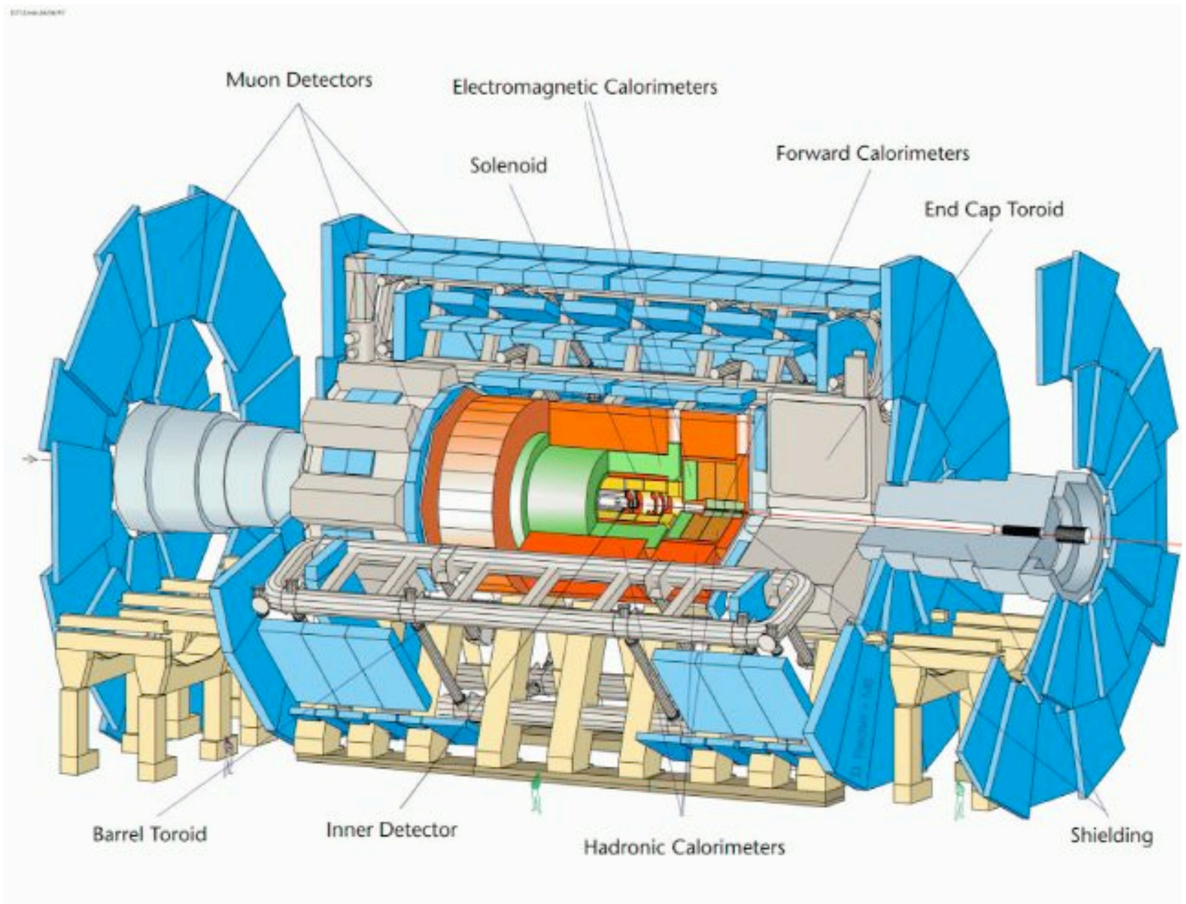
ATLAS: Software and Physics Preparations

Beate Heinemann

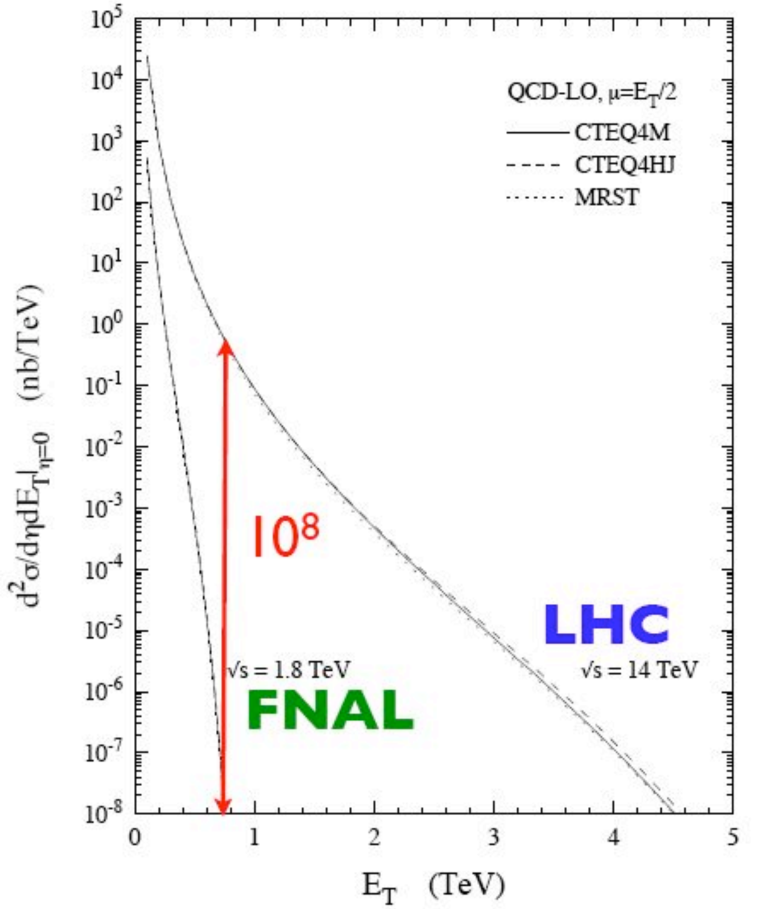


DOE review at LBNL, March 22nd 2007

Physics Opportunities at LHC

Cross Sections of Physics Processes (pb)			
	Tevatron	LHC	Ratio
$W^{\pm}$ (80 GeV)	2600	20000	10
$t\bar{t}$ (2x172 GeV)	7	800	100
gg->H (120 GeV)	1	40	40
$\tilde{\chi}^{+}_{1}\tilde{\chi}^{2}_{0}$ (2x150 GeV)	0.1	1	10
$\tilde{q}\tilde{q}$ (2x400 GeV)	0.05	60	1000
$\tilde{g}\tilde{g}$ (2x400 GeV)	0.005	100	20000
Z′ (1 TeV)	0.1	30	300

Jet Cross Section



- Amazing increase for strongly interacting heavy particles
  => opportunity for discovery with early data!

Challenges for Data Analysis

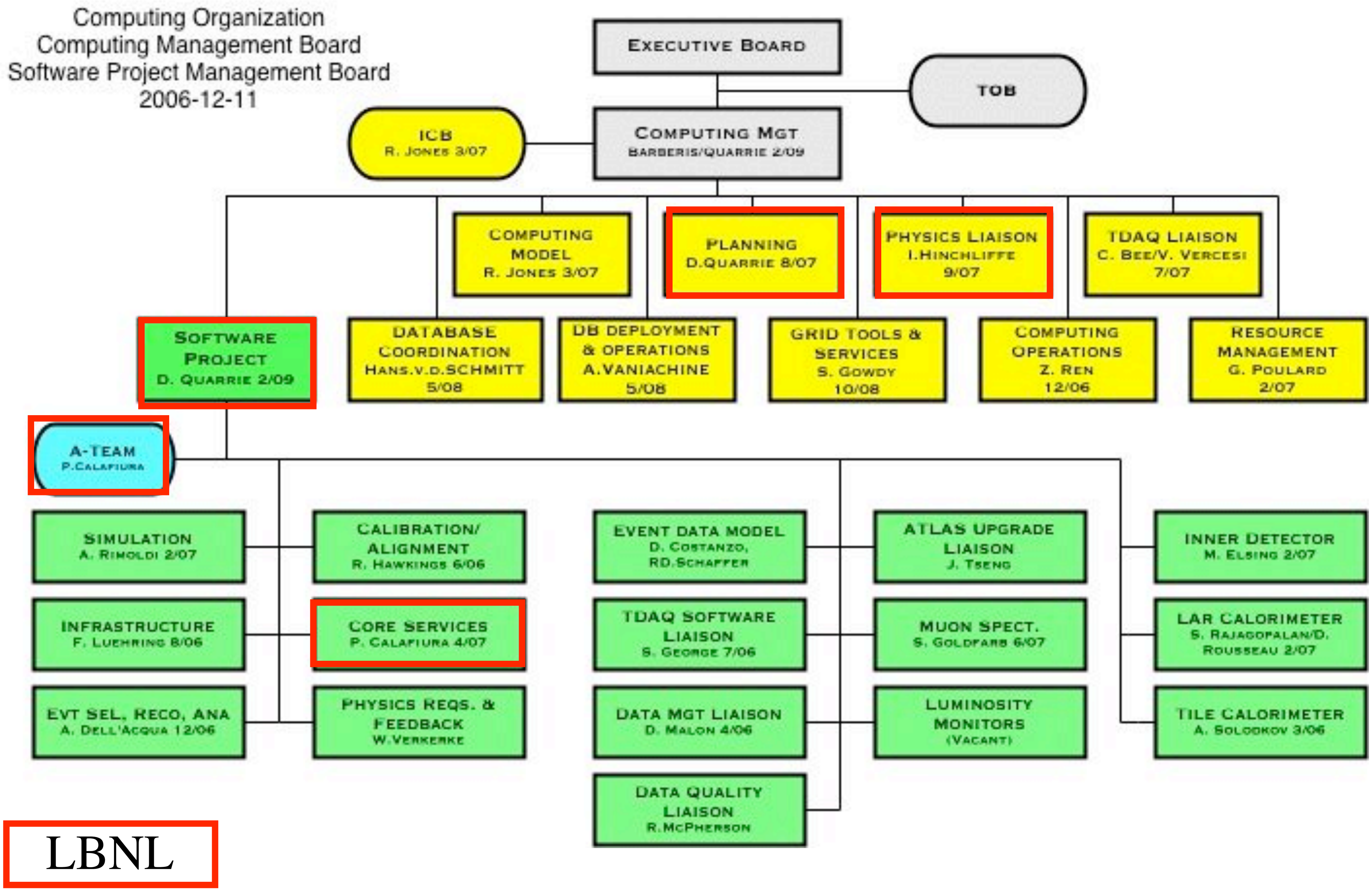
- Access to the data (and MC):
  - Anticipated problems:
    - data retrieval from Grid
    - Software reliability
    - Bugs etc. => validation pivotal
  - ATLAS physics coordinator (Ian H.):
    - "We don't want to look for the data while CMS is looking for the Higgs"
- Understanding of trigger and reconstruction:
  - Trigger, dataset definition, luminosity and data quality:
    - The key to all physics analyses
  - Tracking
    - Electrons, muons, taus, b-tagging
  - Calorimetry:
    - Jets, missing E_T, electrons

