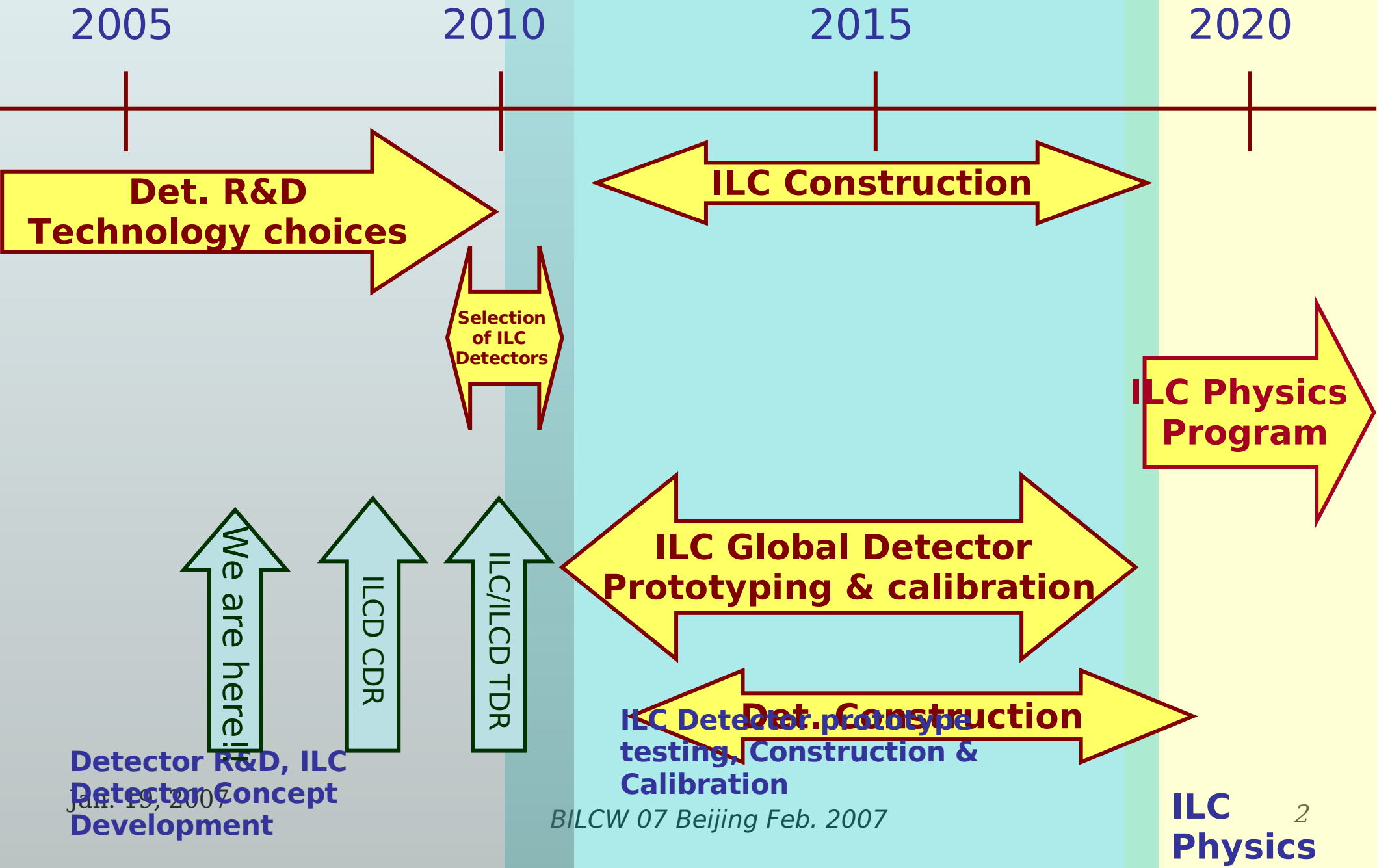
CALICE Data Processing
(From Raw to Physics Data)

Roman Pöschl
LAL Orsay

- The Calice Collaboration
- Calice Testbeam Data Taking
- Data Management
- Event Building and Reconstruction Software
- Pros and Cons ...
- Monte Carlo Tools
- Summary and Outlook

LC Detector Time Line

J. Yu
IDTB 07



Organization of R&D efforts?

Lesson from previous slide

Challenging time line to perform R&D to achieve

- a well motivated choice of detector technologies
- identify and iron out (technical) weaknesses and shortcomings of detector proposals
- Collect and understand data to tune e.g. MC models needed for high precision measurements at the ILC

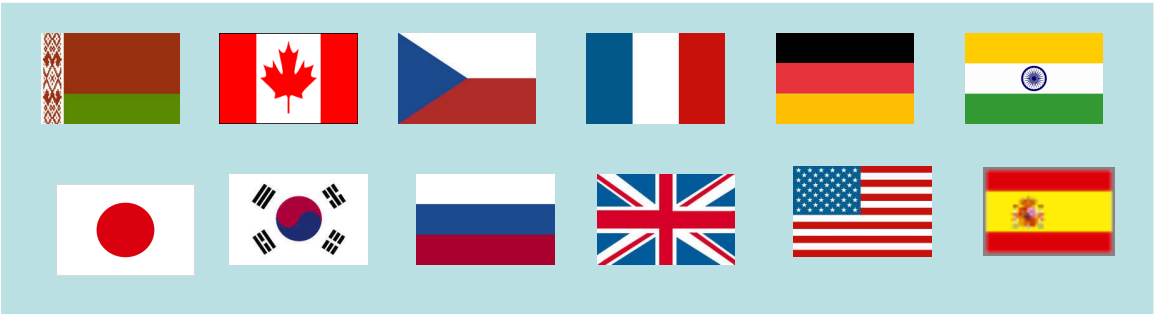
Best way:

Collaborative approach already in the R&D Phase
beyond 'borders' of different concepts

The CALICE Collaboration

Calorimeter for ILC

Calorimeter R&D for the ILC



~200 physicists from 12 Countries 3 Regions

- Integrated R&D effort

- Benefit/Accelerate Detector Development due to common approach

Projects within Calice

First generation prototypes

- W-Si **ECAL** almost complete, in use in testbeam
  (European Project)
- W-Scintillator strip **ECAL** in construction,
  test beam@DESY , **Spring 2007**
  (Asian Project)
- Tile **HCAL** with SiPM (MEPHI/Pulsar) r/o largely ready
  and in use in testbeam

- Digital **HCAL** in plan
  (Advanced) Effort in North America
  recent start up of European Project

