Institut
”Jožef Stefan”
Ljubljana, Slovenija

ATLAS
EXPERIMENT

ATLAS PILE-UP AND OVERLAY SIMULATION

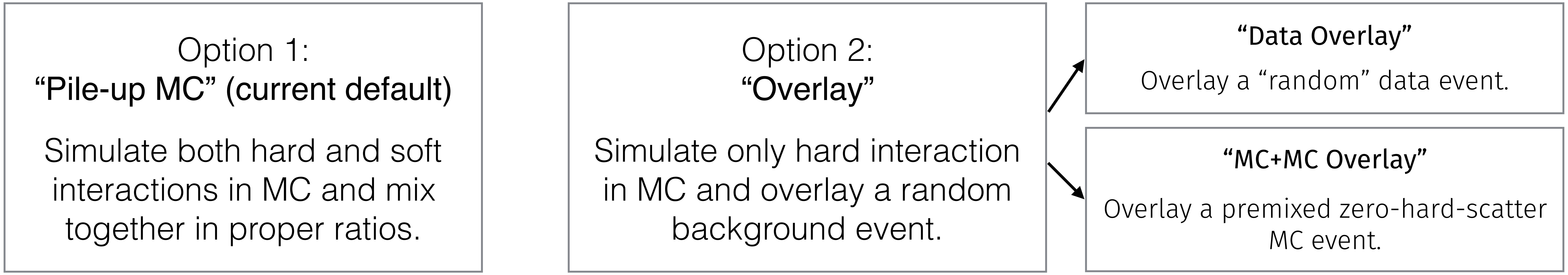
LPCC Detector Simulation Workshop,
June 26-27, 2017

Tadej Novak on behalf of the ATLAS Collaboration

INTRODUCTION

In addition to hard interaction:

- pile-up from other collisions in current and surrounding bunch crossings,
- cosmics, beam-gas, beam-halo, cavern background, detector noise, …



Option 1:
“Pile-up MC” (current default)

Simulate both hard and soft interactions in MC and mix together in proper ratios.

Option 2:
“Overlay”

Simulate only hard interaction in MC and overlay a random background event.

“Data Overlay”

Overlay a “random” data event.

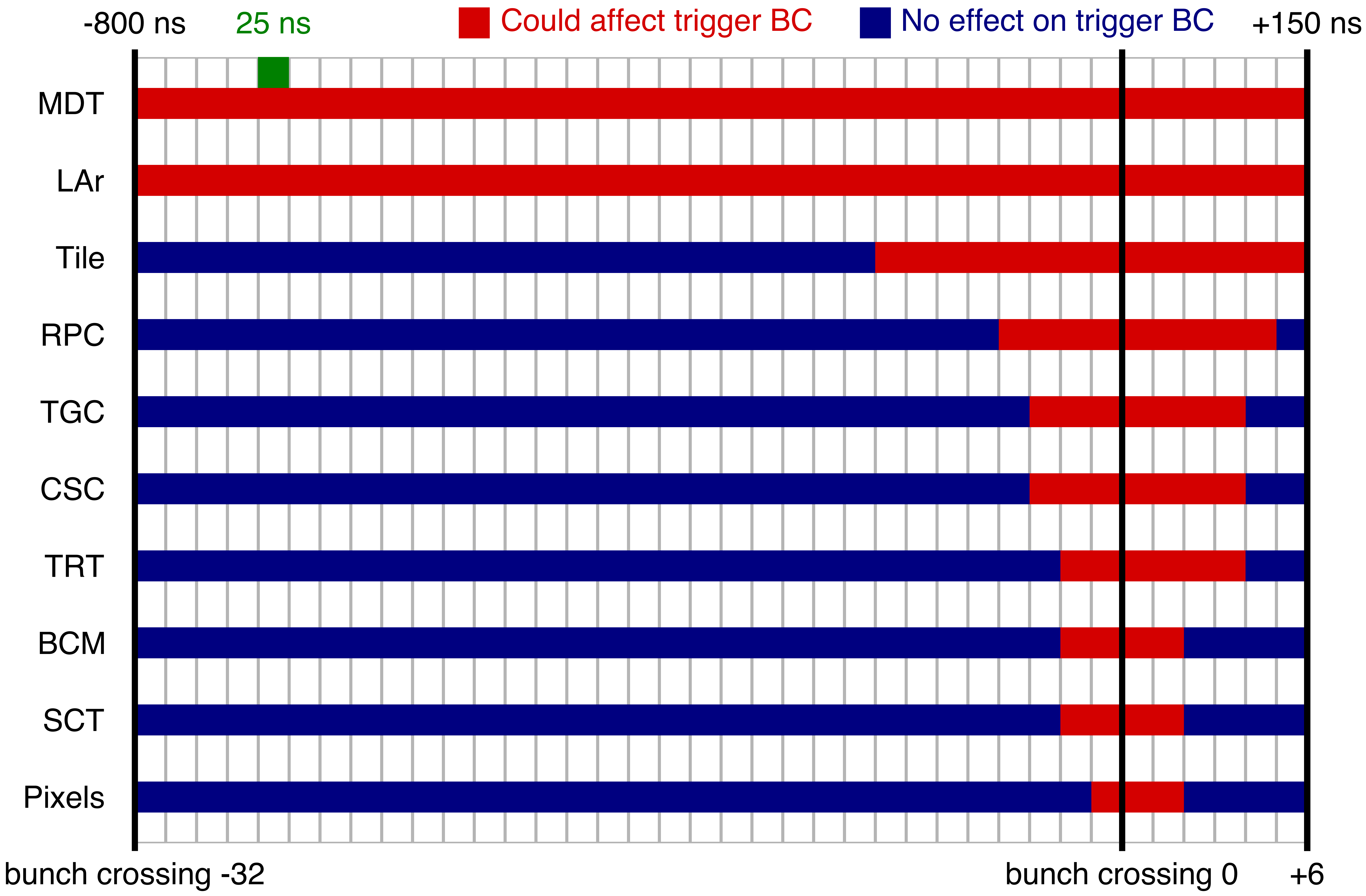
“MC+MC Overlay”

