BNL Contribution to ATLAS

Software & Performance

S. Rajagopalan

April 17, 2007

DOE Review

Outline

- Contributions to Core Software & Support
  - Data Model
  - Analysis Tools
  - Event Data Management
  - Distributed Software
  - Software Infrastructure
    - Including validation effort
- Contributions to Application Software
  - Calorimeter
    - Including EM & Hadronic Calibration
    - Calorimeter database support
  - Muons
  - Trigger
  - Combined Reconstruction
    - e-gamma, Jets, Taus and Missing ET

Leadership roles in ATLAS

Calorimeter Performance Coordinator (SPMB)

- S. Rajagopalan (2003 - 2007)
- H. Ma (2007 -)
  - Calorimeter Cosmic Commissioning (since 2005)
  - Calorimeter Database (since 2003)

Analysis Tools Coordinator (SPMB)

- K. Assamagan (2005 - 2007)

Trigger Jet/Tau/EtMiss Coordinator (TAPMCG)

- K. Cranmer (2006 -)

Trigger Menus (TAPMCG)

- S. Rajagopalan (2007 -)

Distributed Data Management Operations

- A. Klimentov (2006 -)

Software Effort Contribution (snapshot)

- Core Software & Support (9 FTE)
  - Including infrastructure support, validation and physics analysis tools
  - NOT including production support and facility operation.
  - NOT including BNL based OSG or University-RPM funded personnel.
  - S. Panitkin, T. Wenaus, M. Nowak, A. Klimentov, T. Maeno, A. Undrus, S. Ye, P. Nevski [0.5], S. Snyder [0.2], S. Rajagopalan [0.1], H. Ma [0.1], K. Assamagan [0.4], K. Cranmer [0.2]
- Sub-System and Combined Reconstruction Software (5.4 FTE)
  - D. Adams, H. Ma [0.4], S. Rajagopalan [0.4], K. Cranmer [0.3], F. Tarrade [0.2], A. Cunha* [0.2], A. Patwa [0.1], S. Snyder [0.3], F. Paige [0.3], G. Redlinger [0.1], K. Assamagan [0.3], D. Damazio [0.5], S. Kandasamy, H. Chen [0.3]
- CERN Based personnel:
  - D. Damazio, A. Klimentov, P. Nevski, M. Nowak.

Core Software: Data Model

BNL has been playing a significant role in the Data Model Effort.

- S. Rajagopalan (EDM infrastructure), K. Cranmer (Event Management Board)
- K. Assamagan, H. Ma, S. Snyder, M. Nowak, T. Maeno have all contributed

Event Summary Data (ESD):

- Computing Model: 0.5 MB/event (perhaps 0.7 MB early days)
- Current size: > 1.5 MB/event!
- Plan to keep a full copy at U.S. Tier 1 center.

Analysis Object Data (AOD):

- Computing Model: 100 kB/event.
- Current Size: > 200 kB/event (of which Truth is 40%)
- Plan to keep copy at Tier 1(full copy) and Tier 2.

Derived Physics Data (DPD):

- Recent ideas – structured ROOT tuples.
- Perhaps can expect it to be 25 kB/event?
- Depends on the analysis, will have several copies

Core Software: Analysis Tools

- AOD is a reconstruction output used as input to a first stage physics analysis.
  - Proposal for Derived Physics Data providing greater interactive analysis capability
- Proposal for a Structured Athena Aware Ntuples (K. Assamagan)
  - “Structured” in how data is saved in ROOT trees
  - Used for Derived Physics Data (DPD)
  - BNL Analysis Tools Meeting: Technical proposal & implementation.
  - Since then: ATLAS AOD Task Force
    - Build on the BNL meeting, involving a broad user community
- BNL is involved in the data format of the DPD and a providing similar access to a ROOT or an Athena based analysis
  - K. Assamagan, K. Cranmer, S. Rajagopalan, S. Snyder
- BNL is also involved in the development of EDM for DPD data and in the development of common tools for Analysis
  - EventView is popular among physicists providing common suite of analysis tools.

Core Software: Event Data Management

- Key Personnel: M. Nowak, S. Panitkin
- Design and implementation of schema evolution for event data
  - Introduction of a parallel persistent data model with type versioning and creating infrastructure of transient <-> persistent converters.
  - Substantial I/O performance improvements, up to 20x for reading speed in extreme cases. Actual reading speed improved from about 0.5 MB/sec to 2-5 MB/sec.
- Work as LCG/POOL project:
  - Implementation and integration of the new POOL Collections. The main goal was to merge the various database collection packages (Oracle, MySQL, SQLite) into one relational Collection package, where CORAL layer (part of POOL) takes care of database specifics.
- Interest in file based event selection tags using xrootd
- Navigation across files

Core Software: Distributed Software

- BNL has taken a lead role in the development of a grid-based production and distributed analysis tool (PANDA).
  - T. Wenaus, T. Maeno in close collaboration with U.T.Arlington
  - It is a scalable workload system, tightly couple to DDM, highly automated requiring little personnel intervention.
  - Launched and prototyped since 2005, it is now continuously used in production (~30% of total ATLAS jobs handled by PANDA in 2006)
  - PANDA extended to all grid flavors: OSG and LCG.
    - PANDA critically dependent on DDM (managing placement/replication of file based event data).
  - Distributed Analysis has similar requirements as production:
    - pAthena, a simple front-end, is popular with physicists
- Support from OSG to provide an experiment-neutral application

DDM Operations

- A. Klimentov chairs the ATLAS DDM Operations Group, whose role is:
- Distributed Data Management Operations Group
  - The group includes Tier-1 and Tier-2 reps from 50 centers
  - Main activities
    - Day by day users and production data management
    - Set up system for automatic data replication to ATLAS Tier Centers (AOD files, validation samples, streaming test data)
    - Conduct ATLAS wide data transfer functional tests
      - Successful test in replicating 3-5 GB files between T0 and BNL Tier 1/U.S. Tier 2
    - Evaluate SW technology (like file catalog)
    - Support Users (via Savannah)
    - Develop GUI and I/F for data transfer control and monitoring
- SW Integration Working Group
  - Develop and maintain the system for task requests (in production since 2/2006)
  - Propose and implemented the concept of datasets (approved and accepted by the collaboration)
  - Propose the definition and implementation of Logical and Physical File Names
  - Develop the system to support users and physics groups data transfer requests

Core Software: Software Infrastructure

- Key Personnel:
  - A. Undrus, S. Ye, P. Nevski, D. Damazio
- Maintenance of cvs repositories
  - Full Suite of software libraries maintained at the Tier 1 center.
- Nightly Builds
  - Nightly build system developed and deployed by A. Undrus, used at CERN.
- Validation infrastructure
  - Poor validation infrastructure have resulted in long (~months) time to validate a production release.
    - Several problems are found – sometimes after extensive production has already run – Problems that could have been caught much earlier.
  - BNL has taken a lead role in establishing a robust infrastructure.
  - Post processing of validation tests and web-based displays of problems for easy navigation are now being developed at BNL.

