The Grid in/for CALICE

Roman Pöschl
LAL Orsay

- – Introductory remarks
- – The vo calice
- – Activities in the vo ilc
- – Summary, Outlook and Input for Discussion

ILC Software Workshop LAL Orsay – May 2007

Introductory Remarks

Why the using Grid??

Neither ILC nor Calice have an 'experimental center' like CERN, DESY etc. and maybe will never have

World wide distributed R&D effort requires distributed computing

- Easy sharing of data by common data storage accessible by everyone from everywhere

- Exploiting the Grid allows for quick data processing, e.g. Several reconstruction terations for calice testbeam data

- Large simulation effort to come for the ILC requires large computing ressources
  Again no experimental center, the potential experimental centers like DESY and CERN et al. have identified the Grid as computing platform
 General strategic decision by HEP community and science politics to invest in Grid computing

- Exploring the Grid can be regarded as an engineering/R&D effort for the ILC just as hardware development or simulation studies (which in turn demand significant computing power)
  Software (and computing) infrastructure is part of the ILC Project !!!!
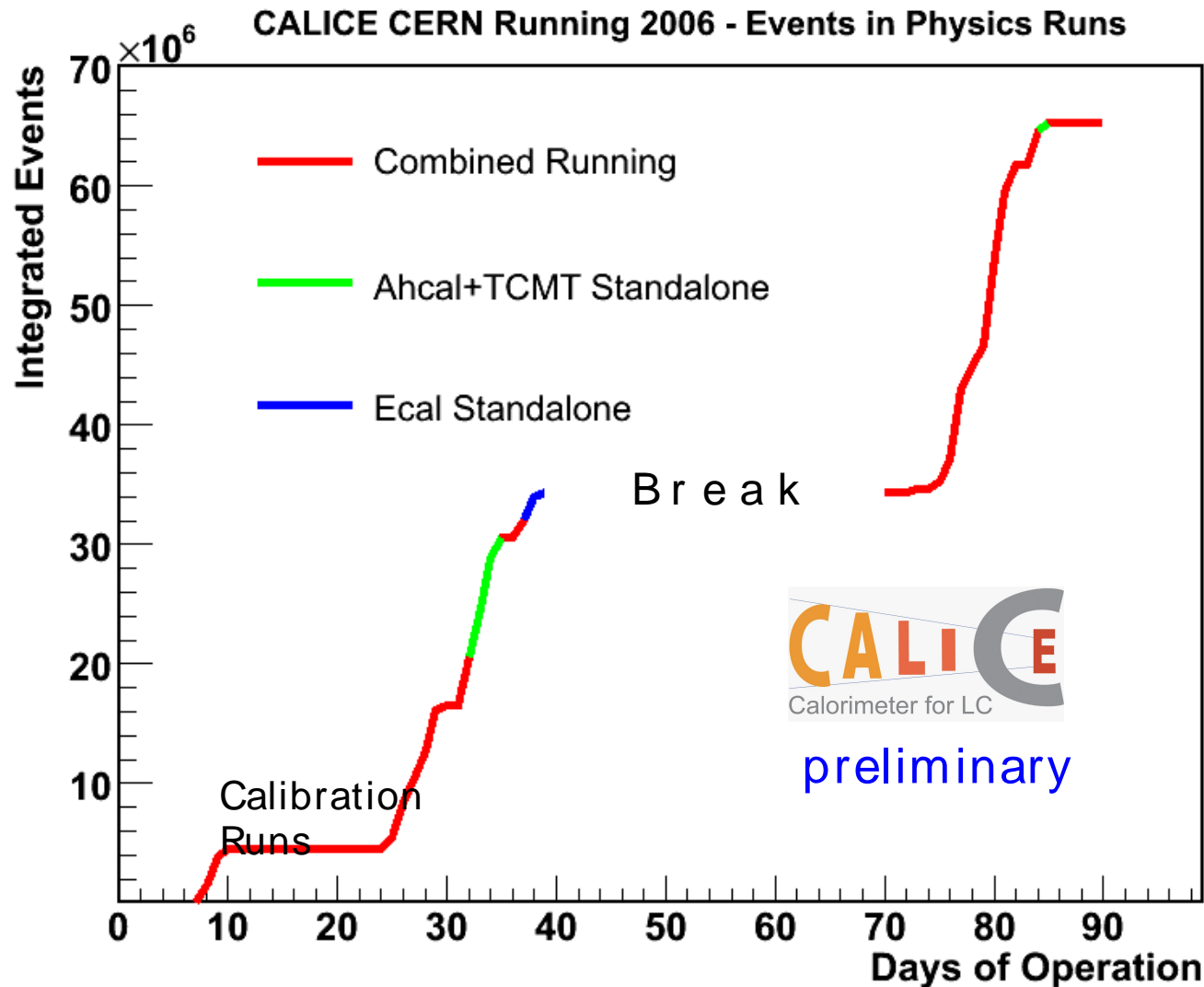
The Calice Mission