Overlay a premixed zero-hard-scatter MC event.

Pile-up MC method mostly used so far at ATLAS

- Data Overlay being used for some studies, specialized analyses, and Heavy-Ion,
- MC+MC Overlay being researched as an alternative to Pile-up MC.

ATLAS SENSITIVE TIME WINDOW

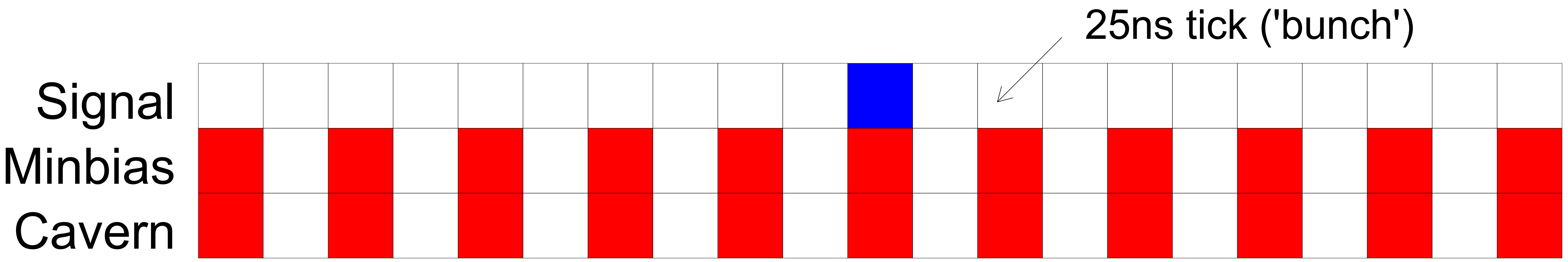


OPTION 1: SIMULATING PILE-UP

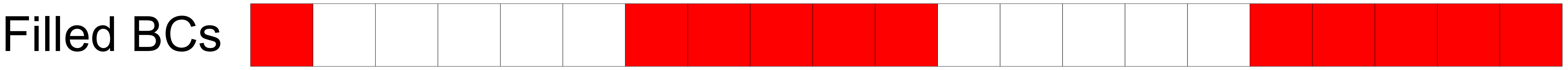
1. Run the event generation for “minbias” for single pp interactions:
    - Pythia8, A3 tune, NNPDF23LO PDFs,
    - inelastic (non-diffractive) and single/double diffractive.
2. Run GEANT4 on each minbias event to simulate detector energy/time “HITS”.
3. Combine multiple (thousands!) of HITS events during digitization:
    - use representative number of interactions per bunch crossing, shifted in time, to reproduce in-time and out-of-time pile-up,
    - sample bunch spacing/pattern within the sensitive time window of ATLAS detectors [-800, 150] ns.
4. Add model of detector noise separately.
5. No cavern background is added by default.

SIMULATING PILE-UP: VARIABLE BUNCH STRUCTURE

Example of a pile-up model with fixed 50ns spacing between filled bunches:



In reality the structure of filled and empty bunch crossings can be more complicated:

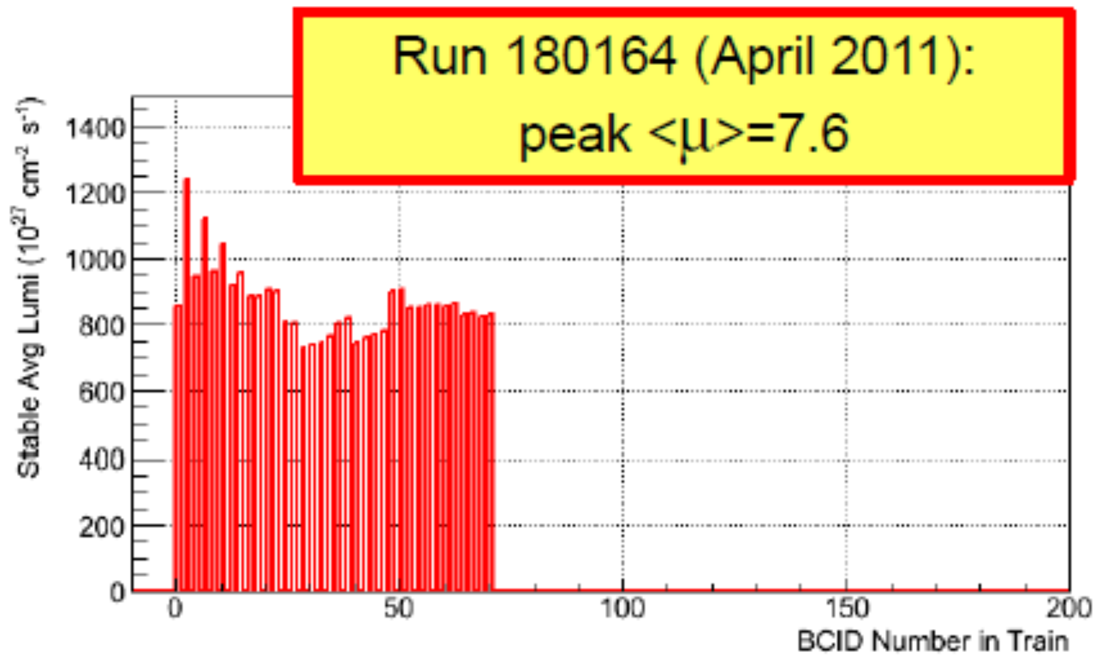


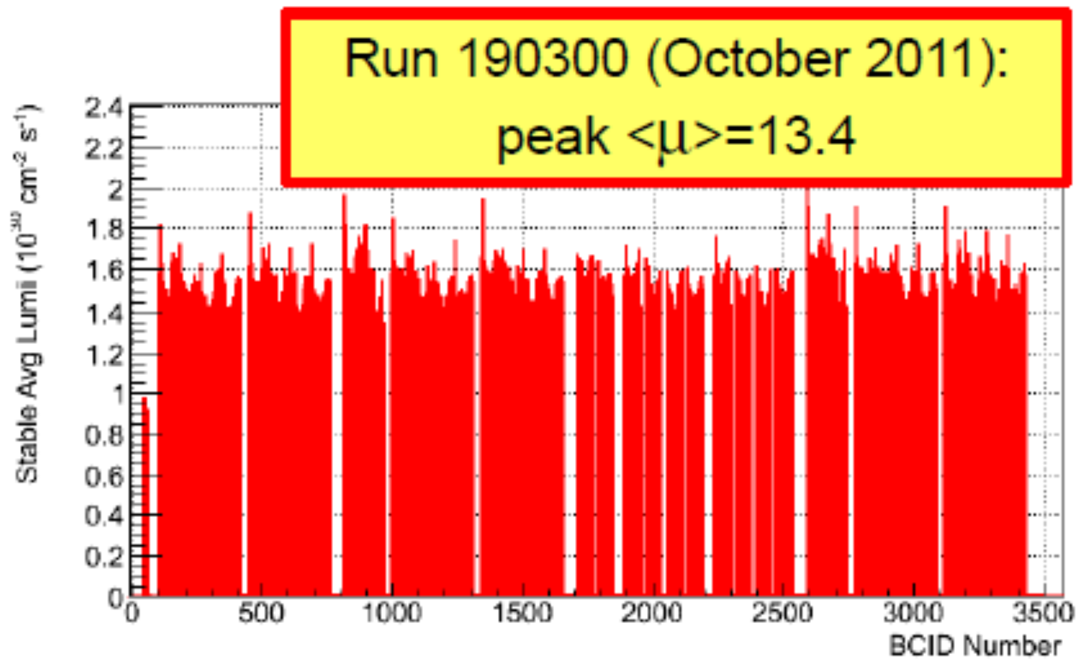
The pile-up/detector response is affected by the position of the triggering BCID in the bunch train. This is modelled in the simulation:

- Patterns can be up to 3564 elements in length and wrap-around if required.
- The triggering bunch crossing is picked from the filled bunch crossings in the pattern, with a probability proportional to the relative luminosities of each bunch crossing.

