CMS Simulation in the HL-LHC Era

V. Daniel Elvira (Fermilab) for the CMS simulation group

CMS Simulation in Numbers

Tens of billions of events produced in Run 1/Run 2:

- CPU time/event range 15 sec – 3 min (1 min ave.)
- Memory usage/event ~2 GB
- 85% of total computing resources (production and analysis of simulated events), 40% G4 module
- Simulation + Reconstruction effort ~$100 M (Geant4, simulation, reconstruction code development and maintenance in experiments)
  - CMS simulation (hardware lifecycle, maintenance, operation): $5-10 M /year

HL-LHC computing needs 10-100 X larger depending on simulation/reconstruction solutions implemented

Challenge Areas

- Simulation complexity of the HL-LHC CMS detector
- Modeling of pileup interactions
- Computing performance
- Fast simulation

CMS Simulation in the HL-LHC Era

Authors: S. Banerjee, D. Elvira, M. Hildreth, V. Ivantchenko, K. Pedro, I. Osborne, S. Sekmen, L. Sexton-Kennedy

Introduction

For its size and complexity, the data produced and handled by modern HEP experiments earned a place in the world of what is known as Big Data. For example, the CMS experiment produced, reconstructed, stored, transferred, and analyzed more than 10 billion simulated events during Run 1 in 2009-2013. The amount of data collected and stored by the LHC

http://hepsoftwarefoundation.org/cwp-whitepapers.html

The LH-LHC CMS detector

New vertex (pixel) and tracker detectors, upgraded calorimeters and muons systems

- Tools for detector description
  CMS uses Detector Description Language (algorithmic description in XML files)
  - Revisit for improved description of complex shapes ? Opportunity for collaboration across experiments
- G4 physics improvement followed by CMS validation using test beam and collider data
  - HGC (end-cap) with silicon as active material, fast timing system (LYSO crystal scintillator)
- HL-LHC detector computing performance monitoring
  - Preliminary measurements (standalone CMS simulation): 25-70% increase in CPU/event

Pileup Interactions

High luminosity environment ➔200 pileup events

- I/O and local network bandwidth issues
  Merge thousands of min-bias interactions with one hard-scatter event (in-time, out-of-time contributions)
  - Run 1/2 solution: pre-mixing of peripheral interactions in site with high local network bandwidth, then distributed for global production to other sites
- Data Mixer is a potential evolution
  Same machinery to overlay a hard-scatter simulated event with real collider min-bias data (or data-on-data)
  - Detector specific ➔detector specific (collaboration across experiments difficult)
- Pileup not such a big issue for simulation but for reconstruction

Computing Performance

Significant improvement of G4 CPU performance through the life of the LHC: 35% with improved physics!

- Test/integrate/monitor CP for new versions of G4
  Highly optimized, no hotspots ➔ large CPU savings unlikely
- Test/integrate VecGeom (GeantV geometry library) to G4-based CMS application
  Preliminary, based on standalone CMS application: 5% CPU gain (scalar mode, solids only, without navigator)
- Test GeantV full engine prototype and libraries by 2018 -if it provides significant CPU time savings (factor of 2-5 promised)
- HEPCloud (hybrid system – CPUs, coprocessors/accelerators, HPC facilities, commercial resources)

For the HL-LHC CMS experiment, reconstruction (not simulation) is likely to be the largest consumer of computing – due to pileup

Fast Simulation

May be used more frequently to save computing resources

- Implement upgraded detectors in a more configurable and flexible way
- Explore GeantV as an alternative framework for FastSim
  FastSim embedded in framework – common, concurrent infrastructure to build libraries of parameterized quantities from full simulation, tools for Fast-to-Full simulation tuning and comparison
- Explore ultra fast self-tuning non-parametric simulation based on lookup tables – map generated and simulated events

(Opportunity for collaboration across experiments)