Final goal:

A highly granular calorimeter optimised for the Particle Flow measurement of multi-jets final state at the International Linear Collider

Imaging calorimeter

Scint. Strips-Fe TCMT

Si-W ECAL

Scint. Tiles-Fe AHCAL
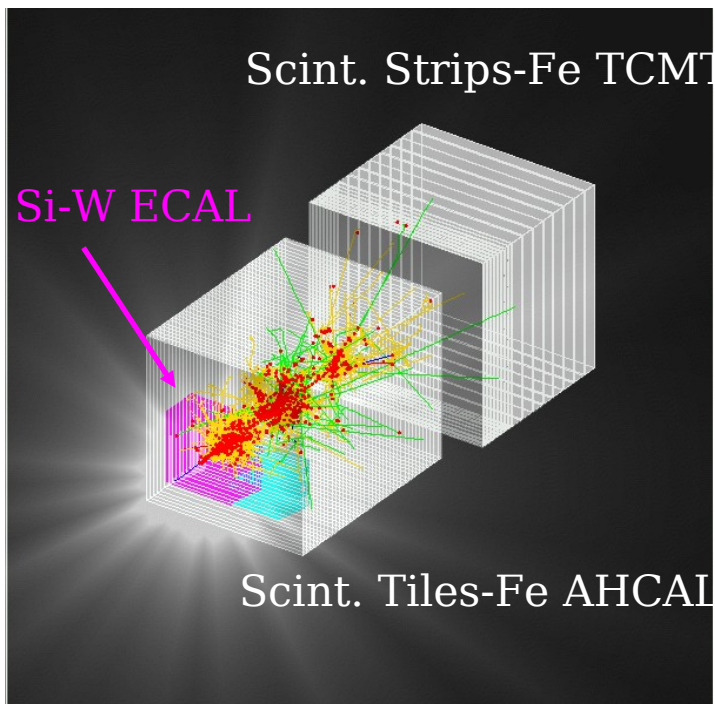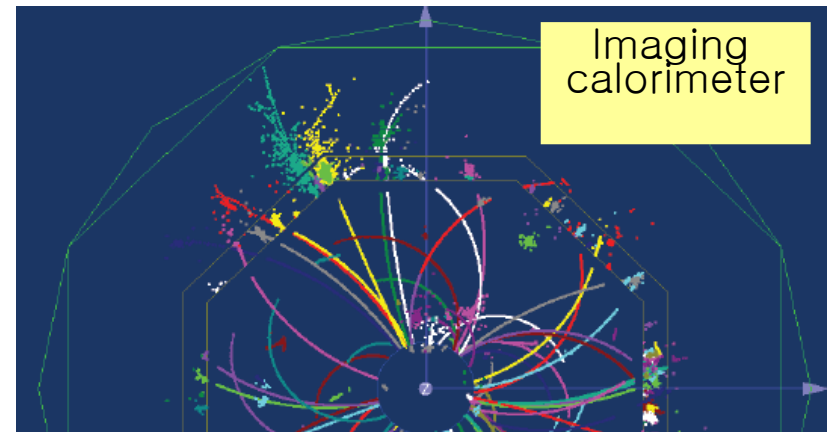
Intermediate task:

Build prototype calorimeters to
• Establish the technology
• Collect hadronic showers data with unprecedented granularity to
  - tune clustering algorithms
  - validate existing MC models

3

E. Garutti ECFA Valencia

CALICE – CERN Data taking 2006

CALICE CERN Running 2006 - Events in Physics Runs

- Combined Running
- Ahcal+TCMT Standalone
- Ecal Standalone

B r e a k

Calibration
Runs

CALICE Calorimeter for LC

preliminary

~65 Mio Events
in 'Physics' Runs
+
O(35 Mio). Muon
Calibration Events)

Efficient and fast
way of data distribution
and processing ?

The virtual organisation – vo calice

Hosted by DESY:

Page for registrations https://grid-voms.desy.de:8443/voms/calice/

Virtual Organization Membership Service

The calice VO

Administration » Users » List of users

ADMINISTRATION
USERS
  LIST OF USERS
  SEARCH FOR USERS
  CREATE A NEW VO USER
GROUPS
  LIST OF GROUPS
  SEARCH FOR GROUPS
  CREATE A NEW GROUP
ROLES
  LIST THE ROLES
  SEARCH FOR ROLES
  ADD A NEW ROLE
GLOBAL ACL

There are 17 users in /calice :

DN		
/C=UK/O=eScience/OU=Birmingham/L=ParticlePhysics/CN=nigel watson	edit	remove
/C=UK/O=eScience/OU=Cambridge/L=UCS/CN=david ward	edit	remove
/O=GermanGrid/OU=DESY/CN=Roman Poeschl	edit	remove
/C=UK/O=eScience/OU=Imperial/L=Physics/CN=gidon moont	edit	remove
/C=UK/O=eScience/OU=Imperial/L=Physics/CN=anne-marie magnan	edit	remove
/DC=org/DC=doegrids/OU=People/CN=Guilherme Lima 269451	edit	remove
/C=UK/O=eScience/OU=RoyalHollowayLondon/L=Physics/CN=pasquale-fabrizio salvatore	edit	remove
/C=UK/O=eScience/OU=RoyalHollowayLondon/L=Physics/CN=michele faucci giannelli	edit	remove
/O=GRID-FR/C=FR/O=CNRS/OU=LLR/CN=Goetz Gaycken	edit	remove
/DC=cz/DC=cesnet-ca/O=Institute of Physics of the Academy of Sciences of the CR/CN=Petr Mikes	edit	remove
/DC=cz/DC=cesnet-ca/O=Institute of Physics of the Academy of Sciences of the CR/CN=Jaroslav Zalesak	edit	remove
/O=GermanGrid/OU=DESY/CN=Vladislav Balagura	edit	remove
/C=UK/O=eScience/OU=Manchester/L=HEP/CN=david bailey	edit	remove
/O=GRID-FR/C=FR/O=CNRS/OU=LPSC/CN=Jean-Yves Hostachy	edit	remove
/O=GermanGrid/OU=DESY/CN=Marius Groll	edit	remove
/O=GermanGrid/OU=DESY/CN=Erika Garutti	edit	remove
/O=GRID-FR/C=FR/O=CNRS/OU=LPSC/CN=Laurent Morin	edit	remove

