

(Conditions) Data Handling in the CALICE Testbeam



Roman Pöschl
DESY Hamburg
CALICE Collaboration



Data Taking and Processing
in a (big) Nutshell

LCWS05 Stanford/USA March 2005

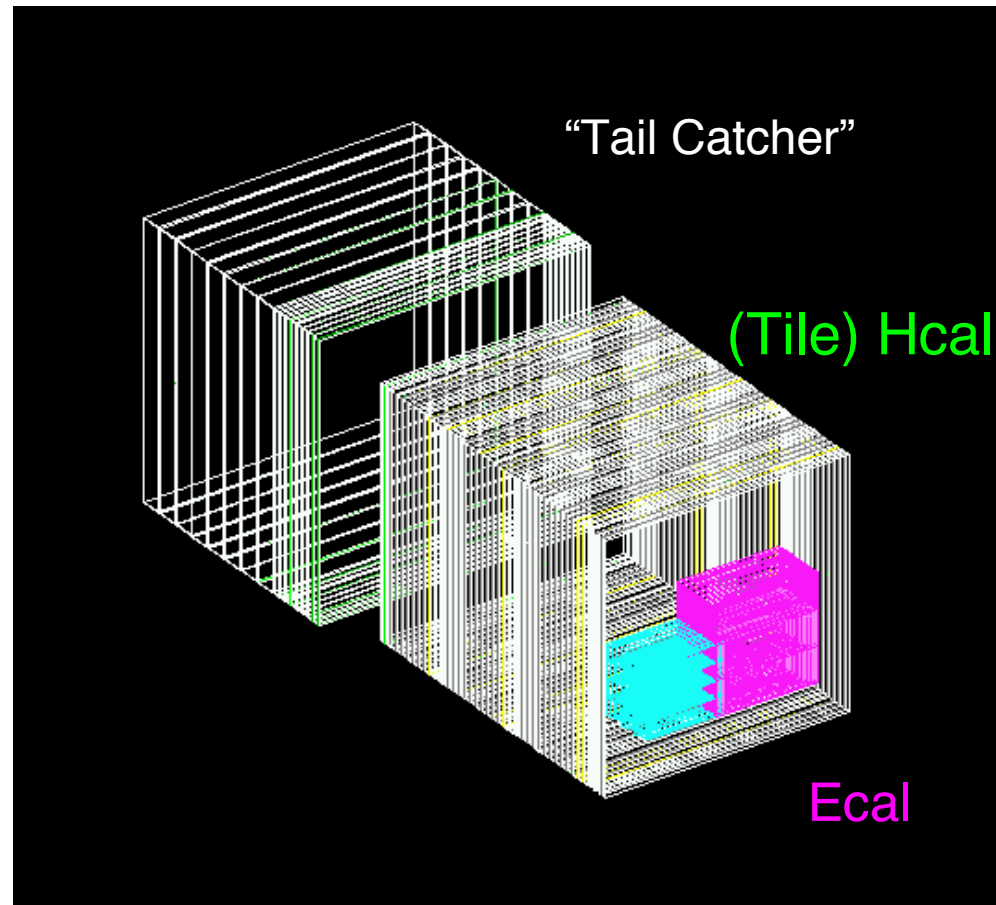
Introduction

CALICE collaboration is preparing/performing large scale testbeam
(See CALO session)

Primary sites DESY and FNAL

- O(20000) calo cells to r/o
~ HEP experiment of the 80's
- Data processing from
Raw Data to final
Clusters in
user friendly way
- Handling of Conditions Data

