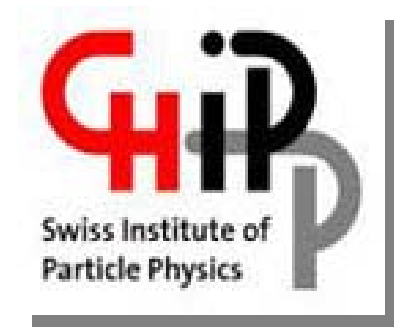
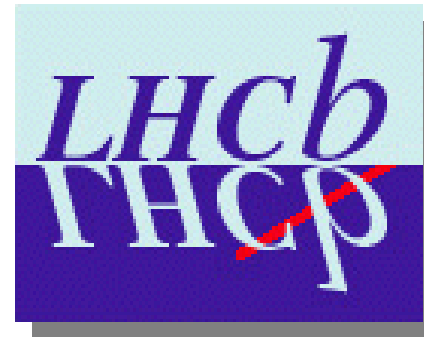
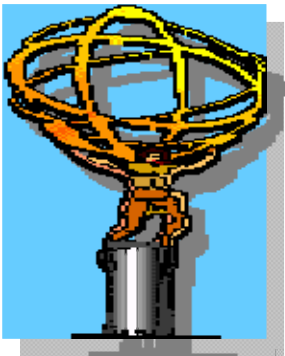


Report from the CHIPP Computing Board



Christoph Grab

Leysin, Sep.2, 2011

- Our Swiss Tier-2 Cluster located at CSCS
“operates in production mode” !