Software

LBNL Software Activities

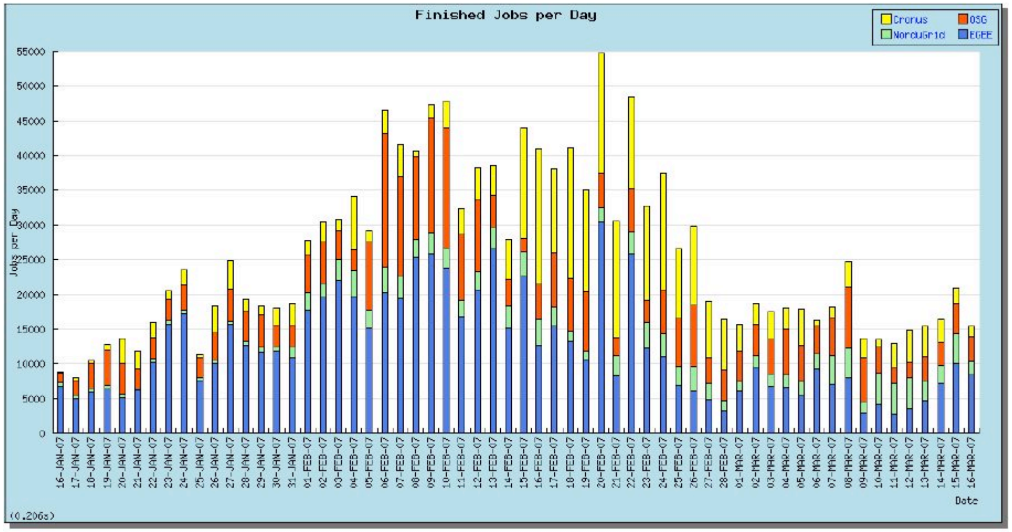
- Core software
  - Project leader: D. Quarrie (until 02/09)
- Pixel software
- Tracking validation
- Alignment
- Data preparation: Streaming Test
- Luminosity Task Force
  - Chair: M. Shapiro
- CSC notes:
  - overall Physics Coordinator: I. Hinchliffe (until 09/07)

Computing Organization



ATLAS Core Software

- Is in commissioning phase
- Current Activities:
  - Detector Commissioning
  - Computer System Commissioning (CSC)
  - Service Challenge
  - Physics Sample Production
  - Calibration and Alignment Challenge
  - Cosmic Ray Tests
- Core Software effort dominated by SW professionals



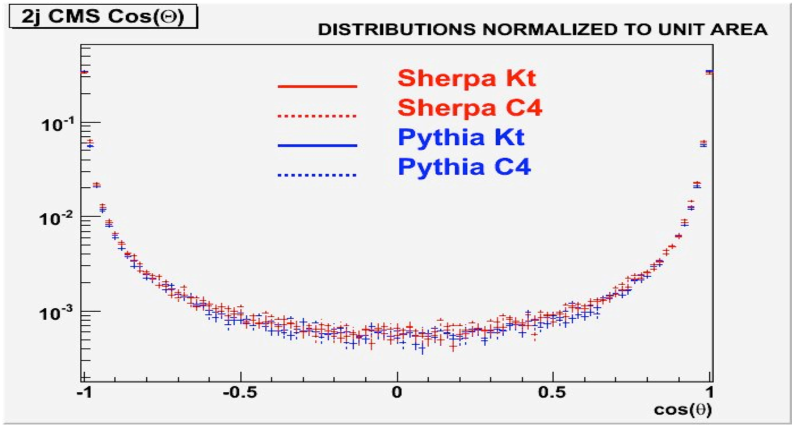
1 job=50 simulated events, 250 reco events, 5K evgen events

LBNL Major Contributor to ATLAS Core Software

- Unique collaboration with NERSC
  - Strong team of software professionals with real-world HEP experience
  - Cross fertilization among many cutting-edge projects
- LBNL members playing leadership roles:
  - Dave Quarrie: Software Project Leader since 1999
  - Paolo Calafiura: Chief Architect; Heads Framework project
  - Charles Leggett: Gaudi core (joint development with LHCB)
- Other important technical roles:
  - Job Configuration
  - Usability
  - Performance Profiling and Performance Optimization

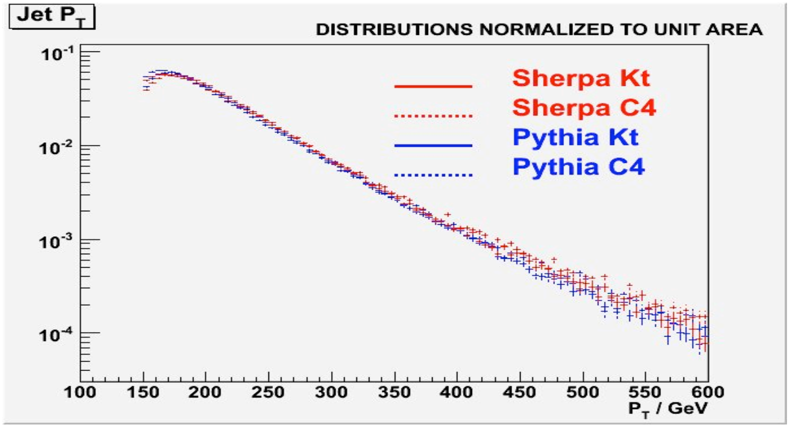
Monte Carlo Event Generators

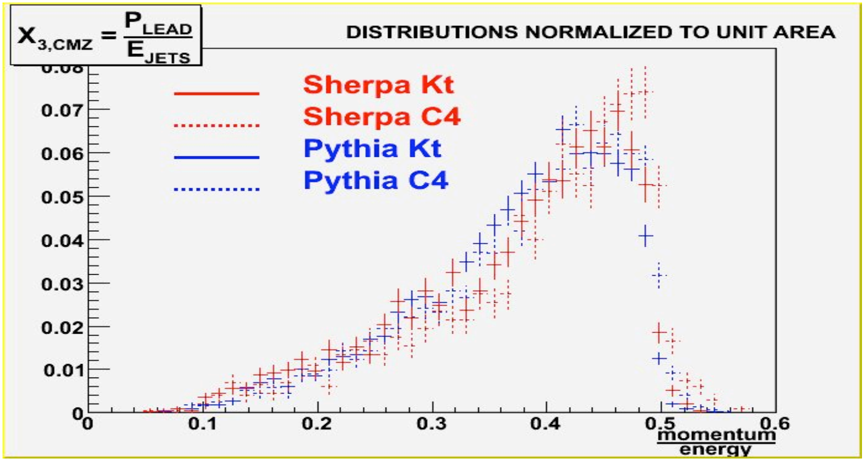
- **LBNL responsible for basic support of MC generators**
  - **Common interface software**
  - **Maintenance of Herwig, Pythia, Jimmy, Sherpa, Alpgen, MC@NLO**
  - **Validation of new releases**



- **Example: Comparison of Sherpa and Pythia multijet production**

plots by: Joe Virzi (grad student)





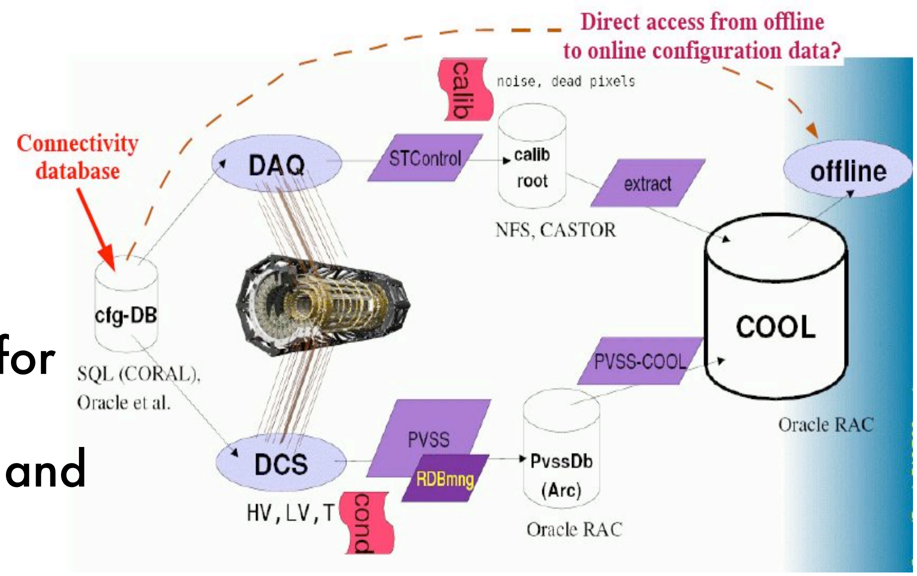
Computing Systems Commissioning (CSC)

- Series of independent well defined tests of computing and software
  - Complements ATLAS hardware commissioning
- Generation and simulation of 10M event sample
  - Representative of Early Physics
  - Analysis Strategies explored using MC samples
  - Physics notes:
    - Homework assigned by physics coordinator to ensure analyzers commission the computing system
- Calibration and alignment challenge
  - LBNL involvement: ID geometry, software and alignment
- Tier 0/Tier 1 Production tests
  - LBNL involvement: inclusion of Data Quality
- Test ATLAS computing model
  - Streaming Test (LBNL leading this effort, see later)

Pixel Detector Software

- Connectivity Database:
  - vital for operation of detector:
    - e.g. voltages, connections, etc. for each module
  - Status:
    - Infrastructure mostly completed
    - Connectivity template worked out for ongoing "connectivity test"
    - Actual DB content for current tests and final detector are being filled in

Beringer, Gaponenko, Zdrazil



- Calibration DB:
  - Needed for offline analysis:
    - e.g. ToT, noise, etc. for each pixel
  - Status:
    - Setup completed
    - Will be filled when data come
- Pixel software:
  - Geometry, bytestream converter, clustering,...

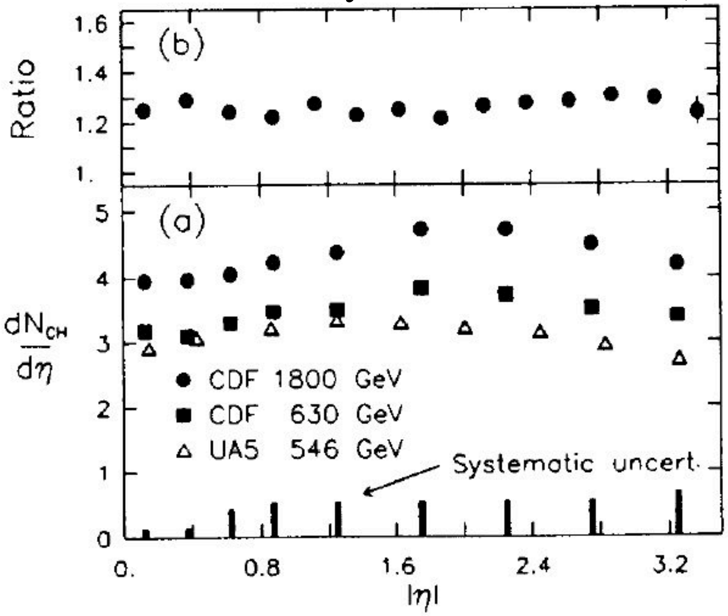
Yao, Zdrazil

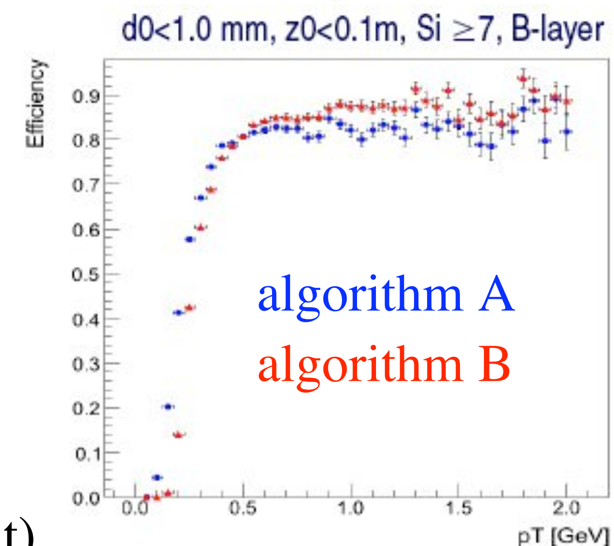
Storage	Definition	Unites	Typical range
1B	threshold	30e	2000-5000 e
1B	dispersion	3e	80 - 600 e
1B	noise	3e	0 - 600 e
1B	timewalk	30e	2000-8000 e
float	A for ToT		0-300
float	B for ToT		
float	C for ToT		
1B	P1 dispersion of ToT	1/100	+-100
1B	P2 dispersion of ToT	1/1000	+-100

