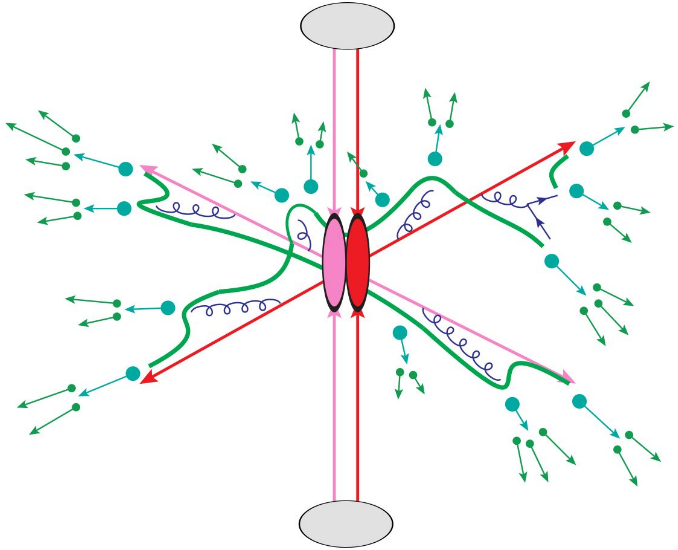
Validating new ePIC MC with “old” data

Andy Buckley,
University of Glasgow

ePIC User Meeting 2026, Glasgow
14 July 2026

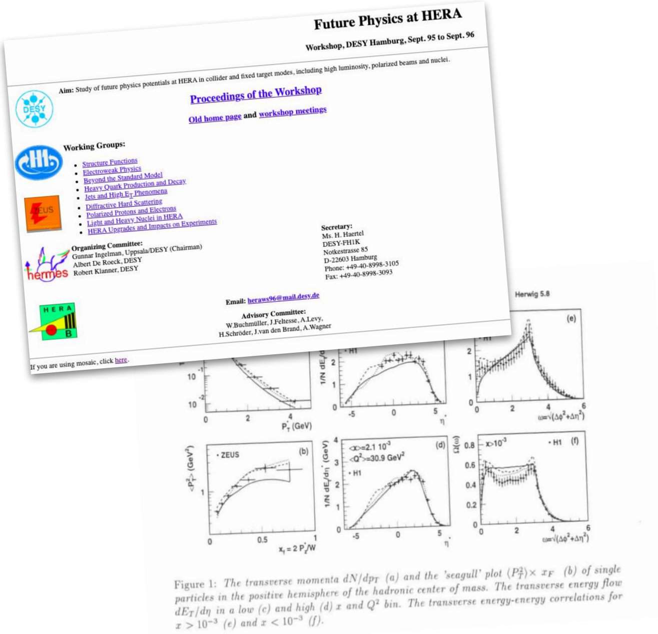
MC process validation

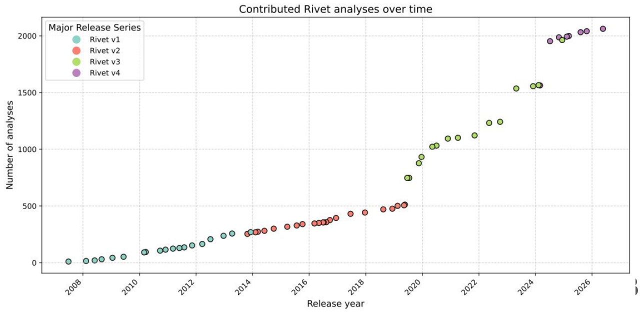
- **Frank has covered the new inclusive DIS and photoproduction processes**
  - Modern MCs like Sherpa (and Pythia, Herwig, maybe external hard-process gens coupled to shower gens) operate at **high formal precision**
  - Loop, perhaps multi-loop, matrix elements, matched to (multiple) real emissions beyond leading order, and parton shower (~resummation) + EW radiative corrs
  - Plus **“informal” high precision** in e.g. non-perturbative: MPI, hadronisation, hadron decays (⇒ via *tuning*)
  - **Many state-of-the-art physics models are *only* in MCs**
- **Precision models need systematic validation**
  - **Against *real* data, not just version-to-version regression**



Not the first rodeo

- **We've been here before. HERA experiments preserved analysis code in HZTOOL for MC validation**
  - And comparison; often the best language for communication between data and different theory groups is via analysis code and MC samples
- **In run-up to LHC, HZTOOL was reinvented as Rivet**
  - Essential to early-LHC validation and tuning of 7 TeV pp physics, and PDF/energy extrapolation from HERA and Tevatron MPI models
  - Now a core component of LHC operations: > 2000 analyses collected, submission by experiments as standard
  - Hidden technical magic to make this scale





Rivet analysis code

Rivet analysis code is intentionally simple, and built+run easily: `rivet -a MY_ANALYSIS`

Each Rivet analysis is a plugin inheriting from `Rivet::Analysis`.

Setup phase: Book histograms, declare support algorithms etc.

Histograms are automatically managed and linked to reference data where available.

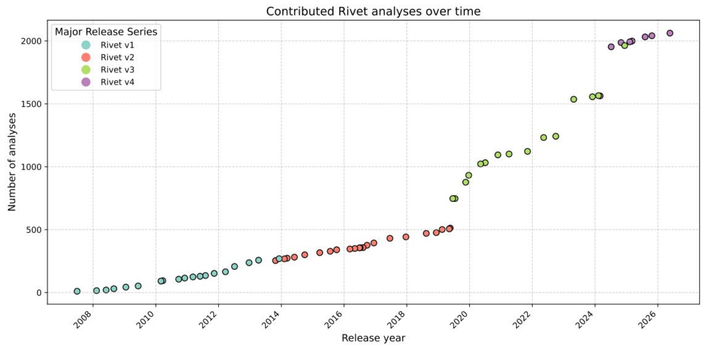
Main event loop: Retrieve event data, apply selections, fill histograms.

Final normalisation step: Scale histograms based on event weight sum.

```cpp
#include "Rivet/Analysis.hh"
#include "Rivet/Projections/FastJets.hh"

namespace Rivet {
  class MySimpleAnalysis : public Analysis {
  public:
    RIVET_DEFAULT_ANALYSIS_CTOR(MySimpleAnalysis);

    void init() {
      FastJets fj(FinalState(), JetAlg::ANTIKT, 0.4);
      declare(fj, "Jets");
      book(_h_jetPt, "Jet_Pt", 50, 0, 500);
    }

    void analyze(const Event& event) {
      const Jets& jets = apply<FastJets>(event, "Jets").jetsByPt();
      for (const Jet& jet : jets) _h_jetPt->fill(jet.pT()/GeV);
    }

    void finalize() {
      normalize(_h_jetPt);
    }

  private:
    Histo1DPtr _h_jetPt;
  };

  RIVET_DECLARE_PLUGIN(MySimpleAnalysis);
}
```

Rivet analysis coverage

- **Version 4.1.3 → 2,045 analyses**
  - `docker run -it hepstore/rivet-tutorial`



- **Official support from experiments is crucial**
  - Preservation of analysis logic as standard
  - Original analysis teams know it best; papers are never quite complete/unambiguous
  - Still imperfect! We monitor paper coverage ⇒ **Analysis wishlist for EIC…?**

Rivet analysis coverage (no searches, no heavy ion)

Rivet analyses exist for 1519/4238 papers = 36%. 237 priority analyses required.

