CALICE Data Processing
(From Raw to Physics Data)

Roman Pöschl
LAL Orsay

- Calice Testbeam Data Taking
- Data Management
- Event Building and Reconstruction Software
- Summary and Outlook

CALICE Software Meeting 28/11/07
The Three Pillars of Calice Software

ILC Software  GRID  Database

This is the concept – All the rest are implementation details

In the following I will outline how these tools are employed and work together
CALICE Testbeam Data Taking

CALICE collaboration is preparing/performing large scale testbeam Data taking in Summer 2006/2007

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
CALICE “TIER 0” – Infrastructure in the Control Room

Gigabit Uplink
- High Speed Connection to the outside world
- Serves all Calice Control Room Computers

caliceserv.cern.ch
- Online Monitoring
- Grid Transfers

Disk Array

DAQ Computer

Well organized setup of computing
Thanks to B. Lutz

Picture courtesy of C. Rosemann DESY
CALICE - CERN Data taking 2006/2007

~200 Mio Events in 'Physics' Runs
+ O(50 Mio). Muon Calibration Events

Efficient and fast way of data distribution and processing?

Roman Pöschl
Data Handling and Processing

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

Data Transfer using Grid Infrastructure
Speed up to 240 MBit/s

Mass Storage

in2p3 Lyon fallback site

DESY dcache (backbone)

~16 TByte for 2007 2/3 binaries 1/3 converted files

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

Data Transfer using Grid Infrastructure
Speed up to 240 MBit/s

Mass Storage

in2p3 Lyon fallback site

DESY dcache (backbone)

~16 TByte for 2007 2/3 binaries 1/3 converted files

Grid UI for direct access

Grid CE for massive processing

MC Production started in UK

- Raw Data are (usually) available ~20 Min. after Run End
- Delay of Converted Files (usually) < 1 day

CALICE is the first HEP Experiment which uses the grid systematically for real data!!!
The Virtual Organisation - vo calice

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

VO Manager: R.P./LAL, Deputy: A. Gellrich/DESY
## Institutes which provide Grid 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  
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. Liverpool | Resources Provided (not yet exploited)  
Univ. Regina | Offer Received

- Most of the sites have been involved in recent data and MC processing  
Smaller Problems at Manchester and KEK (about to be solved)
Conversion to LCIO

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

DAQ Data Files/Types

- Configuration Data
  - + Slow Control Data
  - Conditions Data...

- Event Data
- Trigger Data

事件构建

进阶处理

在LCIO converter中进行事件构建
运行稳定，跨越所有3.0亿事件

LC事件
初步检查数据完整性

转换2G字节'原生'文件
80-100 Hz在标准Grid WN

条件数据...
...被LCCD处理

Remark: LCIO和ILC软件框架不被需要来分析Calice数据，但使用它提供了重要输入，未来ILC软件发展
- General ILC Concept for low level data handling

Roman Pöschl

CALICE Software Meeting

9
Calice Software
Three main packages

Contributions by groups from DESY, Imperial, LAL, LLR, NIU, RHUL

calie_lcioconverter ➔ calice_userlib ➔ calice_reco

Current version v04-02-06
converts calice DAQ format into LCIO (LCGenericObjects)
needs DAQ software expert work
MARLIN processors

Current version v04-10
Interface classes to LCGenericObjects (these classes should be defined by LCIO)
utility functions e.g. For TriggerHandling

Current version v04-06
RawData into CalorimeterHits (standard LCIO) TrackerHits
First stages of higher level analysis MARLIN processors

225 classes or functions
Data of four different Calorimeter Prototypes are available in LCIO format

Roman Pöschl  CALICE Software Meeting 10
Example for Data Processing - SiW Ecal

CaliceTriggerProcessor
common to all reco

Input Trigger Mapping

Main Word Search

Trigger History

SimpleHitSearch

Pedestal Calculation

Pedestal Corrections

Noise Calculation (for MC)

CalibrateAndApplyThreshold

Zero Suppression

Calibration

LCIO Stream

RawCaloHits

LCCaloHits

MC Branch:

TBEcalDigitisationProcessor

Smearing of Calo Hits

Adding of cells not appearing in MC

Conditions Data

LCIO File

Calib Const.

Noise. Const.

SimCaloHits

RawCaloHits

Cell Mapping Module /Cell Dimensions

ADC Values

TriggerBits

Event Flags

Trigger Assignment

TriggerDelay

Nominal Mainword

Trigger Fifo

Roman Pöschl

Graphic following template by G. Gaycken
Real Data Branch:

- CaliceTriggerProcessor
  - common to all reco
  - Input Trigger Mapping
  - Main Word Search
  - Trigger History

SimpleHitSearch

CalibrateAndApplyThreshold

Common to Data and MC

- Trigger Fifo
  - Trigger Assignment
  - TriggerDelay
  - Nominal Mainword

- ADC Values

- Cell Mapping Module / Cell Dimensions

- RawCaloHits

- LCCaloHits

MC Branch:

- LCCaloHits

- SimCaloHits

- Noise. Const.

