Report from HEPCCC

Tobias Haas
XIV HTASC
12 June, 2003
INFN-Pisa

HEPCCC Agenda
(4 April 2003 at CERN)

- Guy Wormser: Computing for QCD Calculations
- Nobu Katayama: Computing for BELLE ☞
- Stephan Paul: Computing for COMPASS ☞
- Fabricio Gagliardi: Outcome of the EDG review ☞
- Les Robertson: LCG progress report ☞
- Irwin Gaines: Recent DOE/NSF initiatives on partnerships for global infrastructure
- Tobias Haas: HTASC report with emphasis on "Traveling Physicist problem"

Traveling Physicist Discussion

- Presented HTASC findings/recommendations to HEPCCC
- Well received
- Asked to make brief summary of findings and recommendations
- Circulated within HTASC
- Sent to HEPCCC

HEPCCC/HTASC Discussion on

"Support for the Traveling Physicist"

Summary

An increasing number of physicists travel regularly between their home institutions and various other laboratories or institutes. They find it difficult or cumbersome to gain access to information technology (IT) resources while being guests at host institutions or while being on the road. They use different pieces of methods, tools, software, hardware, tricks and hints to accomplish what they need to do.

On the request of HEPCCC, during its XXIII meeting on 13/14 March at CERN, HTASC held an extended discussion on this subject in order to review the situation. The aim of the discussion was to

- understand the problems people are encountering,
- review policies in place at different sites,
- And collect solutions that are being used.

As a basis for the discussion, presentations were heard from institutions in Japan (KEK), the US (SLAC) and Europe (CERN, KFK, LIP). It was found that the problems encountered in different parts of the world are very similar. However, the solutions pursued and the policies imposed at the different hosting institutions are quite different. In particular, the major problem appears to be the difficulty to obtain information on the particular local situation. For these reasons HEPCCC/HTASC decided to issue a set of recommendations both to institutions regularly hosting traveling scientists and to the home institutes. These recommendations are of a general nature and are intended as guidelines rather than specific suggestions. The problem of gaining access to essential IT resources while traveling would be very much reduced if a large number of institutions followed these recommendations.

Recommendations for institutions hosting traveling scientists:

1. **Wireless Network:** Wireless networking should be available in publicly visited places, in particular in seminar rooms.
2. **DHCP:** IP addresses should be assigned via the DHCP protocol.
3. **Power Outlets:** Seminar rooms should be equipped with a reasonable number of power outlets.
4. **Print Services:** Print services should be available to visitors
5. **Public Workstations:** Even though many people travel with their own personal devices a limited number of public workstations should be provided for those who do not carry a personal device.
6. **Documentation:** Information on what IT services are available to guests and how to use them should be provided on the web in an easily accessible location. HTASC provides a template web page for this purpose.

ımendations for the home institutions of traveling scientists:

1. **Authenticated SMTP service:** A secure authenticated SMTP service should be provided so that mail can be relayed even if the originator is not at the home institution.
2. **WEBMAIL:** A webmail interface to the mail service should be provided.
3. **VPN:** A VPN service should be provided to tunnel insecure protocols.

al:

1. **Firewall:** Firewalls should be opened for secure protocols to a well defined set of hosts (ssh, VPN, CITRIX, afs)
2. **Redundancy:** Even though more general technical solutions may make other less general solutions obsolete one should keep in mind that simpler solutions can be used when more advanced solutions fail.

iHEPCCC Developments

- iHEPCCC was discussed in Oct. 02 and Feb. 03 meetings of ECFA
- Good level of agreement on the matter with some discussion on membership:
  - ⇒ several changes to proposed charter:
  - ⇒ See http://tilde-djacobs.home.cern.ch/~Djacobs/Hepccw3/Attachments/030404_ihepccc_charter_v3.doc
- More concrete role and actions
- Interaction with ICFA/SCIC
- Rationalized membership (still large: 25)
- Idea of a “bureau” that meets more frequently
- Try to co-locate meetings with other related meetings.

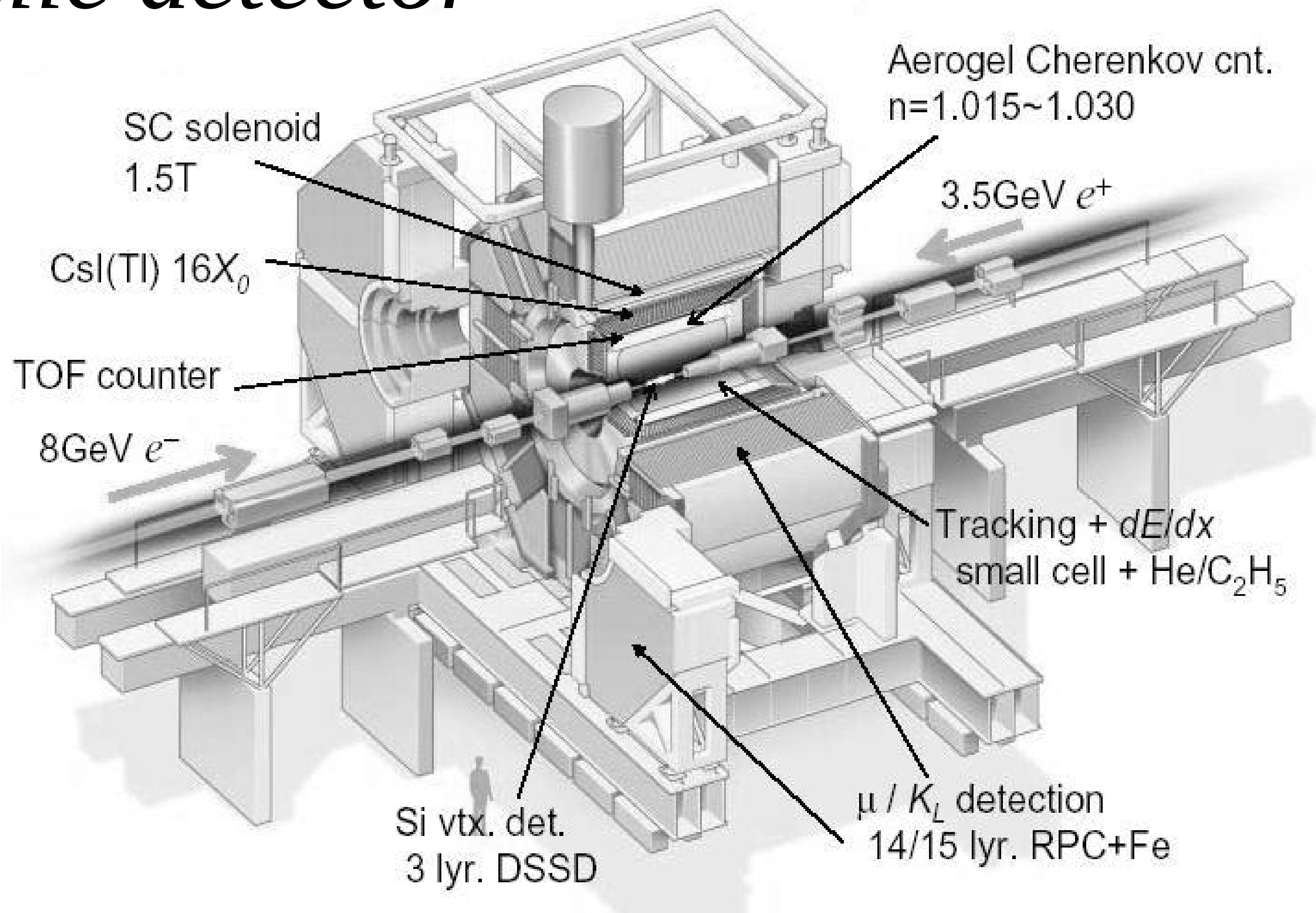
Computing for Belle

HEP-CCC@CERN
April 4, 2003
Nobu Katayama
KEK

Outline

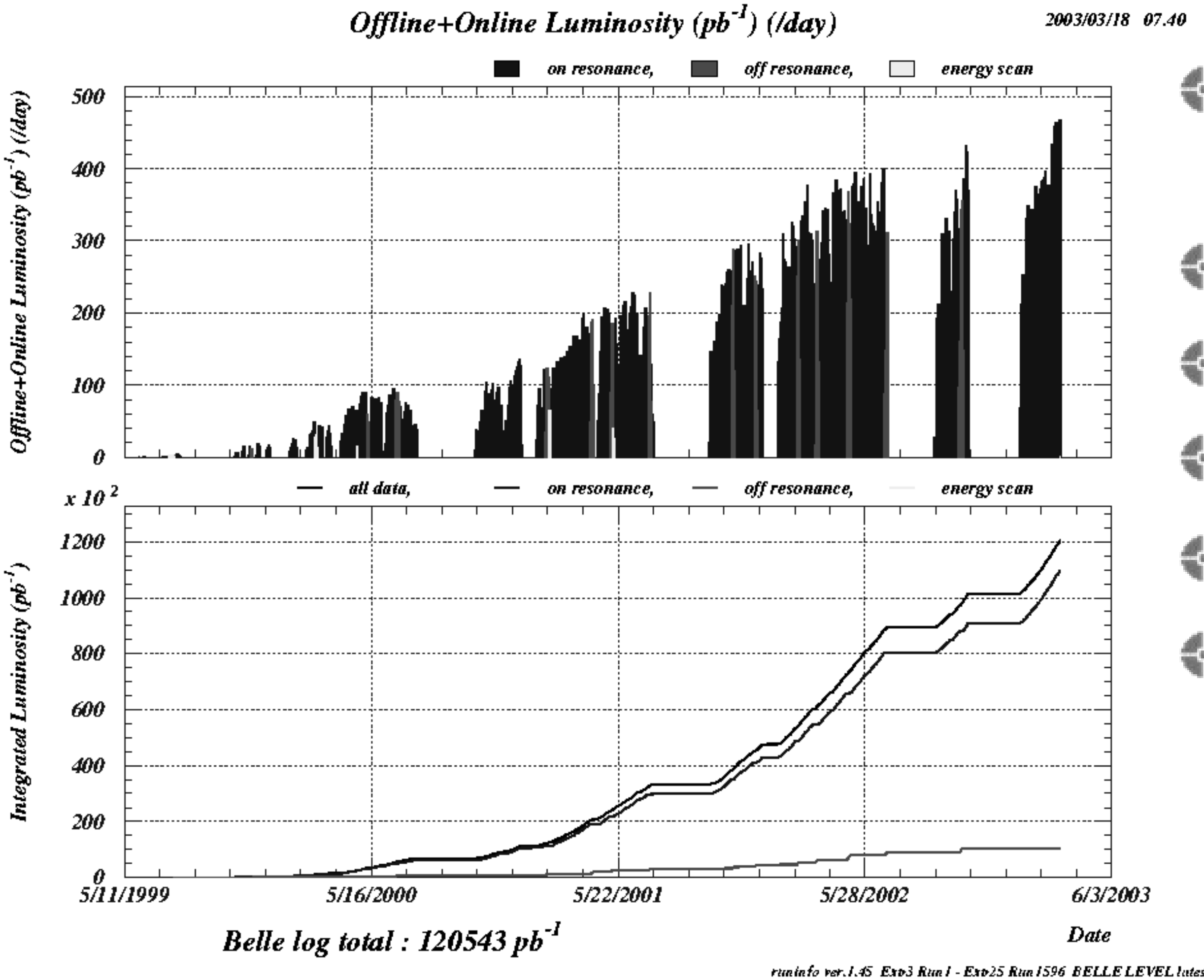
- Belle in general
- Software
- Computing
- Production
- Physics data analysis
- Issues
- Super KEKB

