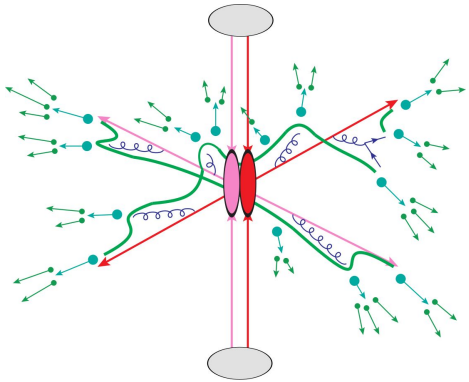
MC event-generator recognition

Andy Buckley
University of Glasgow
for the MCnet Collaboration

HSF/IRIS-HEP software citation & recognition workshop
23 Nov 2022

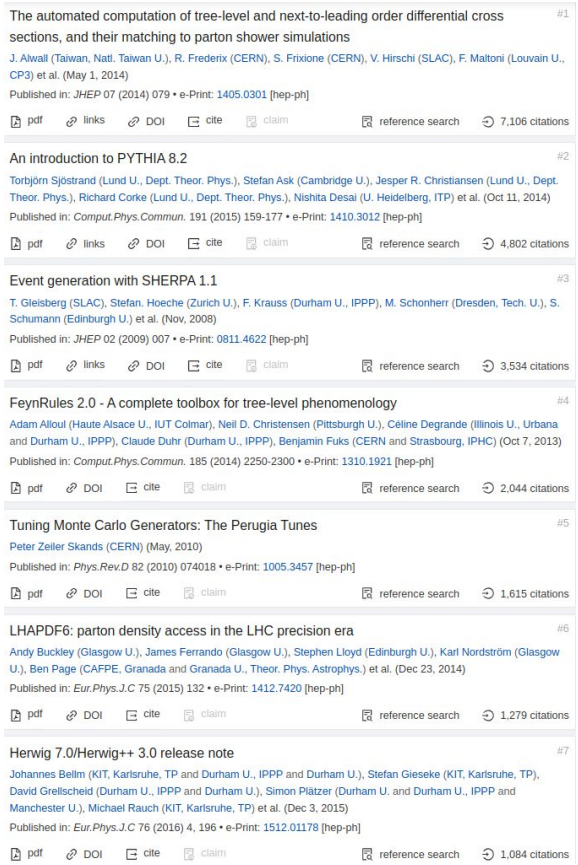
Recognition of MC generator software

- **MCnet represents the main general-purpose MC gens**
  - Herwig, Pythia, Sherpa + MG5_aMC + others
  - Also MC-ecosystem tools (Rivet, LHAPDF etc.)
- **Headline is positive: some (very) highly-cited work**
  - Several 500+ citation papers, typically generator “main papers”
    - see next slide
  - Remaining physics papers (the science substance) far behind
  - So… good, but numbers should be more broadly distributed
    - note: generator citation-magnet papers only every ~10y; can be a lottery for students/RAs to be associated in time
  - technical/maintenance work *not* significantly recognised/rewarded
- **Is citation enough?**
  - Certainly citation counts are not enough as a metric of impact
  - Several instances where citation of the foundational work is “lost” after the first intermediate layer
  - Understandable, but the whole citation network is important: work cited by popular papers should benefit



MCnet top cites

aka Inspire “r mcnet* and topcite 500+” search



The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations #1

J. Alwall (Taiwan, Natl. Taiwan U.), R. Frederix (CERN), S. Frixione (CERN), V. Hirschi (SLAC), F. Maltoni (Louvain U., CP3) et al. (May 1, 2014)

Published in: *JHEP* 07 (2014) 079 • e-Print: 1405.0301 [hep-ph]

pdf links DOI cite claim reference search 7,106 citations

An introduction to PYTHIA 8.2 #2

Torbjörn Sjöstrand (Lund U., Dept. Theor. Phys.), Stefan Ask (Cambridge U.), Jesper R. Christiansen (Lund U., Dept. Theor. Phys.), Richard Corke (Lund U., Dept. Theor. Phys.), Nishita Desai (U. Heidelberg, ITP) et al. (Oct 11, 2014)

