

Accelerating physics impact

Analysis prototyping, preservation
& reinterpretation

Christian Gütschow

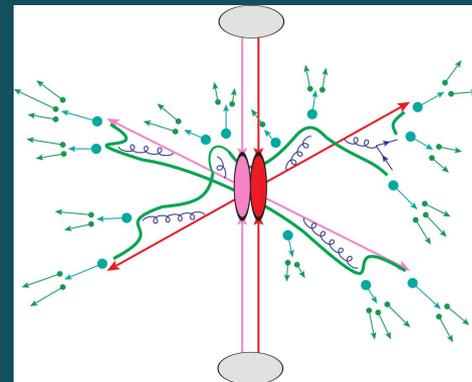
(borrowing heavily from Andy Buckley)

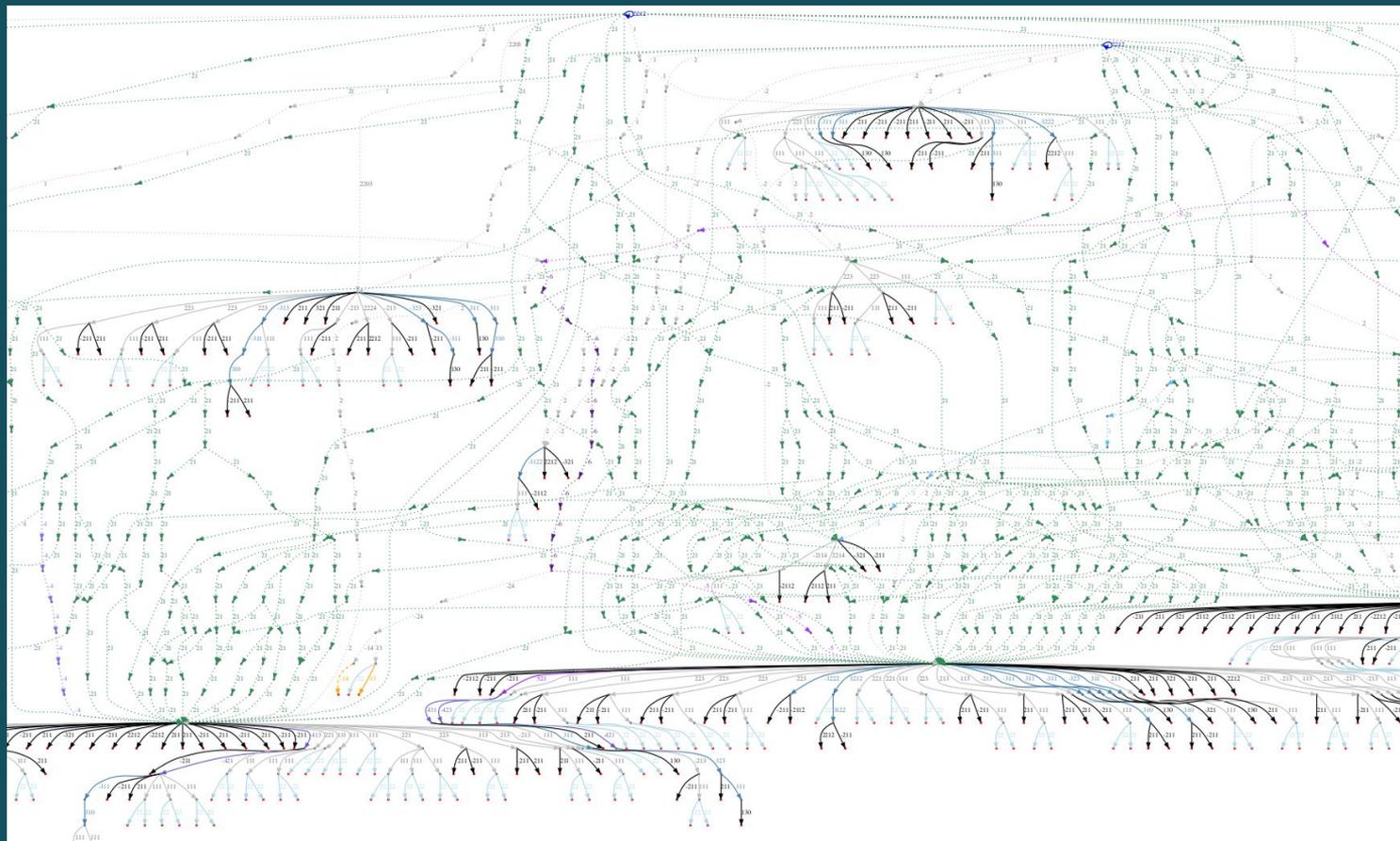
MCnet School, Zakopane

June 2022

What is Rivet?

