ILCDIRAC

DIRAC for the Linear Collider community

S. Poss

CERN, Switzerland

May 12, 2011

Outline

Part I

Context and aim of ILCDIRAC

Context and Aim

ILCDIRAC developed at CERN in the LCD group. The group is dedicated to the study of detectors for linear colliders, in particular detectors for CLIC.

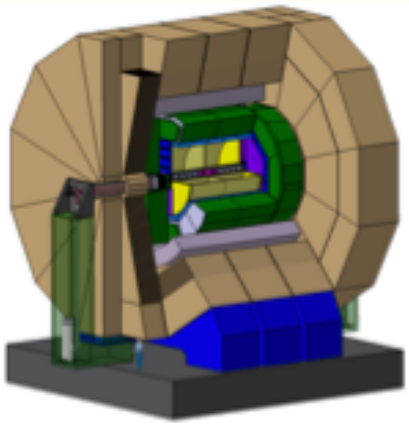
In 2011, CLIC collaboration must provide a Conceptual Design Report (CDR):

- Accelerator and detectors' performance (2 detector models considered)
- 6 physics channels to benchmark detectors, plus relevant background samples

CLIC detectors

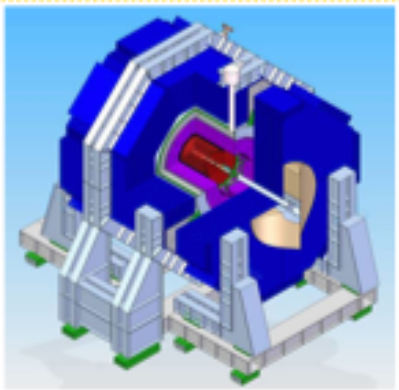
ILD: International Large Detector

"Large" : tracker radius 1.8m
B-field : 3.5 T
Tracker : TPC + Silicon
Calorimetry : high granularity particle flow
ECAL + HCAL inside large solenoid