Belle detector



Integrated luminosity

126fb^{-1} as of today



- May 1999: first collision
- July 2001: 30fb^{-1}
- Oct. 2002: 100fb^{-1}
- July 2003: 150fb^{-1}
- July 2004: 300fb^{-1}
- July 2005: 500fb^{-1}

ß

To Super KEKB

1~10ab^{-1}/year!

Software

Core Software

- OS/C++
  - Solaris 7 on sparc and RedHat 6/7 on PCs
  - gcc 2.95.3/3.0.4/3.2.2 (code compiles with SunCC)
- No commercial software except for LSF/HSM
  - QQ, EvtGen, GEANT3, CERNLIB (2001/2002), CLHEP(~1.5), postgres 7
- Legacy FORTRAN code
  - GSIM/GEANT3/ and old calibration/reconstruction code)
- I/O:home-grown serial IO package + zlib
  - The only data format for all stages (from DAQ to final user analysis skim files)

Reconstruction software

- 30~40 people have contributed in the last several years
  - Total just 0.5 million lines of code (in .cc, without counting comments)
  - 200K line of FORTRAN code (in .F, without counting comments)
- For many parts of reconstruction software, we only have one package. Very little competition
  - Good and bad
- Identify weak points and ask someone to improve them
  - Mostly organized within the sub detector groups
  - Physics motivated, though
- Systematic effort to improve tracking software but very slow progress
  - For example, 1 year to get down tracking systematic error from 2% to less than 1%
  - Small Z bias for either forward/backward or positive/negative charged tracks
    - ⇒When the problem is solved we will reprocess all data again

Analysis software

- Several ~ tens of people have contributed
  - Kinematical and vertex fitter
  - Flavor tagging
  - Vertexing
  - Particle ID (Likelihood)
  - Event shape
  - Likelihood/Fisher analysis
- People tend to use standard packages but…
  - System is not well organized/documented
  - Have started a task force (consisting of young Belle members)

Belle Software Library

- CVS (no remote check in/out)
  - Check-ins are done by authorized persons
- A few releases per year
  - Usually it takes a few weeks to settle down after a major release as we have no strict verification, confirmation or regression procedure so far. It has been left to the developers to check the “new” version” of the code. We are now trying to establish a procedure to compare against old versions
  - All data are reprocessed/All generic MC are regenerated with a new major release of the software (at most once per year, though)

Library cycle (2000~2002)

- Reprocess all data before summer conferences
  - In April, we have a version with improved reconstruction software
  - Do reconstruction of all data in three months
  - Tune for physics analysis and MC production
  - Final version before October for physics publications using this version of data
  - Takes about 6 months to generate generic MC samples
- 20020405 $\rightarrow$ 0416 $\rightarrow$ 0424 $\rightarrow$ 0703 $\rightarrow$ 1003

Data

(Detailed numbers are just for your reference)

Event size (34KB/event on tape)

- Raw data: Typical run (Luminosity:23.8pb^{-1})
  - Accepted 1,104,947 events
  - Accept rate 349.59 Hz
  - Run time 3160.70 s
  - Readout Dead time 228.02 s 6.20% (6.69% intrinsic)
- L3 (online software trigger; fast tracking and vertexing) accepted 59.7% of events
  - Recorded 659976 events Used 24.6[GB] (24604[MB])
- Data size of the sub detector on average/event
  - SVD 13KB, CDC 4KB, ACC 1KB, TOF 2KB, ECL 6KB, KLM 4KB, EFC 1KB, TRG 3KB

DST: Event size/event types

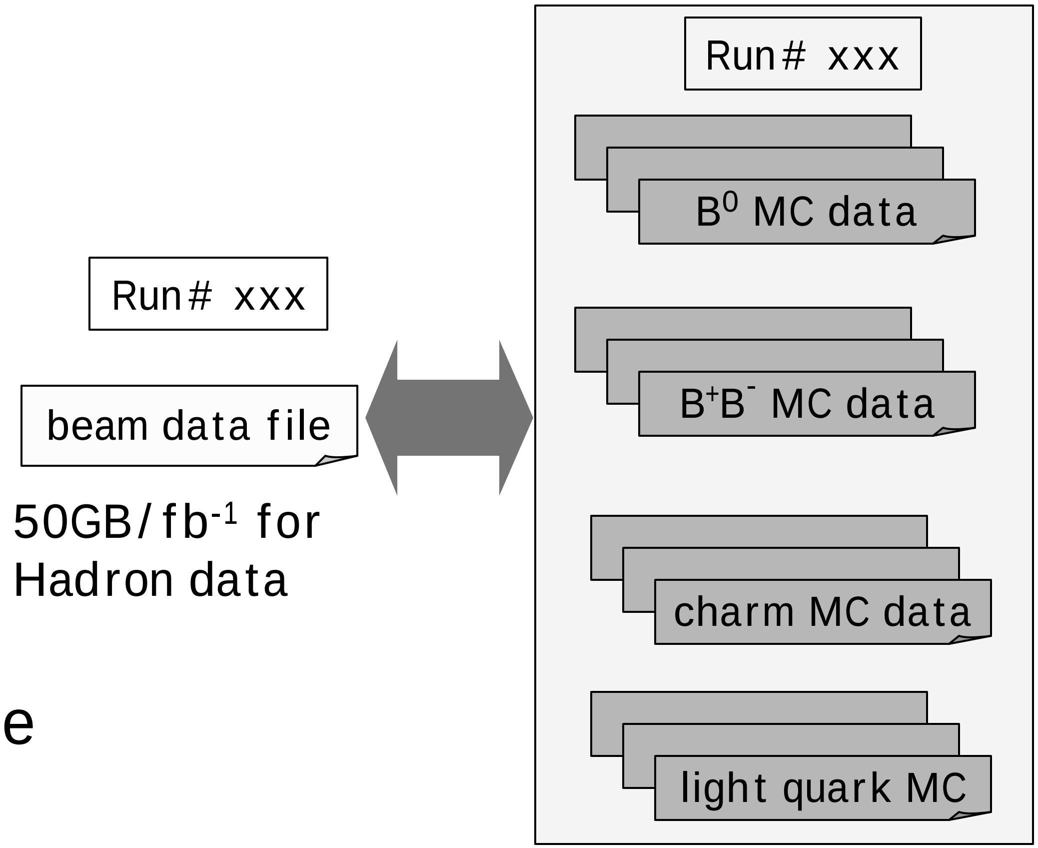
- L4 input (661265)→output (572240) (rate:86.5372%)
  - Level 4 software trigger is a fast tracking, clustering etc.
- Output file 41GB, hardware compressed on tape, 38GB
  - 67KB/L4 passed events
- Bhabha: 47744
  - Barrel Bhabha 28480
- Gamma pair: 7774
- μ pair 13202:
  - Barrel μ pair 9538
- Hadron 198027
- HadronB 91965
- Hadron with J/Ψ candidates 21901
- τ pair 95161
- Two photon 64606

Data size

- Raw data
  - 300TB written since Jan. 2001 for 100 fb^{-1} of data on 1,500 tapes
- DST data
  - 500TB written since Jan. 2001 for 150 fb^{-1} of data on 2,500 tapes, hardware compressed
- MDST data
  - four vectors and vertex and PID info only
  - 5TB for 100 fb^{-1} of hadronic events (BBbar and continuum), compressed with zlib, 12KB/event
  - τ, two photon: 3TB for 100 fb^{-1}

Generic Monte Carlo data

- Mainly used for background study
- Generated for each run, three times as much as real data
- 15~20GB for one million events
  - 100 GB for $1fb^{-1}$ of the real data
  - No “hits” are kept



