ATLAS Simulation using Real Data: Embedding and Overlay

Andy Haas (NYU)
on behalf of the ATLAS collaboration

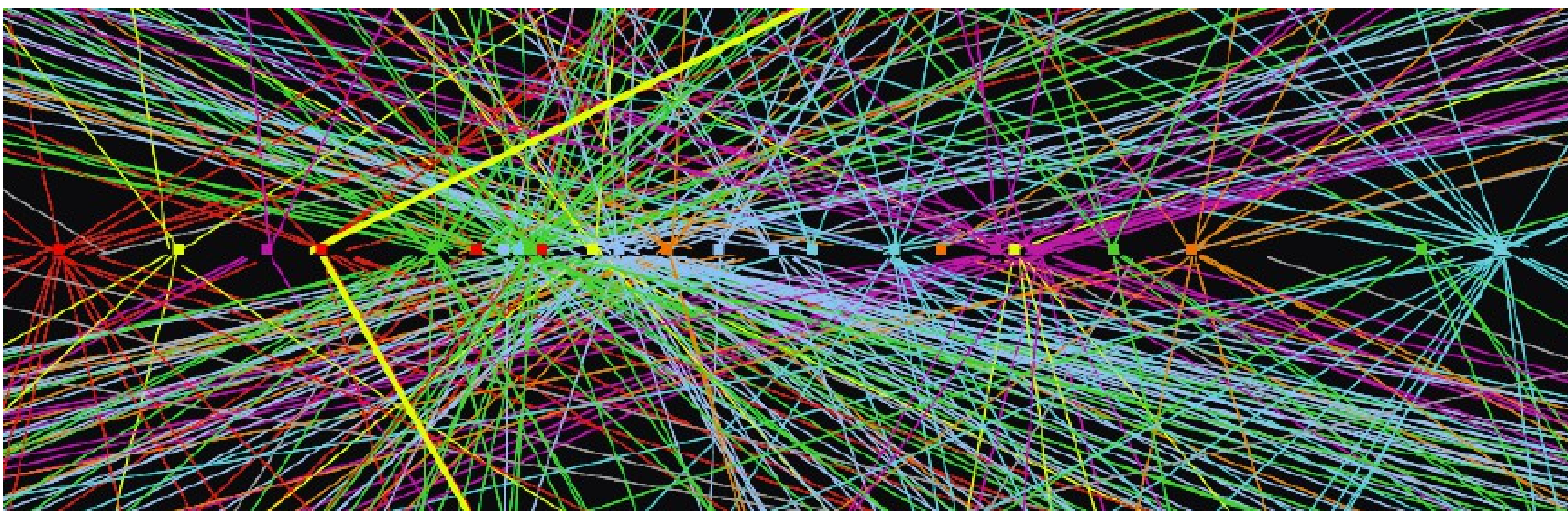
October 11, 2016

ATLAS

CHEP2016

22nd International Conference on Computing in High Energy and Nuclear Physics, Hosted by SLAC and LBNL, Fall 2016