 - Overall reliability and availability above 95%
 - CPU and storage delivered the pledges to WLCG (fig)
- Cluster at CSCS moves to Lugano in May 2012
 - ➔ special measures needed (one time only)



Storage (left).

CPU

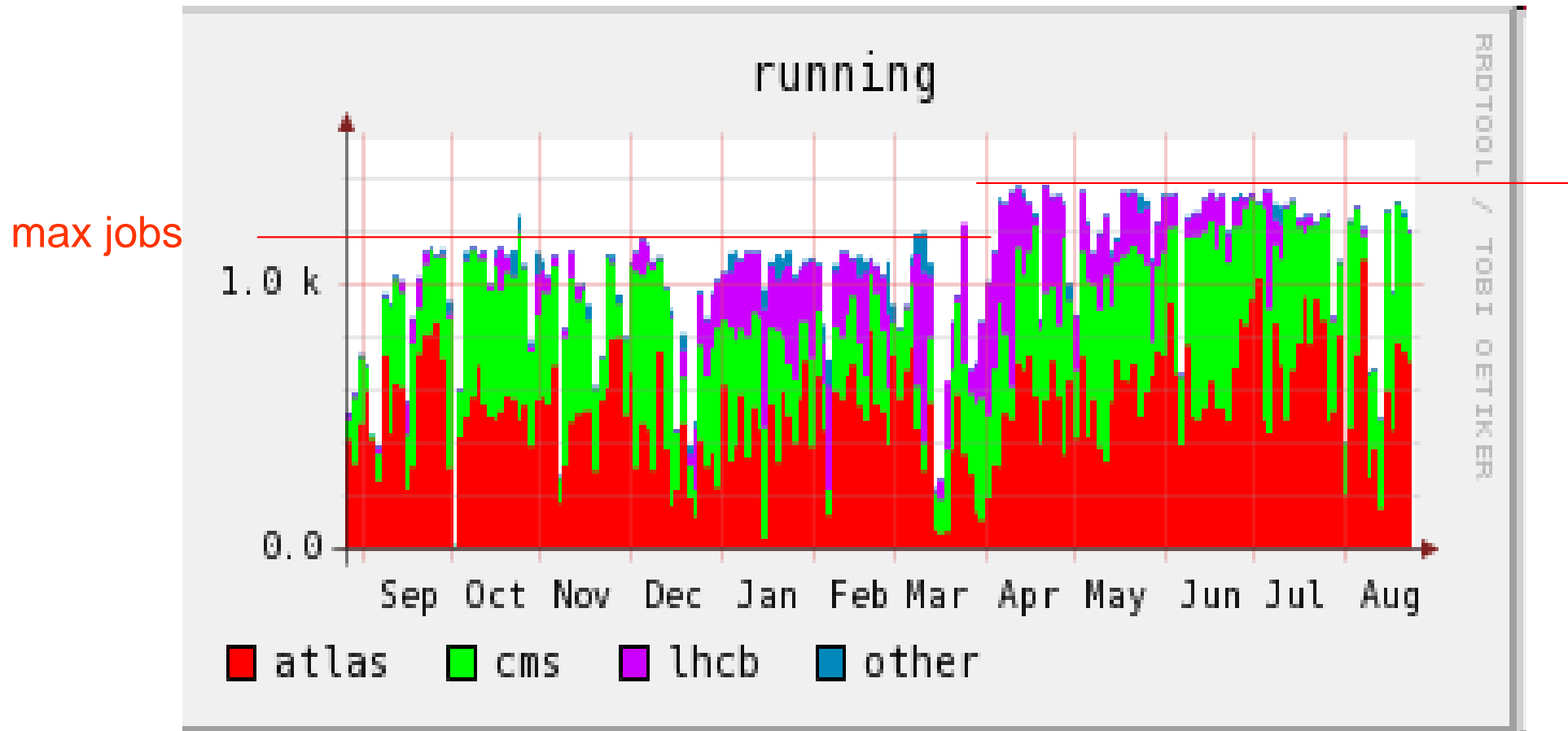
Pledges to WLCG and Capacity (8/2011)