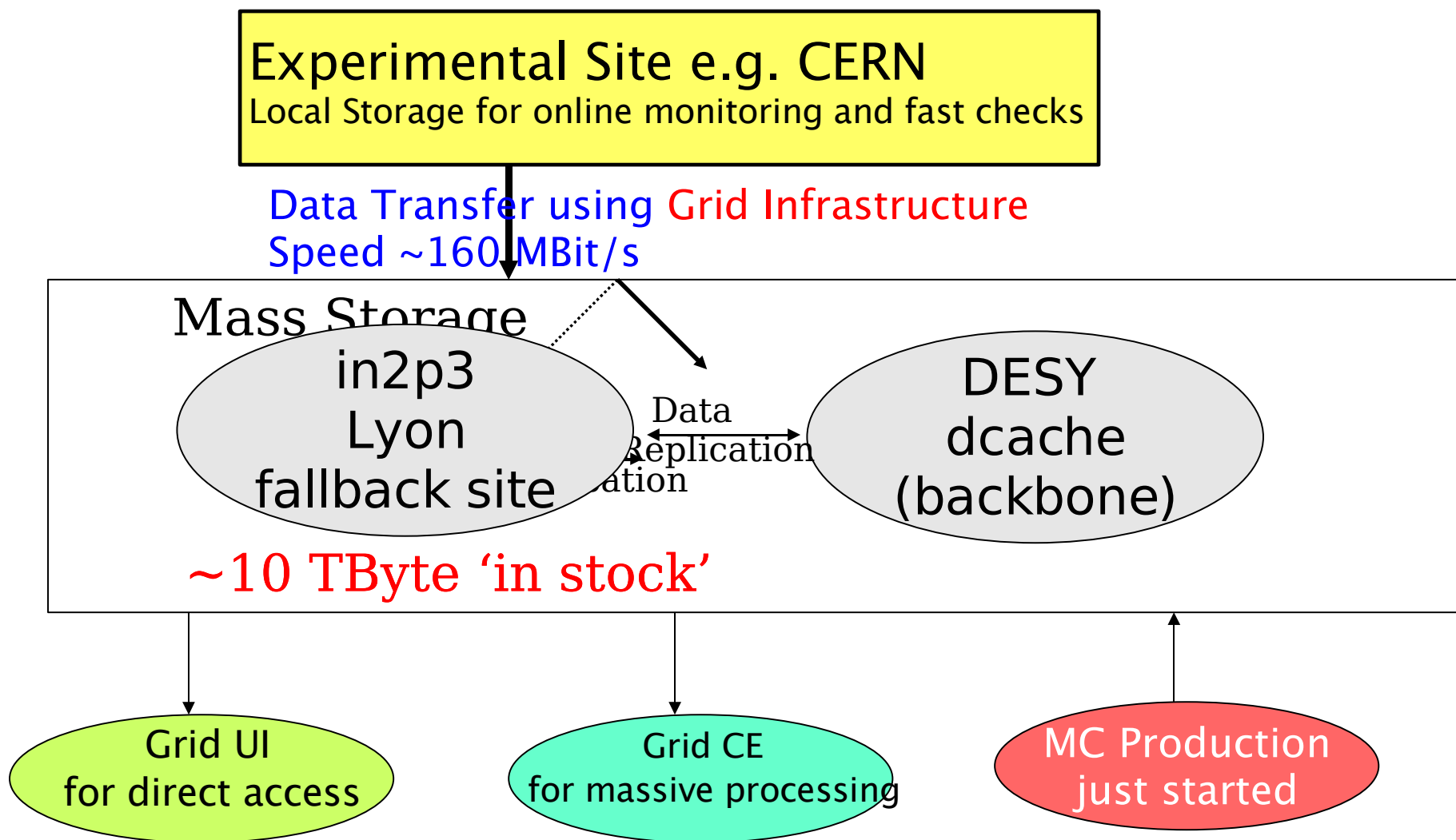You can click on the DN for more details.

VOMS Admin 1.2.16
Release 1
Copyright © 2005 CERN, ELTE

You are logged in as "/O=GermanGrid/OU=DESY/CN=Roman Poeschl"
certified by "/C=DE/O=GermanGrid/CN=GridKa-CA".

42 members ... and counting

VO Manager: R.P., Deputy A. Gellrich DESY

Data Handling and Processing

Experimental Site e.g. CERN
Local Storage for online monitoring and fast checks

Data Transfer using Grid Infrastructure
Speed ~160 MBit/s

Mass Storage

in2p3 Lyon fallback site

Data Replication ation

DESY dcache (backbone)

~10 TByte 'in stock'

Grid UI for direct access

Grid CE for massive processing

MC Production just started

Data available to whole collaboration ~20 Min. – 1h after run end
Data access independant of experimental site
Grid is the only 'environment' where all data are available

