Status of DZERO Reconstruction

Harry Melanson

Fermilab

September 11, 2003

- Overview of progress since last review
- Current issues related to charge of committee:
  - CPU time per event
  - Farm throughput
  - Data tier event sizes

Progress since last review

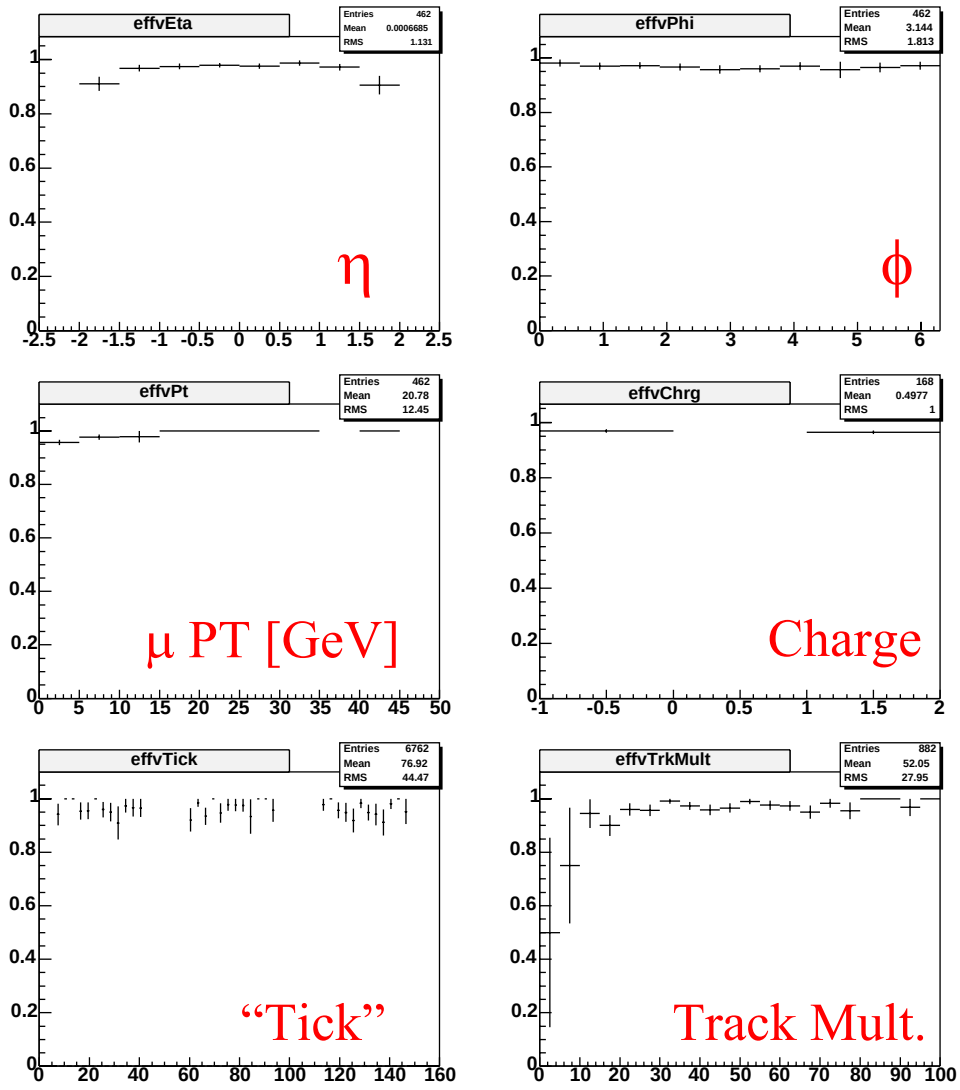
Highlights only

- First pass in situ alignment of tracker (using magnet off tracks)
  - Improved momentum resolution (Z mass width)
  - Improved muon / cal cluster track matching
- Use of online calibrations via offline database servers
- Support for “real” detector issues
  - Non-linear corrections for calorimeter electronics
  - Fixes for various intermittent detector problems
  - New clustering algorithms to deal with calorimeter noise

More new features

- Inclusion of preshower detectors in physics analyses
- Complete migration to "compact data tier" (thumbnail)
- **Major revision** of tracking algorithm
  - Efficiency from ~ 75% to ~ 98 %
  - Reduction of fake rate
  - Ability to reconstruct down to 500 MeV
  - Ability to reconstruct large-impact-parameter tracks
    - K_s, Λ, $\gamma \rightarrow$ e+e-, ...