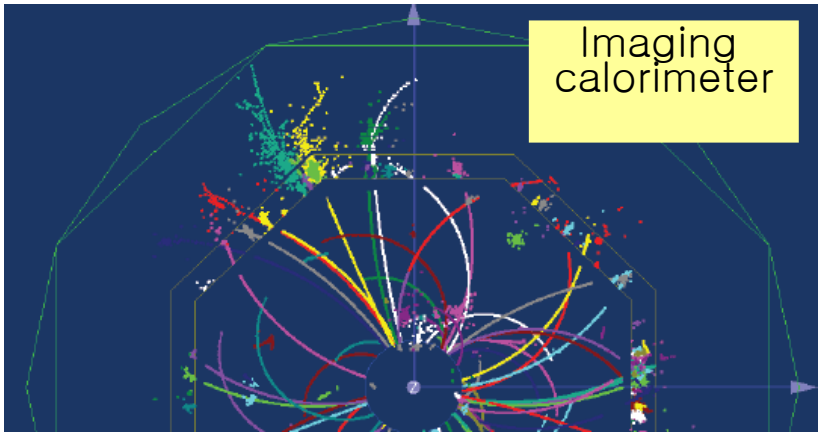
Projects benefit from

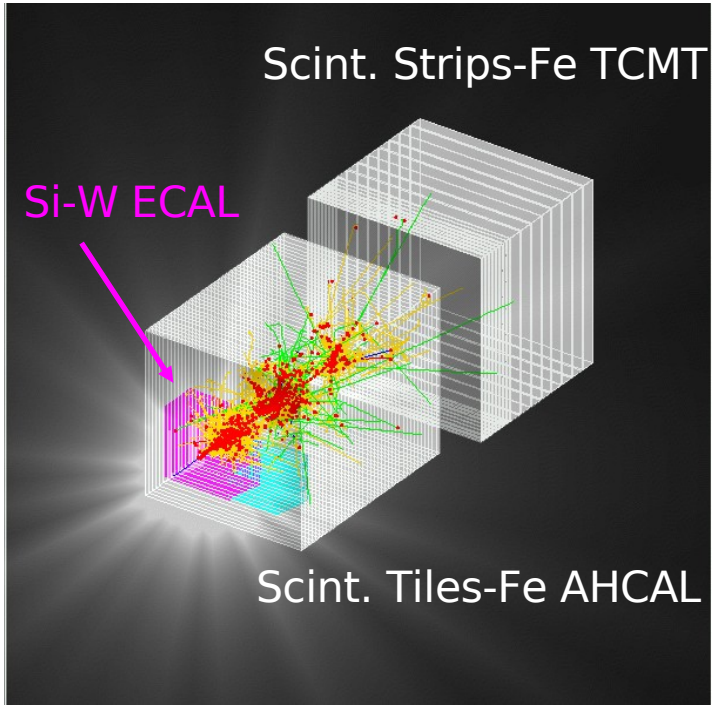
Common DAQ
Common Software
Common infrastructure, e.g. DESY testbeam
Common testbeam planning

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





Intermediate task:

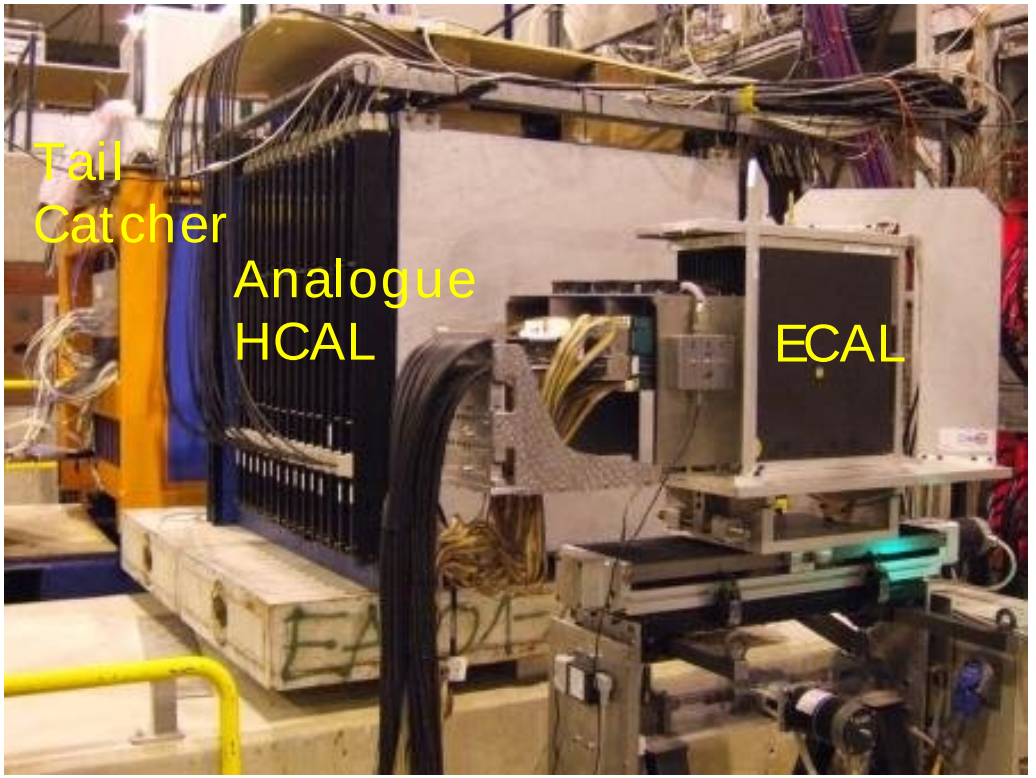
Build prototype calorimeters to
- Establish the technology
- Collect hadronic showers data

with unprecedented granularity to
- tune clustering algorithms
- validate existing MC models

CALICE Testbeam Data Taking

CALICE collaboration is preparing/performing large scale testbeam
Data taking between 1.August and 31.October 2006

Testbeam Setup at CERN



Testbeam program poses software/computing "challenges"

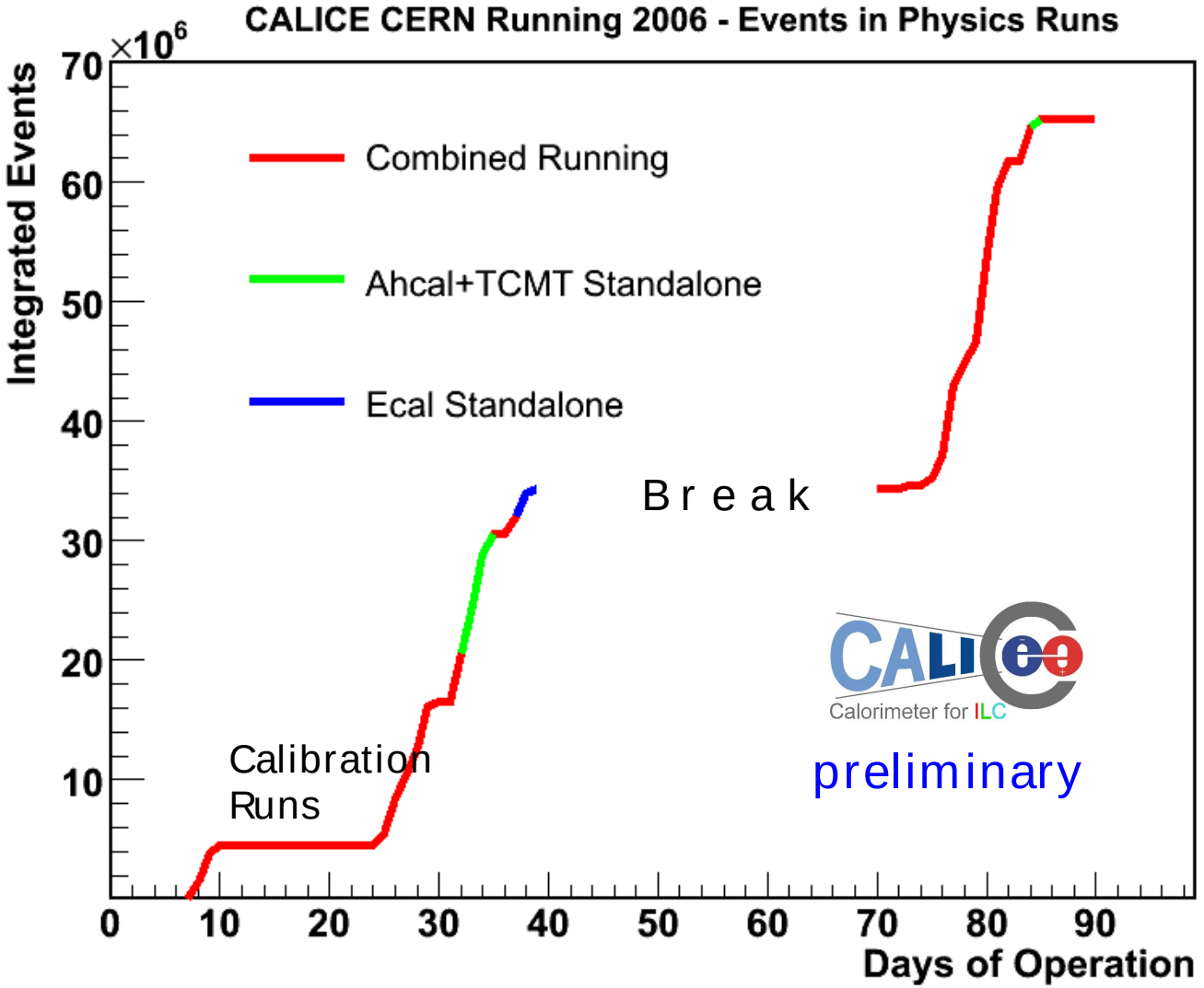
- Data processing from Raw Data to final Clusters in a coherent way

