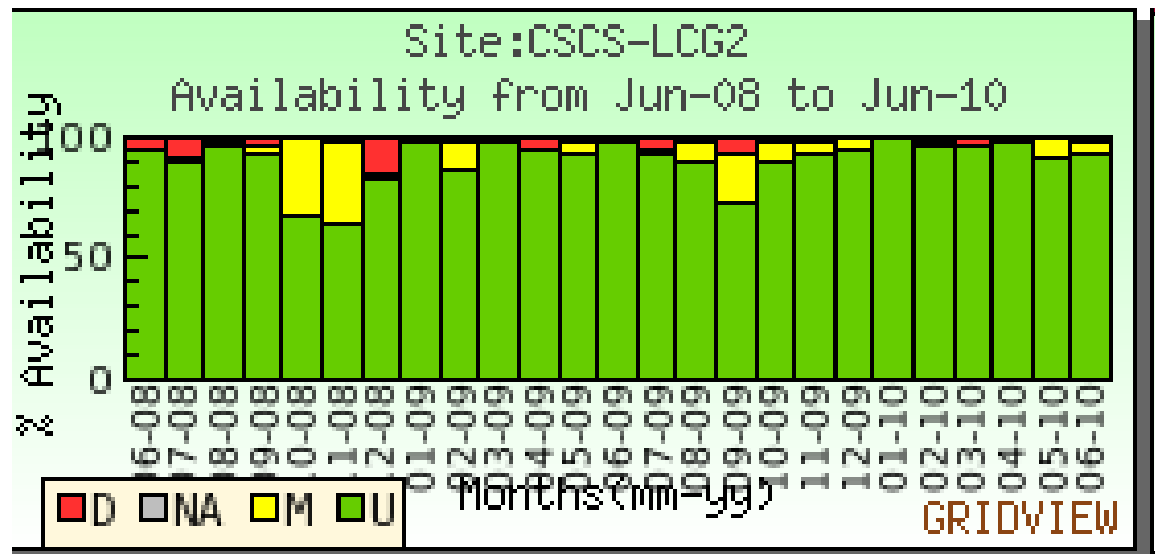
D: down
M: maintenance
U: up

Network

Swiss Tier-2: Network

No changes since 2009

- **Network traffic**:
  - **routing via SWITCH : two redundant lines >10Gbps to CERN and Europe**
  - transfer rates reached up to 10 TB / day   from FZK (and CERN)

  ➔ *Presently still sufficient bandwidths …*

- Provisions are there to increase bandwidths if needed within months

SWITCHlan Backbone: June 2010

Status of the Swiss Tier-3 Centres

Swiss Tier-3 Efforts

- **Large progress** seen over last year for all 3 experiments.
  upgrades in progress nearly everywhere !

- **Close national collaboration between Tiers:**

  - Tier-3 contacts are ALSO experiment's site contacts for CH Tier-2.

- **ATLAS** : operates the Swiss ATLAS Grid → federation of clusters at
  - Bern uses local HEP + shares university resources
  - Geneva operates local cluster
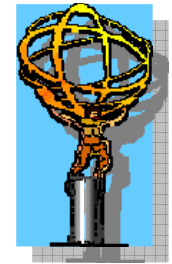
- **CMS** : ETHZ + PSI+ UZH run a combined Tier-3
  - located at and operated by PSI IT

- **LHCb :**
  - EPFL : operates huge local cluster → is DIRAC site
  - UZH uses local HEP + plans to share university resources

Site (#users)	Nr cores	CPU (kSI2k)	Storage (TB)	Comments (8/10)
ATLAS BE (12) GE (~55)	200+300sh 268	~600 462	100 177	BE: standing Atlas production; GE: identical SW-environment to CERN; direct line to CERN.
CMS ETHZ, PSI, UZH (21)	224	~1200	270	GRID SE + UI :direct GRID access.
LHCb EPFL (40) UZH (6)	480 48+sh	~2200 ~400	75 20	EPFL is DIRAC site; test, debug jobs. UZH: most jobs on grid
Total Tier-3		~5000	640	cf: Tier-2: 2500 kSI2k (est), 800 TB

- Tier-3 capacities : similar (>=) size in CPU as Tier-2
- Note: CPU numbers are estimates with large uncertainties !!!
- conversion factors: 1 kSI2k ≈ 4-7 HS06 ≈ 5 Cint06; (rough approximations !!)

