Rivet analysis preservation
What, why and how?

Louie Corpe (CERN),
On behalf of the Rivet development team

With major thanks to A. Buckley and C. Bierlich
whose content has been heavily relied up for this talk

GDR-QCD Seminar
1 July 2021

Who am I?

- I’m Louie, a Research Fellow at CERN, working on the ATLAS experiment
- Have been part of the Rivet collaboration since 2017, my contributions have mainly been in the storage and portability of auxiliary analysis information, such as correlation of uncertainties across bins
- Disclaimer: I am NOT a heavy-ion expert, but I will still be able to give you an overview of Rivet, and point you existing HI plugins and further information
- Fear not: if I can’t answer your question, I can point you in the direction of someone who can !

Outline

- The Rivet toolkit and project
- Origins and generic analysis preservation
- Early applications: tuning
- Retooling for precision at the LHC
- Rivet for heavy ions
- The future of Rivet
- Wrapping up… and writing your own Rivet Analysis

Outline

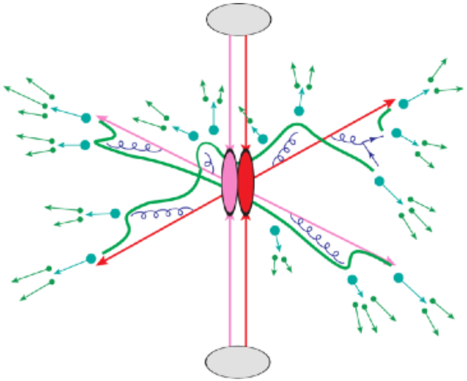
- **The Rivet toolkit and project**
- *Origins and generic analysis preservation*
- *Early applications: tuning*
- *Retooling for precision at the LHC*
- *Rivet for heavy ions*
- *The future of Rivet*
- *Wrapping up… and writing your own Rivet Analysis*

What is Rivet?

- The "LHC standard" monte carlo analysis toolkit

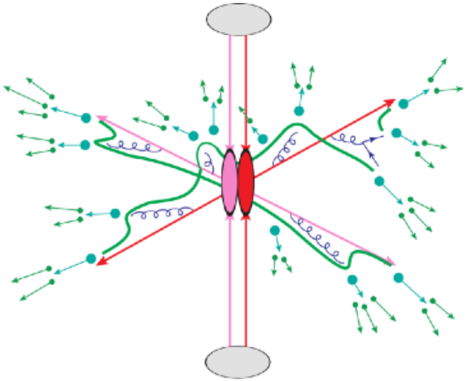
- More broadly, Rivet can be thought of as a language facilitating communication between:
    - 1. experiment & pheno.
    - 2. pheno & pheno.
    - 3. experiment & experiment.
    - 4. experiment & future experiment.

- As any language, the point is to ensure common definitions

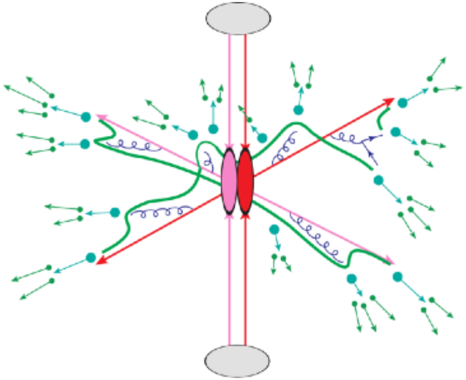


Concretely, What is Rivet?

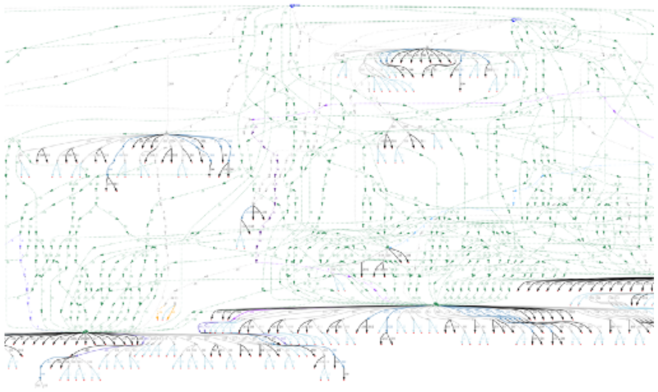
- A software project to preserve the logic of HEP data analyses and further expt-pheno collaboration
- Containing:
  - An event loop
  - Physics object / observable calculators
  - Fiducial / generator-independence emphasis
  - Integration with HepData
  - Transparent weight-stream handling
  - 1000+ analyses!
- Rivet now sits at the centre of a web of analysis reinterpretation tools, linking experiment to theory