Country	Federation	Physical CPU	Logical CPU	HEPSPEC06	CPU Pledge	Total Online Storage (GB)	Disk Pledge
Australia	University of Melbourne	41	164	1,927	4,000	523,748	400,000
Austria	Austrian Tier-2 Federation	58	340	2,845	5,057	458,616	420,000
Belgium	Belgian Tier-2 Federation	336	1,914	16,088	9,600	280,000	1,190,000
Sweden	SNIC Tier-2	1,249	1,249	7,419	7,870	1,714,124	920,000
Switzerland	CHIPP	212	1,390	13,465	13,550	1,056,008	975,000
Israel	IL-HEP Tier-2 Federation	192	1,040	80,582	4,000	4,389,668	560,000
Italy	INFN T2 Federation	2,330	10,411	94,379	84,000	5,020,523	5,900,000
Japan	ICEPP, Tokyo	288	1,152	16,531	12,000	1,199,997	1,000,000

We delivered what we pledged in 2011 !

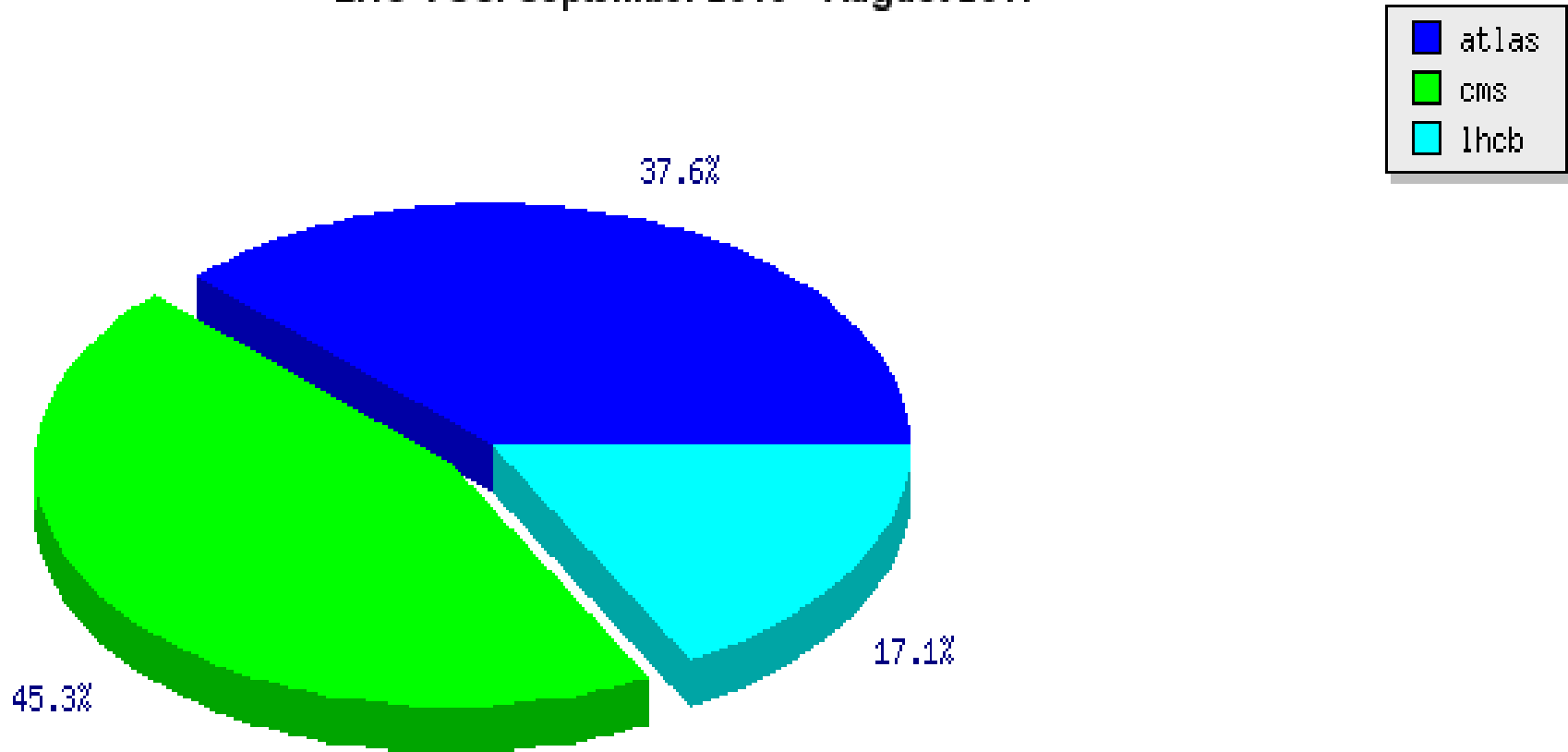
- All tier's are constantly monitored and listed e.g. at:
http://gstat-wlcg.cern.ch/apps/capacities/pledge_comparison/