Tracking

- Early physics measurement:
  - dN/dη for charged tracks (already in 900 GeV data)
- Desirable to go to low momenta:
  - CDF went down to 50 MeV!
  - Standard ATLAS software requires p_T>500 MeV
- Main focus:
  - Understand efficiency and fake rate of tracks as function of p_T
    - Develop optimal cuts
  - Develop algorithm to track below 500 MeV avoiding too many fake tracks

CDF, Phys. Rev D41, 2330 (1990)

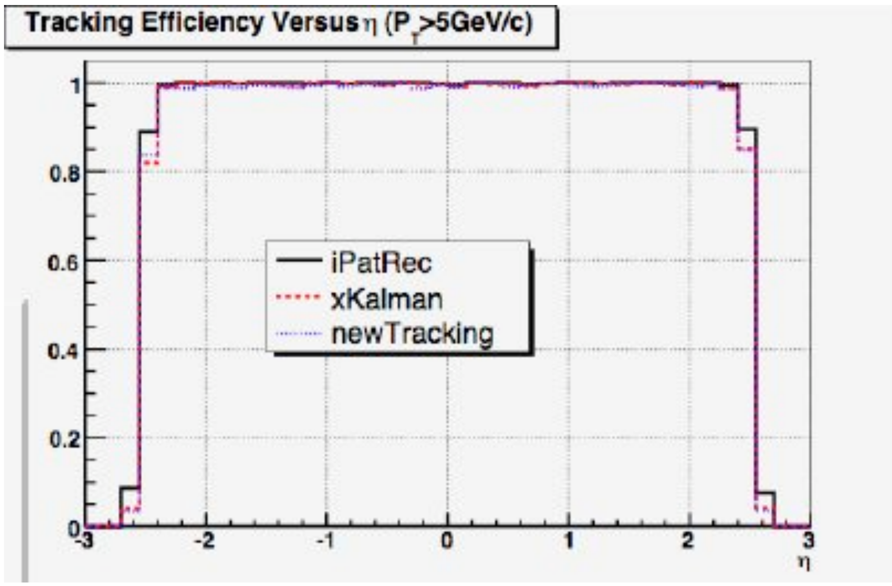


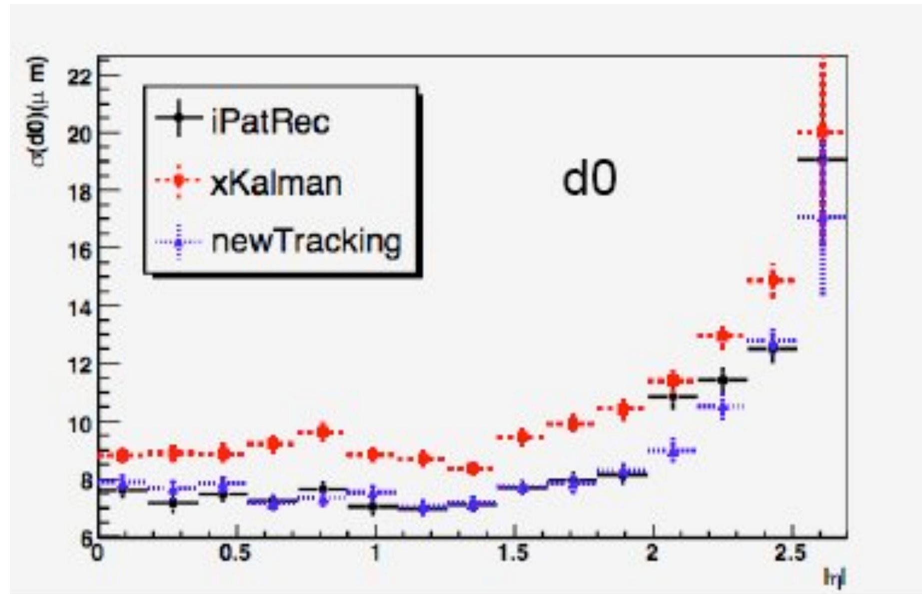


M. Leyton (grad student)

Tracking Validation

S. Zenz (grad student)

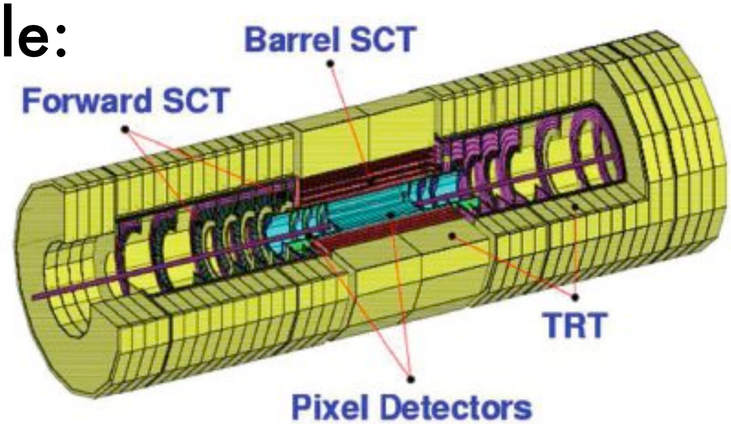




- Validation of latest software releases
  - Comparison of different algorithms
- Commission the algorithms early during development phase to avoid surprises later

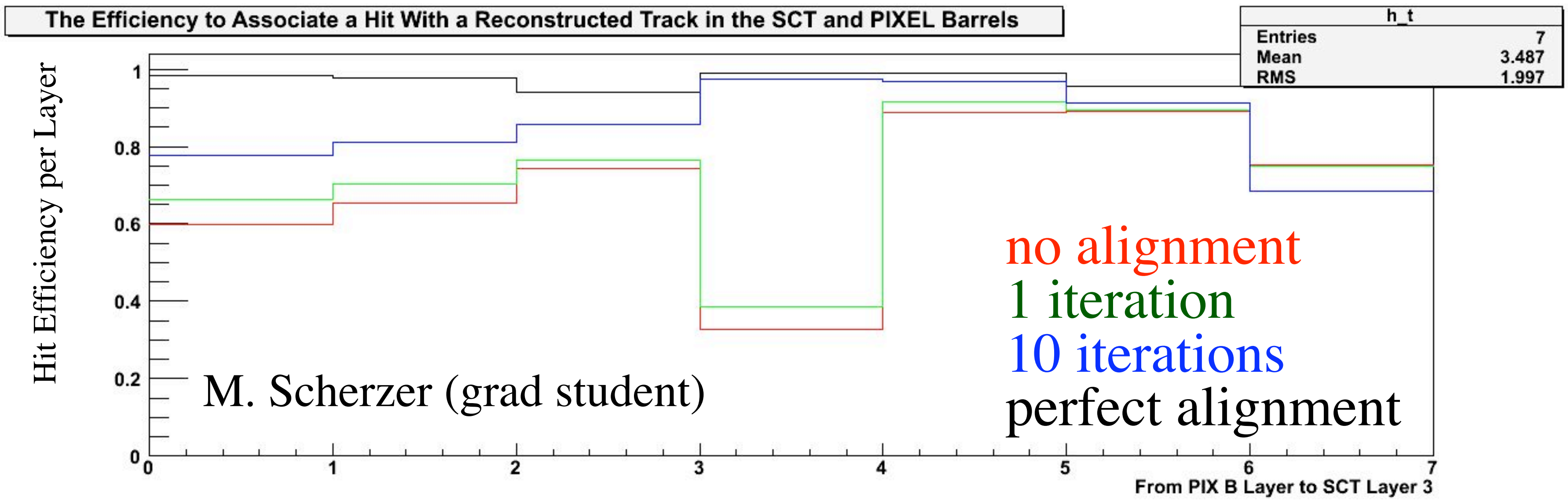
Silicon Detector Alignment

- Not trivial:
  - 1,744 pixel modules and 4,088 SCT modules
  - 6 degrees of freedom per module:
    - 3 translations and 3 rotations
  - 34992 degrees of freedom
    - +possibly module deformations



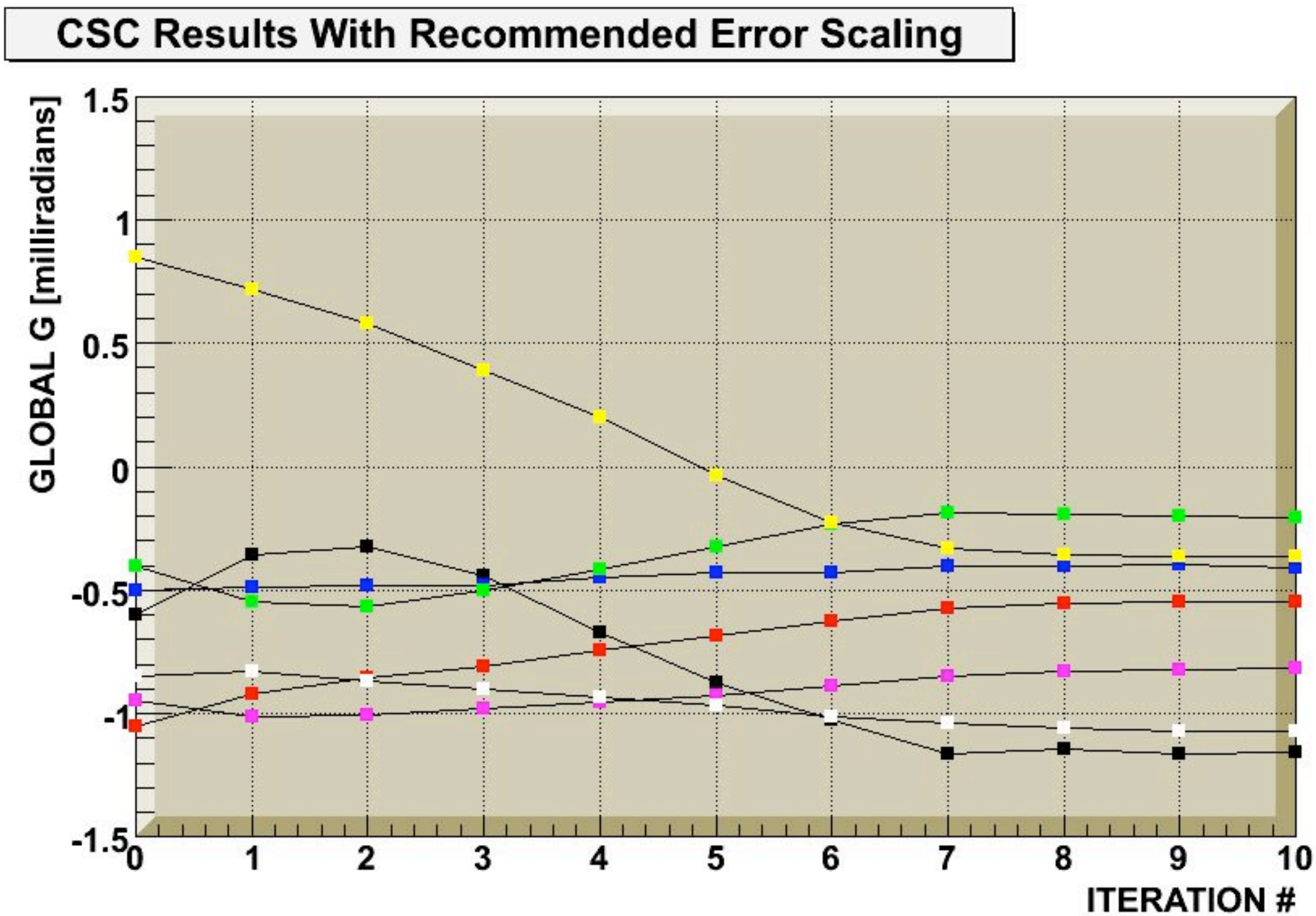
- Two-fold effort:
  - CSC "data" (T. Golling, M. Scherzer):
    - Understand how to constrain large effects with little data: barrel by barrel, disk by disk
      - 7 layers in the barrel and 2x12 disks in forward region
      - Leave module to module for the time being
  - Pixel cosmic data (R. Madaras, W. Yao, M. Zdrazil):
    - Try to actually make an alignment and compare to survey