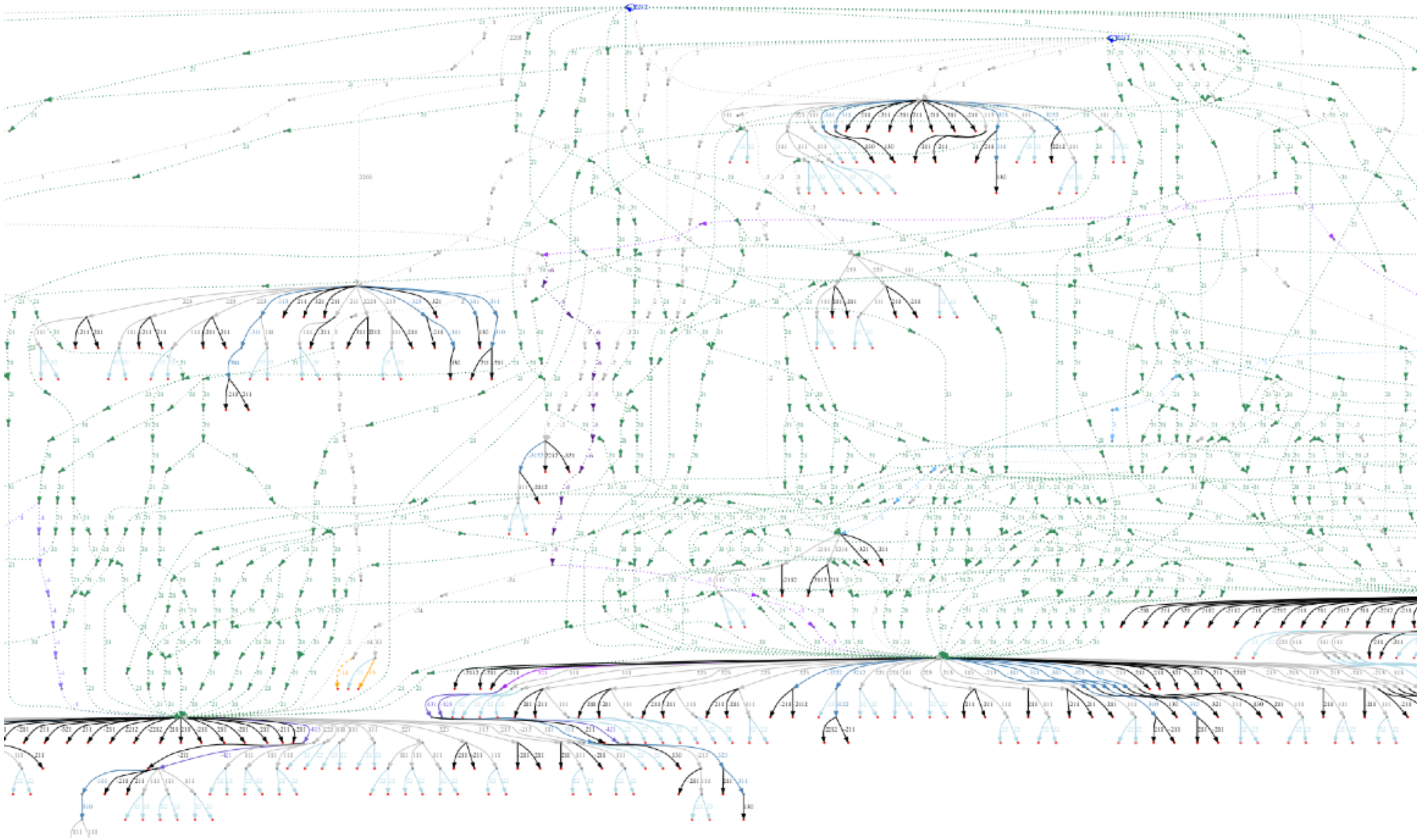
Monte Carlo Event Generators (MCEGs)



- **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"
  - 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 MCEGs !**
  - The main mechanism for translating theory to experimental signatures, from QCD to BSM
  - Generally very complex modelling and output



Just part of an *LO* top event!

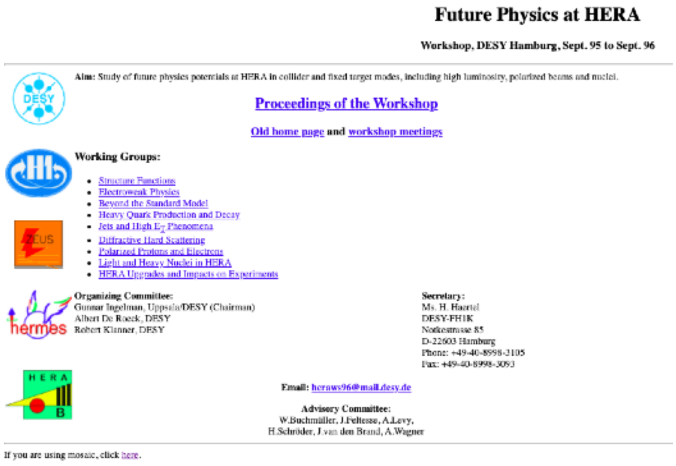


Outline

- *The Rivet toolkit and project*
- **Origins and generic analysis preservation**
- *Early applications: tuning*
- *Retooling for precision at the LHC*
- *Rivet for heavy ions*
- *The future of Rivet*
- *Wrapping up… and writing your own Rivet Analysis*

From HZTool to Rivet

- **The idea of preserving experimental analyses for MC validation was born out of HZTOOL (1990s)**
  - HERA (H1 and ZEUS) DIS and Photoproduction
  - Probing low-x, semi-perturbative physics: DIS with $Q^2 \sim 4$ GeV^2; jet $p_T \sim 5$ GeV; diffraction
  - Many “state of the art” models only in MCs
  - Much confusion about comparing like-with-like between generators, experiments, and analyses
    - Hard to find apples-to-apples comparisons!
  - HZTool (Fortran) for cross-experiment comparisons of similar measurements modulo cut differences
- **Direct line to Rivet, 10 years later: “HZ mark two”**
  - UK e-science funding; adopted by EU MCnet network



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.

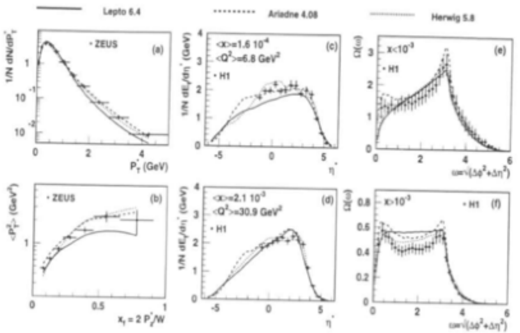
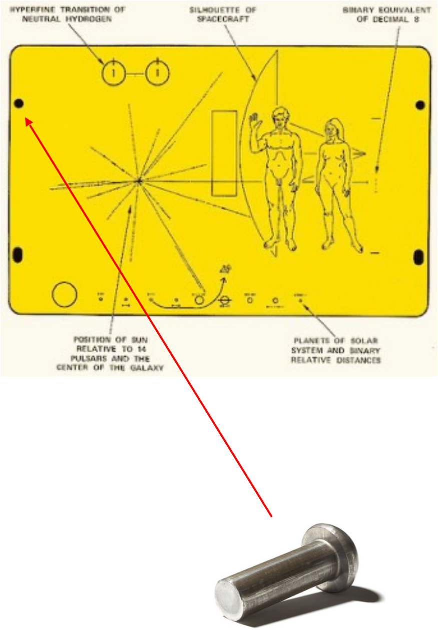


Figure 1: The transverse momenta dN/dp_T (a) and the 'seagull' plot $\langle P_T^2 \rangle \times x_F$ (b) of single particles in the positive hemisphere of the hadronic center of mass. The transverse energy flow $dE_T/d\eta$ in a low (c) and high (d) x and Q^2 bin. The transverse energy-energy correlations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

Lessons learned from HZTool

- **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...**
  - Model dependent observables are bad for MCEG. Might also be unphysical !| eg use of partons, bosons, etc. direct from the event graph
  - Frequently unphysical, depend on approximations. May not even exist!
  - HZTool full of cryptic "if HERWIG, if PYTHIA, if …" code
  - Adding a new generator meant patching ~all analyses!
  - ⇒ predict "real" observables, from well-defined final states