number of batch jobs running on CE



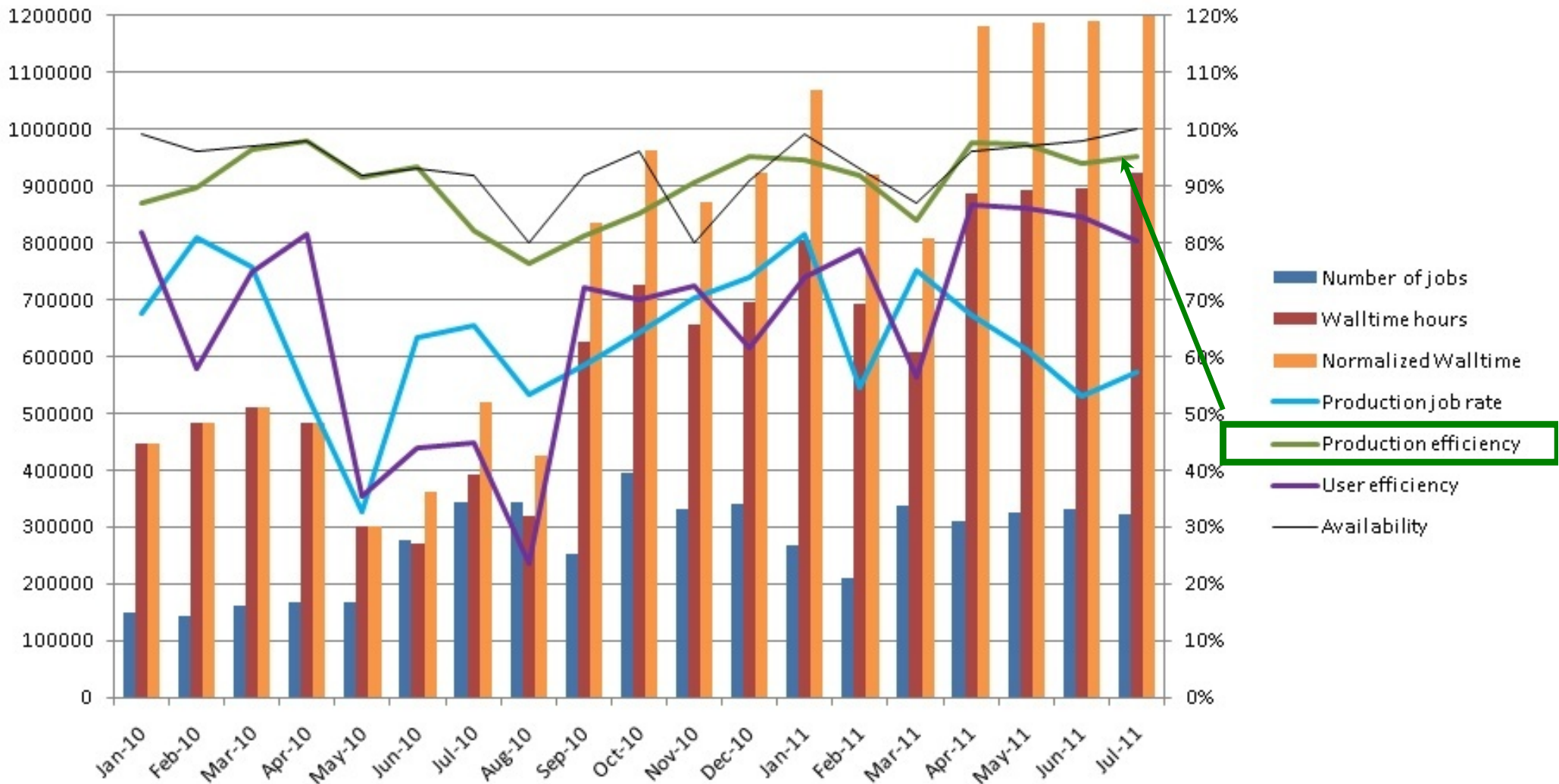
Reasonable exploitation and distribution of CPU to experiments

CSCS-LCG2 Normalised CPU time (kSI2K) per VO LHC VOs. September 2010 - August 2011



Reasonable distribution of CPU to experiments

Cluster Performance 2010-2011



Overall high efficiency and availability achieved !
 Note: resource increase due to upgrade ...

- Yearly planning of hardware upgrades and replacements at Tier-2 for the next phase, taking into account the experiments “Computing requirements documents” (aka “wishlist”).
- Planning for the move to Lugano (USI) in May 2012
- Note: all planning is discussed extensively within the CHIPP computing board (CCB) with representatives from CSCS.

Swiss Tier-2: Issues for Planning (1)

Resources	2011 HAVE	2012	2013	Comments 1 kSI2k ~ 180-250 HS06
Tier-2 CPU originally planned	13488 HS06	17400 as pledged	22100 ?	2013 Is not pledged yet
Tier-2 CPU wished	13488	20140	25000 ?	Expts wish increase in 2012
Tier-2 storage originally planned	1090	1090 as pledged	1400	2013 Is not pledged yet
Tier-2 disk wished	1090	2101	2500 ?	Expts wish increase in 2012

- "Appetite grows with the eating ..."

The "wished" numbers are from the experiment's contacts.

Cannot be met with present level of funding – Q: what is realistic ?

Detailed numbers per experiments and links given at :

<https://wiki.chipp.ch/twiki/bin/view/LCGTier2/ChippCBMeeting20110822>

<https://wiki.chipp.ch/twiki/bin/view/LCGTier2/ResourceOverviewT2T3>

Shortterm issues (2011-2012)