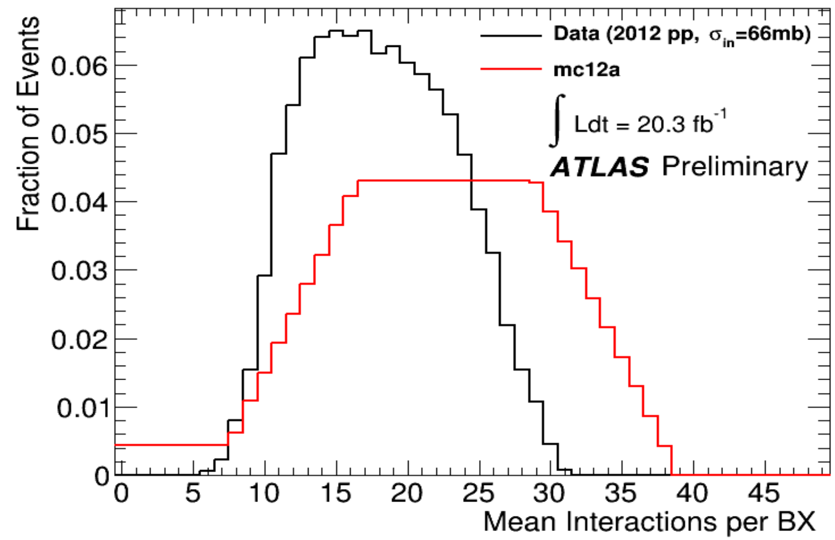
NEW YORK UNIVERSITY



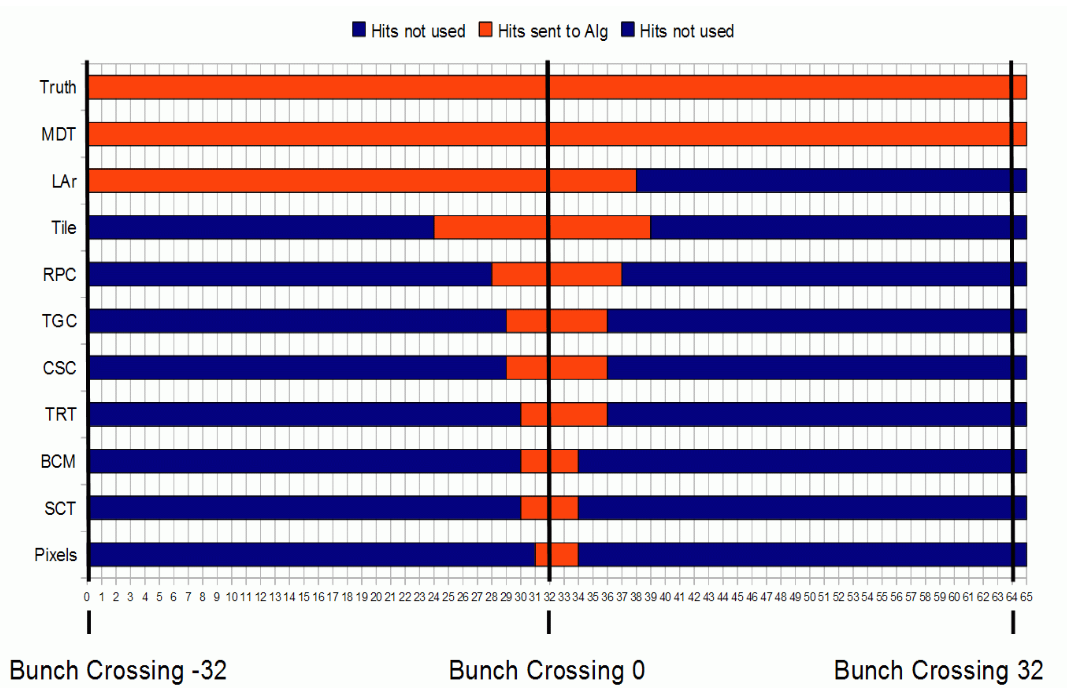
https://indico.cern.ch/event/505613/contributions/2230824/

Pileup and Detector Noise

In each LHC filled bunch crossing, there's soft pp collisions called “pileup” Run1: ~20, Run2: ~40, HL-LHC: ~400



ATLAS sub-detectors are also sensitive to many bunch crossings before and after the triggered event



“Cavern background” (CB) is a gas of neutrons/gammas from collision byproducts that lives for many microseconds and creates muon chamber hits

Each sub-detector also has noise (thermal, electronic, etc.)