50GB/fb^{-1} for Hadron data

300GB/fb^{-1} for 3 sets of generic MC

Computing

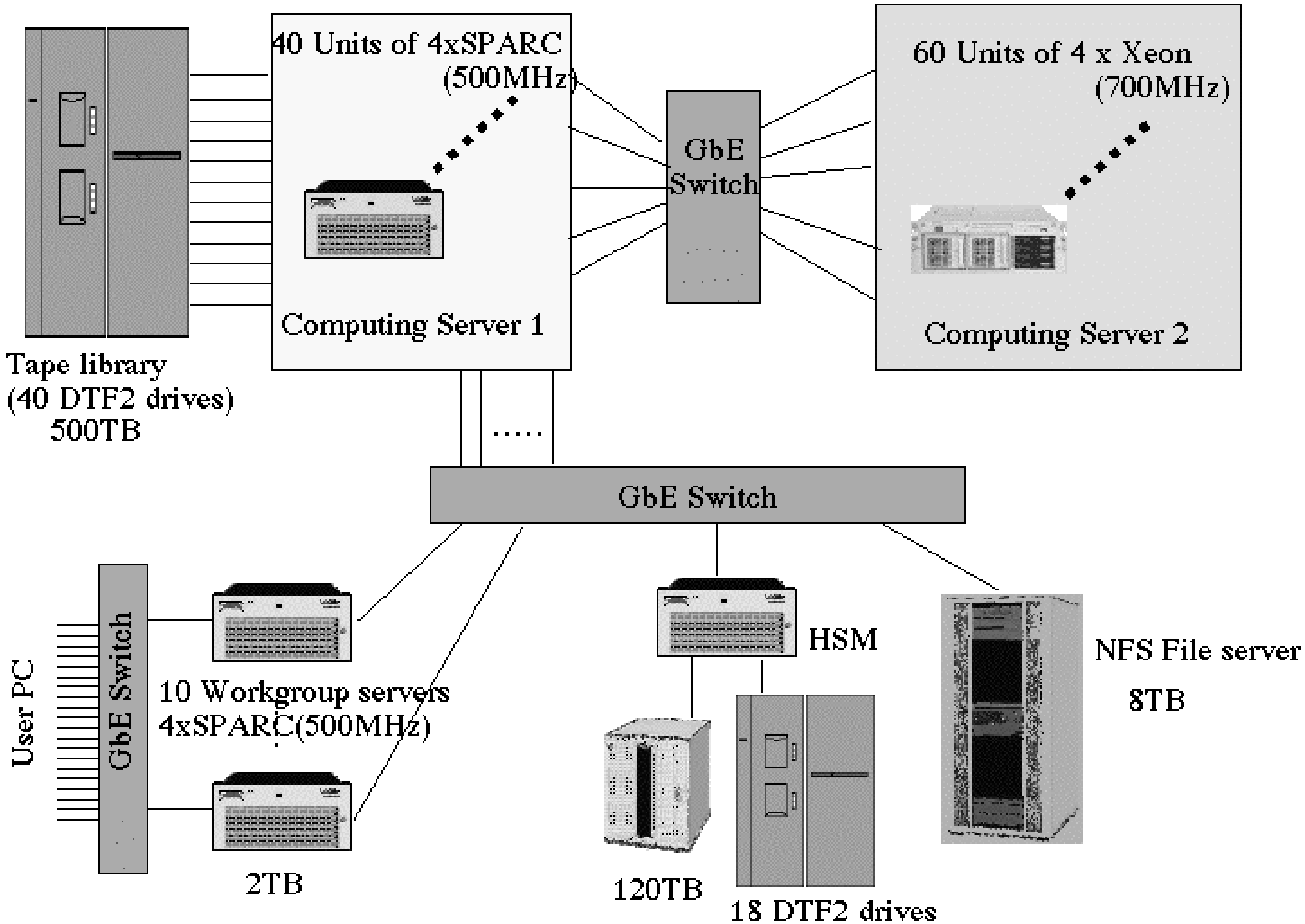
(My) Computing requirements

- All valid data can be reprocessed in three months
- Generic Monte Carlo events of the order of 3~10 times the integrated luminosity can be generated in six months
- All Hadron MDST as well as lots of MC MDST files can stay on disk
- CPU power should not be the bottle neck for physics analysis

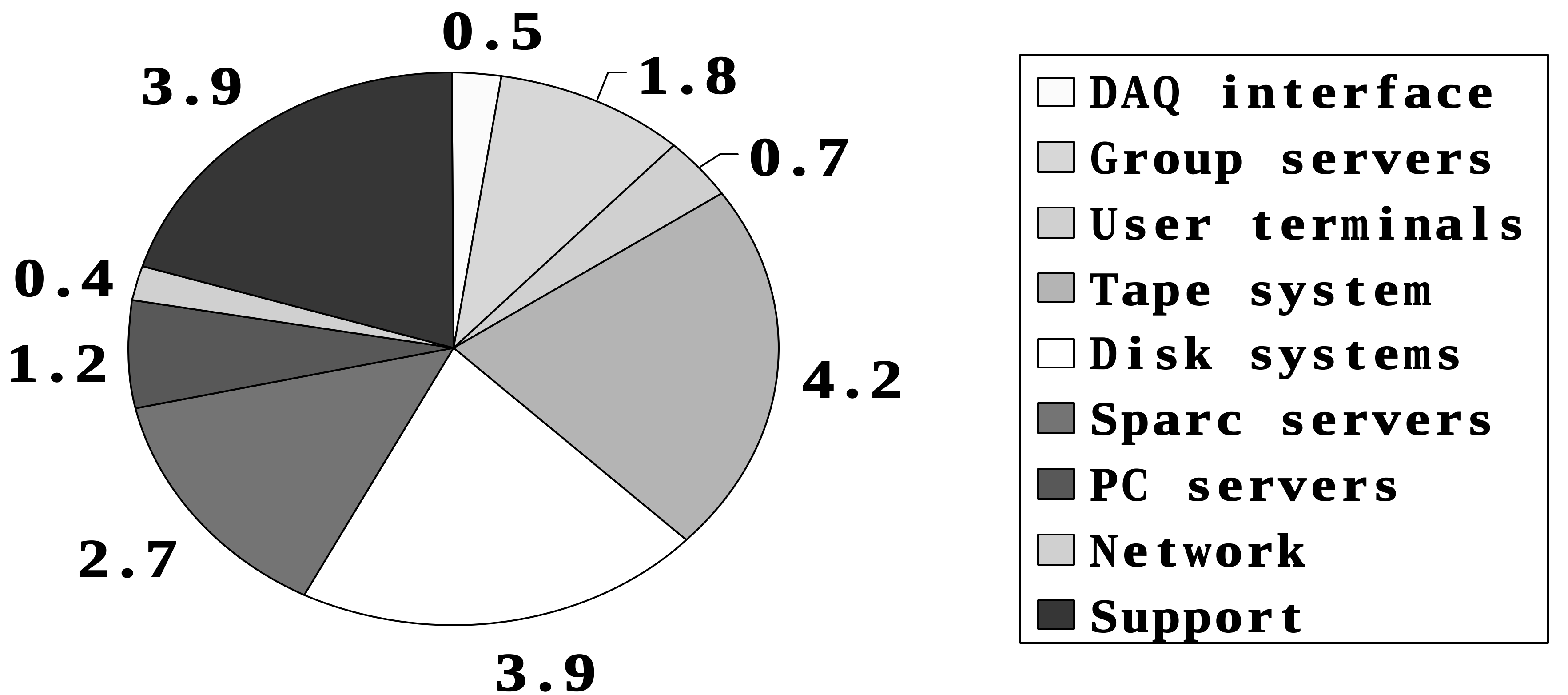
Computing Equipment budgets

- Rental system
  - Four → five year contract (20% budget reduction!)
    - 1997-2000 (25Byen;<20M euro for 4 years)
    - 2001-2005 (25Byen;<20M euro for 5 years)
- Belle purchases
  - KEK Belle operating budget 3M Euro/year
  - Of 3 M Euro, 0.4~1Meuro/year for computing
    - Tapes(0.2MEuro), PCs(0.4MEuro) etc.
  - Sometimes we get bonus(!)
    - so far about 1M Euro in total
- Other institutions
  - 0~0.3MEuro/year/institution
  - On the average, very little money allocated

New rental system(2001-2005)



Rental system: total cost in five years (M Euro)



Sparc CPUs

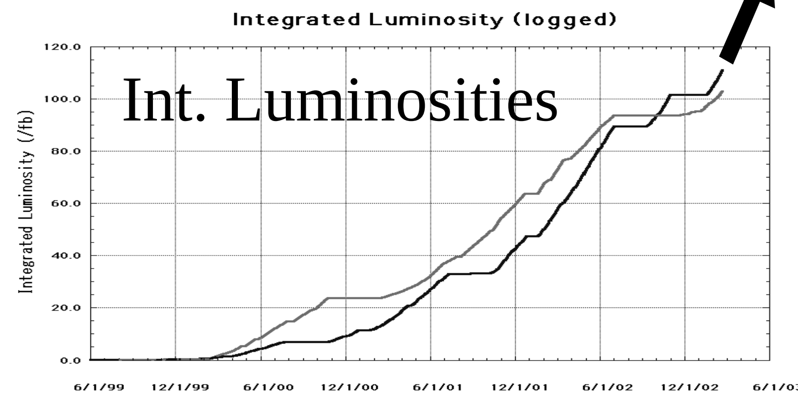
- Belle's reference platform
  - Solaris 2.7
- 9 workgroup servers (500Hz, 4CPU)
- 38 compute servers (500Hz, 4CPU)
  - LSF batch system
  - 40 tape drives (2 each on 20 servers)
- Fast access to disk servers
- 20 user workstations with DAT, DLT, AITs

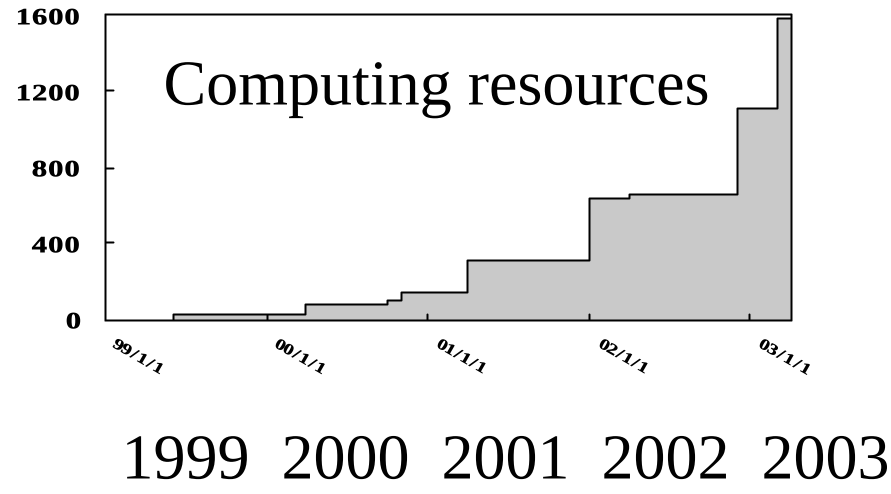
Intel CPUs

- Compute servers (@KEK, Linux RH 6.2/7.2)
- User terminals (@KEK to log onto the group servers)
  - 106 PCs (~50Win2000+X window sw, ~60 Linux)
- User analysis PCs(@KEK, unmanaged)
- Compute/file servers at universities
  - A few to a few hundreds @ each institution
  - Used in generic MC production as well as physics analyses at each institution
  - Tau analysis center @ Nagoya U. for example