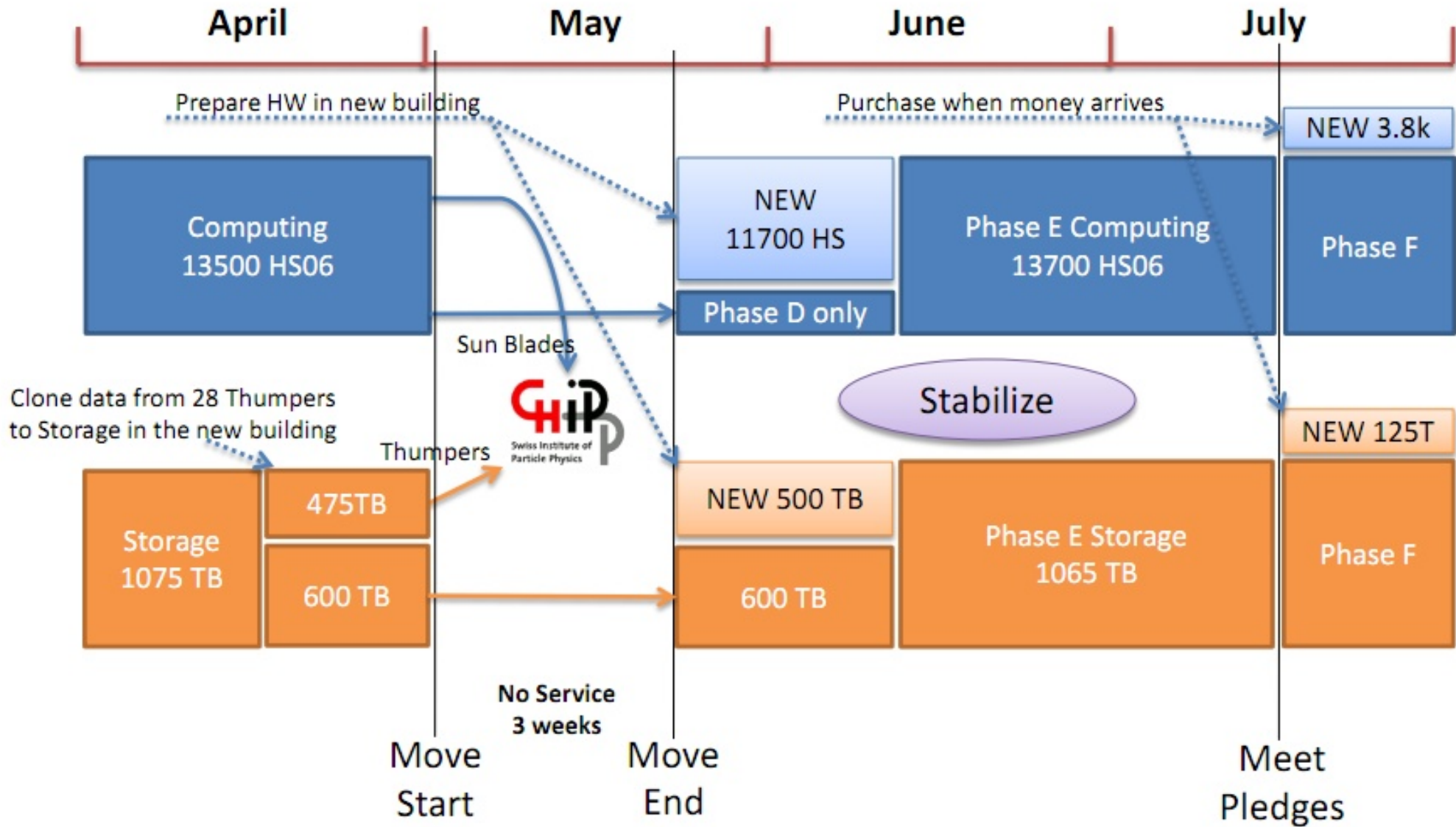
● Moving from Manno to Lugano (USI) in May 2012:

- ➔ Note: resources for 2012 from SNF/FORCE arrive AFTER the move
- ➔ this means, we fail to meet the WLCG pledges (due 1.4.2012) for a few months.
- ➔ solution: ask for delay at WLCG
- ➔ + compensate with one-time investments (early by Uni+ETH, and by SNF in 2012)
 - ➔ allows to meet 2012-pledges in July and next 2013-pledges in Apr.2013

● Plans on how to cope with the moving:

- ➔ consolidate older HW: keep 2 generations only at USI
 - ➔ pass older HW on to Swiss Tier-3s for re-usage
- ➔ replace CPU a bit earlier (footprint SUN M9 switch, heat density 25kW/rack !)
- ➔ invest into new CPUs directly at USI, run old SUN blades only at Manno
- ➔ replace older disks (thumpers) due to cooling (air < - > water) and space issues
- ➔ “Wishlist of experiments has grown → to satisfy ALL exceeds budget.
so try to satisfy the pledges for 2012 (is less than expt’s wishlist ...) – be realistic

Plan for Move and Phase E



- Additional investment sought before Jan 2012: ideally ~ **265 kCHF**
 - To replace old CPU and old storage
 - e.g. by contributions from Unis and ETH/EPFL

- Request investments from FORCE/SNF for 2012 :
 - Request in Sep. 2011 for 2012:
 - additional HW to meet the 2012 pledges ~ **365 kCHF**.
 - HW resources to be spent in Dec 2012 to meet the 2013 “pledges” (be AHEAD of pledge date) ~ ideally **550 kCHF**
 - **To satisfy all expt’s wishes: add another \geq 400 kCHF**
 - **Cannot fulfill ALL \rightarrow how much is affordable ?**

- Planned investments (FORCE/SNF) for longterm 2013 - 2016 :
 - Request for replacements and additions :
 - hardware for 500 – 700 kCHF / year
 - ONE FTE person to support Tier-2 operation at CSCS (at least)
 - recurring power costs, if ETH/CSCS does not continue MoU as is
 - additional costs for EGI contributions (if SBF does not fund).

Note: all numbers are today’s; will change with time, and possibly \$ vs CHF ??

**Include running experience to optimize the cluster usage
(conditions are changing) :**

- **CPU or storage : which should be invested in ?**

- Ratio of CPU/Storage is different in ATLAS (6), CMS (13) and LHCb (4200)!
- storage is more expensive

- **How to split** resource sizes between ATLAS, CMS and LHCb?

Up to now we used a model of 2:2:1 according to personell.

Question: is this still true ? Present user numbers at tier3 ~70:50:40.

- **Layout/configuration of nodes?**

- memory/core requirements is growing and uncertain
- ATLAS >> 3 GB/core, CMS (2-3 GB), LHCb (<=2 GB)

- **Usage of bandwidth for and size of scratch ?**

- Requirements different in ATLAS (>5 GB/s), CMS and LHCb (<=1 GB/s)
- How to dimension scratch space (very expensive) ?

- **Solution / optimization need to be found within CHIPP/CCB !**

some additional issues ...

- SWITCH has agreed to continue acting as leading institution, representing Switzerland towards EGI - under the condition that all resources required are provided by users.
thanks to SWITCH - and thanks for the CHIPP support.
- SwiNG continues as coordinating body.
- Support for international Grid operation (EGI)
 - Request for resources contributions was sent to SBF, asking for
 - 1.5 FTE for national/international support in Switzerland
 - operational contribution to EGI to be taken over by SBF
 - European membership fee (~88 kEuro / year)
 - no matching funds required from SWITCH
- Request for resource contributions for 2013-2016:
 - a request to SBF is being prepared by SWITCH + SWING
 - try to get similar coverage as 2012.

- **Network traffic:**

- ➔ **Still: present bandwidths are sufficient ...**
- ➔ Internally at CSCS now : 4 Gbps; will be 10 Gbps at Lugano
- ➔ externally now to SWITCH 10 Gbps.

- **routing within Switzerland via SWITCH :**
two redundant lines >10Gbps to CERN and Europe

- ➔ If needed, we can upgrade within months ...!

Still not a real issue ...

Swiss Tier-3 Efforts

- Tier-3 have become indispensable for final physics analysis !!
- **Large growth** seen over last year for all 3 experiments.
T3 took and will take the phased-out HW from tier-2
- **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: upgrade in Q4/2011



- ➔ **LHCb** :
 - EPFL : operates large local cluster
 - UZH uses local HEP + shares university resources



Summary: Swiss Tier Efforts (Q3/11)