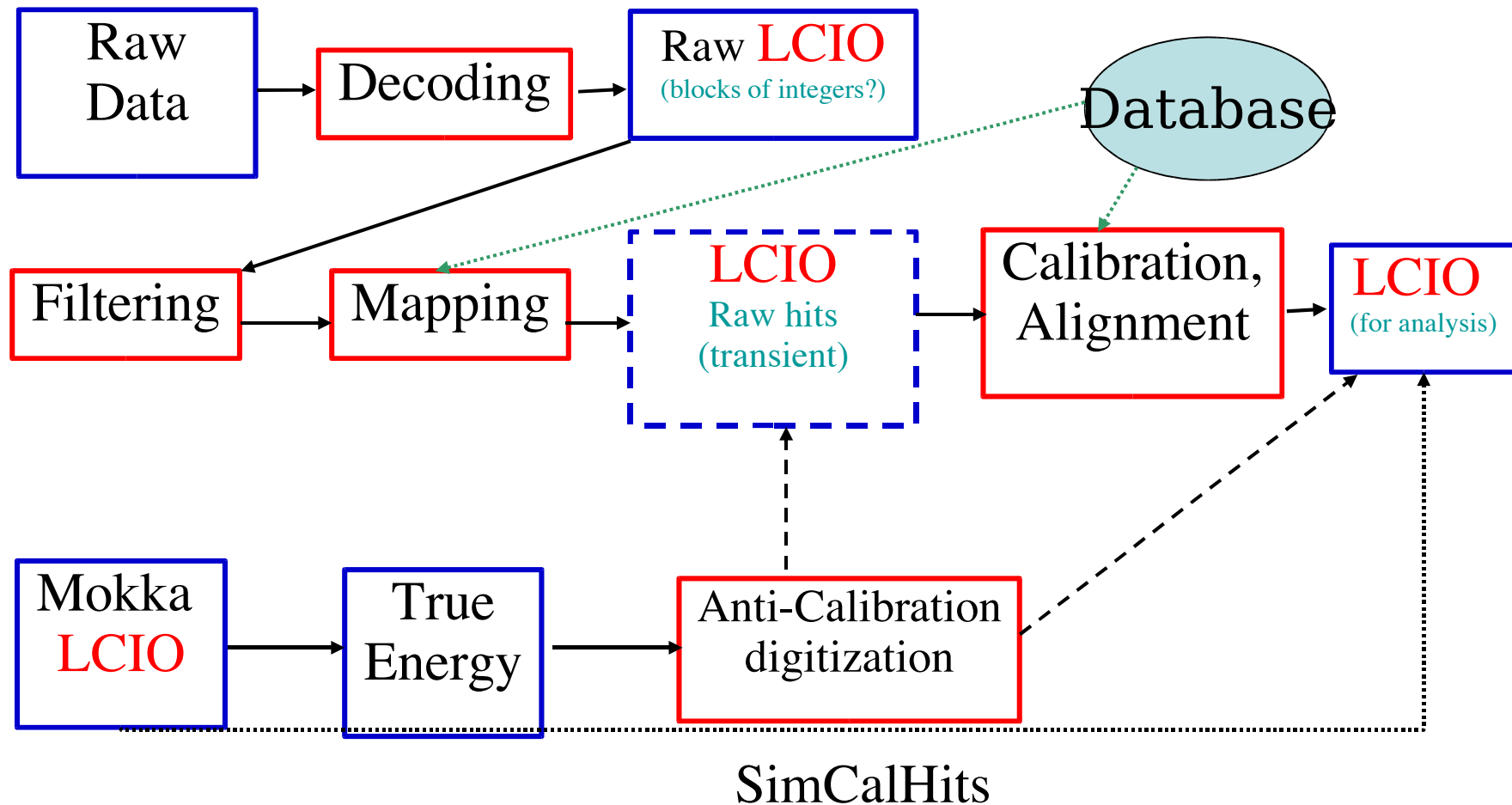
What happens after
the data have
tumbled out of the Detector ?