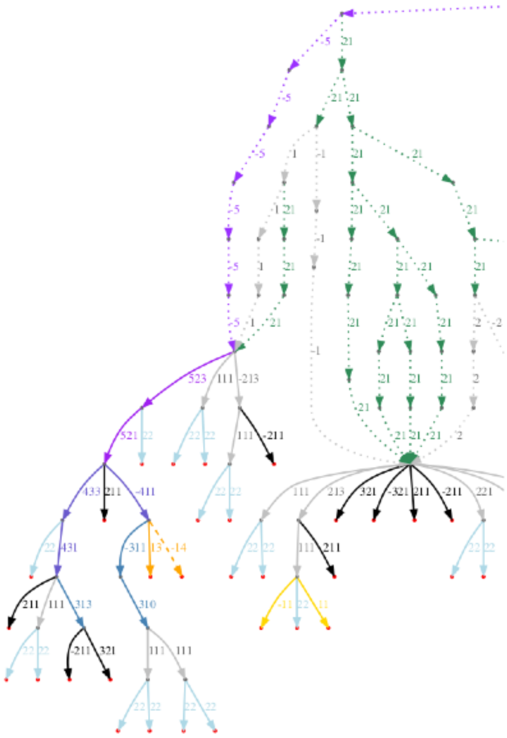
    aka Observables should really be ***observable!***
- **Standardisation: boring but important**
  - (physical) event format conventions, statuses, PDG particle numbering, etc.
- **Scalability**
  - Lots of expensive operations are repeated: sharing calculations is essential



Physically safe analysis methods

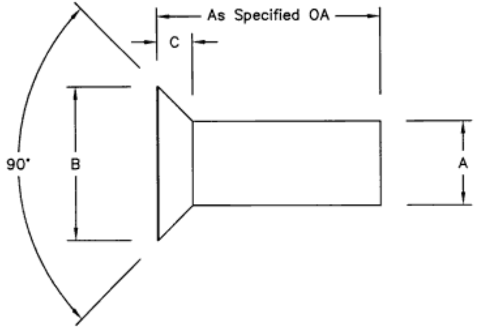
Avoiding unstandardised event-graph features was pragmatic, but led to some genuine physical insights:

- **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
  - Dependencies:
    - yoda (histograms),
    - HepMC (event format),
    - FastJet (jets and event shapes)
    - ***No generator dependencies***
- **Embeddable**
  - OO C++ library, Python wrapper, sane user scripts
  - Generator independence: communication via HepMC
  - Analysis routines factorised, and loaded as "plugins"

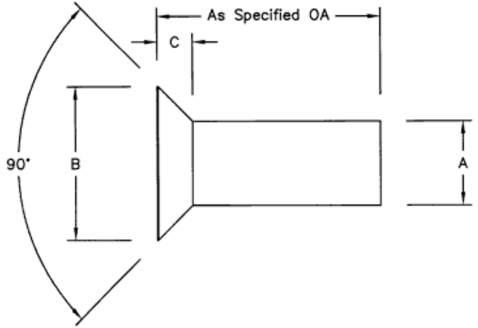


Designing the Rivet

- **Efficiency and Scalability**
  - Avoid recomputations via “projection” caching system
  - Event properties calculated once, should not be calculated again.
  - “Final states“ re-usable across many analyses.
  - Very scalable!
  - Division of tasks
    - Common functionality supplied by Rivet,
    - Analyses as pluggable modules by users.

    - Experiments validate analysis correctness,
    - Rivet dev team keeps the code running with updates.

- **Physical**
  - **Measurements primarily from final-state particles only**

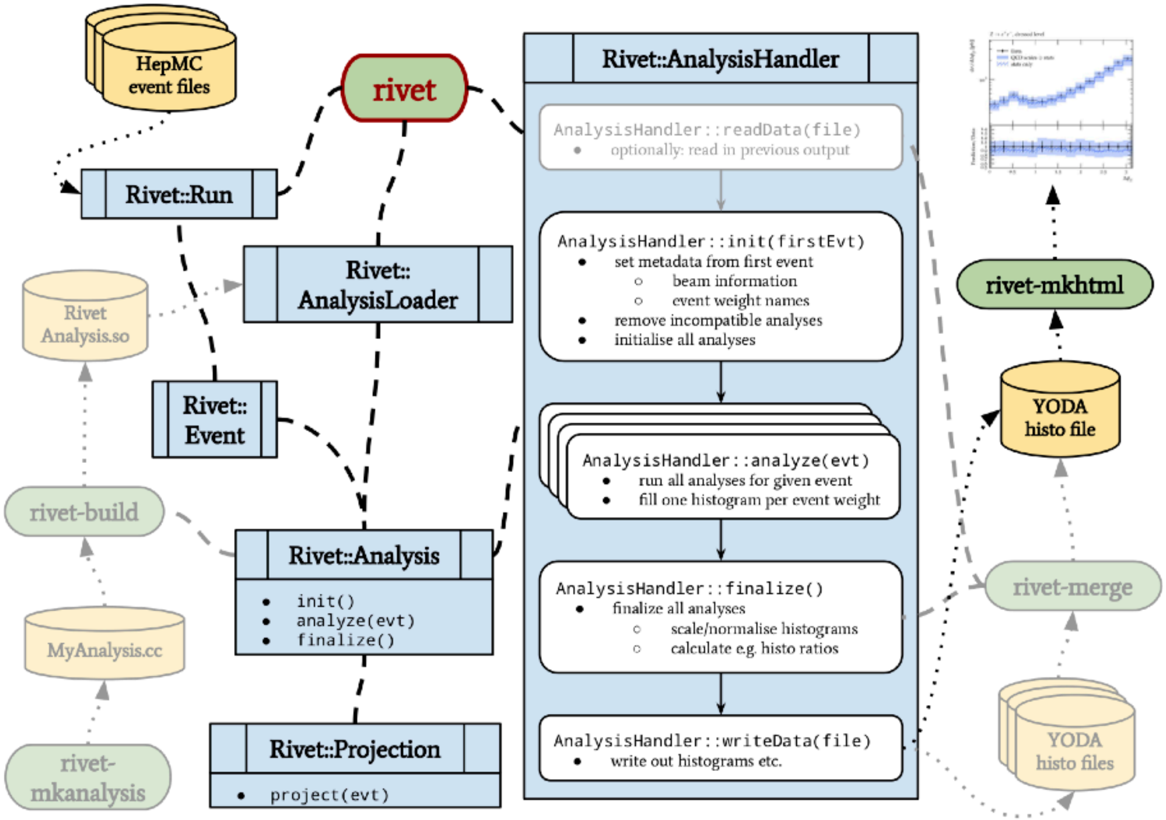


The result

https://rivet.hepforge.org

- **2010: Rivet v1 for LHC Run 1**
- **2019: Rivet v3.1.0 (arXiv:1912.05451)**
- Streamlined set of tools from analysis coding to event processing to plotting (and other applications)

