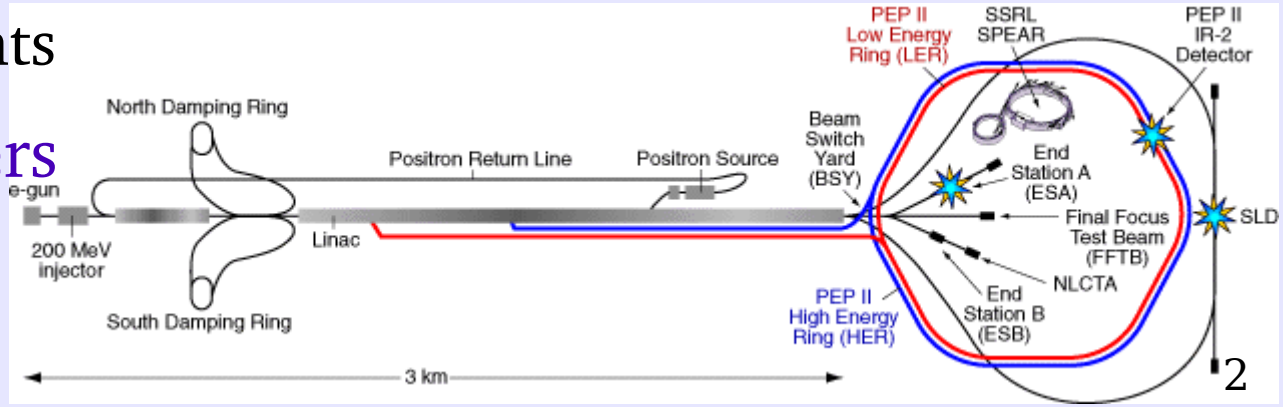
The New *BABAR* computing and Analysis Model

Chih-hsiang Cheng
Lawrence Livermore National Lab
For the *BABAR* Computing Group

DPF 2004, UC Riverside
August 30

The *BABAR* experiment

- *BABAR* is the detector for the PEP-II *B* factory at SLAC. Its main purpose is to study CP asymmetries in *B* decays and other B, charm, and tau physics.
  - ~80 institutes in 10 countries.
  - Operating since 1999
  - 30–100 million $B\overline{B}$ pairs/year
    - rate will double in a couple of years.
- Current Run 1–4 analysis data sample
  - >3 billion data events
  - >3 billion MC events
- > 100 journal papers

