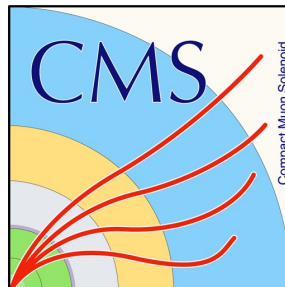


Generators, Tools and New Techniques in LHC experiments

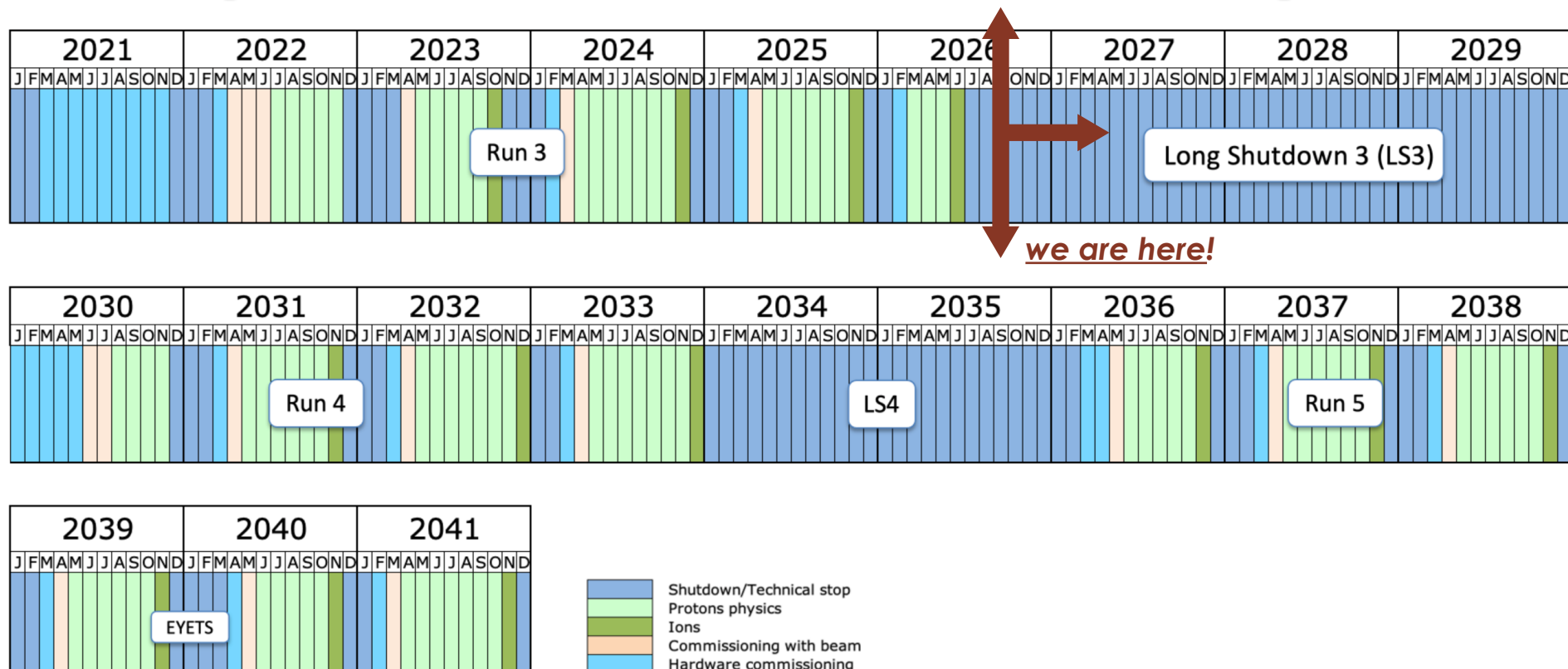
Francesco Giuli (Link Campus University and INFN Rome 2)

QCD@LHC26 (Oxford, UK)

30/09/2026



Improvements for Run 4 and beyond

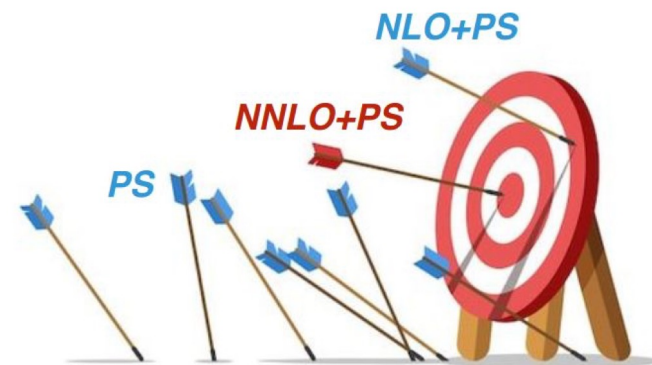


Last update: November 24

- **By summer 2030**, we need to be harmed with High-Lumi MC samples
- Expect *lots of developments* on this timescale

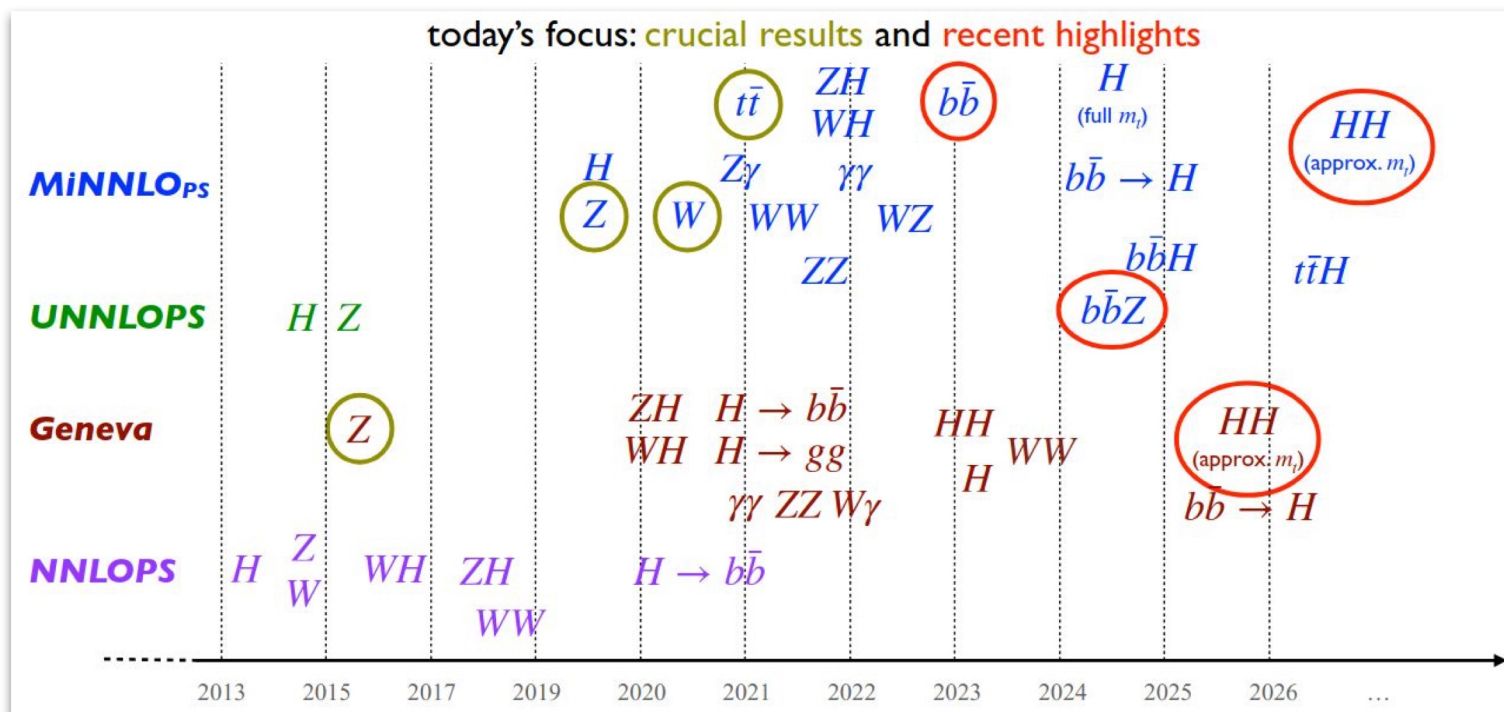
Contents

- **Improvements in perturbative accuracy**
 - NNLO pQCD Matrix-Element (ME)
 - Towards NLL Parton Shower (PS)
- **Addressing MC production bottlenecks**
 - Negative weight fraction reduction
 - On-the-fly reweighting
 - GPU/ML acceleration
 - Rethinking the MC request workflow
- **MC generator (mis)modelling** – impact on objects and analyses
- **Top-quark precision measurements** – bb4l, the "recoil" business
- **Improvements in non-perturbative (NP) modelling and uncertainty quantification**
 - New tuning methodologies offering principled parameter uncertainties
- **HL-LHC event projections and needs**



Improvements in perturbative accuracy

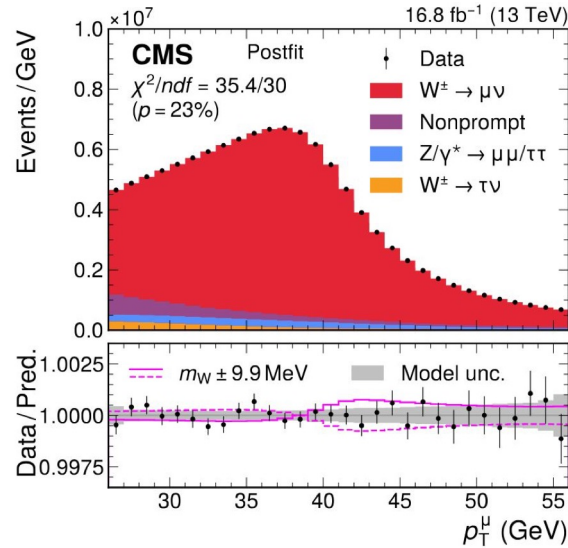
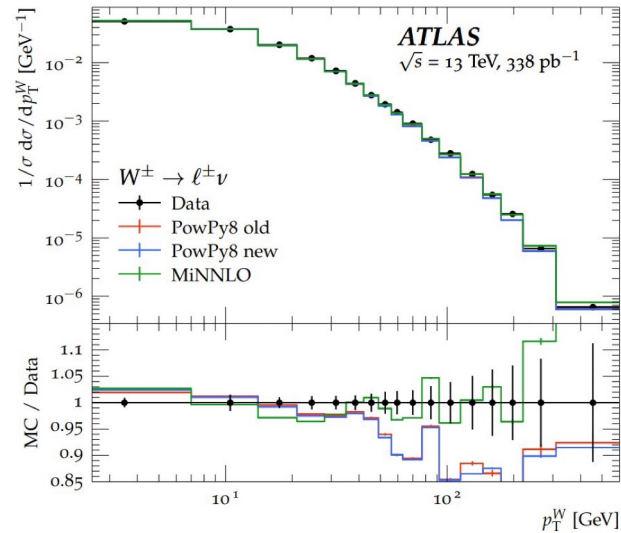
Towards NNLO with MiNNLOps