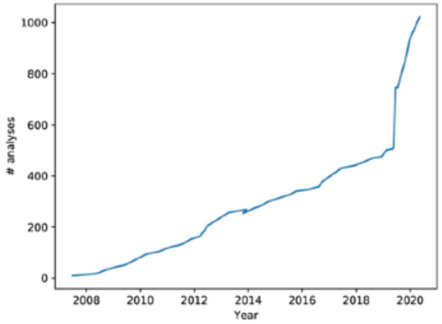
- And a key gateway to connect your analysis to theory (and back again)



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

- Part of publication procedure for ATLAS and CMS
- Many new initiatives and buy-in from ALICE
- Ongoing efforts for RHIC experiments and EIC
- Analysis contributions from NuSea, LHCb, LHCf, TeVatron, UA5, NAXX ... (and many, many more)
- Standard for MCnet event generators, more adding support.

What’s the benefit for experimentalists?

- **Preservation: Store your analysis once, and others will maintain it.**
- **Reproducibility: What happens when your student graduates?**
- **Ensure that your results are used.**
- **Don’t leave it to theorists to re-implement your analysis!**
- **“Do upon others…”: Generate MC tunes using other people’s work!**
- **Can I be sure that the framework will live on?**
  Yes! Large investment by HEP community and MCEG authors. O(1000) analyses already implemented. Active dev team, open to new improvements: If a feature is needed, we can look into it!

What’s the benefit for theorists?

- **A library of validated measurements to test your models against**
  - Does your generator’s latest version fix feature X in a spectrum, without breaking feature Y ?
  - Find out in minutes!

- **A common testing ground for different models**
  - Apples to apples comparisons to other generators

- **Boost your model’s credentials:**
  - **Use RIVET as a model-development tool**
  - Easy to convince experimentalists to use a new model if you can show that it leads to a clear improvement in data/MC agreement
    - While still agreeing in the rest of the body of experimental data
    - Avoid single-observable models and overfitting

What’s the coverage ?

https://rivet.hepforge.org/rivet-coverage

Rivet analysis coverage

Rivet analyses exist for 991/5661 papers = 18%. 189 priority analyses required.

Total number of Inspire papers scanned = 9441, at 2021-05-17

Breakdown by identified experiment (in development):

Key	ALICE	ATLAS	CMS	LHCb	Forward	HERA	e^+e^- (≥ 12 GeV)	e^+e^- (≤ 12 GeV)	Tevatron	RHIC	SPS	Other
Rivet wanted (total):	259	320	427	246	16	503	715	536	1131	454	62	1
Rivet REALLY wanted:	36	42	83	8	0	13	1	0	5	1	0	0
Rivet provided:	**26/285 = 9%**	**175/495 = 35%**	**91/518 = 18%**	**16/262 = 6%**	**8/24 = 33%**	**9/512 = 2%**	**180/895 = 20%**	**305/841 = 36%**	**58/1189 = 5%**	**8/462 = 2%**	**4/66 = 6%**	**112/113 = 99%**

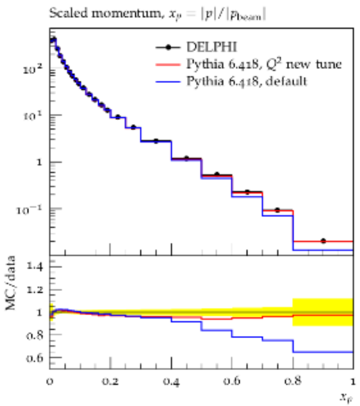
Outline

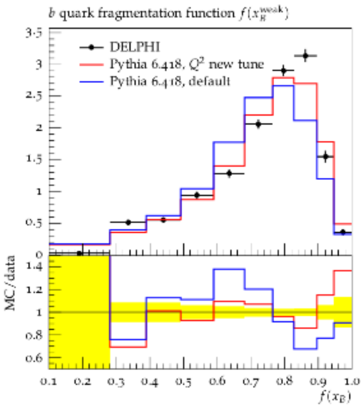
- *The Rivet toolkit and project*
- *Origins and generic analysis preservation*
- **Early applications: tuning**
- *Retooling for precision at the LHC*
- *Rivet for heavy ions*
- *The future of Rivet*
- *Wrapping up… and writing your own Rivet Analysis*