Swiss NGI – EGI   and CHIPP relations

Roles of EGI, NGI, Swing, CHIPP (ref)

- Up to 1.4.2010 the overall "operation, R&D etc." for the GRID was done by EGEE (Enabling Grid for E-sciencE)

- In 2010,EGEE-III was succeeded by EGI, the "European Grid Initiative",
  - "EGI provides a more sustainable way to centrally coordinate, evolves and operates the current European grid infrastructure"
    to guarantee its long-term availability for performing research and innovation..."
  - EGI depends on and is supported by the NGI (National Grid Initiatives), set up per country to manage the national grids
  - Status: MoU signed by SWITCH (participate in FP7 project InSPIRE-EGI)

- Switzerland's official NGI is SWITCH (is LI ad interim till 12.2011)
  - SWING (Swiss National Grid Initiative) is destined to assume this role

- Swiss NGI is obliged to provide services for community :
  - CHIPP/CSCS has to provide its share (essentially the same as now), as provided now to WLCG (through EGEE): support, operations etc.
  - compensating resources expected from SER and EGI

For reference only ...

- **Collaborations in/with EU initiatives :**
  - Operation of European GRID services : transition EGEE-III ➔ **EGI.eu**
  - NGI represent countries in EGI.eu
  - Switzerland is represented by ONE leading institution
    - SWITCH agreed to act as leading institution;
      - ➔ SWITCH coordinates the Swiss contributions to EGI.eu
   - after 31.12.2011 SWiNG will become the official SWISS NGI

- **Main user base in Switzerland:**
  - Dominated by particle physics ➔ CHIPP
  - CHIPP represented by ETHZ (through chair of CCB, a.p. C.Grab)
  - Theor. chemistry represented by UZH (through CCC chair K.Baldridge)
  - other communities directly through SWITCH/ETHZ/UZH

- *Collaborations have been put on solid grounds with official, legal agreement documents between : EGI - NGI(SWITCH) – CHIPP(ETH)*

Summary

- **Swiss tier-2 reliably operates and delivers the Swiss pledges to the LHC** experiments in terms of computing resources since Q2/2005
  - Reached the originally planned "LHC-startup size" in Q1/2010
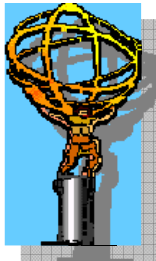 - Tier-3 centres strongly complement Tier-2