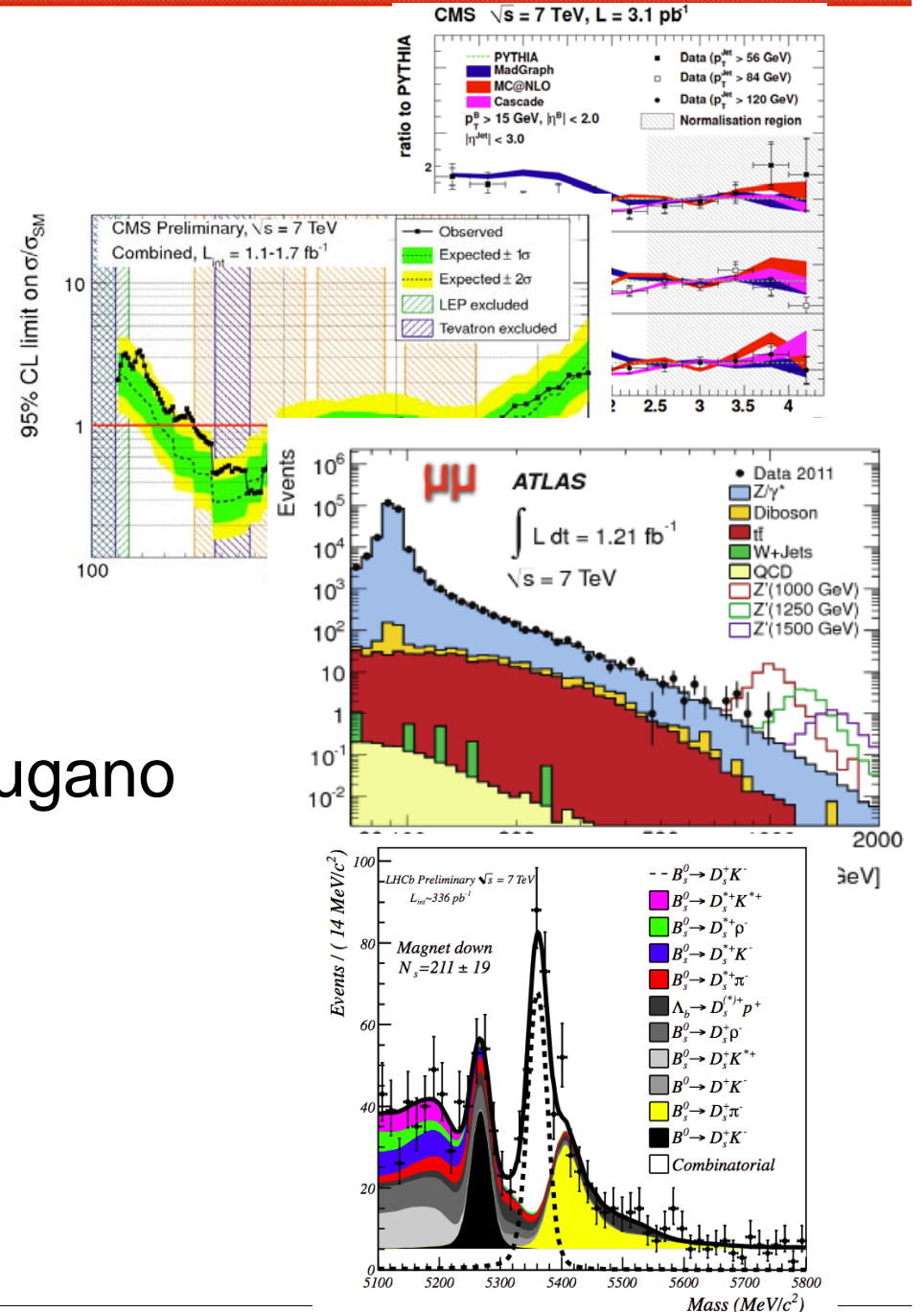
Site (#users)	Nr cores	CPU (HS)	Storage (TB)	Comments 1 kSI2k ~ 180-250 HS06
Swiss Tier-2 ✓	1392	~ 13488 HS	1110	1152_HT+240
ATLAS BE (11) ✓ GE (~57) ✓	500+300sh 428	~5000 HS 6100 HS	200 205	BE: standing Atlas production; GE: direct line to CERN. TB-upgrade planned
CMS ETHZ, ✓ PSI, UZH (48)	216	~3271 HS	270	GRID SE + UI :direct GRID access.
LHCb EPFL (30) ✓ UZH (12) ✓	464 48+96sh	~13900 HS 835+847sh	24 75	EPFL is DIRAC site; UZH:MC production; shared
Total Tier-3		~36000 HS06	775	

