Data/MC comparison & tuning

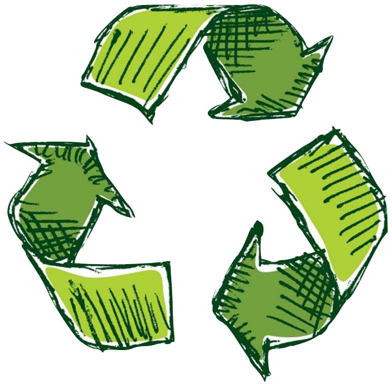
The why & how of Rivet analysis preservation

Andy Buckley,
University of Glasgow

CHACAL 2024
Johannesburg, 22 Jan 2024

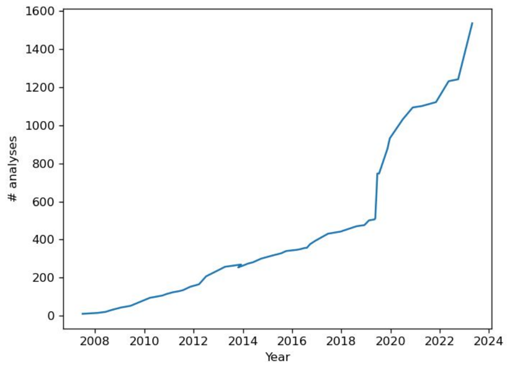
Outline

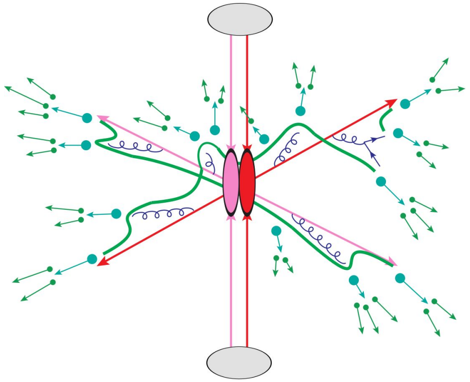
- ❖ **The Rivet toolkit and project**
- ❖ **Origins and generic analysis preservation**
- ❖ **Early developments**
  - ➢ Soft QCD & MC tuning
  - ➢ Refining fiducial definitions
- ❖ **Retooling for precision at the LHC**
- ❖ **Rivet for heavy ions and EIC**
- ❖ **Beyond the Standard Model: searches and Contur**
- ❖ **The future of Rivet**
- ❖ **Joining the party**



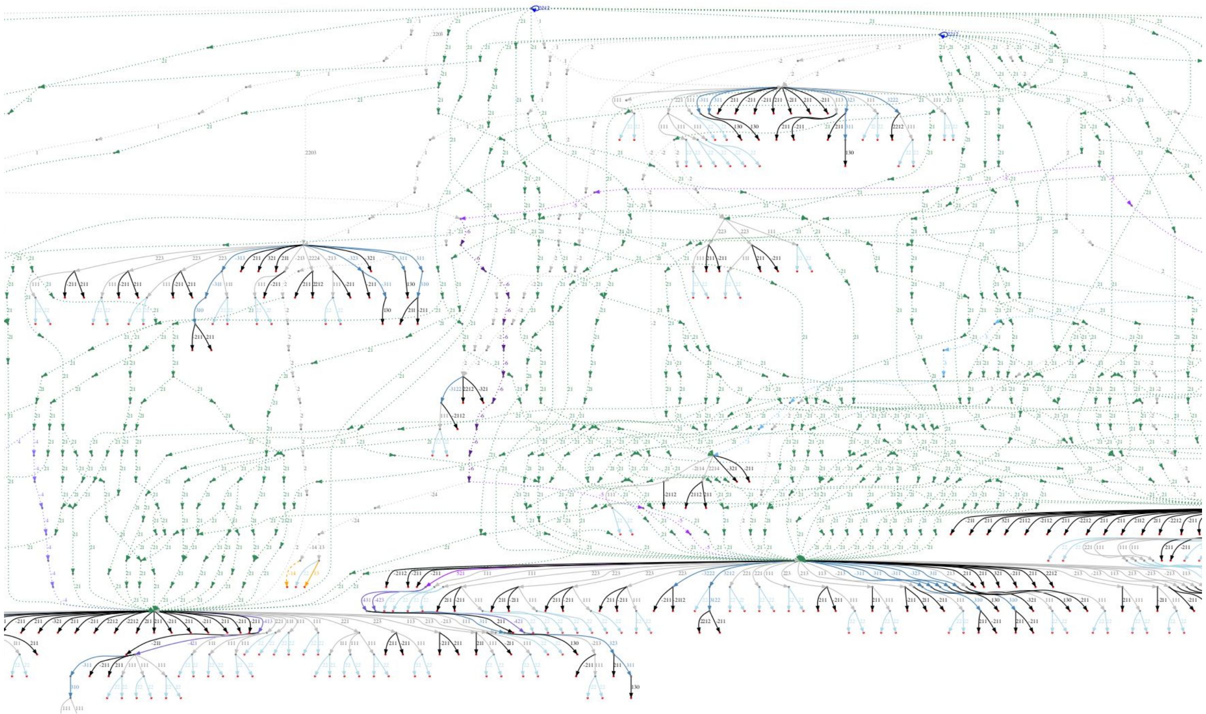
What is Rivet?

- **The "LHC standard" MC analysis toolkit**
- **More broadly a project to preserve the logic of data analyses and encourage expt-pheno collaboration**
- **Code-wise, a C++ core and Python tools**
  - Fiducial / generator-independence emphasis
  - Integration with HepData
  - Transparent HepMC weight-stream handling
  - 1500+ analyses!
- **Central to a community of analysis reinterpretation tools, linking experiment to theory. Collider, some cosmic-ray**
- **But why? Event loops are trivial…**





Because of this:



We want to avoid physicists all needing to rediscover graph algorithms, conventions, pitfalls, physical/debug distinctions, …

From HZTool to Rivet

- **The idea of preserving experimental analyses for MC validation was born out of HZTOOL**
  - HERA (H1 and ZEUS) DIS and photoproduction
  - Probing low-x, semi-perturbative physics: DIS with Q^2 ~ 4 GeV^2; jet p_T ~ 5 GeV; diffraction
  - Many “state of the art” models only in MCs
  - Much confusion about comparing like-with-like between generators, experiments, and analyses
  - HZTool (Fortran) for cross-experiment comparisons of similar measurements modulo cut differences
- **Direct line to Rivet, 10 years later: “HZ mark two”**

Future Physics at HERA

Workshop, DESY Hamburg, Sept. 95 to Sept. 96

Aim: Study of future physics potentials at HERA in collider and fixed target modes, including high luminosity, polarized beams and nuclei.

