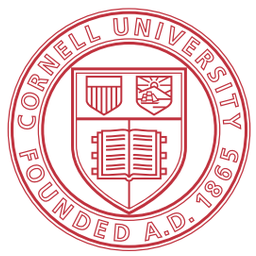
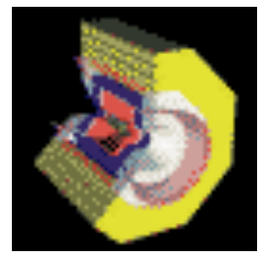
CLEO Analysis and Computing Model

Dan Riley, Cornell University

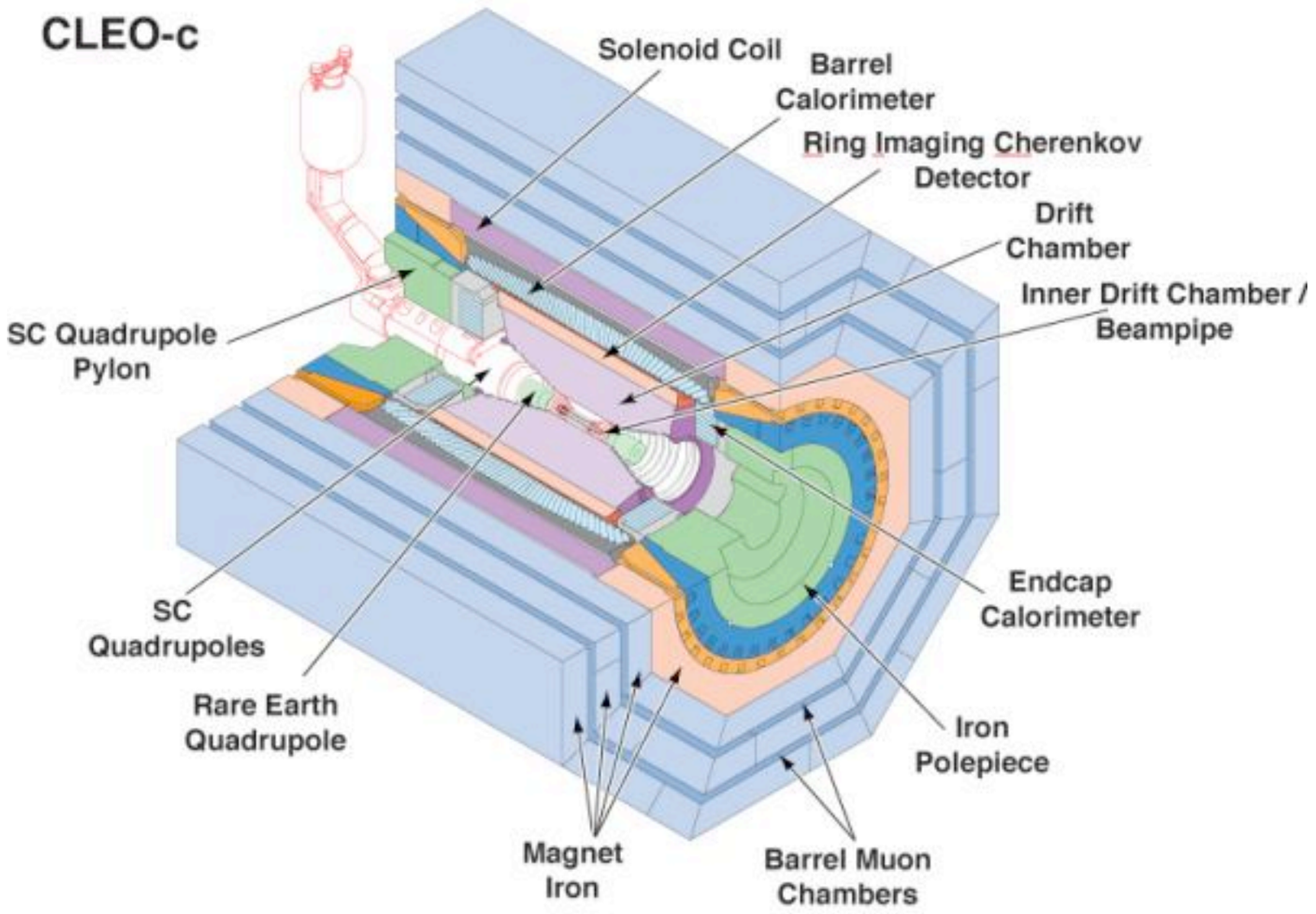
2009-01-26



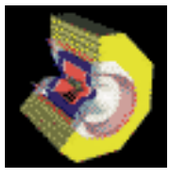
Cornell University
Laboratory for Elementary-Particle Physics



The CLEO Detector



Data collection ended April 1, 2008.