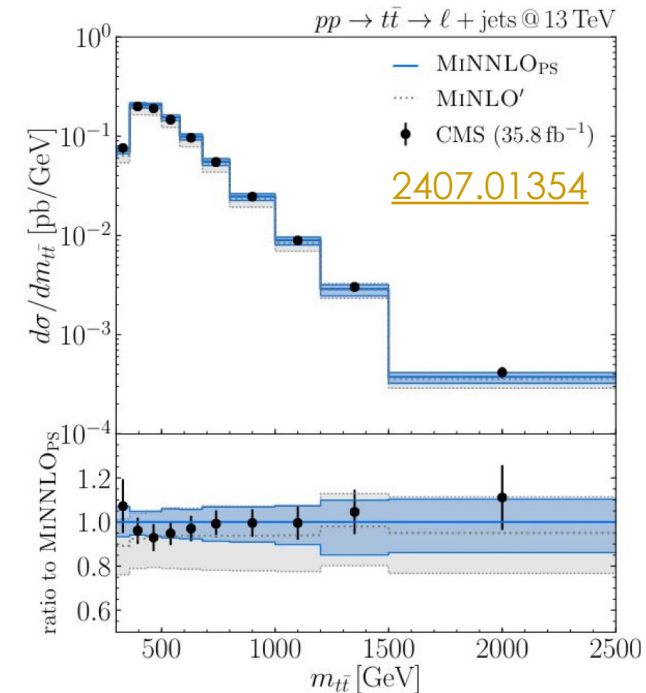
- Overview talk by [Marius Wiesemann](#), $b\bar{b}H$ talk by [Christian Biello](#)
- Ongoing developments:
 - $c\bar{c}, e^+e^- \rightarrow jj, b\bar{b}l\nu l\nu$ (off-shell $t\bar{t}$), W/Z + EW, triboson, H (full m_t, m_b, m_c), $b\bar{b}W$, ...
- MiNNLOps very slow with $O(100)$ PDF replicas – [slides](#)
 - Interested in solving this issue? ☺

Some MiNNLOps highlights

[2404.06204](#)

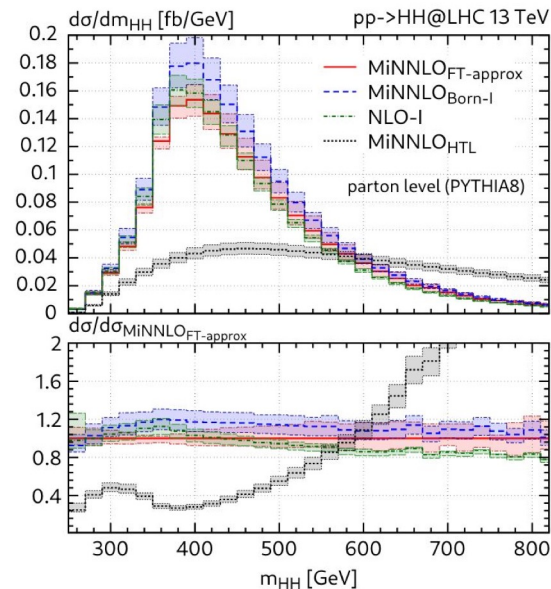
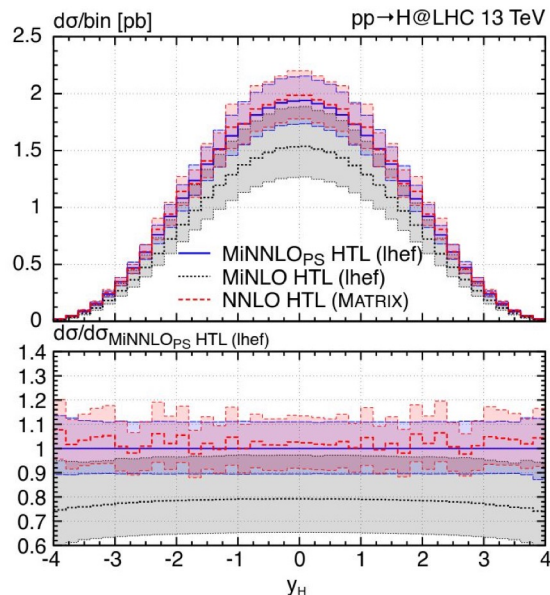


[2412.13872](#)



[2407.01354](#)

[2407.01354](#)

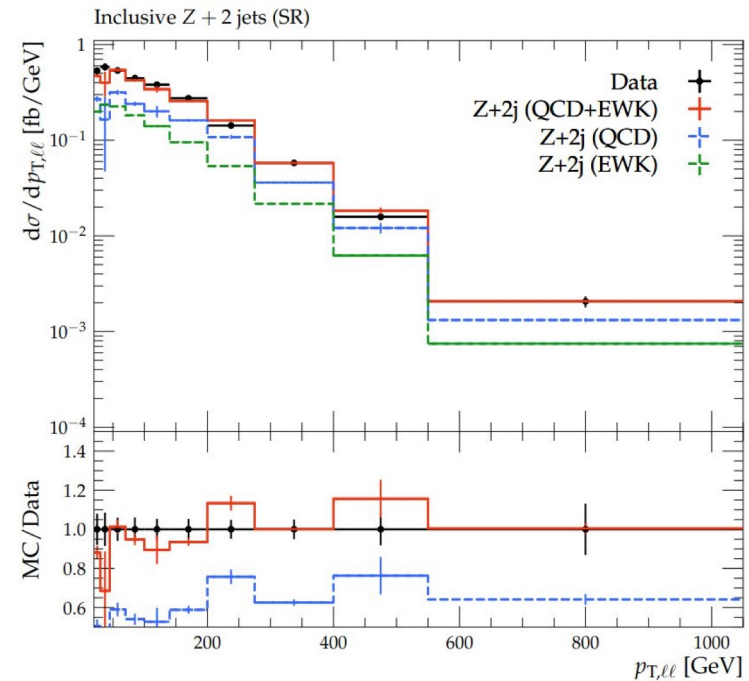
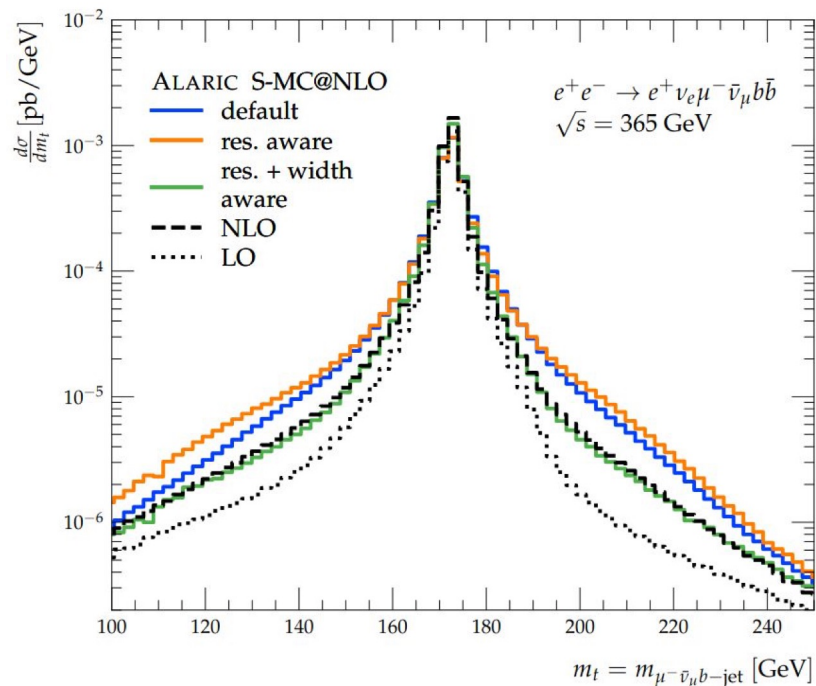


[ATLAS PUB note](#)

[2606.05300](#)

Sherpa 3

- Sherpa3.0.4 validated for use in ATLAS! 😊
- Overview talk by [Frank Krauss](#)
- New CCFM-style non-Sudakov factor for suppressing in the shower
- Visible improvements in V+jets m_{jj} , y_{jj} , p_T etc.



- Resonance and width-aware subtraction and showering in new ALARIC (NLL) PS
- Improved lineshapes
- Sherpa 3 for top modelling?