Proceedings of the Workshop

Old home page and workshop meetings

Working Groups:

- Structure Functions
- Electroweak Physics
- Beyond the Standard Model
- Heavy Quark Production and Decay
- Jets and High E_T Phenomena
- Diffractive Hard Scattering
- Polarized Protons and Electrons
- Light and Heavy Nuclei in HERA
- HERA Upgrades and Impacts on Experiments

Organizing Committee:
Gunnar Ingelman, Uppsala/DESY (Chairman)
Albert De Roeck, DESY
Robert Klanner, DESY

Secretary:
Ms. H. Haertel
DESY-FH1K
Notkestrasse 85
D-22603 Hamburg
Phone: +49-40-8998-3105
Fax: +49-40-8998-3093

Email: heraws96@mail.desy.de

Advisory Committee:
W.Buchmüller, J.Feltesse, A.Levy,
H.Schröder, J.van den Brand, A.Wagner

If you are using mosaic, click here.

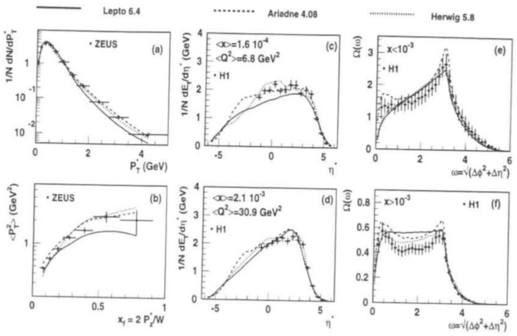
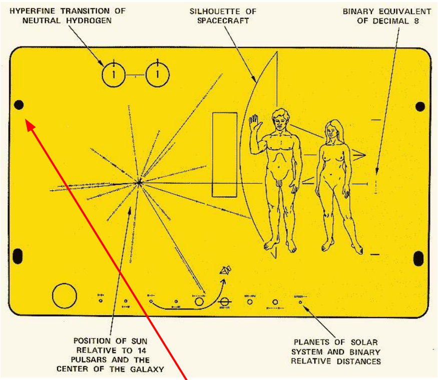


Figure 1: The transverse momenta dN/dp_T (a) and the 'seagull' plot $\langle P_T^2\rangle \times x_F$ (b) of single particles in the positive hemisphere of the hadronic center of mass. The transverse energy flow $dE_T/d\eta$ in a low (c) and high (d) x and Q^2 bin. The transverse energy-energy correlations for $x > 10^{-3}$ (e) and $x < 10^{-3}$ (f).

Lessons learned



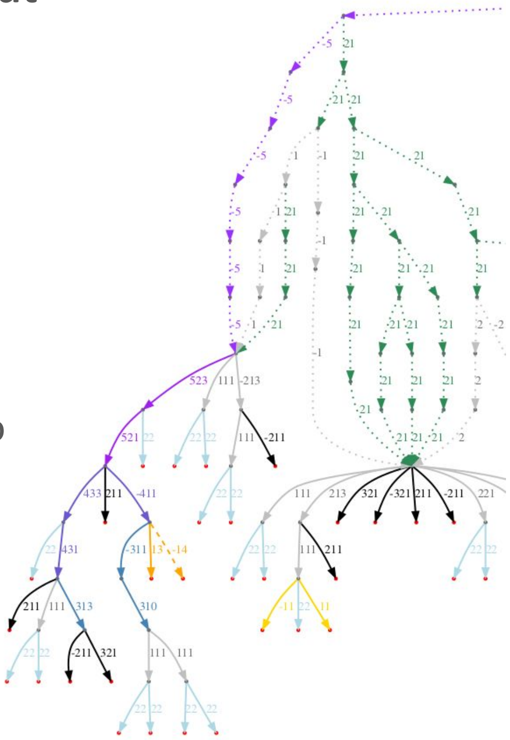
- **A simple/obvious idea, with surprising impact:**
  - **Reproducing a key plot (or not) is *powerful***
    - ⇒ *understand physics, communicate issues, improve MCs*
  - **A *common language* for phenomenology and experiment**
- **But…**
  - "Obvious" to use partons, bosons, etc. direct from the event graph
  - Frequently unphysical, depend on approximations. May not even exist!
  - Scalability of many analyses to new MCs means avoiding gen-dependence
    - **⇒ predict "real" observables, from well-defined final states**

- **Standardisation: boring but important**
  - (physical) event format conventions, statuses, PDG particle numbering, weights…
- **Scalability**
  - Lots of expensive operations are repeated: sharing calculations is essential

Physically safe analysis methods

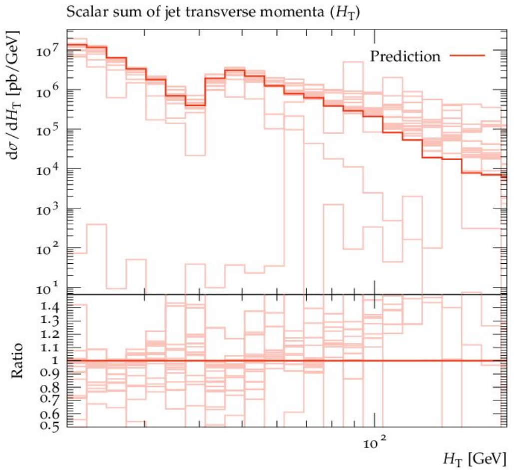
Avoiding unstandardised event-graph features was pragmatic, but led to some physical insights:

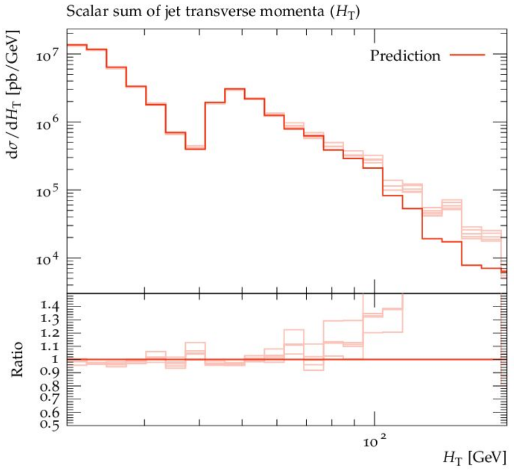
- **[Refining the "fiducial" idea](), defining *unfolding targets***
- **Hadronisation as a "decoherence barrier"**
  use the natural dividing line between the quantum-interfering hard process & semi-classical decays: ~ no tempting partons!
- **Bringing truth tagging closer to reco**
  first releases used *b*-ancestry of jet constituents to set HF labels: too inclusive! ⇒ *associate* the hard-fragmenting, weakly-decaying *B*
- **Promptness/directness tests**
  don't identify a particle "from the hard process"; do it backward. Label as *indirect* via recursive checks for hadron parentage
- **Dressed leptons**
  we now primarily *dress* truth leptons with their photon halo