Tests using misaligned MC



- Compare result of alignment to truly misaligned detector:
  - "as built" geometry
- Large misalignment already seen in hit efficiency plot per layer
  - E.g. 400 μm misalignment in SCT layer 0

MC Alignment Convergence



- Algorithm converges
  - Still understanding whether this result is exactly as expected

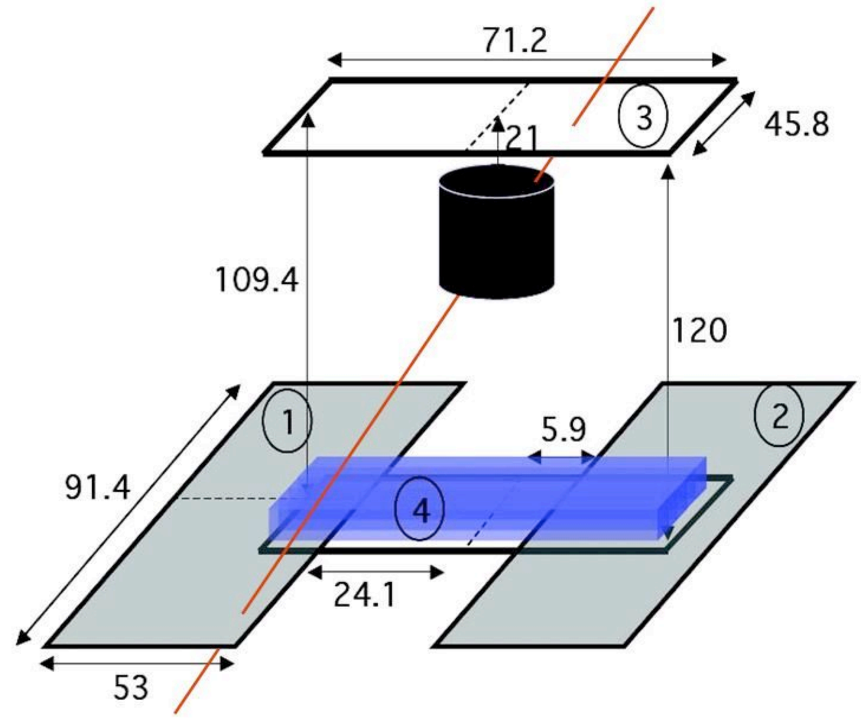
Cosmics in Pixel Endcap A

- Primary goal of “system test” was to operate the detector for the first time
  - Huge success, lots was learned (see talk by Gil)
- Can also use these data to understand the software and the alignment
  - Study noise occupancies, develop rejection cuts
  - Perform full simulation of test setup
  - Perform track reconstruction, study residuals
  - Use event display (ATLANTIS)
  - First look at alignment
    - Compare to precise survey data in endcap
  - Major work to make all this happen:
    - E.g. bytestream converter (Zdrazil), calibration DB (Yao)

Cosmics simulation

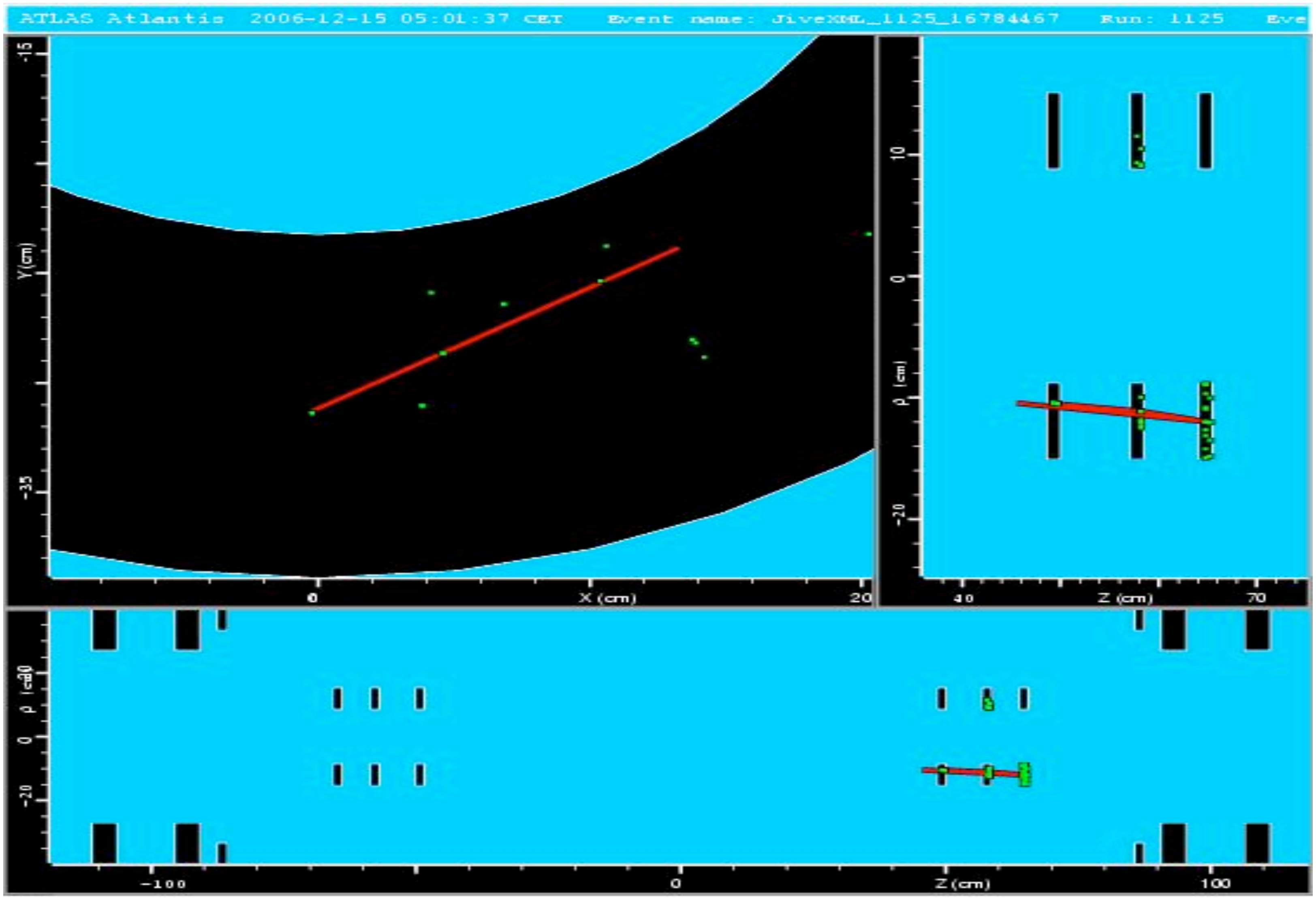
- Toy MC:
  - Muons follow $\cos^2\theta$ distribution
  - Flux: 70 $cm^{-2}s^{-1}sr^{-1}$ for p_T>1 GeV/c

	simulation	Data
top sc.	54 Hz	n.a.
top and bottom sc.	16 Hz	15.7 Hz
Trigger + 3 hits	1.5 Hz	1 Hz

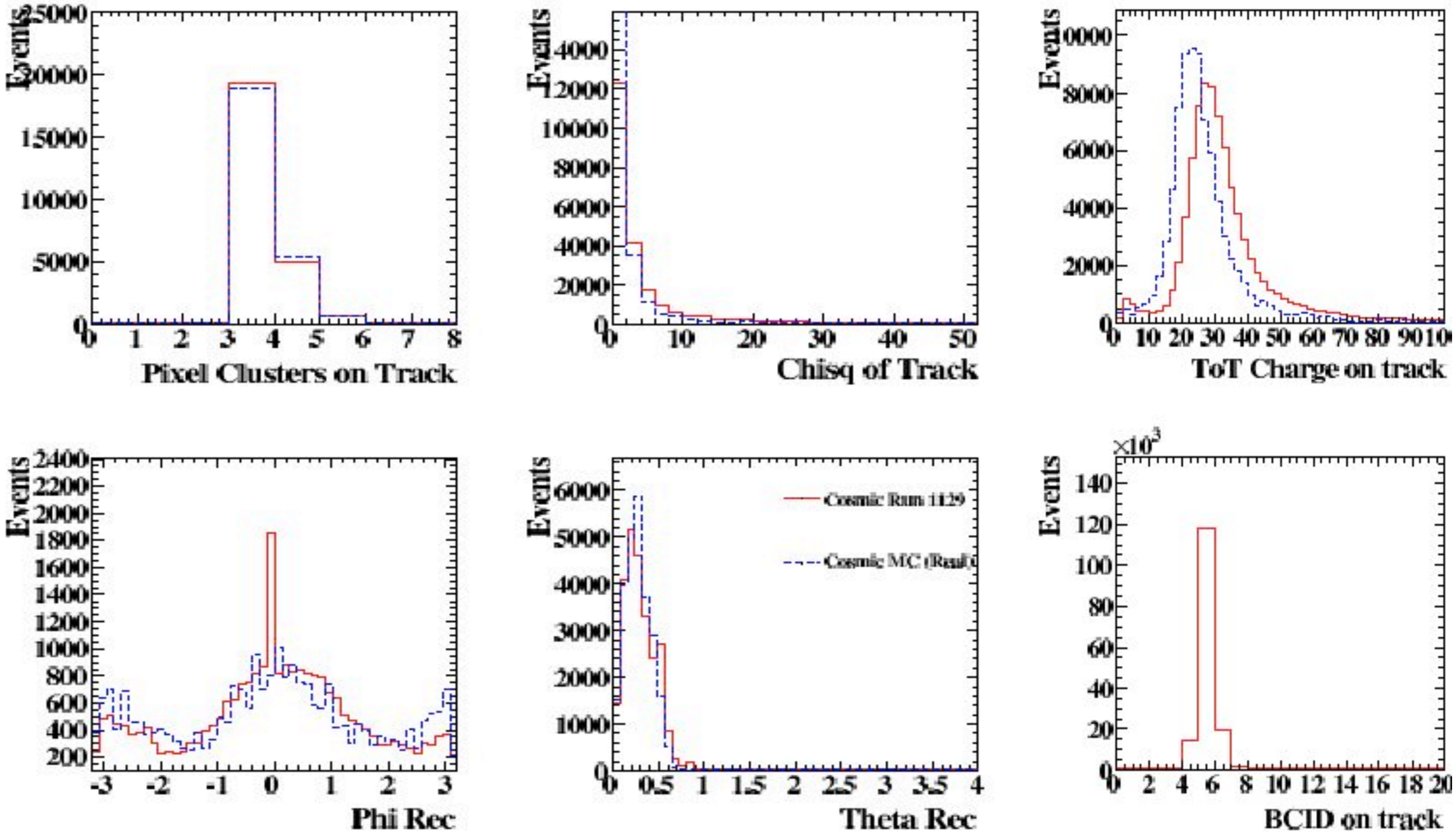


- Commissioning of entire pixel specific software
- Found some problems on the way:
  - E.g. initially efficiency in data low due to phi index definition problem