Support for Calice

Supported by: DESY Hamburg   Hosting, Computing and Storage
           LAL   Computing and Storage
           LLR   Computing and Storage
           DESY Zeuthen   Computing and Storage
           Imperial College   Computing and Storage
           Birmingham   Computing and Storage (not yet tested – at least by me)
           cc in2p3 Lyon   Computing and Storage
           Cambridge   Computing and Storage
           Institute of Physics Computing and Storage
           Prague   (in preparation)
           University College Computing and Storage
           KEK   Computing and Storage
           Manchester   Computing and Storage
           CIEMAT Madrid   Computing and Storage
           Fermilab   Computing and Storage
                              Exploit started between Fermilab and NIU Colleagues
           Univ. Regina   Offer Received

These sites provides ~4000 CPU (not only for calice)
We have ~100 Tbyte Storage at our disposal

- Sites in red are foreseen for a complete storage of calice data

- Most of the sites have been involved in recent data and MC processing

Infrastructure and data availability –
Accessing/Handling the data using grid tools?

Calice data are registered on the Grid

Using the LCG software together with LFC file catalogue
Organized in a unix-like directory structure

e.g. lfc-ls /grid/calice/tb-cern/native/dat

Centralized access to members of virtual organisation 'calice'
e.g. lcg-cp –vo calice lfn:/grid/calice/tb-desy/native/dat/RunXXXXXX.nnn.bin file:<myfile>

Don't need account at particular institute but only grid certificate
Avoids (excessive) duplication of data

The more users the faster remaining pitfalls can be ironed out
and Grid experts can serve our needs !!!

Computing Ressources for Calice for 2007 – CPU Power

Will/Have to use grid for MC production (bulk part of CPU demand)

Estimation:

~45 Mio. Physics Events on tape
Need 'at least' the same amount in MC

'Standard Candle': 1000 20 GeV e– took ~2h on standard Grid CPU (AMD Opteron 2.4GHz)
=> 1Mio. will take 2000 h = 83 days on one CPU

Production of 45 Mio. Events: ~3750 days = 5000 kSi2kd
CPU unit in grid is kSi2k where 1 Opteron is 1.4 kSi2k

May need
– several iterations, 'full productions' with various GEANT physics list, fine angular grid
– reprocessing of 2006 data plus processing of 2007 data (conversion + reconstruction)
  negligible w.r.t to MC production
– lot of freedom for user analysis jobs + private/test simulation

Looks as if to demand 10x5000 kSi2kd (worldwide) seems to be justified
Comparison: ATLAS has used ~35000 kSi2kd in 2006 on german ressources
~70000 kSi2kd in 2006 on french ressources

In general estimation is difficult due to lack of experience

Computing Ressources for Calice for 2007 – Storage

2005/2006 Data ~ 10 TBytes of data (native, lcio converted, reconstructe
Simulation of 2006 data -> ~5 Tbytes
2007 Data will lead also to ~10 Tbytes

End of 2007 we will have ~30-35 Tbytes of data in stock
Lot of it is Hcal calibration data – i.e. Not 'physics' data