Future PS generators

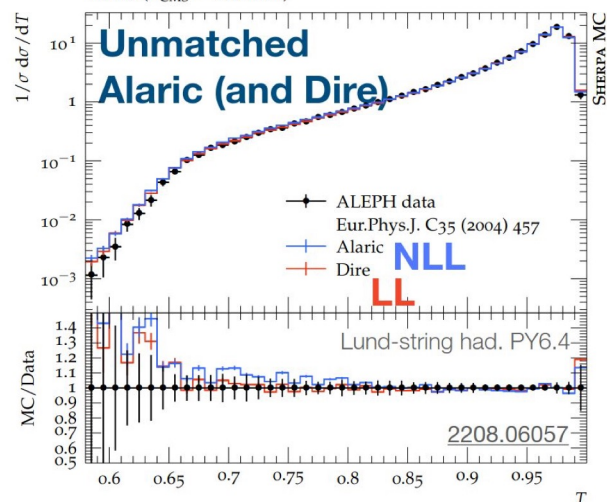
- Since PanScales [2002.11114], PS logarithmic accuracy is being discussed

	Alaric	Apollo	FHP	Panscales
Available for e+e-?	NLL	NLL	NLL	Partial NNLL
Available for pp?	NLL (no numerical proof beyond colour singlet)	No	No	NLL (no numerical proof beyond colour singlet)
Available for DIS?	No	No	No	NLL
Recoil distribution	Global for soft, local pT for collinear	As Alaric's soft map, different ord. variable	Global	Local (PanLocal) and global (PanGlobal) variants
Splitting functions at LO	Soft with special ref. direction, +coll remainder	Antenna functions	AP with sharp partitioning	AP with smooth partitioning
Spin correlations	Yes in e+e-	No	Yes in e+e-	Yes at NLL
Corrections beyond LC	No	No	No	Yes at NLL (NODS)
NLO matching available?	Yes in e+e-, using S-MC@NLO at exact colour	Yes in e+e- on paper, not yet implemented	Yes in e+e-	Yes for CS (dBNLO, P2B or positive weights ESME)
Multi-jet merging?	In e+e-: 5 jets (NLO), in pp: 3 jets (LO)	No	No	No
Particle masses?	Yes in e+e-	No	No	Yes in e+e-
Resonances?	Yes in e+e-	No	No	No
Publicly available?	Python codes, should become part of Sherpa	No, but should become part of Pythia8	No, but should become part of Herwig	Yes, interfaced to Pythia8

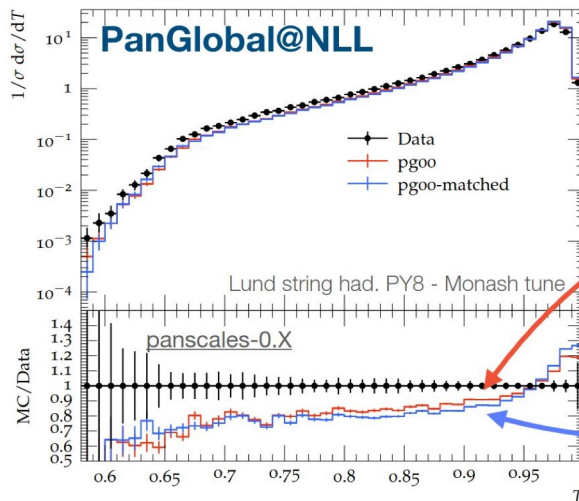
$$\Sigma(\tau) = \sigma_0 e^{-Lg_1(\alpha_s L) + g_2(\alpha_s L) + \alpha_s g_3(\alpha_s L)}$$

- Overview talk on next-gen PS by [Melissa van Beekveld](#)
- Differences in recoil schemes, probability distributions and evolution parameters

Thrust ($E_{\text{CMS}} = 91.2 \text{ GeV}$)



Thrust ($E_{\text{CMS}} = 91.2 \text{ GeV}$)



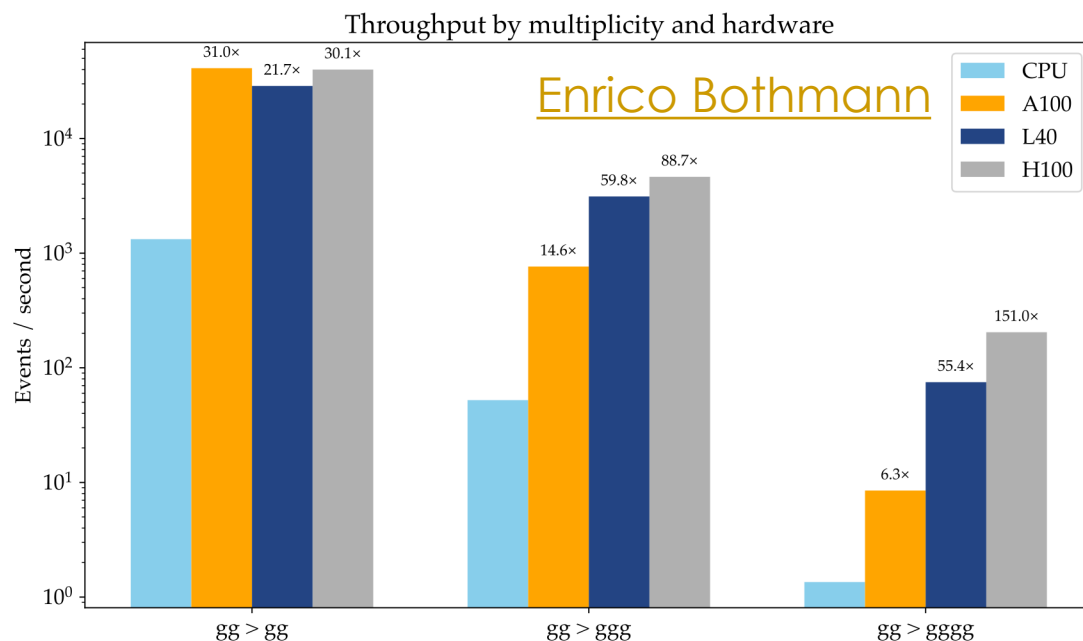
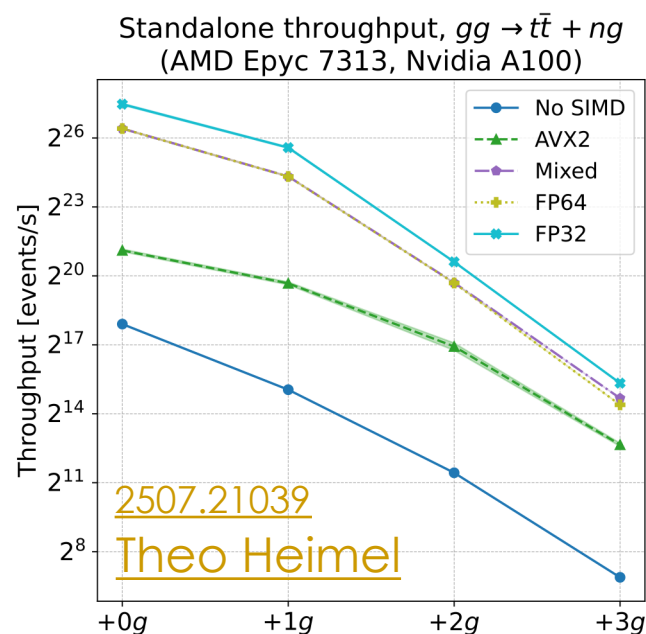
Many features still to be understood

Benchmark exercise
~soon to test them thoroughly

Addressing MC production bottlenecks

Heterogeneous Computing

- It is clear that **GPUs are of interest!** Several groups are looking into it (and vectorized instructions!)
- Performance has been demonstrated in simple, stand-alone setups by the authors



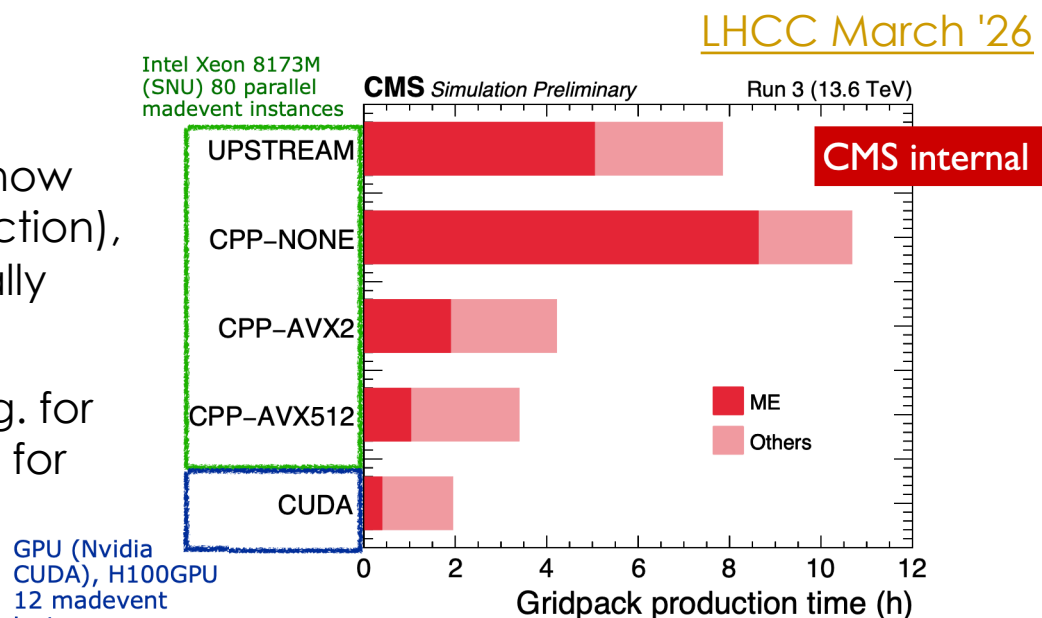
- Now we (experiments) have to show that in our production setups, with our settings, we can get similar results
- This is, as usual, **not trivial at all!** 😊

GPU status

- Ongoing studies for integration and testing of GPU-based Pepper/MG5_aMC
- All this needs to be validated on the Grid

- Eventually, we will need to know:

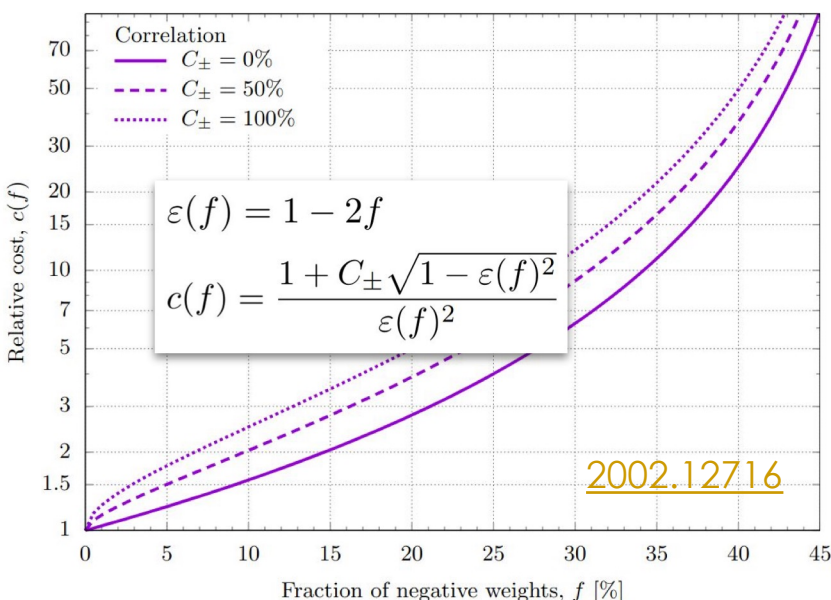
- What GPU queues look like (e.g. how many CPU threads/what GPU fraction), so that jobs can be set up optimally to use the resources
- How to deploy multiple builds (e.g. for Pepper, we need separate builds for different GPUs?)
- Whether we allow “weak” non-reproducibility (do we send jobs anywhere and run them as fast as they can, or direct jobs to specific resources?)