Total number of Inspire papers scanned = 7178, at 2025-08-07

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):	140	173	181	115	16	415	630	72	671	177	61	68
Rivet REALLY wanted:	41	68	75	15	3	17	1	0	9	1	7	0
Rivet provided:	**24**/164 = **15%**	**184**/357 = **52%**	**119**/300 = **40%**	**70**/185 = **38%**	**9**/25 = **36%**	**43**/458 = **9%**	**188**/818 = **23%**	**807**/879 = **92%**	**57**/728 = **8%**	**6**/183 = **3%**	**0**/61 = **0%**	**12**/80 = **15%**

Show greylist Show blacklist Hide rivet

ALICE | ATLAS | CMS | LHCb | Forward | HERA | e^+e^- (≥ 12 GeV) | e^+e^- (≤ 12 GeV) | Tevatron | RHIC | SPS | Other

ATLAS: Measurements of the production cross-section for a Z boson in association with b- or c-jets in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

Inspire: 2771257 · arXiv: 2403.15093 (pdf) · DOI/journal: 10.1140/epjc/s10052-024-13159-w · CDS: 2892678 · HepData: ins2771257 · Report IDs: CERN-EP-2024-081

ATLAS_2024_I2771257

ATLAS: Measurements of electroweak $W^{\pm}Z$ boson pair production in association with two jets in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector

Inspire: 2771256 · arXiv: 2403.15296 (pdf) · DOI/journal: 10.1007/JHEP06(2024)192 · CDS: 2892680 · Report IDs: CERN-EP-2024-082

ATLAS: Measurement of vector boson production cross sections and their ratios using pp collisions at $\sqrt{s} = 13.6$ TeV with the ATLAS detector

Inspire: 2770237 · arXiv: 2403.12902 (pdf) · DOI/journal: 10.1016/j.physletb.2024.138725 · CDS: 2892152 · Report IDs: CERN-EP-2024-069

ATLAS: Measurements of inclusive and differential cross-sections of $t\bar{t}\gamma$ production in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector

Inspire: 2768921 · arXiv: 2403.09452 (pdf) · DOI/journal: 10.1007/JHEP10(2024)191 · CDS: 2891743 · HepData: ins2768921 · Report IDs: CERN-EP-2024-052

ATLAS_2024_I2768921

ATLAS: Observation of electroweak production of W^+W^- in association with jets in proton-proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector

Inspire: 2766986 · arXiv: 2403.04869 (pdf) · DOI/journal: 10.1007/JHEP07(2024)254 · CDS: 2891254 · HepData: ins2766986 · Report IDs: CERN-EP-2024-015

ATLAS: Fiducial and differential cross-section measurements of electroweak $W\gamma jj$ production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector

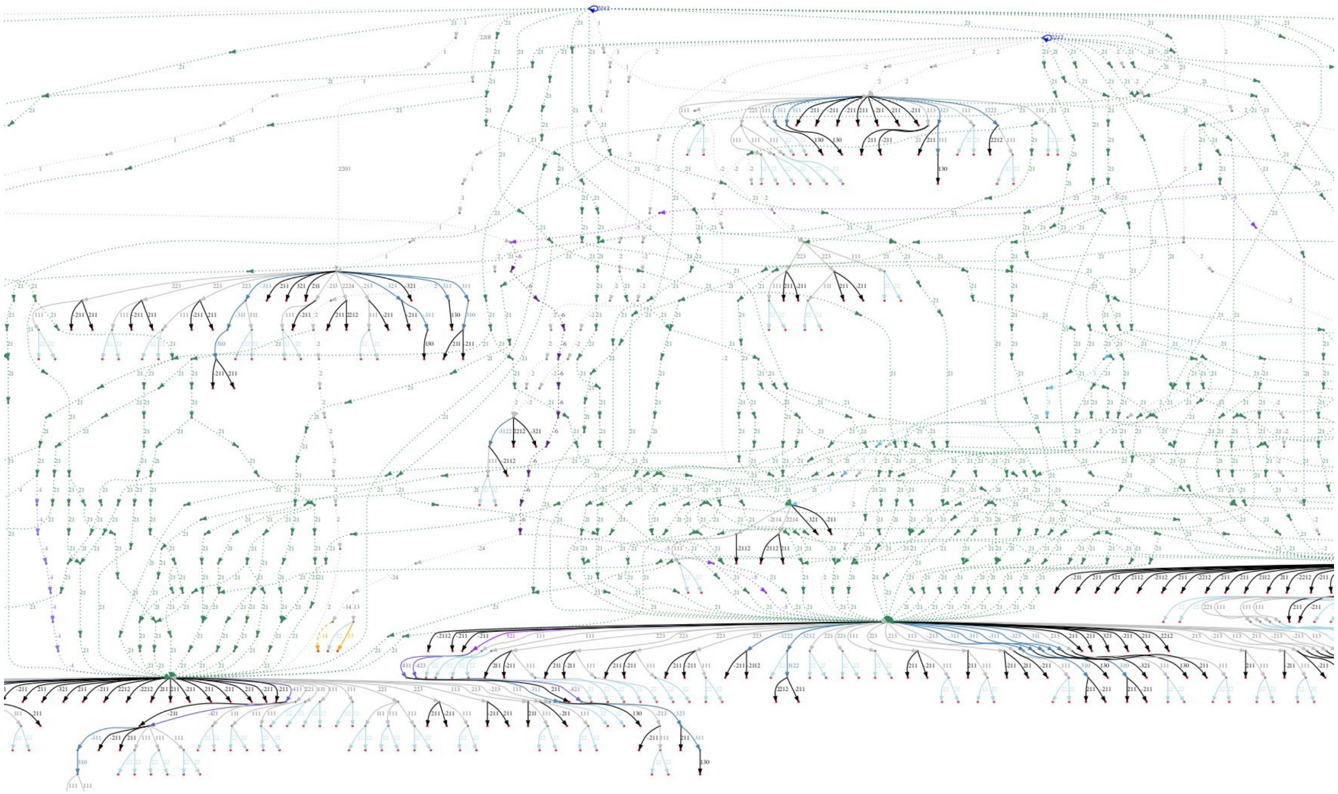
Inspire: 2765267 · arXiv: 2403.02809 (pdf) · DOI/journal: 10.1140/epjc/s10052-024-13311-6 · CDS: 2890900 · Report IDs: CERN-EP-2024-048

ATLAS: Differential cross-sections for events with missing transverse momentum and jets measured with the ATLAS detector in 13 TeV proton-proton collisions

Inspire: 2765017 · arXiv: 2403.02793 (pdf) · DOI/journal: 10.1007/JHEP08(2024)223 · CDS: 2890899 · HepData: ins2765017 · Report IDs: CERN-EP-2024-034

ATLAS_2024_I2765017

Portability… “MC truth” events are not *true*!



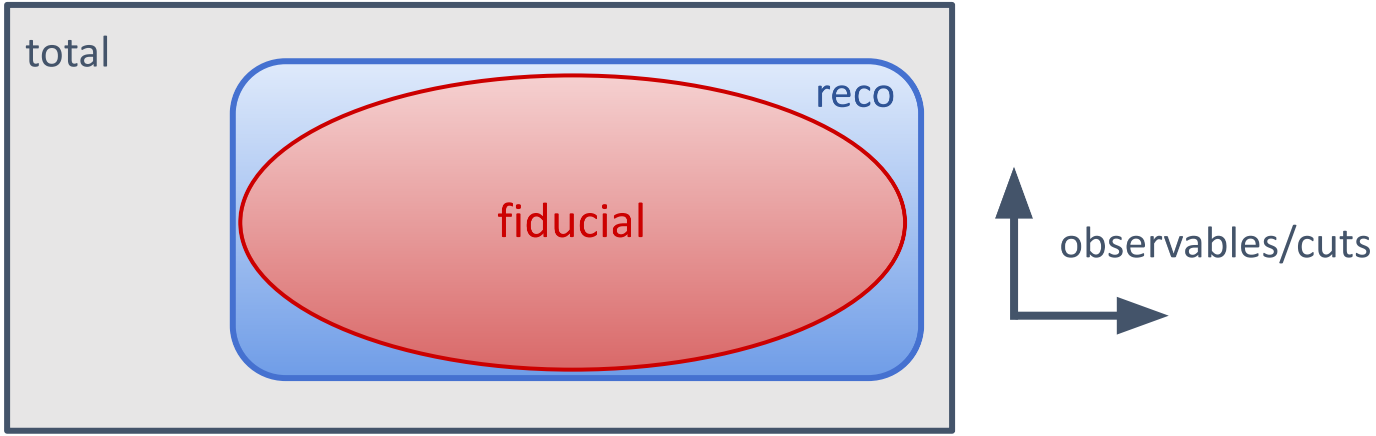
⇐ ~⅓ of an LO tt event

MC events are full of unphysical debug info, kinematic inconsistencies, *ad hoc* structures & representations, etc.!

