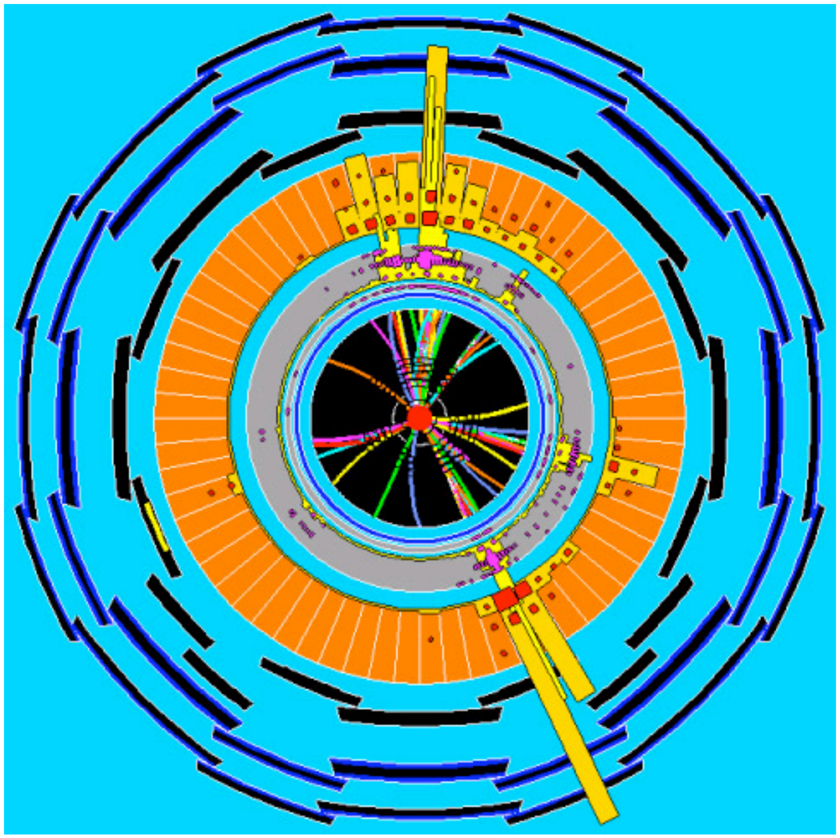
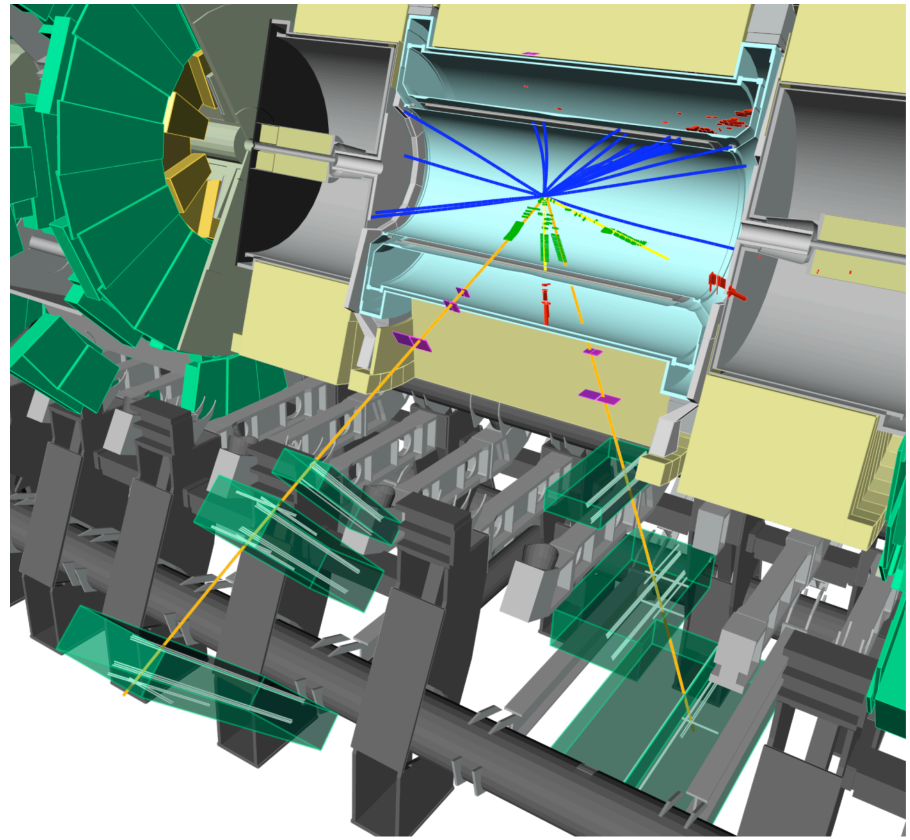