Negative weights

- Negative-weight events are **a killer!** 20% negative weights → 3x effective cost
- We have a bunch of ways to reduce negative weight events
 - Really need a systematic approach to these different methods
- Some come with substantial CPU costs
 - Need to make sure we don't make things worse
- A good, generic solution to this problem could bring down our HL-LHC CPU requirements substantially - **anyone want some everlasting glory?**

SIMD vectorization and GPUs - [2609.23500](#)



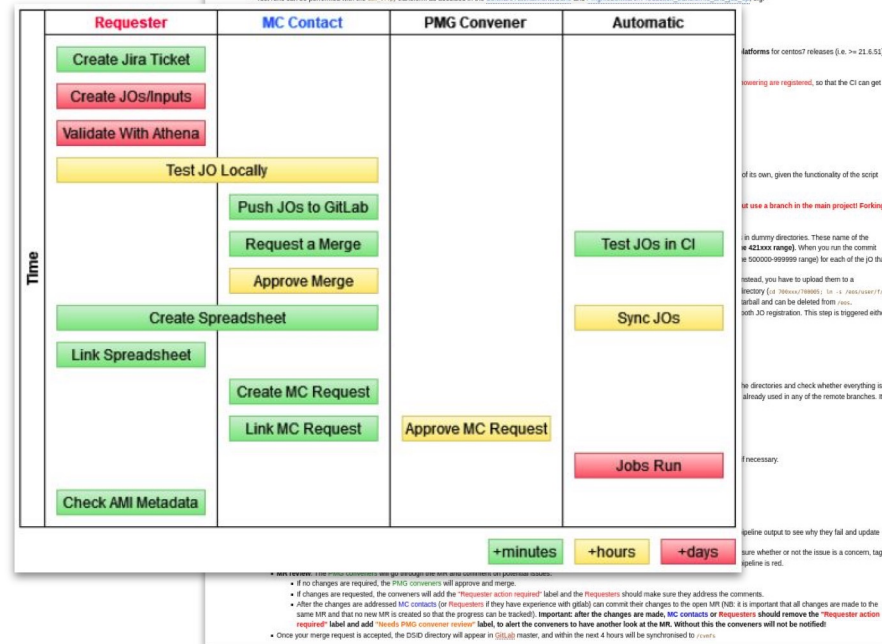
- **Folding**
reduces negative S-events, but at a computational cost
- **Born spreading**
reduces negative S-events, not as well as folding, but at low cost. Potential improvements with introducing Born-level kinematic dependence
- **MC@NLO-Δ**
reduces negative H-events, at a computational cost, and has different "matching shape" as compared to MC@NLO
- **MAcNLOPS**
removes negative H-events, at small cost, but back-and-forth between MG5aMC and Pythia; only proof-of-concept implementation

[Rikkert Frederix](#)

AI/ML in Generator World

- MC requests are substantial work for several people (very often human-time limited)
- Lots of little things to catch in review
 - When not caught in review, they hit the production system and cost us time
- Some automation integrated in the last few years (CI/CD for job options in particular)
- Opportunity for AI/Agents to “fix” the system
 - Already significant progress in the AI support direction by [MadAgents](#) for MG5_aMC - looks like this could be extended “easily”
- To do this, we need **central support!**
 - An open LLM API endpoint, ideally CERN hosted, so that we aren't paying €/CHF
 - [GitLab Duo](#) at 0.25€/MR already cheaper than human effort (~20€/MR, ~0.15 FTE)

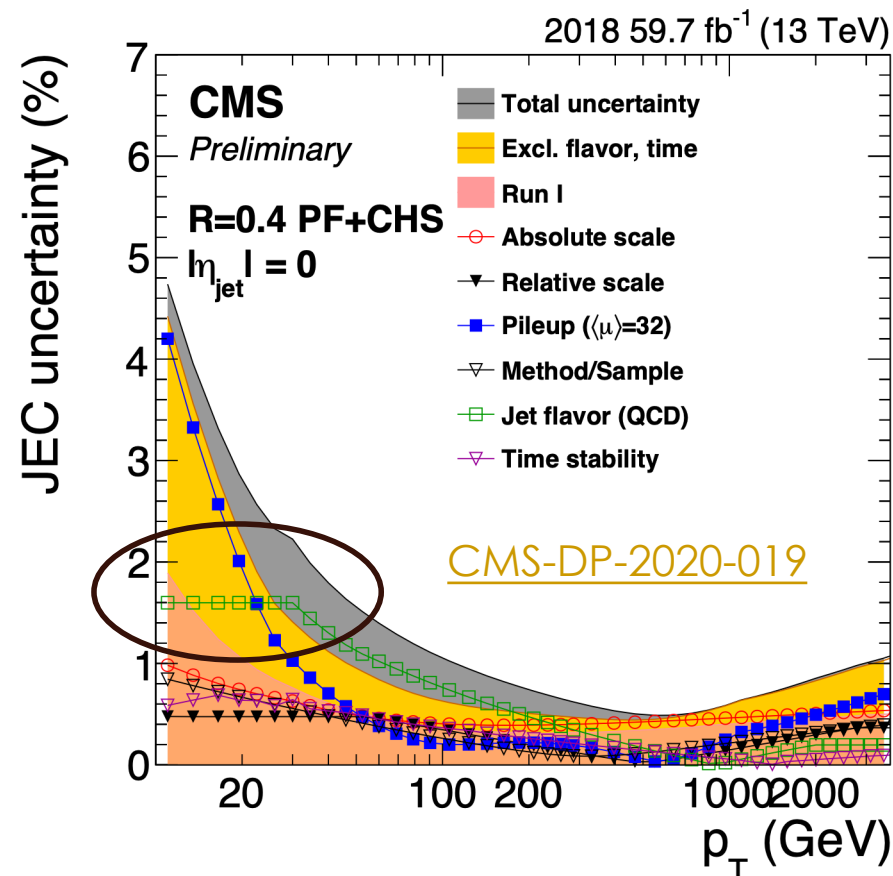
Twiki



MC generator (mis)modelling

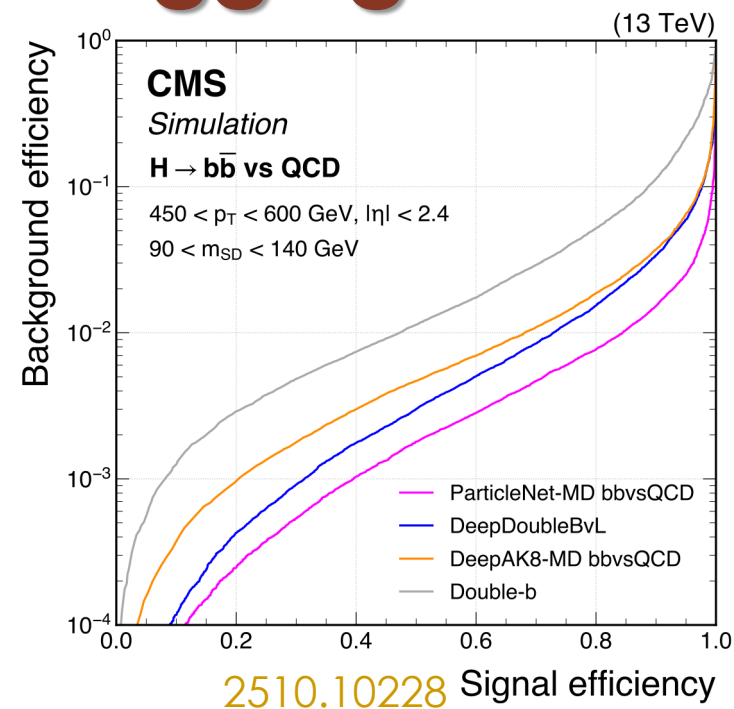
Jet Flavour uncertainties

- Some of these uncertainties are driven by MC simulations
- **Jet flavour uncertainty:** based on the difference between Pythia8 and Herwig in the gluon and quark jet response
- Note: a new approach was discussed at the [2023 Les Houches Workshop](#)
- Change from 2-points systematic to a matrix of systematics, covering the possible generator differences
 - PS models
 - Hadronisation
 - etc...
- Partially implemented by ATLAS already!