Multiweights and re-entry

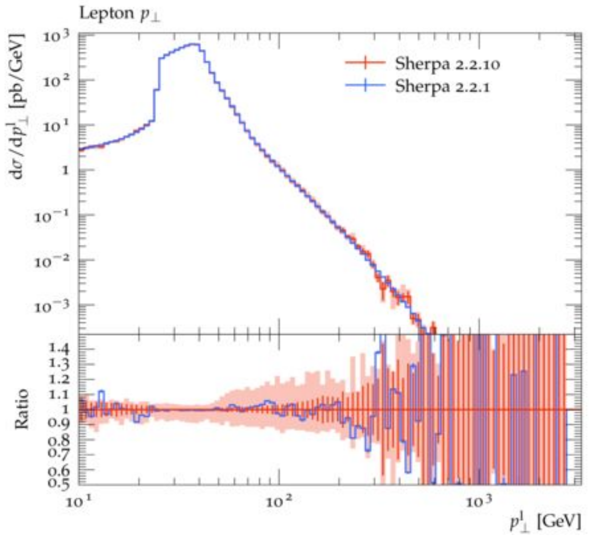
- **MC weight vectors allow expression of increasingly complex theory uncertainties.** But a burden for analysis chains: have to propagate and correctly combine O(200) weight streams!
- **Rivet 3: complex automatic handling of weights**
  ~invisible to users: data objects *look* like histograms etc. but are secretly multiplexed
- **Can now re-call finalisation to combine runs:**
  RAW histogram stage preserves pre-finalize objects
  ⇒ "re-entrant" perfect data-object merging
  Key for e.g. *pA/pp or W/Z* ratios, + BSM recasting
- **Data types are important:** glimpses of a fully coherent separation of semantics from presentation

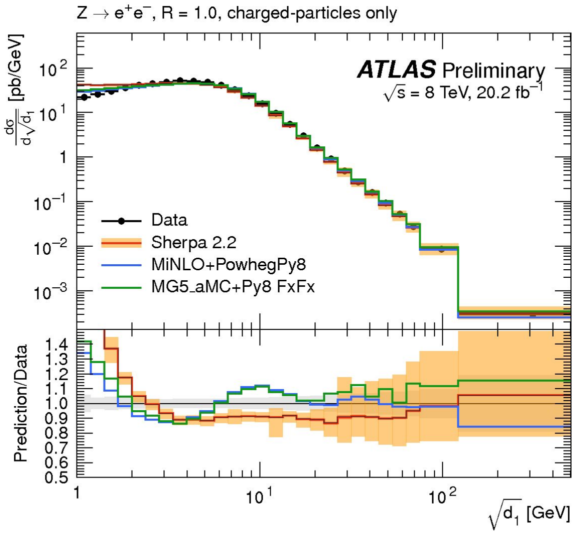


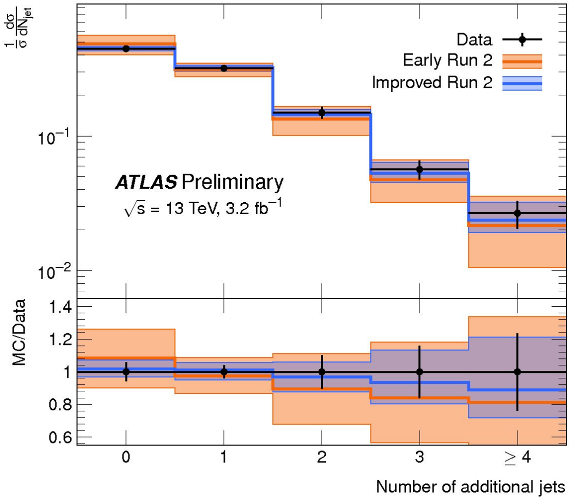


Rivet multiweights in action

ATLAS MC studies have been a significant driver of this feature *(thanks to Chris Gutschow)*





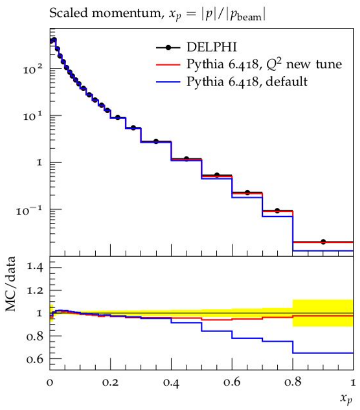


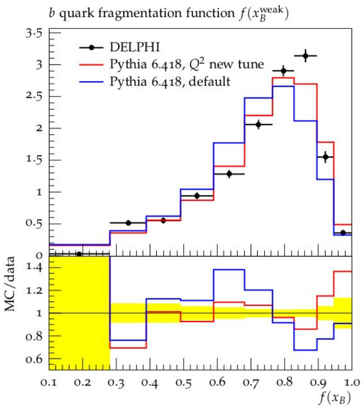
- Weight-naming standardisation underway via MCnet

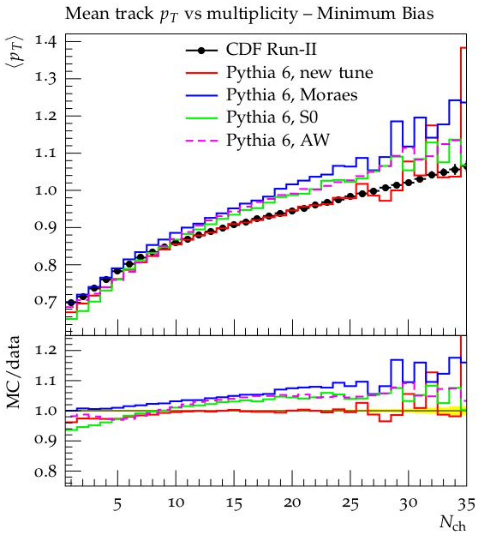
Event generator tuning

Event generators all have dirty secrets. Usually non-perturbative ones… O(30+) parameters