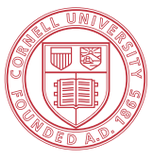
CLEO: Historical & Completion

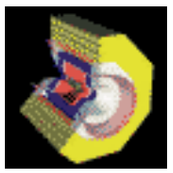
Historical:

- CLEO I, II, II.V, III: Υ; CLEO-c: ψ
- Nearly 500 Publications (~100 CLEO-c publications)
- Peak of 8 active physics topic working groups (with little overlap)
- Typically over 50 active analyses

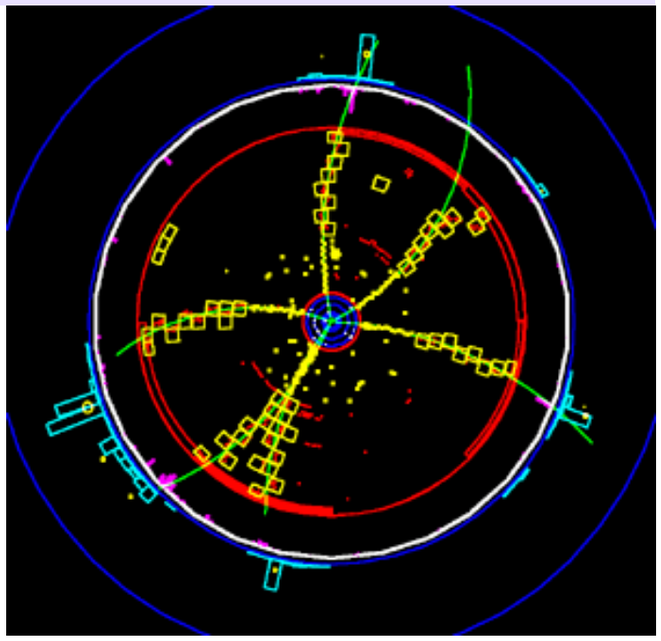
Completion:

- 20 publications expected in 2009
- 3 active physics topic working groups
  - *D Hadronic, Leptonic and Semi-Leptonic, Charmonium*
- 30-40 active analyses





Data Storage



Event Size

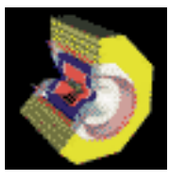
- ~20 kB/event raw data
- < 10 kB/event analysis data

Integrated Size

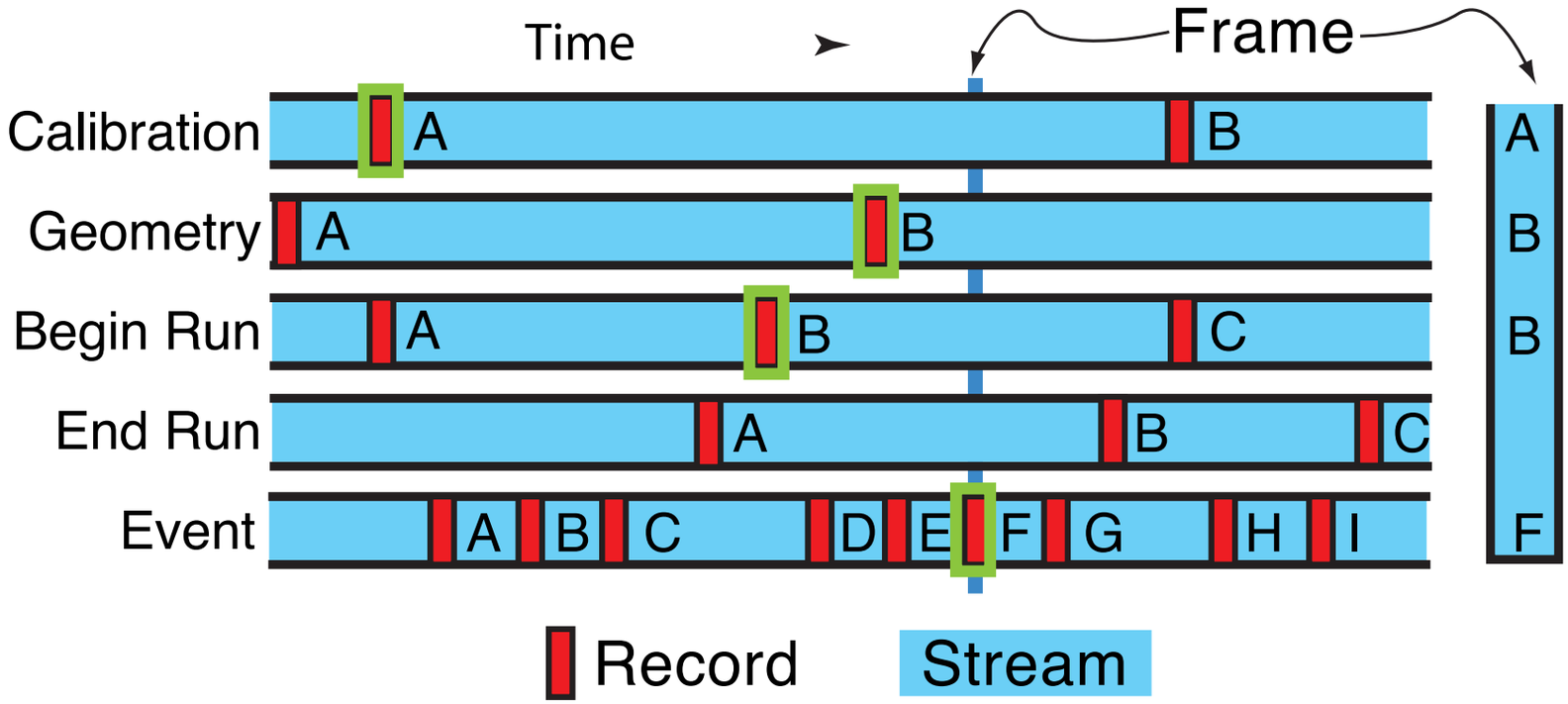
- > 80 TB raw data
- ~40 TB CLEO III and CLEO-c analyzed data + Monte Carlo
- Accessed via network file system
  - *xrootd is supported, but never widely deployed*