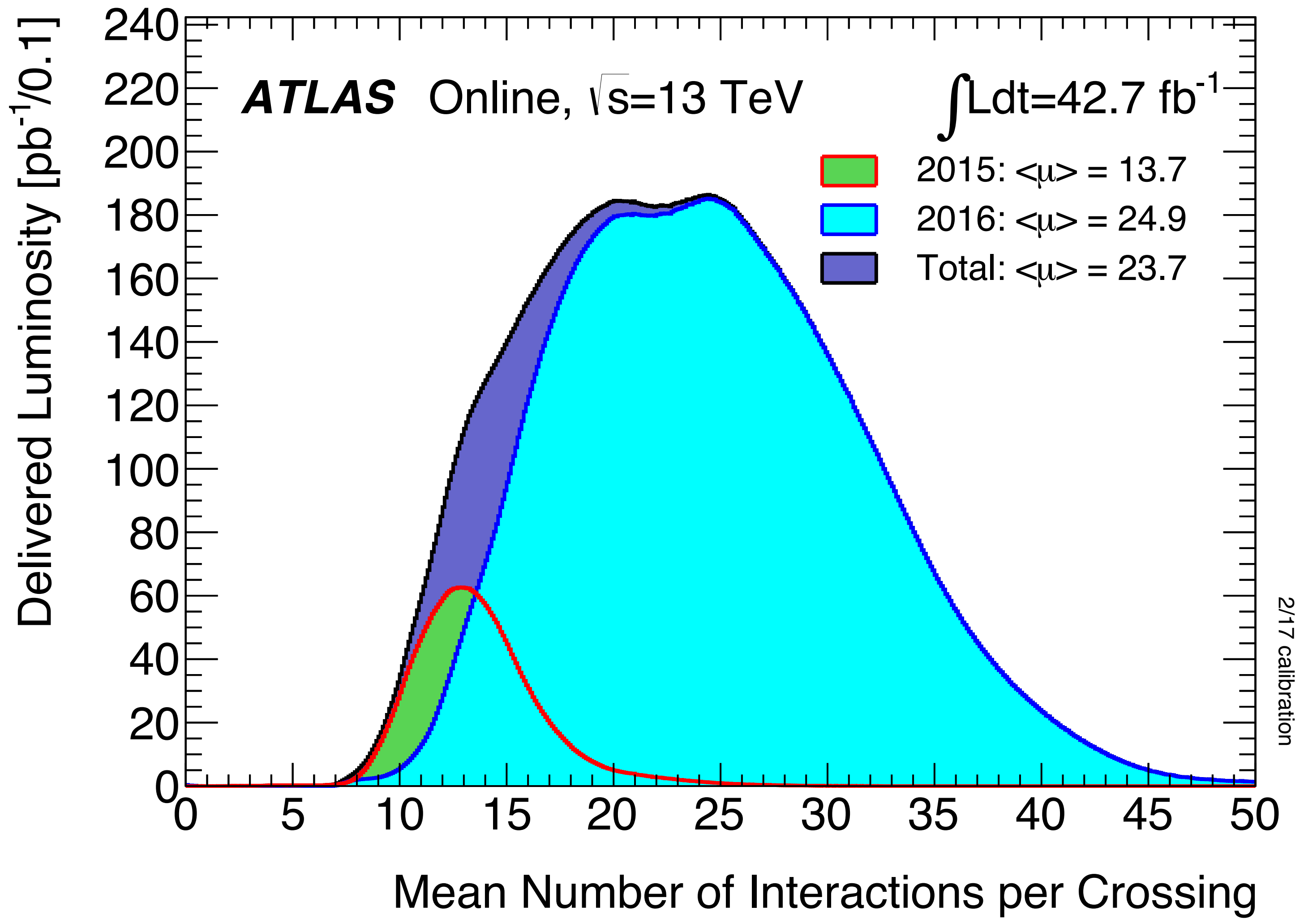
SIMULATING PILE-UP: VARIABLE BUNCH LUMINOSITY

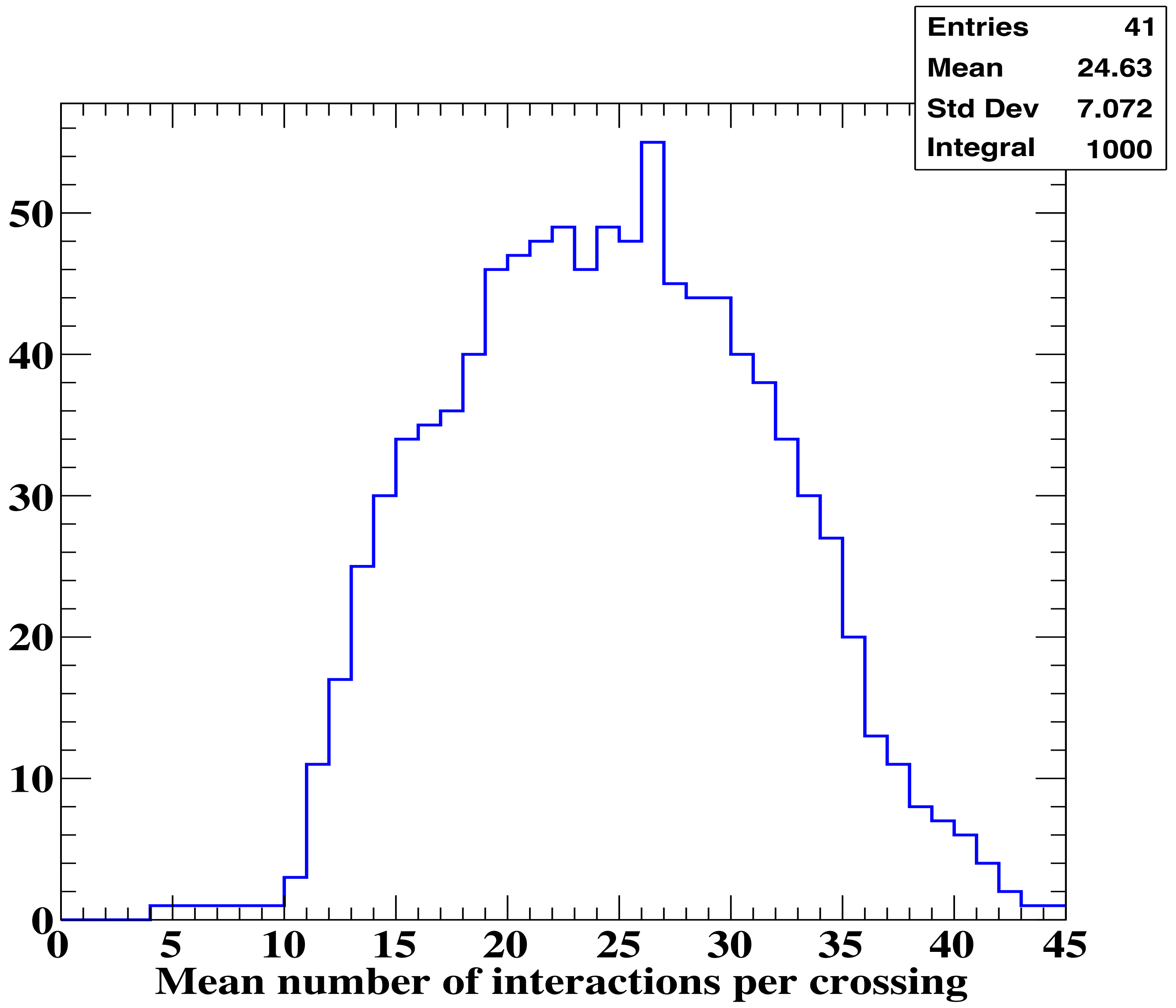
- Well known that <μ> varies over time.
- μ can also vary greatly from BCID to BCID in data. We usually set same mu for each BCID, as so far no analysis has shown sensitivity to this effect.