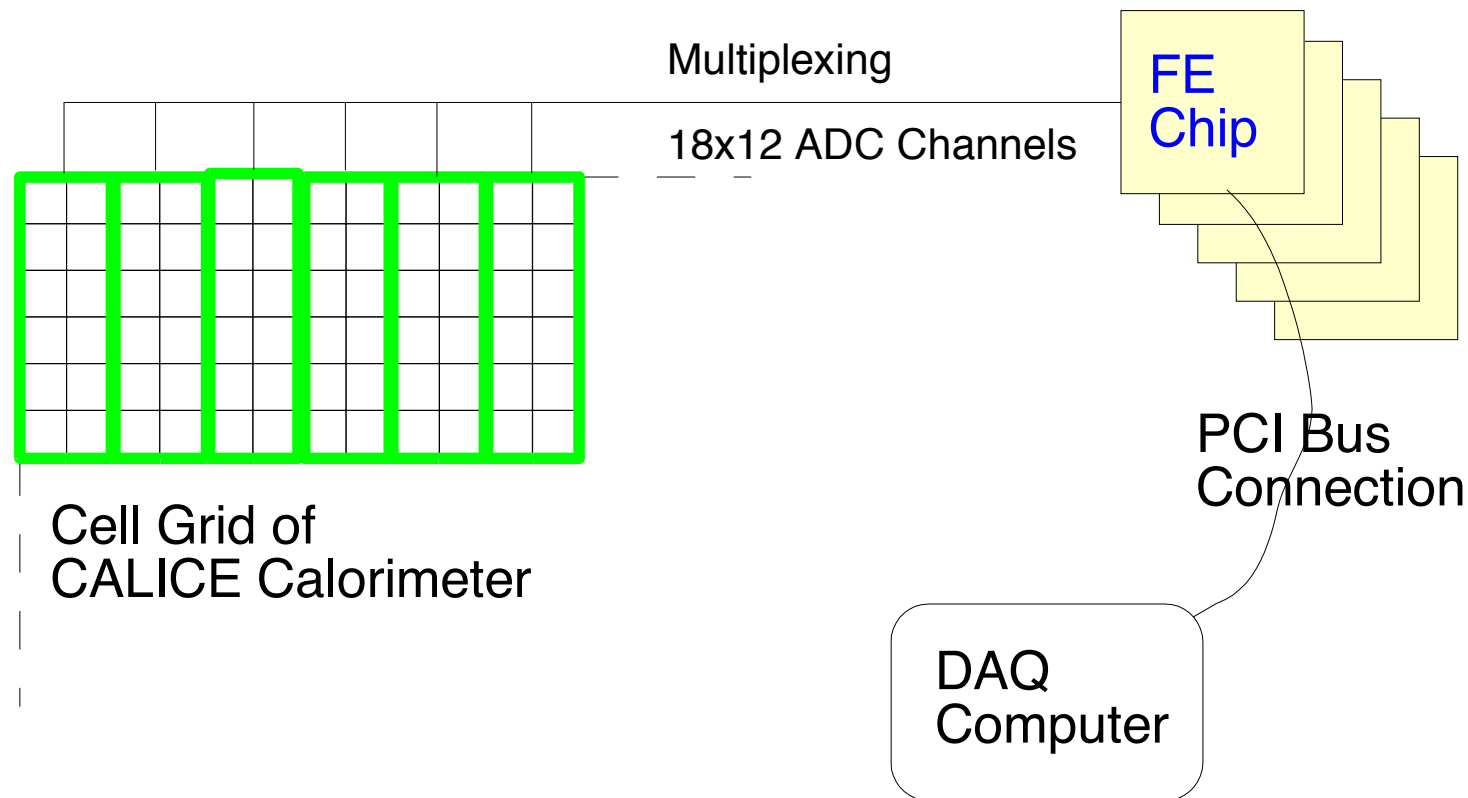
Simulated testbeam setup

Dataflow in CALICE Testbeam

LCIO as backbone of Testbeam Analysis



CALICE DAQ Scheme – Poor Man's View



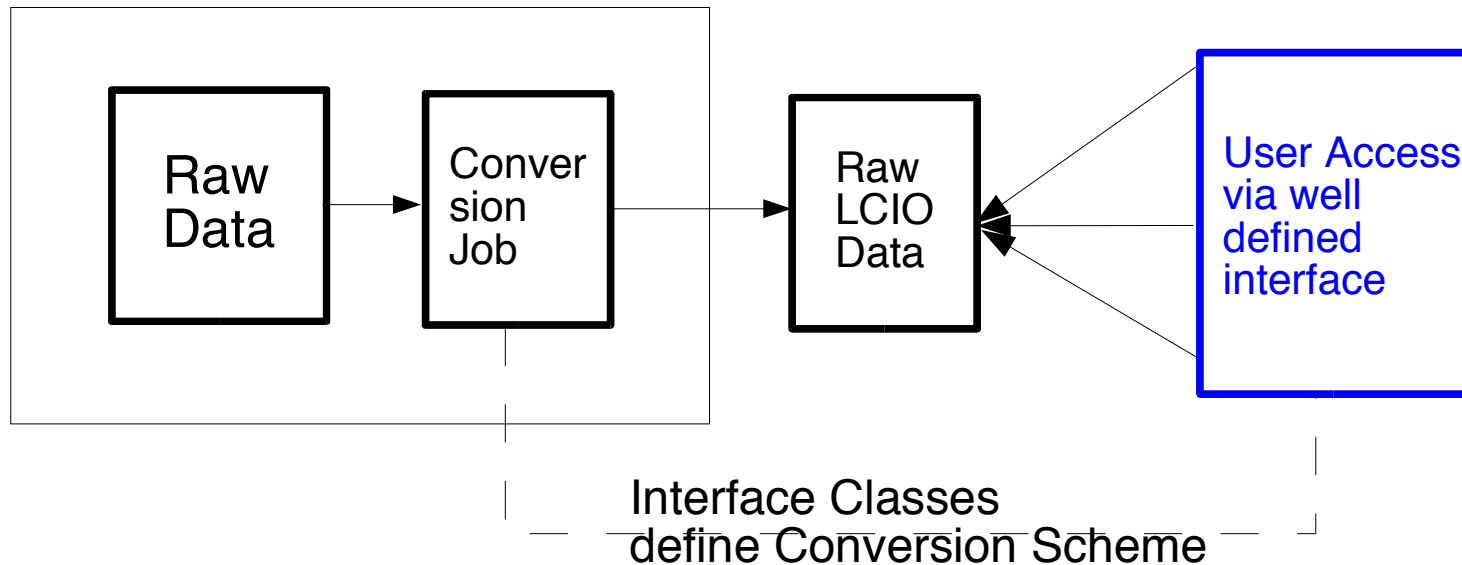
- DAQ is organized 'hardware friendly'
Data received and stored as sequence of 32 bit integers
Needs expert knowledge to analyze the data
- 'Many' people should get involved in calibration/monitoring
- Provide data in 'user friendly' format

Raw Data Conversion to LCIO

Collaboration: G. Mavromanolakis, D. Ward, P. Dauncey, F. Gaede, R.P.

Main Strategic Decision: Raw Data should be available in LCIO Format

Requirements: - 'Intelligent' Conversion from Raw Data to LCIO Raw Data
- Provide all Info on Raw Data also in LCIO Raw Data
in a user friendly way