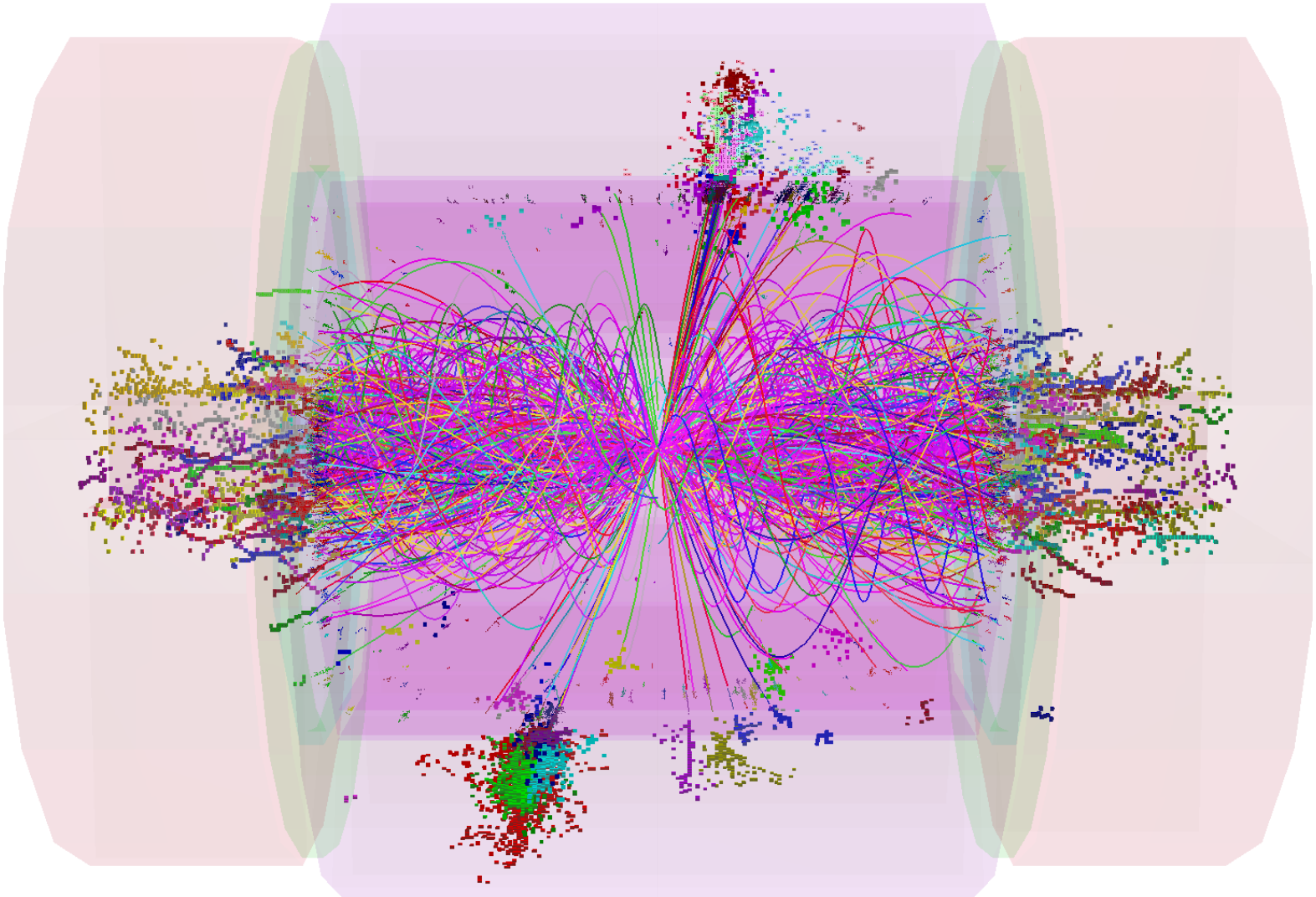
SiD: Silicon Detector

"Small" : tracker radius 1.2m
B-field : 5 T
Tracker : Silicon
Calorimetry : high granularity particle flow
ECAL + HCAL inside large solenoid