- ❖ **First systematic hadron collider "tunes" of PYTHIA6 by Rick Field for CDF ~ 2001**
  - ➢ Tune A, Tune D, Tune DW, etc. etc.
- ❖ **Limited datasets, variation by hand**
  - ➢ Rivet and its analyses were a game-changer
  - ➢ You only know a model is incapable when you've scanned its whole param space… and then the argument is over
- ❖ **The "Professor" tunes, 2008…**







The Professor method

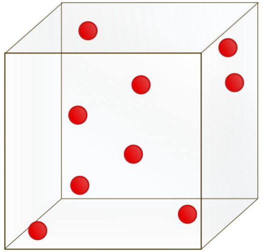
Tuning was historically brute force & inspiration

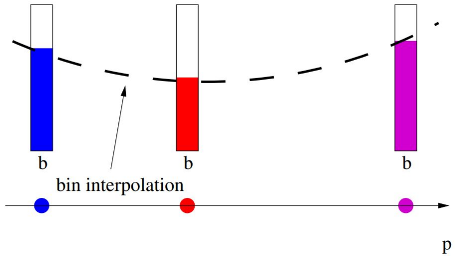
Professor method is an assistant, to aid convergence:

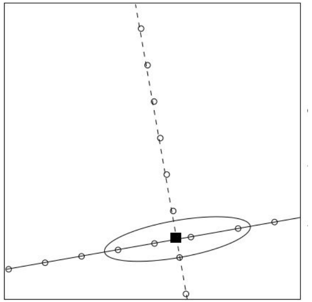
1. Sample (user-)param vectors $\boldsymbol{p}_n$ (from a hypercube/sphere)
2. Generate MC run-sets for beams, processes, etc. at each pt
3. Run *in parallel* on big batch/grid facilities, output histos
4. Build *surrogate models* $\text{bin}_b(\boldsymbol{p})$ from {$\boldsymbol{p}$}, e.g. conventionally a 3rd/4th-order polynomial in $\boldsymbol{p}$. *[Can also interpolate MC errs…]*
5. Use the surrogate models to make a surrogate GoF ⇒ optimize!

Expertise and inspiration still very useful!

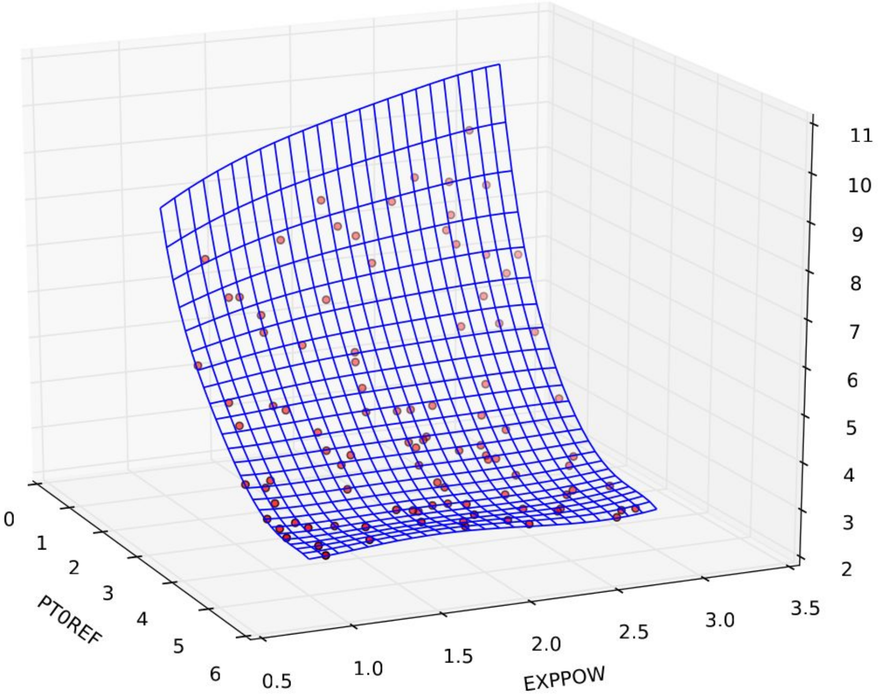
What about machine learning? Sure, fine: easy adaptation.
But if polynomials work — maybe via a change of variables — they are simple and robust