- The “LHC standard” MC analysis toolkit
- **More broadly a project to preserve the logic of HEP data analyses and further expt-pheno collaboration**
- Code wise, a C++ core and Python tools
 - Fiducial / **generator-independence** emphasis
 - Integration with **HepData**
 - Transparent **weight-stream handling**
 - **1000+ analyses!**
- **Central to a community of analysis reinterpretation tools, linking experiment to theory**
- **But why?**

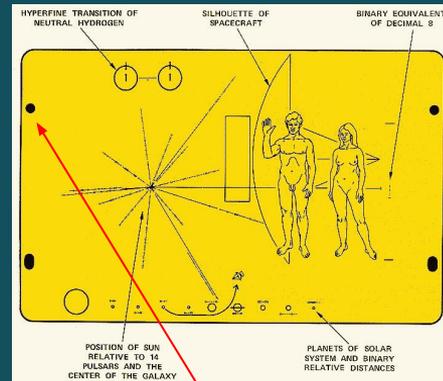




We want to avoid physicists needing to repeatedly rediscover graph algorithms, conventions, pitfalls, physical/debug distinctions, ...

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 pheno and experiment
- **But...**
 - “Obvious” to use partons, bosons, etc. direct from the event graph
 - Frequently unphysical & depend on approximations. May not even exist!
⇒ **predict “real” observables, from well-defined final states**
- **Standardisation:** boring but important
 - (physical) event format conventions, status codes, PDG particle numbering, weights ...
- **Scalability**
 - Lots of expensive operations are repeated: sharing calculations is essential



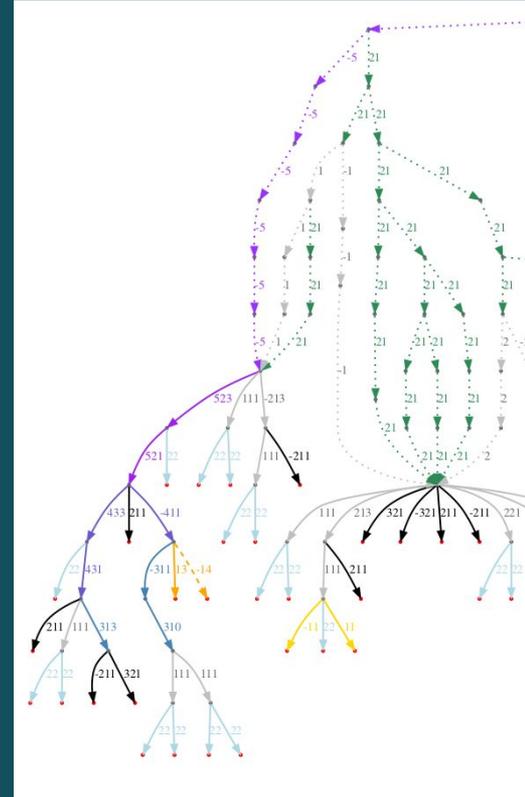
Common pitfalls

- **Colour triplets are not final-state particles**
 - MC event generators do not guarantee the physicality of a "final state top"
- **Electroweak scale particles (W, Z, H) are not final state particles**
 - Focus on the leptons and hadrons for the decay channel you care about
- **Hidden vetoes**
 - all important cuts should be reflected in the fiducial cross section definition
 - e.g. a veto on isolated photons in a dilepton analysis may make no difference to the result when running on a SM sample which is LO in the electroweak coupling, but what happens if more precise calculation is used which may include EW radiation?
- **Missing energy and neutrinos**
 - explicit use of neutrino flavour and momentum is very problematic, especially when there's more than one neutrino in the event
 - better to use the particle-level missing transverse momentum instead, which correctly accounts for possible additional (BSM or other) sources of missing momentum

Physically safe analysis methods

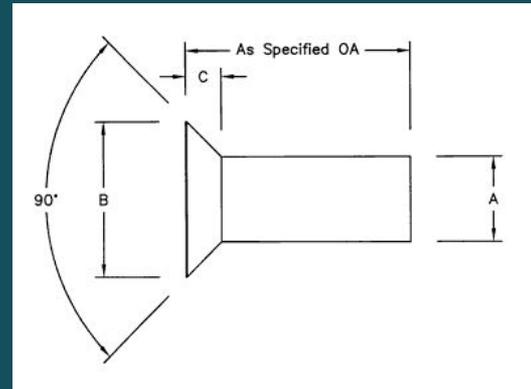
Avoiding unstandardised event-graph features was pragmatic, but led to some genuine physical insights:

- **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



Designing the 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
 - Analysis routines factorised, and loaded as “plugins”
- **Efficient**
 - Avoid recomputations via “projection” caching system
- **Physical**
 - **Measurements primarily from final-state particles only**

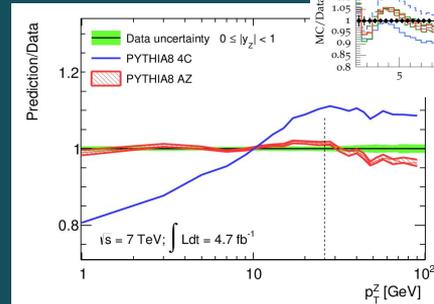
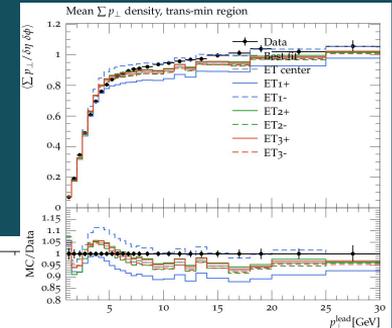
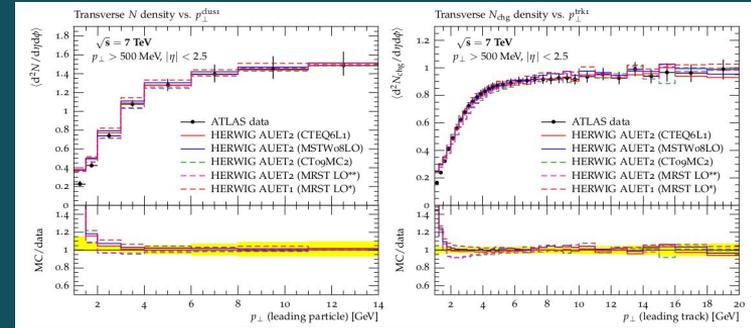