- Handling of Conditions Data Detector Configuration Calibration, Alignment etc.

- Comparison with simulated data
'Physics' Output

O(15000) calorimeter cells readout by Calice DAQ
No Zero Suppression
r/o speed 5 Mbyte/s continously

CALICE - CERN Data taking 2006



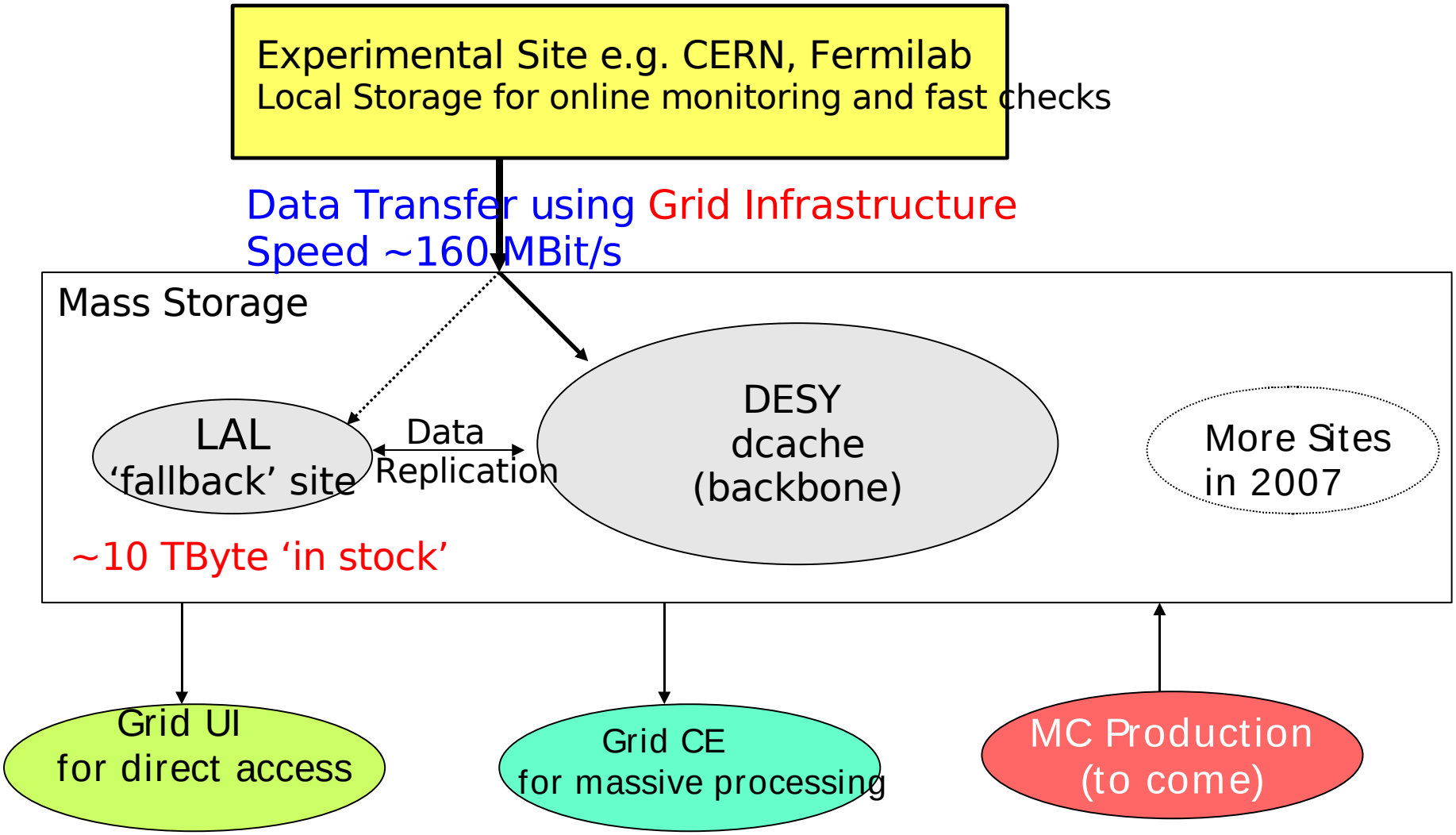
~65 Million Events in 'Physics' Runs

+

O(35 Million) Muon Calibration Events)

Efficient and fast way of data distribution and processing ?

Data Handling and Processing



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

Why using the Grid ?

Neither ILC nor Calice nor other R&D projects
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 iterations for calice testbeam data