Belle PC farms

- Have added as we take data
  - ’99/06:16 4CPU 500MHz Xeon
  - ’00/04:20 4CPU 550MHz Xeon
  - ’00/10:20 2CPU 800MHz Pen III
  - ’00/10:20 2CPU 933MHz Pen III
  - ’01/03:60 4CPU 700MHz Xeon
  - ’02/01:127 2CPU 1.26GHz Pen III
  - ’02/04:40 700MHz mobile Pen III
  - ’02/12:113 2CPU Athlon 2000+
  - ’03/03:84 2CPU 2.8GHz Pen 4
  - …
- We must get a few to 20 TFLOPS in coming years as we take more data





Disk servers@KEK

- 8TB NFS file servers
- 120TB HSM (4.5TB staging disk)
  - DST skims
  - User data files
- 500TB tape library (direct access)
  - 40 tape drives on 20 sparc servers
  - DTF2:200GB/tape, 24MB/s IO speed
  - Raw, DST files
  - generic MC files are stored and read by users(batch jobs)
- 35TB local data disks on PCs
  - zfserv remote file server
- Cheap IDE RAID disk servers
  - 160GB $\times$ (7+1) $\times$ 16 = 18TB @ 100K Euro (12/2002)
  - 250GB $\times$ (7+1) $\times$ 16 = 28TB @ 110K Euro (3/2003)

Data access methods

- streams of data (no objects)
  - DTF2 Tape:200GB/tape
    - Files are managed by software written by Fujitsu
  - Other tape formats: no direct read/write from tapes
  - Disk
    - Just use UNIX file system calls
- index_io: pointer to events in MDST
  - saves disk space for skim files
  - started to use from last fall
- zfserv: simple data server (TCP/IP)
  - can access data files over the network (without NFS)
  - accessing PC local disks from other computers

Production

Network/data transfer

Other issues

File problem

- More than 10 thousand runs have been recorded
- Each run has a unique run number (and experiment number)
- For each run, there are many different data(MC) files
  - 24 generic MC files (4 ´ 3/0.5)
  - Several skim files
  - ~20 types of physics skim files (for each of Hadron data and 24 MC file)
  - different version of library
- Total number of files are now more than one million
  - Size of the files ranges from KB(index skim files) to 30GB (raw/dst files)
- Started to think about managing them…
- Any good idea?

Human resources

- KEKB computer system+Network
  - Supported by the computer center (1 researcher, 6~7 system engineers+1 hardware eng., 2~3 operators)
- PC farms and Tape handling
  - 2 Belle support staffs (they help productions as well)
- DST/MC production management
  - 2 KEK/Belle researchers, 1 pos-doc or student at a time from collaborating institutions
- Library/Constants database
  - 2 KEK/Belle researchers + sub detector groups

Management and Budget

- At Belle, one person is in charge of computing, software and production
- Budget: Belle management requests to KEK management every year how much we need
  - No arrangement to share computing (and other) costs among collaborators for now
  - Like CERN, if we need it, we may have to change

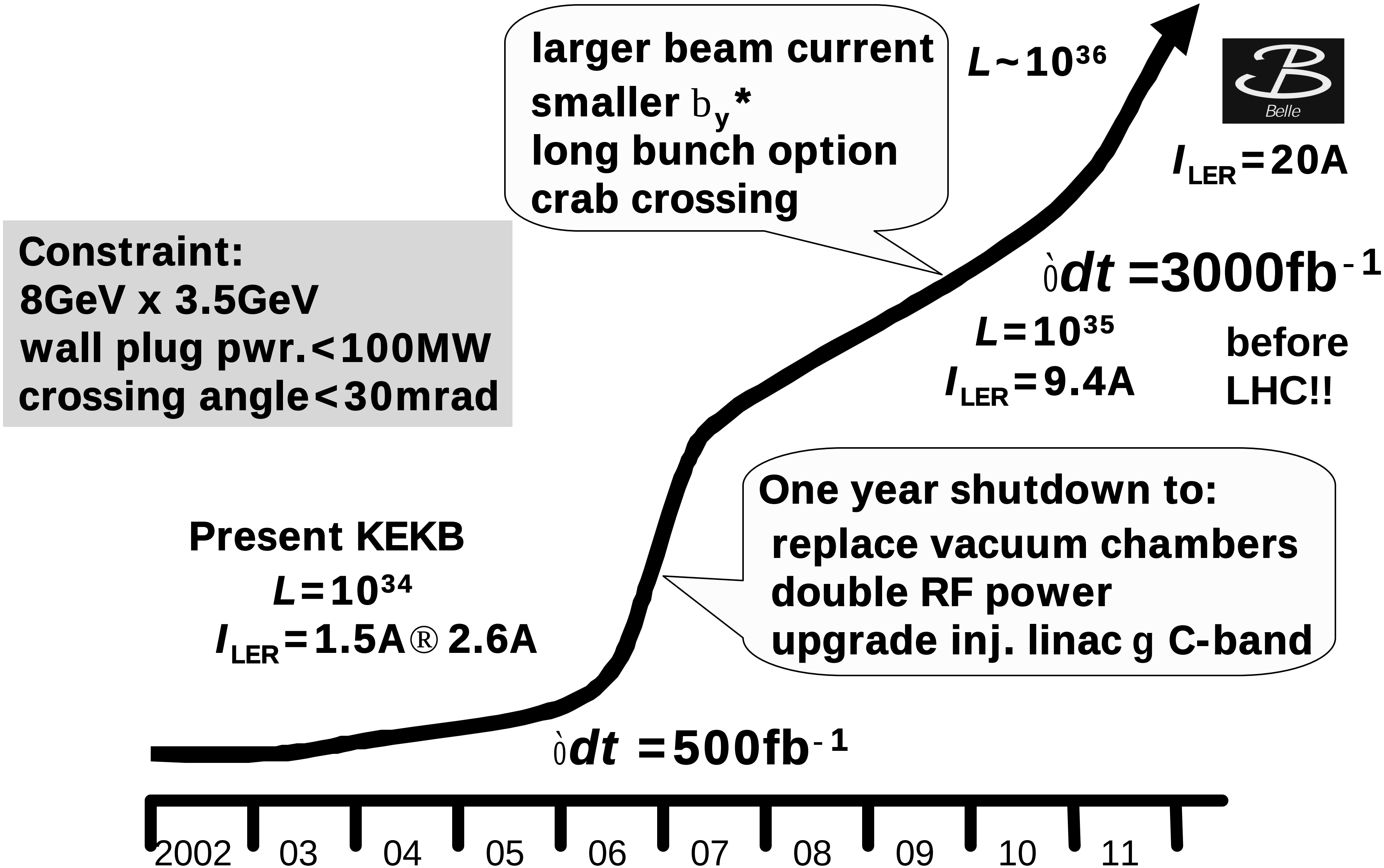
Short term plans (Summer '03)

- Software/constants updates by the end of March
  - No change since last year
  - Less systematic errors (tracking)
  - Finer ECL calibration
- Generic run dependent MC as we take data
  - Run dependent signal MC production ?
- Reprocess all data starting from April for the summer
  - More physics skim during the DST production
- Standardize more physics tools/skims
- 568 CPU LSF licenses on PC
  - Users can use CPUs in the PC farms

Long term plans

- More CPU for DST/MC production
- Faster turn around
- Distributed analysis (with local data disks)
- Better constants management
- More man power on reconstruction software and everything else
  - Reduce systematic errors, better efficiencies

KEKB upgrade strategy



Super KEKB

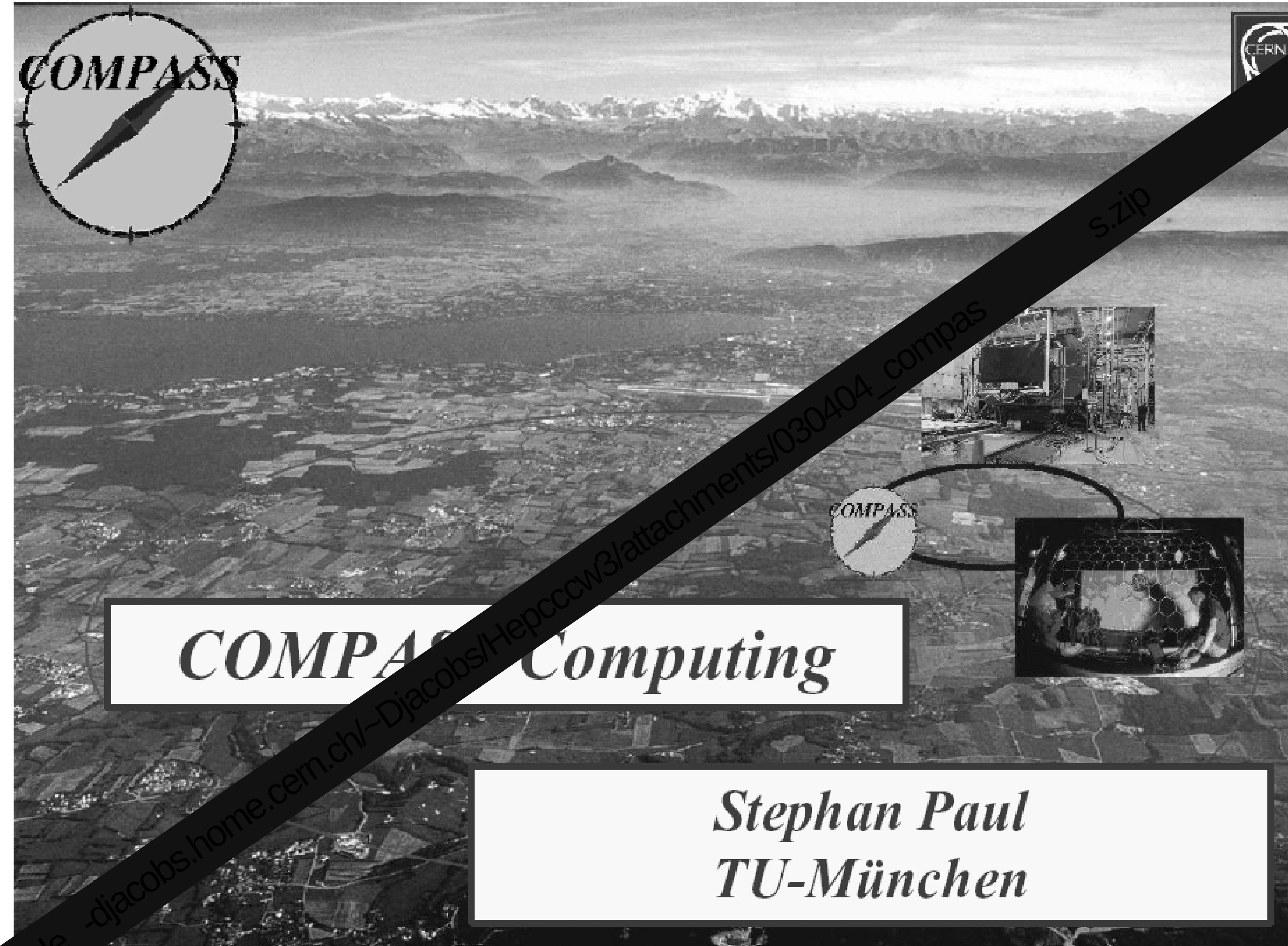
- KEK hopes to upgrade the KEKB accelerator to achieve 10^{35}~10^{36} $cm^{-2}s^{-1}$ luminosity, 1,000~10,000 fb^{-1}/year
- It will cost ~200M Euro to upgrade the machine and the detector
- However, I think S-KEKB/S-Belle need an LHC class computing facility to analyze the data
- Our traditional computing model might fail at some point even if Moore's law holds for coming years as complexity of the physics data analysis certainly increases