More tuning...

It's getting hard to remember now, but pre-LHC the soft QCD uncertainties were *huge*

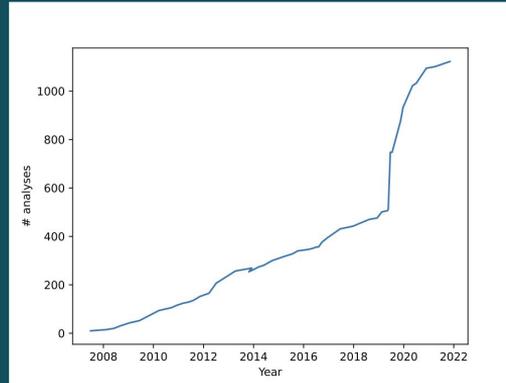
- Factor x2 uncertainty on 7 TeV σ_{tot} !
- Feed in to underlying event, pile-up, etc.
 - Tuning an essential task: better tunes \Rightarrow better analysis designs, better limits, ...
 - Impact: LEP and Tevatron analyses published for ~ 10 years suddenly got used! And cited...
 - ATLAS AMBT, AUET, 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, ...



The state we're in

- **Version 3.1.0 crossed the 1000 analysis mark**

A steady flow of analysis submissions, plus the occasional deluge of (mainly hadronisation) routines from Herwig!



- **Official support from the LHC experiments is crucial**

preservation = just part of how we do science; but still some way to go! Coverage monitoring:

- **“New” features since the v1 vision:** systematics multiweights, “perfect merging”, heavy ions, detector smearing functions, analysis options

Rivet analysis coverage (no searches, no heavy ion)

Rivet analyses exist for 1003/4078 papers = 25%. 141 priority analyses required.

Total number of Inspire papers scanned = 6382, at 2022-05-13

Breakdown by identified experiment (in development):

Key	ALICE	ATLAS	CMS	LHCb	Forward	HERA	e ⁺ e ⁻ (≥ 12 GeV)	e ⁺ e ⁻ (≤ 12 GeV)
Rivet wanted (total):	82	104	134	218	16	424	682	507
Rivet REALLY wanted:	17	33	64	8	0	12	1	0
Rivet provided:	14/96 = 15%	159/263 = 60%	94/228 = 41%	17/235 = 7%	8/24 = 33%	34/458 = 7%	183/865 = 21%	323/830 = 39%

Show greylist

Show blacklist

ALICE **ATLAS** CMS LHCb Forward HERA e⁺e⁻ (≥ 12 GeV) e⁺e⁻ (≤ 12 GeV) Tevatron RHIC

ATLAS: Measurement of the energy asymmetry in $t\bar{t}$ production at 13 TeV with the ATLAS experiment and interpretation

Inspire: 1941095 · arXiv: 2110.05453 (pdf) · CDS: 2783829 · HepData: ins1941095 · Report IDs: CERN-EP-2021-181

ATLAS_2021_11941095

ATLAS: Measurement of b -quark fragmentation properties in jets using the decay $B^{\pm} \rightarrow J/\psi K^{\pm}$ in pp collisions at $\sqrt{s} = 13$ TeV

Inspire: 1913061 · arXiv: 2108.11650 (pdf) · DOI/journal: 10.1007/JHEP12(2021)131 · CDS: 2779272 · HepData: ins1913061 · Report IDs: CERN-EP-2021-123

ATLAS_2021_11913061

ATLAS: Measurement of the production cross section of pairs of isolated photons in pp collisions at 13 TeV with the ATLAS experiment

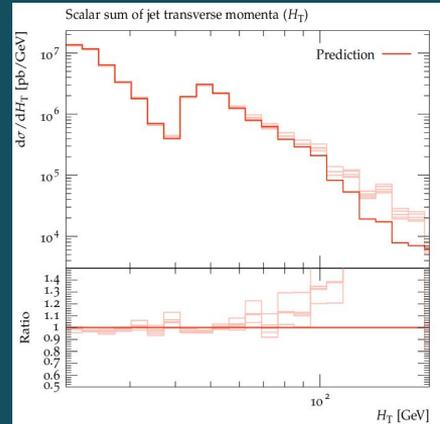
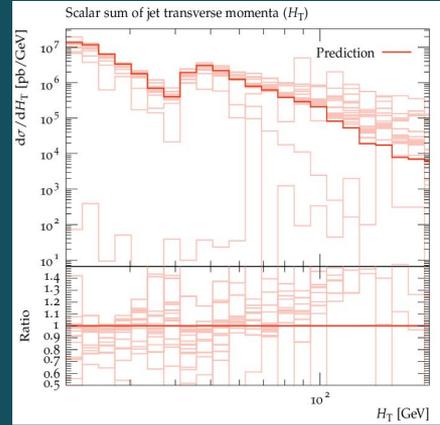
Inspire: 1887997 · arXiv: 2107.09330 (pdf) · DOI/journal: 10.1007/JHEP11(2021)169 · CDS: 2776245 · HepData: ins1887997 · Report IDs: CERN-EP-2021-105