Tracking Efficiency



Tracking (Efficiency-Isolated)

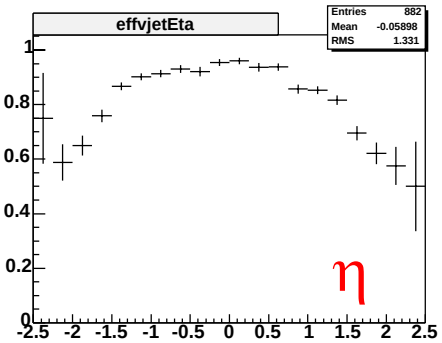
Tracking efficiency as measured in data using ***isolated*** muons

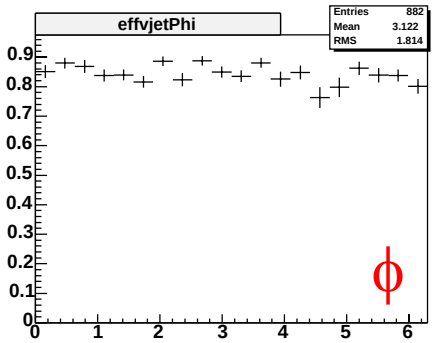
Avg effic. ~ 98 %

Plots from offline monitoring tool, **highest lumi** run to date (4.36E31)

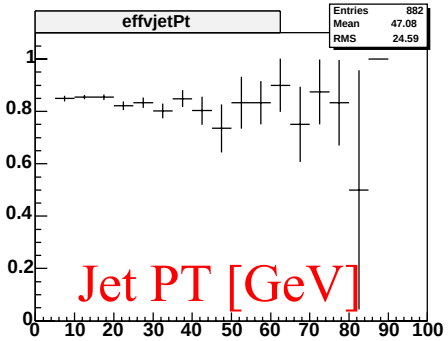
Tracking Efficiency

~ 95% in central region



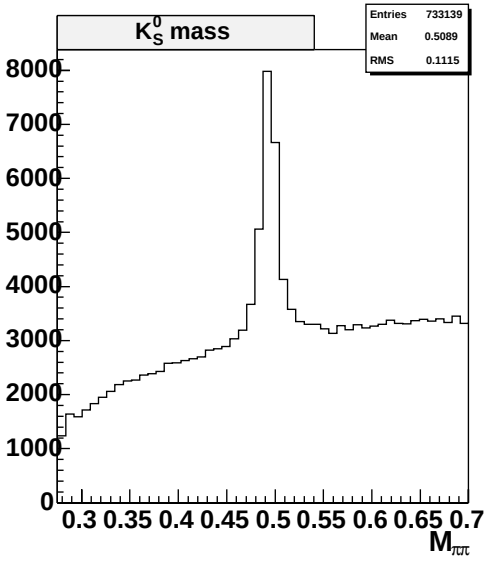


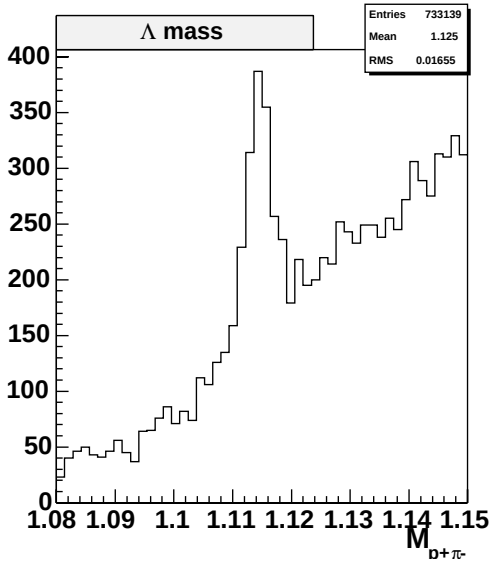
Tracking efficiency as measured in data using ***non-isolated*** muons