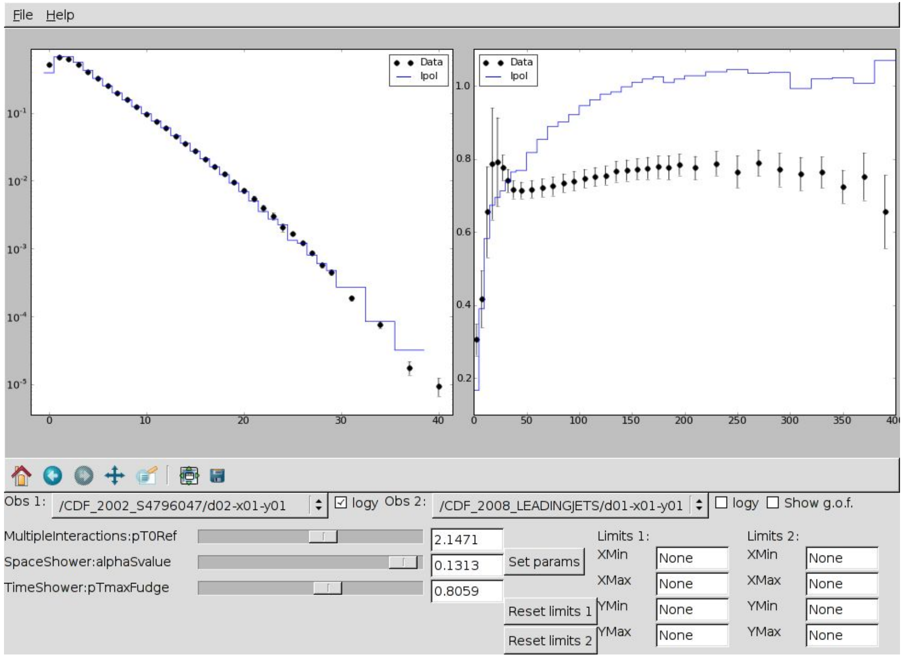




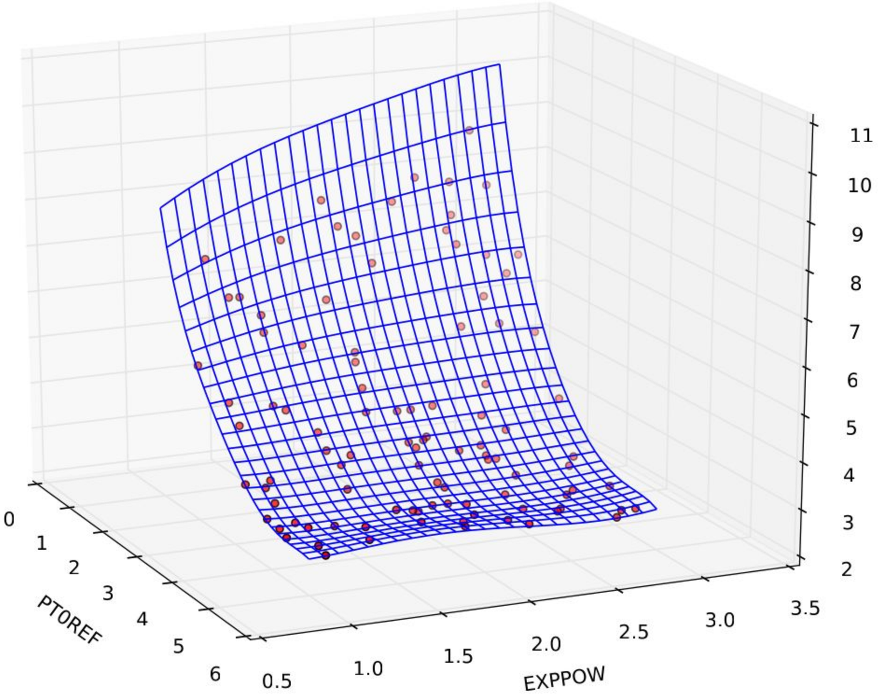


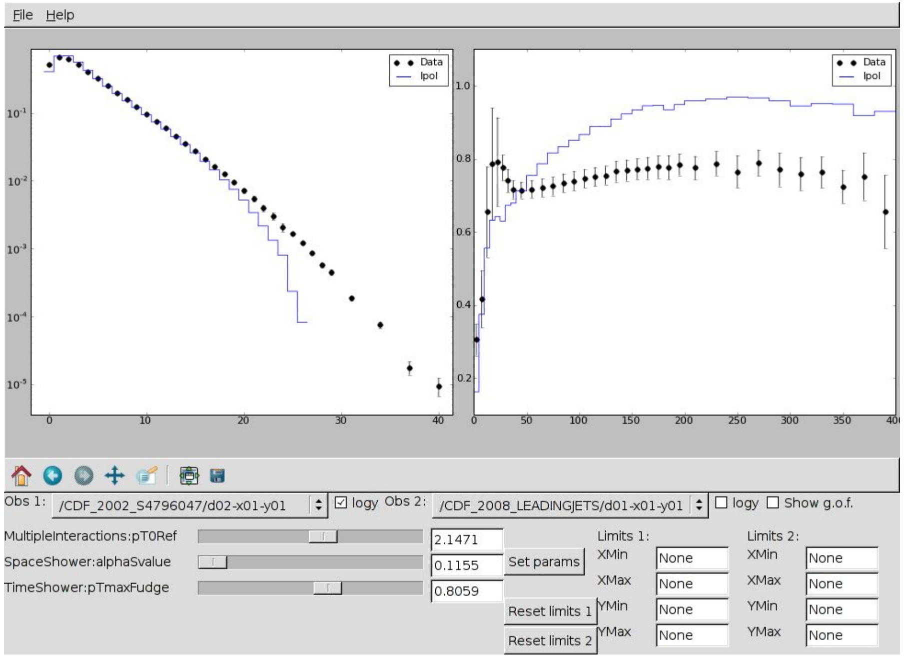
Professor highlights





Professor highlights

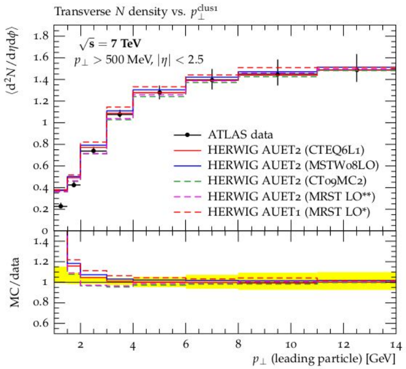


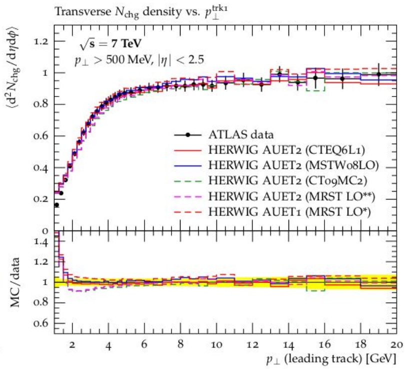


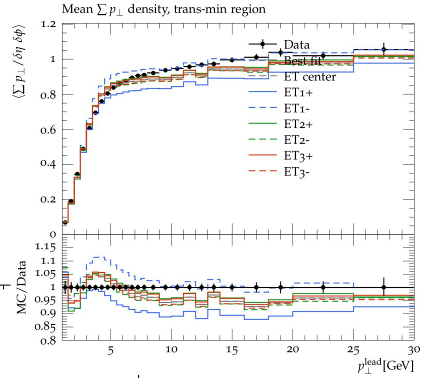
More tuning history

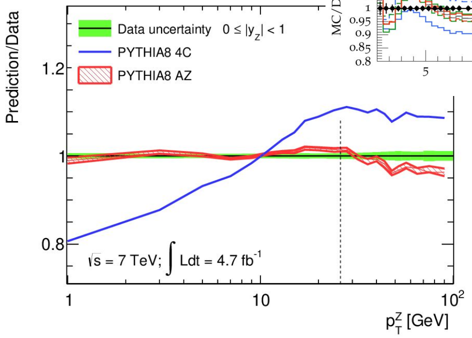
Pre-LHC, the soft QCD uncertainties were *huge*

- ❖ **Factor x 2 uncertainty on 7 TeV σ_{tot}!**
- ❖ **Feed in to underlying event, pile-up, etc.**
  - ➢ Tuning an essential task: better tunes ⇒ better analysis designs, better limits, …
  - ➢ Impact: LEP and Tevatron analyses published for ~10 years suddenly got used! And cited…
  - ➢ ATLAS AMBT, AUET, AZ, A14 etc. tunes + CMS
  - ➢ Rapid responses to preliminary data, changes of model (e.g. Py8 for ATLAS pile-up)
  - ➢ Model development: matching & merging, addition of energy evolution & colour-reconnection to Herwig, …









Tactics for tuning

❖ **Factorise parameter space**

- ➢ Historically split hadron flavours and spectra, jet structure, event topologies, underlying event. Max O(10)
- ➢ Approximate but practical. Can also automate some estimate of factorisation groupings through mutual sensitivities

❖ **Weighting, observable balance, and uncertainties**

- ➢ Tuning naturally involves some data types more than others: balance?
- ➢ Also, models not capable for fully describing all data bins: check envelopes, sensitivities, limit ranges… and *weight bins*
- ➢ Custom goodness-of-fit function? Regularise, lose statistical interpretation?
- ➢ “chi2” already does not behave classically: *eigentunes*, room for improvement

❖ **Future work**

- ➢ Heavy flavour, matching/merging, including systematics via weights…

Heavy ions in Rivet

❖ **“Adding heavy-ion support” sounds trivial!**

❖ Actually a stern test, with far-reaching impacts.