ATLAS_2021_11887997

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

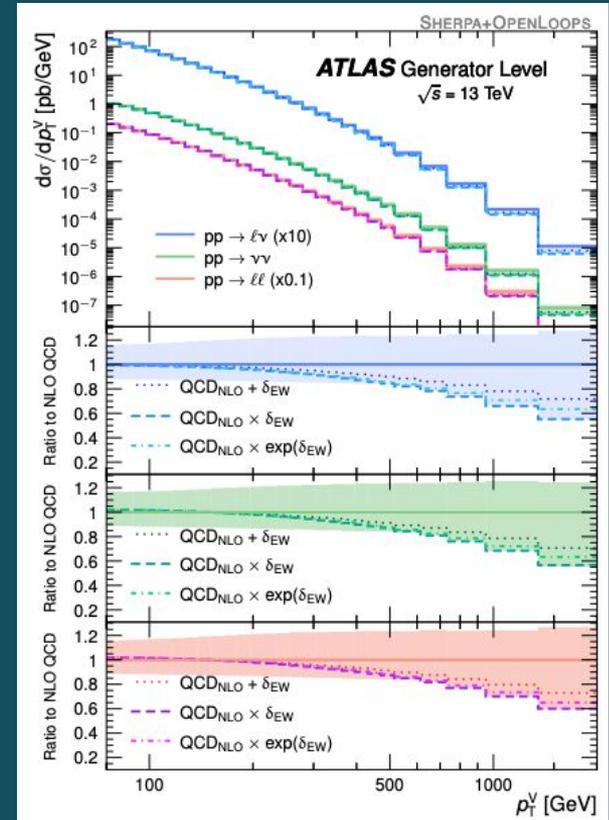
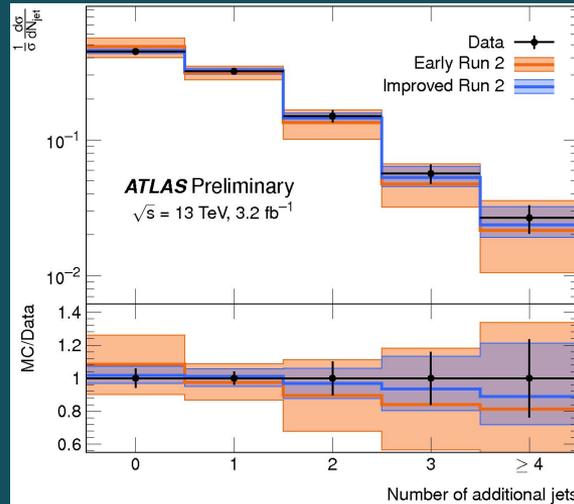
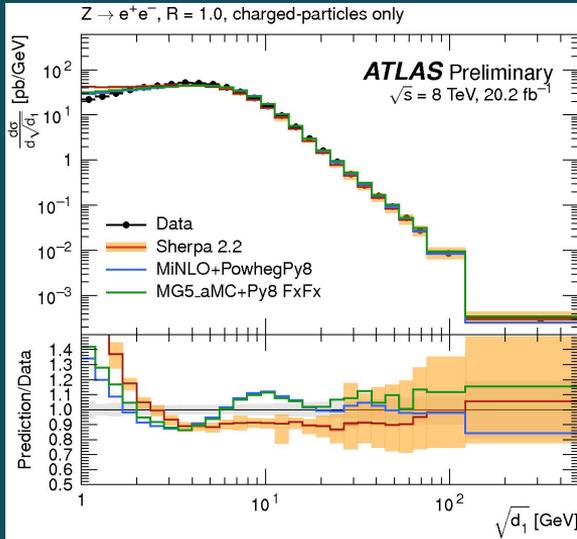
Multiweights and re-entry