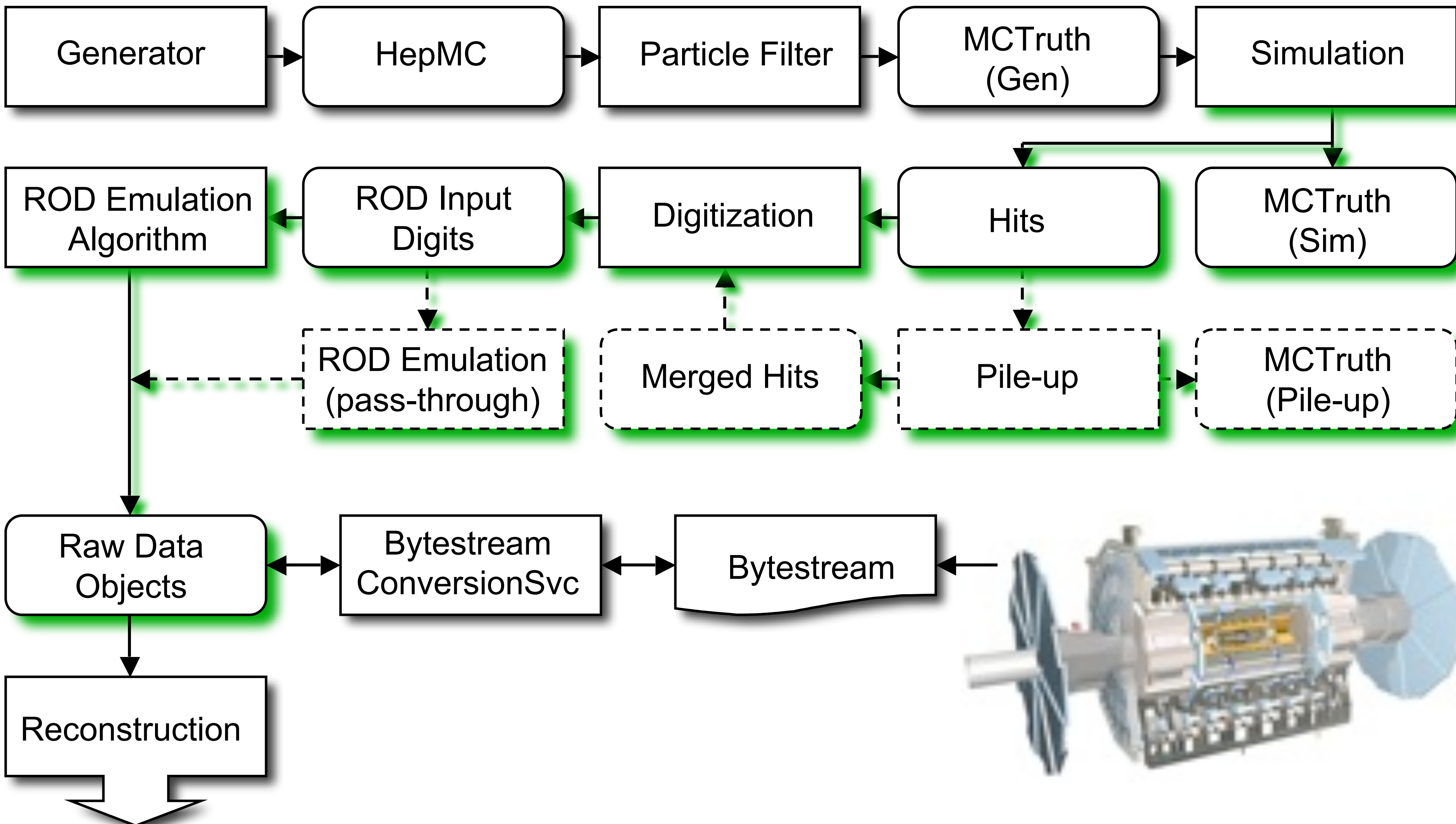
ATLAS Simulation Validation and Computing Performance Studies