Boosted flavour tagging

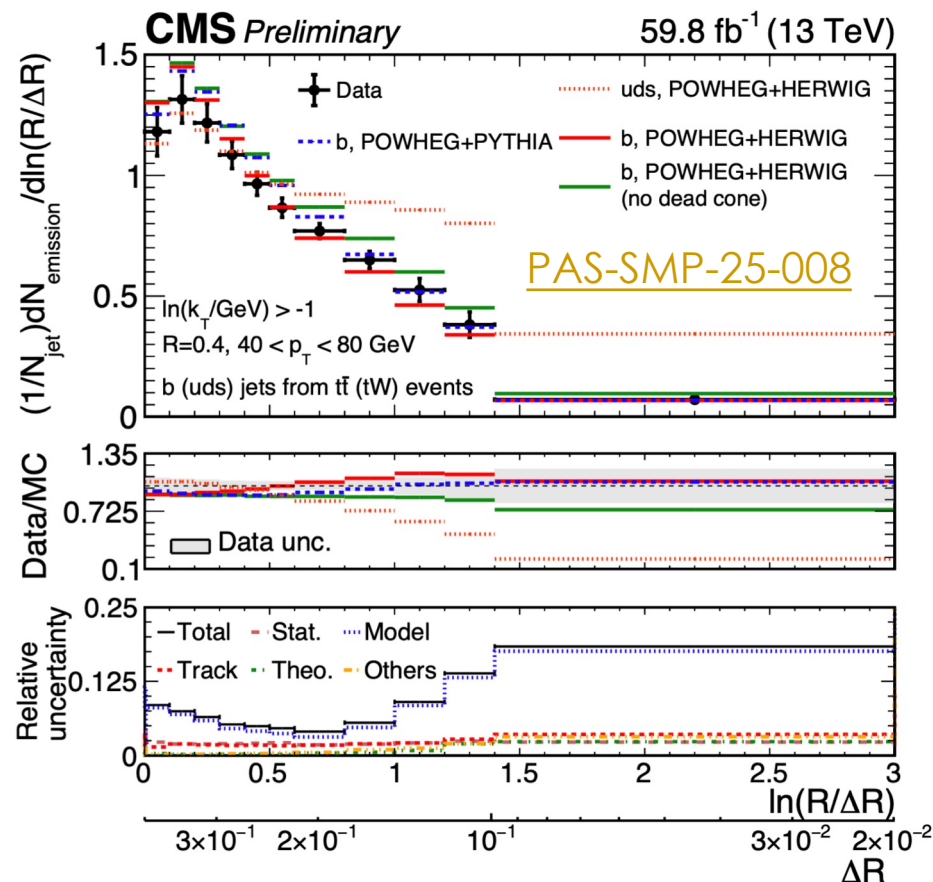
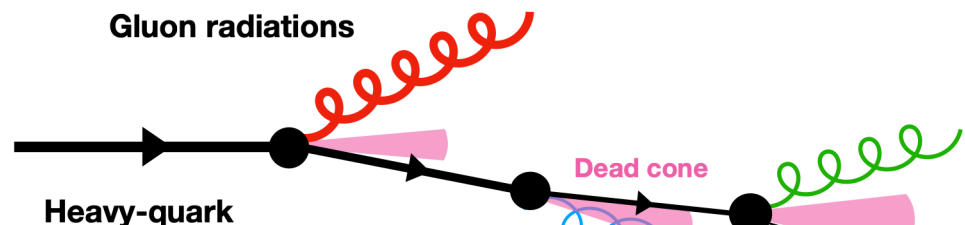
- Recent DNN or Transformer-based flavour taggers show impressive performance in i.e. the identification of boosted $X \rightarrow b\bar{b}$ events vs QCD bkg
- It is essential to calibrate the tagging efficiency using data, to avoid large uncertainties due to hadronization and PS models
- Still, **theory uncertainties** remains **sizeable** in the SF computation
- Obtained varying the b-, c- and light-jets fractions in Pythia8 simulation ←



Uncertainty source	$\langle \Delta SF \rangle$
Statistical	0.063
Theory	
Fraction of jet flavours	0.039
ISR and FSR in parton shower	0.014
Experimental	
Effect of varying sfBDT thresholds	0.048
Effect of applying "reweighting schemes"	0.091
Jet energy scale and resolution	0.008
Integrated luminosity	0.001
Pileup reweighting	0.009

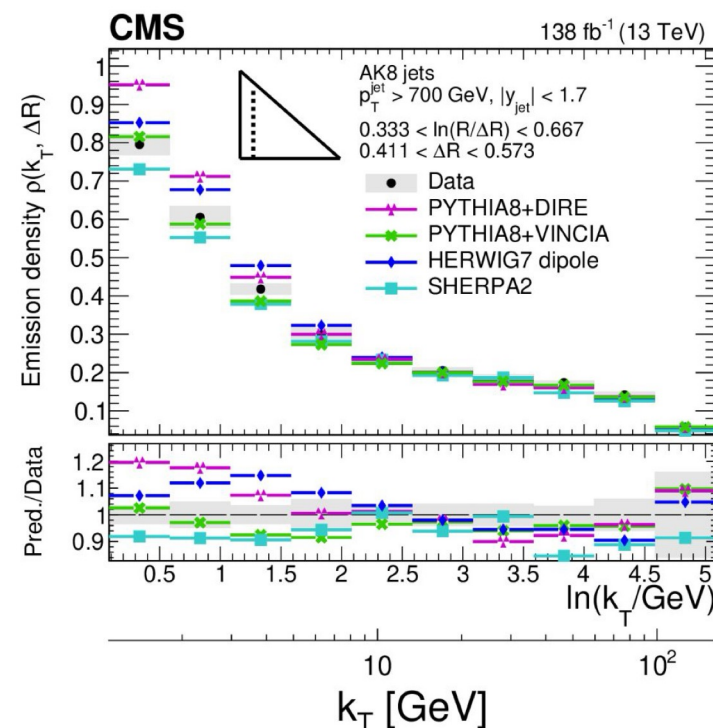
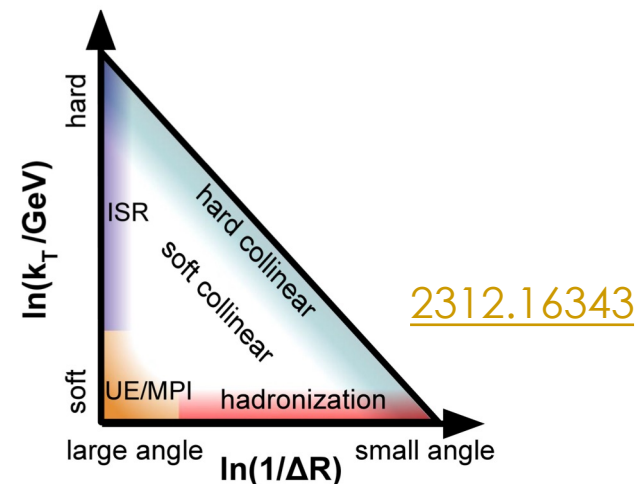
Dead-cone effects in b-jets

- Describes the suppression of collinear gluon emissions inside a cone of $\vartheta_0 \sim m_q/E_q$ around a heavy quark
- Directly related to the harder fragmentation of heavy quarks
- Observed by ALICE back in 2021
- The massless sample isolates the perturbative dead-cone effect from hadronization differences between light and heavy quarks
- Observed a **4.3 σ deviation** to the massless sample (but agrees with the nominal sample)
- Modelling uncertainties dominates



Lund Jet Plane measurements

- Lund Jet Plane measurements are key to understand the modelling of our generators for jet substructure observables
- **Mismodelling** is evident in **the low k_T and large ΔR regions**, particularly sensitive to UE/MPI and hadronization effects
- **Different dipole shower implementations** have a significant impact
- **Generator tunes** also have a large impact
- Particularly sensitive to **different $\alpha_s(m_Z)$ FSR** values



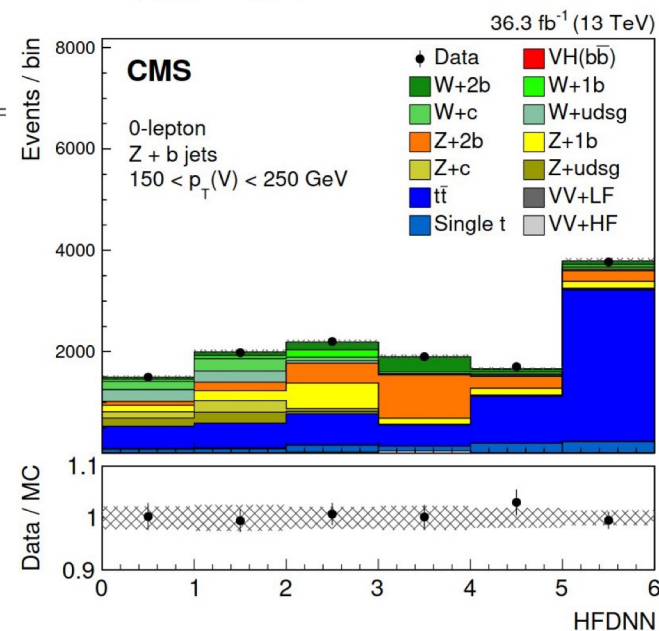
Higgs analysis: $VH \rightarrow bb/cc$

- V+Heavy Flavour jets main irreducible background for $VH \rightarrow bb/cc$ analyses
- **MC modelling** is a **real challenge**, requiring several data-driven corrections
- LO multijets (up to 4) and NLO (up to 3) MG5_aMC@NLO samples represent the baseline in ATLAS/CMS but:
- **Know issue** with the **modelling of the low $\Delta R(bb)$** region for the NLO FxFx sample
- **Negative weights** in the NLO sample reduce the effective statistics
- LO samples **less accurate** and **reweighting** to higher order predictions needed

	$\Delta\mu$
Background (theory)	+0.043 – 0.043
Signal (theory)	+0.088 – 0.059
MC sample size	+0.078 – 0.078
Simulation modeling	+0.059 – 0.059
b tagging	+0.050 – 0.046
Jet energy resolution	+0.036 – 0.028
Int. luminosity	+0.032 – 0.027
Jet energy scale	
Lepton ident.	
Trigger ($\vec{p}_T^{\text{miss}}$)	

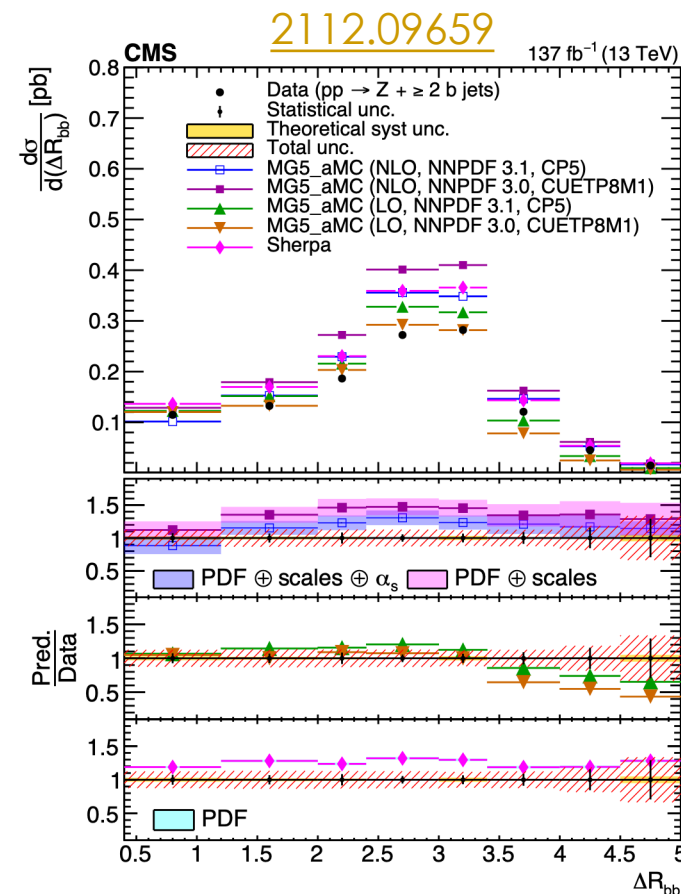
V+jets MC modelling uncertainties have a large impact on the measurement