Backbone will remain desy mass storage
All data will/should be available there

Each 'calice country' should provide ~10 Tbytes for calice
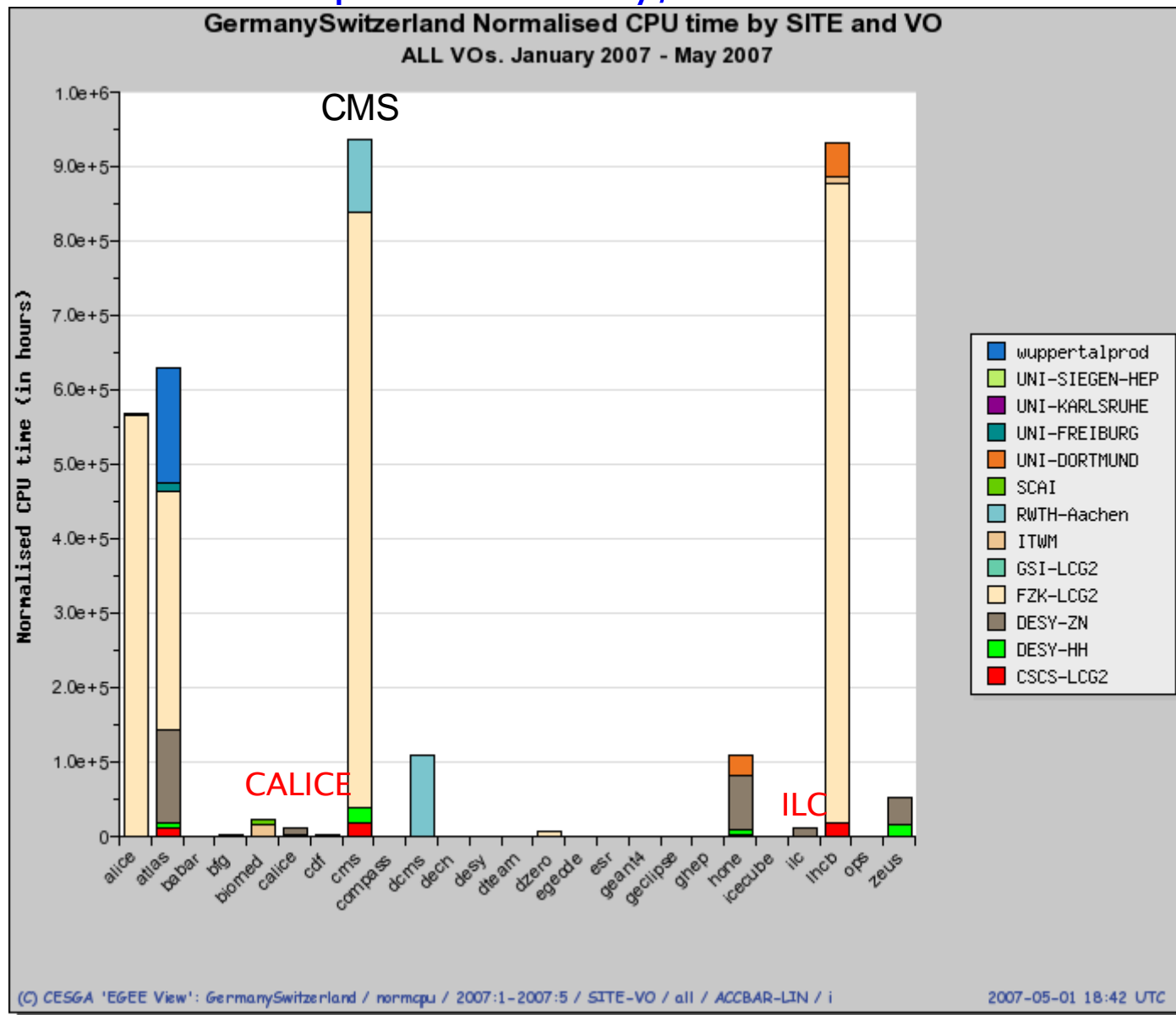
Replication of important data
Reduced dependency on desy availability
Reduction of network traffic
Storage for user analysis output