~ 85% for all jets

Resonances

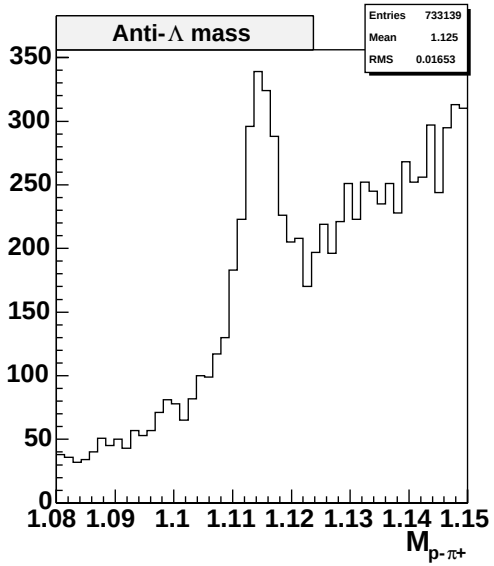




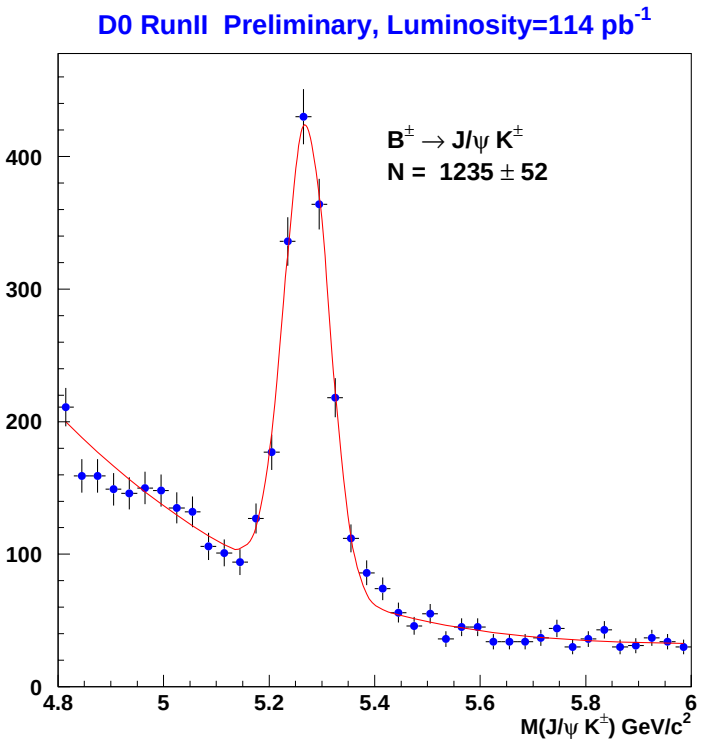
From 117 K events of Run 179934 (highest lumi run)