2026 ATLAS-CMS FTAG workshop



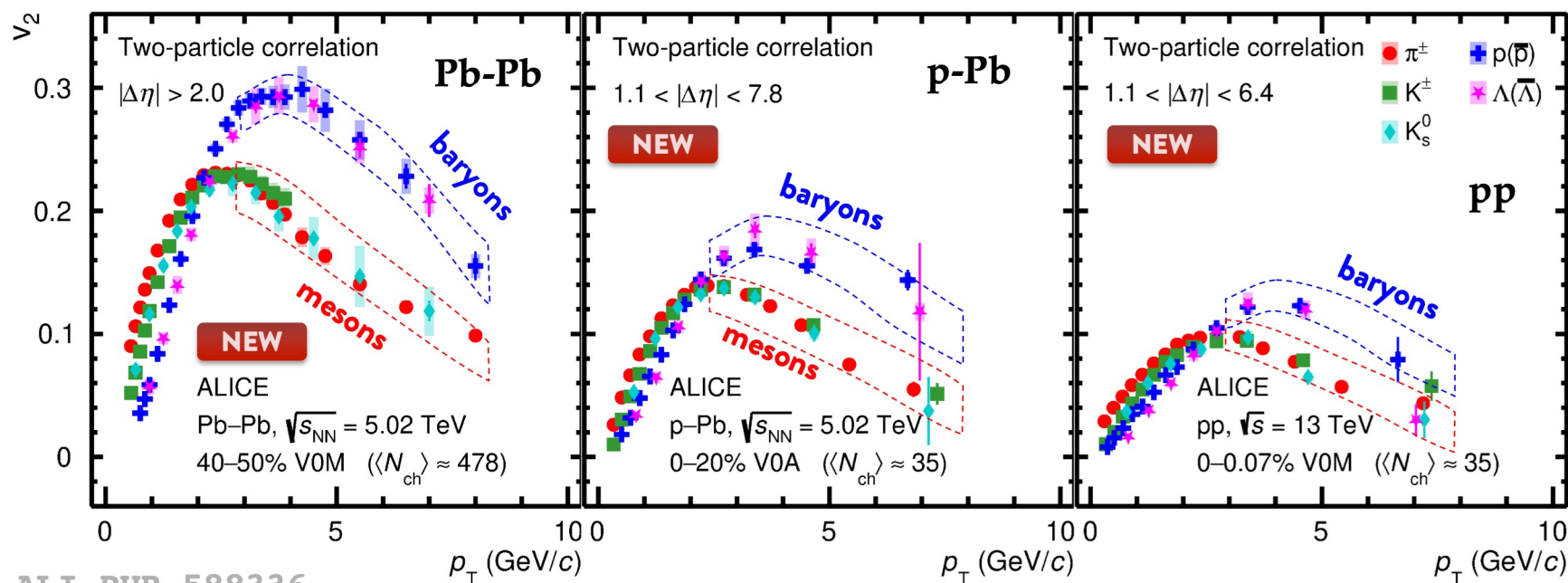
Gluon splitting into HF quarks

- A great deal of work devoted to HF jet algorithms that work at NNLO
- Sizeable difference between the HF jet algorithms applied at the LHC and the theorist algorithms – workshop
- **Soft HQ** can lead to **hard jets being tagged as HF**
- NNLO predictions have to be corrected for this effect – currently done using PS MC
 - Gluon splitting to heavy quarks in the PS underestimates the rate in the data
- MEC in the PS MC may correct the rate
- Possible improvements:
 - Use of NLO HQ ME from OpenLoops into the Sherpa MEPS framework
 - Have a look at HQ splittings in the LJP, understanding the bad-modelled phase space
- Dedicated session during the upcoming LHC MC WG plenary meeting



Elliptic flow

2411.09323



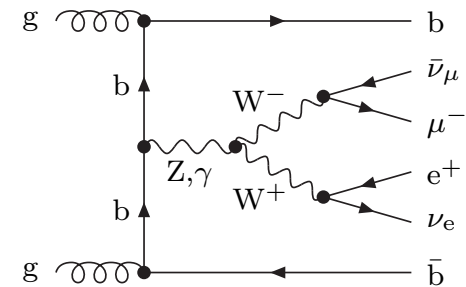
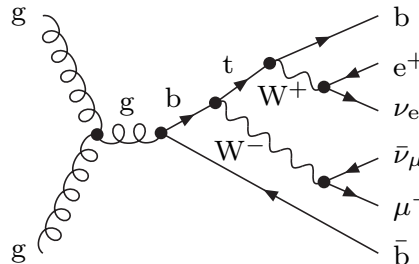
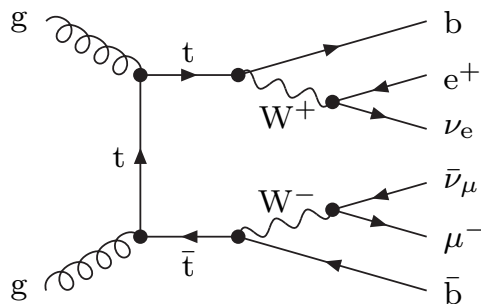
ALI-PUB-588336

- Stress test for partonic collectivity in small systems
- **Baryon/meson** grouping also observed in Pb-Pb, high-multiplicity p-Pb and pp
 - Connection to geometry less clear in pp and p-Pb: sub-nucleonic structure?
 - What works: can be described via phenomenology of hydrodynamics and coalescence
 - What doesn't work (yet): open challenge to QCD-inspired models such as PYTHIA

Top-quark precision measurements

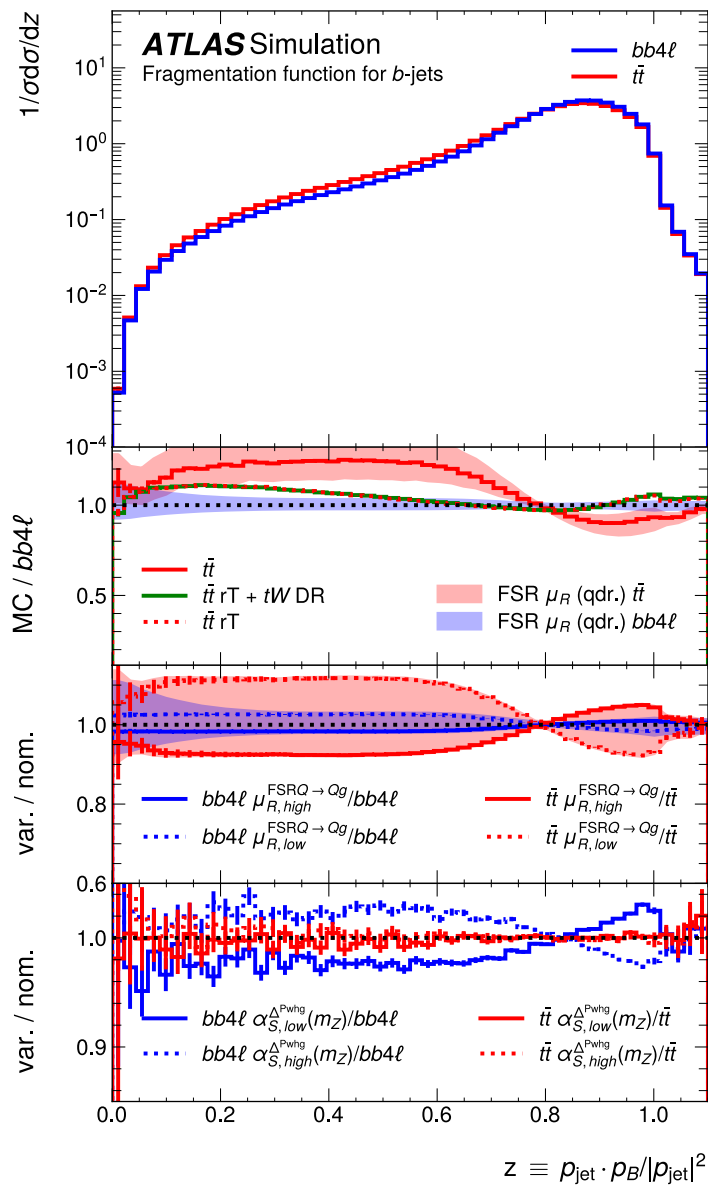
bb4l

- Process implemented in POWHEG-BOX-RES
- Describes the $b\bar{b}l\bar{l}'\bar{\nu}_l\nu_{l'}$ setup and can be interfaced with a PS
- Includes double-, single- and non-resonant diagrams (and corresponding NLO QCD corrections)

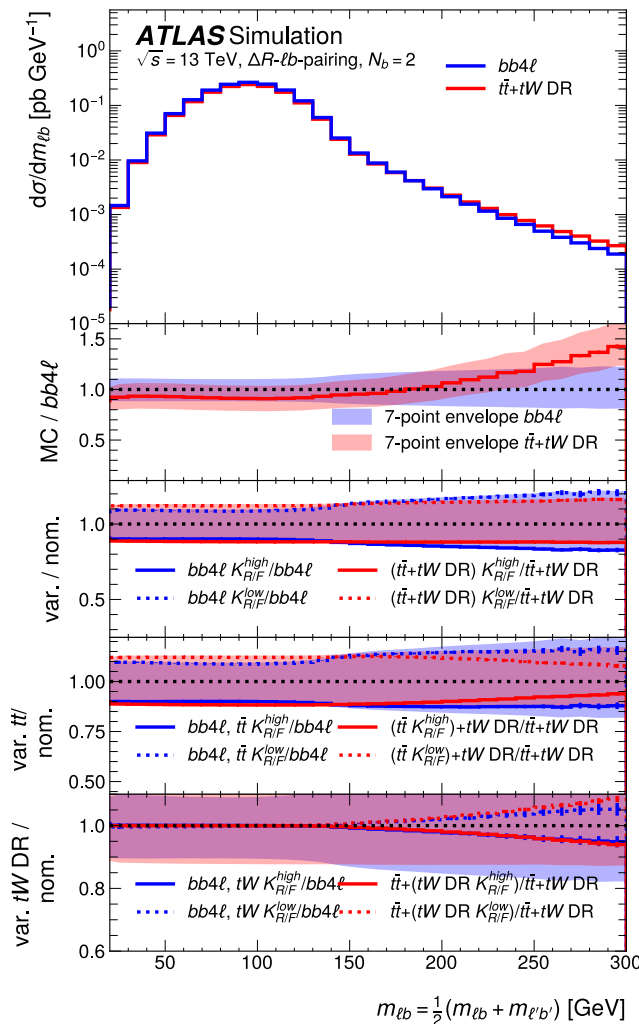


- Dedicated ATLAS PUB note: [ATL-PHYS-PUB-2025-036](#)
 - Example distributions for modelling uncertainties reduction bb4l and $t\bar{t} + tW$
 - Dedicated matching in bb4l and related uncertainties
 - Uncertainties influenced by NLO top-quark description in bb4l
 - Powheg hard process generation: resonance histories, ME scale, width corrections