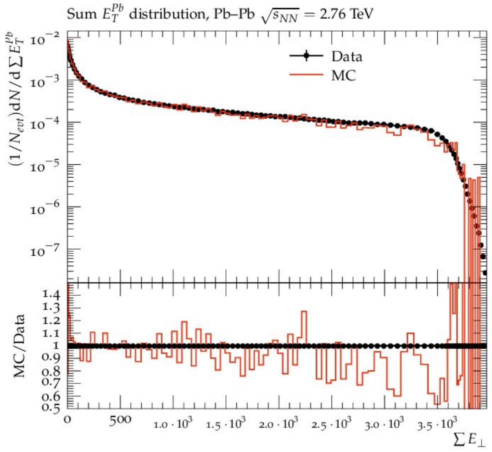
  ➢ HI observables often require centrality calibration curves: we need a 2-pass run. That wasn’t planned

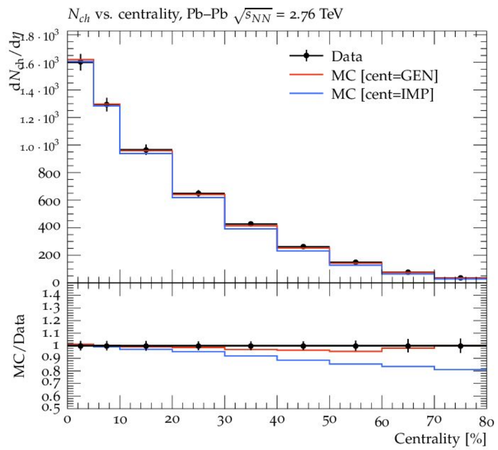
  ➢ And event/event correlations… centrality-binned!

  ➢ Need swappable definitions: few HI generators are general-purpose enough to do
  e.g. both forward E_T *and* jet quenching

❖ Paper: https://arxiv.org/abs/2001.10737

❖ *HI MC standards are also in flux: having a common tool enables discussion on common standards*

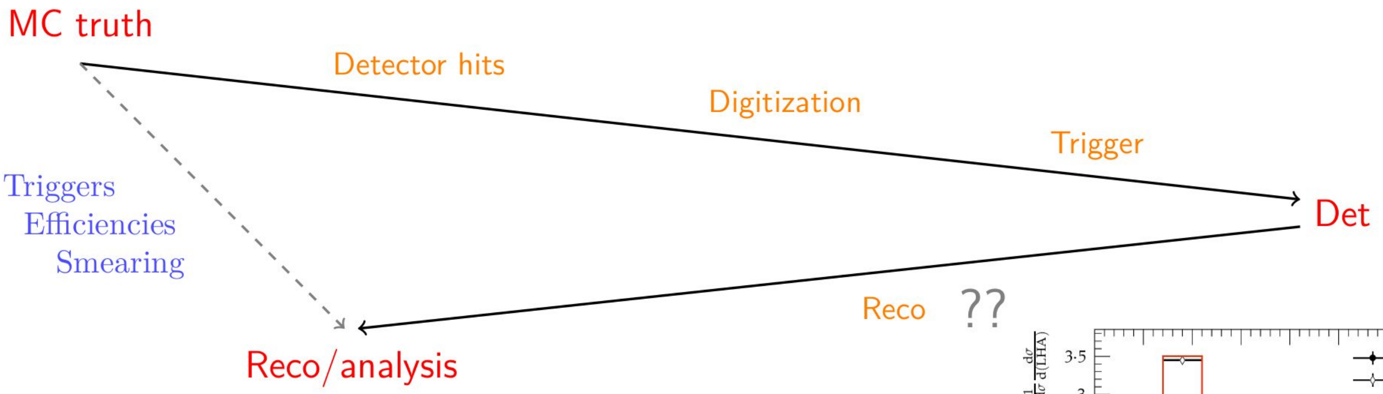




BSM searches: preserving detector+reco

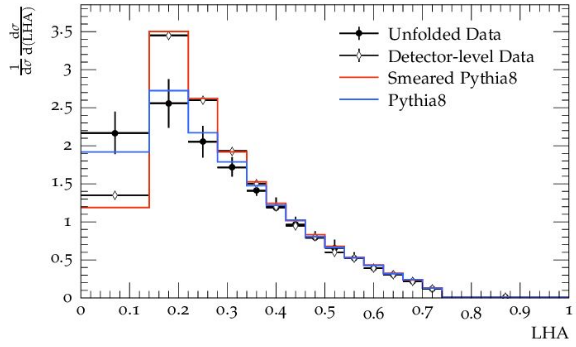
❖ **Detector smearing built on Rivet's projection system — for reco-level analyses**

➢ developed based on Gambit ColliderBit experience: no need for "full fast-sim"



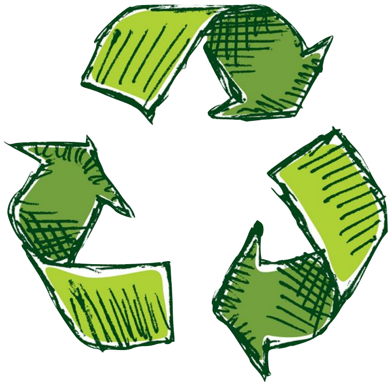
➢ like Delphes, but more flexible & can be *analysis-specific* ⇒ MA5 "SFS" mode

➢ flexibility allows e.g. "tuned" jet-substructure smearing



Summary

- Rivet arose from HERA experiment/MC author collaboration. Other packages with different emphases
- Lightweight analysis preservation has spurred many other experiment/pheno activities, e.g.
  - MC development
  - Tuning
  - PDF studies, EFT studies, global BSM fits…
  - Heavy-ion methods
  - And teaching / UG projects
- “Rough” tuning was important in the early LHC era
  - Arguments to restart in targetted configurations
  - Understand data / model gaps
  - Fairly reduce and estimate model systematics
- Preservation is an accelerator for analysis impact: experiment-theory studies, fun collaborations!