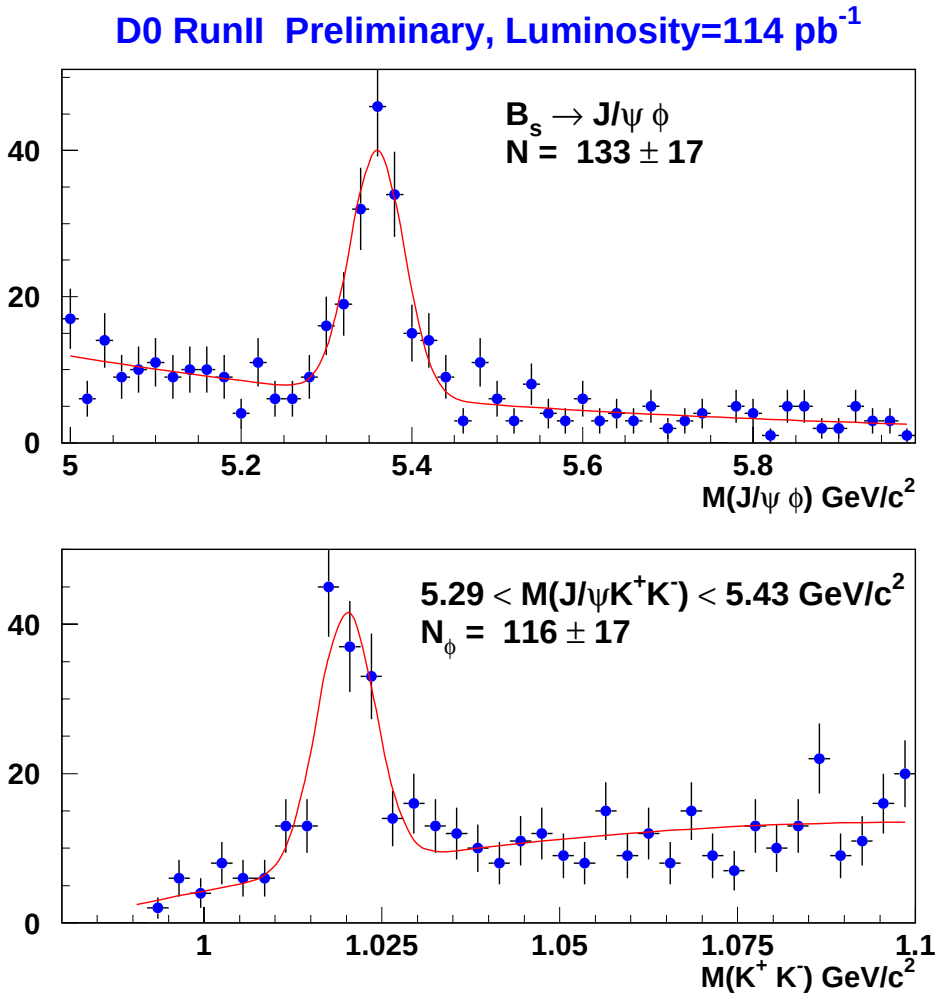
< 1 hour data taking

More plots from offline monitoring tool



Resonances

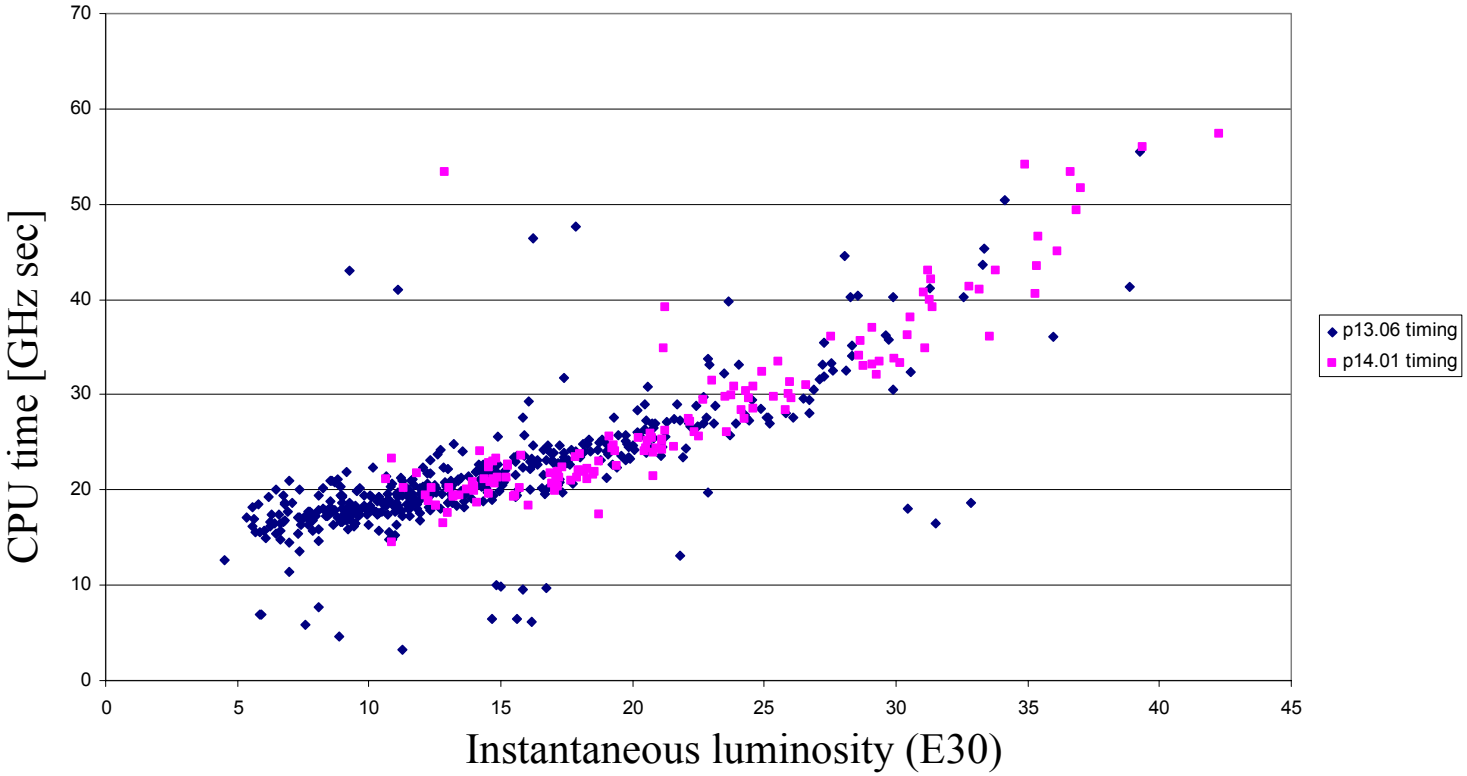




The fun has begun...