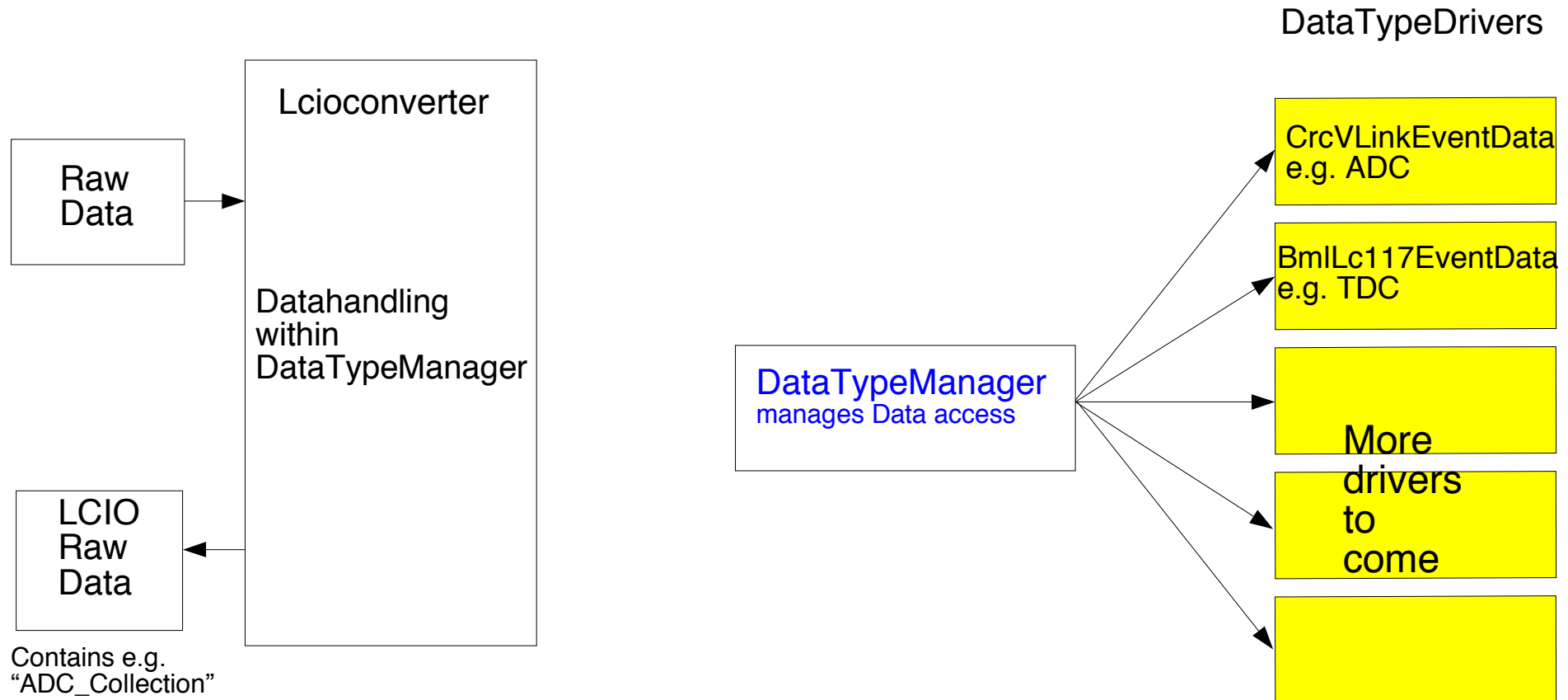


Interface is to be completely decoupled from online software

Dedicated Interface Classes are defined using LCIOGenericObjects

Structure of the LCIOConverter

Documentation: http://www-flc.desy.de/store/hcal/simsoft/calice_soft/lcioconverter/doc/



Conversion happens in a modular way
Drivers for new datatypes can be added w/o touching
existing (working) code
Drivers use interface classes for conversion

Data Storage Issues

4 weeks of CALICE Ecal data taking ~ 250 Gbyte (unconverted) data

- Infrastructure for mass storage setup at DESY (dCache Pool)
Roughly 5 Pbyte diskspace available
- Recommended access via Grid-Tools
Not very much experience gathered yet
Mastering this is essential for sharing of CALICE and ILC data
Efforts to establish a CALICE VO hosted by DESY (?)
- dccp tool installed on DAQ Computer in Control Room at DESY
⇒ Immediate copy of data into dCache possible
- dCache is common effort of DESY and Fermilab !
⇒ Similar 'storage' environment at Fermilab