Avoid physicists needing to rediscover graph algorithms, MC conventions, and physical/debug distinctions, …

Focus on fiducial phase-space

Regardless of detector efficiencies and resolutions, there are **uninstrumented kinematic regions that we don't measure *at all*:**
We do not have 4π detectors, and can't reconstruct particles down to zero p_T



A fiducial phase-space is a set of "truth level" object-selection criteria that align well with the sensitivity of the real detector + reco ⇒ the most robust theory-data comparison

Focus on fiducial phase-space

Regardless of detector efficiencies and resolutions, there are **uninstrumented kinematic regions that we don't measure *at all*:**
We do not have 4π detectors, and can't reconstruct particles down to zero p_T

Designing validation for physicality makes it portable, and also feeds into better analysis design. Modern, gold-standard approach

A fiducial phase-space is a set of "truth level" object-selection criteria that align well with the sensitivity of the real detector + reco ⇒ the most robust theory-data comparison

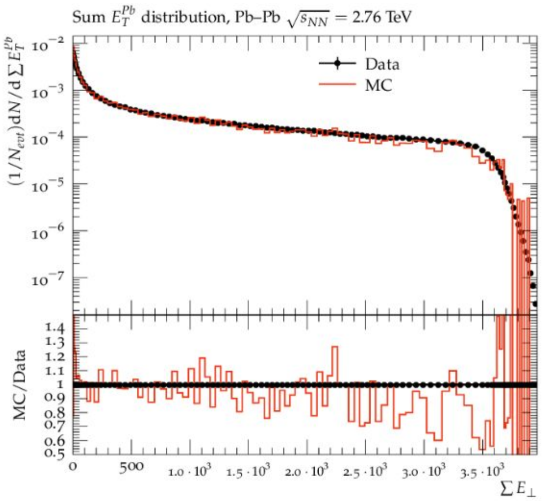
Ion-physics in MC analysis

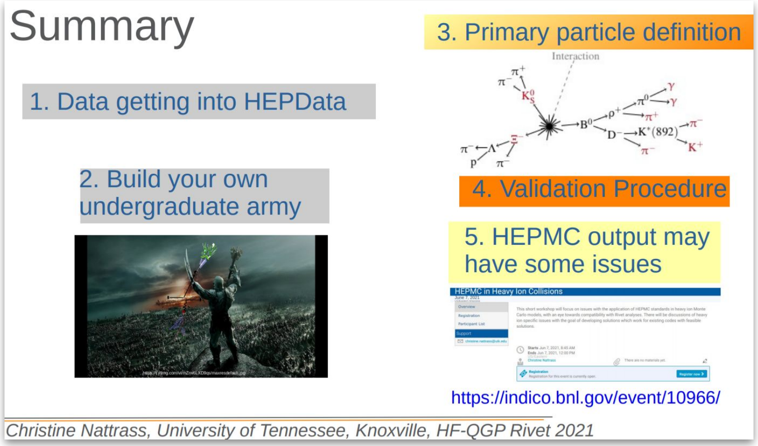
- **Heavy-ion analysis:** high-complexity multi-scale event modelling, no current tools that can do everything. Multi-pass processing, external calibration of e.g. *centrality*

- Specific tools in Rivet include flow cumulant calculations, and centrality "calibration" inputs

- Some really nice community-led initiatives grew up around tools, spurred standardisations, collaboration between HEP/nuclear communities, and drive modelling developments

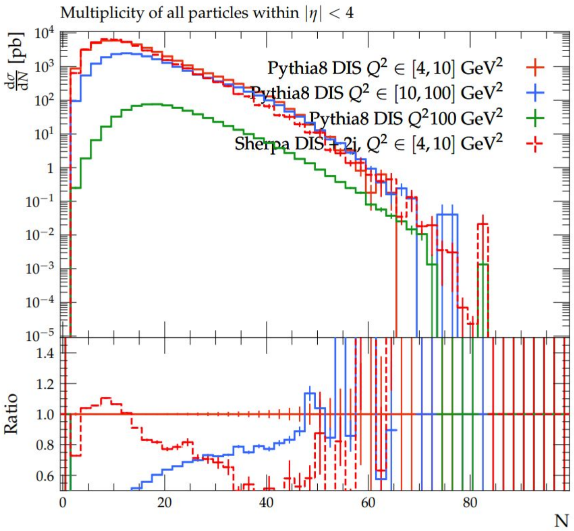
- **Still room for more RHIC analysis →ePIC**

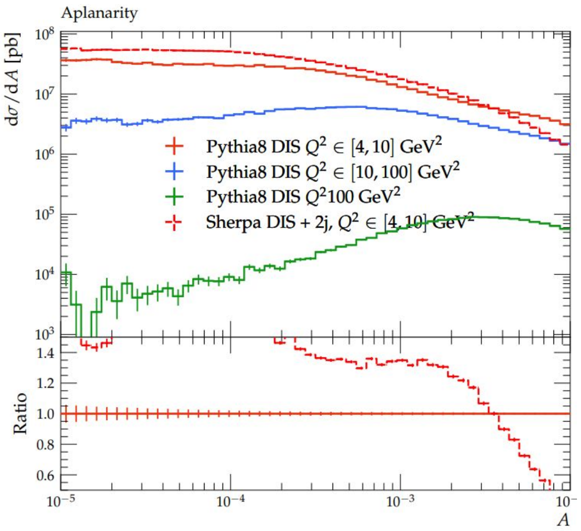


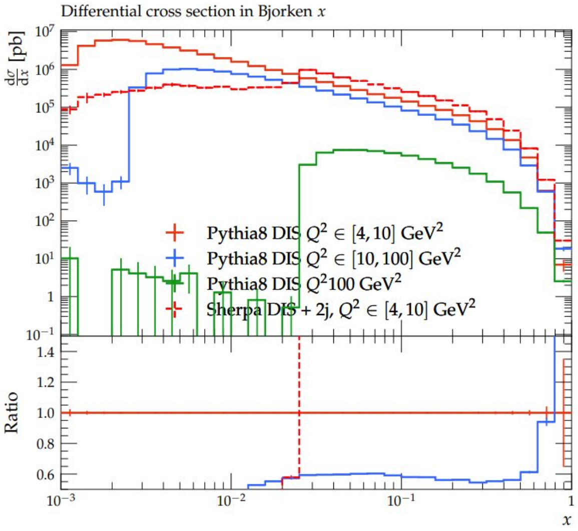


Starting ePIC DIS validation

- New `MC_DIS_EIC` test analysis from Ilkka Helenius; 35 differential distributions, in p_T, pseudorapidity, x, DIS kinematics, event shapes, jet splittings, identified hadrons
  — Initial MC samples from Ilkka (Pythia8) and Peter Meinzinger (Sherpa)

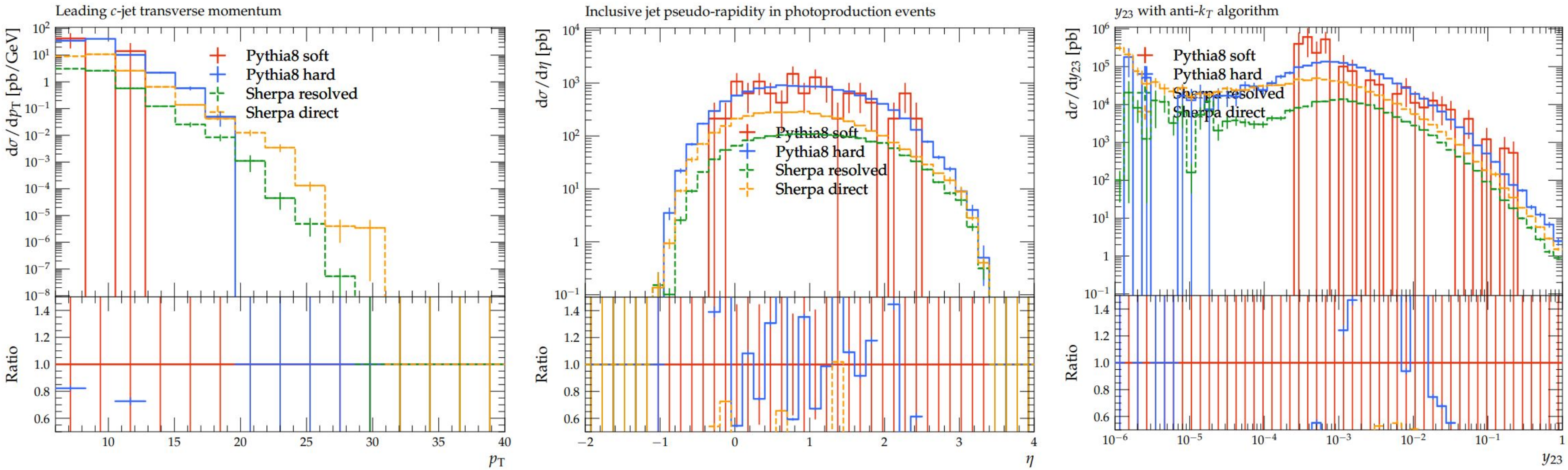






Starting ePIC photoproduction validation

- Built-in MC_PHOTOPRODUCTION (Anjelo Narendran, Peter Meinzinger); Pythia8 and Sherpa samples from Ilkka and Peter again