Event Display: Atlantis



Comparing Data and Simulation



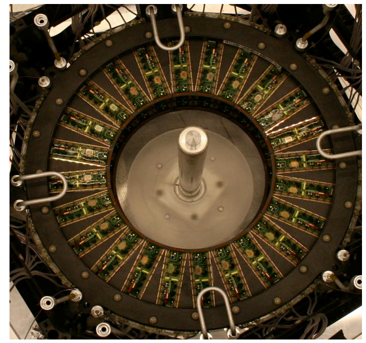
Data
MC

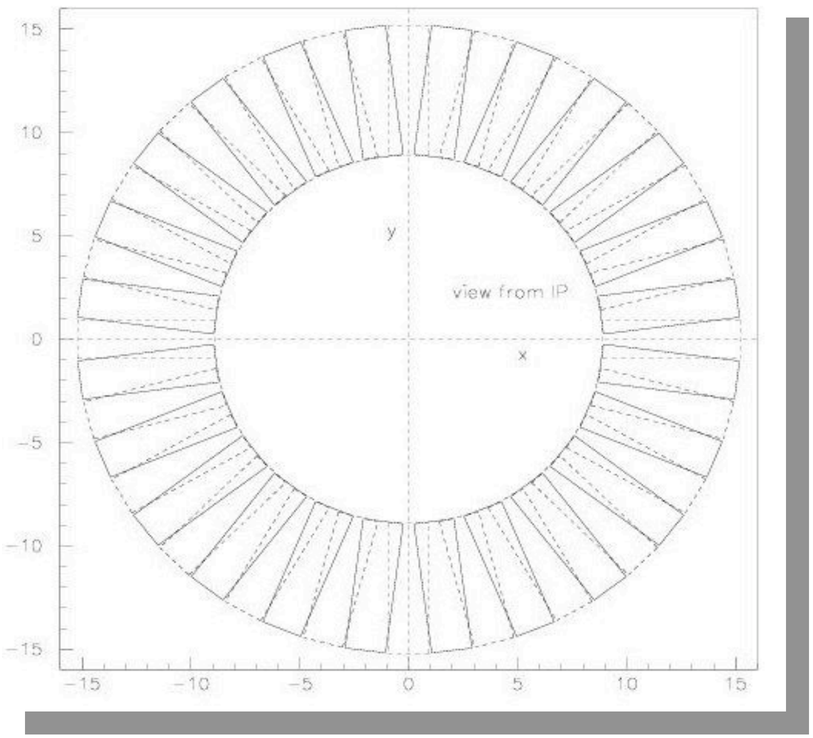
Data Sample	Tracking Rate/event (%)	Overlap Fraction($\frac{\geq 4hits}{\geq 3hits}$)(%)
Data (Nominal)	2.83 ± 0.01	23.4 ± 0.2
Data (Survey)	2.80 ± 0.01	23.7 ± 0.2
Ideal MC	≈ 6	≈ 28
Realistic MC (dead modules)	3.9 ± 0.1	24.6 ± 0.9

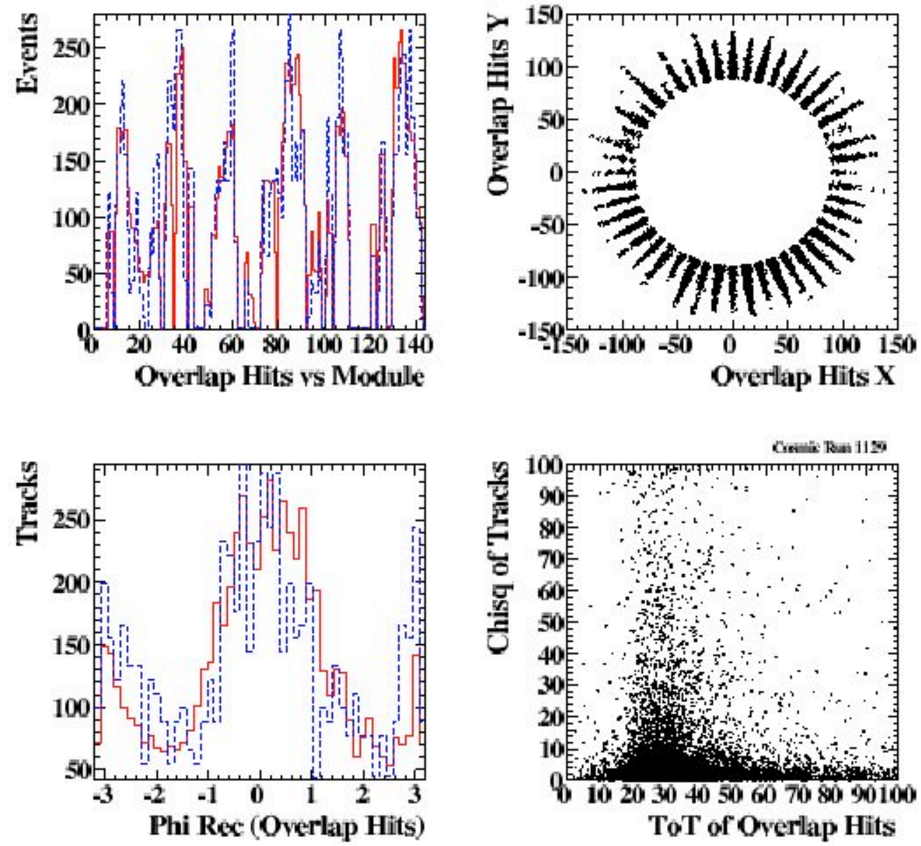
Total: 883,710 events

Overlap hits

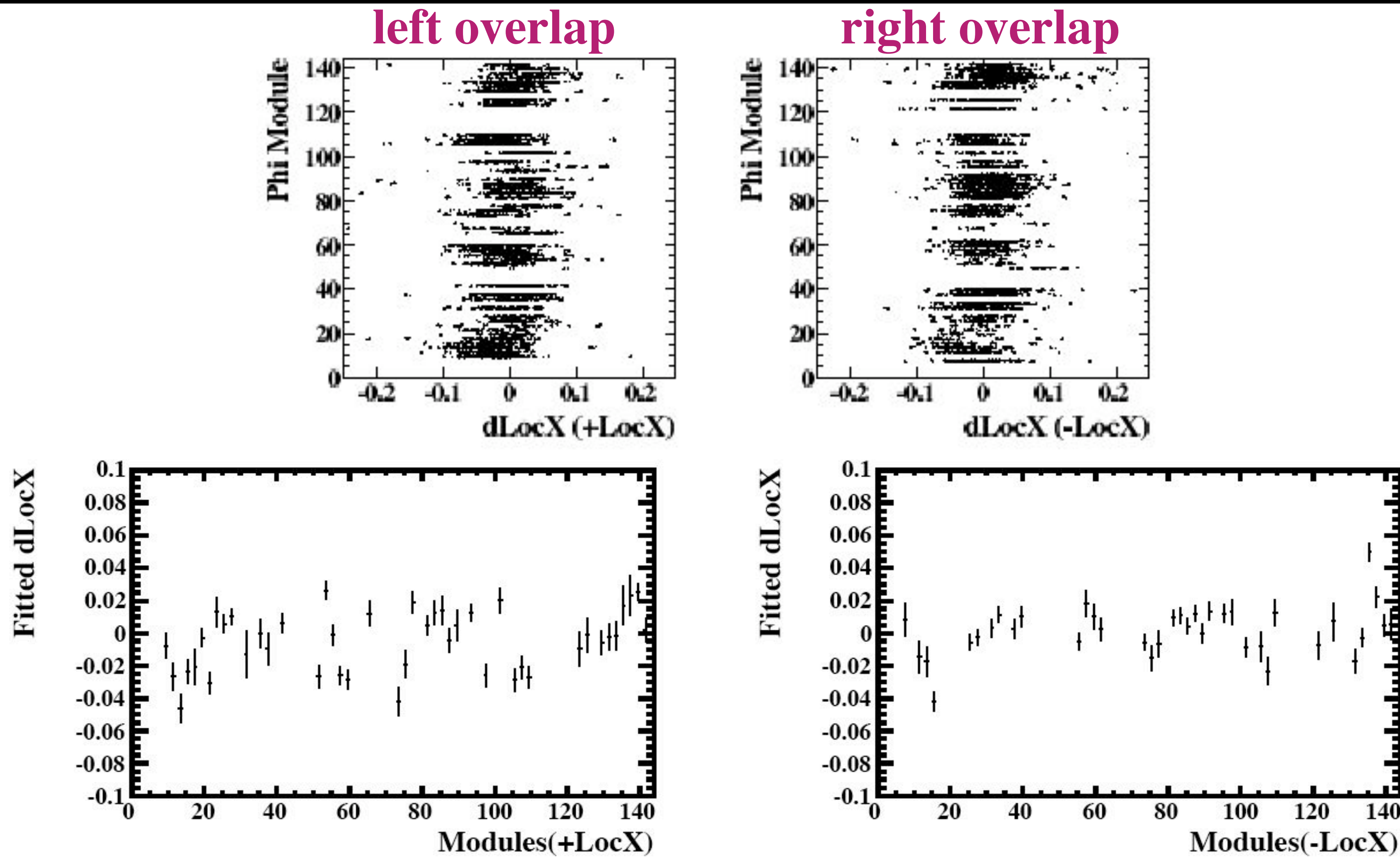
- Modules are on both sides of pixel disks
  - overlap hits very useful for alignment





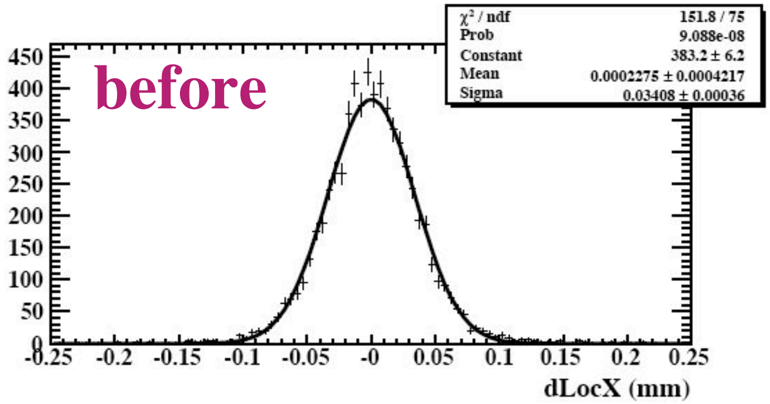


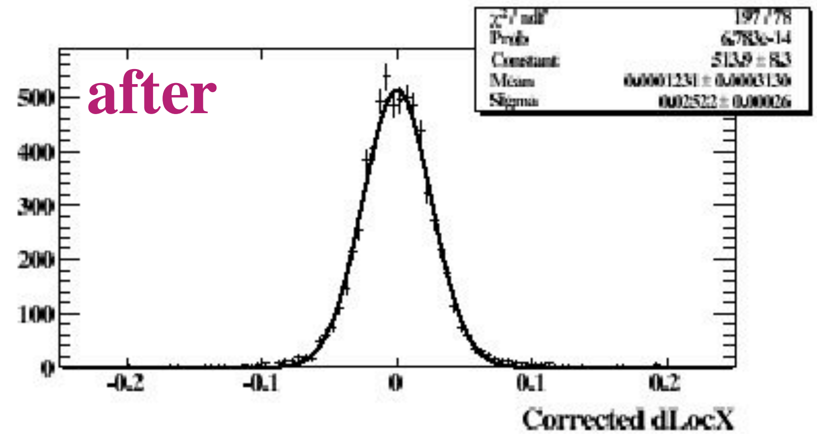
Residuals + Alignment Constants

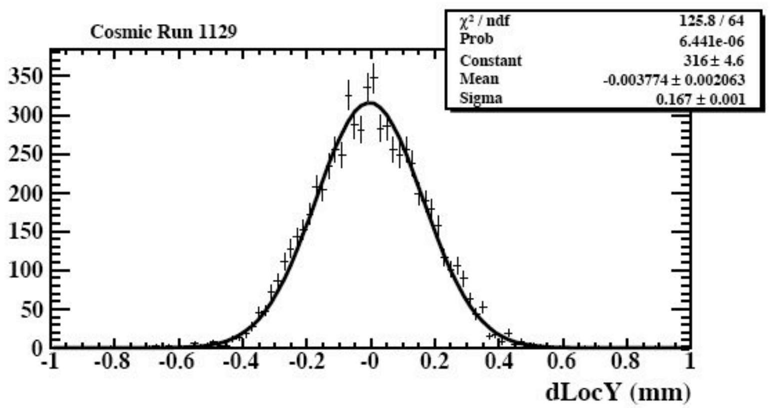


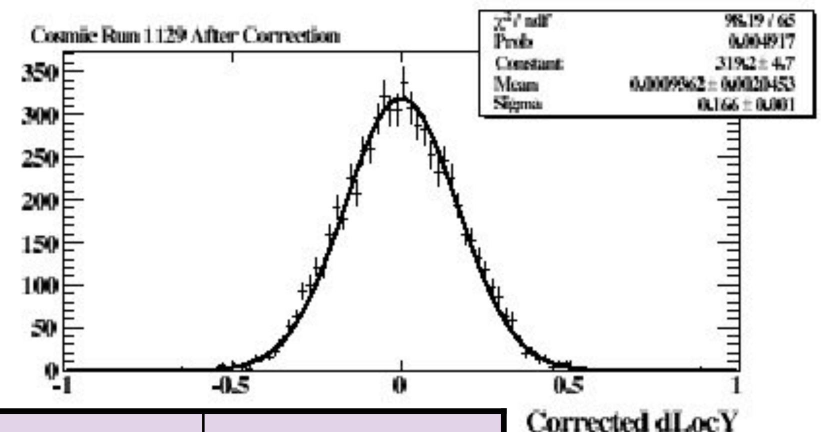
- Shifts generally less than 20 μm!