Some Statistics – EGEE Accounting

CPU Consumption Germany/Switzerland between 01/07 and 05/07

CMS 10^6 kSi2kh
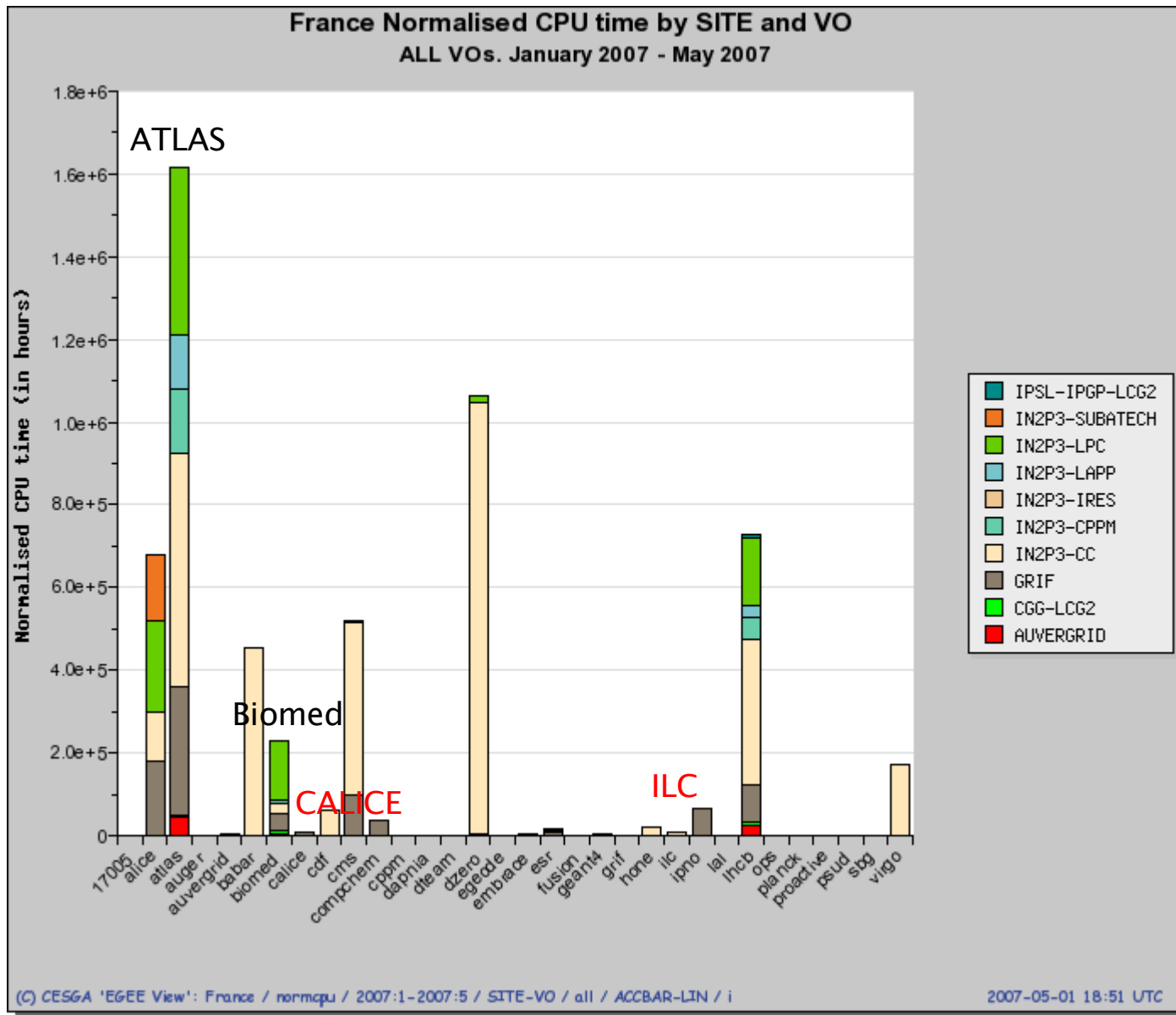
ILC 10^4 kSi2kh

Calice 10^4 kSi2kh

1 Opteron hour ~ 1.4 kSi2kh, Example 8x10^5 kSi2kh means that an Opteron 2.4 Ghz calculates 580000 hours without interruption

