Run3/4 simulations

(still qualitative) input from PWG-JE

Leticia Cunqueiro,Tatsuya Chujo CERN 3rd April 2019

How is the current MC/Data motivated?

Currently, no precise motivation. In the past we had 8M events per pthard bin. When we started working on substructure we multiplied X5, since 5 was the typical number of bins in the substructure observable.

We don't have a precise quantification of the **minimum** number MC events (or of MC jets per data jets in given bin) that is needed for the stable unfolding, but this can be toy-tested in some prototype analysis by:

*smoothening data according to a factor ~10-100 increase in stats and testing unfolding performance
*splitting the response to test by how much we can reduce the #jetsMC/#jetsData keeping unfolding stable

Which signal biasing techniques are already used and which other techniques could be used?

Our reco signal covers a wide range: p_T^{rec} goes from 10 GeV to 200 GeV
We need wide particle-level coverage in the response for the unfolding
We use $p_T$hard bins to enrich the high p_T region.

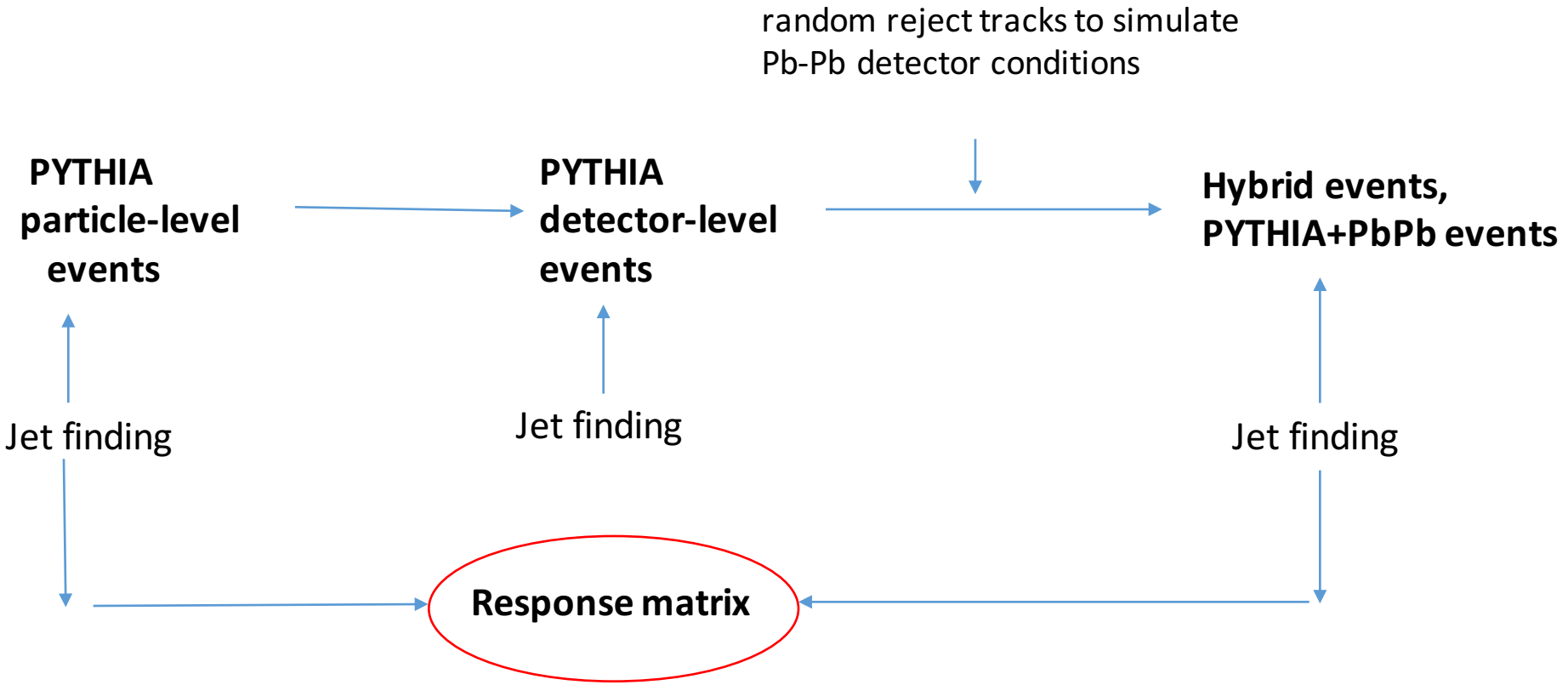
We do fine binning in pthard to reduce the outliers:
pthardbin_loweredges=(5 7 9 12 16 21 28 36 45 57 70 85 99 115 132 150 169 190 212 235)
pthardbin_higheredges=(7 9 12 16 21 28 36 45 57 70 85 99 115 132 150 169 190 212 235 -1)

Since $p_t$hard bins overlap, can the binning still be optimized?

Do we need full simulation of the background event or could it be replaced by a simplified version?

In the jet working group we do not simulate the full Pb-Pb event. We take PYTHIA events at detector level (worsening tracking conditions as in Pb-Pb) and we superimpose those to real Pb-Pb events, to study the heavy ion background effects.

Additional problem for Run3 we'll have to think about: we currently repeat the embedding process over full data for each pthard bin (this means 20 trains over full data to run the embedding)



Are there areas where full simulation could be replaced by fast simulation?

The rejection and smearing of tracks according to eta/phi maps should be accurate, let's test it.

How easy it is to fast-simulate the Emcal response? Was this ever tried?
For calorimeter cells, interesting approaches are on the radar, like CaloGun

Note that apparently the EMCal reco time does not stand out compared to other detectors-
->Hybrid approaches? Fast Tracking and Emcal via Geant?
Complex, tracks need to be propagated to simulate the showers but possible with multiple engines?

Which detectors/secondaries really need to be simulated for your analysis?

CENTRAL Barrel +EMCAL/DCAL