Simulated Pileup/Noise vs. Data Overlay

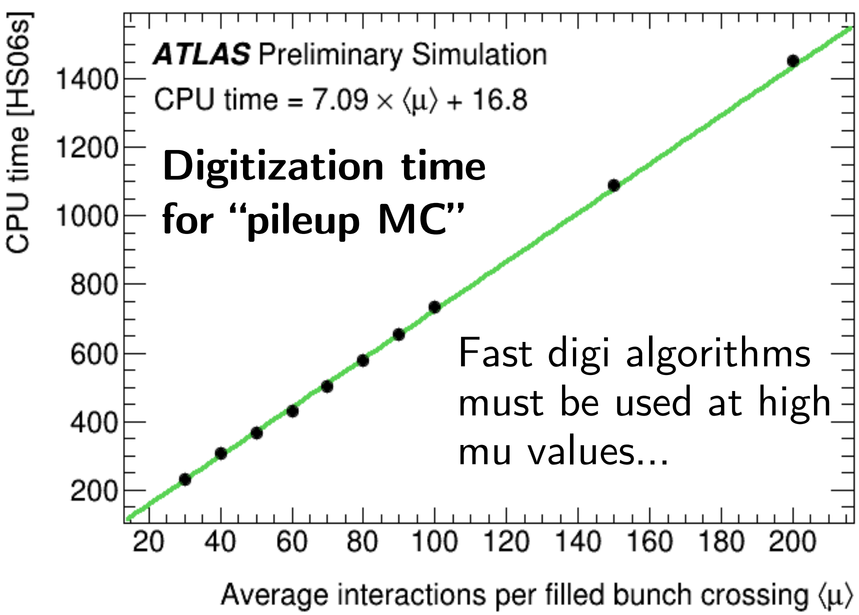
- **Overlay simulation is an alternate way of modeling pileup and detector noise during digitization**

Option 1:
"Pileup MC" (current default)

Simulate hard interaction in MC, mix with simulated pileup MC during digi, and add emulated detector noise

Option 2:
"Overlay MC"

Simulate hard interaction in MC, overlay "random" RAW data event during digi to include pileup, CB, noise



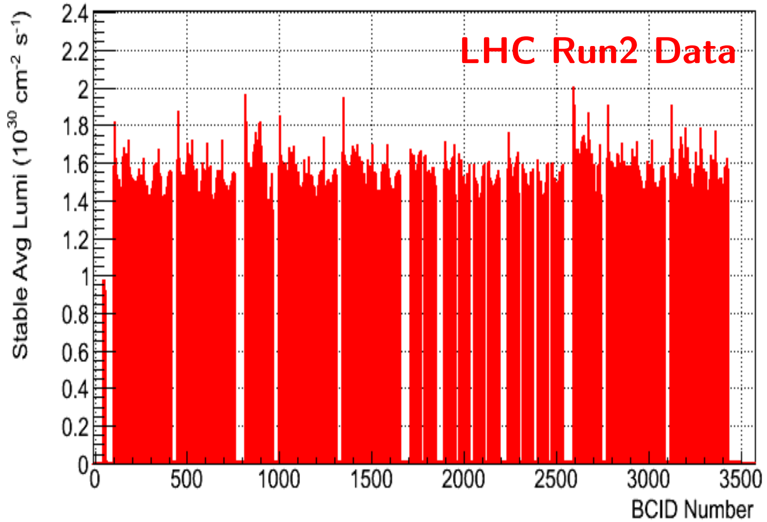
Overlay method used successfully in the past at BaBar and D0

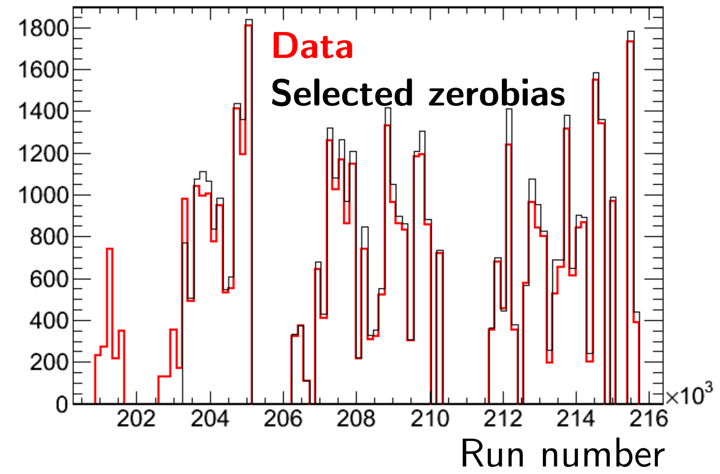
Inherently fast, since the pileup background is data and thus already digitized