Bonus: Professor tuning tutorial

from HSF tuning workshop 2023

Sampling and generating

purple = command shell

- **Working environment: Docker**
  - $ docker pull hepstore/prof2-tutorial
  - Run the container interactively, with cleanup and a mapped-in dir:
    $ docker run -it --rm -v $PWD:/host hepstore/prof2-tutorial
  - Optional! ⇒ # apt-get update && apt-get install vim (or emacs-nox)
  - Note: for now these Dockers are amd64 architecture: slow on M1 Macs, will be improved asap

- **Producing the inputs factorises from the tuning**
  - Image built on Rivet+Pythia 8.3. Docs: https://pythia.org/latest-manual/Welcome.html
  - Can't assume a particular generator, batch-farm interface, etc. ⇒ sample with script+cfg templating
  - Look in the tmpl/ directory: template files for MB Py8+Rivet jobs

  - # nano tmpl/mbrun.sh ⇒ reduce number of events if you want a quick local run
  - # prof2-sample -t tmpl/mbrun.sh -t tmpl/py8mb.cmnd -n 20 tmpl/paramranges.dat
  - # ls scan/*
  - Check the contents, values, etc. — is it clear what's going on?

  - And run, e.g.: # for i in 000*/mbrun.sh; do nice -5 bash $i & done
  - It's quite plausible to generate small samples like this on a laptop! But multiple processes, multiple energies, different cuts, and in particular *far* more expensive matrix elements -> cluster/Grid/HPC

Inspecting, interpolating, tuning

- **Once the run is finished, you can inspect the outputs written into each run dir**
  - A good idea to clean out unnecessary data:
    # for i in 001*; do (cd $i; yoda2yoda -M "/RAW.*" mb.yoda tmp.yoda; mv tmp.yoda mb.yoda); done
  - # cd /work && prof2-envelopes -d /usr/local/share/Rivet/ scan/ or use the pre-prepared mc/ dir now
  - Copy output back to the host to view: # cp -r envelopes /host/
  - Similar with Rivet plotting: # rivet-mkhtml-mpl scan/001*/mb.yoda -o /host/rivet-plots

- **Might as well immediately build a surrogate interpolation ("ipol")**
  - # prof2-ipol -h
  - # prof2-ipol mc/ Unfortunately, prof2-residuals is currently broken…
  - Use the ipol-listing tool to generate a starter weights file:
  - # prof2-ls -w ipol.dat > weights0.dat
  - # cp weights{0,1}.dat && nano weights1.dat
  - Edit to cut out bad bins, tweak the fit toward things you care about… this is the creative bit!

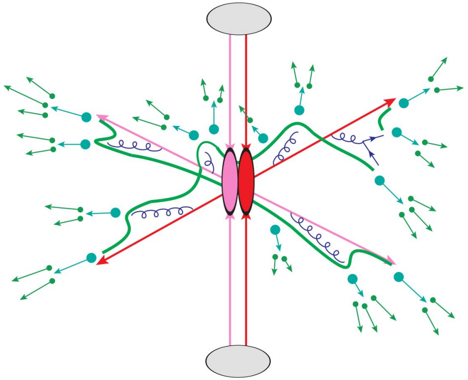
- **And… tune!**
  - # prof2-tune -d $(rivet-config --datadir) -w weights1.dat
  - Plot the output: # rivet-mkhtml-mpl tunes/ipolhistos.yoda -o /host/rivet-plots-tune1
  - And iterate! Unfortunately the eigentunes script needs a fix, so no demo: dev help is welcome!!

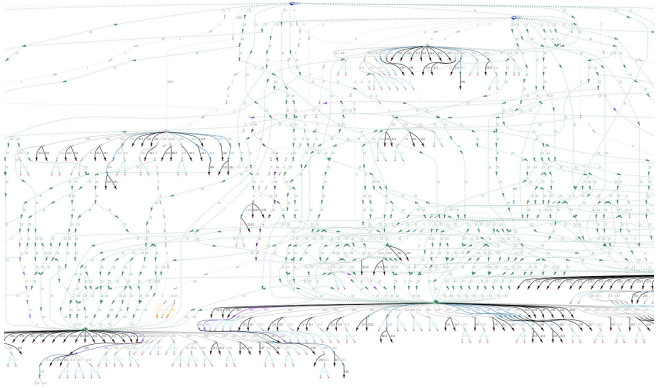
Thanks for coming!

Backup slides

MC generation

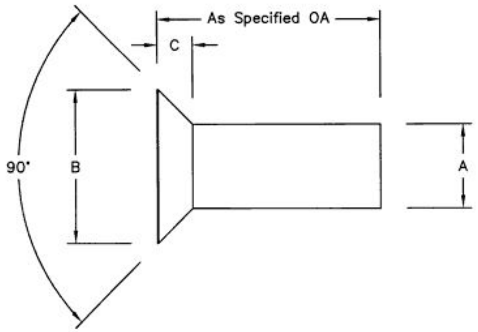
- **MC generation is where theory meets experiment**
  - The fundamental *pp, pA, AA* collision, *sans* detector
- **Components of an "exclusive" event-generator chain:**
  - QFT matrix element sampling at fixed-order in QCD
  - *Dressed* with approximate collinear splitting functions, iterated in factorised Markov-chain "parton showers"
  - FS parton evolution terminated at Q ~ 1 GeV: phenomenological hadronisation modelling
  - Mixed with multiple partonic interaction modelling
  - Finally particle decays, and other niceties
- **Modern HEP is hostage to shower MCs!**
  - The main mechanism for translating theory to experimental signatures, from QCD to BSM
  - Generally very complex modelling and output