Computing for super KEKB

- For (even) 10^{35} luminosity;
  - DAQ:5KHz, 100KB/event⇒500MB/s
  - Physics rate: BBbar@100Hz
  - 10^{15} bytes/year:1PB/year
  - 800 4GHz CPUs to catch up data taking
  - 2000 4GHz 4CPU PC servers
  - 10+PB storage system (what media?)
  - 100TB MDST/year online data disk
- Costing >70 M Euro?

Will Grid help?

- Just started learning
- Started to extend our traditional, centralized computing
  - Remote institutions connected over super-SINET (1Gbps dedicated lines)
- Functionalities we want
  - replication
  - single login
  - batch system
  - parallel processing
  - fast data transfer
  - data/file management


COMPASS
COMPASS Computing
Stephan Paul
TU-München

Outline

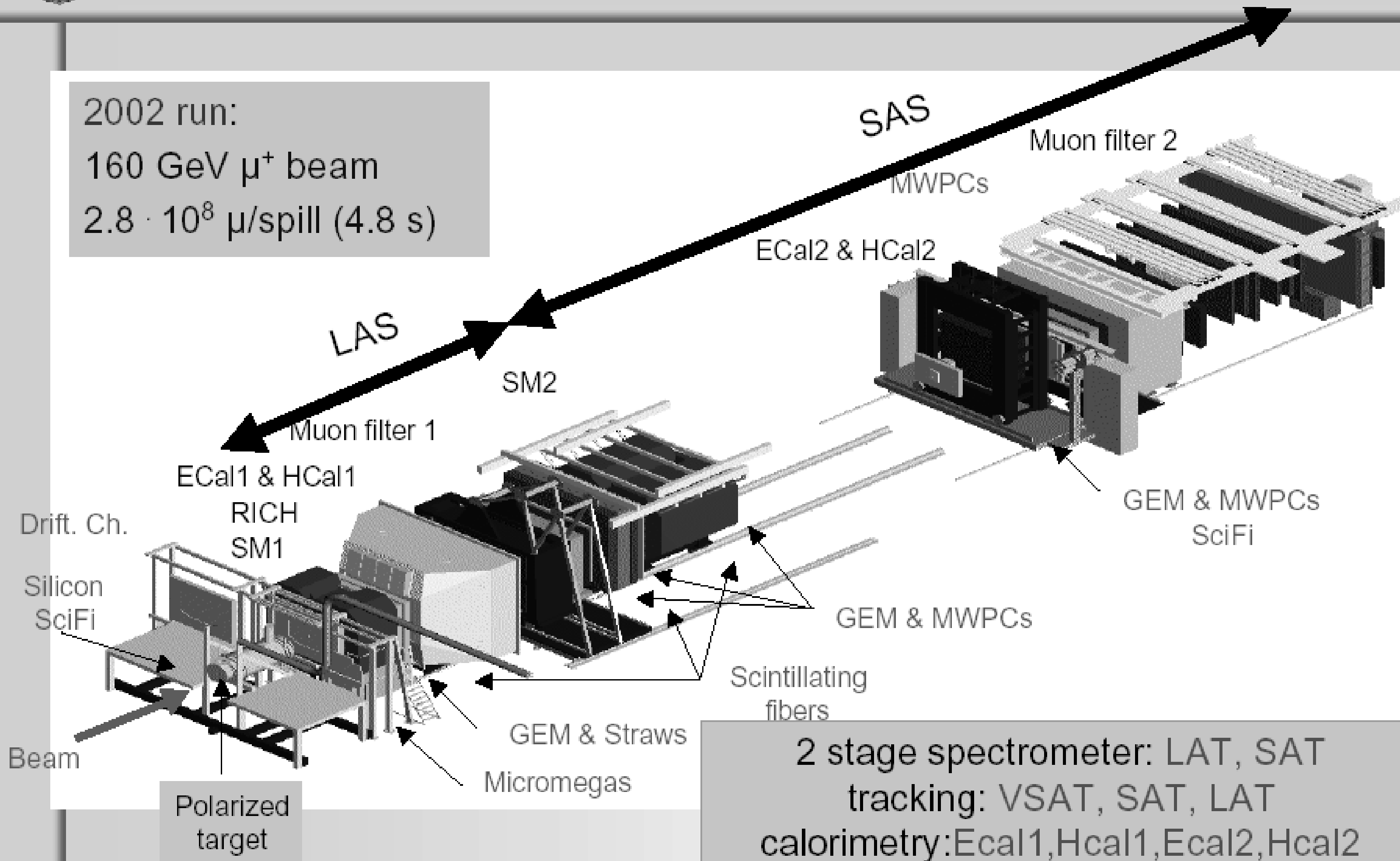
- COMPASS Detector and DAQ
- Central Data Recording and Event Database
- Event Reconstruction
- Data Volume
- Analysis Model

COMPASS Status:

2001 → COMPASS *Detector commissioning*
2002 → COMPASS *data taking* - **Analysis commissioning**
2003 → COMPASS *data taking*

COMPASS Detector

2002 run:
160 GeV μ^+ beam
$2.8 \cdot 10^8$ μ/spill (4.8 s)



2 stage spectrometer: LAT, SAT
tracking: VSAT, SAT, LAT
calorimetry:Ecal1,Hcal1,Ecal2,Hcal2
PID: RICH-1, μ–wall1, μ–wall2
Polarised Target

Compass Offline

- 40-60 MB/s
  - Central Data Recording (CDR) is fundamental
- $5 \cdot 10^9$ events, many detectors
  - Complexity
- 300 TB
  - CDR from the experiment to the Computer Centre
  - Automated storage system at CERN (Castor – 100GB tapes)
- ~100 days of data taking
  - DST production has to be done near tape silo
  - Many CPU's are used in parallel - achieve "quasi-online" analysis (future)
  - CPU system usable for further analysis during "off data taking" periods
- Analysis Strategy
  - produce DST and filtered mDST centrally
  - export mDST to different home institutes for final analysis (network)
  - <u>MonteCarlo processing in the home institutes</u>

Central Data Recording

Design value (35MB/s)

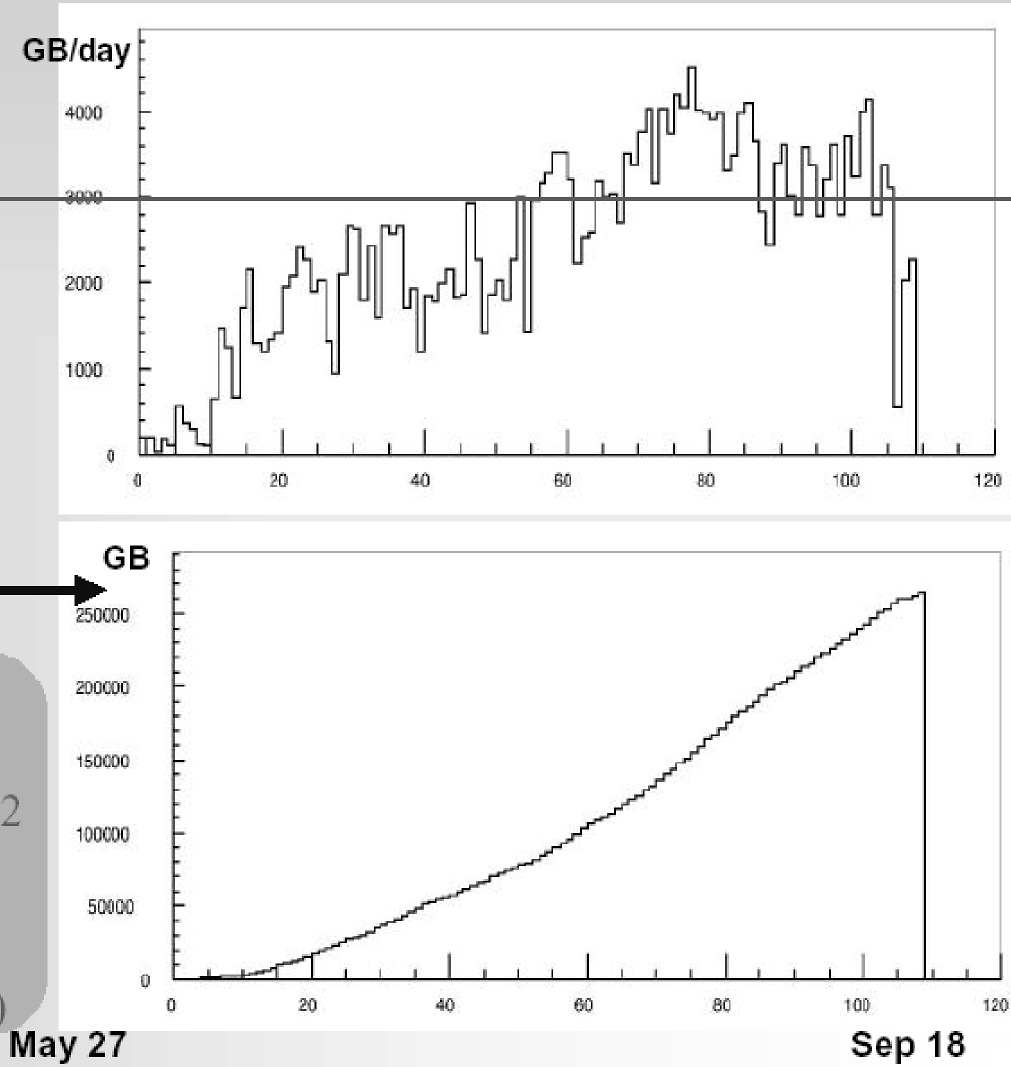
260 TByte in ~100 days

Compare to BaBar:

- 1 TByte/day
- 662 TByte in 1999-2002

Or CDF:

- 500 TByte/year
- (U.S. year >> 100 days)



Data in Database

2001:

Detector commissioning: 65 days, 13 TByte
Target polarized longitudinally: 14 days, 15 TByte total: **28 TByte**

2002:

Target polarized longitudinally: 57 days, 173k spills
Target polarized transversely: 19 days, 52k spills total: **260 TByte**

2002 on average 22k triggers/spill → 5 Gev (3.8Gev, 1.2Gev)
online: ~260,000 files (~100 files (1GB)/ run)

- ✓ After transfer to CCF ~2000ev/s are formatted into Objectivity/DB
- ✓ Different data bases for raw events and special events
- ☞ Planned: result of event online-reconstruction added to tape based DB federation with raw data
- 2002: offline reconstruction (calibrations were/are not available)

Objectivity phased out at CERN and replaced by Oracle9i

Migration to Oracle

CERN Contract with Objectivity terminated; they have chosen Oracle as new DBMS

PAST (Objectivity/DB 6.1.3)

In the ODBMS:
- **Metadata & Associations**
- **Raw Data**
- **Reconstructed Events**
- **Conditions**

In plain files:
- **Conditions**

DB overhead
~30% of Raw Data size
(turns into ~6% on tape)

FUTURE (ORACLE 9i)

In the ORDBMS:
- **Relations (metadata)**

In plain files:
- **Raw Data (orignal DATE files)**
- **Reconstructed Events (8% of raw)**
- **Conditions**

DB overhead:
~0.3% of Raw Data size

Event Reconstruction

Event Reconstruction: 3-steps

- Pattern recognition and (charged) track reconstruction
- Particle ID & calorimeter reconstruction (neutrals)
- Basic event analysis (vertexing, kinematics)

☞ Data Filter steps possible at intermediate levels

Task (to come):

- "quasi real time" processing of ~ 10k events/15 seconds (~750 evts/sec.)
- reconstruction must be efficient and fast
  - goal: 100-200 msec/event for charged track reconstruction
  - 100 msec/event for other processing

  ☛ parallel processing on ~200 processors

Status:

- Average time to reconstruct one event: 400ms/ev (1.5 GHz standard PC)
  (typical: 300ms/ev; during the first 400ms of a spill up to 1300ms/ev)
- 5Gev • 400ms/ev = 2Gs / 200CPUs = 10 Ms/CPU @ 100% efficiency
  ⇨ 115 days on 200 CPUs

Event Reconstruction

- **Compass Computing Farm:**
  - 100 dual Processor PIII **Linux PCs**
  - Mainly for **event reconstruction** and (Mini)-DST production
  - maintained and operated by CERN staff
  - Part of CERN LSF BATCH-system

- Process as much in parallel as possible:

  1 run : ~ 100 data files

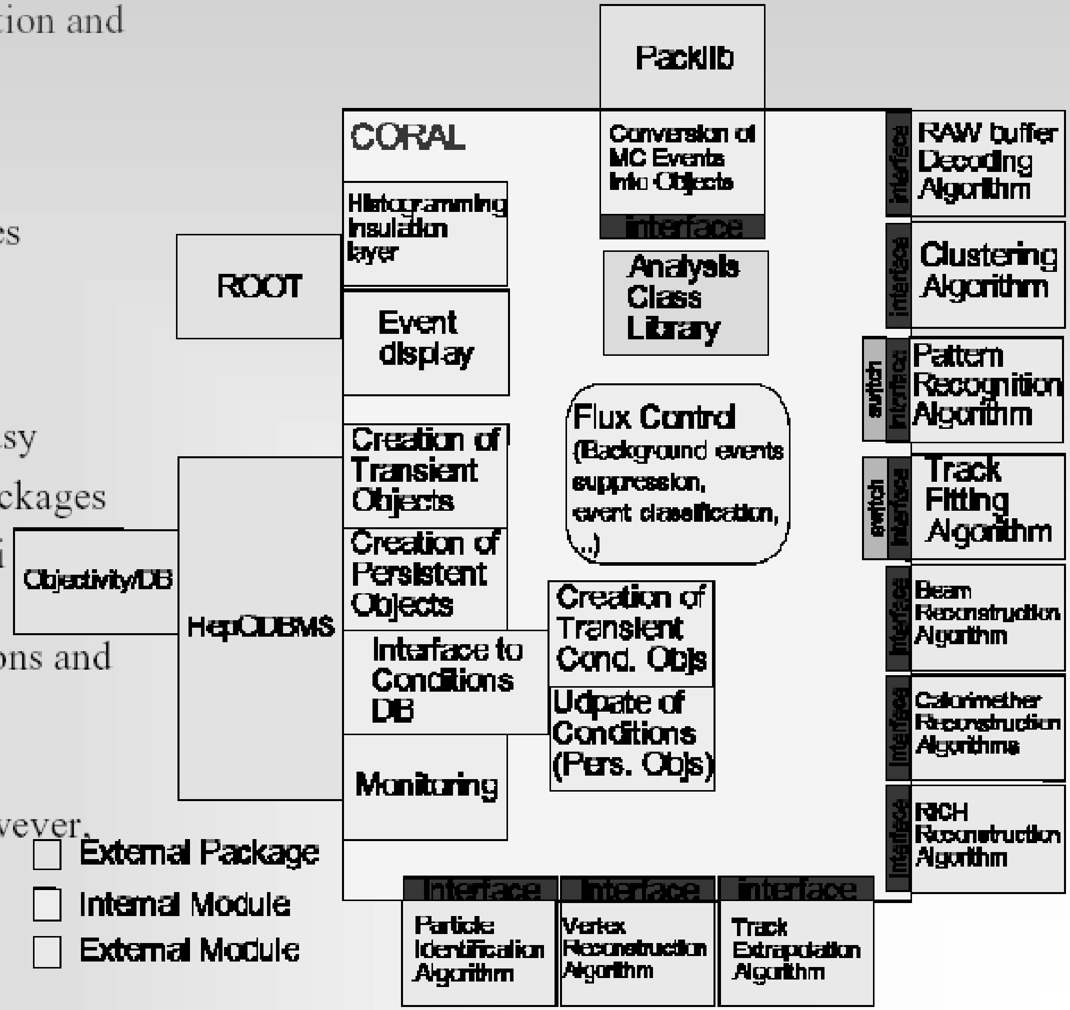
  : ~ 350-400 batch jobs running at the same time

- Today: about 20% of 2002 data are being processed for physics analysis (since January)

CORAL - Event Reconstruction

= COMPASS Reconstruction and AnaLysis Program

- Modular architecture
- Following OO techniques
- Fully written in C++
- Written from scratch
- Defined interfaces for easy exchange of external packages (e.g. OB/DB → Oracle9i
- Access to event, conditions and calibration data bases
- Not as user friendly, however, as it may look



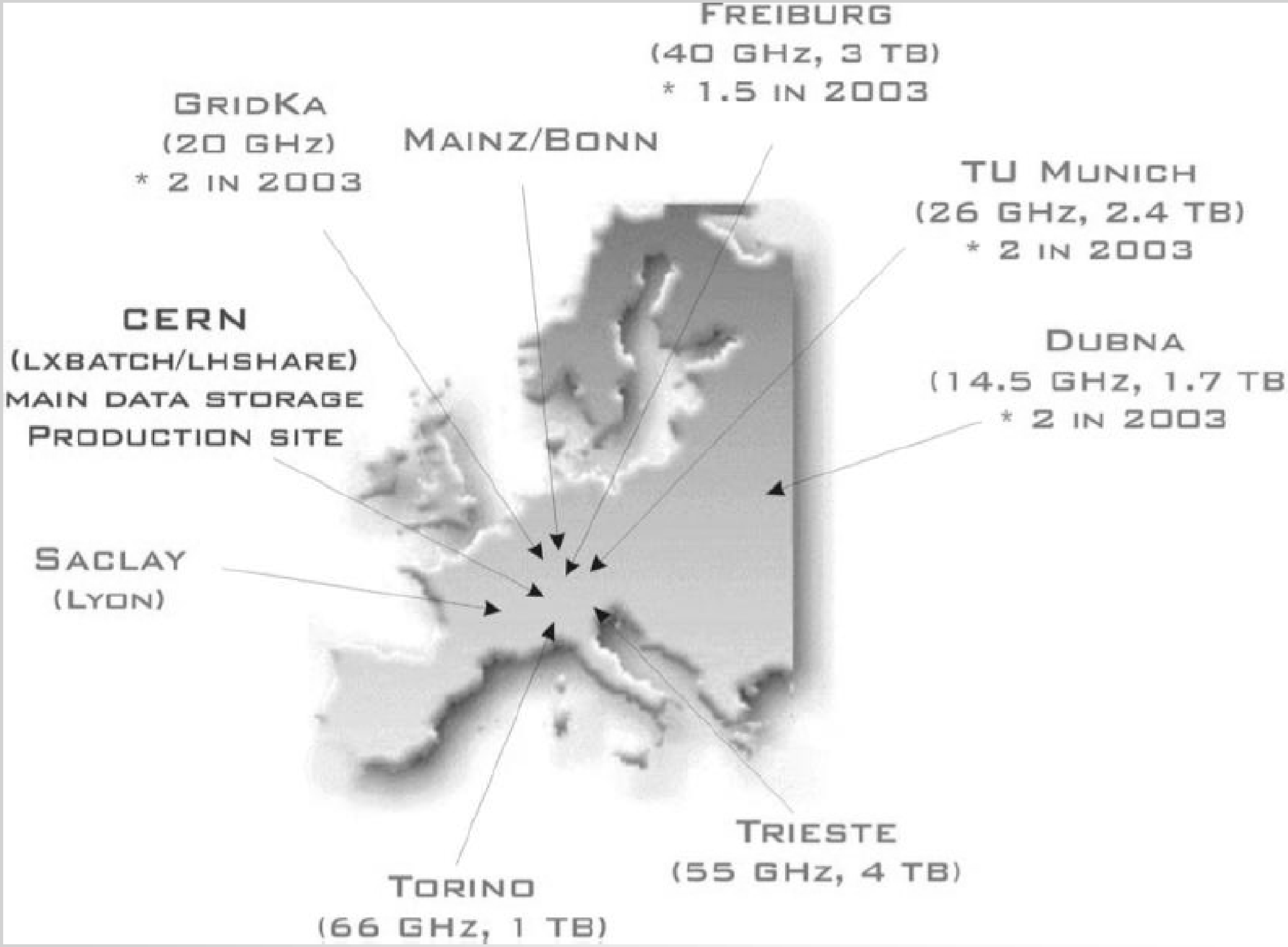
PHysics Analysis Software Tools

- Data structure and file format for storage of reconstructed events and analysis relevant experiment parameters