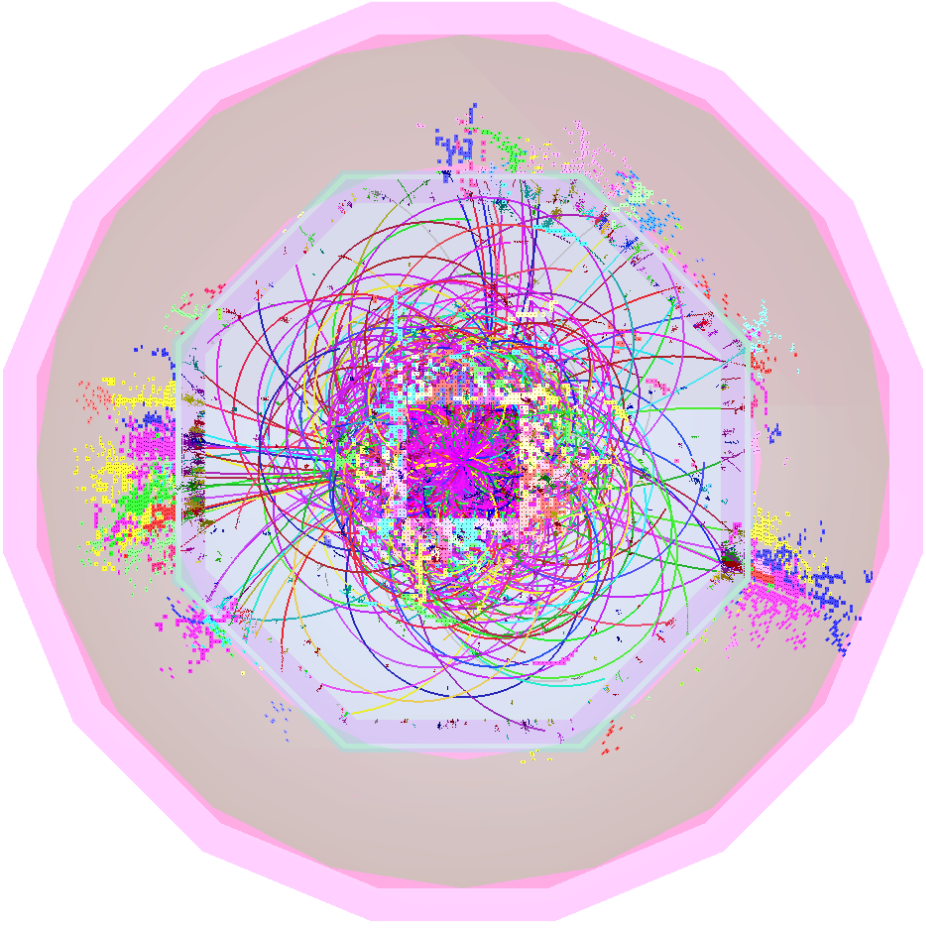
An event at CLIC

$Z \rightarrow q\bar{q}$ event with all background added

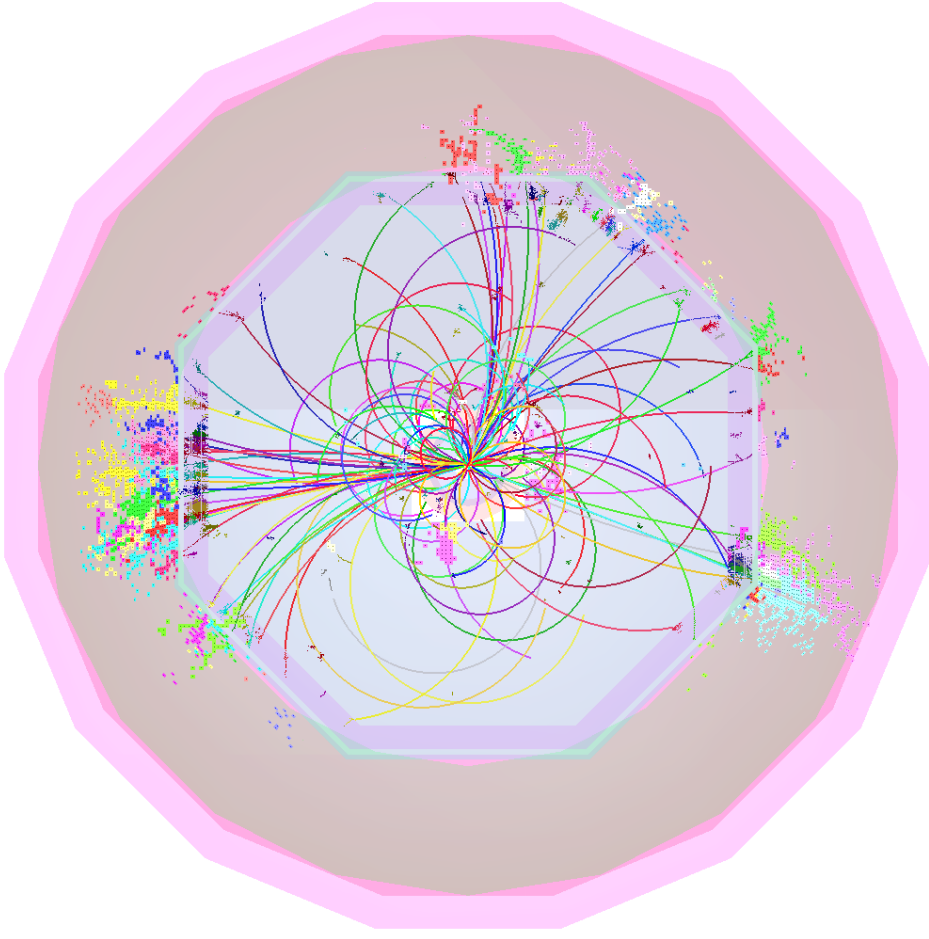


An event at CLIC

Before cleanup



After cleanup



More on the CLIC CDR

Use of the ILC software framework for generation, simulation, reconstruction:

- Generation: Whizard and PYTHIA
- Simulation: Mokka (ILD) and SLIC (SiD)
- Reconstruction: Marlin (ILD) and LCSIM (SiD)

Depending on event type and application, time per event varies from 100s to 600s. But 5 million events to process! $\rightarrow$ 95 years on one computer $\Rightarrow$ need to use the GRID

The ILC VO

- Dedicated to linear collider community
- 2 VOs actually: EU/Japan and US, discussions to merge them
- EU ILC VO hosted in DESY (Hamburg, Germany)
- 126 members in the EU/Japan VO

The DESY group developed their own production system, but was not yet fully functionnal when needed $\Rightarrow$ CERN group decided to use DIRAC.

I used to be in LHCb and learned about DIRAC there.

Part II

ILCDIRAC

DIRAC systems in ILCDIRAC

ILCDIRAC makes use of:

- Configuration, Core, Framework: the basics
- DataManagement: Dirac File Catalog (DFC) is used instead of Lcg File Catalog (LFC)
- Workload Management: job handling

DIRAC systems in ILCDIRAC

- Workflow Management: job submission and application handling (see C.B.Lam's talk about workflow usage)
- Transformation system: benefit from existing interface for job submission and monitoring

ILCDIRAC specifics (1)

Scripts added in Core for application registration: adding to CS and upload to GRID.

ILCDIRAC specific Utilities:

- Software management: we download tar balls from the GRID containing the software and install them in the local area (plans to use the Shared Area)
- data handling: application of ILC specific conventions on Logical File Names (LFNs), registration in DFC.

ILCDIRAC specifics (2)

Specific interface needed: 6 different applications for generation, simulation and reconstruction that all have different input parameters, different input files, etc. $\rightarrow$ ILCJob class, based on LHCbJob class, inherits from Job class of DIRAC.

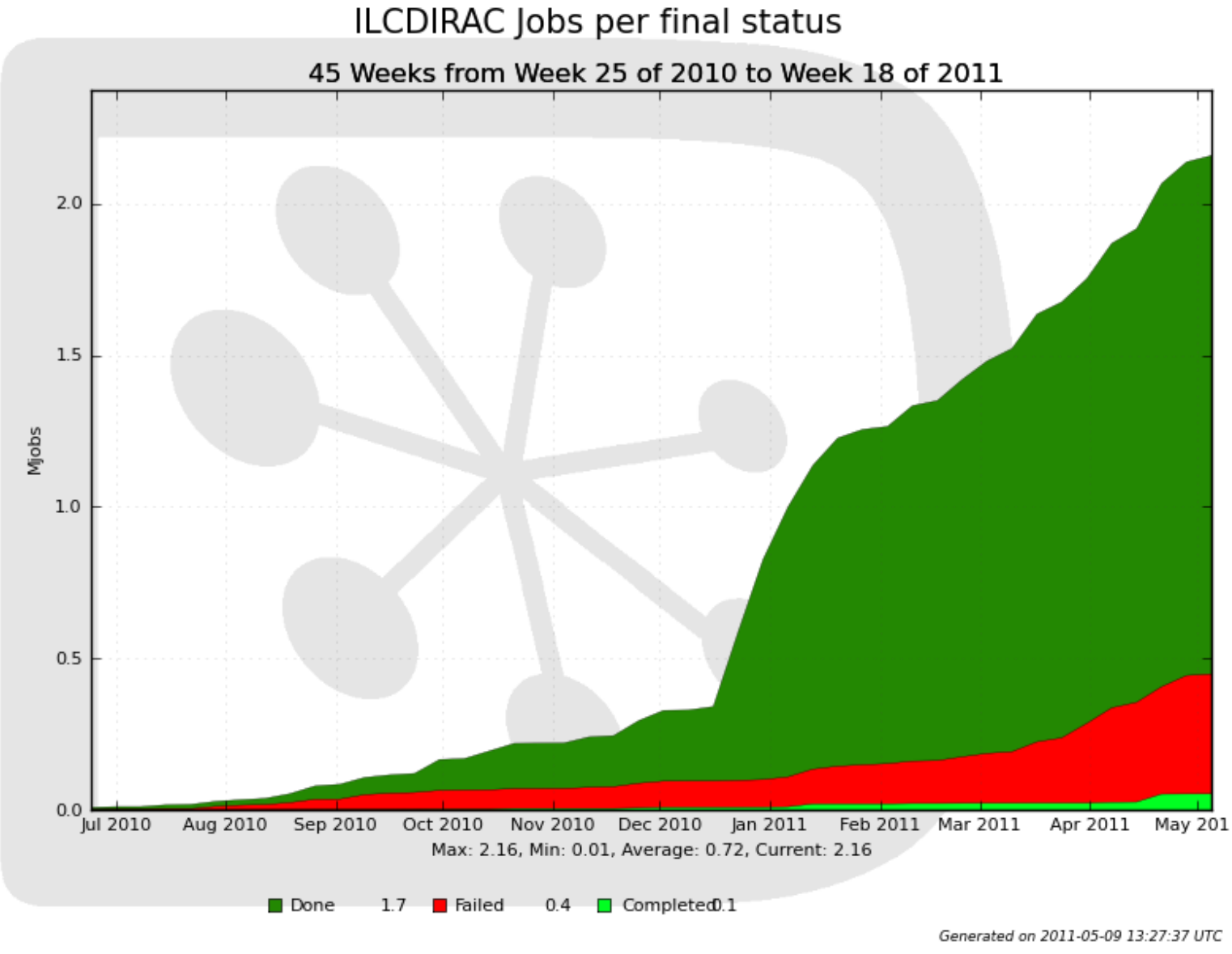
Whole new set of Modules for handling of ILC use case, i.e.