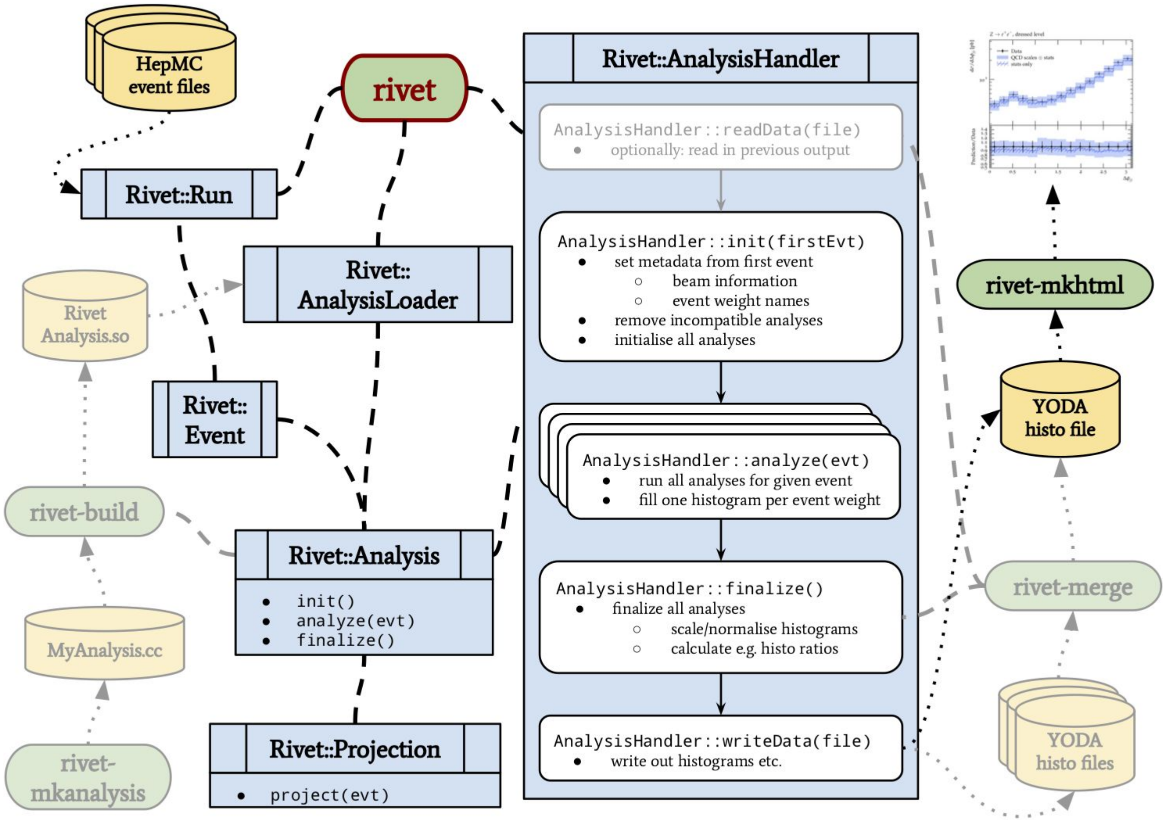
Designing Rivet

- **Ease of use**
  - **Big emphasis on "more physics, less noise"!**
  - Minimal boilerplate analysis code, HepData sync
  - Event loop and histogramming basically familiar
  - Tools to avoid having to touch the raw event graph
- **Embeddable**
  - OO C++ library, Python wrapper, sane user scripts
  - Generator independence: communication via HepMC
    - Note HepMC3 HI-support efforts
  - Analysis routines factorised: loaded as "plugins"
- **Efficient**
  - Avoid recomputations via "projection" caching system
- **Physical**
  - **Measurements primarily from final-state particles only**



The result

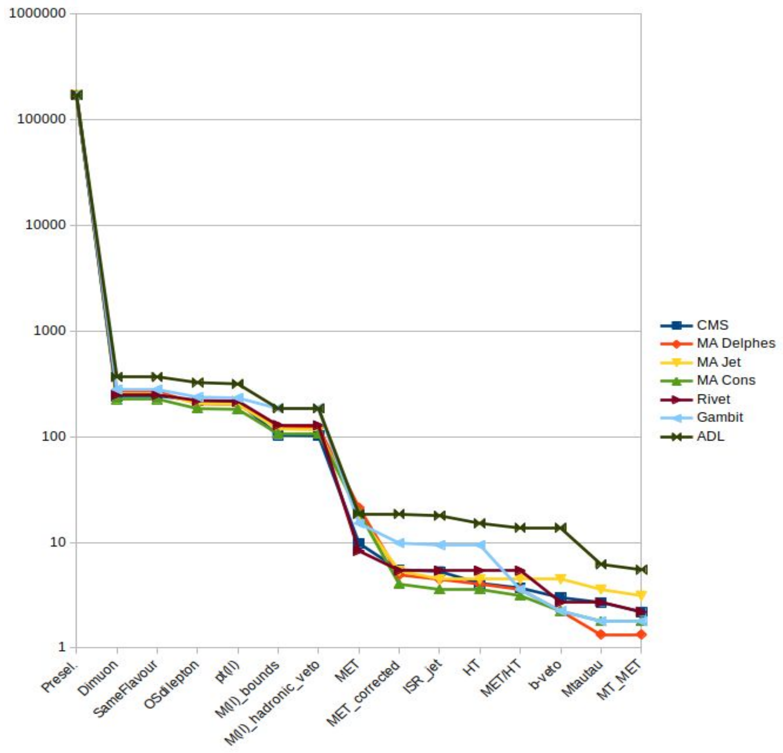
- **As of Rivet v3.1.0**
  arXiv:1912.05451
- Streamlined set of tools from analysis coding to event processing to plotting (and other applications)
- And a key gateway to connect your analysis to theory (and back again)
- Let's review some of the early impacts…



Rivet and BSM-search recasting

- **Rivet's main emphasis *isn't* BSM direct searches, but there's no reason not to**
  - lots of experiment experience and support
  - efficient scaling-up to hundreds of analyses, with distinct phase-space specific detector/efficiency functions
  - can we do for BSM preservation what we did for measurement analyses?

- **Friendly competition, mainly from/with MA5**
  - all good tools, all geared to getting your analysis into pheno studies asap
  - but ours is best, obv… ;-)



Les Houches 2019 CMS soft-lepton recasting-tools comparison

The future of Rivet

- **Vision: Rivet as a standard for "truth-level" observables, across collider physics**
- Not just standalone, but as a library in pheno & experiment frameworks, too:
  standard MC definitions (cf. CMS), seamless systematics handling, etc.
- At its core: a physics-oriented system for physicists to **compare MC predictions to one another and to data, on many simultaneous observables, in myriad ways**
  We don't know all the use-cases yet!
- **Challenges:**
  - Extension of HepData and other community infrastructure for ever more precise data. Even our compressed data format is struggling with the volume of analyses and data.
    Work needed on multiweight-oriented data format and tools
  - Improved, modernised visualisation and exploration
  - Connections to global (BSM) fitting tools
  - Preserving MVAs: BDT and NN in vanilla C++