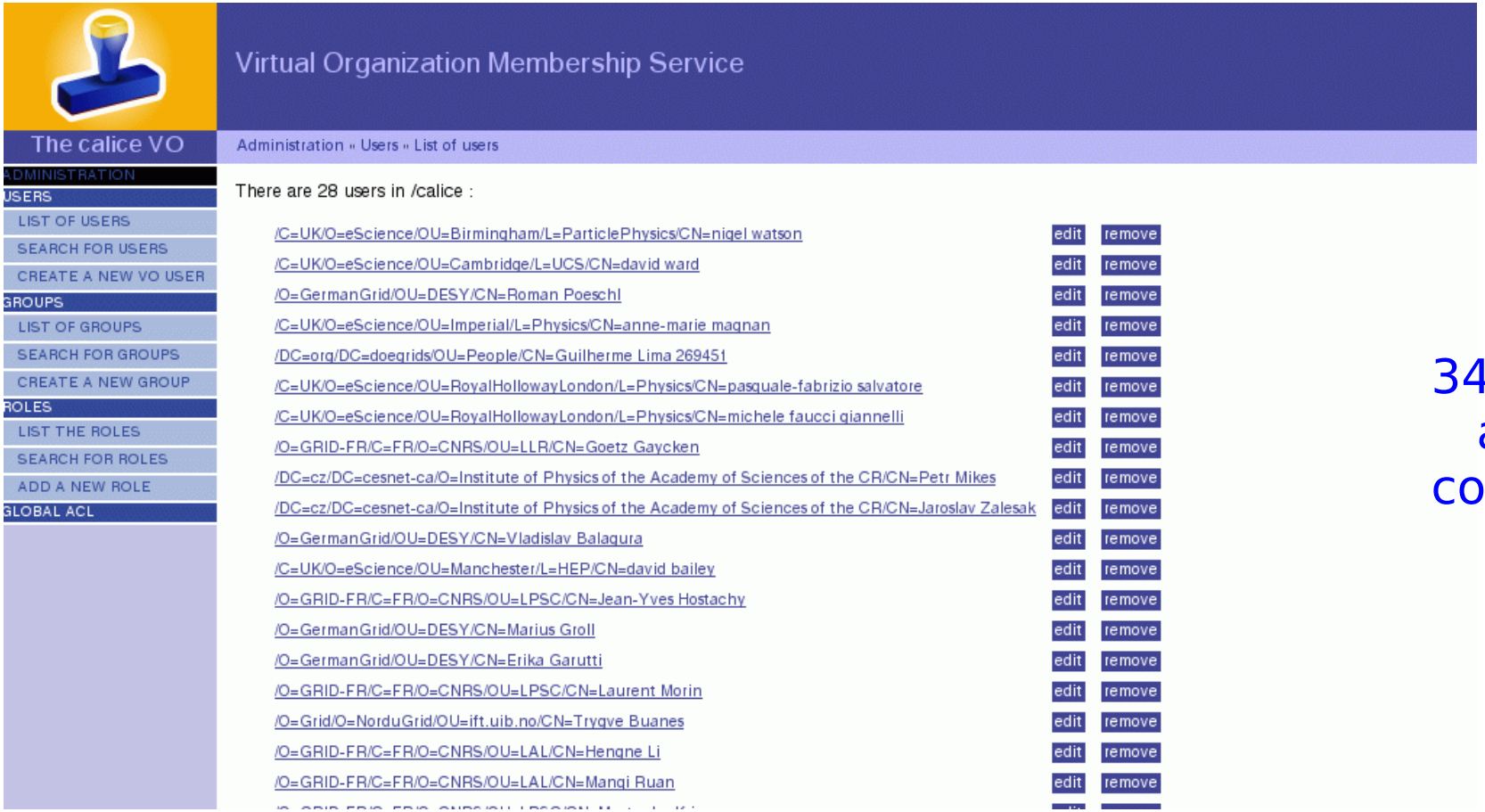
- Large simulation effort to come for the ILC requires large computing ressources
  Again no experimental center, the potential experimental centers like DESY, CERN Fermilab et al. have identified the Grid as computing platform
  General strategic decision by HEP community and science politics to invest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

Hosted by DESY:
Page for registration is https://grid-voms.desy.de:8443/voms/calice



34 Members
and
counting ..

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

The Grid in/for Calice

Large Data Volume => Significant Computing Ressources required
Decentralized Organization <=> Decentralized Computing

Virtual Organization calice

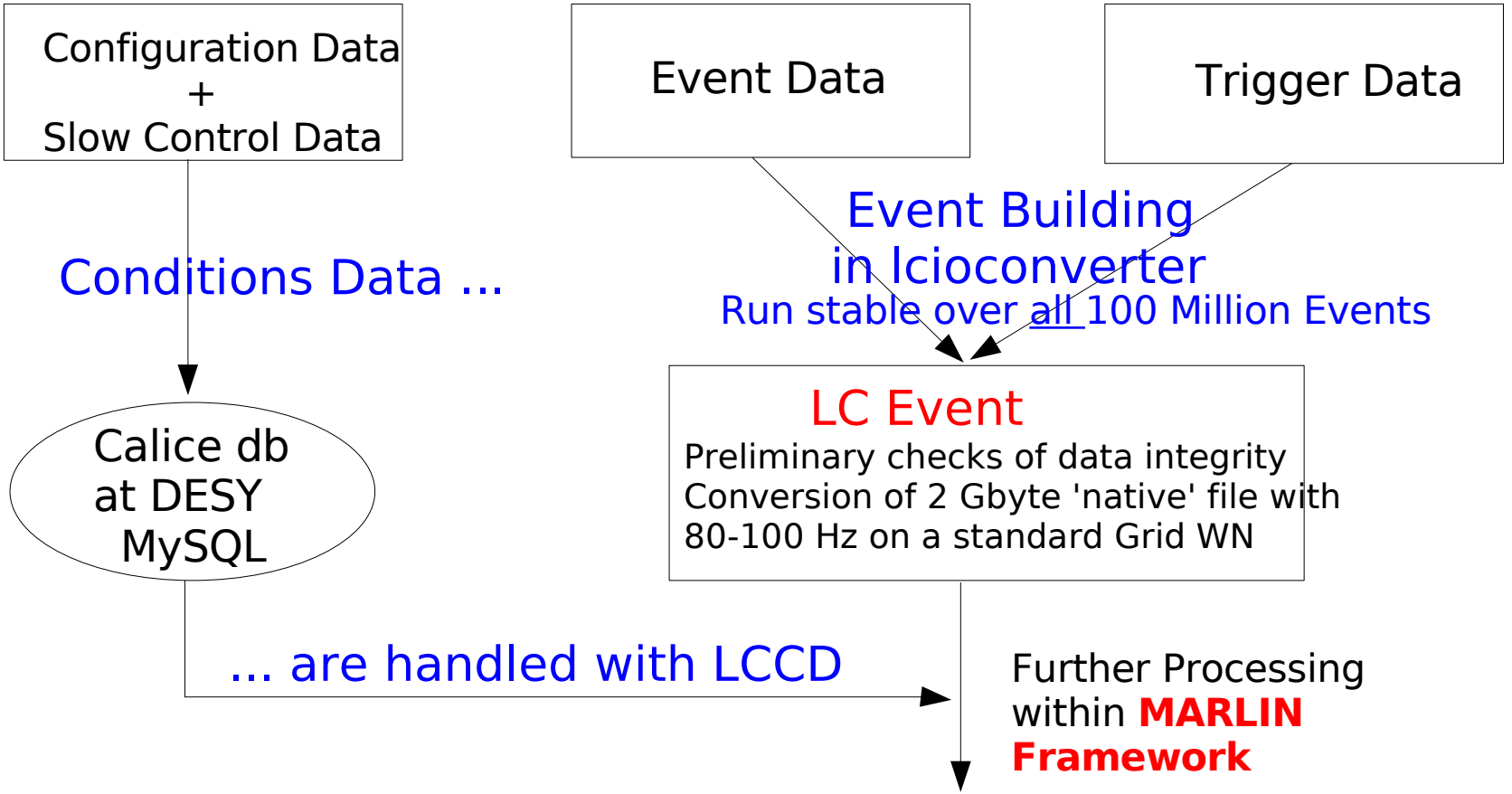
Supported by:	DESY Hamburg	Hosting, Computing and Storage
	LAL	Computing and Storage
	LLR	Computing and Storage
	DESY Zeuthen	Computing and Storage
	Imperial College	Computing and Storage
	cc in2p3 Lyon	Computing and Storage
	Cambridge	Computing and Storage
	Institute of Physics	Computing and Storage
	Prague	(in preparation)
	University College	Computing and Storage
	KEK	Computing and Storage
		(In preparation)
	Manchester	Computing and Storage (in preparation)
	CIEMAT Madrid	Computing and Storage
	Fermilab	Offer Received
	Univ. Regina	Offer Received