Application Software: Calorimeter

- Significant participation in the development of calorimeter software since the early days, primary contributions in:
  - Calorimeter Reconstruction and data model
    - S. Snyder, H. Ma, S. Rajagopalan
  - EM Calibration
    - S. Snyder, S. Rajagopalan
  - Hadronic Calibration
    - F. Paige
  - Database support for LAr calorimeter
    - H. Ma, S. Kandasamy
  - Cosmic Ray Commissioning
    - H. Ma, F. Tarrade

Calorimeter Cluster Level Corrections

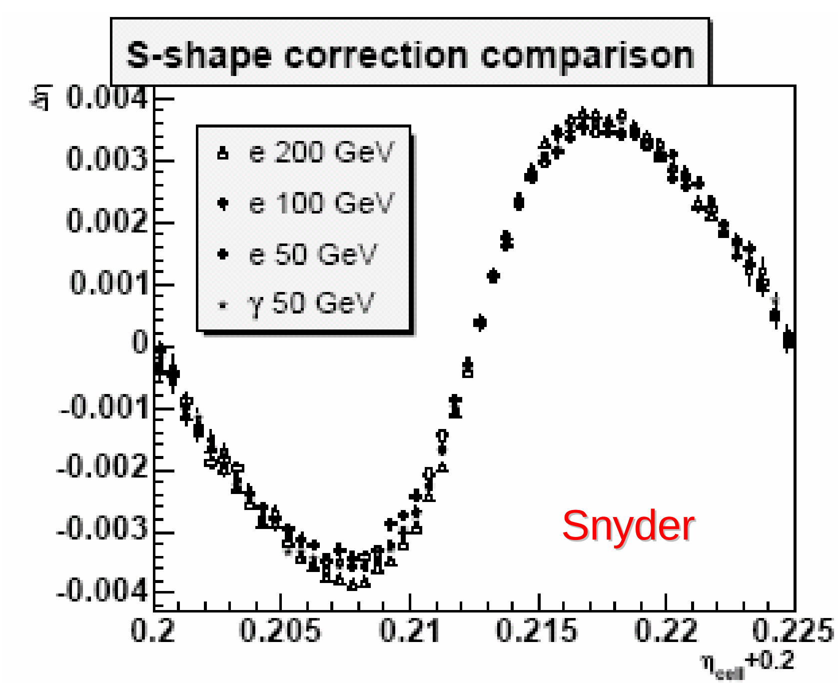
- Two clustering algorithms are used:
  - Sliding Window algorithm producing EM clusters for different cone sizes: 5x5, 3x5, 3x7 etc.
  - A 3-d nearest neighbor algorithm (topological clustering)
- Series of corrections applied to reconstructed EM clusters:
  - Eta and phi position corrections
  - Energy modulations vs eta, phi
  - Lateral out of cone energy corrections
  - Longitudinal corrections including upstream matter & leakage
  - Gap corrections, if relevant
  - Correct for residual HV effects and pathological cells.
  - Overall energy scale
- BNL contributed to the derivation of several of these corrections and the overall software implementation

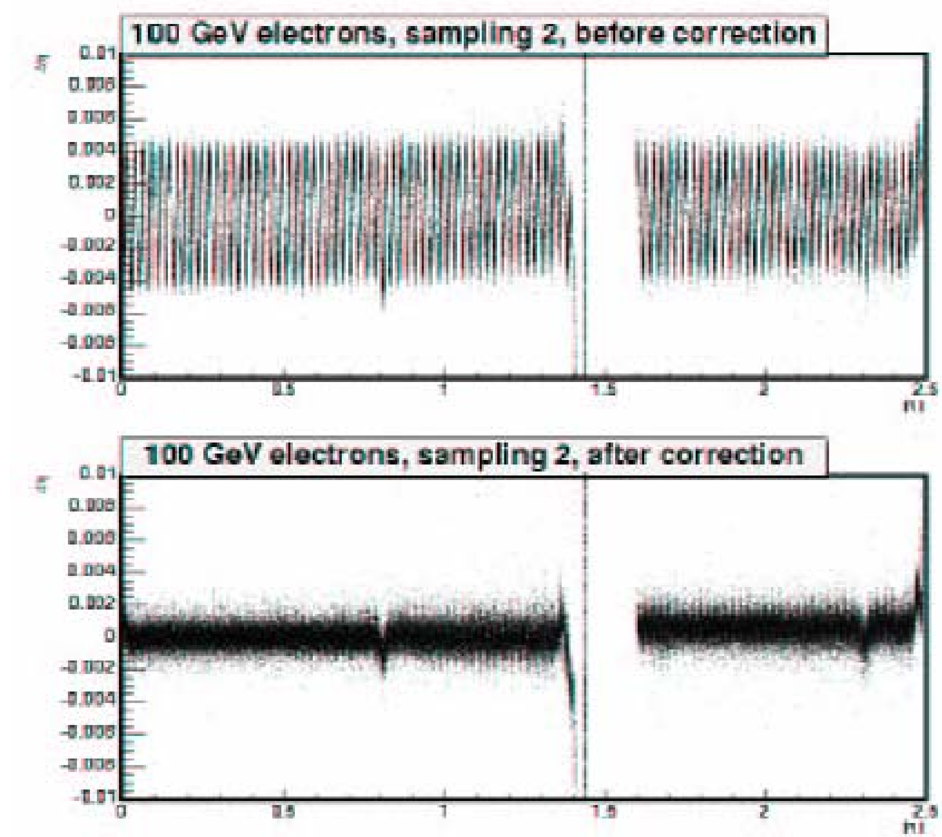
S-shape corrections

Finite granularity of middle sampling (0.025x0.025) not small compared to shower width