See also: <http://www.dcache.org/>

Handling of Conditions Data

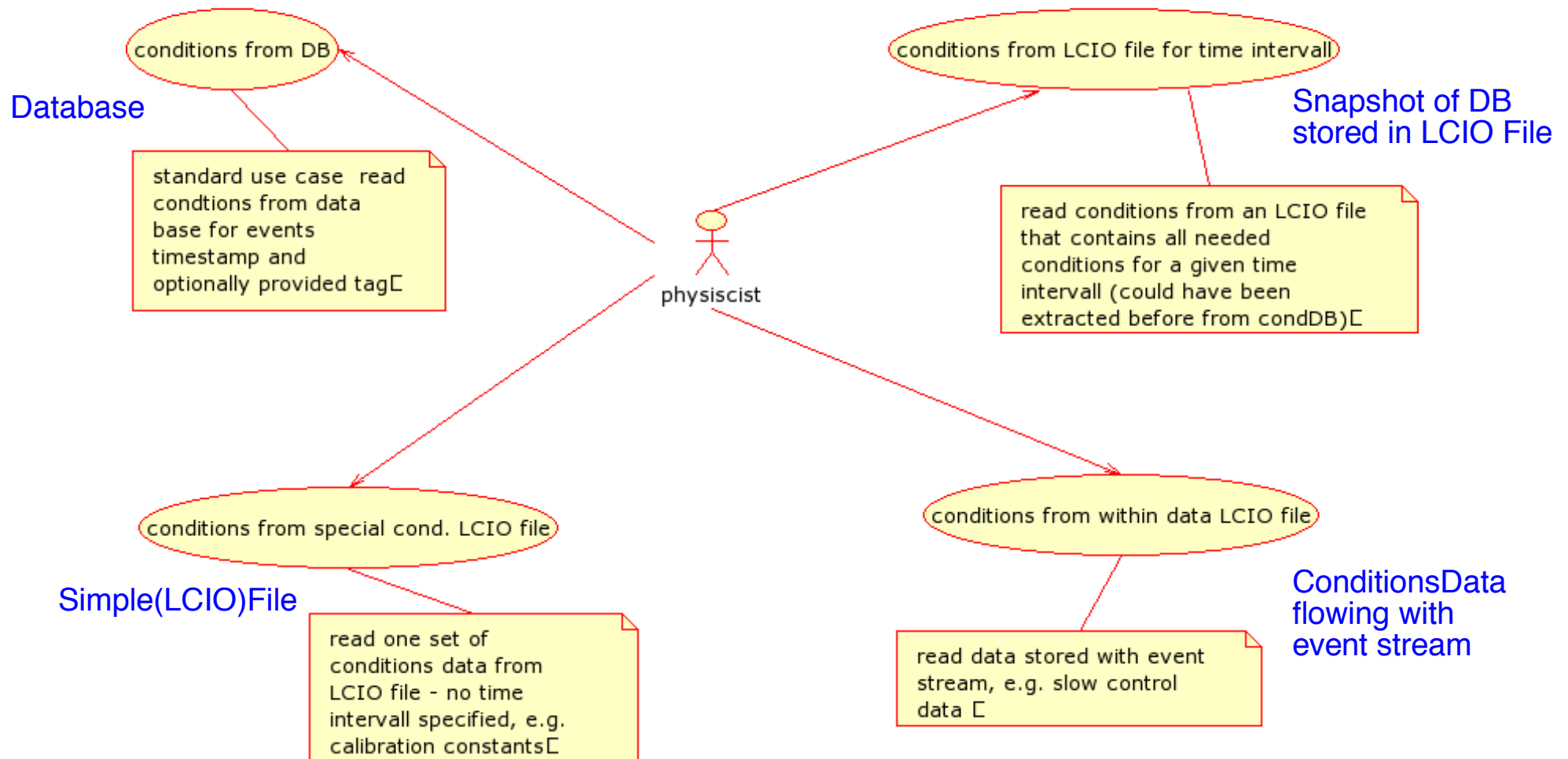
Application of LCCD for Testbeam Data Processing

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

Author Frank Gaede, DESY
- Conditions Data:
 - all data that is needed for analysis/reconstruction besides the actual event data
 - typically has lifetime (validity range) longer than one event
 - can change on various timescales, e.g. seconds to years
 - need for tagging mechanism, e.g. for calibration constants