- **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

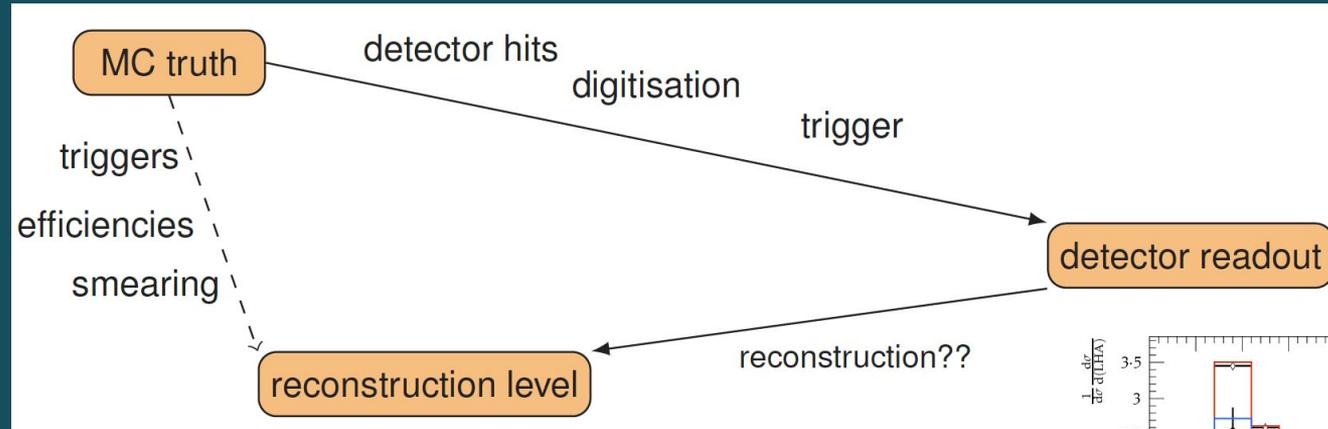
ATLAS MC studies have been a significant driver of this feature



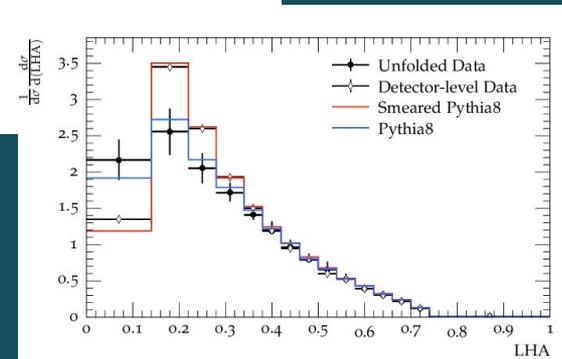
Weight-naming standardisation via MCnet ([arXiv:2203.08230](https://arxiv.org/abs/2203.08230))

Detector emulation

- **Detector smearing built on Rivet’s projection system — for reco-level analyses**
 - developed based on Gambit ColliderBit experience: no need for “full fast-sim”



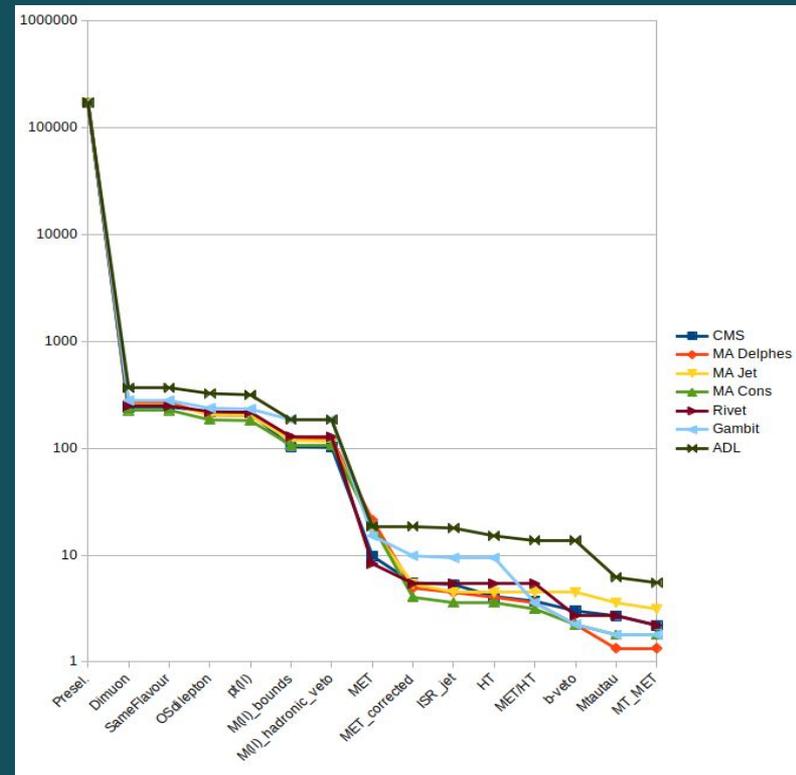
- like Delphes, but more flexible & can be analysis-specific \Rightarrow MA5 “SFS” mode
- flexibility allows e.g. “tuned” jet-substructure smearing, systematics studies, ...