Zach Marshall (Caltech/Columbia)
For the ATLAS Core Sim Group
CHEP09, Prague
23 March 2009

Simulation Overview



Monte Carlo Overview

- ➔ Each Monte Carlo job is divided into three parts
- ➔ Generation: Pythia, Herwig, etc produce a single pp scatter event, with rapid hadronization and decay
  - – Won’t be covered in this talk
- ➔ Simulation: all particles from the generater are run through a full detector simulation
  - – “Hits,” energy depositions (with position and time) are kept
- ➔ Digitization: all hits from the simulation are run through electronics simulation
  - – Digitization output can be translated to look identical to what comes off the detector
  - – At this stage, we overlay pile-up: additional minimum bias, beam gas, beam halo, and cavern background events
- ➔ Reconstruction and trigger run in the same way for both Monte Carlo and real data

ATLAS Simulation Basics

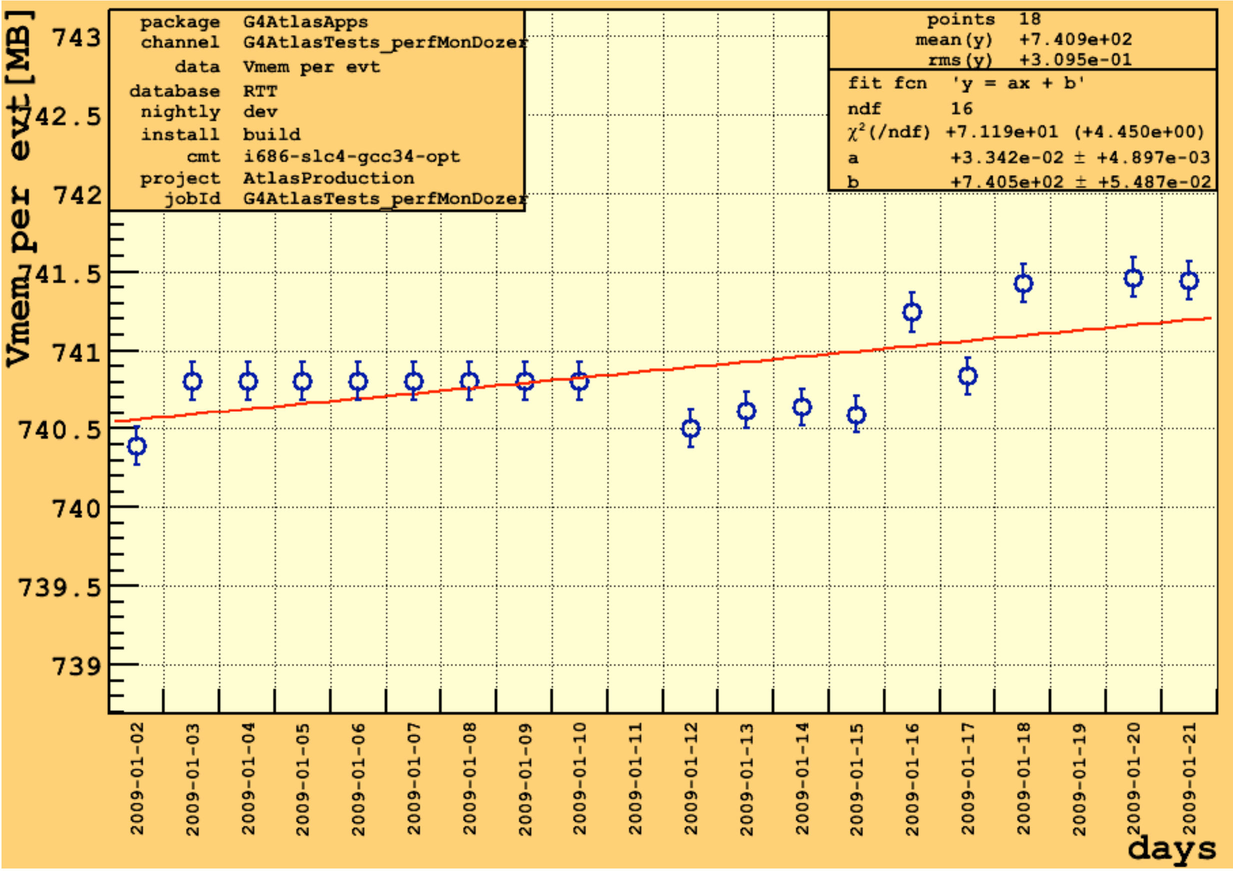
- Simulation based on the Geant4 toolkit
  - Recently transitioned from 8.3.patch02 to 9.1.patch03
  - Now testing 9.2 (candidate for first data)
  - Using CLHEP 1.9.4.2
- Default configuration is 32-bit libraries built with gcc3.4.6 on SLC4
  - Also supporting SLC5, gcc4.3, 64-bit libraries, several LCG versions, attempting a port for Mac OS X…
- Production is run on the Large Computing Grid
  - >500 M events run last year
- Fresh releases built every ~month
  - Patches for production every ~week, or as needed

