ECHEP Generators area update

Andy Buckley, Marek Schoenherr

24 Mar 2020

Generators work package

- **Challenge: current MC high-precision bulk sample generation alone is on track to saturate HL-LHC compute budget. Mix of method & logistics approaches needed**

- **Main issues & possible activities:**
  - low efficiency in (particularly NLO) ME phase-space sampling
    - machine learning -> better proposal functions
  - high rates of negative weights in NLO ME/shower matching (kills statistical power)
    - new matching schemes, post-hoc methods?
  - general computational efficiency and match to Grid architectures
    - comprehensive performance profiling from expt MC-campaign perspective
    - reconfigure workflows to use HPC systems
    - investigate any role (and technical feasibility) for GPU acceleration
    - decay filtering / re-hadronisation efficiency

- **Need to be realistic about goals for 6 months of part-time effort**

Activities summary

- **Machine learning for improved phase-space sampling**
  - Kicked off by Bendavid (2017); realistic use-cases from Hoeche et al (2020), Schumann et al (2020) for V+*n*j production
  - No improvements achieved for V+>2j though, thus not trivial
  - ECHEP 6 month prospect: nope! Can any “prospecting” be usefully done?

- **Addressing negative-weight rates**
  - Study https://arxiv.org/abs/2002.12716 scheme; characterise/profile in new aMC@NLO/Py8 when publicly available
  - Alternative schemes? Some discussion of post-hoc reweighting (i.e. overgenerate at ME/LHE level and filter: save event-dressing & sim/reco CPU)

- **Accuracy vs efficiency**
  - Highest accuracy not needed for every topology
  - Survey where accuracy needed and where speed-ups can be achieved

Activities summary (2)

- **General computational efficiency**
  - Cannot expect order of magnitude gains across the board, but experiment expertise (and incentives) may help
  - Obvious: run profiling on “all” gens… need >superficial: triage for next iteration

  - Some interventions may be more logistical than technical
    - Event-sharing — what barriers vs potential gains? *Not really factor 2…*
    - HPC: huge MPI scaling benefits NLO high-multiplicity event gen. New HDF event format designed by FNAL et al -> changes to workflow, what impact?
    - Accelerators: GPUs visited many times, mismatch to MC shower algs and event trivial ||… still true? Barrier reduced by new toolchains like SYCL?

  - Aside from high-tech samples, specific hadron/decay configurations also a large CPU sink for expts. Origins from all SHG elements: better forced hadr/decay?