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



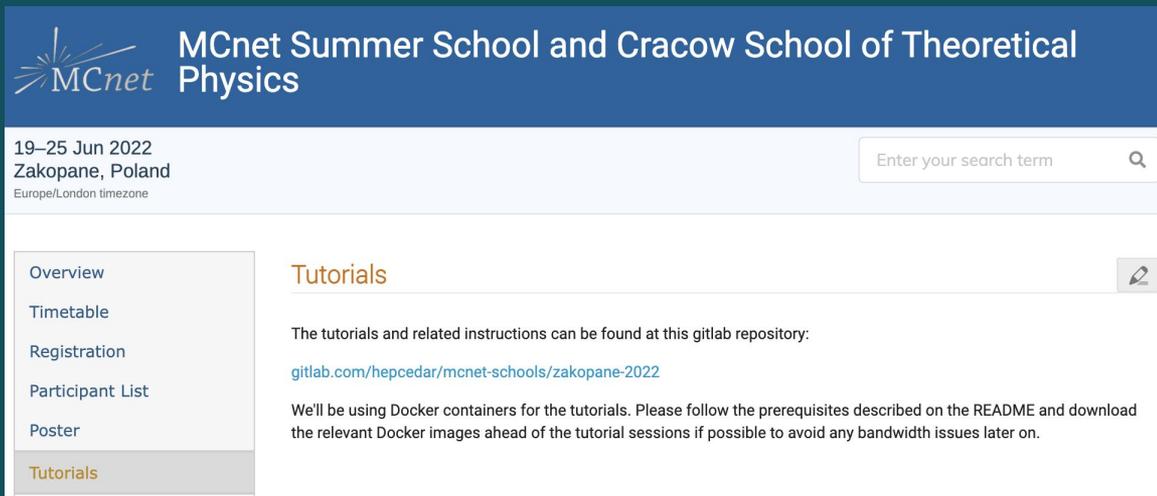
Getting and using Rivet

An analysis that's immediately available to the pheno community is 10x more useful \Rightarrow payback!
In the past, key analyses were ignored due to the barrier to entry

As either a “user” or analysis author,
the barrier is lower than ever:
we recommend using our
[Docker images](#) to get started

Tutorials available from the
[Rivet website](#), a walkthrough in
the [Rivet3 paper](#) or this
MCnet school! :)

[Link to the gitlab repo on Indico!](#)



The screenshot shows the website for the MCnet Summer School and Cracow School of Theoretical Physics. The header is blue with the MCnet logo and the text "MCnet Summer School and Cracow School of Theoretical Physics". Below the header, the dates "19–25 Jun 2022" and location "Zakopane, Poland" are listed, along with the note "Europe/London timezone". A search bar is present on the right. A navigation menu on the left includes "Overview", "Timetable", "Registration", "Participant List", "Poster", and "Tutorials" (which is highlighted). The main content area is titled "Tutorials" and contains the text: "The tutorials and related instructions can be found at this gitlab repository: gitlab.com/hepcedar/mcnet-schools/zakopane-2022". Below this, it states: "We'll be using Docker containers for the tutorials. Please follow the prerequisites described on the README and download the relevant Docker images ahead of the tutorial sessions if possible to avoid any bandwidth issues later on."

Rivet quick start

Rivet can be used as a library (e.g. in a generator interface) or directly from the command line:

```
rivet -a MC_JETS -o output.yoda.gz events.hepmc
```

Produces a YODA file which is ASCII based - open your favourite editor and look at your histograms!

Compile a custom routine using: `rivet-build RivetMY_ANALYSIS.so MY_ANALYSIS.cc`

Produce a web album with histogram comparisons:

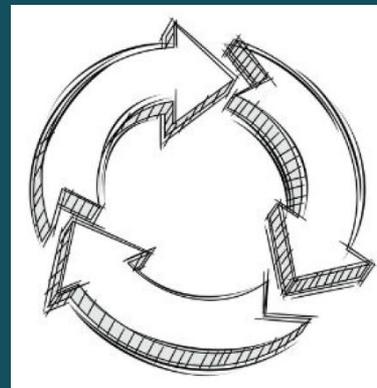
```
rivet-mkhtml --errs -o my_album prediction1.yoda.gz prediction2.yoda.gz
```

All Rivet helper scripts come with a `--help` option for all Rivet scripts!

**Self-guided Rivet tutorial on gitlab with more in-depth discussion
and some physics examples to play with!**

Summary

- Rivet arose from HERA experiment/MC author collaboration, in time for the LHC
- Like HZTool, its existence has spurred many other experiment/pheno activities, e.g.
 - MC development, validation and quality control
 - Tuning
 - PDF studies, EFT, global BSM fits...
 - Heavy-ion methods
 - And teaching / UG projects
- An accelerator for analysis impact: immediate entry to many theory studies. Lots of fun collaborations!
- As we head into another LHC era, there will surely be more use-cases for analysis recycling. **Join in!**



Tutorial schedule

- Monday:
 - Getting set up with Docker
 - Quick intro to Rivet
 - Getting started with the generators
- Tuesday:
 - Generator tutorials
 - Common project: W mass
- Thursday:
 - Parallel options: Contur, Madgraph
- Friday:
 - Machine learning in HEP
 - Discussion of common project