- Software infrastructure for
  - Access to reconstructed event's data (10^3evts/s)
  - Analysis code development
  - Event filtration tools
  - n-tuple writing for further analysis
  - ROOT library available
  - COMPASS library (e.g. vertex-fitting, constraint fitting)

- Easy portability (no CERN platform dependence)
- Easy to use (even for senior physicists)
- Data compression (x2)

COMPUTING Distribution



Conclusions

- COMPASS : running experiment with high data rates/volume
- 'Small' size collaboration (210 persons) spread over Europe (mostly)
- COMPASS is in its first year of "real" data analysis
  Challenges are enormous: calibrations, data handling, etc.
  … but very interesting physics ahead and thus its worth all the efforts
  (and sleepless nights)
- Tools for event reconstruction and (Mini- & Micro-) DST analysis exist but still need continuous improvements and adaptations to reality
- Still weak on tools to submit & monitor and manipulate & handle output of 200+ batch jobs per hour (effort was underestimated by many of us)
  Presently under evaluation: ALIEN (ALICE development)
- Organizational structure of analysis group:
  1. Established a core group at CERN with regular WORK(!)shops
  2. Setup shifts for DST production
  3. Lot of work done in home institutes
  4. Communication to be optimized (still travel intensive)

Report from DataGrid Project Review

Fabrizio Gagliardi

Project Leader

Fabrizio.Gagliardi@cern.ch

Major Review Goals

- Important to get approval for a number of variations from original plans:
  - refocus on production testbed releases driven by applications (HEPCAL)
  - synchronization with LCG timeline and plans
  - multiple testbeds (development, application)
  - financial status of the project
  - M/W development plans
  - dissemination activity
  - support for future EU projects (EGEE)

DataGRID project priorities refocused

After initial middleware development and testbed deployment, effort has been refocused on quality and stability

- Quality Policy Statement published
  - http://eu-datagrid.web.cern.ch/eu-datagrid/WP12/default.htm
- List of priorities defined at a project retreat
  - http://documents.cern.ch/age?a021130
- Followed-up at the last project conference
  - http://www.tomiexpress.hu/datagrid/

- Show-stoppers found by users on the application testbed were the highest priority
- Incremental improvements driven by the needs of the applications (HEPCAL)

Atlas (August and Dec/Jan) & CMS (Dec) Evaluations
(*DETAILED PAPER IN PREPARATION*)

- RESULTS
  - Atlas software was used in the EDG Grid environment
  - Several hundred simulation jobs of length 4-24 hours were executed , data was replicated using grid tools
  - Results of simulation agreed with 'non-Grid' runs
- OBSERVATIONS
  - Good interaction with EDG middleware providers and with WP6/8
  - With a substantial effort it was possible to perform the jobs
  - Showed up bugs and performance limitations (fixed or to be fixed in EDG 2.0)
  - **We need EDG 2.0 release for use in large scale data challenges**

- RESULTS
  - Could distribute and run CMS s/w in EDG environment
  - Generated ~250K events for physics with ~10,000 jobs in 3 week period
- OBSERVATIONS
  - Were able to quickly add new sites to provide extra resources
  - Fast turnaround in bug fixing and installing new software
  - Test was labour intensive (since software was developing and the overall system was fragile)
  - **EDG 2.0 should fix the major problems providing a system suitable for full integration in distributed production**

Review Conclusions

- Difficulties arise from finding balance between support of the current s/w and effort devoted towards advance solutions and migration to new emerging standards
- Important progress made in functionality and performance of software and testbed(s)
- Pioneered Grid technology adopted by many projects including LCG for one of the largest scientific enterprises to date
- Exploring further Grid major deployment activities in FP6
- Fulfilling its role of EU Grid flagship project

EU reviewers feedback

- Congratulations for a good review.
- Good presentations and no "Murphy's law“ for the demos. An impressive job.
- This success reflects the interest of all the partners involved.
- Congratulates the project management for taking the risk of concentrating on production quality.
- Would like to see the promise fulfilled of no relevant loss of functionality by the end of the project.

3rd year schedule

March

D6.6,8.3,9.3,10.3 evaluation reports (**rescheduled**)

D7.6 Security design report

May

EDG 2.0 release deployed

subsequent improvements based on application feedback

Project conference in Barcelona

June

D11.6 Report of the 2nd annual conf. and industry Grid Forum workshop

July

D9.4 EO application platform interface

September

EDG 2.x release deployed

final testbed

D1.6,2.5,3.5,4.5,5.5,6.7 sw and doc.

Final project conference in Heidelberg

December

D11.7 Report on final project conference

D11.9 Report on contributions to international standards

D1.7,2.6,3.6,4.6,5.6,6.8,7.7 Final evaluation reports

D8.4,9.5,10.4 Application demos and final reports

D12.19 Third annual report

Early 2004

Final project review

Conclusions

- Important milestone passed
- Major re-orientation of the project accepted
- EDG M/W being released to LCG for LCG-1 release
- Need to develop further plans with LCG and in view of future project EGEE
- Need to accommodate other applications (in agreement with LCG)
- Plan long term support of EDG developments (after 2003)
- Major opportunity for further EU funding (EGEE)
- EDG was launched by HEPCC, they can be happy and proud
- We hope to repeat the same success with EGEE!

LHC Computing Grid Project - LCG

Status Report - HEP-CCC

CERN, 4 April 2003

Les Robertson – LCG Project Leader

Information Technology Division
CERN – European Organization for Nuclear Research
Geneva, Switzerland

les.robertson@cern.ch

LCG

LCG Fabric

– Tier 0+1 Preparation
– Computing Data Challenges
– Physics Data Challenges at CERN
– Technology Tracking
– Regional Centre information exchange

CERN

Events during the last few months

Data Challenges

- In December 2002 the ALICE-IT Computing Data Challenge reached ~300 MB/s sustained (for 7 days) dataflow from an emulated DAQ system into the HSM system CASTOR (disk and tape) with peak values of 350 MB/s. The goal for the MDC4 in 2002 was 200 MB/s.
- In January ATLAS used 230 testbed nodes successfully for an Online Computing Data Challenge (postponed from October 2002) to test event building and run control issues

Technology tracking

- PASTA III complete (see LCG/PEB web page) – presented at CHEP

Organisation

- Fabric – Grid Deployment re-organisation at CERN
- LCG ←→EDG consolidation
  - hardware resource allocation and planning improved

Funding

- No further investment in tape infrastructure for Phase 1
  - all CERN tape drives upgraded to STK 9940B
  - if necessary Computing Data Challenges may take equipment from the production systems outside beam time
- HEP wide availability of ORACLE licenses

Next 3 months

- Migration of equipment into the re-furbished vault in Building 513
  - all new systems
  - STK Tape Silos
  - selected servers
- (Maybe) 1 GByte/s IT Computing Data Challenge in April
  - takes advantage of a period of overlap of “upgraded” 9940Bs with old equipment
  - ‘CDR’ system at → 50 cpu server + 50 disk server + 50 tape drives
  - target → 1 GByte/s into CASTOR and on to tape

Grid Deployment

Recent Progress

- Define LCG-1
  - The Grid Deployment Board agreed on a set of recommendations for th LCG-1 service on 6 February
  - Sufficient to define direction and issues to be addressed; used in planning and deploying
- Pre-production Pilot Cluster is available
  - worker nodes managed by FIO group - preparing for full integration of physics production (LXBATCH) and LHC Grid
  - configured as minimum but can move batch nodes between Pilot and LXBatch as needed
  - integrating LSF, addressing NFS vs AFS issues
- Deployment schedule ->
  - LCG-0 deployed to CERN, RAL, CNAF + Legnaro(T2), Taiwan, FNAL
  - Russia, BNL, Tokyo in preparation
  - This is actually ahead of proposed schedule

LCG Ramp-up Schedule

	Date	Regional Centre	Experiment(s)
Pilot – LCG-0 - started Feb 1			
0	15/2/03	CERN	All
1	28/2/03	CNAF, RAL	All
2	30/3/03	FNAL	CMS
3	15/4/03	Taiwan	Atlas,CMS
4	30/4/03	Karlsruhe	All
5	7/5/03	IN2P3	All
6	15/5/03	BNL	Atlas
7	21/5/03	Russia(Moscow),Tokyo	All
LCG-1 Initial Public Service Start – scheduled for July 1			

Tier 2 centres will be brought on-line in parallel once the local Tier 1 is up to provide support

Recent Progress (2)

- Certification process defined (January)
  - This has been done – agreed common process with EDG
  - Have agreed joint project with VDT (US):
    - VDT provide basic level (Globus, Condor) testing suites
    - We provide higher level testing
  - Expect to get HEPCAL test-cases from GAG
  - Need to pull in other expertise
    - E.g. EDG WP8/loose cannons
- Need <u>much more effort</u> on devising & writing tests
  - Real effort currently is only 2 people

- Packaging/configuration mechanism defined
  - Group (EDG, LCG, VDT) have documented an agreed common approach
  - Now will proceed with a staged implementation
    - Basic for LCG-1 in July, and more developed later