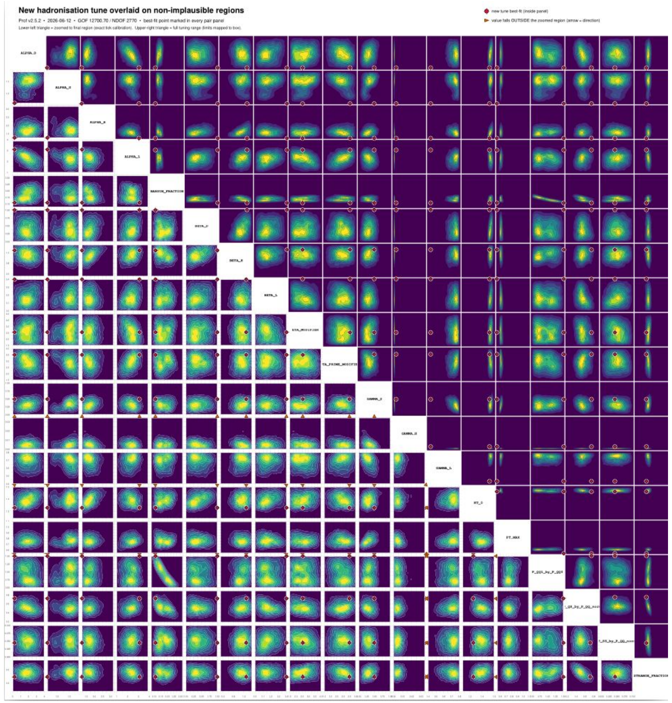


Legacy analysis push

- **HERA: H1, ZEUS**
  - Narendran / Gutschow / Meinzinger via MCnet/LPCC short-term studentship
  - Implemented ~10 new HERA measurements in Rivet for EIC-relevant hadronisation studies
  - Validating and extending tuning workflow through reproduction of an existing Sherpa LEP tune ⇒ dedicated HERA hadronisation tune

- **HERA: HERMES?**
  - What are we missing, how can we get it?

- **What else?**
  - RHIC, COMPASS, JLab, …?



Credit: Anjelo Narendran, Moritz Pabst, cf. https://arxiv.org/pdf/2602.22324

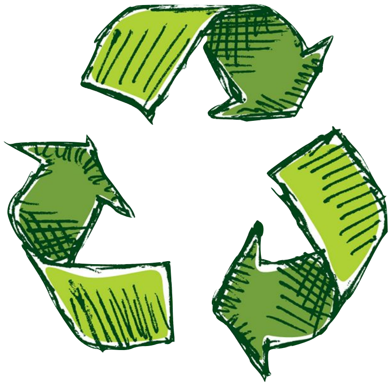
Starting ePIC exclusive and ion validation

- **What is wanted / needed?!**
  - MC generation has to go to the HepMC format
  - Opportunity to extend to variation weighting, etc.
  - Use of LHE → using general PS/hadr gens for higher-order effects? Precision in ME?
  - Ion gens?

- **JLab via York (AB, Barry Tan, Kayleigh Gates et al)**
  - Helicity asymm in $p\gamma \to p\pi^0$ https://link.springer.com/article/10.1140/epja/s10050-023-01123-3
  - Beam recoil in $K^+\Lambda$ production https://journals.aps.org/prc/pdf/10.1103/1q5m-x849
  - Beam asymmetry in π^0 and η production https://arxiv.org/pdf/1701.08123
  - Where's the data?! Need a resource like hepdata.net…

Summary

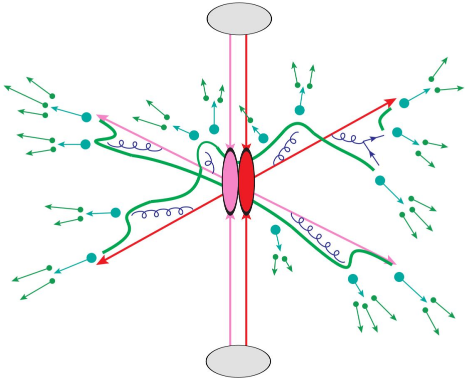
- Precision of particle-level MC generation has leapt forward through the HEP LHC era
- Getting the most from ePIC physics requires systematically embracing the modern tools, methods, and ecosystem
- But pp high-energy focus so far means work is needed for EIC
  - Validation against HERA inclusive physics
  - JLab etc. exclusive and semi-inclusive physics
  - RHIC etc. ion physics
  - Tuning of models, exploring and reducing uncertainties
- **Rivet toolkit from LHC is used by MCnet generator teams, has the features needed for ePIC validation and tuning**
  - Analyses and legacy porting underway. **Input on requirements needed. Get junior scientists/students enthused and involved**



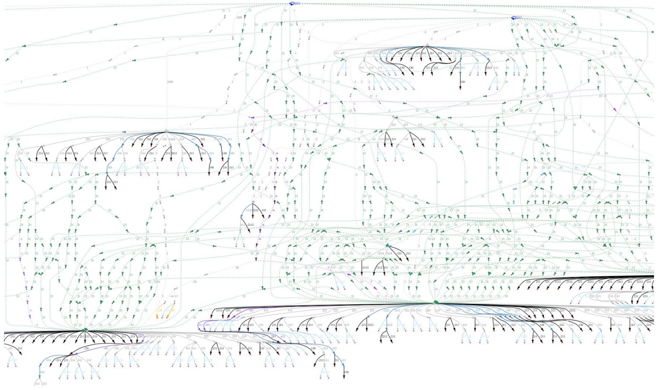
Backup slides

MC generation

- **MC generation is where theory meets experiment**
  - The fundamental *pp, pA, AA* collision, *sans* detector



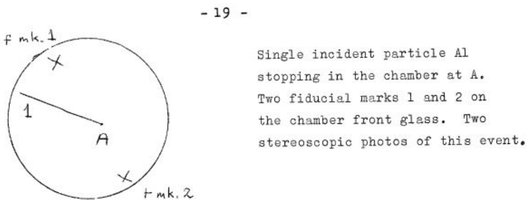
- **Components of an “exclusive” event-generator chain:**
  - QFT matrix element sampling at fixed-order in QCD
  - *Dressed* with approximate collinear splitting functions, iterated in factorised Markov-chain “parton showers”
  - FS parton evolution terminated at Q ~ 1 GeV: phenomenological hadronisation 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

Fiducial analysis

- **More simple & obvious consequences:**
  - "**Say what you see**": **don't report what you couldn't see!**
  - More specifically: *do* correct for detector biases, but **minimise extrapolations** beyond experiment acceptance
  - Done by aligning **unfolding target** (usually MC) definition with reco-level acceptances and selection cuts
  - Take "safe" shortcuts, e.g. use hadron decay histories in place of reco, but **don't rely on partons from interfering amplitudes**: hadronization is a decoherence barrier
  - Result is **"best estimate of what could be seen by a perfect detector"**: **don't fill unseen phase-space with model-dependent assumptions**



- **Analysis lifetime is maximised by not being model-specific**
  - E.g. HH-production signal-strength at HL-LHC has ~40% theory uncertainty from m_t scheme. No theory resolution in sight. But **fiducial cross-section is unaffected**

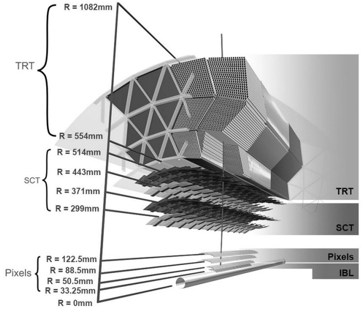
What is a final-state particle?