We are contributing to the PHYSICS !
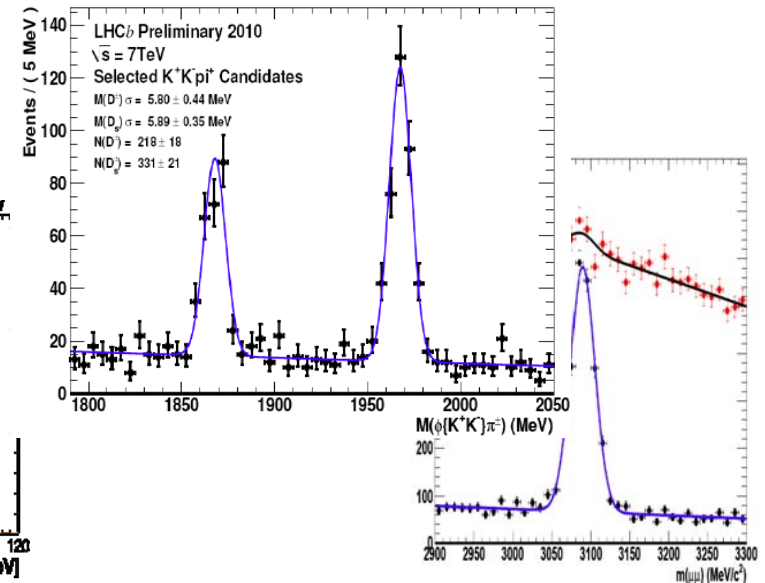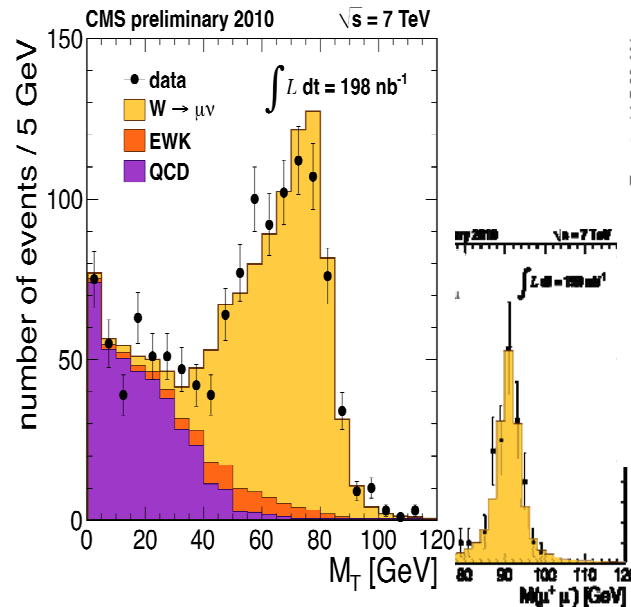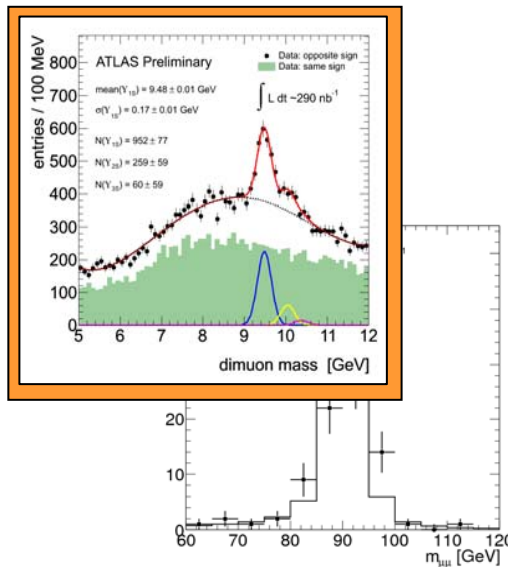
C.Grab (ETHZ)   [chair CCB]
D.Feichtinger (PSI)
L.Sala (ETHZ)

M.Goulette, S.Gadomski, (UNI Ge)
S.Haug (UNI Bern)

R.Bernet (UNIZH)
Y.Amhis (EPFL)