- Based on the real <μ> from data, a PDF is prepared for simulation to choose the values for each event:



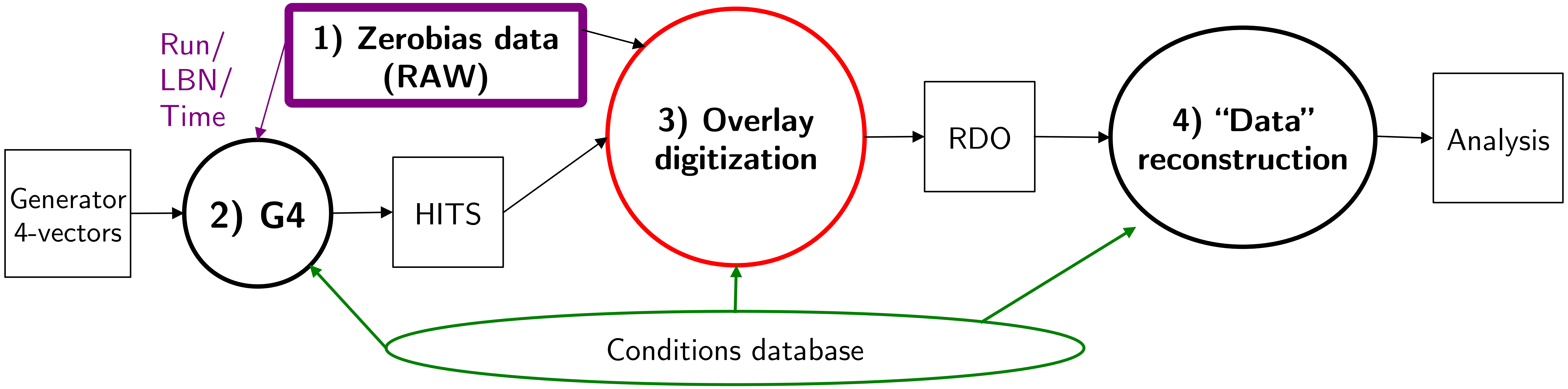


SIMULATING PILE-UP: DIGITIZATION

- For each hard-scatter event, make a selection of in-time and out-of-time pile-up events.
- To reduce memory usage one filled bunch crossing is processed at a time for all sensitive sub-detectors and then immediately discarded.
- Only read in / cache the parts of each event that are needed (e.g. discard HITS in silicon strips outside [-50,50] ns).
- Generating huge samples of minbias background is expensive!
  - 40M minbias events simulated for 2015 & 2016 (20M “low-pt”, 20M “high-pt”),
  - “low-pt”: no AntiKt6Truth jets with $p_T > 35$ GeV.
- Simulated minbias events are reused:
  - across various MC samples,
  - “low-pt” out-of-time minbias events within a MC sample.