Remember…

We only “see” *stable final-state* particles:

- **electrons:** stable
- **muons:** stable (τ_0 = 2.2 μs, mean decay after ~ 1 km at 20 GeV!!)
- **taus:** unstable (τ_0 = 0.3 ps, mean decay after ~ 1 mm at 20 GeV!!)
- **neutrinos:** stable (but invisible)
- **photons:** stable
- **hadrons:** unstable→ more hadrons, photons, charged leptons & neutrinos
- **quarks and gluons:** unstable — are they even real?? → hadrons → jets
- **W, Z, Higgs, top:** unstable, varying degrees of objective reality!

Certainly FS needs to be post-hadronisation.
Best if visible to the detector: often use $c\tau_0$ > 10 mm.
Lots of discussion: ideas and issues are still evolving.

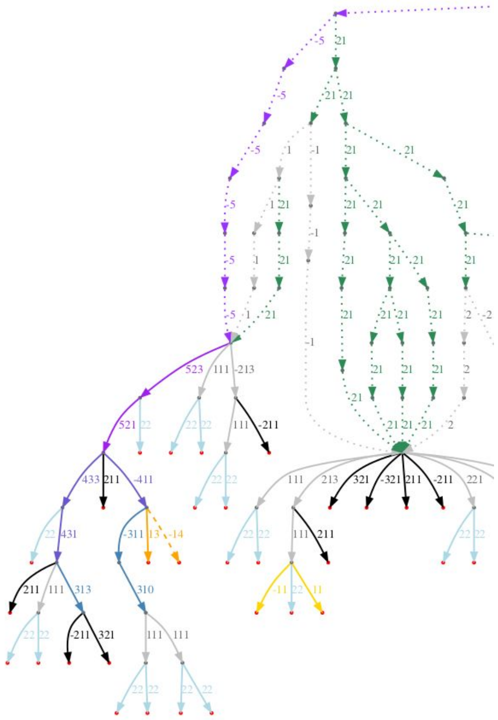


What is directness?

Not all final-state particles are equally significant:

- Hadrons are ten-a-penny! Even *c* and *b* hadrons can be produced fairly easily from $g \to bb$ splittings
- Similar for photons: a high-energy photon direct from the hard process is a significant EM-interaction event, but *lots* of low-energy ones from $\pi^0 \to \gamma\gamma$
- It's harder to make electrons and muons come out of hadron collisions: needs an internal electroweak process ⇒ weak coupling. But common in heavy-hadron semileptonic decays, with neutrinos.
- Leptonic taus are indistinguishable from $e/\mu+\nu$

- **⇒ we define *directness* "backwards": recursively eliminate anything from a hadron.** Only leptons and photons can be unambiguously direct

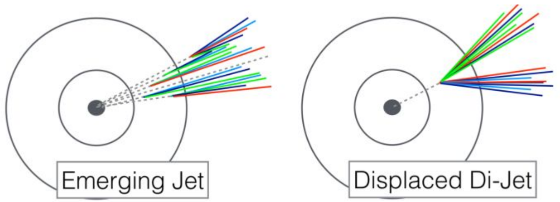


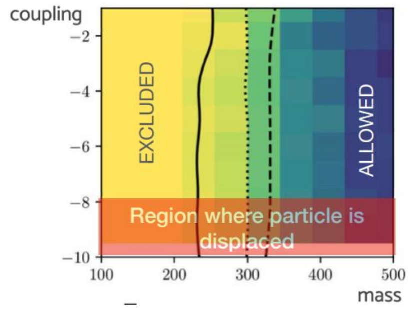
Often see "prompt" used to mean direct… but misleading since not a timing or displacement thing, unlike *b*-physics usage

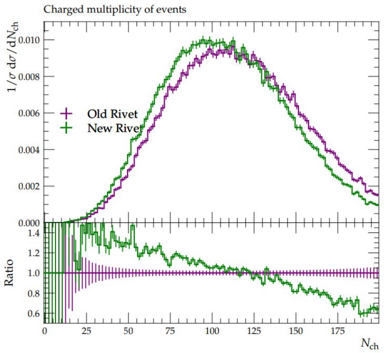
What is promptness?

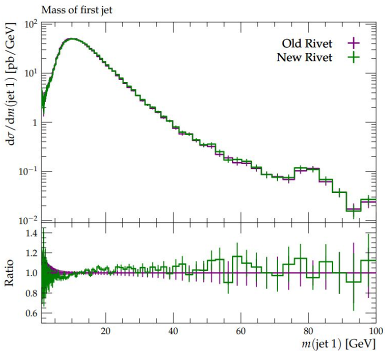
The LHC programme seeks/scrambles to cover more obscure signatures ⇒ **new focus on long-lived (BSM) particles: LLPs.**

- Lots of challenges in analysis, e.g. jet constituents have to be filtered by origin as well as momenta/ID
- Traditional **$c\tau_0$ definition of stability isn't enough…**
  - exponential distribution still peaks at zero; we care about **specific particles, not the species average**
  - species average requires the species to be known, and a **database lookup**… only SM by definition! What about BSM LLPs?
  - **Use $c\tau$ rather than $c\tau_0$ for fiducial FS?**
- Rivet mainly requires "correct" generator settings; v4.2 will finally implement a **lifetime-based final-state definition**. (Sophie McNeill, MSc/summer project)



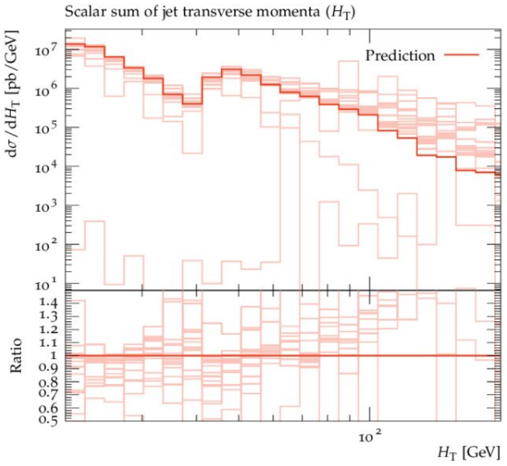


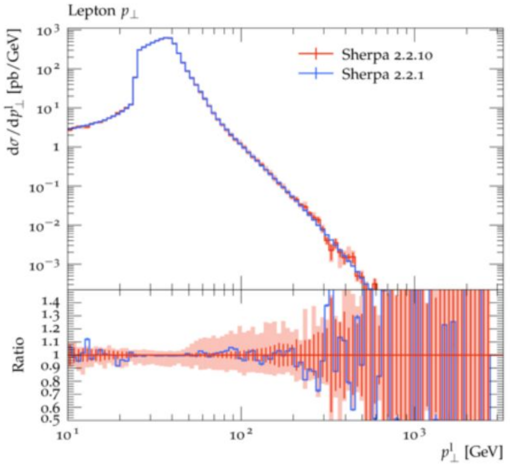




Rivet multiweights and re-entry

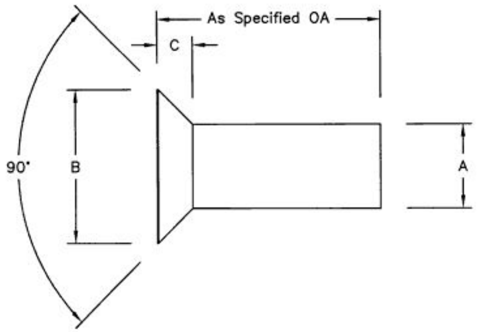
- **MC weight vectors allow expression of increasingly complex theory uncertainties.** For algorithms & observables that permit reweighting. But a burden for analysis chains: have to propagate and correctly combine O(200) weight streams!
- **Rivet 3: complex automatic handling of weights**
  ~invisible to users: data objects *look* like histograms etc. but are secretly multiplexed
- **Can re-call finalisation to perfectly 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, etc. etc.