Published in: *Comput.Phys.Commun.* 191 (2015) 159-177 • e-Print: 1410.3012 [hep-ph]

pdf links DOI cite claim reference search 4,802 citations

Event generation with SHERPA 1.1 #3

T. Gleisberg (SLAC), Stefan. Hoeche (Zurich U.), F. Krauss (Durham U., IPPP), M. Schonherr (Dresden, Tech. U.), S. Schumann (Edinburgh U.) et al. (Nov, 2008)

Published in: *JHEP* 02 (2009) 007 • e-Print: 0811.4622 [hep-ph]

pdf links DOI cite claim reference search 3,534 citations

FeynRules 2.0 - A complete toolbox for tree-level phenomenology #4

Adam Alloul (Haute Alsace U., IUT Colmar), Neil D. Christensen (Pittsburgh U.), Céline Degrande (Illinois U., Urbana and Durham U., IPPP), Claude Duhr (Durham U., IPPP), Benjamin Fuks (CERN and Strasbourg, IPHC) (Oct 7, 2013)

Published in: *Comput.Phys.Commun.* 185 (2014) 2250-2300 • e-Print: 1310.1921 [hep-ph]

pdf DOI cite claim reference search 2,044 citations

Tuning Monte Carlo Generators: The Perugia Tunes #5

Peter Zeiler Skands (CERN) (May, 2010)

Published in: *Phys.Rev.D* 82 (2010) 074018 • e-Print: 1005.3457 [hep-ph]

pdf DOI cite claim reference search 1,615 citations

LHAPDF6: parton density access in the LHC precision era #6

Andy Buckley (Glasgow U.), James Ferrando (Glasgow U.), Stephen Lloyd (Edinburgh U.), Karl Nordström (Glasgow U.), Ben Page (CAFPE, Granada and Granada U., Theor. Phys. Astrophys.) et al. (Dec 23, 2014)

Published in: *Eur.Phys.J.C* 75 (2015) 132 • e-Print: 1412.7420 [hep-ph]

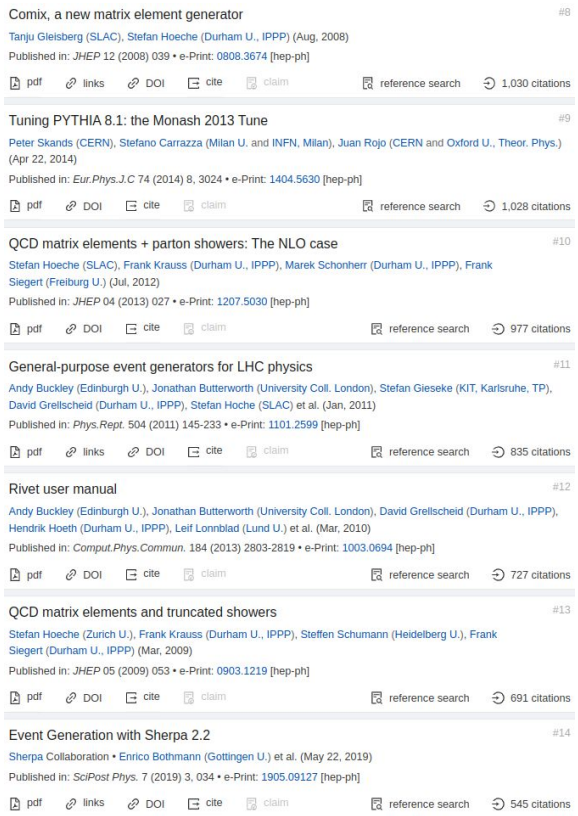
pdf DOI cite claim reference search 1,279 citations

Herwig 7.0/Herwig++ 3.0 release note #7

Johannes Bellm (KIT, Karlsruhe, TP and Durham U., IPPP and Durham U.), Stefan Gieseke (KIT, Karlsruhe, TP), David Grellscheid (Durham U., IPPP and Durham U.), Simon Plätzer (Durham U. and Durham U., IPPP and Manchester U.), Michael Rauch (KIT, Karlsruhe, TP) et al. (Dec 3, 2015)