Slow Progress

- Delivery of middleware –
  - We have a working set ("LCG-0") that is in use now
  - Deadline for delivery of new EDG middleware by end April
  - Milestone was originally March 1
- Identify operations and call centres – Milestone was February 1
  - 2 candidates for operations centres – hopefully this should be clarified soon, possibly as a collaboration
  - No clear candidate for a support centre –
    -- the LCG CERN group will have to set up a basic support service

Test-beds and services

- Agreement with EDG and key Regional Centres to separate test-beds and later merge EDG and LCG production services
  - The only way to deal with the scarce support resources

- March – July:
  - Very limited participation of CERN in the EDG applications test-bed (access to Castor, user interface)
  - EDG core sites will run either EDG applications testbed or LCG pilot, unless they have resources for both

- Once LCG-1 is established:
  - There will be a single production system for LHC – LCG-1
    - CERN (and others) will support non-LHC EDG partners on LCG-1
    - EDG will maintain development testbeds

Deployment Staffing

- Staffing of Grid Deployment at CERN left too late – Now we have a serious lack of effort
- EDG testbeds absorbing more effort than expected
  - applications testbed continued after EDG review
  - EDG integration activity absorbing effort at CERN
  - rationalisation of EDG/LCG resources at CERN is a response but not a complete solution
- Infrastructure support, Experiment support (both grid experts, and production adaptation) understaffed
- Testing group is badly understaffed
  - had expected to find more tests from EDG
  - hoped that EDG WP8 and GAG would provide packaged tests
  - urgent to find at least 3 more full time people to contribute here
- INFN recruiting now – but do not expect arrivals before July
- Scheduled German recruitment would largely solve the problem – but administrative difficulties at present

Middleware System Support

- Acquiring in-depth expertise in the middleware for LCG-1
  - 2(3) people at CERN
  - building relationships with EDG, Globus, VDT
- Very important activity –
  - problem determination
  - fast fixes
  - expert feedback to middleware owners
- European grid support centre
  - Maarten Litmaath as 1/3 of technical Globus support people (SE, UK, LCG)
  - Will participate in Globus 2.4 release process

Security Group

- Dave Kelsey will lead ongoing security activity
  - Policies
  - Security strategy and plan
  - This is needed urgently – as basis for operational agreements at centres
- Security operational issues:
  - Led by Dane Skow (FNAL), group of site security contacts
  - Gathering issues, constraints, etc.
  - This group will handle daily security issues
- Proposing collaboration on VO management
  - FNAL, INFN, ...

Collaborative Activities

- HICB – JTB
  - GLUE Schema and evolution
  - Validation and Test Suites
  - Distribution, Meta-Packaging, Configuration
  - Monitoring tools (proposed), aspects of ops centres
  - Proposed collaboration on VO tools (led by FNAL)

- GGF
  - Production Grid Management (operations)
  - User Services (call centres)
    - Tools, trouble ticket exchange standards, etc
  - Site AAA (security)
  - Particle and Nuclear Physics Applications area
    - As a forum in GGF to present issues and get collaboration

- Other
  - HEPiX – Fabric, operations, tools, procedures
  - Security – site security contacts
  - Storage Interfaces – SRM

Grid Technology

- LCG is a consumer, not a producer
- Evaluating new technologies, tracking developments, industry
- Proposing the strategy – suppliers, components
- Lobbying with potential, new projects
- Identifying risk, contingency plans

Forging Partnerships to Developing Global Infostructure

Irwin Gaines
FNAL/DOE
http://home.fnal.gov/~gaines/Info.html

Outline

- Principles for forging Partnerships
- Components of global infostructure
- Contributors
- Strategies for moving forward together

Agreement on 5 principles:

- The cost and complexity of 21st Century Science requires the creation of **advanced and coherent global Infostructure** (information infrastructure).
- The construction of a coherent Global Infostructure for Science requires **definition and drivers from Global Applications** (that will also communicate with each other)
- Further, **forefront Information Technology must be incorporated** into this Global Infostructure for the Applications to reach their full potential for changing the way science is done.
- **LHC** is a near term Global Application requiring advanced and un-invented Infostructure and is **ahead in planning** compared to many others.
- U.S. agencies must work together for effective U.S. participation on Global scale infostructure, and the successful execution of the LHC program in a **4 way agency partnership, *with international cooperation in view*.**

Global Infrastructure for Science: Why?

- Scale and complexity of modern science demands world wide involvement
- Scarcity of resources demands efficient resource utilization and decentralization
- Most especially must efficiently utilize scientific effort

Global Infrastructure for Science: What?

- Tools (and deployment) that facilitate location independent scientific participation
- Tools (and deployment) that facilitate efficient utilization of globally distributed resources
- Tools (and deployment) that allow resources to be directed towards highest priority problems even with a globally distributed architecture
- ➢ Much more than just the Grid!! We need a full collaborative research environment. Not just access to data; instead access to physics!

Recent Activities

- Joint DOE/NSF Physics/Computer Science (4way) meetings (Oct 16, Nov 22 [including Europeans], Feb 7, Feb 14, Feb 27)
- Mar 21 meeting at CERN
- ITR Proposals to NSF
- European meetings to prepare 6th framework proposal EGEE

Partnerships

- International: Europe/US/Asia (Europe in particular putting heavy funding into “Grid”)
- Interagency: Different funding agencies
- Interdisciplinary: Application scientists and computer scientists

➢ Must have broader applicability than just the LHC; but LHC is an ideal technical driver/early adopter

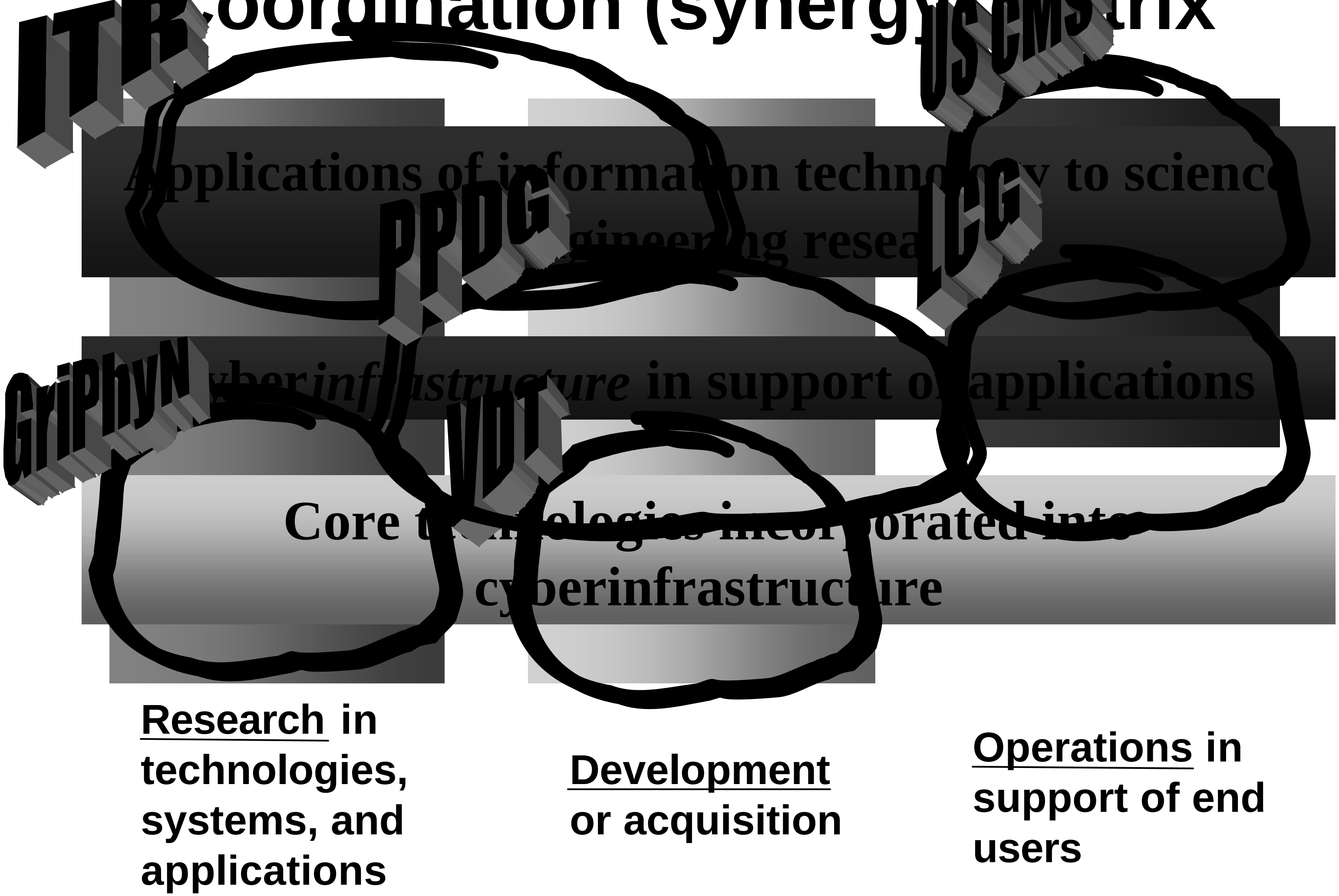
LHC as exemplar of global science

- Project already involves scientists (and funding agencies) from all over the world
- High visibility science
- Experiments already making good use of prototype grids
- Sociological (as well as technical) reasons for decentralized computing systems
- Recognized challenge of accumulating sufficient resources

National Science Foundation

Blue Ribbon Panel on Cyberinfrastructure

Coordination (synergy) Matrix



Contributors

- Funding Agencies: Base Program
- Funding Agencies: LHC Research Program (LHC Software & Computing Projects)
- US Funding Agencies: networks and infrastructure
- CERN
  - Tier 0/1 facilities at CERN
  - Networking and infrastructure
  - LCG Project
- Other collaborating countries funding agencies
- DOE/NSF Computatational Science Research Program

Strategies for Moving Forward

- Agreement on 5 principles (international version)
- Reviewers/liaisons from other side of Atlantic for new round of proposals
- Technical Workshops
- Startup of some common projects (looking to broaden beyond just LHC):
  - Grid Middleware Institute (Computer science side)
  - Open Science Consortium (Domain science side)
- Periodic video meetings to review progress, especially on status of “missing pieces”
- Trans Atlantic meeting of funding agencies?