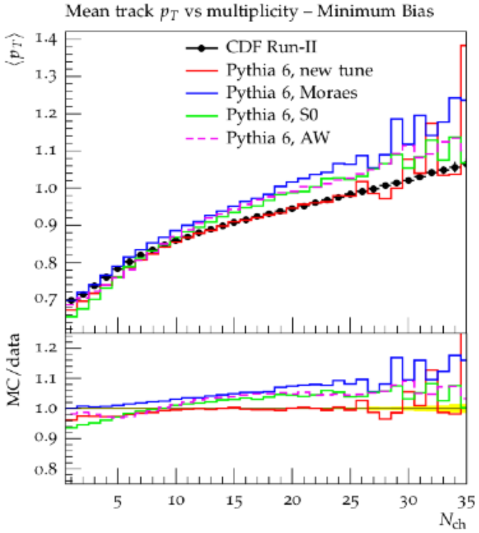
Event generator tuning

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; and…**



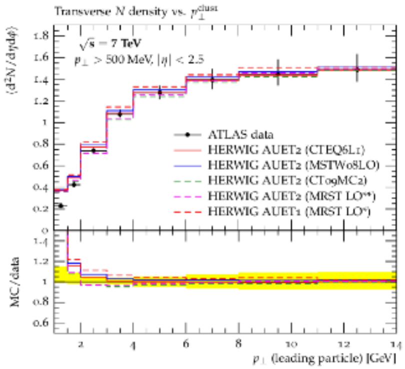


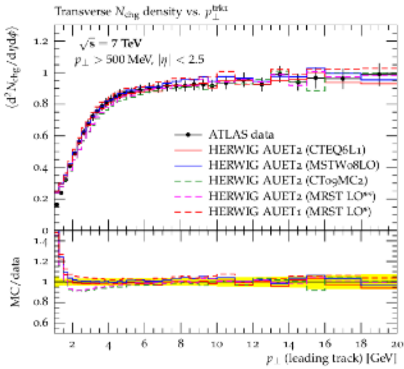


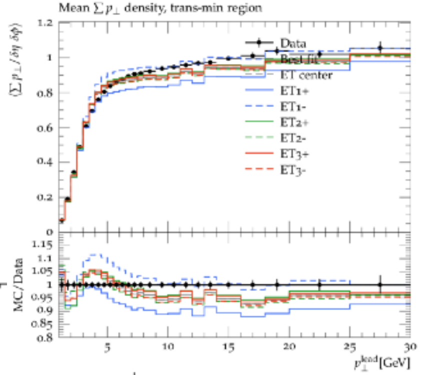
More tuning…

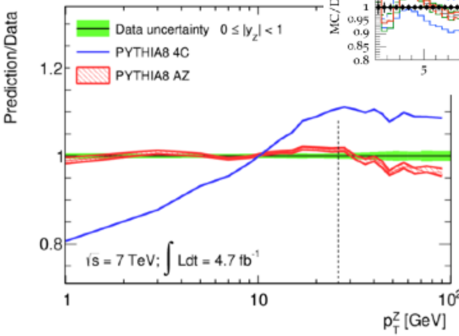
It’s getting hard to remember now, but pre-LHC the soft QCD uncertainties were *huge*

- **Factor x2 uncertainty on 7 TeV σ_{tot}!**
- **Feed in to underlying event, pile-up, etc.**
    - Tuning an essential task: better tunes ⇒ 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, …







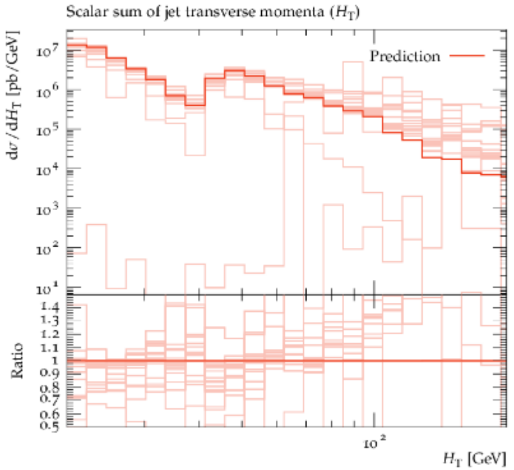


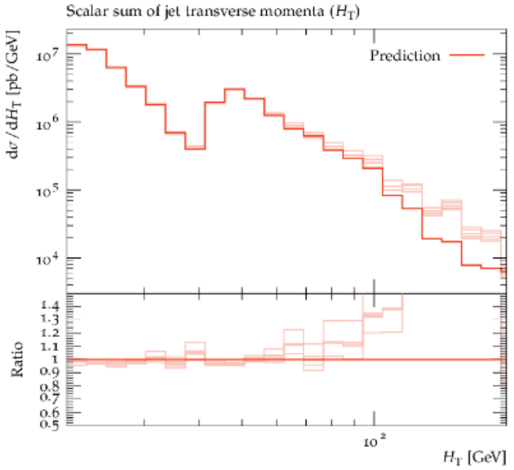
Outline

- *The Rivet toolkit and project*
- *Origins and generic analysis preservation*
- *Early applications: tuning*
- **Retooling for precision at the LHC**
- *Rivet for heavy ions*
- *The future of Rivet*
- *Wrapping up… and writing your own Rivet Analysis*

Multiweights and re-entry

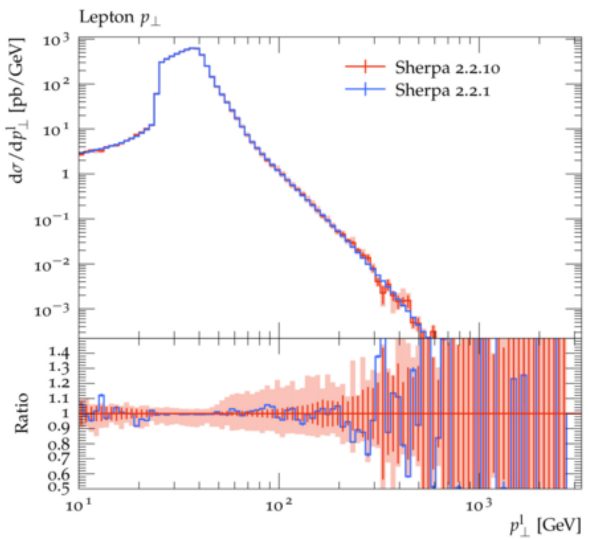
- **MC weight vectors allow expression of increasingly complex theory uncertainties.** But a burden for analysis chains: have to propagate and correctly combine O(200) weight streams!
- **Rivet 3: complex automatic handling of weights** ~invisible to users: data objects *look* like histograms etc. but 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