Current status and issues

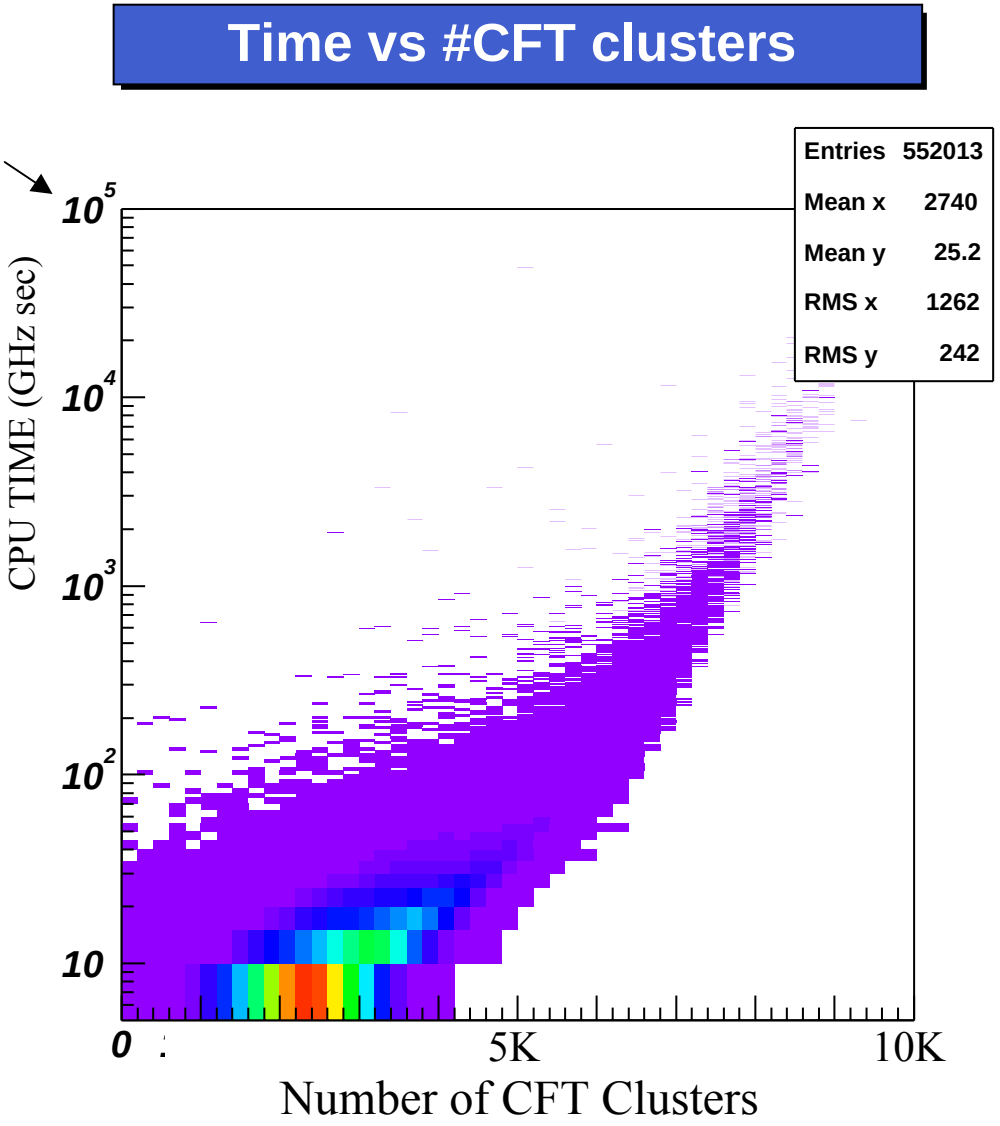
- CPU time per event



- CPU time is measured in 10's GHz sec / event (e.g. "slow")
- Grows with instantaneous luminosity (e.g. "gets slower")
- Current version ~ same as previous version (albeit with higher tracking efficiency)

CPU Time

- Observe some “really slow” events