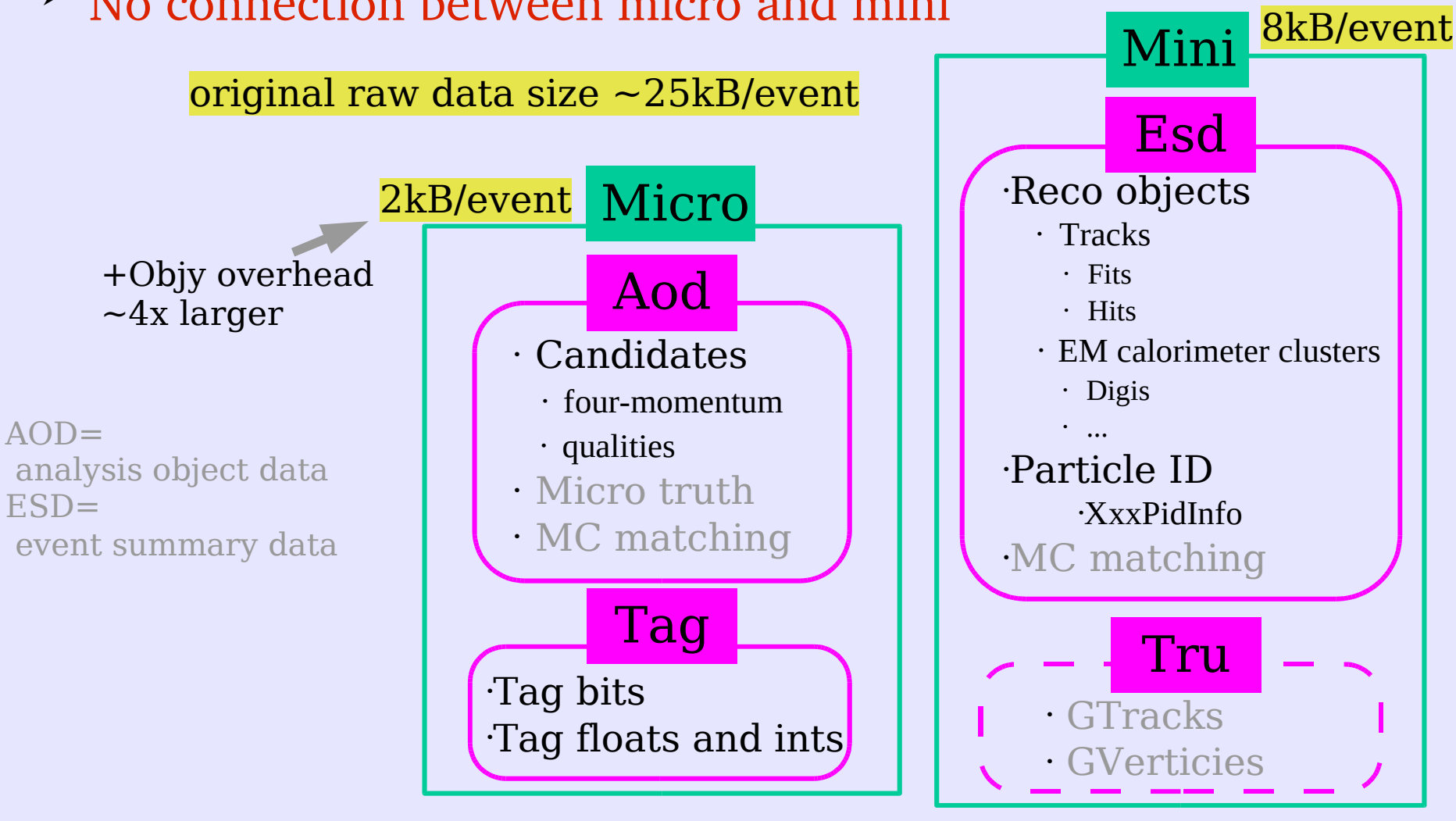


BABAR decided to revise its computing and analysis model in 2002

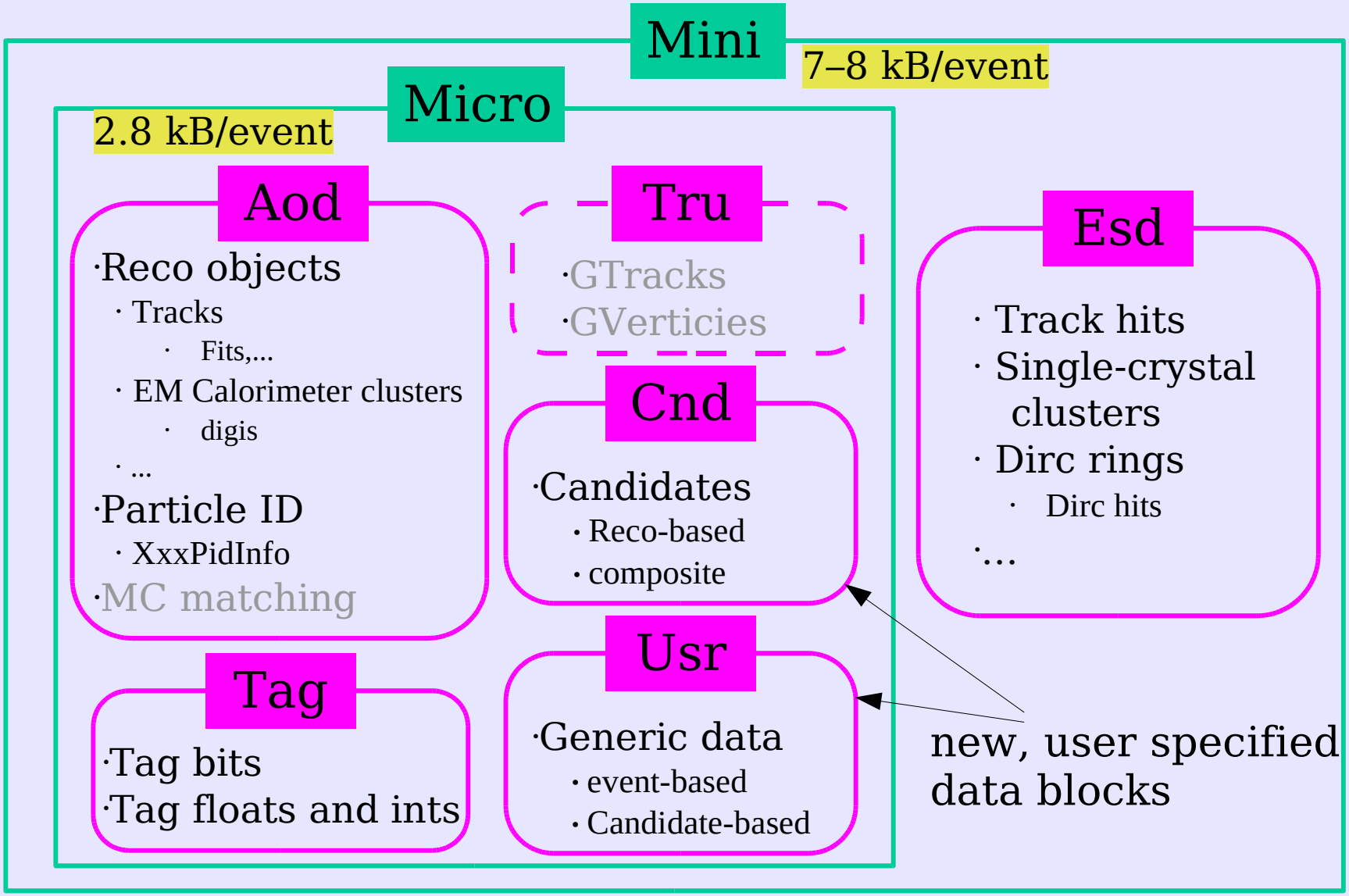
- Data sample is very large and growing quickly.
- Scaling issues with original Objectivity database technology.
  - Datasize/event too large, Federation limits, locking, …,etc.
- Need to distribute computing to sites other than SLAC.
  - maintenance costs too high for small sites
    - export only micro content using ROOT I/O: not a complete solution.
- The implementation of Computing Model 2 (CM2) has been fully deployed since late 2003.
  - New data structure
  - New eventstore
  - Skim and user customization