Simple energy weighted position (η) measurement pulled toward middle of cell

Corrections derived from single electrons (Snyder)





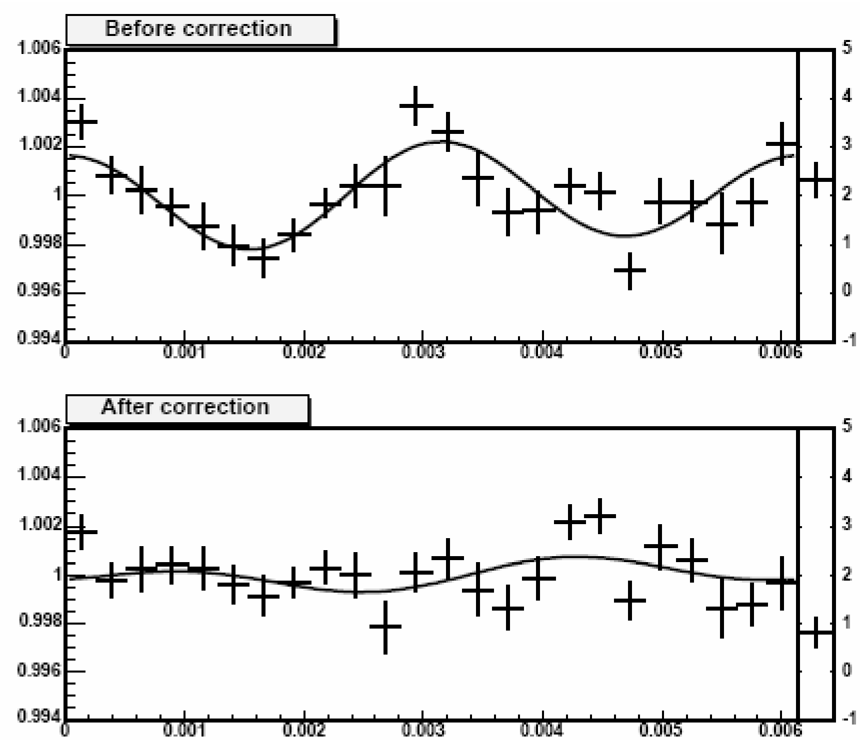
Energy modulation

S. Snyder

Energy modulations as a function of phi

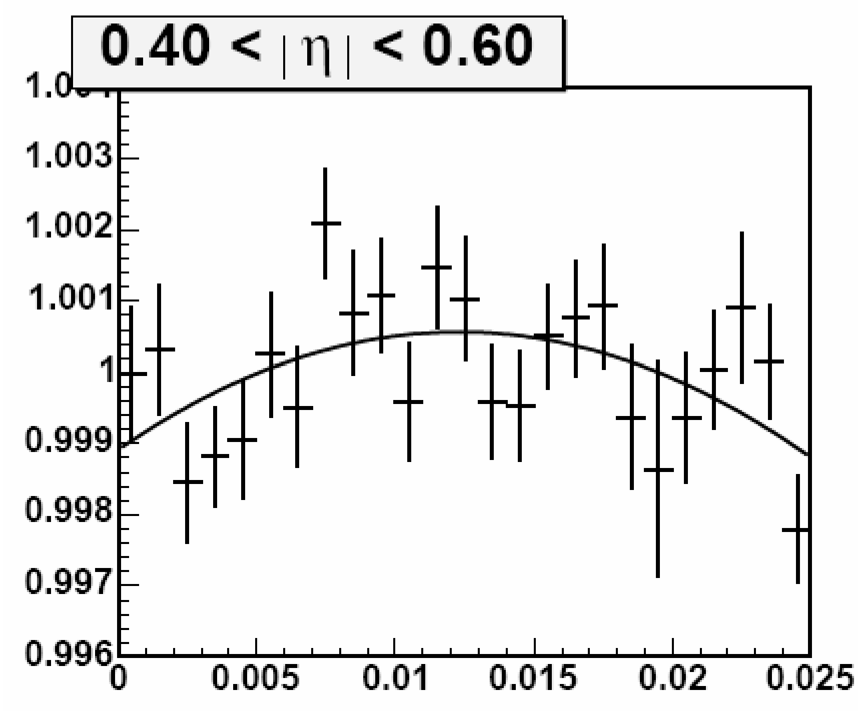