- RawCaloHits

- RawCaloHits

TBEcalDigitisationProcessor

- Smearing of Calo Hits
  - Adding of cells not appearing in MC

Conditions Data

LCIO Stream

LCIO File

Graphic following template by G. Gaycken
Reconstructed LCIO files are entry point for newcomers

... and starting point of high level analysis

Contain e.g. 'familiar' CalorimeterHits

Benefits:

Application of Standard ILC Software allows e.g. for analysis of Ecal data 6 different institutes without major startup problems

Transfer of knowledge into full detector simulations (and vice versa)
First attempts are ongoing
See e.g.
Talk by J.Samson at CALICE Software Meeting 14/11/07
Talk by M.Reinhard at ILC Software Workshop at Orsay 2007

Though not the whole story – Still have to understand fundamentals of detectors -> required digging in raw data (less standard by definition)
Intermezzo – 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

- 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!
Sources of Conditions Data – Use Cases

**LCCD Use Cases**

- **Database**
  - **Conditions from Database**
    - Standard use case: read conditions from database for events, timestamp, and optionally provided tag.

- **Simple (LCIO) File**
  - **Conditions from Special Cond. LCIO File**
    - Read one set of conditions data from LCIO file - no time interval specified, e.g., calibration constants.

- **Snapshot of DB stored in LCIO File**
  - **Conditions from LCIO File for Time Interval**
    - Read conditions from an LCIO file that contains all needed conditions for a given time interval (could have been extracted before from condDB).

- **Conditions Dataflowing with Event Stream**
  - **Conditions from Within Data LCIO File**
    - Read data stored with event stream, e.g., slow control data.
Sources of Conditions Data – Use Cases

- Database
  - conditions from DB
    - standard use case: read conditions from database for events timestamp and optionally provided tag
  - conditions from special cond. LCIO file
    - read one set of conditions data from LCIO file - no time interval specified, e.g. calibration constants

- Simple(LCIO)File
  - read data stored with event stream, e.g. slow control data

- Snapshot of DB stored in LCIO File
  - read conditions from an LCIO file that contains all needed conditions for a given time interval (could have been extracted before from condDB)

- LCCD Use Cases
  - conditions from LCIO file for time interval
  - conditions from within data LCIO file

- ConditionsData flowing with event stream
  - read data stored with event stream, e.g. slow control data
Conditions Data and Systematic Studies

- The running of a database is indispensable for and experiment as big as calice allows e.g. for quick reproduction of running conditions.

- The interface to the database (or better conditions data) may lack convenience. However, all studies needed can be done with the available software.

- As of today systematic studies as e.g. varying the calibration constants do need a re-running of the reconstruction (different collections in raw and reco files). In principle no database access is needed for these. LCCD allows for other sources (see above). => New set of calibration constants can be put into (LCIO) files. The following use cases can be imagined =>
Systematic Studies – Use Cases

User made set of calib constants true to format needed for reco stored in a (LCIO) file

do not to be stored into a db

'Usual' Reconstruction

note that other processors might need the database

Please note that conditions data does not mean automatically literal database access, can e.g. work with a snapshot of the db
It means however using the LCCD interfaces!!!

Hand tailored set of calibration constants

e.g. Flat file

User defined processors to be run on CaloHit collections in reco files (can be standardized)

The latter scenario however might lead to conflicts with other cuts applied earlier in reco (e.g. noise cuts)
A view to the Monte Carlo Branch

- Model for the simulation of the CERN (and DESY)test beam is available (in release 06-04-p03 of Mokka)

TOP

Common effort of groups at RHUL, DESY, LLR, NIU

Will use grid for MC production
Estimation ~ 10000 kSI2kd for simulation of CERN 2006 data
Geometry Definitions: Mokka database <-> Calice database

- Mokka db for Simulation
- Calice db for Reconstruction

Might lead to conflicts if simulated setup different from reality

Needed to simulated run conditions beam energy etc.

Roman Pöschl

CALICE Software Meeting
Geometry Definitions: Mokka database <-> Calice database

Mokka db for Simulation

Might lead to conflicts if simulated setup different from reality

Calice db for Reconstruction

Needed to simulated run conditions, beam energy etc.

Envisaged step Unification i.e. Feed mokka drivers from calice db

E. Garutti
Si-W ECAL
Scint. Strips-Fe TCMT
Scint. Tiles-Fe AHCAL

Roman Pöschl
Summary and Outlook

- **Calice uses 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 a revision of the ILC Software concepts on a 'living' beast
  Consistent application of ILC Software allows non experts for an easy startup

- **Calice uses systematically Grid tools**
  First (and only?) R&D project within ILC effort
  24h/24h 7h/7h during CERN testbeam

- **Database indispensable for data integrity**
  Creates some threshold for users
  However, all studies can be performed with existing (but maybe imperfect) tools
  Users have to be ready to use these tools

- **Different sources of information for simulation**
  Efforts to solve this issue are ongoing

- **Organised software approach compromises 'publication speed'**
  But Calice is experiment with 20000 cells and 230 Collaborators (~HERMES)

- **Lack of Documentation for sure**