- Tier-3 capacities : exceed size in CPU of Tier-2

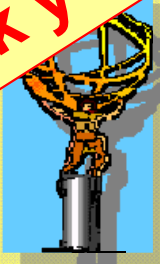
<https://wiki.chipp.ch/twiki/bin/view/LCGTier2/ResourceOverviewT2T3>

- Cluster runs well and reliably !
- We contribute(d) to physics exploitation !
- We deliver as pledged to WLCG !

- 2012 will be different → move to Lugano
- Need additional funding for 2012
- Need to optimize the cluster setup+operation according to gained experience ..



Thank you ...



Coordinates the Tier-2 activities
representatives of all institutions and experiments
included also the tier-3 expert

S.Gadomski, A.Clark (UNI Ge)

S.Haug, H.P.Beck, G.Sciacca (UNI Bern)

C.Grab (ETHZ) **chair CCB**

D.Feichtinger (PSI) **vice-chair CCB**

L.Sala (ETHZ), F.Martinelli

R.Bernet (UNIZH)

Y.Amhis (EPFL)

P.Fernandez, M.De Lorenzi (CSCS)

Additional Investment in 2011 (opt)

- for HW to be spent in the NEW building, in “Phase E”
 - ➔ replace SunBlades compute nodes from PhaseC (11520 HS06), because:
 - depend on M9 network switch with footprint incompatible to new building
 - heat density is too high for new cooling capacity (25 vs 10 KW of cold)
 - blade’s rack size has non-standard measures new building.
 - ➔ Replacement for CPU will cost 345 kCHF, providing 11776 HS06.
 - ➔ Need replace all Thumpers from PhaseB (504 TB) because the hardware is very old (>4 years) and it’s actual capacity is already degraded.
 - ➔ Replacement for storage will cost 360 kCHF, providing 500 TB.
 - ➔ Requires three new IB network switches, new GPFS licenses, and electrical installation, that sum a total of 80 kCHF (30+30+20)
 - ➔ Need to pay 30 kCHF to protect the compute nodes in the new building with UPS. Only Service Nodes are covered by default.
- ➔ This makes ~ 815 kCHF to spend before January 2012 for PhaseE, have new HW ready in new building before moving (+copy data ..).
- Have already 550 kCHF from PhaseD, ie. need additional 265 kCHF.
Thus, cluster size will remain at 13700 HS06 of CPU and 1100 TB of Storage.

Details on Request for 2012 (opt)

- Invest in **HW to meet our 2012 pledges** (to be spent in July) → **365 kCHF**
 - 3800 HS06 of CPU → additional 480 cores for ~ 105 kCHF.
 - 125 TB of Central Storage (added to dCache) for ~ 85 kCHF.
 - 96 TB of Scratch, to keep up with bandwidth of new CPU expansion for ~85 kCHF
 - 3 new Service Nodes, for CE expansion, network switch, GPFS licenses and electrical installationc (45+10+15+20) for ~90 kCHF
 - Provides a total of
 - 17500 HS06 of Computing. (pledged: 17400, wished: 20140)
 - 1225 TB of Storage (pledged: 1090, wished: 2100)
- Invest in **HW to meet our 2013 pledges before April 2013** → **550 kCHF**
 - 3500 HS06 in new Compute for ~ 95 kCHF.
 - 500 TB in new Central Storage for ~ 330 k CHF.
 - 96 TB of new Scratch (keep up with same bandwidth from nodes) ~ 80 kCHF.
 - Installation, GPFS licenses and network switches should cost 45 kCHF.
 - Provides a total of 21000 HS06; 1750 TB of Storage.
 - satisfy all wishes requires additional ≥ 400 kCHF.
- Cannot afford ALL → need optimisation in 2012/2013 – affordable?
→ **Ask for ~900 kCHF in Sep.2011 request to be spent in year 2012**



Storage (left) with cooling cage.

CPU + SUN infiniband switch (right)

- SUN M9 Switch for Infiniband connections



Tier-2 LCG Services – Phase D