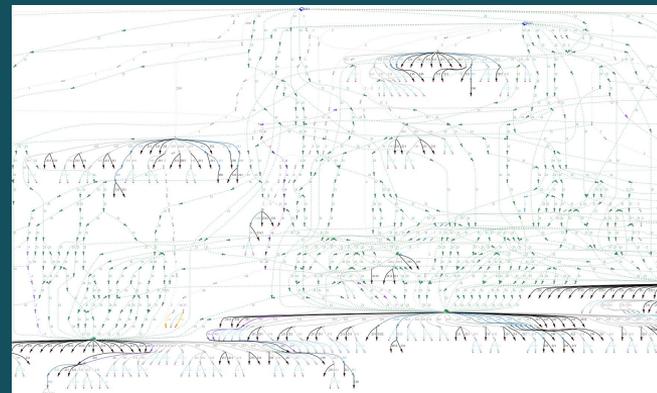
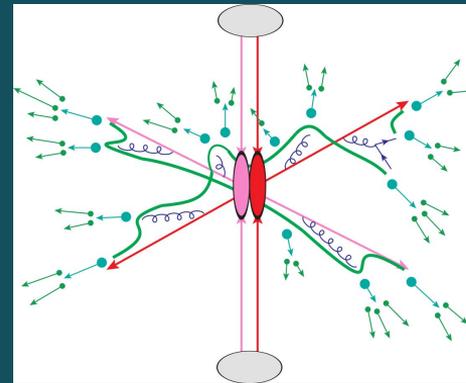
Backup

MC generation

- MC generation is where theory meets experiment
 - The fundamental pp (etc.) collision, sans detector

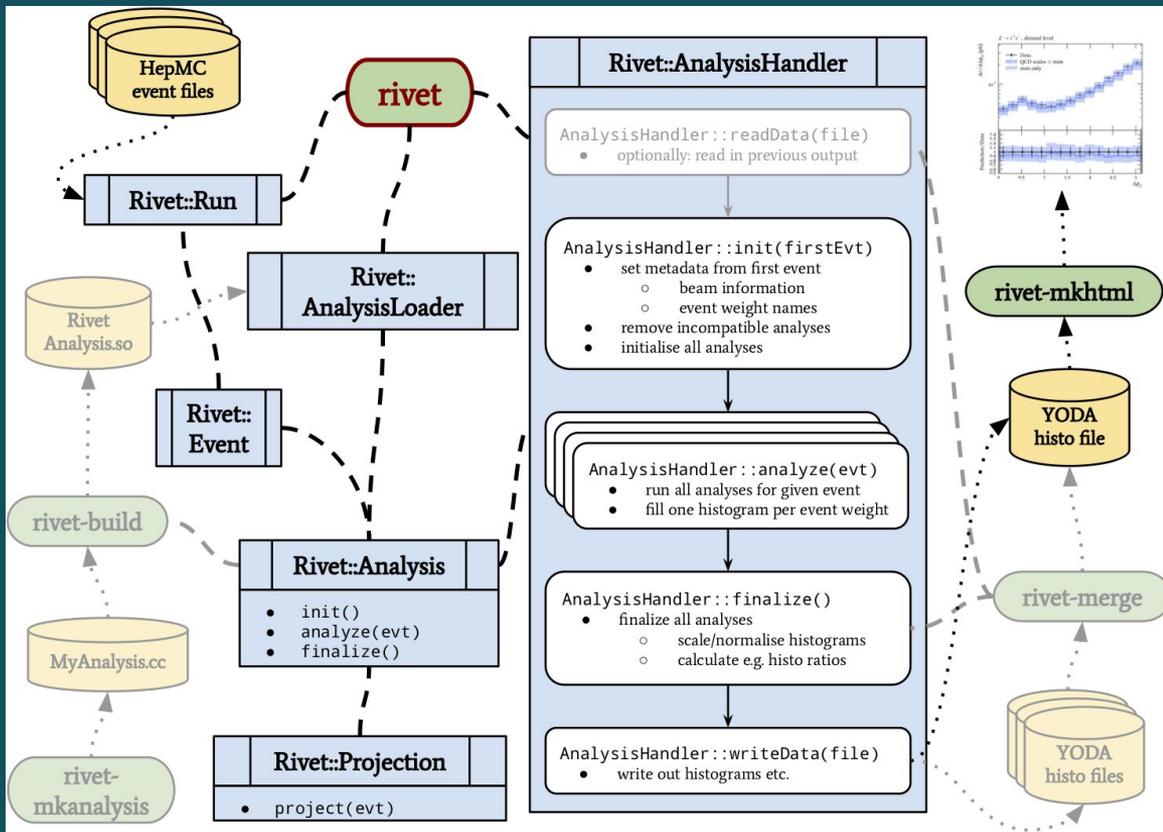
- Components of an “exclusive” event-generator chain:
 - QFT matrix element sampling at fixed order in QCD etc.
 - Dressed with approximate collinear splitting functions, iterated in factorised Markov-chain “parton showers”
 - FS parton evolution terminated at $Q \sim 1$ GeV: phenomenological hadronisation modelling
 - Mixed with multiple partonic interaction modelling
 - Finally particle decays, and other niceties

- Modern HEP is powered by shower MCs
 - The main mechanism for translating theory to experimental signatures, from QCD to BSM
 - Generally very complex modelling and output



