

Contribution ID: 229

Type: Parallel

Raw Data Reduction in the CMS Experiment for Run-3 and Phase-2

Thursday 10 July 2025 08:50 (20 minutes)

Reducing event and data sizes is critical for experiments at the LHC, where high collision rates and increased detector granularity rapidly increase storage and processing requirements. In the CMS experiment, a recent development to address this challenge is the "Raw" format: a new approach for recording silicon strip data in which only the reconstructed cluster's barycenter and average charge are stored, rather than the analog-to-digital converter counts from every strip. This format was successfully deployed online during Run-3 for PbPb collisions at CMS, achieving an event size reduction by nearly a factor of two and enabling CMS to record almost all hadronic minimum bias PbPb collisions.

To further enhance Raw', we optimized the number of bits used to encode the cluster barycenter and total charge, using tracking efficiency and resolution as benchmarks. Comparing standard RAW with Raw'shows that refining the bit precision yields stronger compression while maintaining similar performance.

Additionally, we introduce a lossless compression strategy that encodes distances between clusters instead of their absolute positions within a detector module. Unlike absolute positions, the distribution of these distances is peaked around zero, effectively reducing entropy of that variable. Consequently, LZMA compression becomes more efficient, allowing even stronger data reduction than the current Raw'algorithms without losing information integrity.

Lastly, we discuss projected data sizes for Phase-2 and explore extending these techniques to other CMS detectors, notably the High-Granularity Calorimeter, which is anticipated to generate a substantial fraction of future data.

Secondary track

Author: COLLABORATION, CMS

Presenter: COLLABORATION, CMS

Session Classification: T12

Track Classification: T12 - Data Handling and Computing