Production

- Reconstruction at Cornell
- Monte Carlo production offsite (principally U. Minnesota for CLEO-c)



Unified Data Model



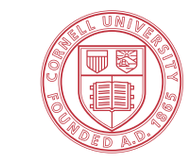
A **Record** holds all data that are related by lifetime

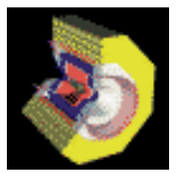
e.g., the Event Record holds Raw Data, Tracks, Showers, etc.

A **Stream** is a time-ordered sequence of Records

A **Frame** is a collection of Records describing the state of the detector at an instant in time

All data are accessed via a common type-safe interface to the Frame





Uniform Access Methods

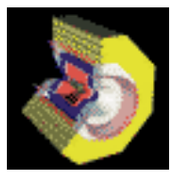
Common access methods for Frame objects

- Common C++ software framework using dynamically loaded plugins
- Frame access is independent of data source, data file format, etc.
- "One thing to learn"

```
FAItem< DBEventHeader > eventHeader;
extract( iFrame.record( Stream::kEvent ), eventHeader );
```

Levels of Abstractions

- Data partitioned into "hot", "warm" and "cold" data stores
  - *In practice only "hot" store used for most analysese*
- Hot store includes 4-vectors, showers, a subset of full helices, showers, particle ID probabilities, cluster-track matches, etc.
- Warm and Cold store include the rest of the track helices (less used particle hypotheses), data used in detector diagnostics



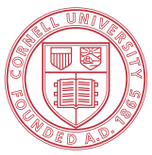
EventStore Grades

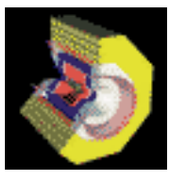
CLEO-c EventStore System

- Specify "grade" and start date of an analysis
  - *e.g., "eventstore in 20040501 physics".*
- Full reproducibility (unless data are lost)
- Data format independent (in principle)
- Adding some CLEO III data
  - *Converting from Objectivity object database to CLEO-c "PDS" format*

EventStore grades support the data analysis lifecycle

- raw data: "daqraw"
- Data quality monitoring: "daqraw" → "raw"
  - *Histograms, ntuples made for calibrations*
- Reconstruction ("pass2"): "raw" → "pass2"
- Post-reconstruction calibrations: "pass2" → "physics"
  - *Post-reconstruction corrections are written to EventStore and thus frozen; analysis jobs do not need access to calibrations and conditions.*



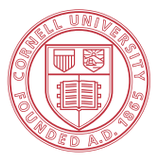


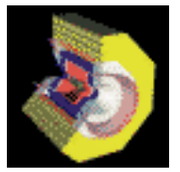
EventStore Sizes

"Collaboration", "Group" and "Personal" EventStores

- Designed as a hierarchical federation
- Data can be exported from a Collaboration or Group EventStore to a Personal EventStore
  - *and vice versa*
- Personal EventStore can be used to create a standalone analysis environment
  - *Interface is a subset of the production system*
  - *Designed to handle subsets of the data*

Simple to repackage CLEO data for redistribution!





Monte Carlo

Generators

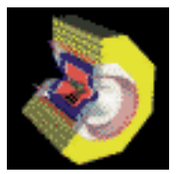
- qq, EvtGen, etc.

Simulation

- GEANT3 based

Calibrations and Conditions

- Required for Monte Carlo production
- Stored in Objectivity database accessed via CORBA
- Uses standard format-independent Frame interface
  - *CORBA plugin could be replaced via flat files or local SQLite database*



Data Preservation Prospect

What CLEO data are worth preserving?

- Υ data mostly superseded by B factories
- ψ should be superseded by BES III

Level of data preservation

- Raw data requires too much infrastructure
- Analysis level could be CLEO analysis objects or ROOT 4-vectors
  - Hybrid format with CLEO analysis objects directly accessible in ROOT technically difficult
- Precision measurements require accurate simulations, understanding of systematics

No funding model for data preservation!