Automatic Testing

- ➔ No human can keep up with all those builds
- ➔ Nightly, we run three types of tests
- ➔ ATLAS Nightly (Build) Tests (~10)
  - Run for *every* build
  - Very basic (does it run), very short (<10 min)
- ➔ Run Time Tests (~50)
  - Run for select (~10) builds
  - Longer tests (hours), more complete
- ➔ Full Chain Tests (~5)
  - Run for builds prior to release (~5)
  - Full day for a single test
  - Output of each day used as input for the next day
- ➔ ALL of these required to pass prior to release…
- ➔ More on testing infrastructure in other talks

Nightly Testing

- ➔ Automatic testing also provides comparisons from day to day, including notification on failures
  - – Invaluable for those with bad memories (like me)



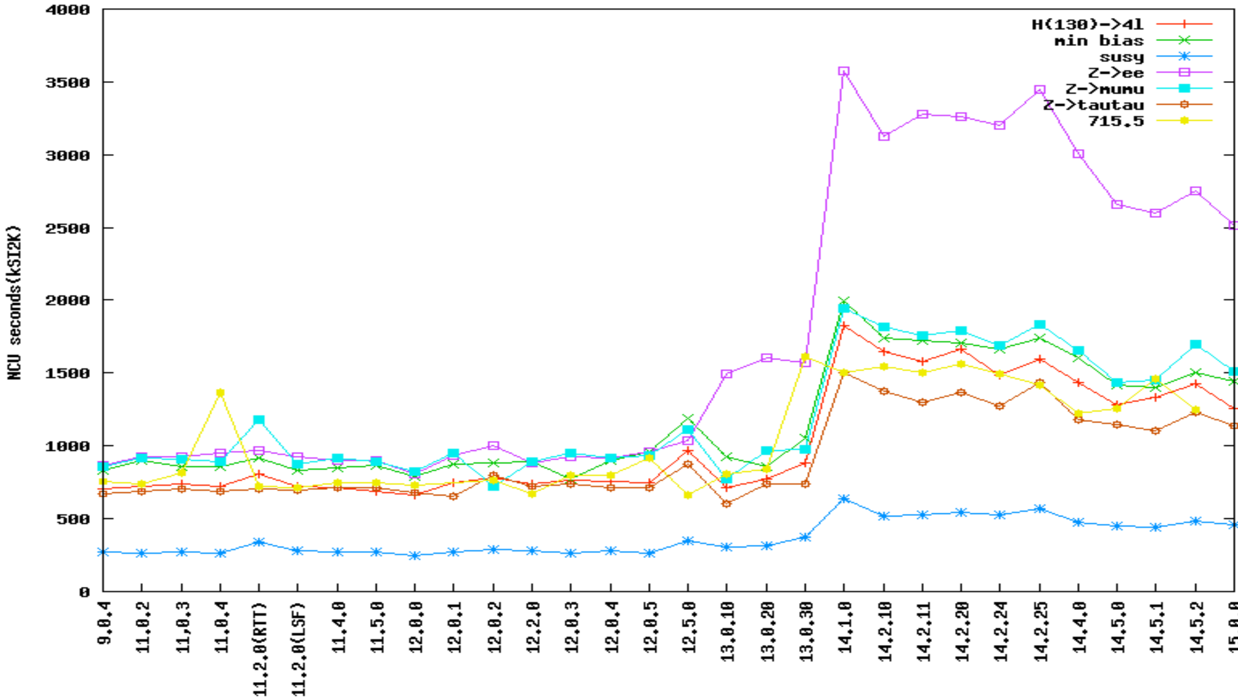
20 day history of memory use for simulation