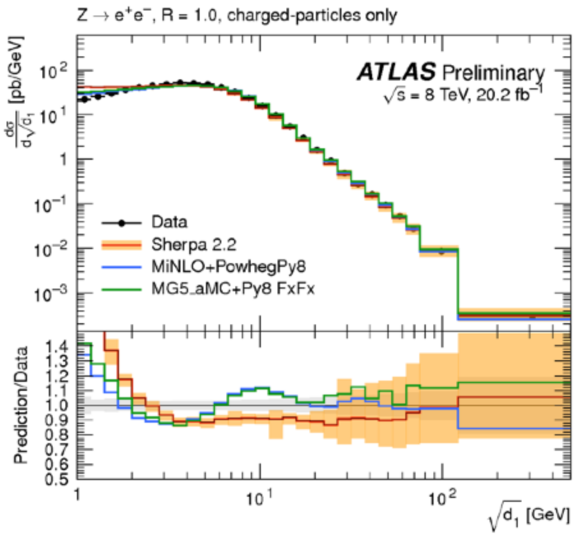


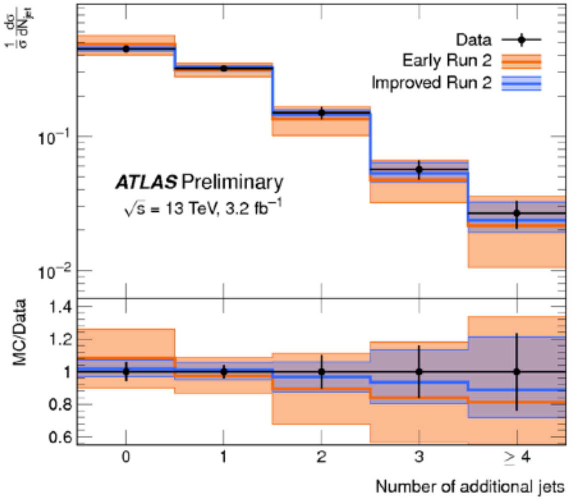


Rivet multiweights in action

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







- Weight-naming standardisation underway via MCnet

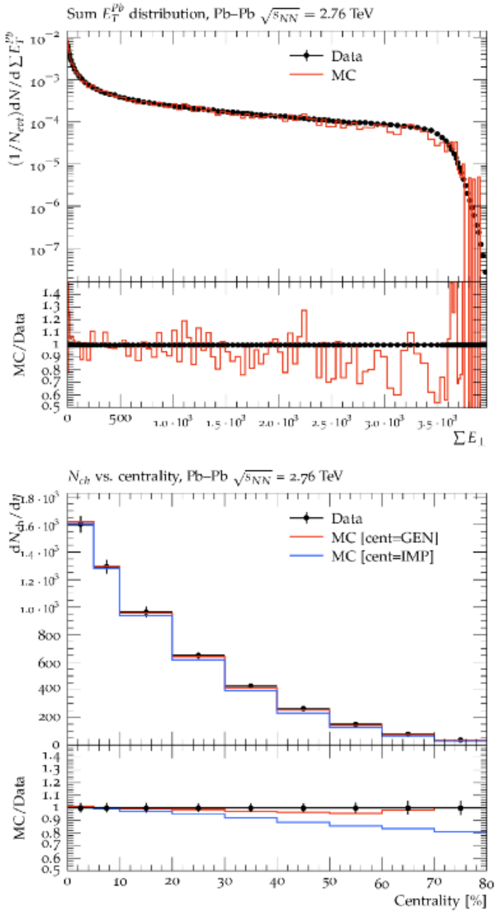
Outline

- *The Rivet toolkit and project*
- *Origins and generic analysis preservation*
- *Early applications: tuning*
- *Retooling for precision at the LHC*
- **Rivet for heavy ions**
- *The future of Rivet*
- *Wrapping up… and writing your own Rivet Analysis*

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!



Rivet for Heavy Ions

- Jan 2020 Paper:
  - https://arxiv.org/abs/2001.10737
  - Huge step towards, involving effort from all sides of the community
  - Now an integral part of core Rivet : not a separate tool
  - Still A work in progress! Improvements possible
  - *Result: Features to allow comparison between heavy ion data and MC*

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

LU TP 20-04
MCNET-20-04

Confronting Experimental Data with Heavy-Ion Models

Rivet for Heavy Ions

Christian Bierlich,[1,2] Andy Buckley,[3] Christian Holm Christensen,[1] Peter Harald Lindenov Christiansen,[4] Cody B. Duncan,[5,6] Jan Fiete Grosse-Oetringhaus,[7] Przemyslaw Karczmarczyk,[8,7] Patrick Kirchgaeßer,[6] Jochen Klein,[9,7] Leif Lönnblad,[2] Roberto Preghenella,[10] Christine O. Rasmussen,[2] Maria Stefaniak,[8,11] and Vytautas Vislavicius[1]

[1]*Niels Bohr Institute, Blegdamsvej 17, 2100 Copenhagen, Denmark*

[2]*Dept. of Astronomy and Theoretical Physics, Lund University, Sölvegatan 14A, S-223 62 Lund, Sweden*

[3]*School of Physics and Astronomy, University of Glasgow, G12 8QQ, Glasgow, UK*

[4]*Dept. of Physics, Lund University, Professorsgatan 1, S-223 62 Lund, Sweden*

[5]*School of Physics and Astronomy, Monash University, Clayton, VIC 3800, Australia*

[6]*Institute for Theoretical Physics, Karlsruhe Institut für Technologie, Wolfgang-Gaede-Straße 1, 76131 Karlsruhe Germany*

[7]*CERN, Esplanade de Particules 1, Geneva, Switzerland*

[8]*Faculty of Physics, Warsaw University of Technology, Koszykowa 75, 00-662 Warszawa, Poland*

[9]*Istituto Nazionale di Fisica Nucleare, Sezione di Torino, Italy*

[10]*Istituto Nazionale di Fisica Nucleare, Sezione di Bologna, Italy*

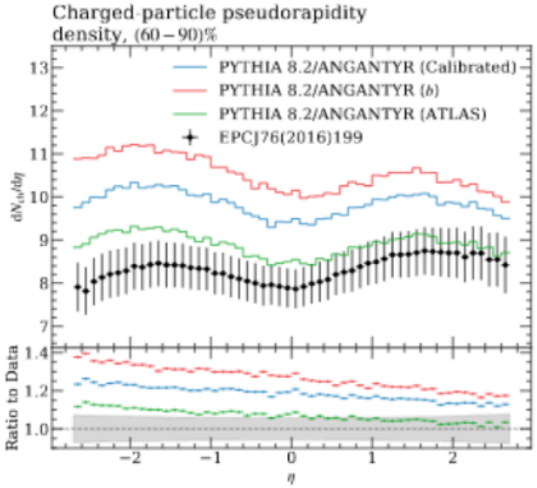
[11]*Subatech – IMT Atlantique, 4 rue Alfred Kastler, 44307 Nantes, France*

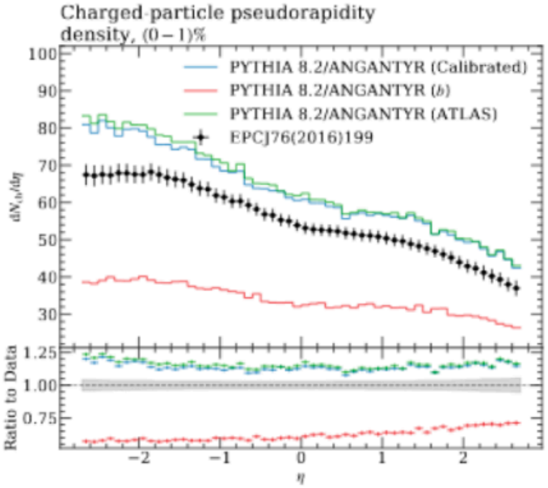
(Dated: January 30, 2020)

arXiv:2001.10737v1 [hep-ph] 29 Jan 2020

Centrality

- Centrality is a key concept in HI physics
- Theory-level definition different to experimental level
- Subtle biases, especially in pA