Acknowledged EGEE project: https://cic.in2p3.fr

Conversion to LCIO

DAQ data types are converted/wrapped into LCIO on the basis of LCGenericObjects

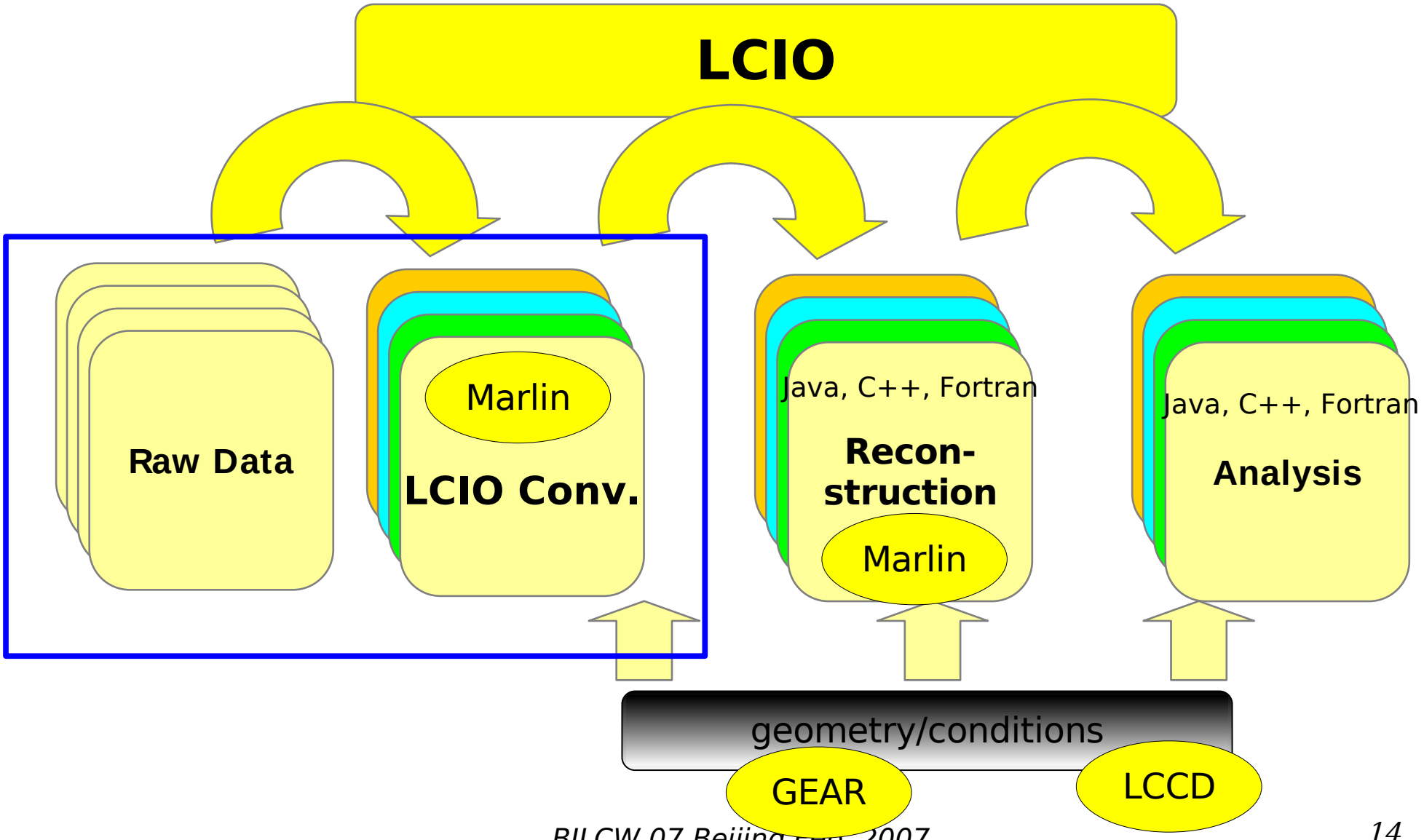
DAQ Data Files/Types



Remark: LCIO and ILC software framework is not needed to analyze calice data but using it delivers important input for future ILC s/w development
-> General ILC Concept for low level data handling

Software Tools for the LDC Study

Originally developed for simulation studies but applied in CALICE TB effort



Conditions Data Handling

- LCCD – Linear Collider Conditions Data Framework:
  - Software package providing an Interface to conditions data
    - database
    - LCIO files

  Author Frank Gaede, DESY

LCCD works and is heavily used within calice

The importance of conditions data (not only) for 'real' data renders the development of a fully functional cd data toolkit to be a fundamental !!! piece of the ILC Software
LCCD is first attempt into that direction

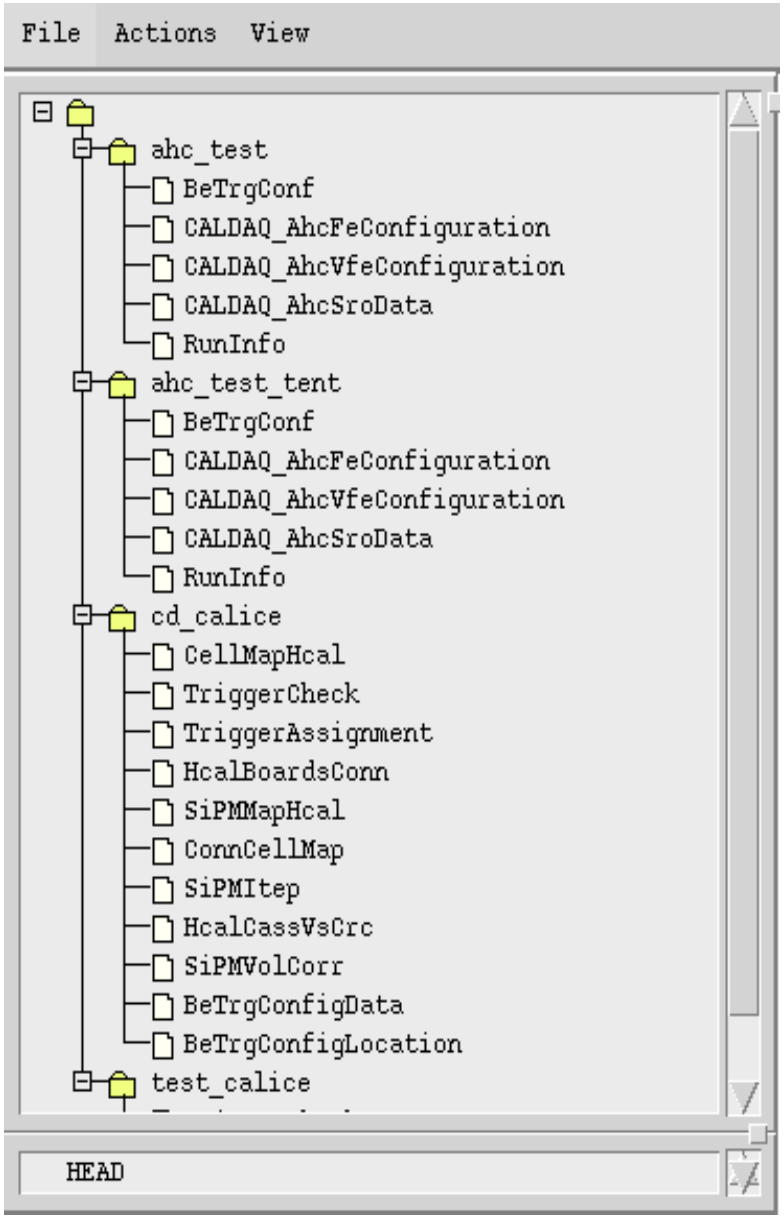
Issues to be addressed:
- Type safety
- Efficient storage and access to conditions data

  Browsing, convenient interfaces
- How to 'distribute' conditions data (e.g w.r.t to grid) ?