Published in: *Eur.Phys.J.C* 76 (2016) 4, 196 • e-Print: 1512.01178 [hep-ph]

pdf DOI cite claim reference search 1,084 citations



Comix, a new matrix element generator #8

Tanju Gleisberg (SLAC), Stefan Hoeche (Durham U., IPPP) (Aug, 2008)

Published in: *JHEP* 12 (2008) 039 • e-Print: 0808.3674 [hep-ph]

pdf links DOI cite claim reference search 1,030 citations

Tuning PYTHIA 8.1: the Monash 2013 Tune #9

Peter Skands (CERN), Stefano Carrazza (Milan U. and INFN, Milan), Juan Rojo (CERN and Oxford U., Theor. Phys.) (Apr 22, 2014)

Published in: *Eur.Phys.J.C* 74 (2014) 8, 3024 • e-Print: 1404.5630 [hep-ph]

pdf DOI cite claim reference search 1,028 citations

QCD matrix elements + parton showers: The NLO case #10

Stefan Hoeche (SLAC), Frank Krauss (Durham U., IPPP), Marek Schonherr (Durham U., IPPP), Frank Siegert (Freiburg U.) (Jul, 2012)

Published in: *JHEP* 04 (2013) 027 • e-Print: 1207.5030 [hep-ph]

pdf DOI cite claim reference search 977 citations

General-purpose event generators for LHC physics #11

Andy Buckley (Edinburgh U.), Jonathan Butterworth (University Coll. London), Stefan Gieseke (KIT, Karlsruhe, TP), David Grellscheid (Durham U., IPPP), Stefan Hoche (SLAC) et al. (Jan, 2011)

Published in: *Phys.Rept.* 504 (2011) 145-233 • e-Print: 1101.2599 [hep-ph]

pdf links DOI cite claim reference search 835 citations

Rivet user manual #12

Andy Buckley (Edinburgh U.), Jonathan Butterworth (University Coll. London), David Grellscheid (Durham U., IPPP), Hendrik Hoeth (Durham U., IPPP), Leif Lonnblad (Lund U.) et al. (Mar, 2010)

Published in: *Comput.Phys.Commun.* 184 (2013) 2803-2819 • e-Print: 1003.0694 [hep-ph]

pdf DOI cite claim reference search 727 citations

QCD matrix elements and truncated showers #13

Stefan Hoeche (Zurich U.), Frank Krauss (Durham U., IPPP), Steffen Schumann (Heidelberg U.), Frank Siegert (Durham U., IPPP) (Mar, 2009)

Published in: *JHEP* 05 (2009) 053 • e-Print: 0903.1219 [hep-ph]

pdf DOI cite claim reference search 691 citations

Event Generation with Sherpa 2.2 #14

Sherpa Collaboration • Enrico Bothmann (Gottingen U.) et al. (May 22, 2019)

Published in: *SciPost Phys.* 7 (2019) 3, 034 • e-Print: 1905.09127 [hep-ph]

pdf links DOI cite claim reference search 545 citations

pre-MCnet, PYTHIA 6.4: 12,500

MG automation: 7,000

Pythia8: 4,200

Sherpa 1.1: 3,500

FeynRules: 2,000

Perugia tunes: 1,600

LHAPDF6: 1,300

Herwig 7: 1,100

Comix ME gen: 1,000

Monash2013 tune: 1,000

NLO MEPS: 1,000

Slightly odd distribution?

Tunes overvalued?

Hon mention: FastJet: 5,000

Experiment MC gen citation

- **Experiment citation culture for MC is pretty good**
  - e.g. Pythia 8: 556/645 ATLAS publications, 654/771 CMS, 306/355 LHCb, 150/286 ALICE
  - Should it be closer to 100% for most? Historic offset cf. PY6? Still, not bad
- **Several influencing factors:**
  - MC community primed for start of LHC, much noise and agitation at experiments in run-up and early phase ⇒ culture change
  - Official GUIDELINES document publicised aggressively early on, since ~2007 (also online https://www.montecarlonet.org/guidelines)
  - Many MCnet gens & tools write out citation instructions & ref the guidelines on the command-line
    - Maybe helps for pheno community, but most experimentalists never run the MC
  - Experiments (at least ATLAS) have developed standard texts to describe generator samples, and these are checked in extensive (sometimes painful!) internal review
    - These do better than just catching the main “magnet” papers, but rely on the knowledge and continual effort of curators & editorial reviewers (journals don’t catch missed citations, MCnet does not particularly check & chase)