- Associated with very high occupancy in fiber tracker
- Currently under intense investigation
- Many events appear to be multiple interactions + “exceptional” detector noise

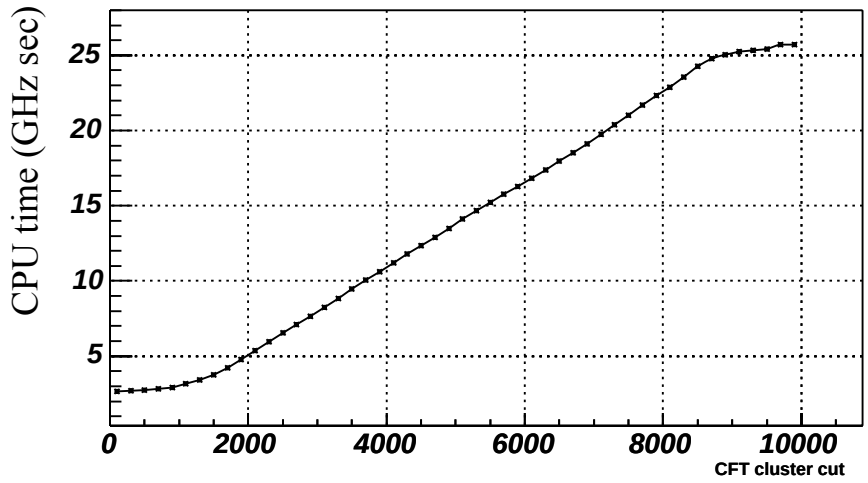


CPU Time

10% gain in processing speed by skipping 0.05% of events

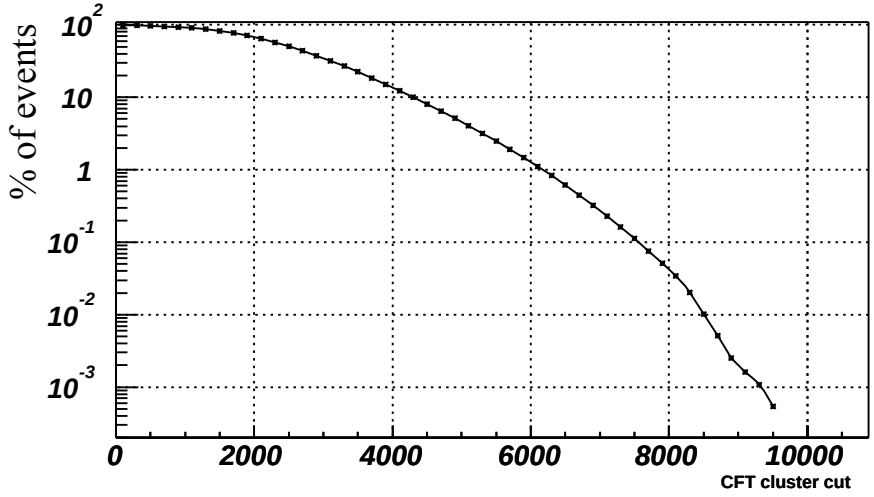
20% gain by skipping 0.1%

Etc...