  BTW.: LHC does have some headache with that!

Testbeams are ideal environment to develop a working Conditions Data Handling before ILC starts

CALICE Database Hosted by DESY



First attempt to visualize Conditions Data
(S.Schmidt, M.Schenk, R.P.)

Trigger Info: Assignment of triggerbits
Trigger Configuration
Info to validate Trigger information

Calibration Data

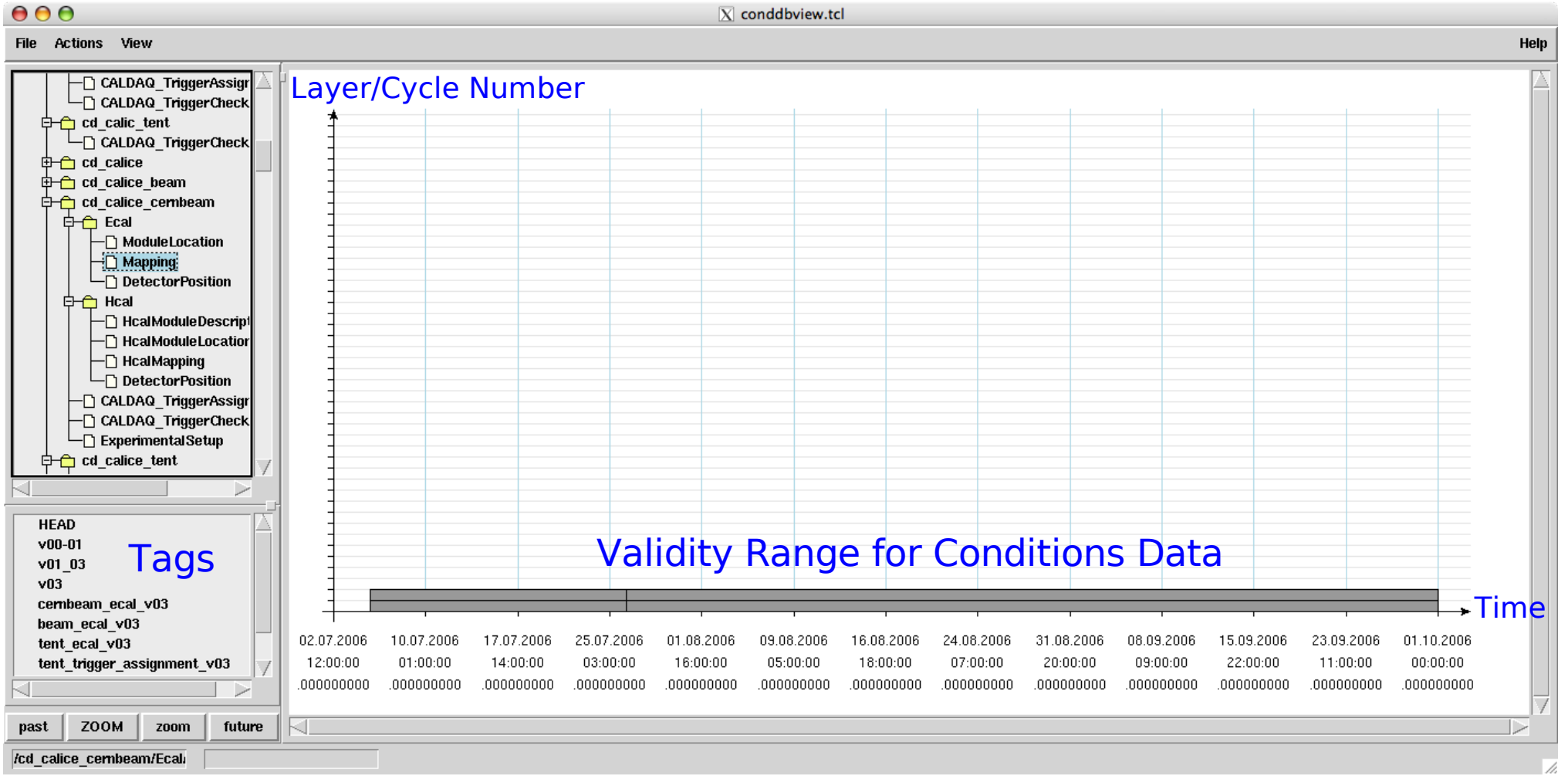
Cell Mappings: Relation electronic channel and
geometrical channel
i.e. Cabling of devices

Hardware configuration during data taking.

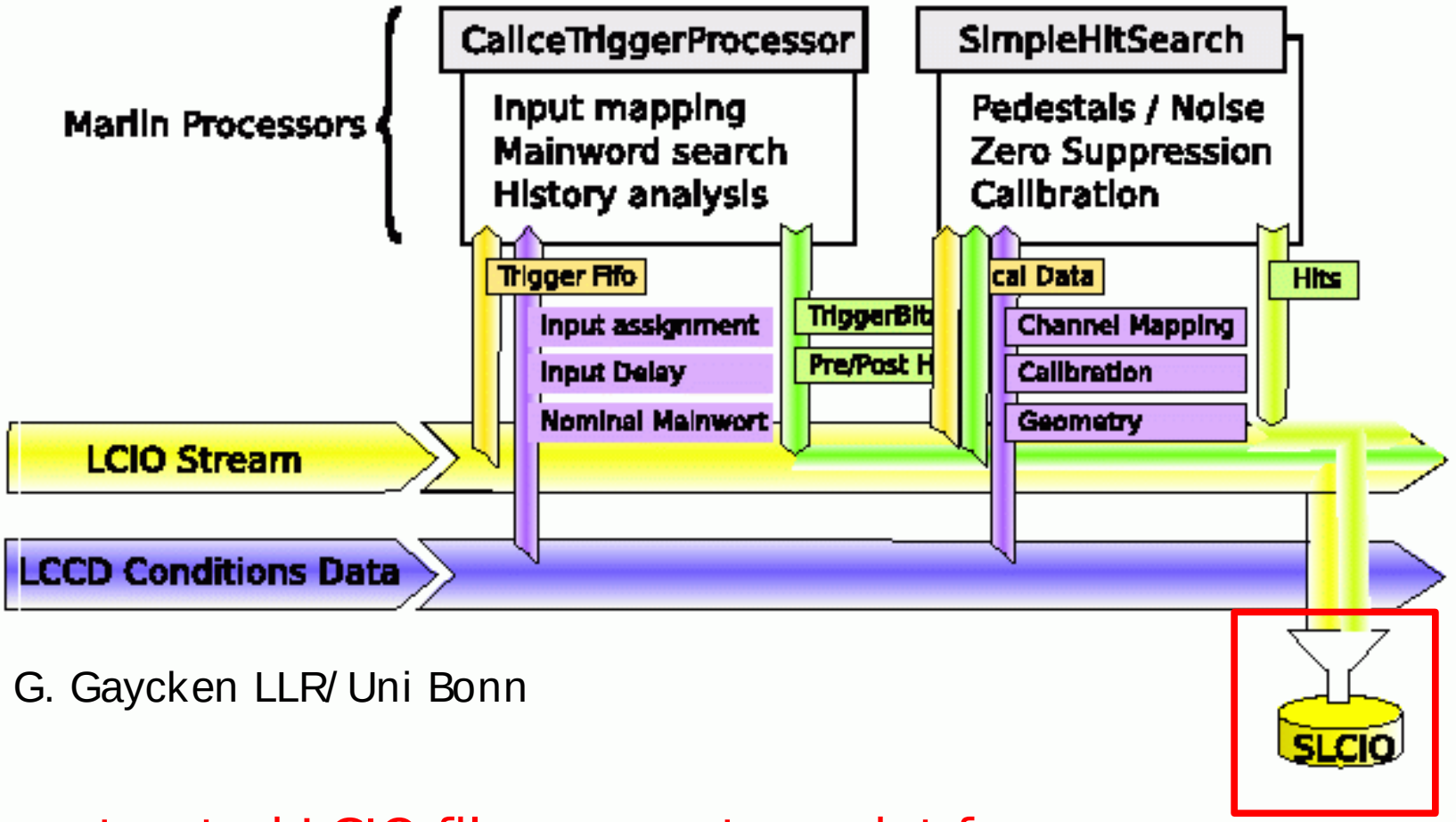
CondDBMySQL package of Lisbon ATLAS Group
Validity Time stamp and tagging