Derived for different eta positions

0.1 to 0.2% effect

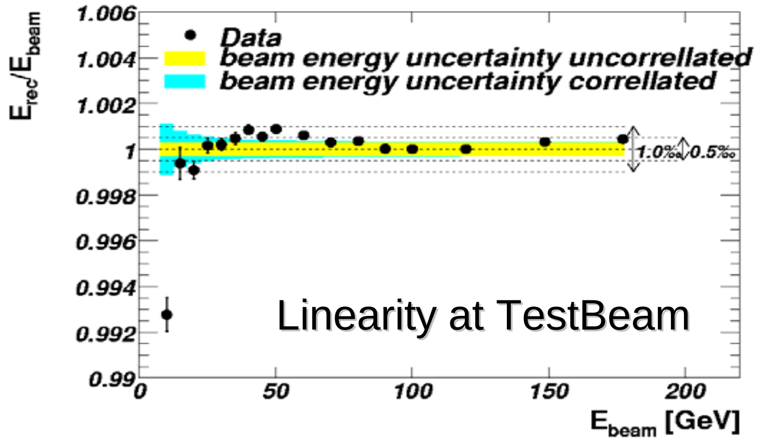


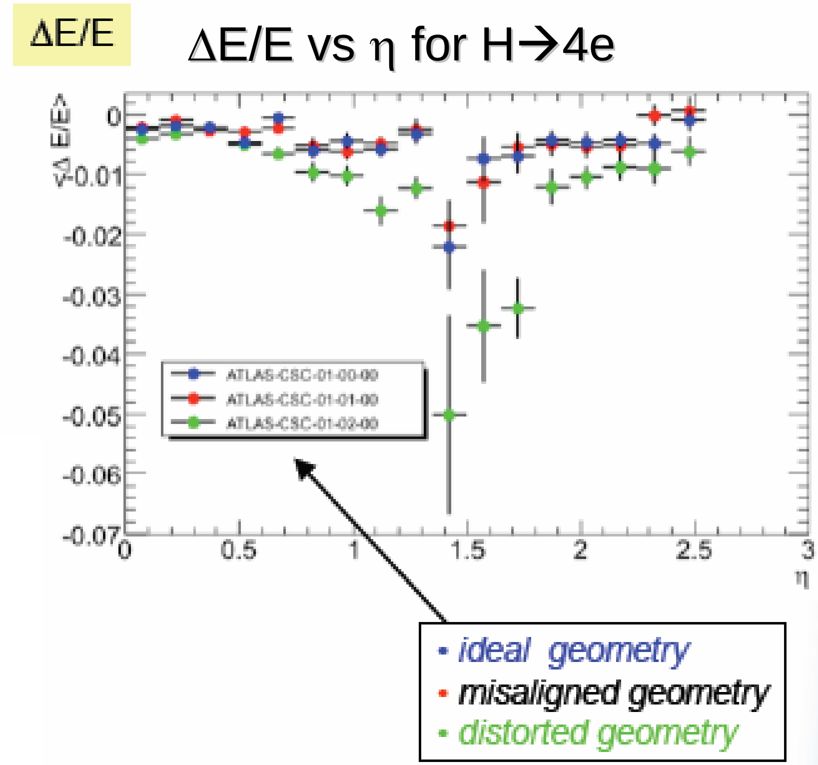
Energy modulations as a function of eta

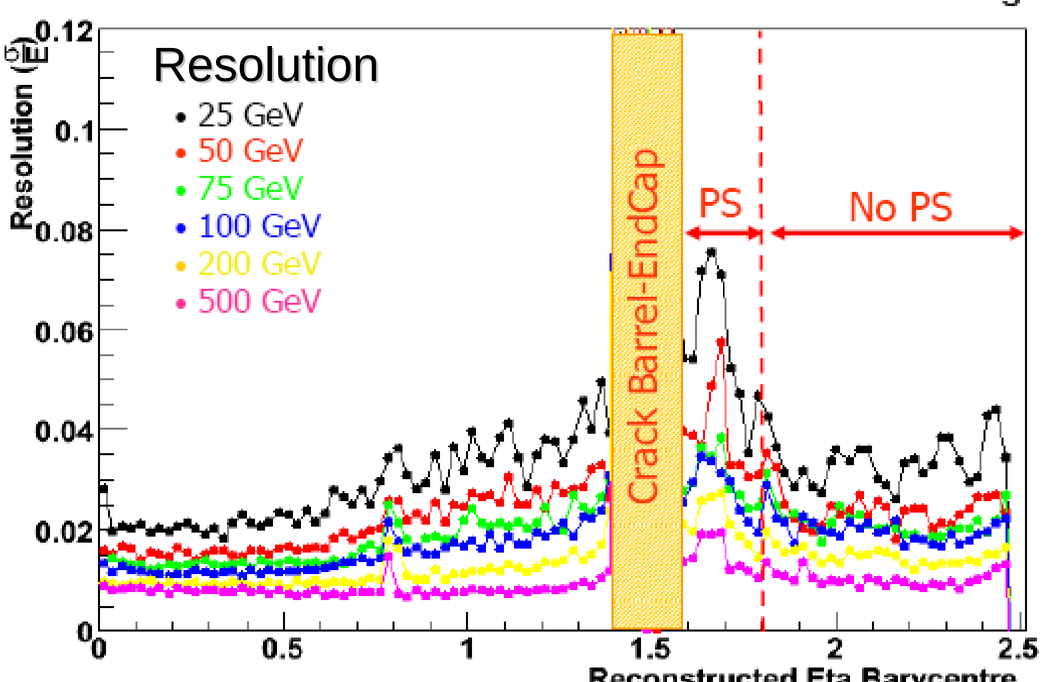
Derived for different cone sizes and eta bins



