

HSF tools review

LHAPDF, HepMC, Rivet, and so-on

Andy Buckley

University of Glasgow

HSF MC generators meeting

24 June 2021

MC-tools review

❖ Tools to support and build on MCs are an important part of the collider-pheno ecosystem

- Different/overlapping developer pool — more experiment involvement
- But typically *partial/complete absence of funding!* Not sustainable? End of MCnet...

❖ I'll give a high-level review of the status for:

- **LHAPDF**
Library for PDF-value access. v6 redevelopment -> pure interpolation library. Huge MC CPU fraction — largely fixable by better call strategy. Opportunities for speed increases: more caching, less flexibility (?), GPUs; need for precision ipols
- **HepMC (and LHE, etc.)**
Standard library for event-graph representation. v3 rewrite addressed some issues, increases flexibility. Performance still an issue, need supporting tools
- **Rivet**
Analysis-preservation tool, used in MC validation, development, tuning, and BSM reinterpretation. Tech opportunities, mainly needs more expt submission
- **YODA, Professor, mcutils/heputils**
More small “support” and downstream tools, largely unsupported

LHAPDF

LHAPDF v6 redeveloped ~2012

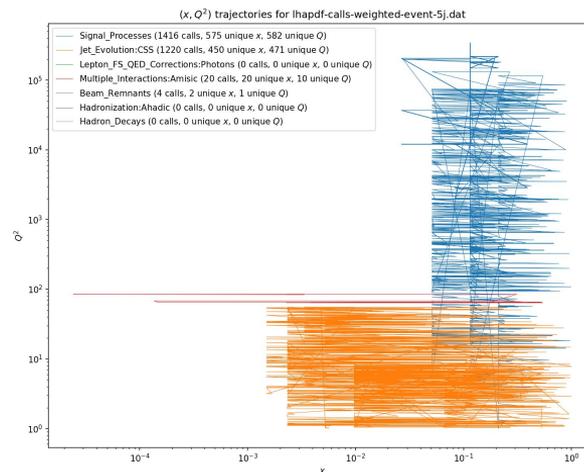
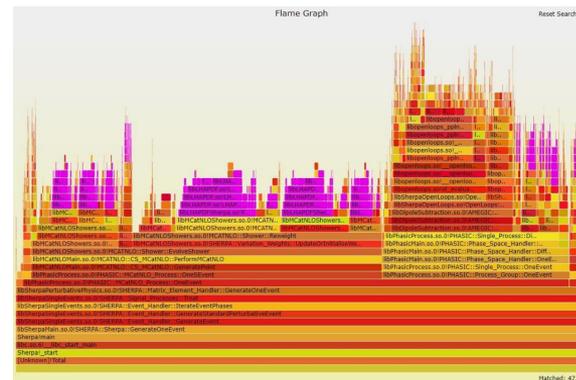
- Purely interpolation, unlike v5
- Standard data format for all PDF-set data; MPI-friendly

Profiling Sherpa in experiment-like NLO V+jet setups:

- main CPU sinks = PDFs & transcendental functions
- 10% gain from Intel math lib, but already in LHC use
- 2020: LHAPDF 6.3 introduced thread-safe caching...
- needs to be *called* efficiently, too: impl in Sherpa. MG5?

Next activities:

- current GSoC student on better interpolation: 5th & 7th order polynomials, Chebyshevs. Actively look for ways to cache, e.g. pre-computation of knot derivatives in init
- MCnet placement on GPU ipol starting July, UK SWIFT-HEP part-time RA on MC/LHA from Oct



HepMC ecosystem

Data formats

- now native support for textual v2 & v3, zipped text, ROOT... others? Improvements? Heavy-ion interop?
- schema evolution, e.g. “event number → long” discussion

Parton-level events also important

- Textual LHE format, now (non-standardised?) HDF5?
- HepMC3 can embed LHE. No plan for the H5 info... yet
- Neither lib-development institutionally supported

Manipulation suite(s)

- Still inconvenient to filter events and files (cf. mcutils)
- Being able to write code or use cmd line utils to filter, cut, concatenate each event format would be amazing
- Again, “grunt work” requires paid effort

Rivet

❖ Rivet v3 released in June 2019, now 3.1.4

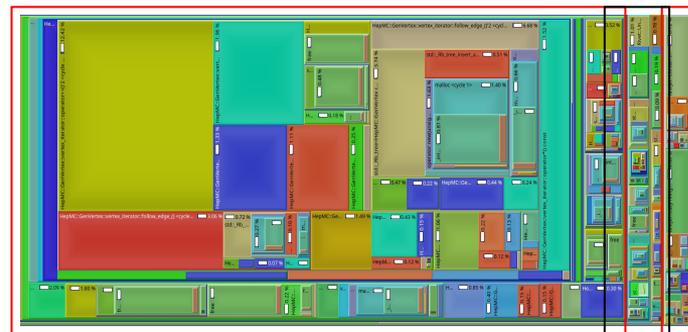
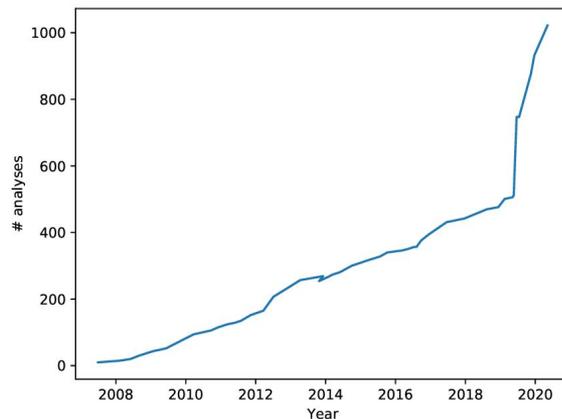
- automatic MC systematics multiweight handling
- “perfect” re-entrant run-merging
- passing optional parameters to analyses
- 2-pass heavy-ion observable machinery