  - Easy data access and distribution

In CM1 only tag/aod were easy to use

- To get esd/rec, one had to go back to original raw data files.
- Partial solution: the “mini”
  - No connection between micro and mini

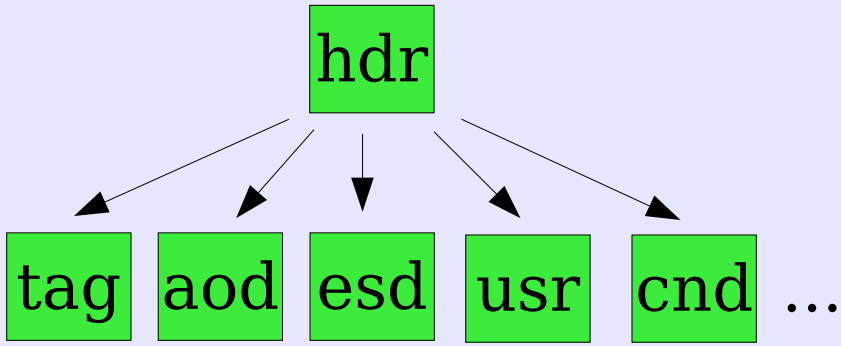


Objy Mini format extended to form CM2 micro and mini data format



CM2 eventstore based on ROOT I/O

- Each component is a TTree
  - can be in different files.
- The header (hdr) TTree points to the location of other components.
- Accessible from both *BABAR* framework and ROOT macros.
- CM1 micro interface to transient objects preserved; added access to persistent and underlying reconstruction objects.