FSR μ_R , α_s and scale variations



- Larger variations in $t\bar{t} + \text{PS}$
- No α_s variations in $t\bar{t} + \text{PS}$, but in $bb4l + \text{PS}$



Good agreement
between ME
scale uncertainty

Slightly larger
 $bb4l$ uncertainties
in the tail

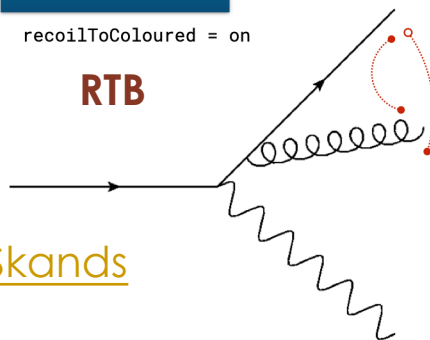
Recoil in $t\bar{t}$ events

- Uncertainty related to how the recoil for the 2nd gluon emission is handled in dipole shower introduced in recent ATLAS precision top measurements

PYTHIA

recoilToColoured = on

RTB



Peter Skands

$g - t$ dipole treated as
 $g - b$:

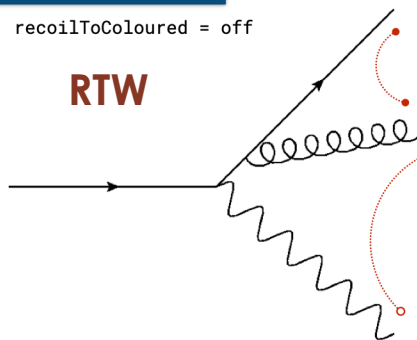
Phase space & recoils set by b

Affects b fragmentation

PYTHIA

recoilToColoured = off

RTW



$g - t$ dipole treated as
 $g - W$:

Phase space & recoils set by W

b fragmentation more "normal"?

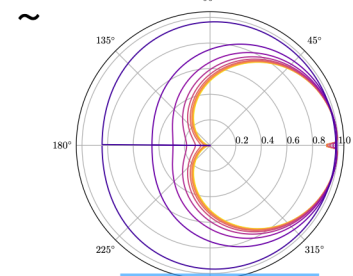
recoilToTop

UserHook

Suppresses radiation
in W hemisphere

RTT

⊗ Correction factor



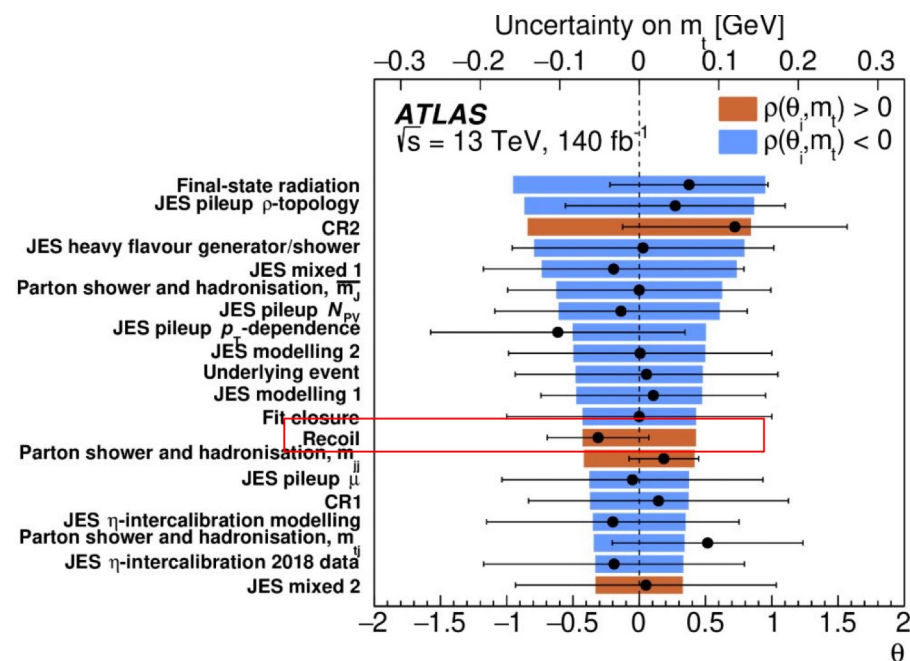
~equivalent

- Result in different momentum reassignment/phase space coverage
- Effect on W and b-jet p_T , massive impact on b-hadron kinematics (via fragmentation)
- Present only in top-events → cannot be tuned @LEP ☹

ATLAS PUB note

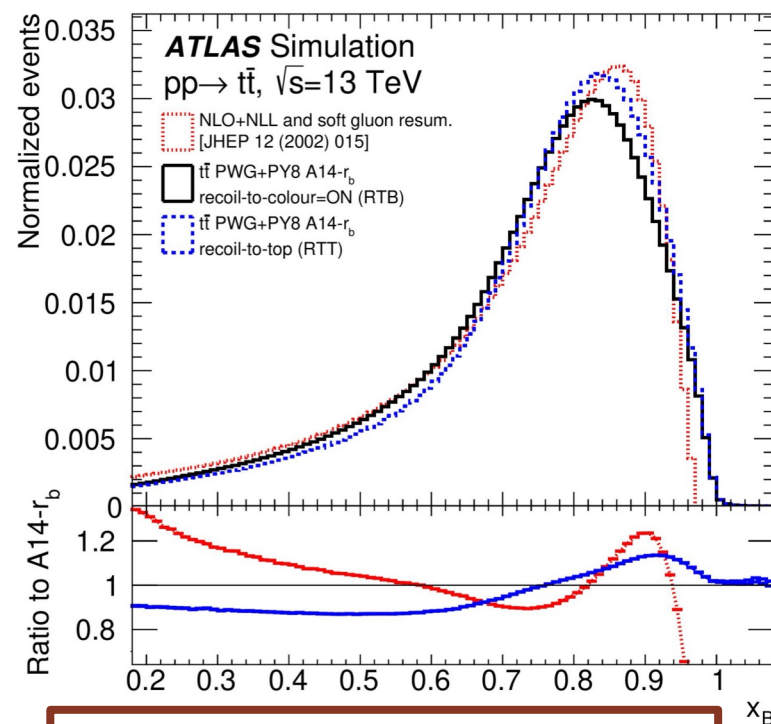
How it is treated in different generators

- Dipole showers: Pythia8 uses RTB, while Herwig and Sherpa RTW
- Angular showers: no ambiguity, recoil collectively shared by the whole system (similar to RTW in practice)
- VINCIA: no ambiguity, dipole-antenna shower
- **Limiting factor** for present/future precision top measurements
 - Top mass with SMT: 250 MeV (RTB vs RTT but with b-fragmentation reweighting)
 - Top mass boosted, large R-jets: 70 MeV (constrained by additional observable $m_{t\bar{t}}$ in the LH fit)
 - Top mass using $J/\psi \rightarrow \mu\mu$: 1.07 GeV (out-of-the-box)
- Currently a huge, analysis-dependent effort required to constraint its impact
- We would like to find a **common approach**, together with the theory community, to remove this limiting factor for top measurements



What can we do?

- Use **data** to constrain the uncertainty → can we remove this 2-points uncertainty?
 - Jet shapes, substructure, other observables i.e. colour-flow sensitive ones → no smoking guns, very little sensitivity
 - In-situ constraint via additional observable
 - Performing a dedicated measurement with a newly designed observable - **any suggestion?**
- Use **theory** to constrain the uncertainty → can we rule out one of the two variations?
 - NLO+NLL predictions available: significant differences in shape, but Mellin moments in better agreement with RTB
 - Now we have also NNLO predictions: some work performed on them for but nothing conclusive
- Use a **“better” MC generator**:
 - VINCIA could have been a solution, but it is not a short/medium term solution

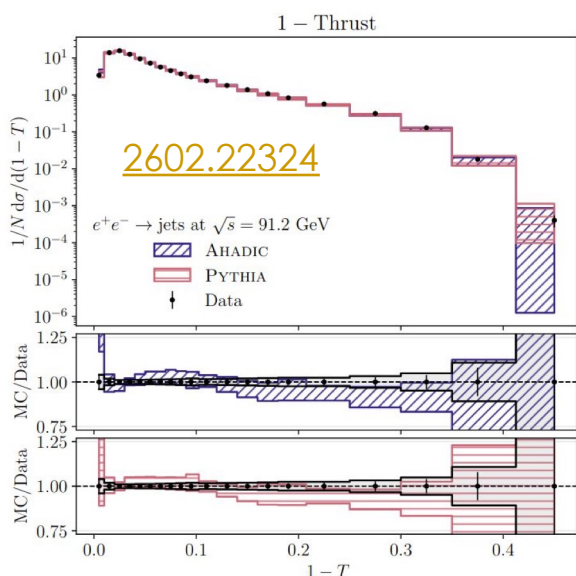


No uncertainties on them for a fully quantitative statement

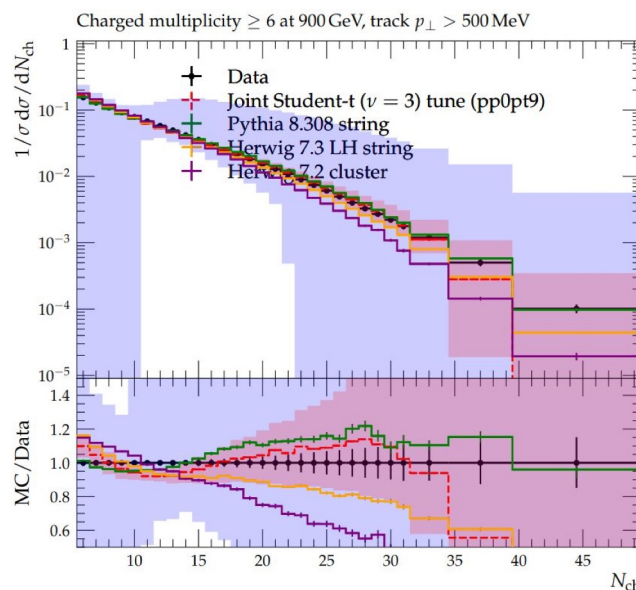
Improvements in non-perturbative modelling