1 MB change well within the resolution from day to day

Simulation Benchmarking

- Run a "representative" sample of events, to:
  - Check for problems throughout the detector
  - Check performance for most physics groups
  - Provide some rapid handle on any problems
- Single particles and full events
  - $e^{+/-}$ 5, 50, 100 GeV of p_T
  - $\mu^{+/-}$ 5, 50, 200 GeV of p_T
  - $\pi^{+/-}$ 5, 50, 200 GeV of p_T
  - Z> e^+e^-, $\mu^+\mu^-$, $\tau^+\tau^-$
  - Minimum bias
  - H(130)>ZZ>4l
  - SU3 SUSY
  - Jets with leading parton p_T between 35 and 70 GeV
- 3000 events, 300 hours of computing time

Simulation CPU Performance



Times benchmarked by release. Recent jump from changing physics descriptions.

Sample	Generation	Simulation	Digitization
Minimum Bias	0.0267	551.	19.6
$t\bar{t}$ Production	0.226	1990	29.1
Jets	0.0457	2640	29.2
Photon and jets	0.0431	2850	25.3
$W^{\pm} \rightarrow e^{\pm}\nu_e$	0.0788	1150	23.5
$W^{\pm} \rightarrow \mu^{\pm}\nu_{\mu}$	0.0768	1030	23.1
Heavy ion	2.08	56,000	267

Times per event in kSI2K sec (divide by 3 for a modern CPU)

Full events take ~10 minutes each

Simulation Optimizations

- ➔ New platform (slc5/gcc4.3) gives 20% improvement in CPU performance
  - Expected to become the default platform before first data
- ➔ Migration to Geant4 9.2 gives 10% improvement in CPU performance
  - First attempt scheduled for this month
- ➔ Some code profiling is underway
  - Removing some hot spots (string comparisons)
  - Easing others (MANY B-Field value queries)
- ➔ Generally, philosophy is to have as accurate as possible a full simulation
  - Means sacrifice of some computing resources
  - But fast simulations can be used for higher statistics
  - Makes simulation-based detector studies more “realistic”

Fast Simulations

- Several flavors of fast simulation exist for ATLAS
  - Shower libraries for low energy EM particles (“Fast G4”)
  - Parameterization of the calorimeter (“ATLFAST-II”)
  - Tracking using a simplified geometry (“ATLFAST-IIF”)
  - Parameterization of final physics objects (“ATLFAST-I”)
- Different use-cases require different granularity
  - Final simulation strategy makes best use of all flavors of fast simulation within the limits of grid resources

Sample	Full Sim	Fast G4 Sim	ATLFAST-II	ATLFAST-IIF	ATLFAST-I
Minimum Bias	551.	246.	31.2	2.13	0.029
$t\bar{t}$	1990	757.	101.	7.41	0.097
Jets	2640	832.	93.6	7.68	0.084
Photon and jets	2850	639.	71.4	5.67	0.063
$W^{\pm} \to e^{\pm}\nu_e$	1150	447.	57.0	4.09	0.050
$W^{\pm} \to \mu^{\pm}\nu_\mu$	1030	438.	55.1	4.13	0.047
Heavy ion	56,000	21,700	3050	203	5.56