Source: http://www3.egee.cesga.es/gridsite/accounting

Some Statistics – EGEE Accounting

CPU Consumption in France between 10/06 and 01/07
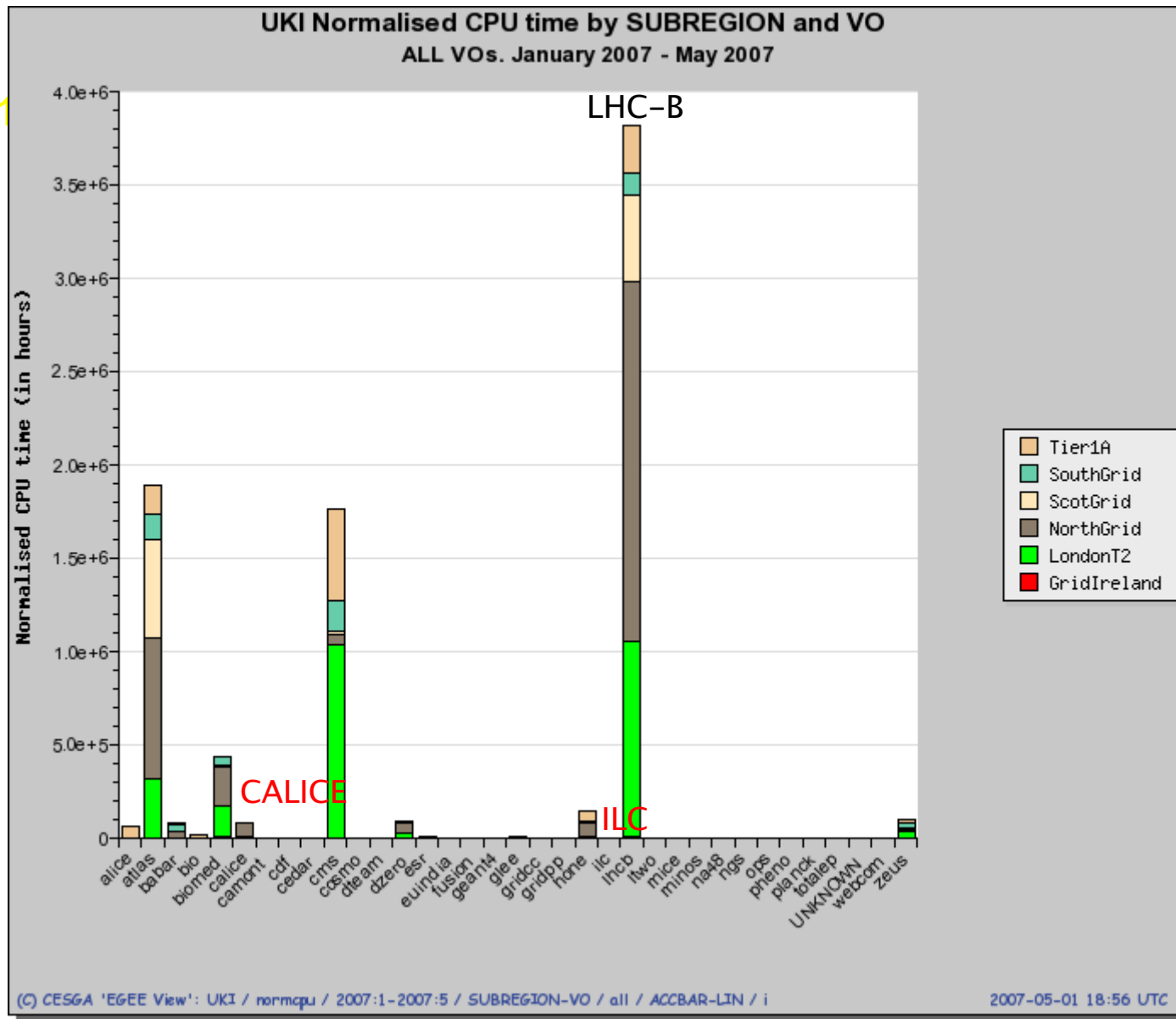

ATLAS 1.6×10^{6} kSi2kh

ILC 10^{4} kSi2kh

Calice 10^{4} kSi2kh

Some Statistics – EGEE Accounting
CPU Consumption in UK between 01/07 and 05/07

LHC–B 4×10^{6} kSi2kh

ILC 4.5×10^{3} kSi2kh

Calice 8.3×10^{4} kSi2kh

Conclusion

– Grid activities in ILC business have only started
  Although I don't know about all ILC grid activities
  Efforts in Noth America !?

– Calice uses Grid tools systematically
  Grid tools seem to be suited for Calice needs

– Important that we show our presence to be able to
  allocate ressources at an early stage

– Have to get trained to use the 'new' tools to be ready when
  they are really needed

– Need to identity needs for ILC with next 1½ years
  Structured processing as for LHC necessary, if yes when?