- About 50 HS06s per event, with only slight mu growth

Background Data Preparation

- "Zerobias" data: trigger fires 1 LHC turn after high-pt L1 EM trigger
  - Next event in the same BC position
  - Proportional to luminosity in each BC
- Prescaled to keep ~10 Hz in Run2
  - ~3 Hz is selected at HLT to have a jet with pT>40 GeV *
  - ~100 M events/year
- No zero-suppression (except in silicon)
  - ~2 MB/event compressed
- Offline, zerobias data are sampled from lumi-blocks in the desired time-period to reproduce the luminosity profile of a high-pt trigger (account for dead-time, prescales, mix of HLT jets, etc.)

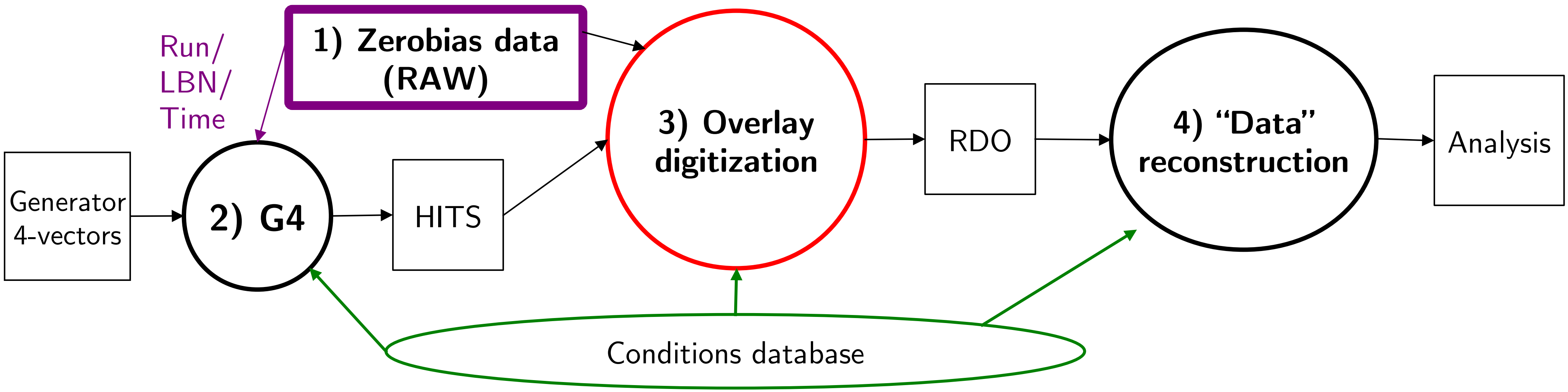




*Allows us to reuse background events with "interesting objects" about 10x less frequently than "soft" events

Overlay Steps

1. Start with input zerobias lumi-weighted RAW events, ordered in time
2. Simulate a hard-scatter G4 event, with conditions matching each selected data event (beamspot/tilt, alignments*, magnetic fields, etc.)
3. Overlay each zerobias data event with matching G4 event at the detector channel level, then digitize combined signals
4. Reconstruct the combined event as data (with some MC conditions... where simulation can not model data accurately, e.g. drift radius vs time for straw tracker / muon tubes. We thus use MC constants for digits that have some energy contribution from simulated hits, and data constants elsewhere for background digits – not ideal for digits with energy from background and simulation.)



*Alignments are updated at the initialization of each G4 job of 100 events.
Using real-data detector alignments does lead to overlapping G4 volumes in some cases.
The geometry description (of dead material) has been updated to eliminate all major ones...