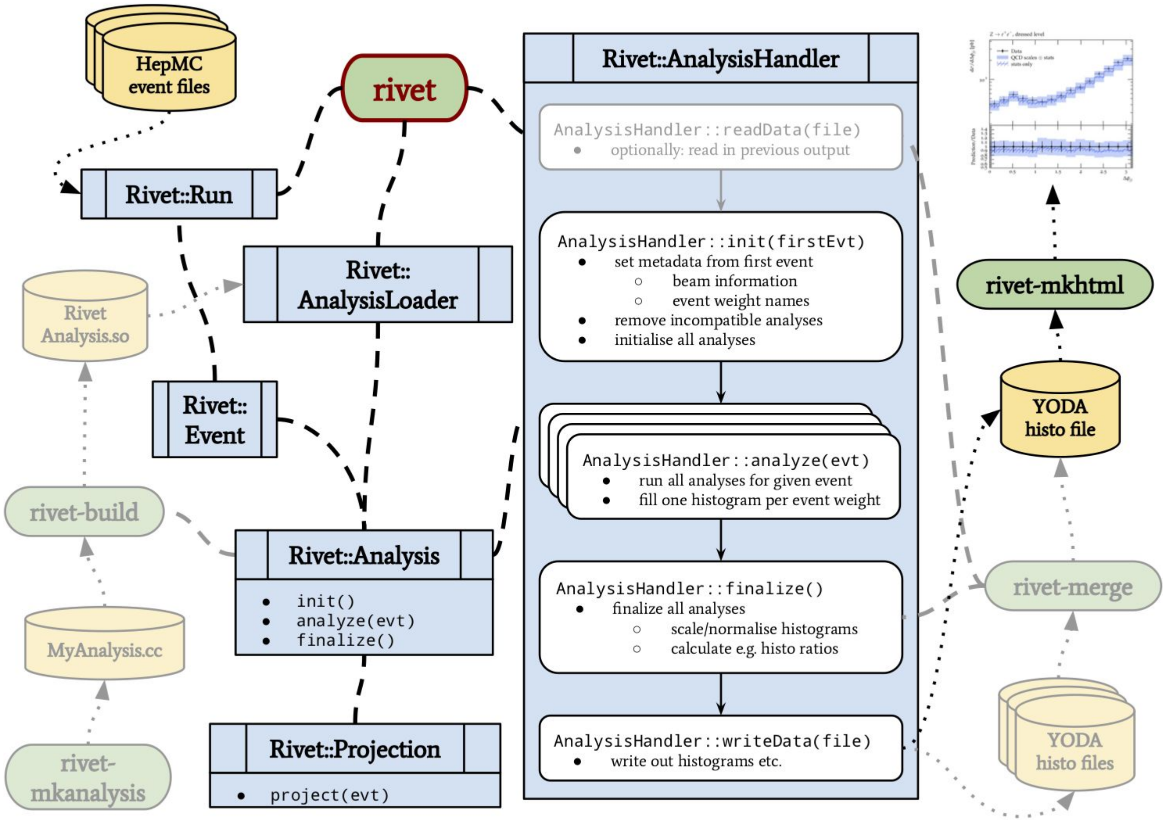
Designing Rivet

- **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**



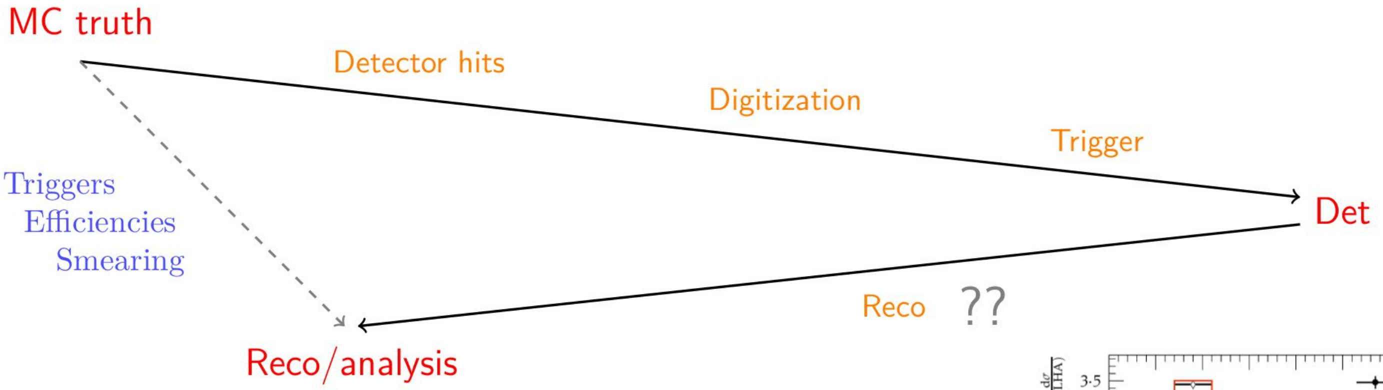
The result

- **Rivet v3 structure**
  arXiv:1912.05451
- Streamlined set of tools from analysis coding to event processing to plotting (and other applications)
- And a key gateway to connect data analysis to theory (and back again)

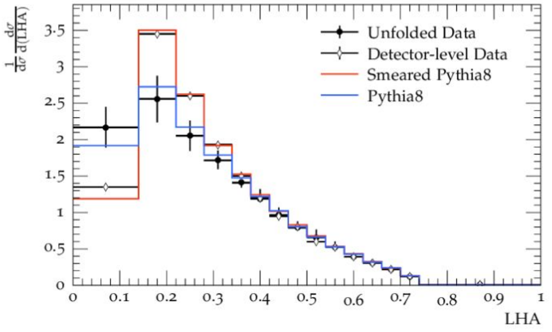


Detector emulation

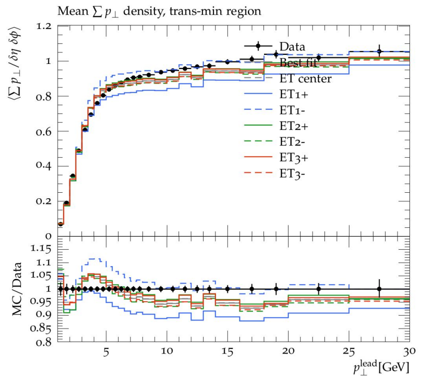
- "Detector smearing" is valid for many reco-level analyses (also in GAMBIT, MA5): reco is calibrated back toward MC truth, so go direct and skip the unknowns



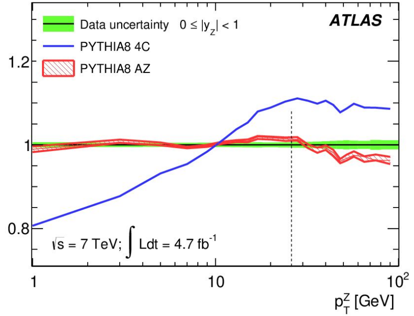
- capture key efficiencies cf. Delphes, but analysis-specific and less "simulation theatre"
- flexibility allows e.g. "tuned" jet-substructure smearing, systematics studies, whatever…



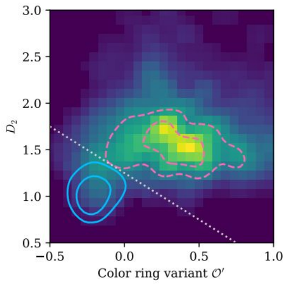
Applications: from tuning to BSM

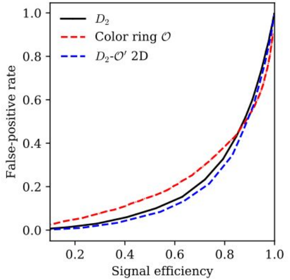


- **Pre-LHC huge QCD uncertainties: MC tuning via Rivet analyses**
- **Tunes revealed gaps in data and in modelling**
  - Better tunes ⇒ better analysis, better results ⇒ better MC
  - Impact: LEP and Tevatron analyses published for ~10 years suddenly got *used*! And cited…
    - ⇒ ATLAS and CMS tunes, tune uncertainties
    - ⇒ Rapid responses to preliminary data
  - Model development: matching & merging, addition of energy evolution & colour-reconnection to Herwig, …