OPTION 2: OVERLAY PILE-UP FROM DATA

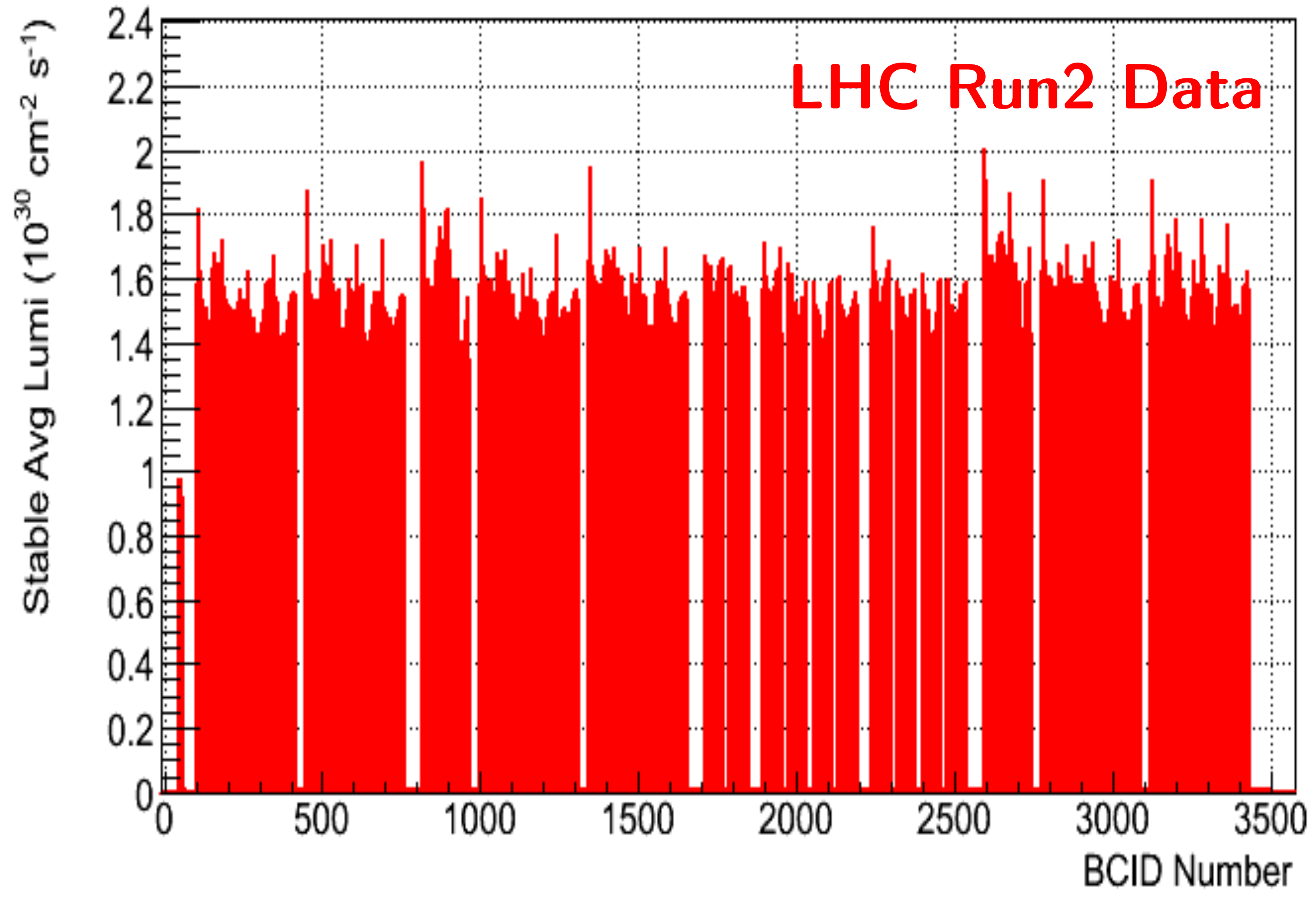
1. Start with input zerobias lumi-weighted RAW events, ordered in time.
2. Simulate a hard-scatter G4 event, with conditions matching each selected data event (beamspot/tilt, alignments, magnetic fields, etc.).
3. Overlay each zerobias data event with matching GEANT4 event at the detector channel level, then digitize combined signals.
4. Reconstruct the combined event as data.