Rivet for HI includes:

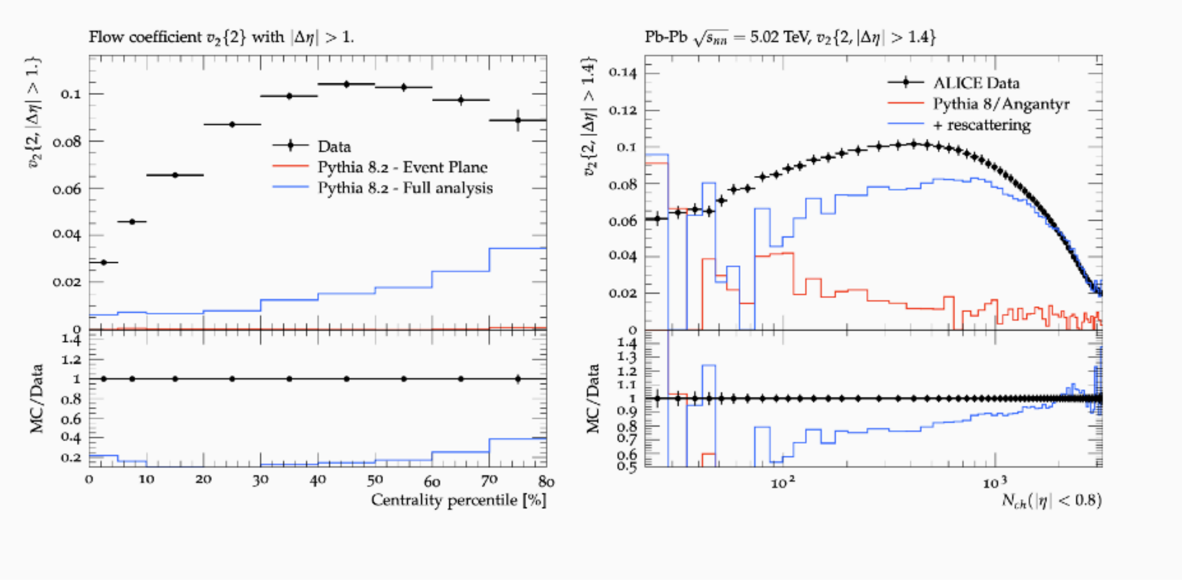
- Centrality calibration
- Analysis options to select calibration

Flow observables

Rivet for HI includes:

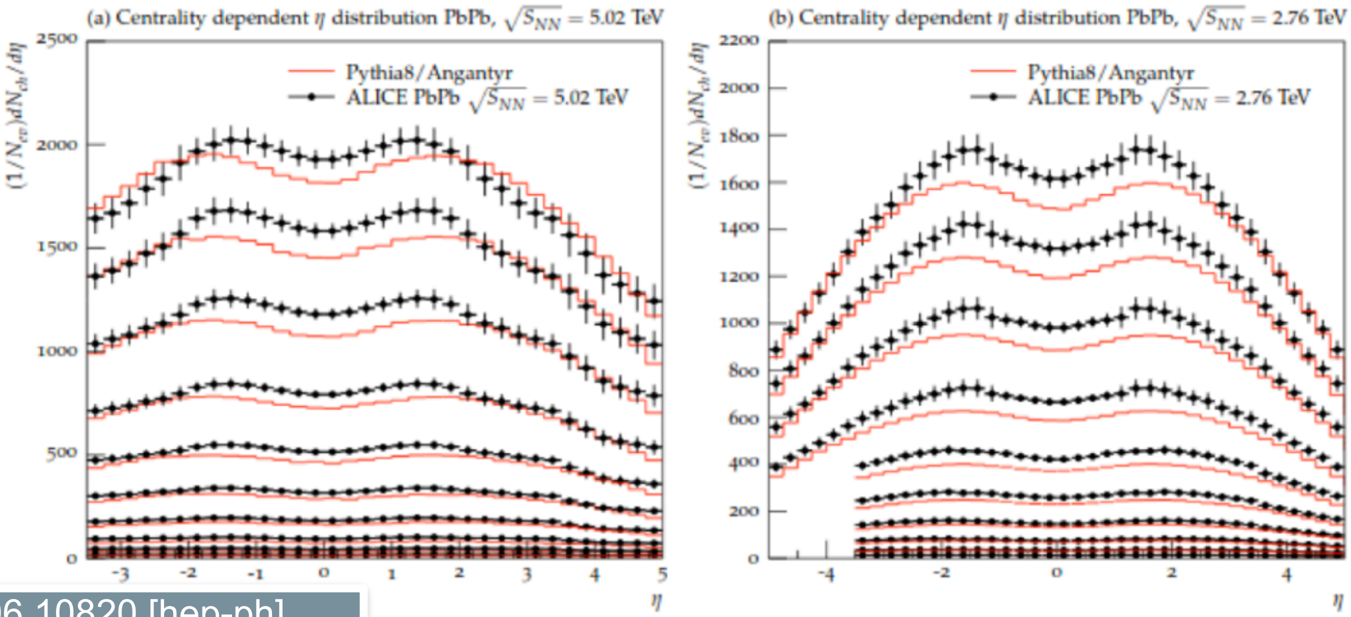
- Generic framework for calculating flow observables, energy correlators and cumulants