Access via LCCD Interface

Conditions Data in CALICE Database



Data Processing and Reconstruction



G. Gaycken LLR/ Uni Bonn

Reconstructed LCIO files are entry point for newcomers

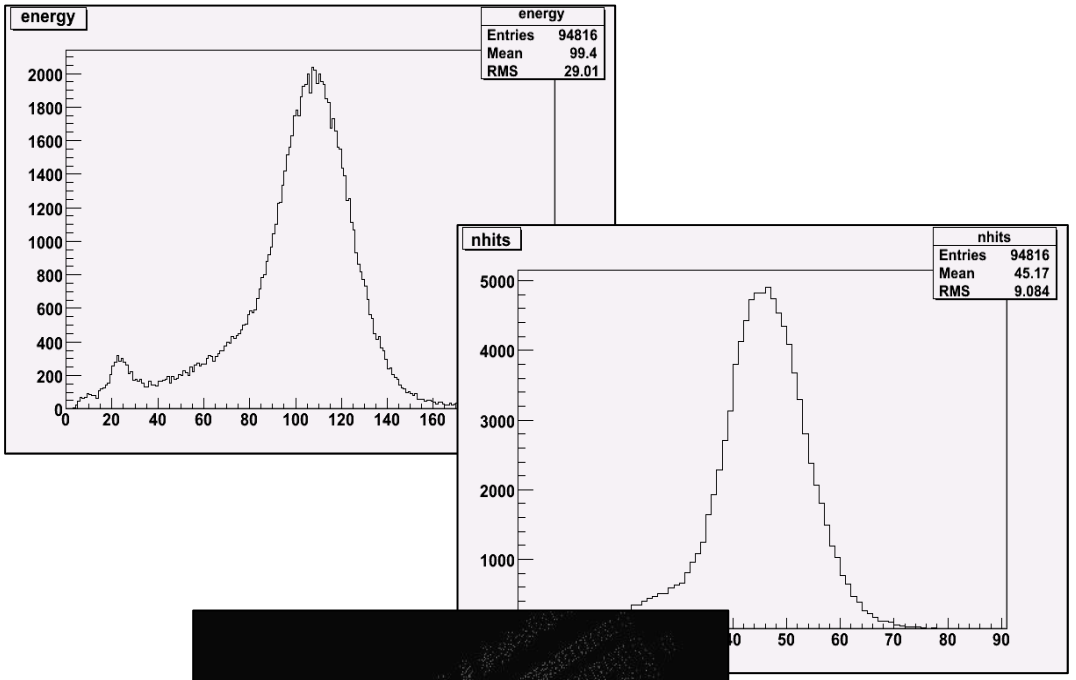
... and starting point of high level analysis

Contain 'familiar' CalorimeterHits

Though not the whole story – Still have to understand fundamentals

AHCAL Reconstruction Chain – S.Schmidt/DESY

- MappingI
  - ADCBlocks → CaliceHits1
- PedestalCalibration
  - CaliceHits1 → CaliceHits2
- GainCalibration
  - CaliceHits2 → CaliceHits3
- InterCalibration
  - CaliceHits3 → CaliceHits4
- SaturationCorrection
  - CaliceHits2, CaliceHits4 → CaliceHits5
- MIPCalibration
  - CaliceHits5 → CaliceHits6
- MappingII
  - CaliceHits6 → CalorimeterHits



Calibration steps modularized in MARLIN processors

Pros and Cons using ILC Software for (Calice) Testbeam Data

Pros

Benefit from existing tools/features for/of ILC Software
e.g. LCEvent allows to gather information on event

Newcomers can work in one software framework for testbeam and physics studies

Define at an early stage of the ILC R&D the needs for a complete data processing
Coherent s/w concept at time of ILC Detector TDR
Not just guesswork!!!

Cons

Need to wait for converted files
No quick turnaround in particular during development of DAQ and tests
Needs tight communication between DAQ and s/w developers

Overhead generated by usage of ILC Software
- Slower program execution?
- Profiling of ILC Software needed

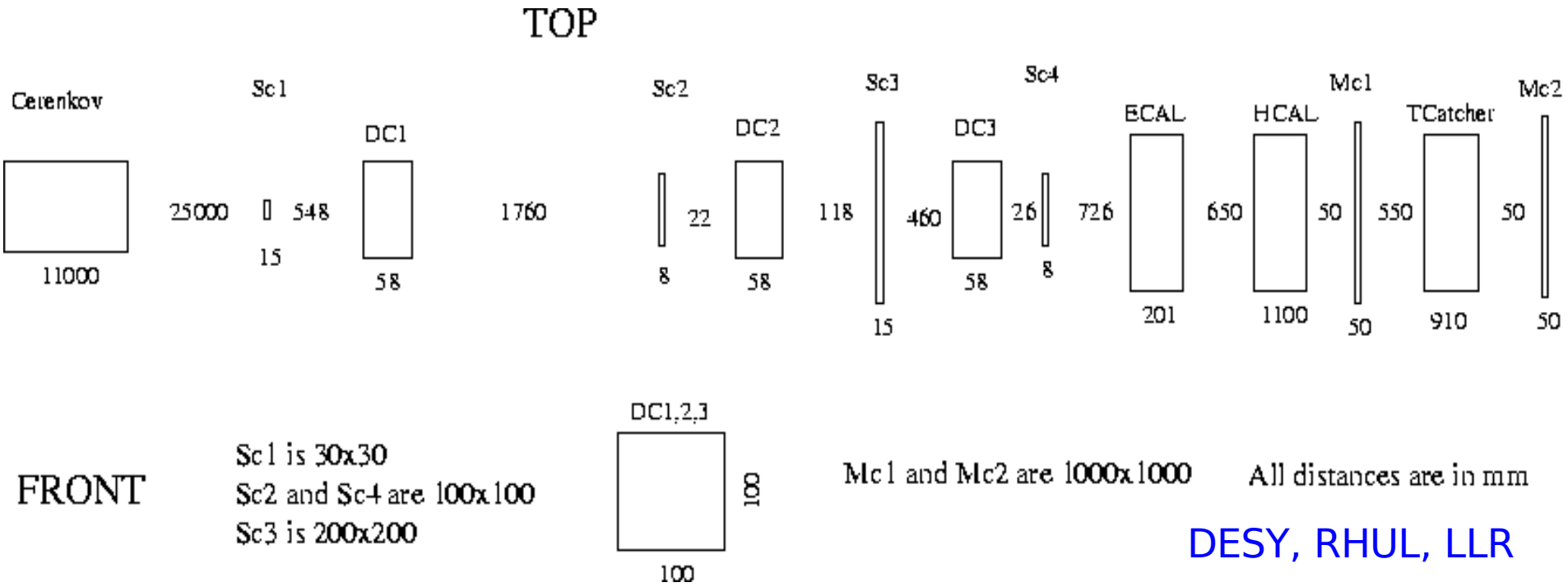
Source of (potential) errors unclear

It's in the spirit of the (LDC) CDR!!!!