Thanks

P.Fernandez, P.Oettl, F. Georgatos [CSCS]
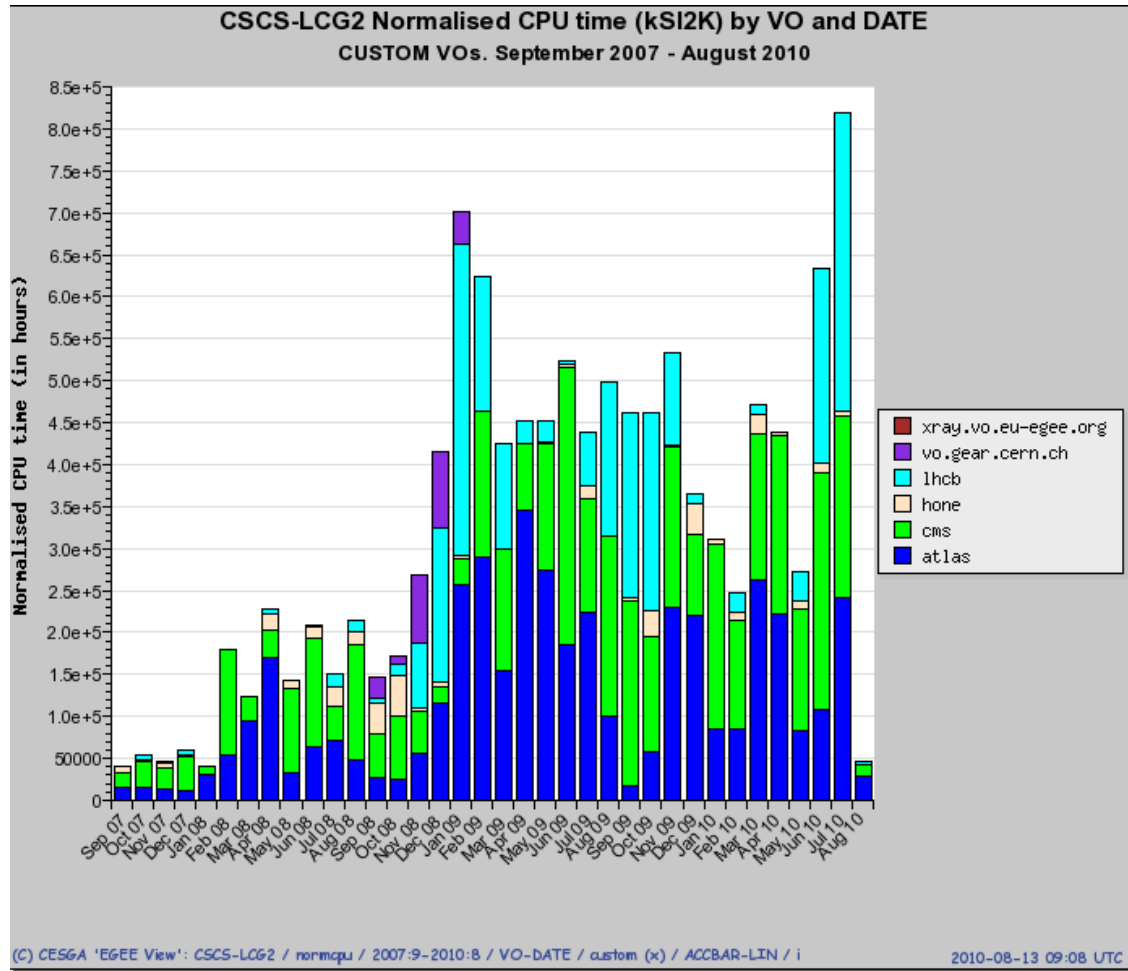
and many more …

System Phase B operational since Nov 2008

System Phase C operational since Mar 2010

❖ **CPUs:** 768 cores ~2500 SI2K
X6275 blades, twin node configuration
2x (2x E5540 @ 2.53GHz, 24GB RAM)

❖ **Storage : ~ 800 TB**
 - 27 X4500 - thumpers
 - 10 X4540 "thors"

Normalised CPU time
per month

ATLAS

CMS

LHCb

H1

others

- Shares between VOs varies over time (production, challenges...)

EPFL: Tier 3 Status

Hardware and Software

- Intel Xeon Cluster, 64-bit, SL4 (Beryllium)
  - 58 worker- and 2 interactive-nodes with $2 \times$ quad-core (2.5 GHz)
  - $3.6 + 53.8 + 17.5$ Tb of storage in RAID6/5/0 configuration

- Machines identical to those in the LHCb pit

- Uses SLC4 & SLC5 binaries of LHCb software

- Storage managed by one node - need more

Current Status and Operation

- Migrating to SLC5

- Used mainly for user analysis and MC generation

- Some extra hardware required for heavy analysis

- GRID better for analysis but not all users have certificate

Zurich HEP Cluster:

New Intel Cluster, 64-bit, SLC5
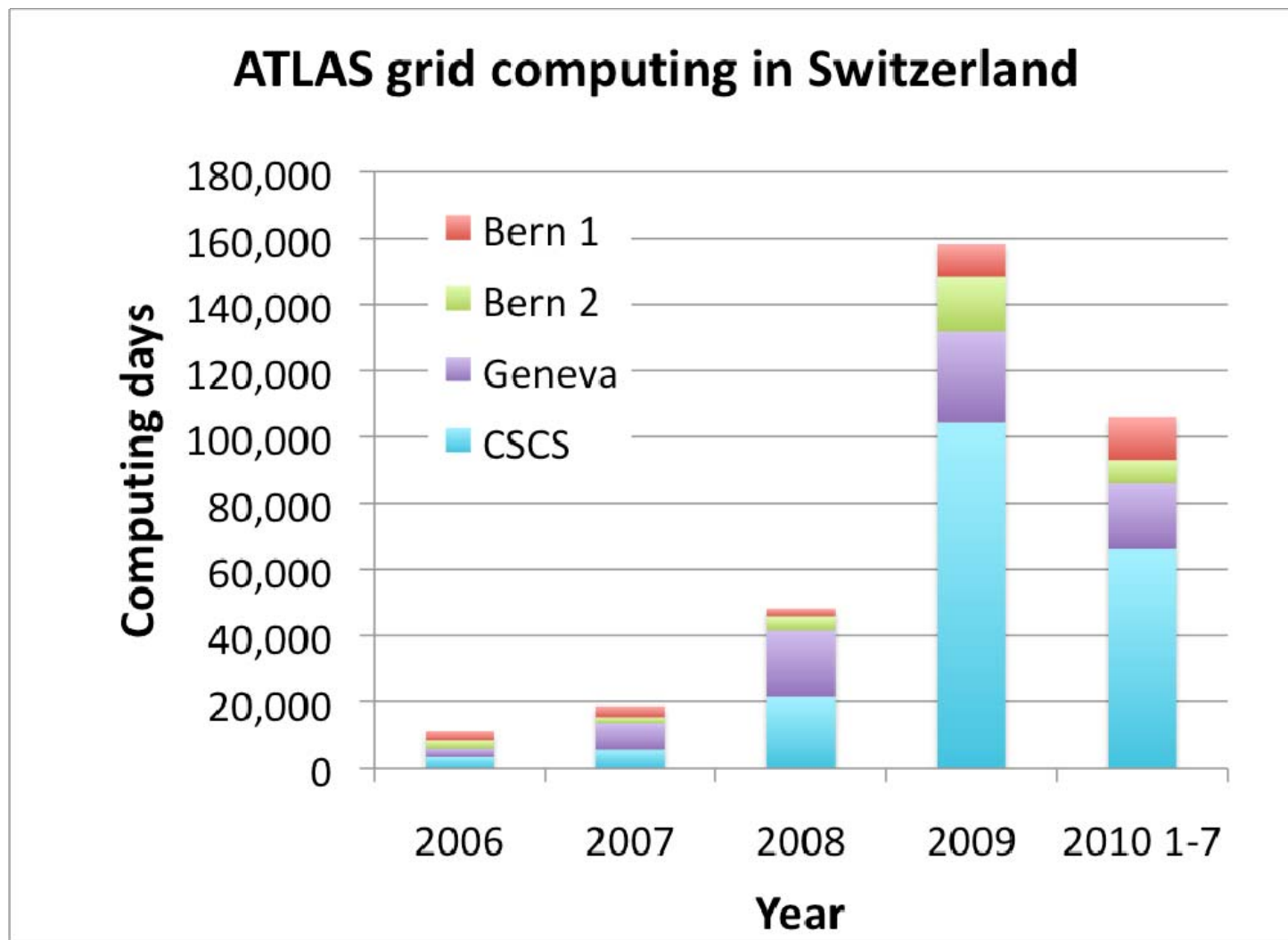
Local LHCb Software Installation

Future: DIRAC installation, LHCb Dirac Site

Hardware: CPU: ~500 HS06,   Disk: ~20 TB

Usage: Mainly code development and test jobs.

Most jobs are still going to the Grid.

Zurich Schrödinger Cluster:

ATLAS grid computing in Switzerland

some 0.8% of ATLAS computing in 2009/10