Residuals before and after alignment







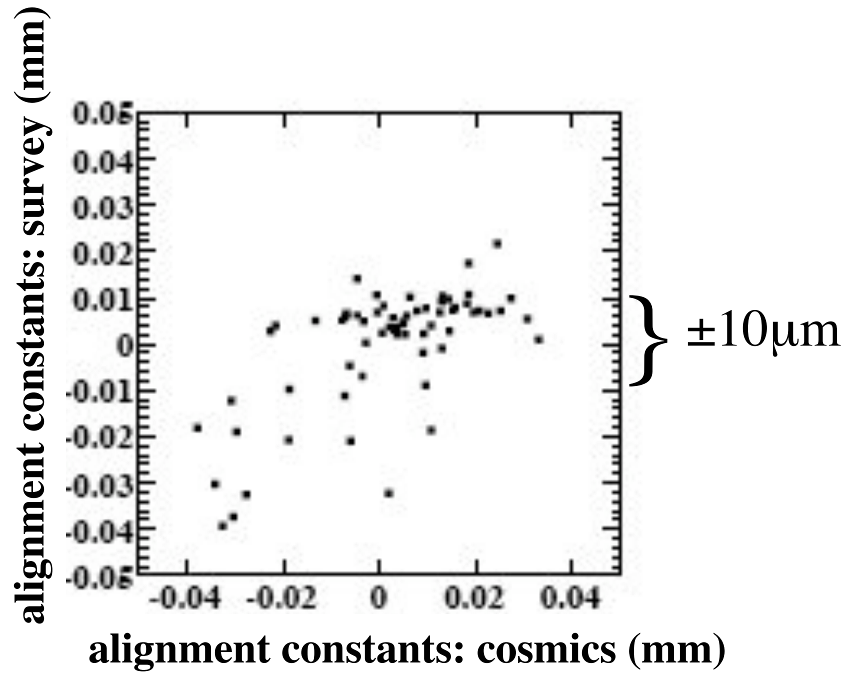


	Before	After	MC
σ(dX)	24 μm	17 μm	16 μm
σ(dY)	117 μm	117 μm	? μm

NB: resolution=√2 of fit shown

Comparison to Survey

- Survey by using SmartScope:
  - 36 measurements per module
  - High precision
    - 1 μm (front) , 5 μm (back)
- Correlation between survey and in situ constants observed:
  - The detector went in a plane from California to Switzreland between!
  - Work on understanding the remaining differences continues



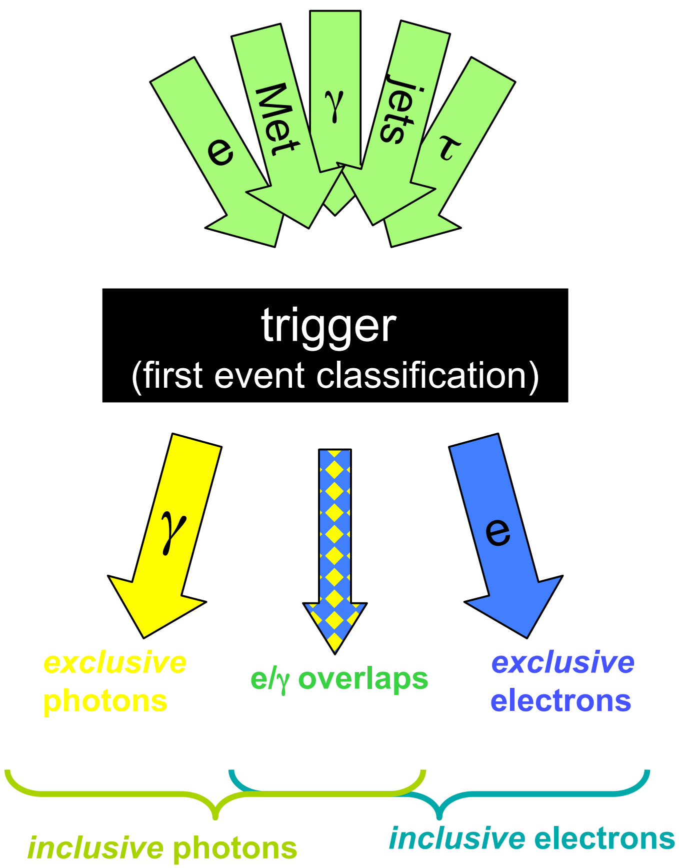
Even without any alignment the precision is 24 μm!

Streaming Test, Luminosity and Data Quality

Data Streaming

- Data get sorted based on which trigger they passed:
  - Faster data access for users
    - User has to run only over relevant stream rather than all ATLAS data
- Data streams had not been foreseen in ATLAS until 2006
  - Main objection to streams is overlap between streams, i.e. writing events twice
  - First studies by J.-F. Arguin showed that overlap rather small
    - Now 5 streams are foreseen: e, μ, jets, τ/MET, γ
- Large scale streaming test now ongoing:
  - Provides also mock data sample for physics studies
  - People: Hinchliffe, Holloway, Shapiro, Tompkins, Zenz

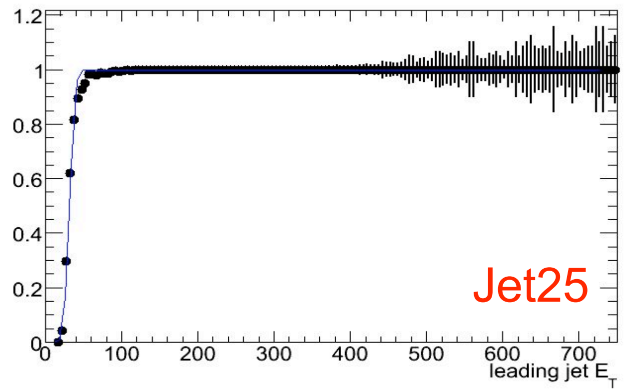
Streaming Test

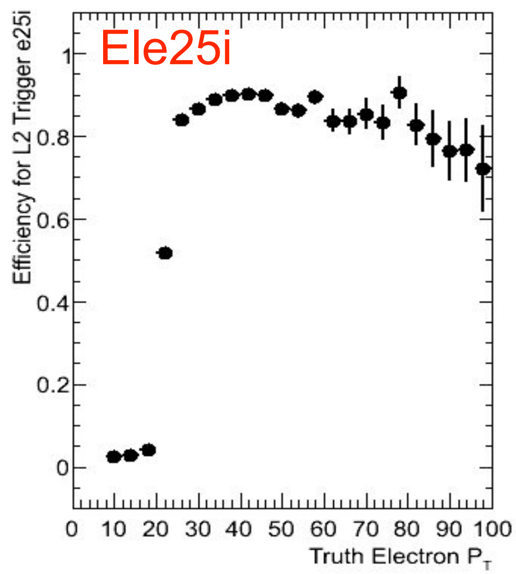


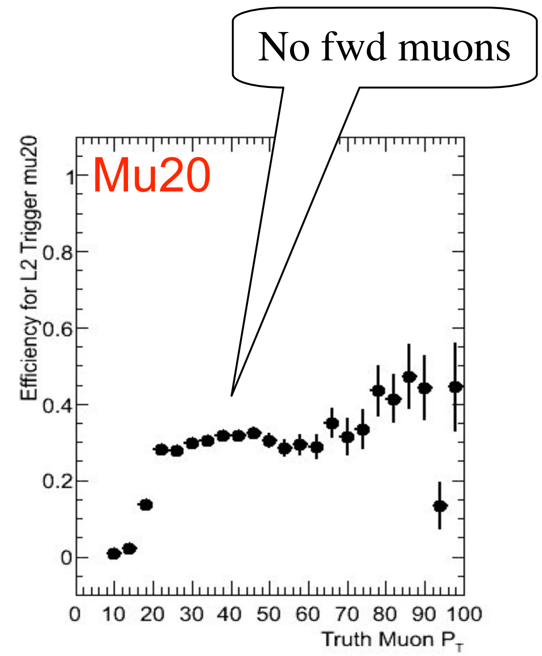
- **Streaming test creates mockup data sample corresponding to 18 pb^{-1} of real data:**
  - Physics processes mixed according to their cross sections (30 SM processes)
  - Trigger decisions applied
  - Prototype luminosity/conditions database is generated in parallel
- **What can we do with these data:**
  - Test of different streaming scenarios
    - Inclusive streaming: events are written twice
    - Exclusive streaming: events are written once
      - all events triggered by >1 trigger go into "overlap stream"
  - Test computing aspects:
    - data handling, offline trigger simulation, luminosity calculation, dataset bookkeeping,...
  - Practice doing physics
    - Measuring e.g. W or ttbar cross sections or backgrounds to new physics

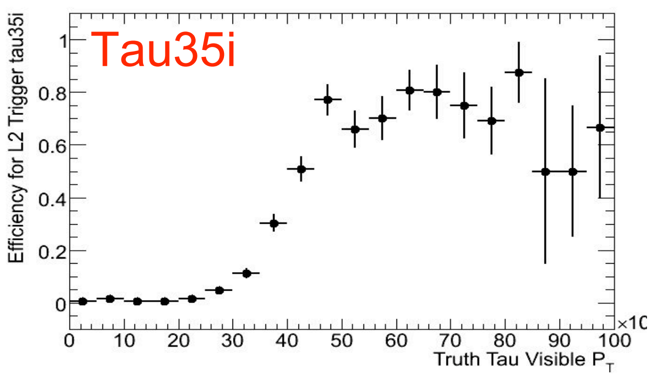
Trigger Validation

- efficiency plots for various trigger thresholds vs. MC truth information: electrons, muons, jets









This looks okay now but interesting problems were found and fixed on the way

Streaming Status/Outlook

- Validation of the trigger and mixing are complete
- 110 raw datasets are available to users
  - Reconstruction is beginning now using release 12.0.6
- Looking at streaming data is mandatory “homework” for each physics group
- The next iteration (updated simulation, more data) will be part of the ATLAS “full dress rehearsal”
- LBL group will
  - apply optimized SUSY search cuts to the sample
  - measure the top-pair cross section with all official/realistic tools

Luminosity and Data Quality

- Luminosity task force work completed:
  - Chair: M. Shapiro
  - Accepted recommendation (in short):
    - Record luminosity in blocks of order 1 minute
- Implications on Data Quality:
  - MetaData task force recommends for detectors to report their validity as function of LB#
    - E.g. HV trips
  - Work ongoing in setting up framework to propagate this recommendation to Data Quality monitoring (Heinemann):
    - Online and offline (Tier-0)
    - Tests of framework ongoing during Tier-0 tests and during M2 weeks (LAr, Tile+Muon combined running)