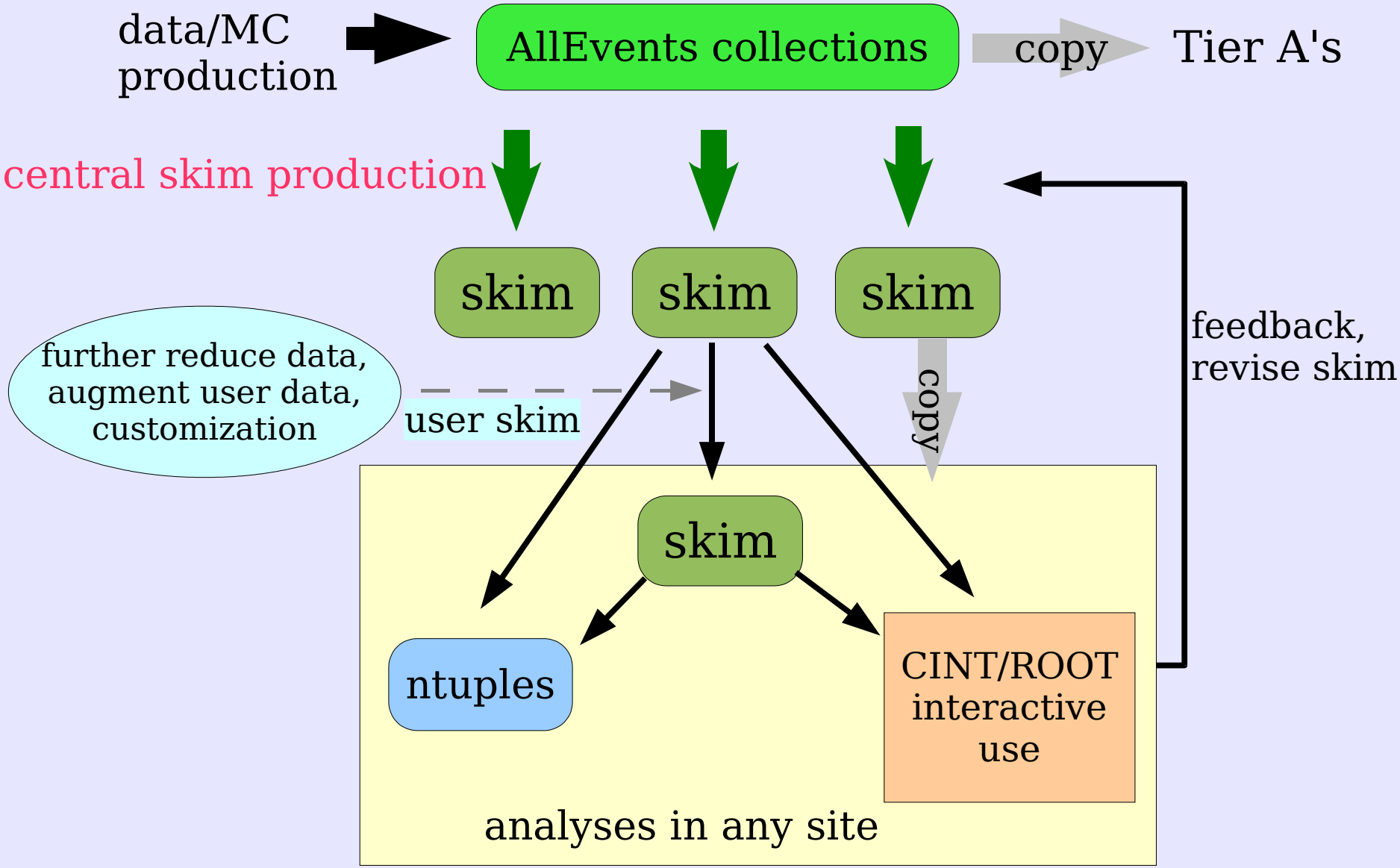


- New user customized components
  - Usr: bool, int, float, etc.
  - Cnd: composite candidates.
- Data can simply be carried home on your laptop.