Alternative approaches under investigation:

- Flag / skip events - next release + 1 ("p14.05.00")
- Raise P_T threshold
- Do SMT to CFT tracking only
- Change clustering algorithm



Number of CFT Clusters

CPU Time

Other speed-ups in the pipeline:

- 10% - faster matrix classes in new tracking algorithm. Next release – p14.04.00
- 8 % - faster magnetic field class. Needs to be tested with MC generation.
- 10% - faster propagator at the expense of loss of high impact parameter tracks – under study by physics groups.

  May require reprocessing of "B physics" samples with alternative cuts (actually being done now anyway)
- 10%? - Rewrite of hot spot in tracking. Initial studies showed 10% speed up; current tests show little change.

Farm Efficiency

- Current throughput numbers
  - 1.5 M events / day (was 2 M)

- Current bottlenecks
  - Production efficiency – Fraction of wall time resulting in output file (measures crash rate)
  - Dead time – Fraction of time RECO is idle (no CPU being used – swapping, etc.)
  - Unaccountable losses – node needs to be rebooted, power outage, etc.

Category	P13	P14
Efficiency	98.4%	89.2%
Dead time	2.6%	21.6%
Losses	2.8%	5.7%

Correlated with high luminosity runs

Farm Efficiency

Category	P13	P14
Efficiency	98.4%	89.2%
Dead time	2.6%	21.6%
Losses	2.8%	5.7%

- Solving high occupancy problem should alleviate dead time (swapping) and improve losses
- May also improve efficiency
- After addressing this problem, will go after remaining (as was done with P13)

Data tier event sizes

- Current DST size
  - "Low lumi run" – **223 KB**
  - "High lumi run" – **335 KB**

DST size sensitive to detector conditions

- DST permits reprocessing
  - Can redo alignment
  - Cannot redo clustering
  - Rest of RECO can be redone

Target: **150 KB** (or less)

- Next release (p14.04.00)
  - "Low lumi run" – **176 KB** (20% reduction)
  - "High lumi run" – **288 KB** (14% reduction)

Reducing DST size

"Low lumi"

Chunks	size/event	percentage
L3Chunk	62645	17.1
FPSClusterChunk	40131	11.0
CPSClusterChunk	37747	10.3
Calt42Chunk	27549	7.5
JetChunk	26703	7.3
ChargedParticleChunk	19715	5.4

"High lumi"

Chunks	Size/event	percentage
CPSClusterChunk	117788	19.2
JetChunk	77107	12.5
FPSClusterChunk	67973	11.1
L3Chunk	65571	10.7
ChargedParticleChunk	37611	6.1
Calt42Chunk	37579	6.1

- Preshower detectors (CPS, FPS) still being calibrated
- If we remove PS's from DST (painful) or make their size negligible (preferred)

  DST: **141 KB – 202 KB**

- Next on the list:
  - Improve calorimeter calibration (in progress)
  - Rethink L3Chunk (maybe hard)
  - Start removing functionality (very painful)

Currently under development

Thumbnail size

- Current size: **27 KB**
- Independent of instantaneous luminosity
- Goal: **15 KB** (or less)
- Current content:
    - Thumbnail – 55%
    - CalDataChunk – 38% (all cal cells)
    - CalNadaChunk – 2% (hot cells)

- The "cal chunks" have been extremely valuable, allowing physics analyses to deal with calorimeter problems.

- Removal of these chunks gets TMB to **16 KB**.

Conclusions

- **Tremendous progress** has been made over the last year
  - Alignment, calibration, db servers
  - Detailed detector understanding
  - Algorithm tuning
  - **Major** new tracking algorithms

- **Current performance** is the **limiting factor** in
  - Timely reconstruction of collider data
  - Ability to reprocess large samples with new versions
  - Storage requirements at the DST data tier (also impacts reprocessing strategy)

Outlook

- **CPU time**
  - A few 10%'s in pipeline
  - Addressing high occupancy:
    - Several approaches under investigation
    - Speed up by 20% (or more)
    - Should significantly improve farm efficiency as well
  - Adding up (linearly), can see ~ 40% speed up within 1 – 2 months
- **DST size**
  - Next release: 176 – 288 KB
  - After calibration / rewrite of PS: 141 – 202 KB
  - Some gain after improved calorimeter calibration
  - Further reduction more painful