❖ Development areas

- statistical-object handling, HepData integration
- detector folding/smearing, systematics & bootstrap
- technical developments like H5 formats to enable/speed-up
- embeddability in adaptive-sampling/optm code

❖ Code efficiency profiled in GSoC 2020

- Limit by HepMC: IO, units-copy, graph traversal
Projection-caching makes thread-safety tricky.
- Histo-fill calls for O(100) multiweights: non-trivial



HepMC

Rivet FJ

Rivet analysis coverage and compliance

- ❖ **Semi-active preservation programmes in expts**
 - Various levels of official requirement, but growing
 - Role in MC dev, validation, fitting requires better compliance
- ❖ **Experiment published-analysis coverage tracked at e.g. <https://rivet.hepforge.org/rivet-coverage>**
- ❖ **Status:**

CMS lags ATLAS on core SM analysis coverage: 36% vs 54%. ALICE programme activating, also HI beyond LHC. LHCb have many Rivety observables, but $\lesssim 1$ active person
- ❖ **Critical issue:** analyses submitted with reference data incompatible with that submitted to HepData!

Rivet analysis coverage (no searches, no heavy ion)

Rivet analyses exist for 919/3986 papers = 23%. 146 priority analyses required.

Total number of Inspire papers scanned = 6277, at 2021-05-17

Breakdown by identified experiment (in development):

Key	ALICE	ATLAS	CMS	LHCb	Forward	HERA	e^+e^- (\geq GeV)
Rivet wanted (total):	73	108	131	202	15	448	694
Rivet REALLY wanted:	17	39	62	8	0	13	1
Rivet provided:	14/87 = 16%	145/253 = 57%	82/213 = 38%	16/218 = 7%	8/23 = 35%	9/457 = 2%	171/868 = 20%

Show greylist Show blacklist

ALICE	ATLAS	CMS	LHCb	Forward	HERA	e^+e^- (≥ 12 GeV)	e^+e^- (≤ 12 GeV)
-------	--------------	-----	------	---------	------	---------------------------	---------------------------

ATLAS: Measurements of $W^+W^- + \geq 1$ jet production cross-sections in pp collisions at $\sqrt{s} = 13$

Inspire: 1852328 · arXiv: 2103.10319 · CDS: 2757281 · HepData: ins1852328 · Report IDs: CERN-EP 2021-030

ATLAS_2021_11852328

ATLAS: Measurements of differential cross-sections in four-lepton events in 13 TeV proton-proton collisions

Inspire: 1849535 · arXiv: 2103.01918 · CDS: 2753518 · HepData: ins1849535 · Report IDs: CERN-EP 2021-019

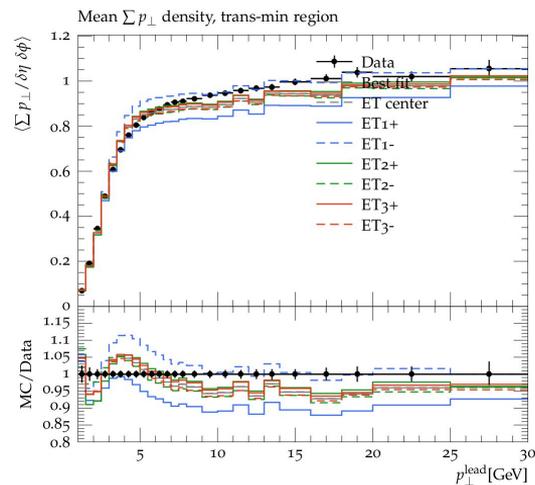
ATLAS_2021_11849535

ATLAS: Observation of photon-induced W^+W^- production in pp collisions at $\sqrt{s} = 13$ TeV using

Inspire: 1822077 · arXiv: 2010.04019 · DOI/journal: 10.1016/j.physletb.2021.136190 · CDS: 2740726 · Report IDs: CERN-EP-2020-165

Other tools

- ❖ **YODA:** statistics library underpinning Rivet. Thread-local (cf. ROOT), lightweight. Being extended for generic-object binned containers, coherent N-dim histograms, correlations, new binary format, ... *no support / GSoC 2020*
- ❖ **Professor:** MC-tuning and general parametrisation/optimisation tool. Rewrite as Apprentice added rational approximants, better correlations, etc. — but no uptake? *Unsupported?*
- ❖ **mcutils/heputils:** general tools to help with MC analysis (AB). HepMC2-specific — integrate search and graph-reduction algs into HepMC3? Standardise physical-reduction / write-out levels?
- ❖ **HepData:** operation supported, but little *development*



Summary

- ❖ Tools are a key element in making MC generators useful. But technical projects \Rightarrow no funding source
- ❖ MCnet/CEDAR role has been crucial in enabling this activity. Ending... to be replaced with?
- ❖ LHAPDF and HepMC crucial to expt MC production:
 - Huge PDF-ipol cost. Restructuring generators will help a lot, but ipol library also needs attention
 - N^3 LO etc. require more precise PDF interpolation: more cost
 - Virtual-photon PDFs from v5 will be needed again (cf. EIC etc.)
 - Need better graph algs, copying, manipulation tools (lib and command-line)... and smaller event-graph options from MCs
- ❖ Don't forget the small tools!