Calorimeter Performance





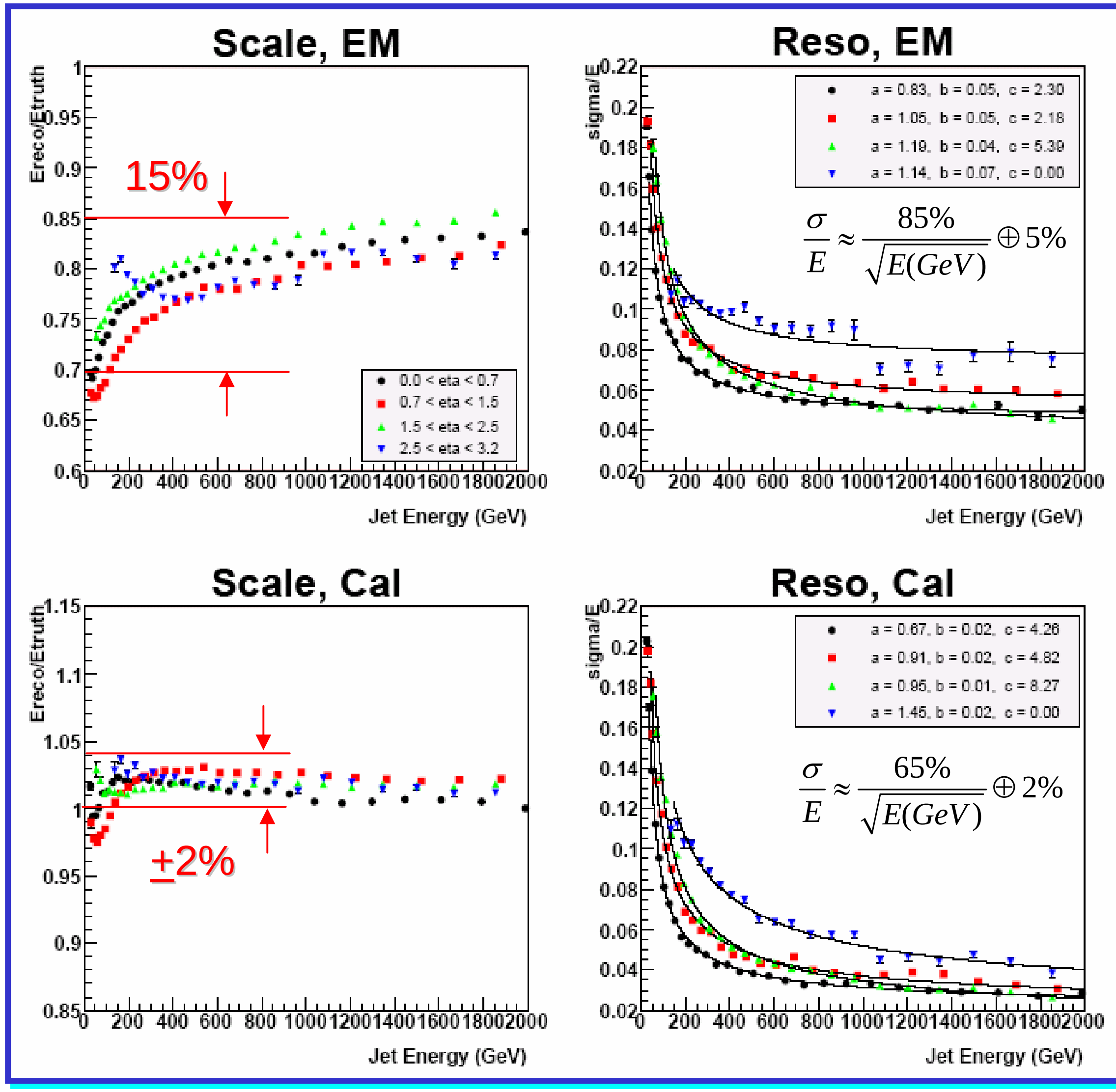


Hadronic Calibration Performance

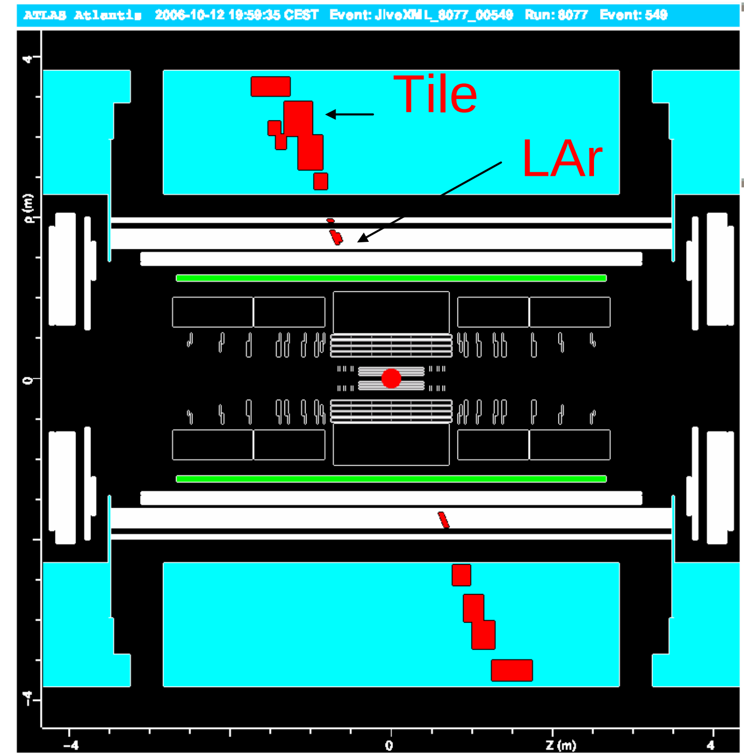
Several calibration schemes under Study. Most developed is:

Calibration derived from observing the density of signal in cone jets (R=0.7). EM Shower are more dense than hadronic shower. This has been derived by F. Paige and is the default in the current reconstruction.

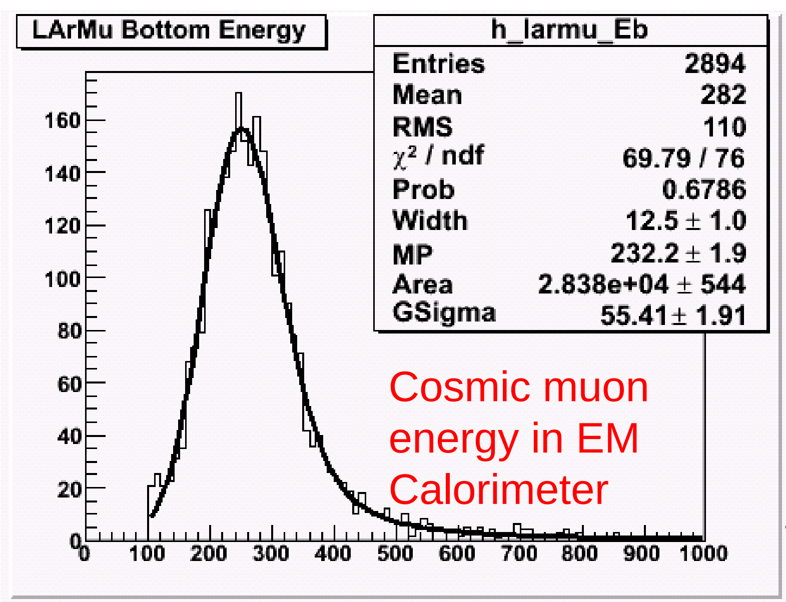
Alternate schemes being developed by other groups.