- Already analyses using them: eg ALICE 2016 I1419244 and ALICE 2019 I1723697



Triggers and Particle Definitions

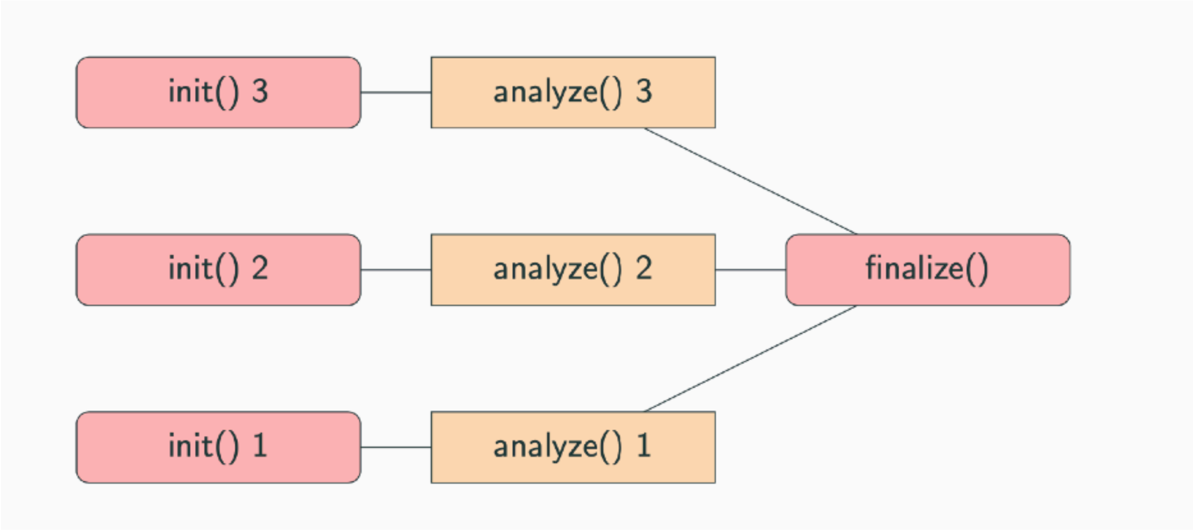
- Getting these right are important, can be ~10% effect, similar to MC accuracy
- Rivet for HI includes ALICE::Trigger and ALICE::PrimaryParticle projections



Angantyr: 1806.10820 [hep-ph]

Run Parallelisation

- Parallelization is necessary but potentially difficult: flavour ratios, R_{AA}, flow…
- Solution: rivet-merge before finalization
- Let analyser implement merging → perfect run combination



Outline

- *The Rivet toolkit and project*
- *Origins and generic analysis preservation*
- *Early applications: tuning*
- *Retooling for precision at the LHC*
- *Rivet for heavy ions*
- **The future of Rivet**
- *Wrapping up… and writing your own Rivet Analysis*

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, 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 fitting tools
  - Preserving MVAs: BDT and NN in vanilla C++

Outline

- *The Rivet toolkit and project*
- *Origins and generic analysis preservation*
- *Early applications: tuning*
- *Retooling for precision at the LHC*
- *Rivet for heavy ions*
- *The future of Rivet*
- **Wrapping up… and writing your own Rivet Analysis**

Wrapping up… and Getting your hands dirty!

I hope I've convinced you that lightweight analysis preservation isn't just some niche interest or admin overreach

An analysis that's immediately available to the pheno community is 10x more useful ⇒

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 R3 paper

Imitation is the highest form of flattery: copy an existing analysis!

https://gitlab.com/hepcedar/rivet/-/tree/release-3-1-x/analyses/pluginALICE
https://gitlab.com/hepcedar/rivet/-/tree/release-3-1-x/analyses/pluginRHIC
https://gitlab.com/hepcedar/rivet/-/tree/release-3-1-x/analyses/pluginHERA
… etc :)

ALICE: Centrality evolution of the charged-particle pseudorapidity density over a broad pseudorapidity range in Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 2.76$ TeV

Inspire: 1394676 · arXiv: 1509.07299 · DOI/journal: 10.1016/j.physletb.2015.12.082 · CDS: 2052525 · HepData: ins1394676 · Report IDs: CERN-PH-EP-2015-257, ALICE-PUBLIC-2015-010

ALICE_2016_I1394676

ALICE: Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions

Inspire: 1471838 · arXiv: 1606.07424 · DOI/journal: 10.1038/nphys4111 · CDS: 2189682 · HepData: ins1471838 · Report IDs: CERN-EP-2016-153

ALICE_2016_I1471838

H1: Diffractive Dijet Production with a Leading Proton in ep Collisions at HERA

Inspire: 1343110 · arXiv: 1502.01683 · DOI/journal: 10.1007/JHEP05(2015)056 · CDS: 1987204 · HepData: ins1343110 · Report IDs: DESY-14-242

H1_2015_I1343110

STAR: Bulk Properties of the Medium Produced in Relativistic Heavy-Ion Collisions from the Beam Energy Scan Program

Inspire: 1510593 · arXiv: 1701.07065 · DOI/journal: 10.1103/PhysRevC.96.044904 · HepData: ins1510593

STAR_2017_I1510593

The first of many!
Thank you for your attention

ZEUS: Diffractive photoproduction of dijets in ep collisions at HERA

Inspire: 763404 · arXiv: 0710.1498 · DOI/journal: 10.1140/epjc/s10052-008-0598-2 · HepData: ins763404 · Report IDs: DESY-07-161

ZEUS_2008_I763404

STAR: Beam Energy Dependence of the Third Harmonic of Azimuthal Correlations in Au+Au Collisions at RHIC

Inspire: 1414638 · arXiv: 1601.01999 · DOI/journal: 10.1103/PhysRevLett.116.112302 · HepData: ins1414638

STAR_2016_I1414638

STAR: Measurement of the central exclusive production of charged particle pairs in proton-proton collisions at $\sqrt{s} = 200$ GeV with the STAR detector at RHIC

Inspire: 1792394 · arXiv: 2004.11078 · DOI/journal: 10.1007/JHEP07(2020)178 · HepData: ins1792394

STAR_2020_I1792394

H1: Tests of QCD factorisation in the diffractive production of dijets in deep-inelastic scattering and photoproduction at HERA

Inspire: 746380 · arXiv: hep-ex/0703022 · DOI/journal: 10.1140/epjc/s10052-007-0325-4 · CDS: 1023604 · HepData: ins746380 · Report IDs: DESY-07-018

H1_2007_I746380

ALICE: Investigations of Anisotropic Flow Using Multiparticle Azimuthal Correlations in pp, p-Pb, Xe-Xe, and Pb-Pb Collisions at the LHC

Inspire: 1723697 · arXiv: 1903.01790 · DOI/journal: 10.1103/PhysRevLett.123.142301 · CDS: 2661740 · HepData: ins1723697 · Report IDs: CERN-EP-2019-033

ALICE_2019_I1723697

https://rivet.hepforge.org/rivet-coverage