Validation and Testing Methods

Various outputs of the overlay simulation are studied, to isolate modeling of the background, hard-scatter signal, and combinations

1) Compare directly reconstructed zerobias input RAW dataset to overlay MC with just a single neutrino, using data conditions and alignments

- Isolates just the pileup/noise background effects, with no signal energy

2) Compare standard to overlay simulation, using MC alignments but data conditions, with no background hits overlaid and no pileup MC collisions added

- Isolates just the hard-scatter MC effects, with no background energy

3) Compare overlay of energy at the digitization step to direct addition of energy/time G4 HITS, with MC alignments and MC conditions:

a) overlay simulation: *signal HITS + (MC HITS → digitization) → overlay*

b) standard simulation: *(signal HITS + MC HITS) → digitization*

- Isolates the overlay digitization algorithms, for channels with background and hard-scatter energy, to see e.g. effects of tracker zero-suppression

4) Compare overlay MC to pileup MC, for various processes

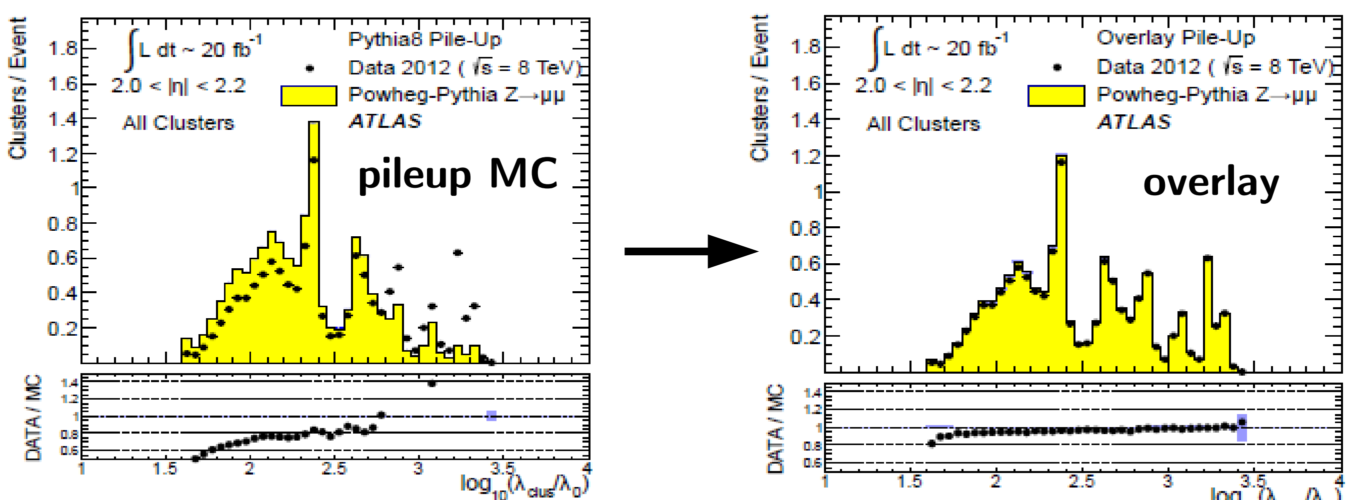
5) Compare overlay MC to data (and pileup MC), with a clean selection, e.g. Z→mumu

Drawbacks compared to "pileup MC"

- Less accurate when combining overlapping background and signal on the same channel for some subdetectors (e.g. silicon), since zerobias data is zero-suppressed in silicon
  - In overlay, a background hit below threshold can not add to a G4 HIT below threshold, to make a hit above threshold
- In some cases the background data does not record enough samples to reconstruct the background pulse precisely
  - For instance, the L1calo trigger uses 7 samples online, but only 5 stored
- MC conditions are used in a few places, even when some energy is coming from background data (overlapping straw-tracker hits, e/gamma shower shapes, scale-factors for JES, etc.)
  - Could use truth to tell when the object is mostly background?
- Potential Geant4 geometry overlaps when using data alignments
- Harder to simulate future high luminosity (multiple overlay possible in most sub-detectors, but challenging in the calorimeters)
- Have to wait until all data is collected before generating simulations
- More challenging to produce – zerobias data preparation, large database access needs during G4/overlay and reconstruction!
- Don't have the background truth information – it's data :)

Advantages compared to "pileup MC"

- Real pileup events from data – no generator tuning needed here
- N_vertex, z_vertex, and inst luminosity match data precisely
  - No event weighting needed, so higher "efficiency" of MC use
- Faster (and less memory) at high luminosity than standard pileup digi
- Realistic mix of inst luminosity variations and bunch structure, in-time/out-of-time pileup contributions (including satellite collisions)
- True detector noise and occupancy, including lumi-weighted variations and correlations between channels (like noise bursts)
- Conditions (beamspot, dead channels, etc.) are from data, including lumi-weighted correlations – e.g. realistic fraction of events where Tile section 19 is dead and pixel module 177 is noisy and beamspot is near (0.61,0.52)...