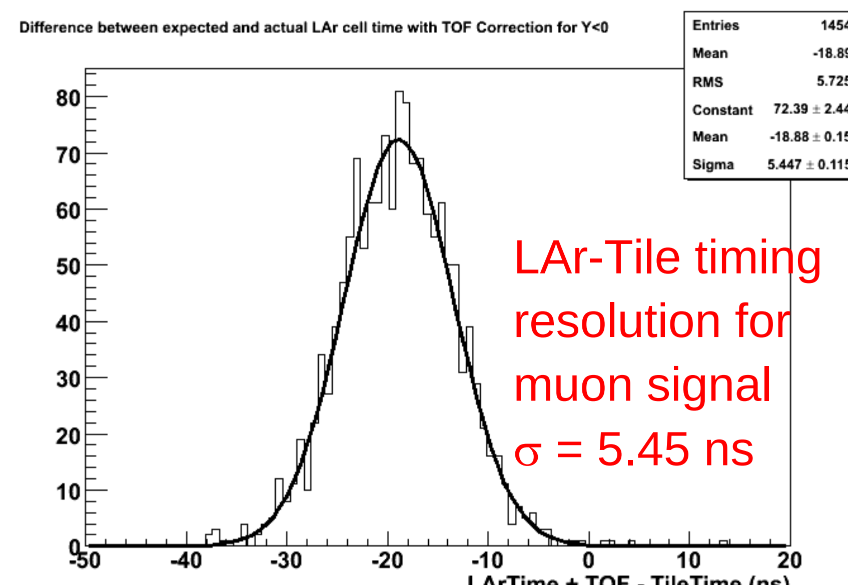


Calorimeter Commissioning Analysis



- H. Ma: LAr calorimeter commissioning analysis co-coordinator
  - Electronics calibration
    - Calibrating 180k channels
  - Cosmic muon data analysis
    - Collecting cosmic muon data since 8/2006
    - Evaluating calorimeter performance
- Integrated detector cosmic tests from now through summer.



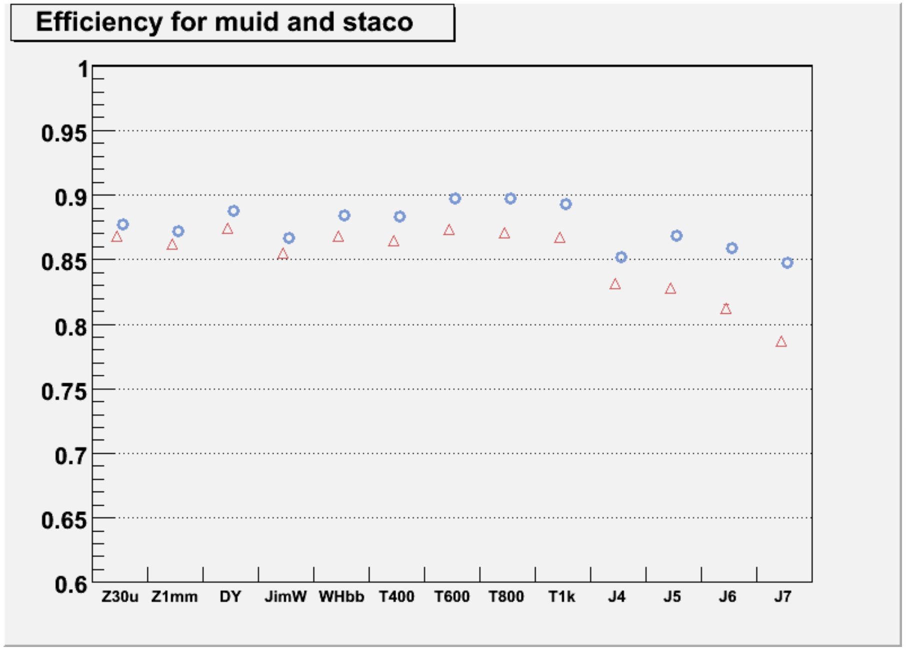


Application Software: Muon Reconstruction

- BNL is primarily involved in the development of :
  - The Muon Event Data Model
    - K. Assamagan
  - Contributions to the CSC reconstruction software
    - K. Assamagan
  - Validation and optimization of the Muon Reconstruction software
    - D. Adams