- **Also very useful for prototyping and pheno**
  - Same features that made analyses quick to use for tuning also useful in analysis prototyping and model scans
  - And re-interpreting model-independent data on BSM

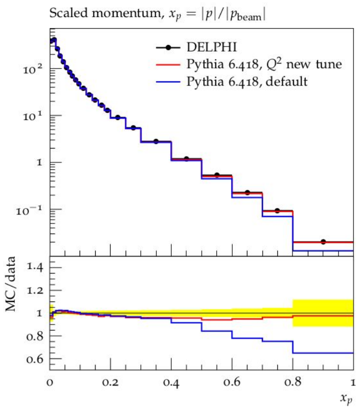


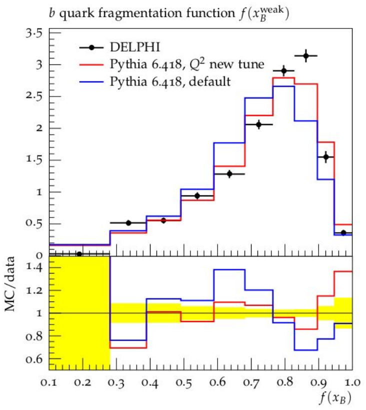


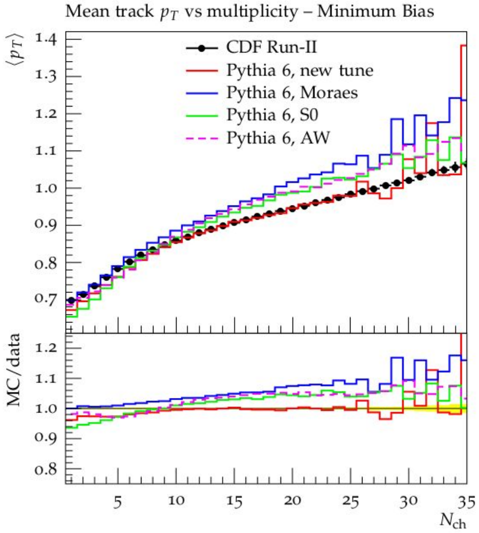
Event generator tuning

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 2008 "Professor" tunes, and many more since, including ATLAS A14 and A3**







The Professor method

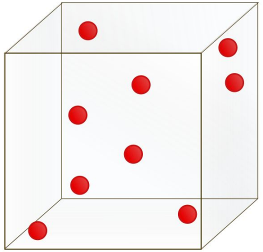
Tuning was historically brute force & inspiration

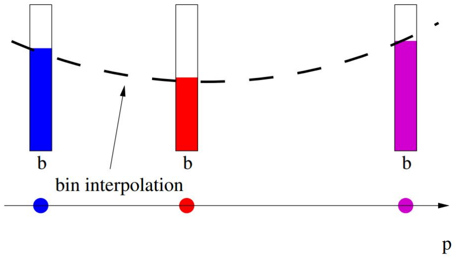
Professor method uses surrogates, to aid convergence:

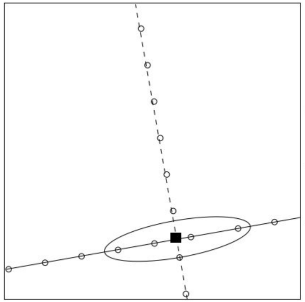
1. Sample (user-)param vectors $\boldsymbol{p}_n$ (from a hypercube/sphere)
2. Generate MC run-sets for beams, processes, etc. at each pt
3. Run *in parallel* on batch/grid facilities ⇒ output histos
4. Build *surrogate models* $\mathrm{bin}_b(\boldsymbol{p})$ from $\{\boldsymbol{p}\}$, e.g. conventionally 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 are still essential!

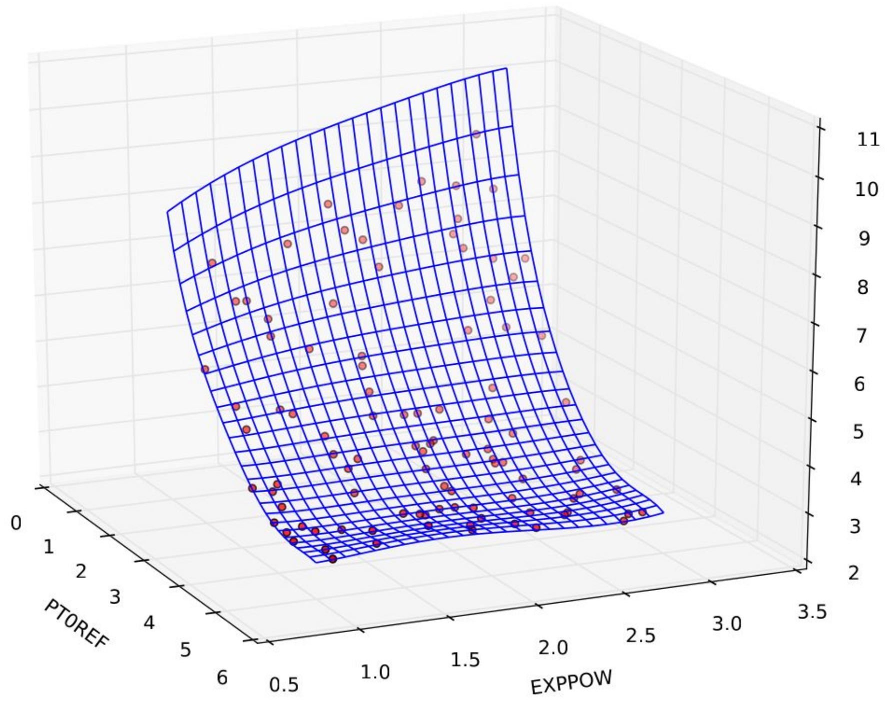
What about machine learning? Sure, fine. But if polynomials work — maybe via a change of variables — they are simple and robust. *Apprentice* also introduced rational approximants.