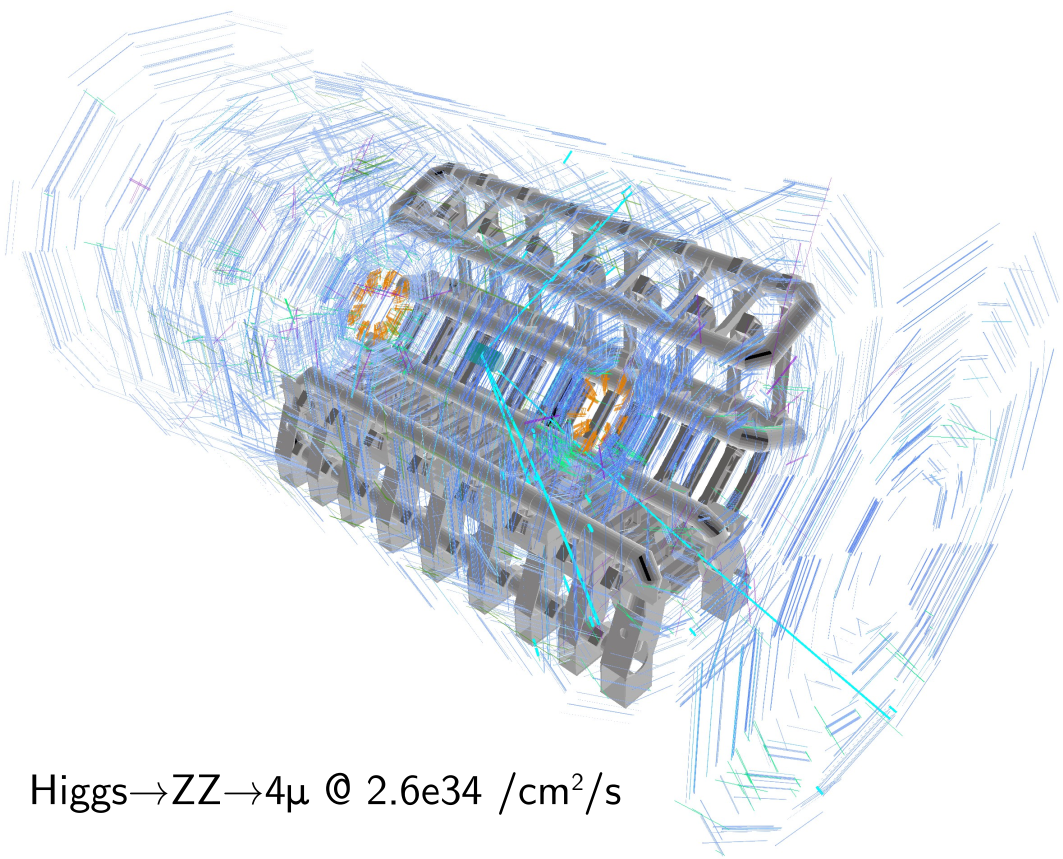


Number of calo clusters vs. depth in calorimeter of cluster center... sensitive to calo noise and pileup distributions due to seeding