BTW: The converted LCIO files can be analyzed on any OS (endianess) and on future 64bit architectures!!!

A view to the Monte Carlo Branch

- Model for the simulation of the CERN test beam is available (in release 06-02 of Mokka)



DESY, RHUL, LLR

Will use grid for MC production

Estimation ~ 5000 kSI2kd for simulation of CERN data

Simulation will be followed by a digitisation step

Realized as Marlin Processors within Digisim Package
A.M Magnan, G. Lima

ILC Detector Testbeam Workshop - IDTB 07

No dedicated talk at this (ACFA) workshop!?

- Workshop: 17/1/07 – 19/1/07 at Fermilab

- Aim of the workshop

  Overview on status of the activities of the various R&D groups/collaborations for the ILC

  Overview on tentative sites for testbeam measurements with some focus on Fermilab facilities

  Identify the future needs of these activities and outline a Roadmap

- ~120 participants from all areas of detector R&D in discussion with maintainers and leaders of the potential test facilities

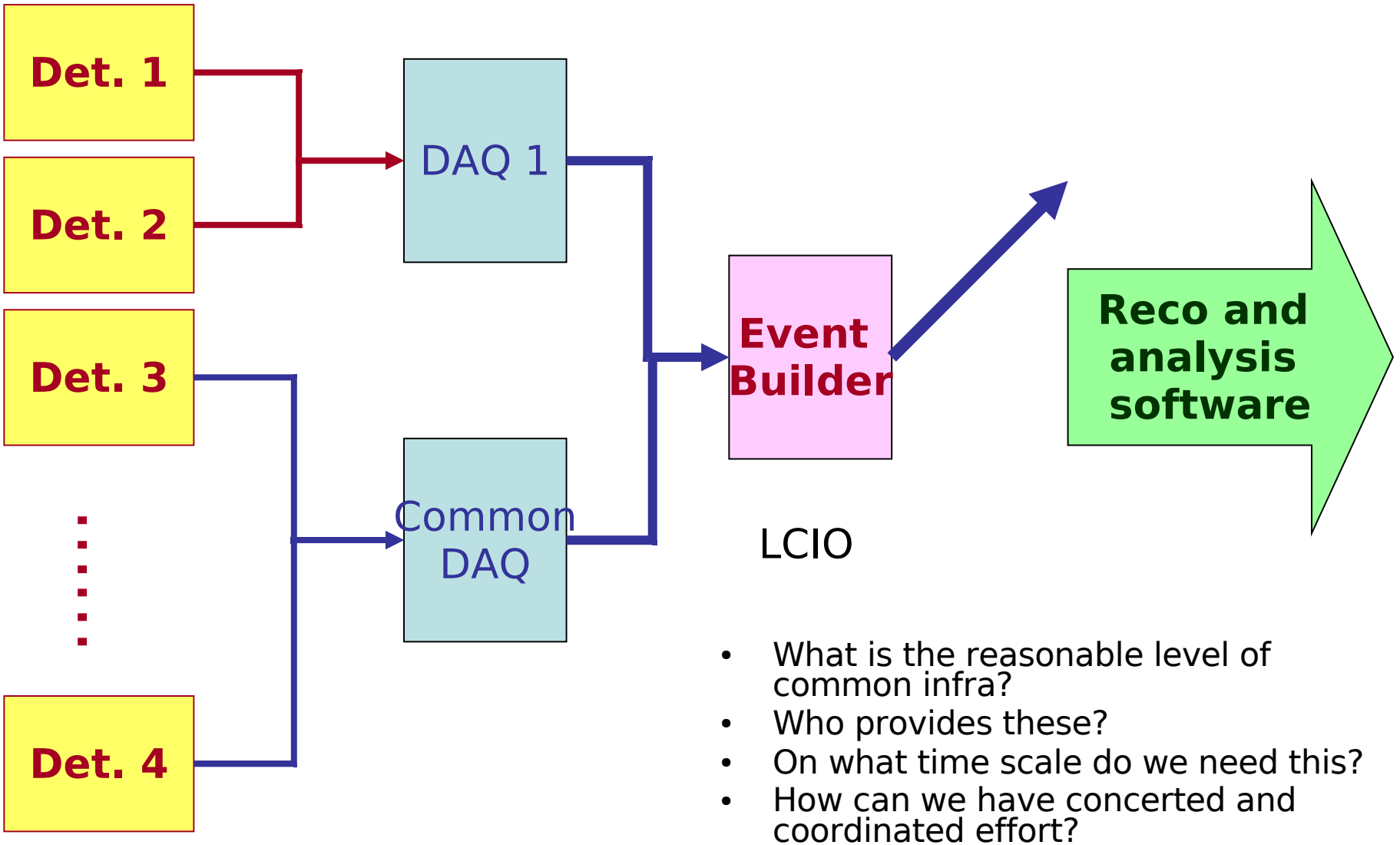
- Starting gun for the write-up of a roadmap document

Notable requests at the workshop

- Large bore, high field magnet (up to 5T)
  - VTX and tracking groups
- ILC beam time structure (1ms beam + 199ms blank)
  - VTX, TRK and CAL electronics
- Mimicking hadron jets
  - VTX, TRK and CAL

- Common DAQ hardware and software
- Common online and offline software
  - Reconstruction and analysis software

Point of Merge for Commonality



- What is the reasonable level of common infra?
- Who provides these?
- On what time scale do we need this?
- How can we have concerted and coordinated effort?
- Do we need this at all?

J. Yu IDTB 07

Summary and Outlook

- Calice uses European ILC Software for processing of Testbeam Data

  ILC Datataking in a (big) nutshell
  Very important input for current and future developments of ILC Software
  Allows for stringent tests of the ILC Software concepts on a 'living' beast

- Calice uses systematically Grid tools

  First (and only?) R&D project within ILC effort
  24h/24h 7h/7h during CERN testbeam
  So far mostly for data management
  CPU consumption still tiny but will grow fast when starting e.g. MC production
  Calice has benefitted a lot from close collaboration with and support by IT Divisions of DESY, CERN and LAL

- Experience with testbeam data clearly reveals the needs for a coherent concept to handle 'low level' data within ILC Software

  (Latest) Next generation R&D projects should be used to develop a complete data processing/handling strategy for the ILC.
  Avoid 'island' solutions and work on an integrated effort
  Point clearly raised at IDTB 07 workshop

CALICE does not only hardware-prototyping but also 'computing prototyping' for the ILC