Muon reconstruction efficiency

D. Adams

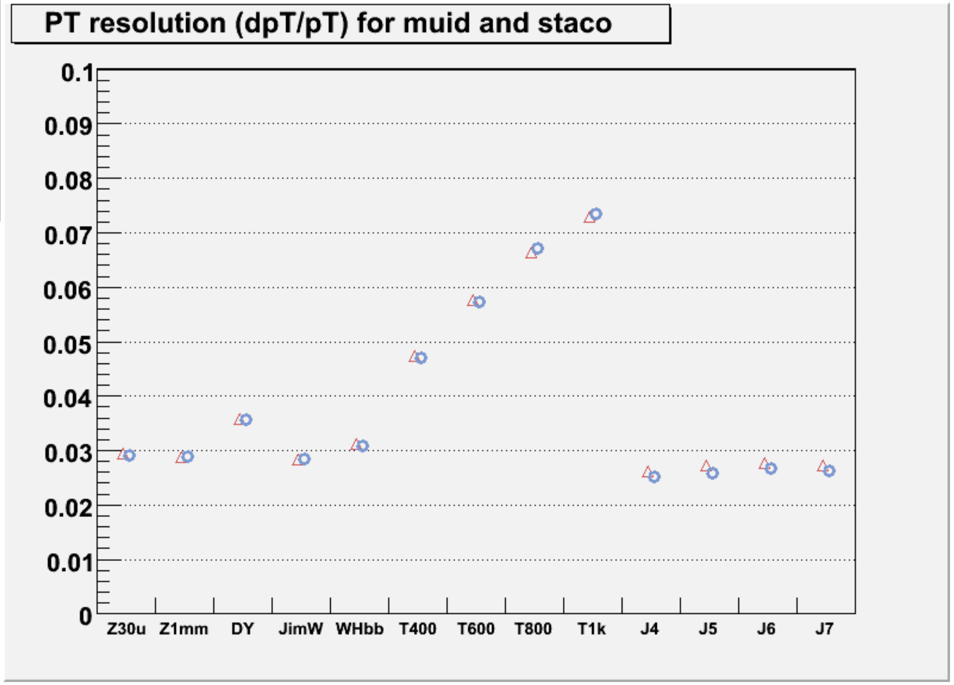


Muon Efficiency for several processes

For PT > 4 GeV and |η| < 2.8

Two primary muon reconstruction programs compared

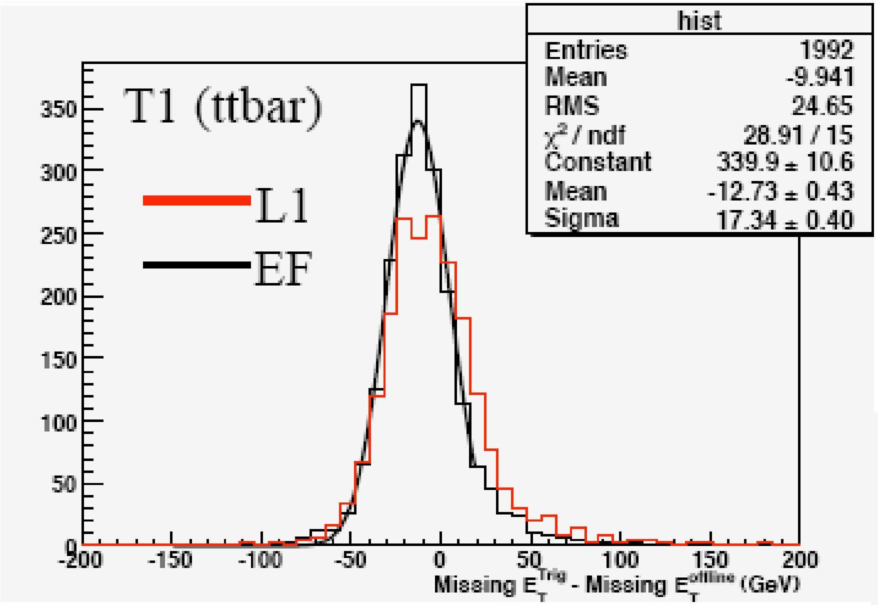
PT resolution in various processes

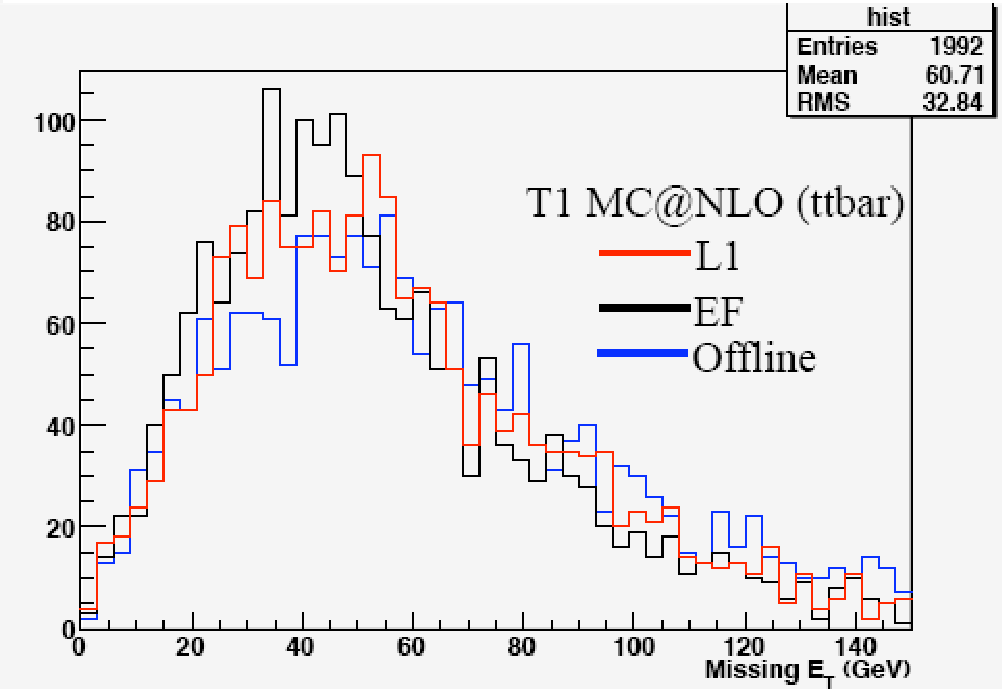


Application Software: Trigger

- Development of e-gamma L2 trigger algorithms
  - D. Damazio
- Development of Missing ET & Jet algorithms for HLT
  - K. Cranmer
- Software infrastructure contributions such as support for DataModel, bytestream, navigation, etc.
  - K. Cranmer, D. Damazio, H. Ma, S. Rajagopalan
- Trigger Menus
  - S. Rajagopalan

HLT Missing ET Resolution for ttbar events





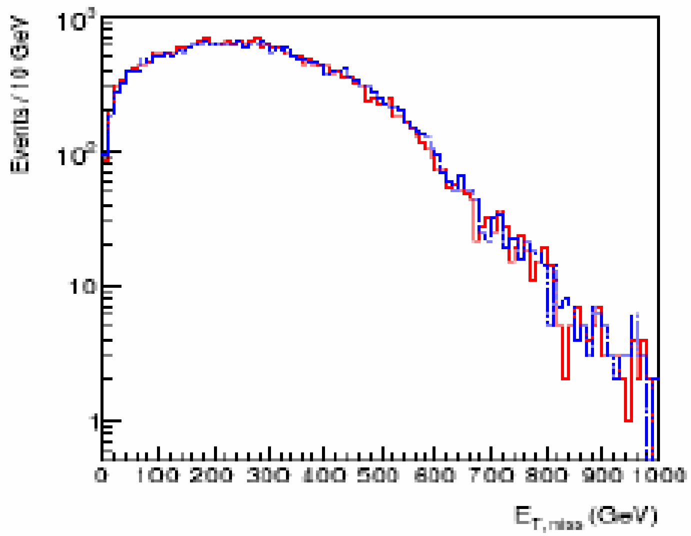
Comparison to Offline:

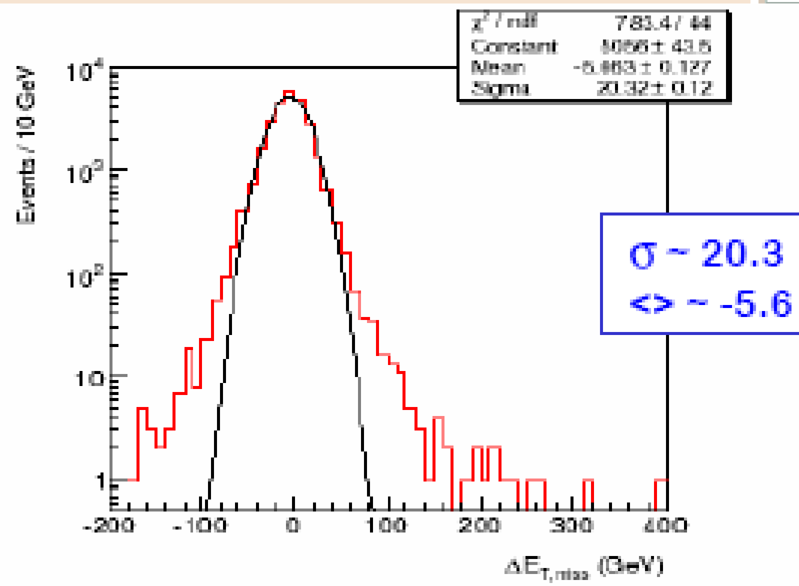
- NO calibration nor noise suppression applied at Trigger (Event Filter) stage yet.
- Good correlation seen between Trigger and Offline.

Combined Reconstruction Software

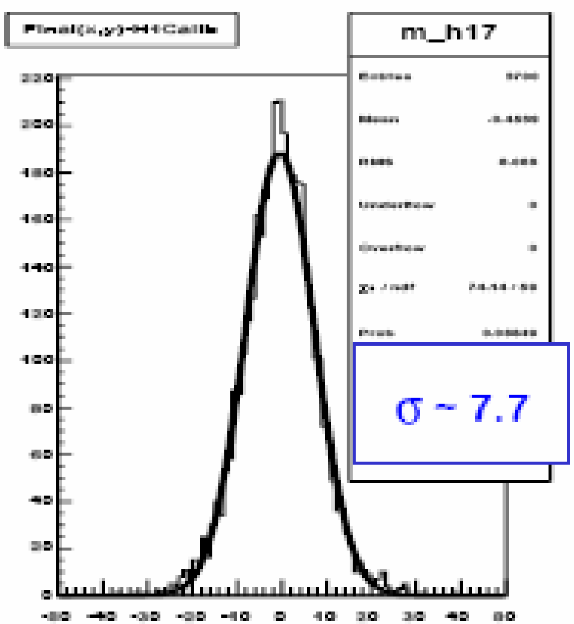
- e-gamma software (K. Assamagan, K. Cranmer, S. Rajagopalan)
  - Design and development of the e-gamma reconstruction software
- Jets (K. Assamagan, K. Cranmer, F. Paige)
  - Optimization of Jet Algorithms
  - Incorporation of hadronic calibration in Jet Algorithms
- Taus (K. Assamagan, A. Cunha, K. Cranmer)
  - Optimization of tau reconstruction algorithms
- Muons (D. Adams, K. Assamagan)
  - Validation of combined muon algorithms
- In all, we have significantly contributed to the overall design of the combined reconstruction algorithms, its Data Model and its subsequent use in Physics Analysis.
  - This knowledge is an asset during analysis of physics data.

Missing ET Performance





Validation of Missing ET in SU3 events (F. Paige)



Missing ET Resolution in Z $\rightarrow \tau\tau$

Major events in FY07

- Integrated Cosmic Ray Test.
- Calibration Data Challenge.
  - Involves our ability to reconstruct a mis-aligned and mis-calibrated detector.
- Full Dress Rehearsal.
  - A full chain test to stress test the mechanics: From writing out data, streaming, reconstruction to distributing it to Tier1/Tier 2 centers and subsequent distributed analysis. 900 GeV commissioning run.
- Each of these tests are designed to stress test the overall ATLAS software preparing us for the data taking phase.

Concluding Remarks

- The BNL group is playing a significant role in the ATLAS software development process.
  - Almost 15 FTE involved in ATLAS specific core software, sub-system & combined reconstruction software and development of physics analysis tools.
- Series of exercises planned this year to ensure readiness for the data taking phase. The main emphasis during the coming year is validating the software and ensuring robust software performance.
- We have built a strong foundation of expertise in the underlying software. This is an asset that will propel us rapidly to take on the challenges of LHC physics.

Calibration Data Challenge

- Demonstrate and commission the calibration ‘closed loop’:
  - Simulate events with an imperfect (i.e. realistic) detector
  - Reconstruct them with imperfectly known calibration constants
  - Improve the calibration using calibration/alignment procedures, re-reconstruct and demonstrate performance improvements
- Exercising various aspects of software and computing model
  - Simulation and reconstruction of a non-ideal detector
  - Calibration algorithm processing in offline software framework
  - Interactions with the conditions database - storage, access, replication
  - Offline production system issues: Bookkeeping, calibration versions
- More ambitious goals:
  - Combining calibration/alignment information from different subdetectors
  - Learning how to do calibration/alignment on ‘real’ samples, with ‘real data’
  - Calibrating under time pressure.

Full Dress Rehearsal

- Complete exercise of the full chain, from Trigger to Distributed Analysis,
  - Generate 10^7 events. Few days of data taking at L = 10^{31} $cm^{-2}s^{-1}$
  - Mix and Filter events to get correct physics mixture as seen at the output of HLT.
  - Pass events through G4 simulation (as-built geometry)
  - Run Level-1 simulation
  - Production bytestream -> emulate raw data.
  - Pass data through HLT nodes, write out events into streams
  - Send data to Tier0, manipulating/merging as expected
  - Perform calibration/alignment at Tier0
  - Reconstruction at Tier0 and produce ESD, AOD, TAG, DPD
  - Distribute to Tier-1 and Tier-2, replicating databases as well.
  - Perform Distributed Analysis using TAG, produce addition group-specific DPDs.
  - Data Quality/monitoring during all stages of processing.