- One module per application
- Module for output data registration in DFC
- Module to concatenate files
- Etc.

Part III

Performance

ILCDIRAC performance



DIRAC File Catalog performance

- All ILCDIRAC productions' and users' data managed using the Dirac File Catalog
- 1 284 200 files registered so far
- 15 seconds to obtain 10 000 files' lfns and replicas

DIRAC File Catalog features

Metadata:

- directory level: event type, production ID, software used, etc.
- file level: ancestors, number of events, cross section, cuts, etc.

Command Line Interface available that allows browsing and operations: get file, set metadata, directory quota, ACLs, size used per site, etc.

Part IV

NEW! Process Production system

Motivations

- Link software versions to event type: with WHIZARD, configurations are set at compilation time, not run time $\Rightarrow$ an old version does not contain a configuration.
- Currently, available processes are stored in a file on the GRID: need better solution.
- Many software revisions to handle.
- Make sure a requested software is available before job submission.
- Mark software as invalid to prevent users from using it.
- If needed, also mark the data as invalid when processed with a faulty software.

Need a dedicated service not available in DIRAC

$\Rightarrow$ Process Production System.

Additional considerations

- Directory in DFC containing the production data,
- What cuts were used to produce a sample,
- Links between productions: a simulation production uses as input a generation production's output,
- Steering files used,
- Dependencies between software,
- Etc.

System being developed, open for discussions, DB diagram in backup. Other communities can be interested.

Underlying database developed with MySQL Workbench, great tool for that!

Part V

Other activities

Other activities

- Review of interface by better use of workflow modules: C. B. Lam (see next talk)
- Prediction of job running times using Case Based Reasoning: E. Hidle (see tomorrow)

Part VI

Conclusion

Conclusion

- ILCDIRAC derived initially from LHCbDIRAC.
- $\sim$ 15 active users: CERN (Ch.), LAL (Fr.), VINCA (RS.), ASCR (CR), IRES (Fr.), LAPP (Fr.), MPI (De.).
- Failure rate observed not due to DIRAC but to high load on disk servers (lots of time outs).
- DIRAC provides a developer friendly environment to add functionality.
- Some documentation missing, but I know work is ongoing (see next talk for some).
- Users are very satisfied with ILCDIRAC.

Part VIII

Backup

Database design for Process Production