arXiv:1603.02934

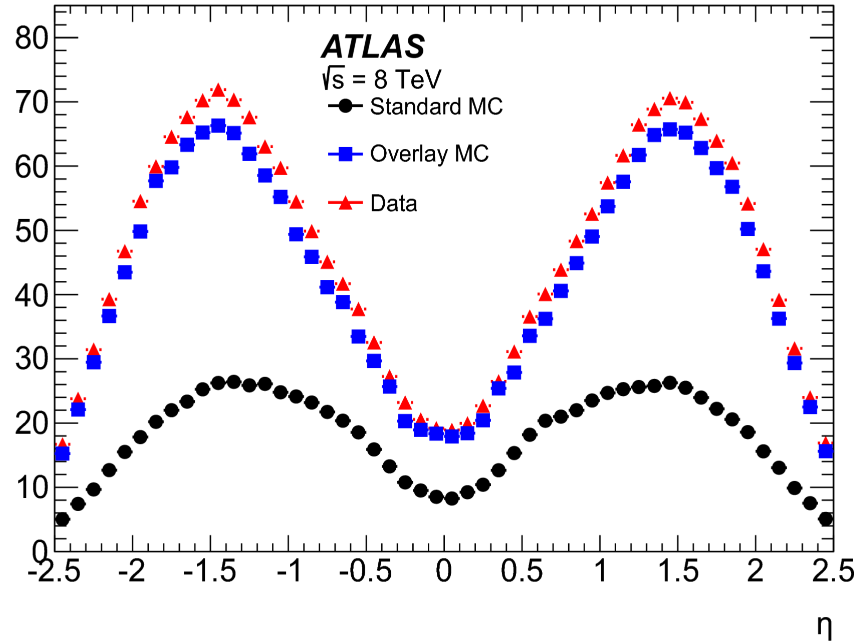
Cavern Background Effects

- Difficult to simulate accurately (low-energy neutrons, shielding, etc.)
  - Not included in standard ATLAS pileup MC
  - Automatically included in data overlay MC
- Critical to model well for some physics analyses, e.g. searches for exotic long-lived vertices in the muon system…

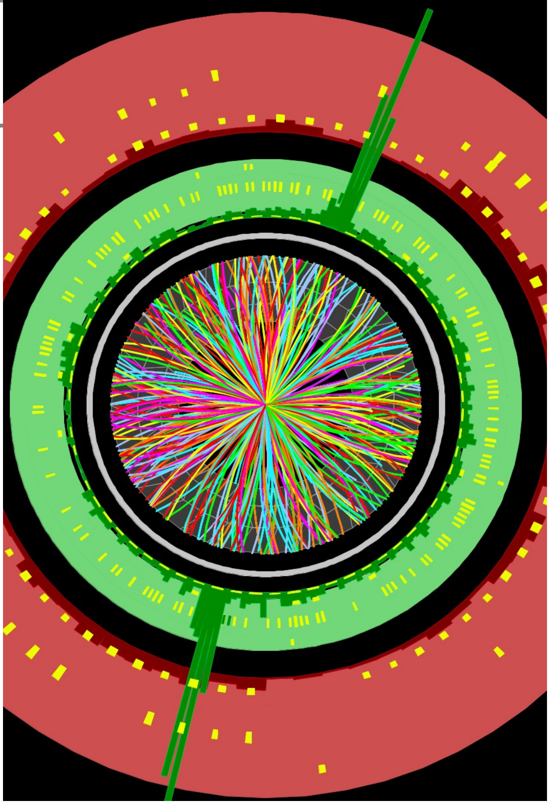


Higgs→ZZ→4μ @ 2.6e34 /cm^2/s

Phys. Rev. D 92, 012010 (2015)
https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/EXOT-2013-12/