The result

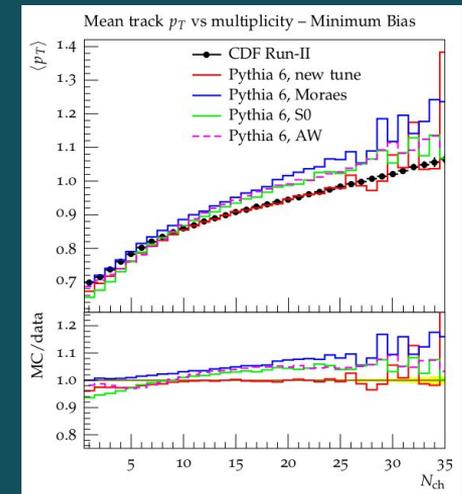
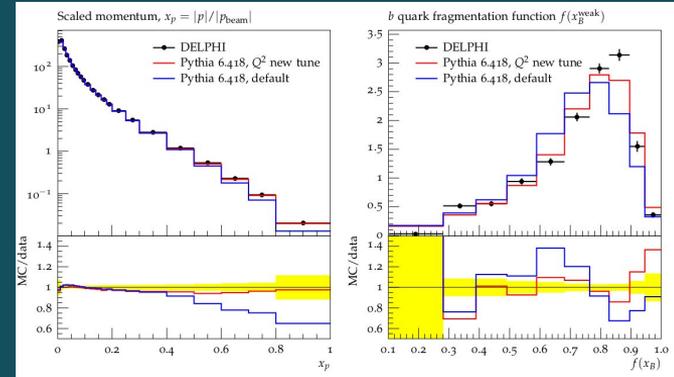
- As of Rivet v3.1.0 [arXiv:1912.05451](https://arxiv.org/abs/1912.05451)
- Streamlined set of tools from analysis coding to event processing to plotting (and other applications)
- And a key gateway to connect your analysis to theory (and back again)
 - integral part of MC validation on the experiments
- Let's review some of the early impacts...



Event generator tuning

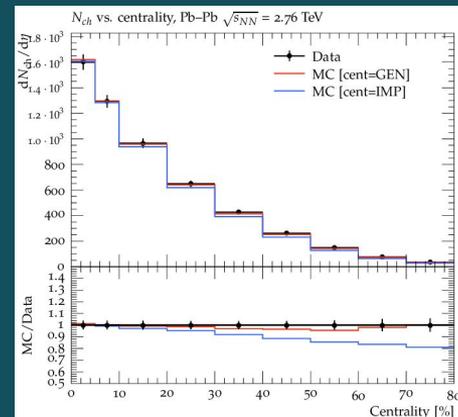
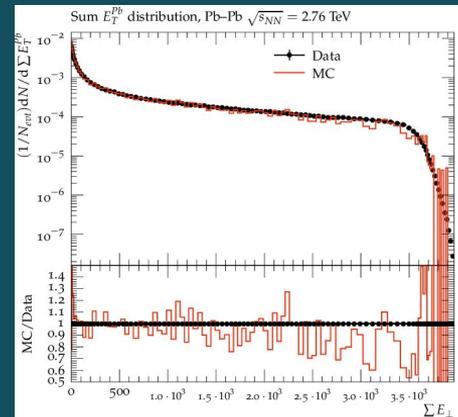
Event generators all have dirty secrets.
Usually non-perturbative ones... $O(30+)$ parameters

- 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 parameter space... and then the argument is over
- The “Professor” tunes, 2008; and ...



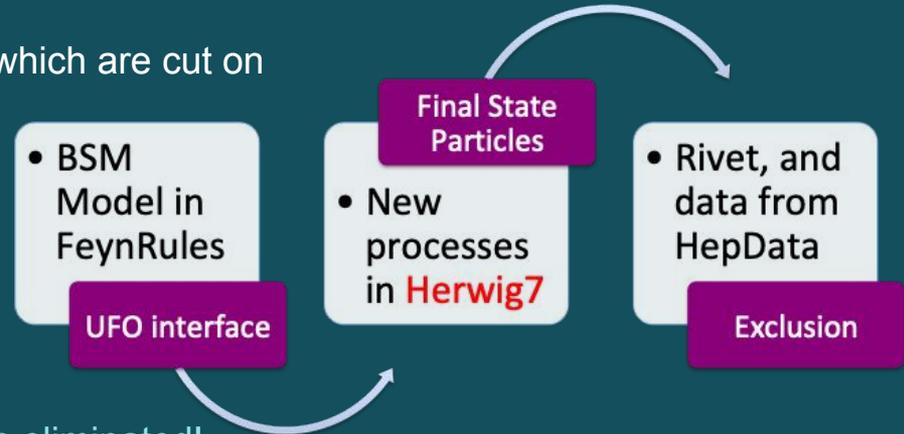
Heavy-ion physics preservation

- “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: [arXiv:2001.10737](https://arxiv.org/abs/2001.10737)
- HI MC standards are also in flux: having a common tool enables discussion on common standards



BSM from “Standard Model”

- **Not being focused on direct searches doesn't mean no interest in BSM!**
- **Particle-level measurements can achieve high level of model-independence**
 - Careful definition of fiducial cross-section
 - Control distributions of “hidden variables” which are cut on
 - Reduce model sensitivity in unfolding
- **Rivet used directly in e.g.**
 - TopFitter top quark EFT fits;
 - at core of ATLAS VH EFT fits; and...
- **Contur is getting particular uptake**
 - Inject signal to “SM” measurements: if it'd been statistically distinct, the model is eliminated!
 - Rivet gives huge “synoptic” coverage: a new result with Rivet code can be in BSM fits within *hours*
 - see Jon's talk for more details



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. CMSSW), 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++