Generator-tools ecosystem

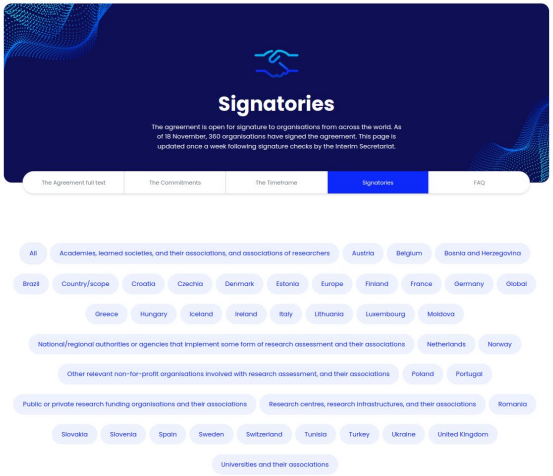
- **MCnet covers more than generators: also tools**
- **LHAPDF: 1300 cites since 2014 (after switch from v5)**
  - Like generators: used in every paper, std cite.
- **Rivet: 730 (2010 paper) + 140 (2020 Rivet 3)**
  - Some underreporting cf. seen plots, use in CMSSW for MC-truth classification, in ATLAS for EFT mapping
  - Many experimental theses cite it and mention in their text, but citations don't appear in the final experimental papers
- **Professor MC tuning tool: 350 (2009 paper)**
  - A specialist tool… but underlies major MC tunes, which are far more heavily cited (ATLAS Py8 A14 has 1,250; PY6 1,050)
- **Others:**
  - HepMC (400+30) + LHE (530): big underestimates by comparison
  - (Contur, etc. most papers are more physics than tools; one technical ref, too early to see effect)
  - How best to reflect foundational technical libraries one step removed?

We also monitor and cite experimental publications entered into the Rivet analysis database!

Policy & technology

- **Citation of *papers* is still the overwhelming norm**
  - Rivet & LHAPDF (at least) also have Zenodo DOIs
  - Maybe also SciPost Codebases, cf. latest Py8 paper?
  - These don't seem to be picked up by Inspire:
    - *the* crucial gateway for community referencing

- **Assessment policy**
  - Agreement on Reforming Research Assessment at https://coara.eu/
    - 350+ signatories from 40 countries, including CERN, CNRS…
    - Plus UKRI, ANR, DFG … only (?) Europeans so far?
  - Declaration on Research Assessment (DORA) https://sfdora.org/
  - General aims to reflect diverse outputs, not just journal publications
    - Does anyone actually dare submit a sw DOI to a high-stakes grant application?! Needs a cultural clean break



- **BibTeX outputs for physics-specifics produced by e.g. Herwig**
  - Is this useful? Again, goes into logfiles most experimentalists never see
  - Could add more dynamic cut-n-paste bib entries in online manuals?
  - What technical support to encourage "full" citation might actually work

Summary / open questions

- **Generators are broadly well-cited: thanks to LHC experiments for efforts!**
  - Citation culture falls off rapidly away from headline generator-release papers
  - More technical, physics-specific papers probably under-recognised
- **And generally a problem with reflecting foundational contributions of software (both tools and generators) built upon to make something flashy**
  - Technical development and physics development are not separable in MC gens
  - Citations from “users” often only refer to the code they directly ran (quite reasonably)
- **Where is the incentive for technical engineering & support work?**
  - Long a problem in generators, esp. as technology & user needs become more complex
  - Career incentives are theory-cultural, esp. for ECRs
  - Main papers only ~1/decade; funders’ research-output eligibility windows more like 5y
- **Are there technical or policy solutions?**
  - Acknowledging citation “impact chains” in standard metrics?
  - Automatic provision of BibTeX/other ref dbs from MC runs & expt samples?
  - Experiments to pioneer direct software-release DOI citation? (Needs recognition in Inspire, etc.)
  - Citing software *from* software, not just publications — Zenodo, SciPost? + Inspire tracking again