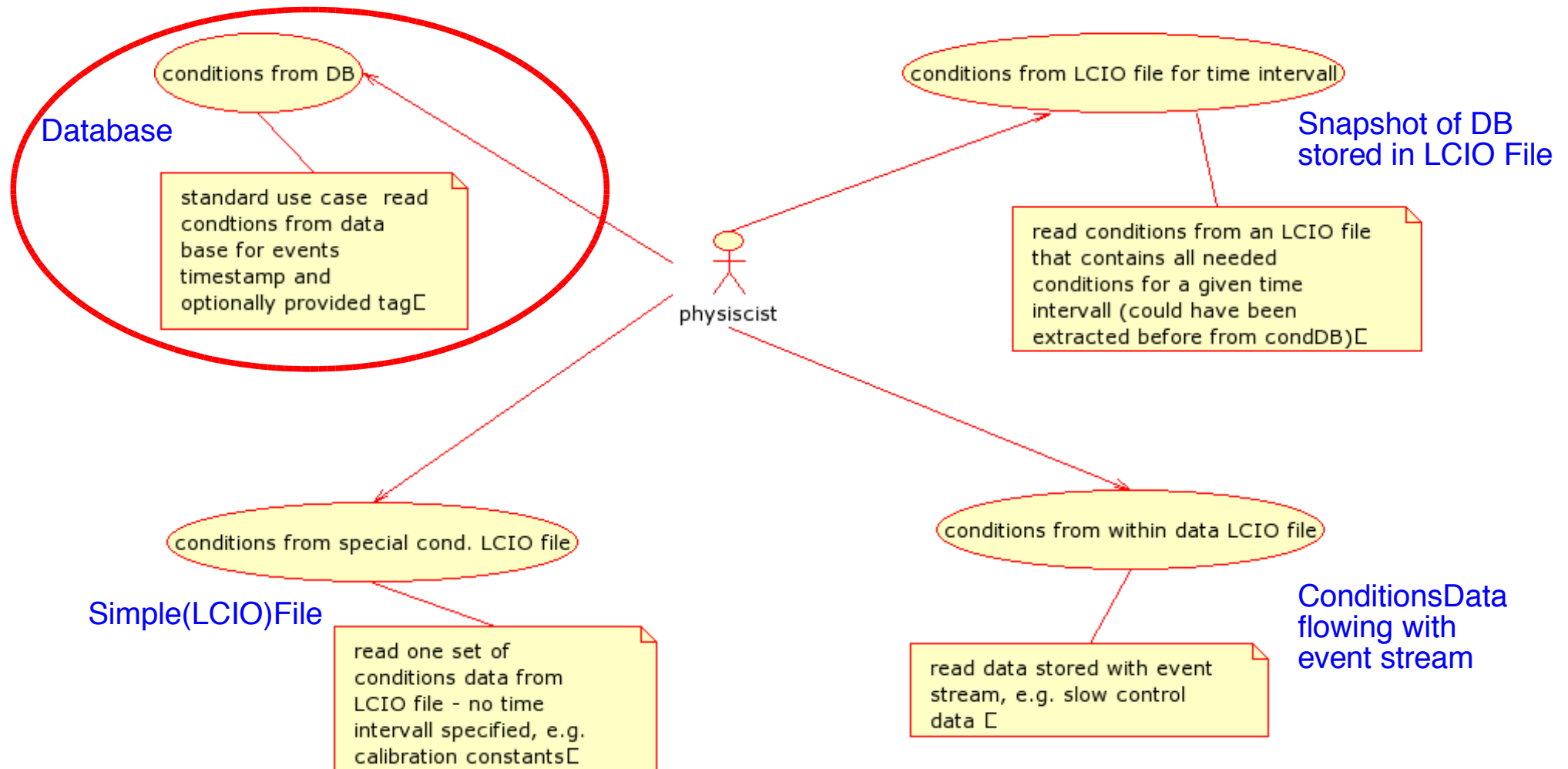
Sources of Conditions Data – Use Cases

LCCD Use Cases



Sources of Conditions Data – Use Cases

LCCD Use Cases

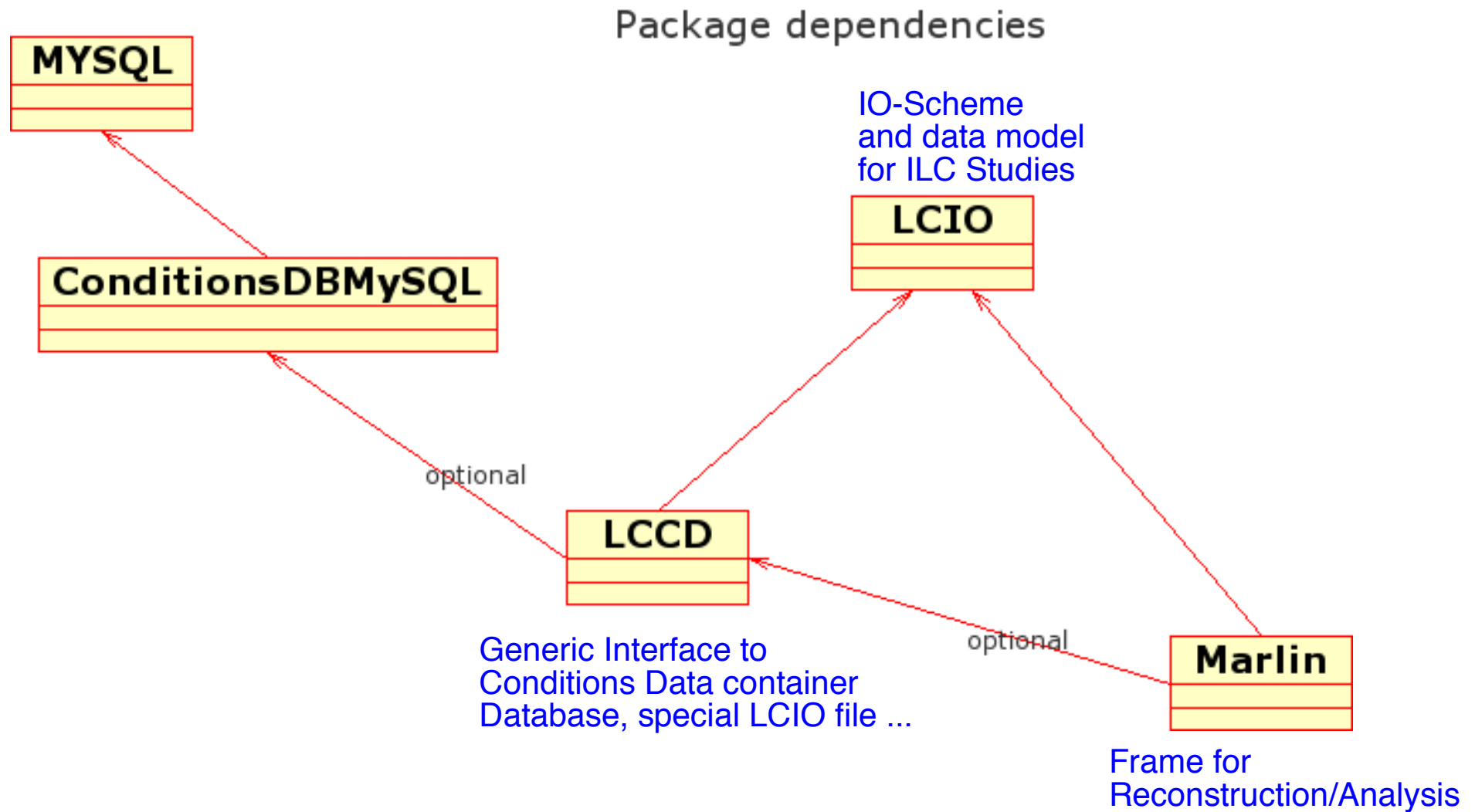


ConditionsDBMySQL – Overview

Interfaced to LCCD by Frank Gaede

- Open source implementation of CondDB API
 - Conditions data interface for ATLAS
- developed by Lisbon Atlas group
- features
 - C++ interface to conditions database in MySQL
 - data organized in folder/foldersets
 - objects stored as BLOBs (binary large objects)
e.g. LCIO objects or std::vector
 - tagging mechanism similar to CVS
 - scalability through partitioning options
 - outperforms implementation based on Oracle

Interplay between (ILC) Software Packages



Conclusion and Outlook

- CALICE Testbeam is first major data taking effort in ILC R&D
- Data Taking has started successfully at DESY
- Whole chain from data acquisition to physics results to be set up
- LCIO will be backbone of Data Processing/Analysis
 - Converter from Raw Data to LCIO exists
- LCCD is primary tool for handling of Conditions Data
 - Conditions data storage in MySQL database
 - LCCD allows for other means
 - Have to test it in 'real life'
- DESY dCache mass storage to be used for data storage
 - Exploit GRID tools for data distribution and access