Analysis data samples produced by centralized skim production

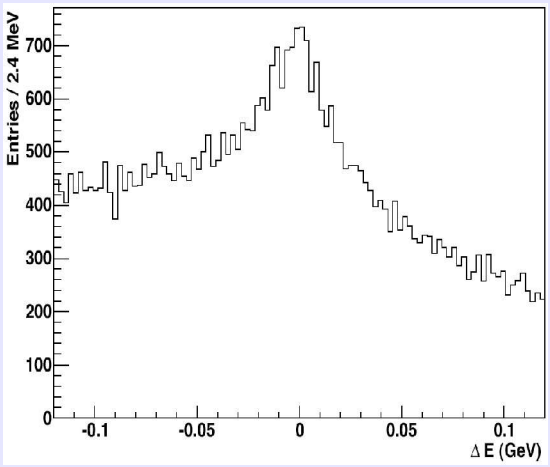
- Skim processes select events of interest for each type of analysis and store output in separate event collections.
- Each analysis group or user may define a skim output for this production in which they may choose to
  - "Deep copy" micro or mini data
  - Add their own customized data (e.g., composite candidate lists or associate calculated quantities with candidate or event)
- Should run frequently (target every 3 months).
- Time-consuming combinatorials only have to be done once.
- There currently O(100) skims defined.
  - Most skim sizes range from a few per mil to a few percent.

New analysis model



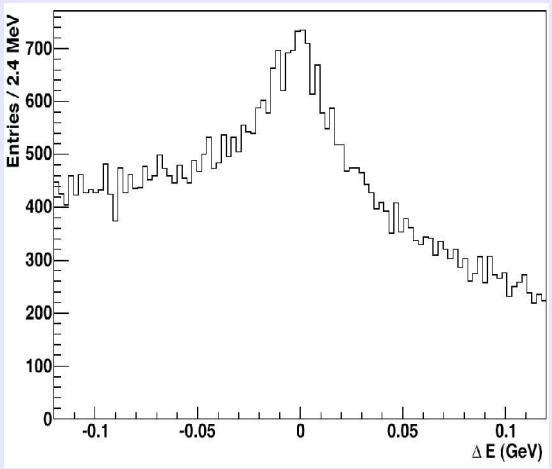
Detector QA and analysis with ROOT macros on the eventstore data

ΔE for B →D$^{(*)}$ X "Semi Exclusive" Selection
43,000 Selected Candidates from 2.5 M Events



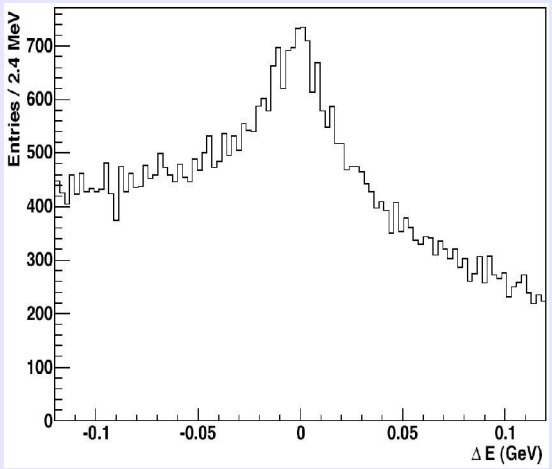
10,000 sec.

Rerun combinatorics and compute ΔE



1,000 sec.

Compute ΔE from persistent composites



10 sec.