Physics

Physics Strategy

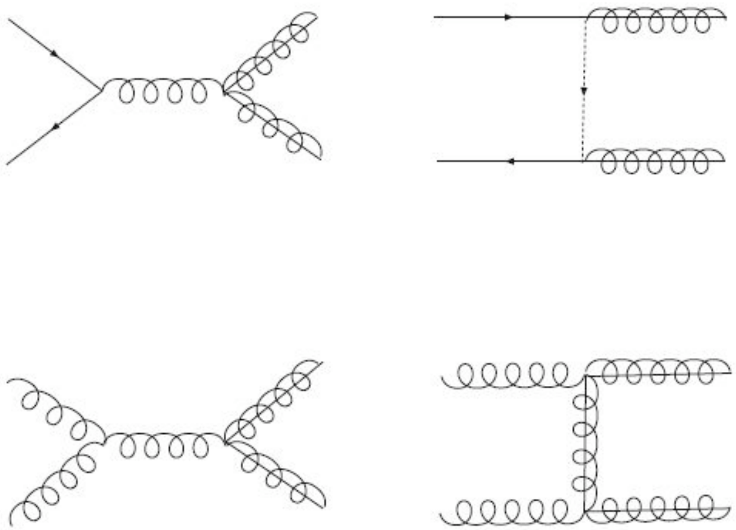
- Focus on early physics:
  - Channels where Tevatron can be beaten very quickly
  - Keep medium/longer term in mind though: e.g. Higgs
- Key ingredients:
  - Understanding of reconstruction software
  - Measurement of backgrounds using data
  - "simple" and robust analyses
- Physics personpower will increase when pixel detector is installed:
  - Workshop planned for August to focus and channel activities
- Group can build on lots of expertise and on major contribution to ATLAS detector construction

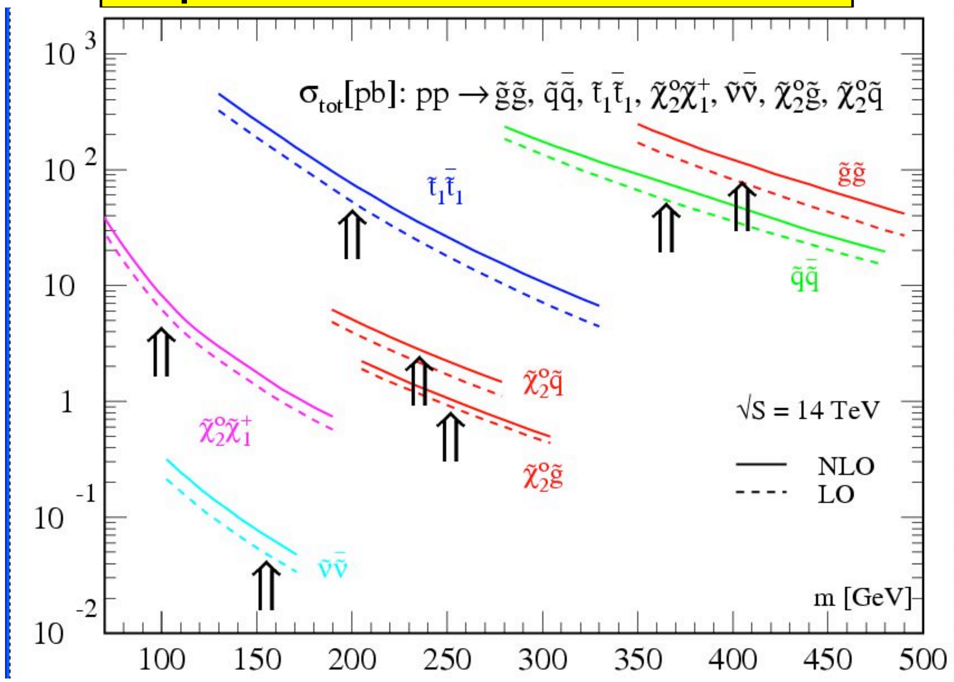
Sparticle production

- Cross-section for producing SUSY particles in proton-proton collisions calculated at NLO
- Production of the *strongly interacting* sparticles dominates: pp→ $\tilde{g}\tilde{g}, \tilde{q}\tilde{q}, \tilde{q}\tilde{\bar{q}}, \tilde{g}\tilde{q}$

- $\sqrt{s}$=14TeV, m_{SUSY} ≈0.5-1.0 TeV → $\sigma_{pp\rightarrow SUSY}$ ≈ 1-100 pb

→ Potential for copious sparticle production at the LHC!

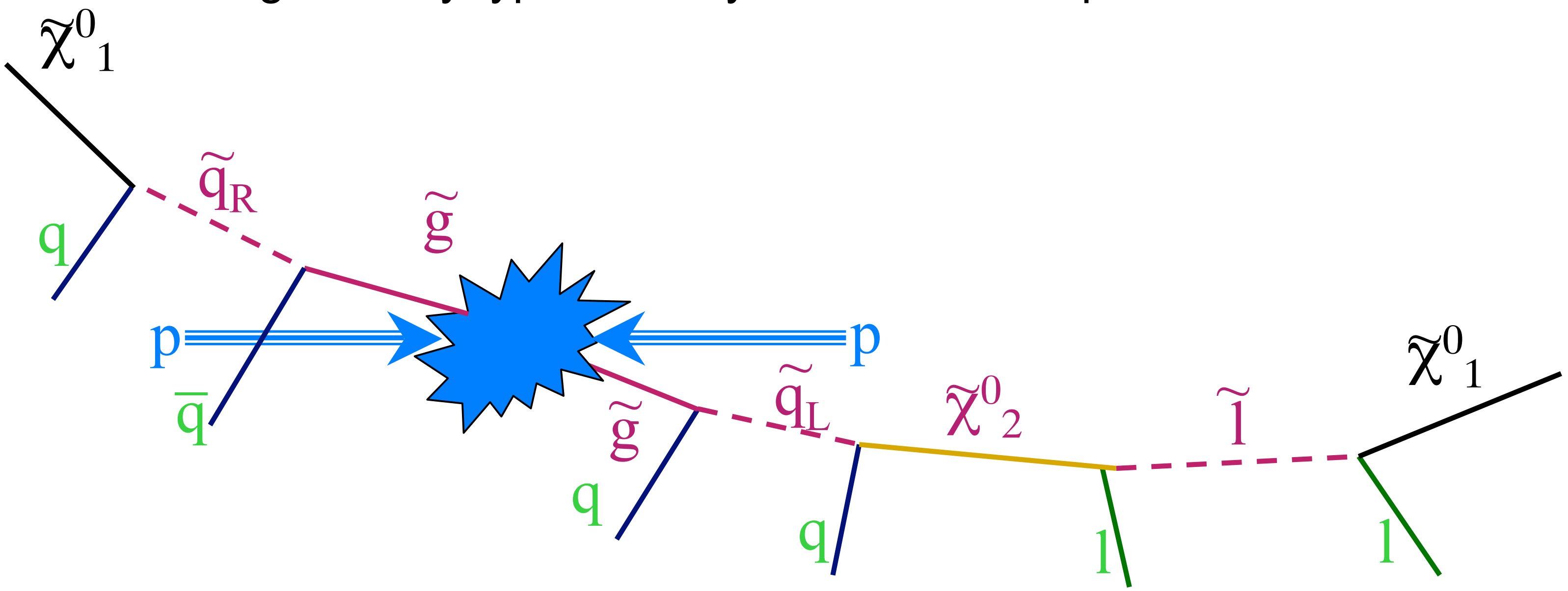




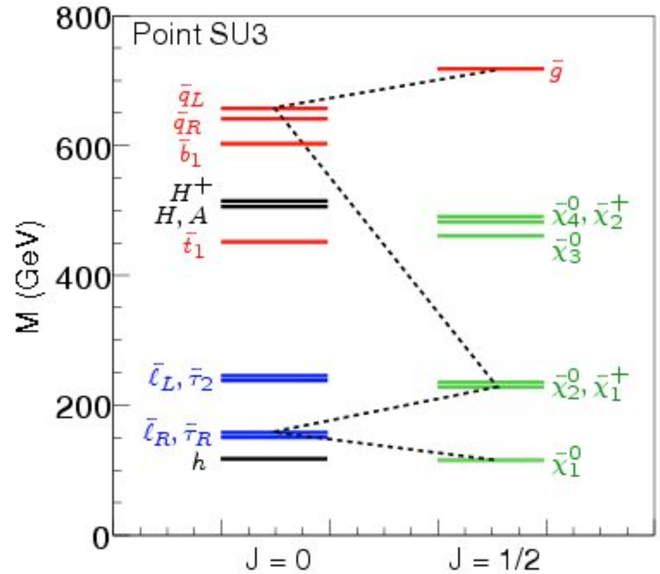
References: Beenakker, Höpker, Spira, Zerwas, 1995, 1997; Beenakker, MK, Plehn, Spira, Zerwas, 1998; Baer, Hall, Reno, 1998; Beenakker, Klasen, MK, Plehn, Spira, Zerwas, 1999; Beenakker, MK, Plehn, Spira, Zerwas, 2000; Berger, Klasen, Tait, 1999-2002; Beenakker, MK, Plehn, Spira, Zerwas, 2006

Sparticle Decays

- Event selection guided by typical decay chain of SUSY particles



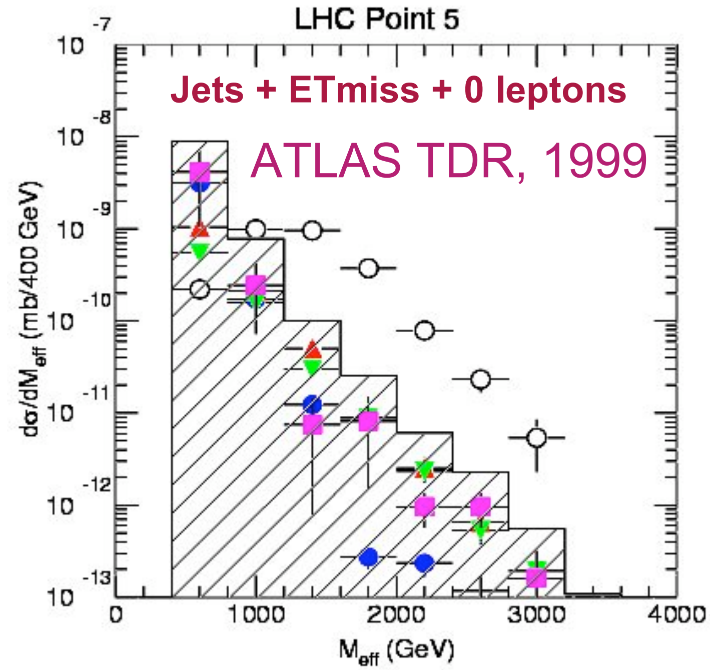
- Two strongly interacting sparticles produced
- If "*R-Parity*" conserved
  - Each sparticle decays into lighter sparticle + SM particle (→ jets + leptons)
  - cascade decays down to stable, undetected LSP (usually neutralino in mSUGRA) → large E_T^{miss}