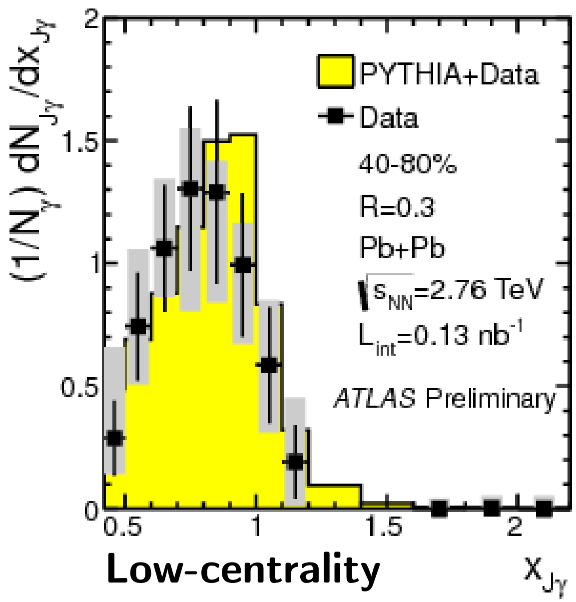


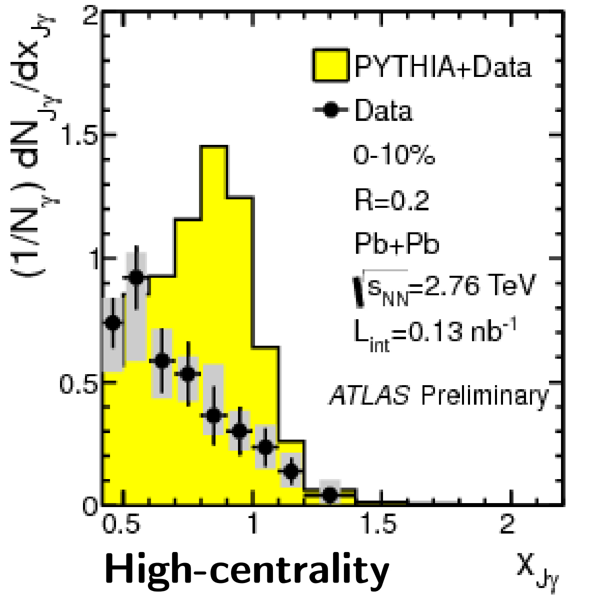
Heavy Ion Collisions



- Very challenging to simulate the "soft" parts of nuclear collisions
  - Generators such as HIJING have limitations
  - Simulating the ~10k particles in G4 is slow
- ATLAS Heavy-Ion group has made extensive use of overlay simulation
  - Record minimum-bias HI events (want a collision, and for HI runs there is <1 collision per BC)
  - Overlay minimum-bias HI event on simulated hard-scatter (pp!) collision at same vertex position
  - Use matching alignments and conditions, just like for pp overlay simulation
- Data shows jet quenching, which is not modeled in Pythia+Data overlay

ATLAS-CONF-2012-121





Embedding

- Embedding takes a data event (e.g. Z$\rightarrow$mumu) and replaces objects with another type of object (e.g. taus) to emulate a related process
- Critical for modeling Z$\rightarrow$tautau background in H$\rightarrow$tautau analysis
- So far this has been done at the reconstruction level at ATLAS, with tracks and calorimeter cells, but this has inaccuracies...
  - tau's are simulated and reconstructed without pileup
  - no cells below zero-suppression threshold in data event
  - can't run L1 trigger simulation

- Working on using overlay MC techniques to perform embedding
  - Recording ~5 Hz of non-zero-suppressed Z$\rightarrow$mumu this year, to use as overlay input (instead of zerobias)
  - Then simulate taus in G4 at same vertex/momenta as recoed muons
  - Overlay the MC taus and Z$\rightarrow$mumu data event, removing digits that were used to form the reconstructed Z$\rightarrow$mumu tracks' hits
  - Reconstruct overlaid event, performing subtraction of the muons' calorimeter cell energies (as in original embedding technique)

Summary

- Pileup, detector noise, and cavern background are difficult to model accurately at ATLAS using simulated pileup and emulated noise
- Overlay simulation is a way of automatically including these backgrounds, using specially recorded data events as input
  - Has some drawbacks and limitations, but also many advantages in terms of speed and accuracy

- ATLAS has used overlay for analyses where cavern background is important, and for many Heavy Ion analyses
  - Also found use recently for Fast-TracK trigger operations, where pre-calculated banks of *data* track patterns must be calculated (from simulation)
- Working to extend the overlay technique to do embedding at ATLAS

- Expect to use overlay simulation at ATLAS more regularly in the future, as the technique matures and production issues (such as conditions access) are improved
  - Pileup will only become more important to model accurately in the future, as instantaneous luminosity increases