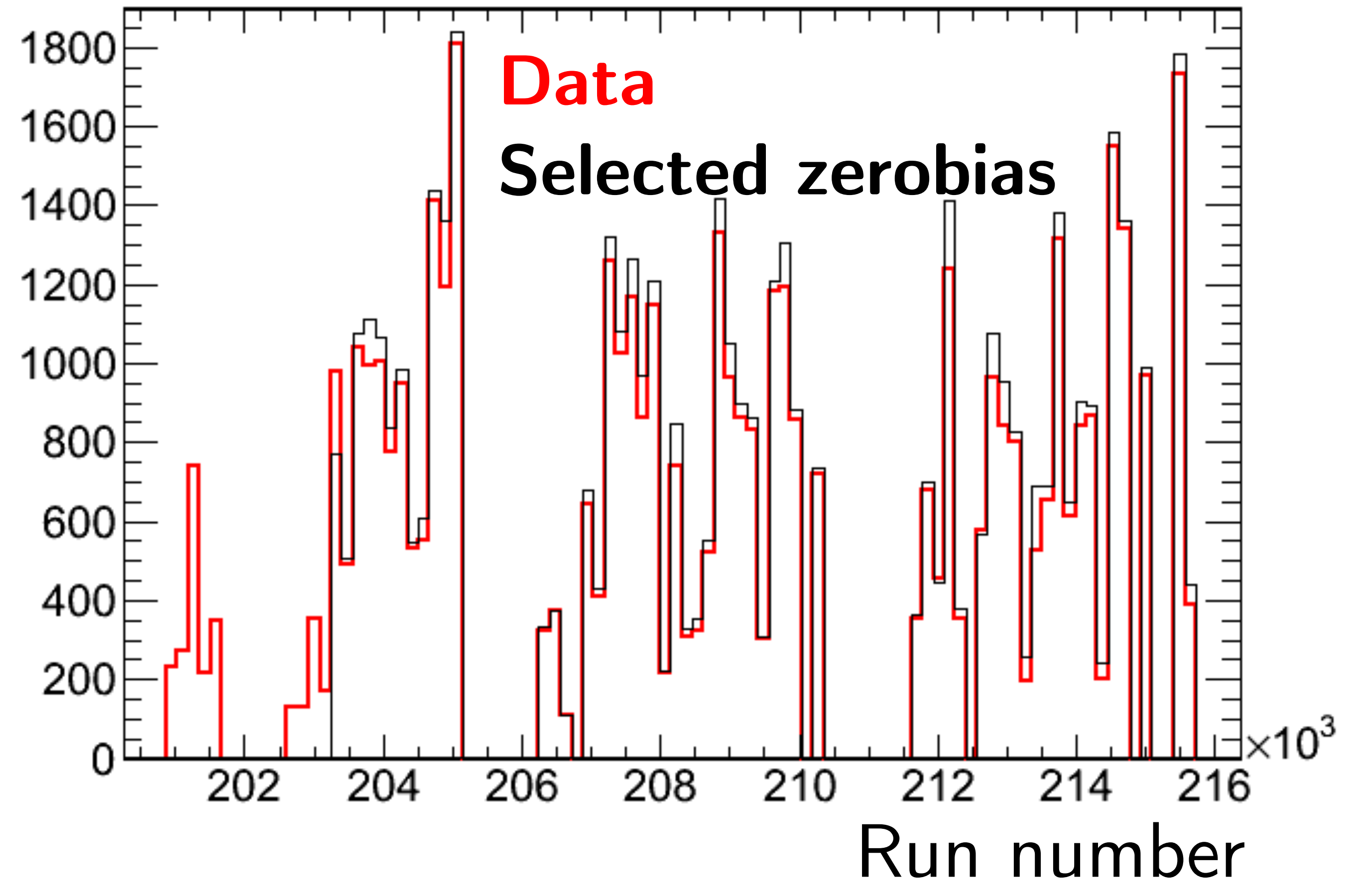


BACKGROUND DATA PREPARATION

- “Zerobias” data: Select event one accelerator turn after a high-pt trigger (e.g. L1_EM14) fires
  - next event in the same BC position,
  - proportional to luminosity in each BC.
- Prescaled to keep ~10 Hz in Run 2
  - ~3 Hz is selected at HLT to have a jet with $p_T > 40$ 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.).





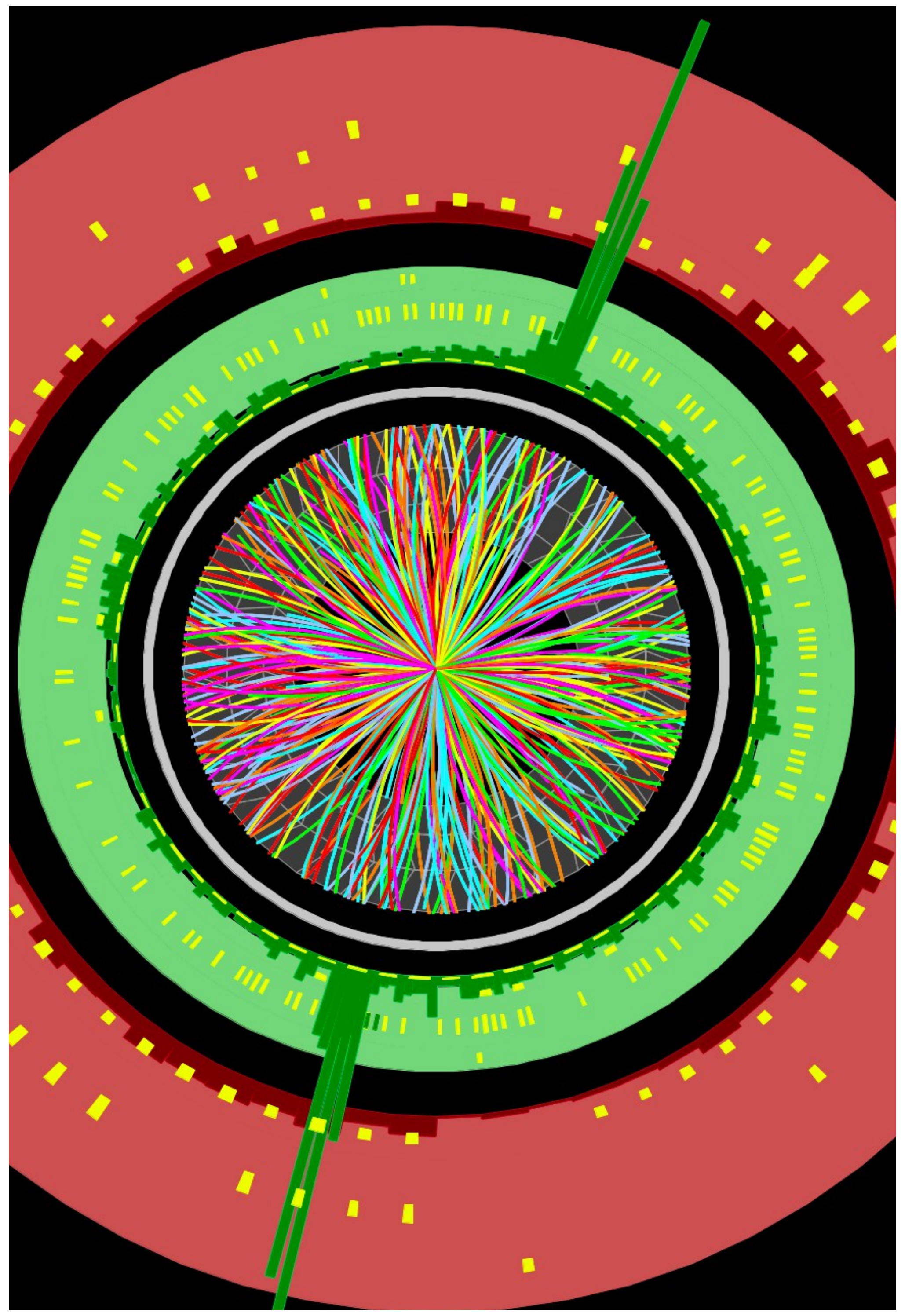
DATA OVERLAY VS. PILE-UP DIGITIZATION

- p-p overlay will get the details of the pile-up correct - using real data:
  - Analyses sensitive to pile-up modelling should benefit.
  - Should help detector performance groups where pile-up modelling is a problem. E.g. for Jet calibration at high pile-up.
- True detector noise and occupancy, including lumi-weighted variations and correlations between channels.
- Conditions (beamspot, dead channels, etc.) will track data more closely.
- Faster jobs with lower memory requirements.