→ Canonical SUSY signature: E_T^{miss}, high-p_T jets, often leptons

Missing E_T + jets

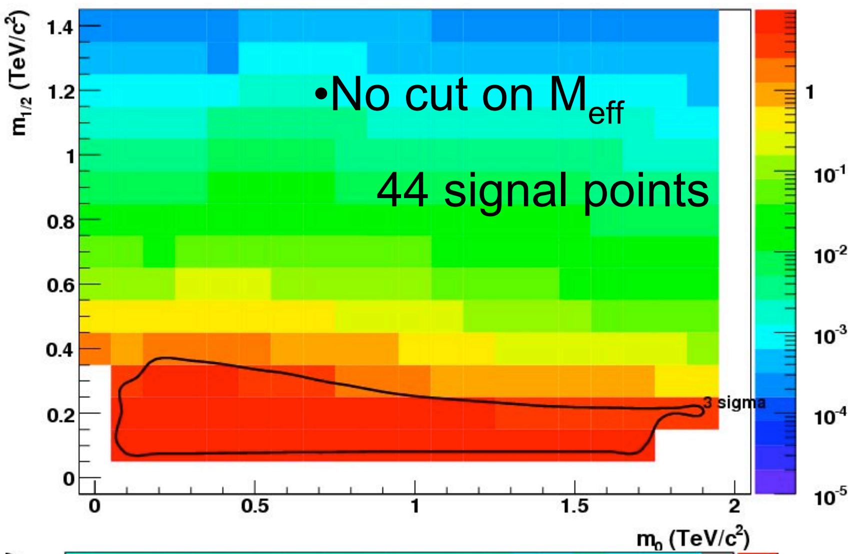
- Event selection (0-lepton case)
  - 4 high E_T jets
  - E_T^{MISS} > 100 GeV
  - Event sphericity
  - No isolated leptons
- Discriminating variable:
  - $M_{eff} = \sum_i |p_T^i| + E_T^{MISS}$
  - SUSY events typically have higher M_{eff} than SM backgrounds
    - $Z(\nu\ \nu)$ + jet
    - $W(\mu\ \nu)$+ jet
    - QCD
    - Top

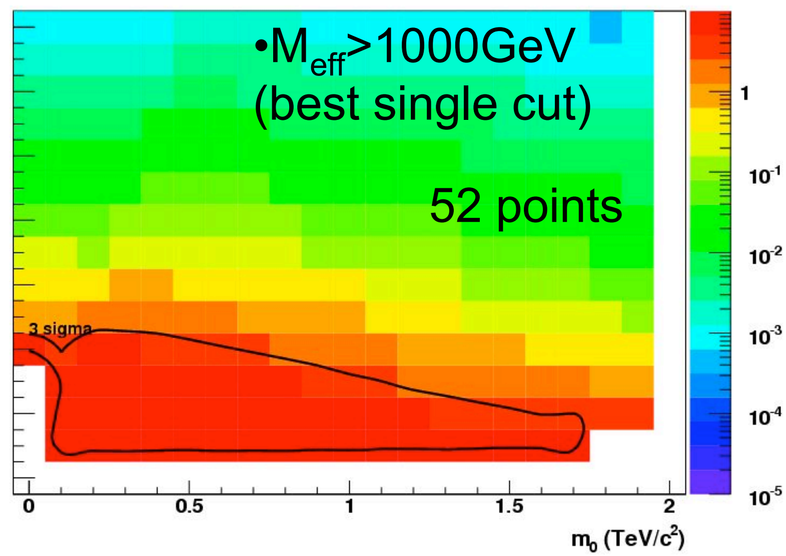


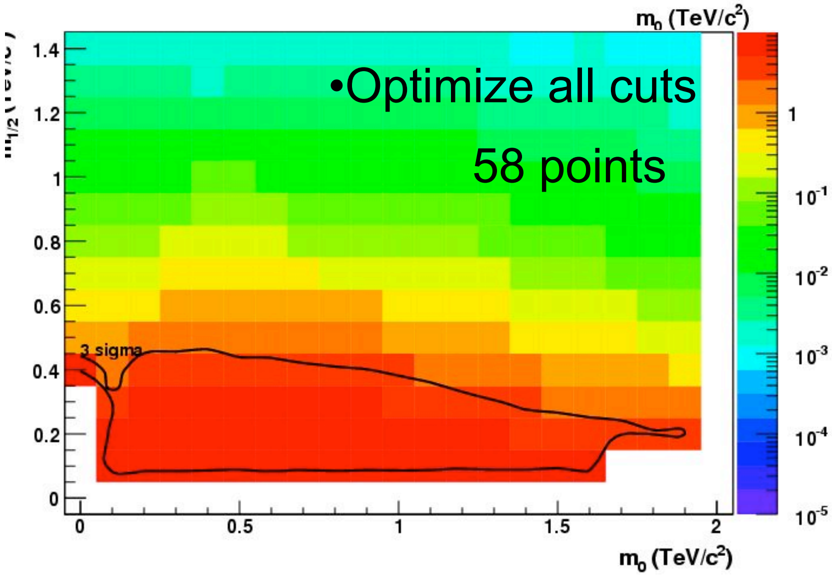
Develop strategy to optimize analysis cuts using MC

Optimization in mSUGRA plane

Significance vs m_0 and $m_{1/2}$



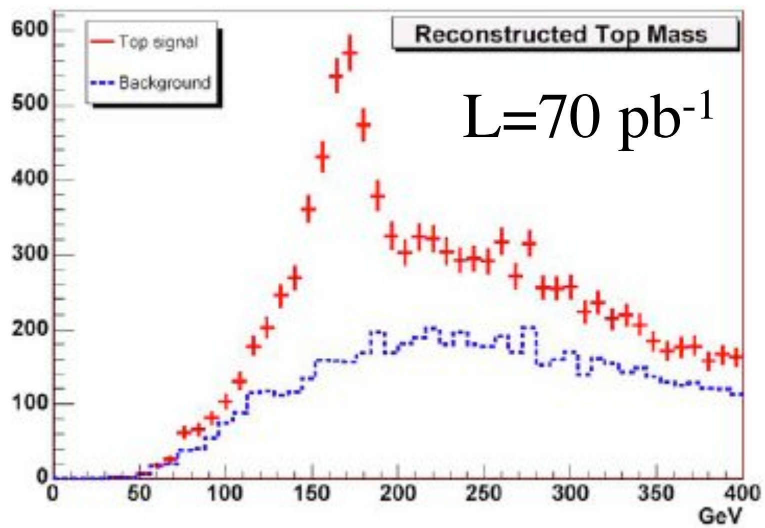




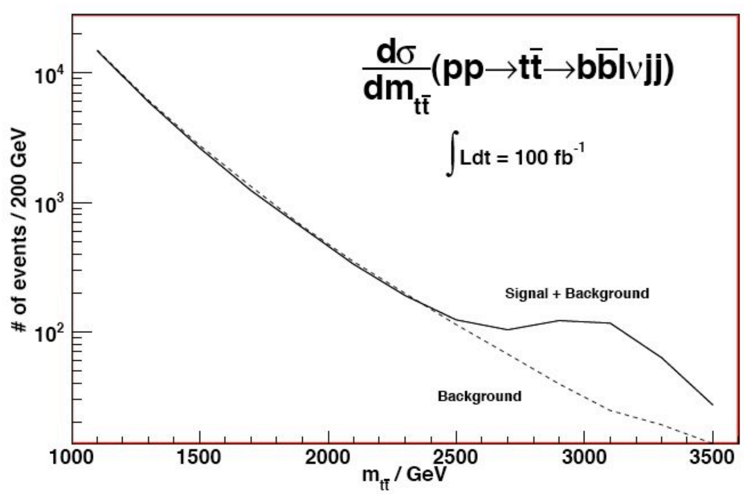
- Optimize to maximize coverage of SUSY parameter space in early data
- 3-sigma contour for 100 pb^{-1}
- Sensitive to 1 TeV mass range

Top Quark

- Top cross section 100 times larger than at Tevatron:
  - Background to it relatively smaller
  - Use streaming test MC to measure MC top cross section
- Excellent calibration sample for
  - Understanding the jet energy scale
  - Understanding b-tagging
- Interesting physics-wise also:
  - Resonance production of e.g. Z′ or KK Gluons?
  - Largest background to e.g. SUSY searches with jets and E_T^{MISS}



J. Virzi et al., hep-ph/0612015



LHC specific Workshop Series

- Focused workshop to bring experimentalists and theorists together to work on LHC specific problems, e.g.:
  - W/Z+jets modelling
  - MC at NLO
  - SUSY model generation
  - Underlying event
  - …
- Topics tbd by committee at request of community
  - Aim at 3-4 workshops per year

<u>LBNL Workshop Series on LHC physics</u>

(C. Bauer, B. Heinemann)

Purpose of workshop series

- The workshop series consists of very focused workshops, which discuss topics of direct relevance to the exploitation and interpretation of the LHC data.
- Each workshop addresses a particular and specific problem, and the participants are expected to collaborate during and after the workshop to help achieve this goal.
- A unique feature of the proposed workshop series is that the topics are at the interface of theoretical and experimental physics. The close collaboration of the participants will hopefully fuel communications between the experimental and theoretical community

Proposed Setup

- There will be 3-4 workshops per year, and each workshop is intended to consist of a small, dedicated working group, rather than have the format of a large meeting. Usually we expect about 1/3 of the participants to be from LBNL, with the rest from universities or other laboratories around the world.
- The length of each workshop will depend on the specific topic, typically around 5 days. There will be a few topical talks, but most of the time will be devoted to collaborations amongst the participants. If possible, the High energy Seminar at LBNL that week should be on a topic directly related to the workshop goal.
- The organizing committee consists of Christian Bauer and Beate Heinemann, together with two scientists (one theorist and one experimentalist) from the university community. The organizing committee will be in place originally for one year, and can rotate after that.
- Proposals for topics can be by the community, but decisions are made by the organizing committee.

Required Logistics

- <u>Office space</u>: All participants should be in offices very close to one another. Ideally we would have two or three offices, which can host all participants of the workshop.
- <u>Writeups</u> containing the results and conclusions of each workshop will be compiled on a workshop series website. They can also be submitted to the archive if the authors want.

Conclusions and Outlook

- Software efforts focused on getting ready for early data
  - Continue to play a leading role on core software design and development
  - Major contributions to pixel and tracking software
  - First in situ alignment of pixel detector with cosmic data
    - Alignment with simulated collision data under study
  - Streaming test critical part of ATLAS data preparation
- Physics efforts focused on early physics:
  - SUSY, top and Z′ production so far
  - Workshop planned for summer with entire group
    - More postdocs and students will start to focus on physics when pixel detector is installed
- Main objective is to make sure LBNL group can fully exploit LHC data
  - Collaborations with other universities on physics and tracking/alignment are starting

Backup Slides

Alignment procedure

- There are $> 10\%$ overlap between adjacent modules (front and back) with $dz = 4.2mm$.
- The extrapolating error is negligible ($< 1\mu m$ in x and $< 5\mu m$ in y).
- Assuming as a rigid body in a disk plane, 4 parameters to describe the module:
  - Shift X_0 in local X (Short pixel)
  - Shift Y_0 in local Y (Long pixel)
  - Shift Z_0 in local Z (Perpendicular to the disk)
  - Rotation α_0 along local Z axis
- Defining Residuals in local frame:
  - $\delta x = x_{odd} - (x_{even} + (dz - \delta Z_0) \cdot tan\theta \cdot cos\phi) = -(\delta X_0 - \delta\alpha_0 \cdot LocY)$
  - $\delta y = y_{odd} - (y_{even} + (dz - \delta Z_0) \cdot tan\theta \cdot sin\phi) = -\delta Y_0$
  - both for left overlap (-LocX) and right overlap (+LocX)
- With relative alignment between adjacent modules, and internal alignment can be derived with all working modules, and 6 parameters remaining for disk global position.

W. Yao

Data Volume

- Expect huge datasets:
  - 200 Hz written to tape
  - Event size
    - 1 MB (ESD): Tier-1
    - 100 kB (AOD): Tier-2, desktops, laptops
  - 20 million events per day:
    - 20TB ESD's / day
      - compressed data format needed for detailed studies of reconstruction algorithms
      - needed for early physics most likely
    - 2TB AOD's / day
      - main format for physics analyses
- Challenge:
  - Get good data reliably and quickly to users