Memory Performance

- Memory is not really an issue for the simulation
  - We use ~850 MB of memory (grid nodes allow 2.4 GB)
  - Memory consumption is quite stable: leaks below 0.25 MB/ev (remainder may be a "feature")
  - Very small dependence on geometry, physics descriptions, channel
- Digitization is more of a concern, particularly with high luminosity pileup
  - Able to run 10^{34} locally without problems, but still issues when running on the grid

Resource	**No Pile-Up**	10^{33} cm^{-2} s^{-1}	3.5×10^{33} cm^{-2} s^{-1}	10^{34} cm^{-2} s^{-1}
Memory Leak [kB/event]	10	270	800	2100
Virtual Memory [MB]	770	1000	1300	3600
Malloc [MB/event]	12	21	40	985

Output File Sizes

- ➔ This is the major limiter for grid production
  - We stopped saving most digi output files to save space
  - Tracker is by far the largest consumer
  - Calorimetry does collection on-the-fly

Collection Name	Size [kb/event]	Percentage of File
Silicon pixel tracker	82	4%
Silicon strip tracker	356	16%
Transition radiation tracker	921	46%
Electromagnetic Barrel Calorimeter	89	4%
Electromagnetic Endcap Calorimeter	104	5%
Hadronic Barrel Calorimeter	29	1%
Hadronic Endcap Calorimeter	22	1%
Forward Calorimeter	42	2%
Calorimeter calibration hits	243	12%
Muon system (all collections)	3	<1%
Truth (all collections)	134	7%
Total	**1987**	**100%**

→ Transition radiation tracker (46%): Recently reduced by ~70% (but not validated yet)

→ Calorimeter calibration hits (12%): Optional

Average for 50 tt events (Varies by channel)

Geometry and Validation

- ➔ Aim for a very realistic a detector description
  - Rechecking this year with engineering drawings
  - Approx. some dead material (wires) as distributed lumps
- ➔ Work toward real detector conditions
  - As installed positions for all detectors
  - Not just the right number of dead channels, but the right dead channels
  - Infrastructure available to recreate conditions for a specific data taking run
- ➔ Good practice from cosmic data taking in 2008!
  - One dead power supply in the calorimetry
  - Several disabled cooling loops in the tracker
- ➔ Validate by weighing the real and simulated detectors
  - Aiming to be within a few tons (<1%)…

Robustness

- ➔ Grid production is the ultimate test
  - About 250M events produced since August with full simulation, about 300M events with fast simulations
  - Available CPU allows order 1M Geant4 events per day
    - Of course more with fast simulation…
    - Still, the limit on production is *disk space* available!!
- ➔ No crashes in simulation reported since October!
  - Includes the transition to G4 9.1 - we have not had a bug reported in production with 9.1 yet!
- ➔ No crashes in low pile-up digitization either
  - Still pushing luminosity limit, and memory consumption is an issue

Physics Validation

- ➔ Once our software runs, we have to check the *physics content* of the output!
- ➔ Dedicated group compares release-to-release
  - Representatives from all detectors and physics groups
  - About 1M events in different channels are used
  - This group provides tests of *computing* motivated simulation modifications (e.g. changing cuts)
- ➔ Other groups compare to test beam
  - And to data, once we have it
  - Limited (interest in) comparisons to cosmics thus far
  - These groups provide most of our physics motivation for simulation modifications (e.g. changing physics models)

Physics Validation (II)

- Once we have good data available, quickly move to testing and tuning of the simulation
- Long check list of things to test
  - Geometry description (esp. thickness of the tracker)
  - Shower shapes for EM and hadronic calorimeters
  - Cavern background and noise (can be taken from data and overlaid on simulated signal events)
- Now we can work to understand what knobs are available to modify detector response
  - And how moving those knobs affects physical observables
  - Same story for fast simulations - some can be tuned directly from data, others need a tuned full simulation before they are able to “tune to data”

Summary and Conclusions

- ATLAS has developed a robust simulation based on the Geant4 toolkit
  - Over 500M events were produced in the last year
  - No failures have been observed in the last six months
- Benchmarking and validation is done at several stages before and after each release
  - Nightly tests ensure software functionality
  - Large scale tests expose rare problems with new options and parameters
  - CPU, memory, and disk space consumption are constantly monitored
- The ATLAS simulation software is ready to face the challenges of data!