- Can not simulate imaginary situations (e.g. Future high luminosity, different energy events, or upgrades to the detector).
- Less accurate when combining overlapping background and signal on the same channel for some sub-detectors (e.g. silicon). RDO contains less information than background HITS.
- Inaccuracies when simulating the trigger from overlay, since not all info used in real trigger is stored in the RAW data.
- Probably can not have as many overlay events as simulated events.
- Have to wait until all data is collected before generating simulations.
- Don't have the background MC truth information.

EXAMPLE: 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.



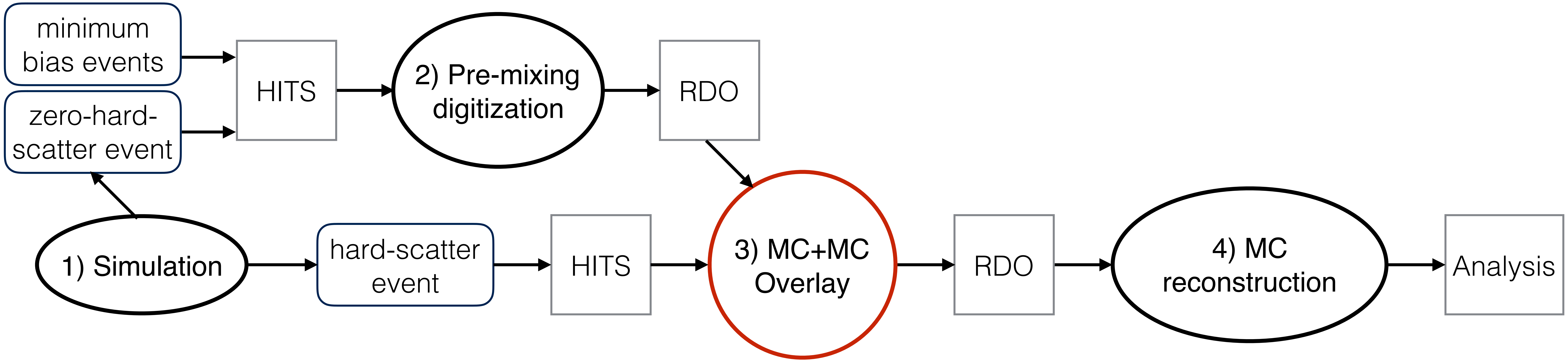
SPECIAL OPTION: SIMULATION USING EMBEDDING

- Embedding takes a data event (e.g. Z→mumu) and replaces objects with another type of object (e.g. taus) to emulate a related process.
- Critical for modeling Z→tautau background in H→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 pile-up,
  - no cells below zero-suppression threshold in data event,
  - can’t run L1 trigger simulation.

- Working on using overlay MC techniques to perform embedding:
  - Recorded ~5 Hz of non-zero-suppressed Z→mumu in 2016, to use as overlay input (instead of zerobias).
  - Simulate taus in G4 at same vertex/momenta as reconstructed muons.
  - Overlay the MC taus and Z→mumu data event, removing digits that were used to form the reconstructed Z→mumu tracks' hits.
  - Reconstruct overlaid event, performing subtraction of the muons' calorimeter cell energies (as in original embedding technique).

OPTION 2: OVERLAY PILE-UP FROM PRE-MIXED MC

1. Simulate a hard-scatter G4 event with usual configuration.
2. Pre-mixing of pile-up events: Digitization of zero-hard-scatter events (e.g. single neutrinos).
3. Digitize simulated hard-scatter event and overlay it on pre-mixed pile-up digits.
4. Reconstruct the combined event as ordinary MC.



MC+MC OVERLAY VS. PILE-UP DIGITIZATION

- CPU requirements for an MC campaign would be substantially reduced.
- Only have to digitize the background dataset once.
- I/O requirements for most production jobs would be reduced.
- A 2k event overlay + reconstruction job would only require one 8GB pile-up RDO file, rather than 40 GB of pile-up HITS.
- Could probably get away with fewer copies of the pile-up dataset.

- Different datasets would have the same pile-up backgrounds.
- Channels with HS and PU deposits just below threshold would be lost.
- RDOs are 4 MB per event compared to 900 kB per event for HITS.
- 1B event dataset would be 4 PB.

CONCLUSIONS

- ATLAS uses several methods for modelling pile-up and other detector backgrounds in our simulation:
  - simulated Pythia8 minbias event selection and digitization is currently used for most simulations,
  - data overlay is an alternate method, currently used for some performance studies, pp and HI physics analyses, and detector upgrade studies,
  - embedding is used for specialised studies, e.g. for $Z \rightarrow \tau\tau$,
  - MC+Mc overlay is being investigated as an alternative to pile-up simulation.

- Pile-up will become increasingly important with larger instantaneous luminosity.
- Working to improve accuracy and speed of all these methods.