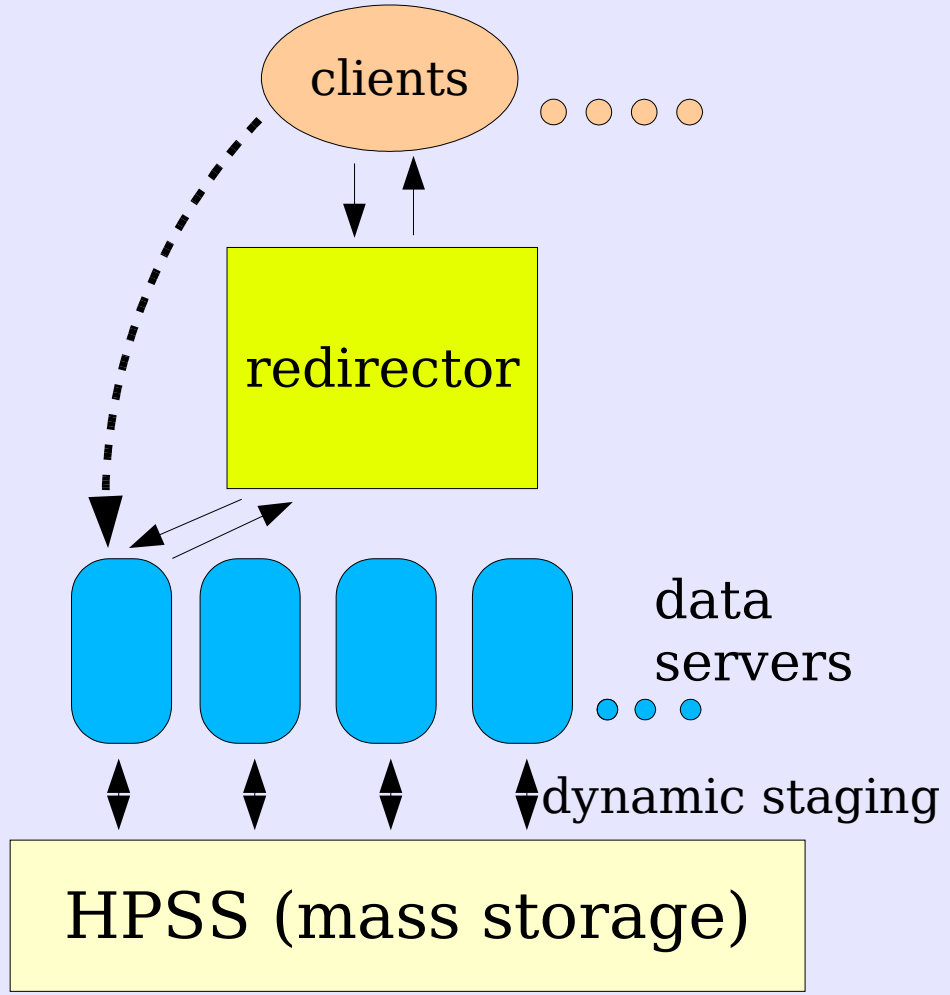
Interactive plot of UsrData ΔE

```
root[0] KanEventSource *input= KanEventSource::micro();
root[1] input->Add("my/Favorite/Collection");
root[2] input->getComponent(KanCompMap::usr)->Draw("SemiEx_DeltaE");
```

Xrootd provides scalable data access

- Xrootd: Extended root daemon.
- At SLAC xrootd serves O(2000) analysis and skim client cpus
  - 30 data servers.
  - 150 TB disc space
  - fault-tolerant clients
  - stable performance

http://xrootd.slac.stanford.edu



Distributed computing

- *BABAR* has 6 Tier A sites (SLAC, CCIn2p3, RAL, GridKa, Infn-Padova and CNAF/Bologna) plus 20+ university sites.
  - They now can all use the same eventstore.
  - Skimmed data split among Tier A sites according to interests of site users.
- All new data are processed in Padova, and output can be back at SLAC in 24 hours.
- Padova and GridKa also contribute to the skim production.
- Universities use standard tools to import their favorite skims.
  - easy to manage

CM2 skim production ramps up in late '03

- Significant new physics code (to define skims) incorporated along with CM2 core code.
- Primary computing resources: > 1000 cpus at SLAC
  - resources at GridKa, In2p3 and Padova also.
- Some resources diverted to convert Objy mini to CM2 format
  - fixed problems with existing mini data and reconstruction bugs;
  - improved calibrations and muon reconstruction.
- We learned to deal with problems from long running jobs and data merging.
- Will finish all existing data in September 04.
- *22 papers submitted to ICHEP in Aug. 04 based on this skim.*

Summary

- Implementation of the new computing and analysis model has been a success.
- Much easier to access and distribute data.
- For the coming run 5 in October, we will try to meet the three-month skim schedule.
- Many analyses can process all data in less than a week.
  - reduced from months of processing time in CM1;
  - much quicker to test new ideas, fix bugs.