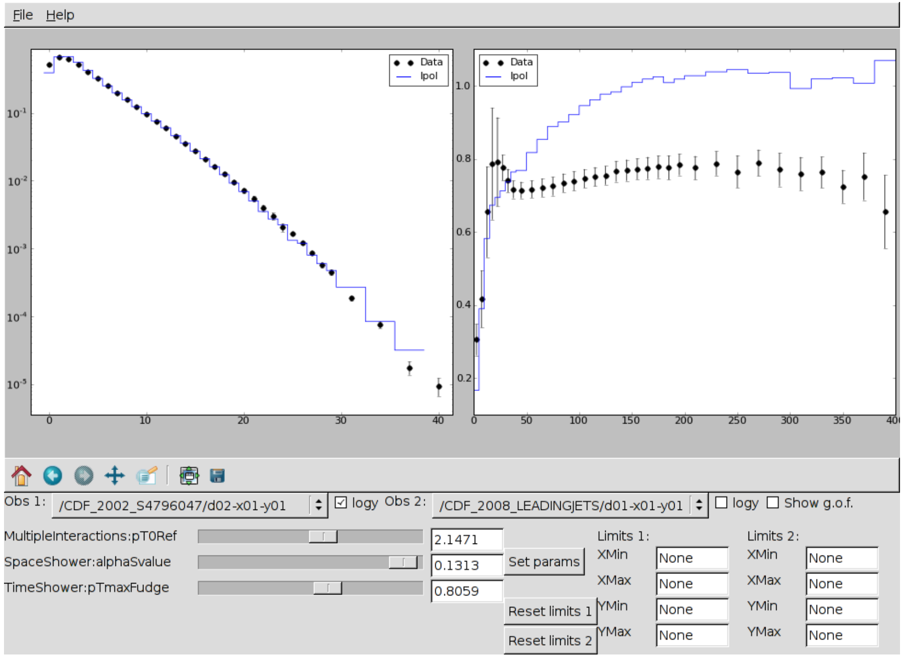




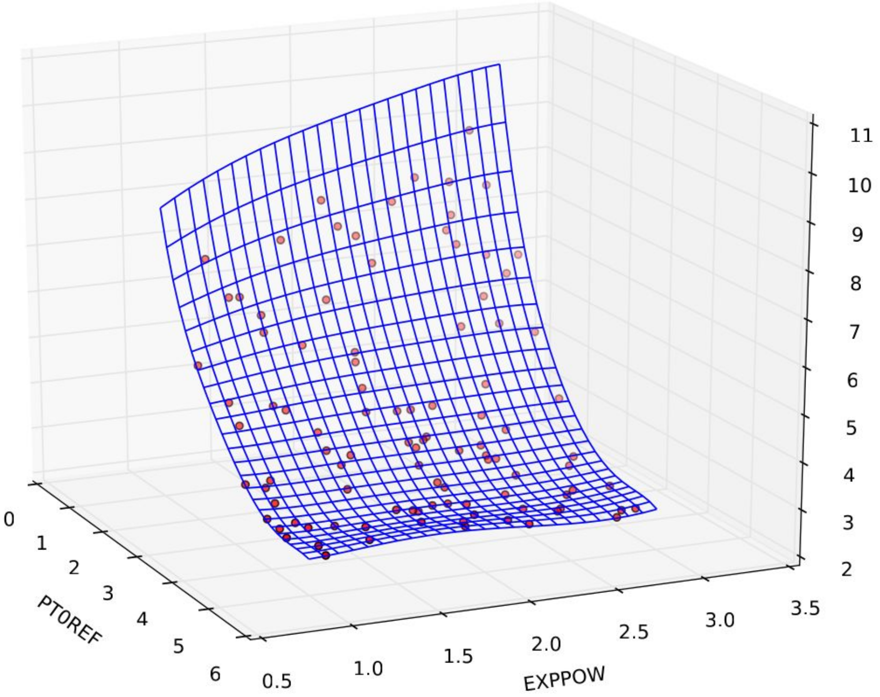


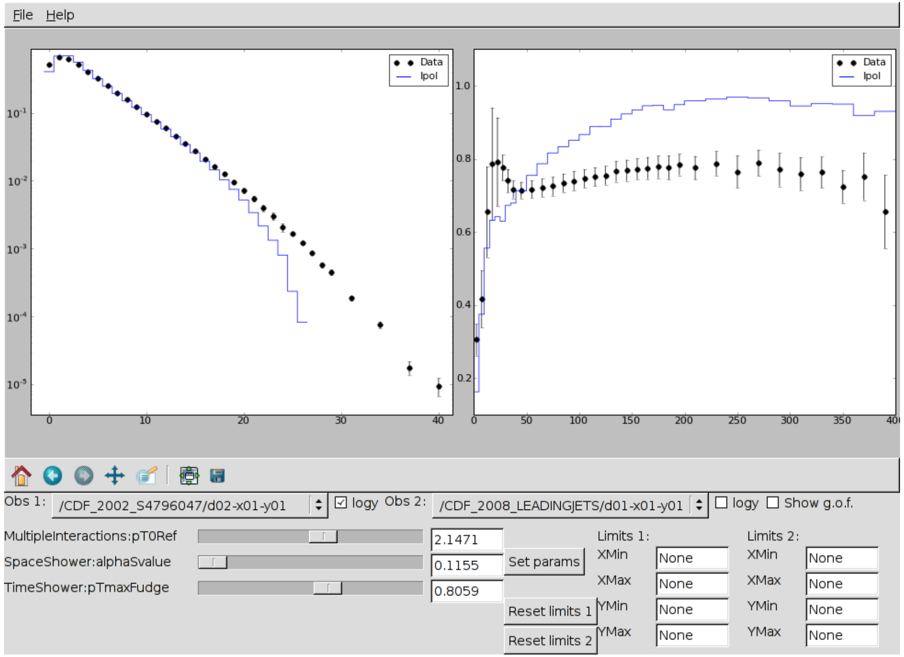
Professor highlights





Professor highlights





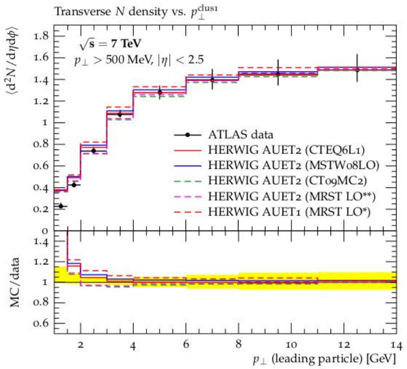
Tactics for tuning

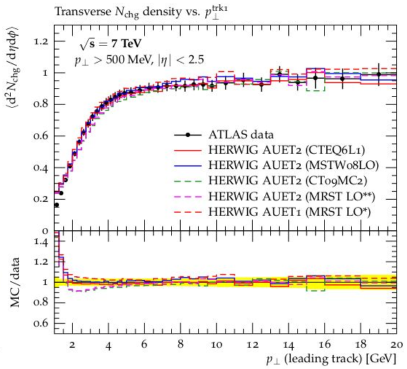
- **Factorise parameter space**
  - Historically split hadron flavours and spectra, jet structure, event topologies, underlying event. Max O(10)
  - Approximate but practical. Can also automate some estimate of factorisation groupings through mutual sensitivities
- **Weighting, observable balance, and 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*, empirical tolerance calculation, still room for obvious improvements
- **Future work**
  - Heavy flavour, matching/merging, including systematics via weights…

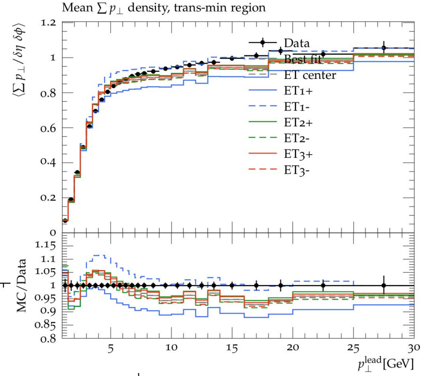
More tuning history

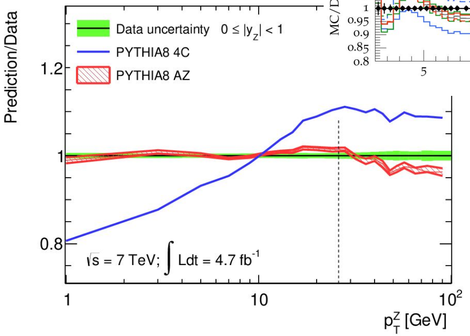
Pre-LHC, the soft QCD uncertainties were *huge*

- ❖ **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, …



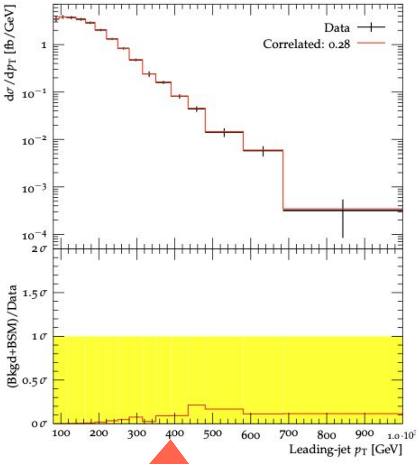




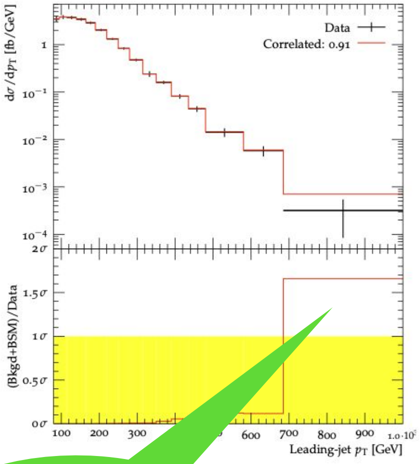


BSM from "Standard Model"

- **Particle-level measurements *can* achieve high model-independence**
  - Careful definition of fiducial cross-section, reduce model sensitivity in unfolding
- **E.g. Contur injects BSM signal into "SM" measurements**
  - Many models already "dead" before a dedicated search ⇒ save *years* of effort (cf. ATLAS EXO)
  - Particularly strong for measurements with complex signatures: mixtures of leptons, jets, MET, …
  - But even e.g. model-independent unfolded MET+jet has near-search power
- **All at truth-level ⇒ *SPEED!***



Signal would have small effect wrt uncertainties, can't exclude it (28 % CL)



Signal would have large effects wrt uncertainties: can exclude at high CL