Underlying Event & Hadronisation Tuning

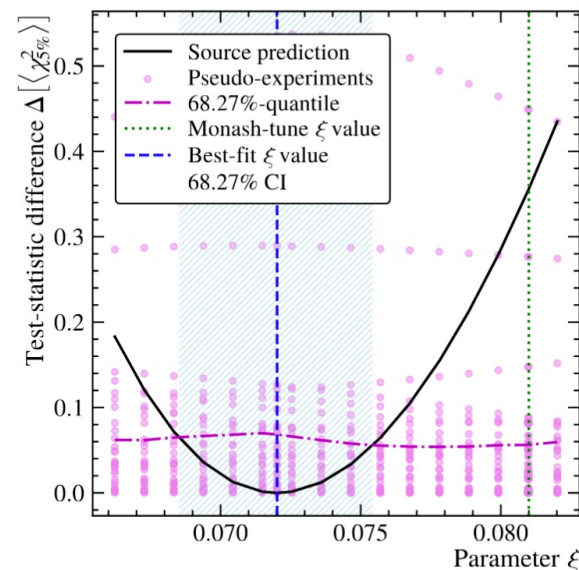
- **Recent explosion of tuning activity** - methodological developments toward reliable tune uncertainties



Sherpa history matching
Steffen Schumann



**Herwig MCMC
posterior sampling**
James Whitehead



**Frequentist
confidence interval**
Micheal Wilkinson

- Plus efficiency developments i.e.

- Pythia fragmentation reweighting – 2308.13459
- Sherpa MPI reweighting – work in progress
- ML-inspired tuning – Tony Menzo

Tuning in ATLAS

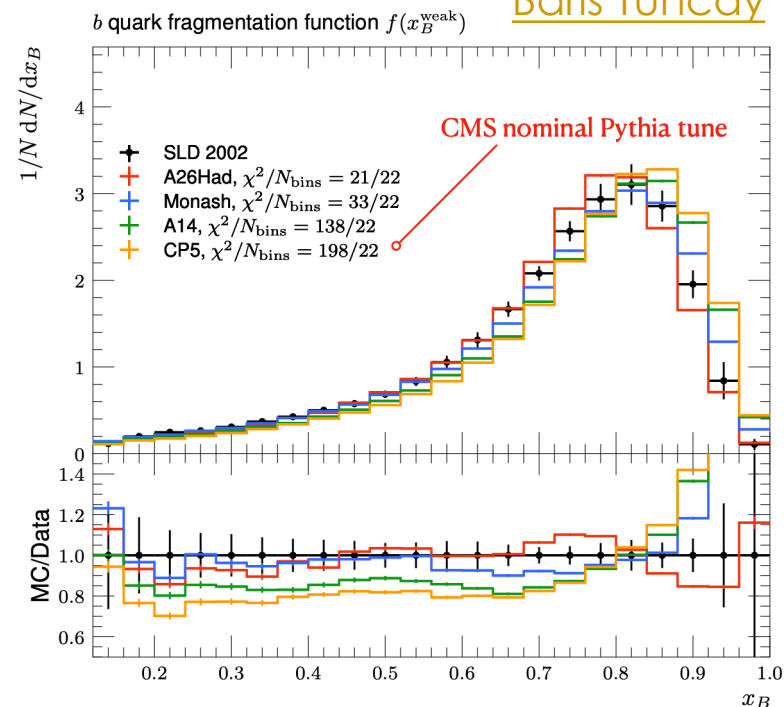
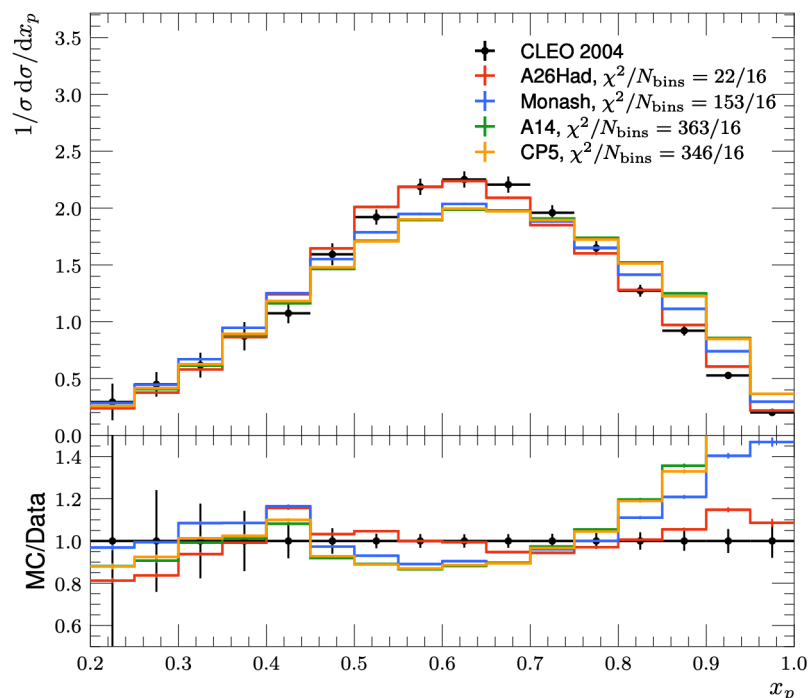
Baris Tuncay

- **A26 Family of Pythia and Vincia Tunes**
 - Hadronisation parameter tuning: A26Had
 - UE/MPI tuning: A26UE
 - ISR Tuning: A26ISR

- Many new ingredients:

- NLO ME, data beyond LEP, MC-friendly PDF sets, new QCD-based CR scheme etc.

D^{*+} scaled momentum using $D^0 \rightarrow K^- \pi^+$



- **Best description of SLD data**
- **Improved peak position and shape distribution in CLEO data**
- Working on a tailored version of Professor including bin-by-bin correlation
- *Still missing MEC/subtraction terms for Vincia*

Tuning in CMS

➤ **Professor2** experience in CMS

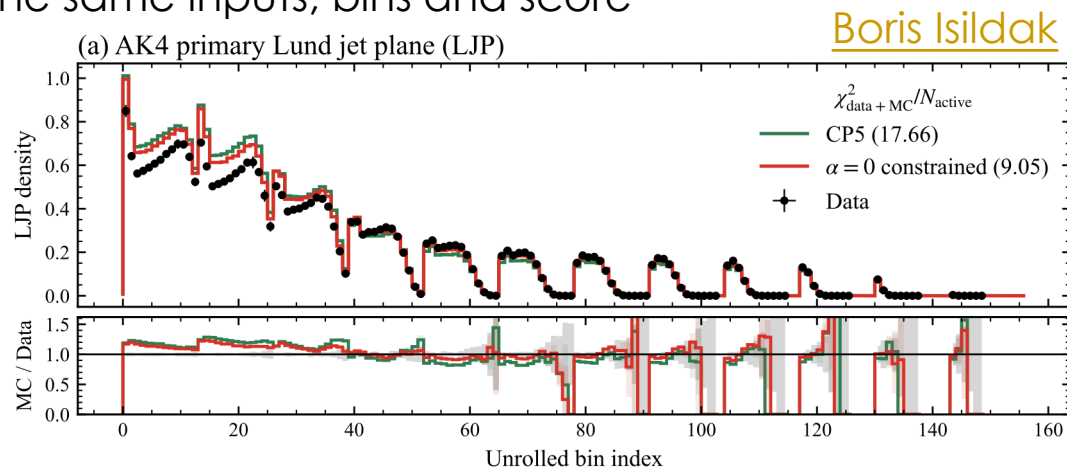
- Fast, parallel-friendly tuning workflow, but the response fit alone is not sufficient – a low predicted score can worsen after full-generator validation

➤ **Jet-substructure methodology**

- Six-parameter Pythia8 study (shower + hadronization) tuned to CMS LJP data
- CP5/CH3 do not describe the substructure observables well
- Constrained ($\alpha = 0$) optimisation improved the LJP fit while holding MB/UE observables at CP5-level performance

➤ **MCNNTUNES evaluation**

- Neural-network alternative to the Professor response
- CMS plans a like-for-like test on the same inputs, bins and score
- Plan:
 - Held-out prediction
 - Pseudo-data closure
 - Validate the selected tune with fresh generator runs
 - Benchmarking against Professor2



Tuning in Alice and LHCb

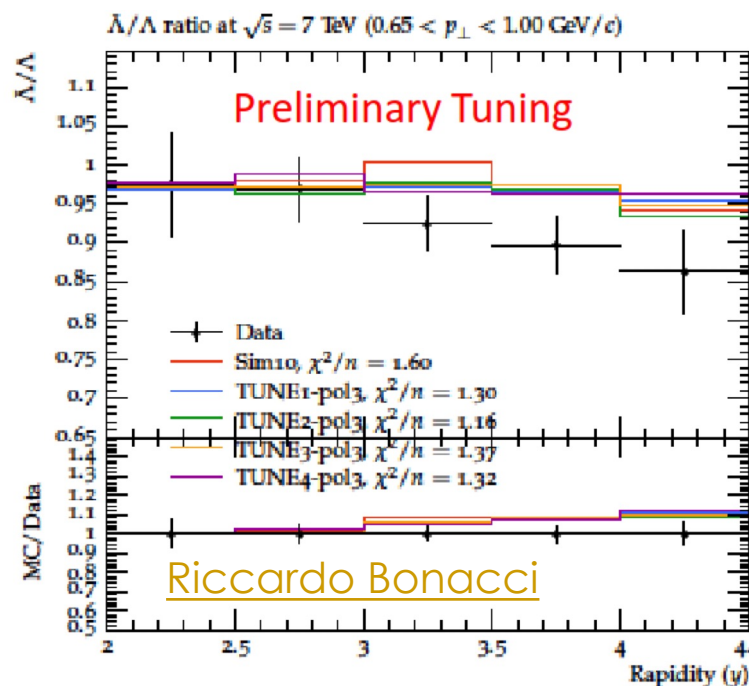
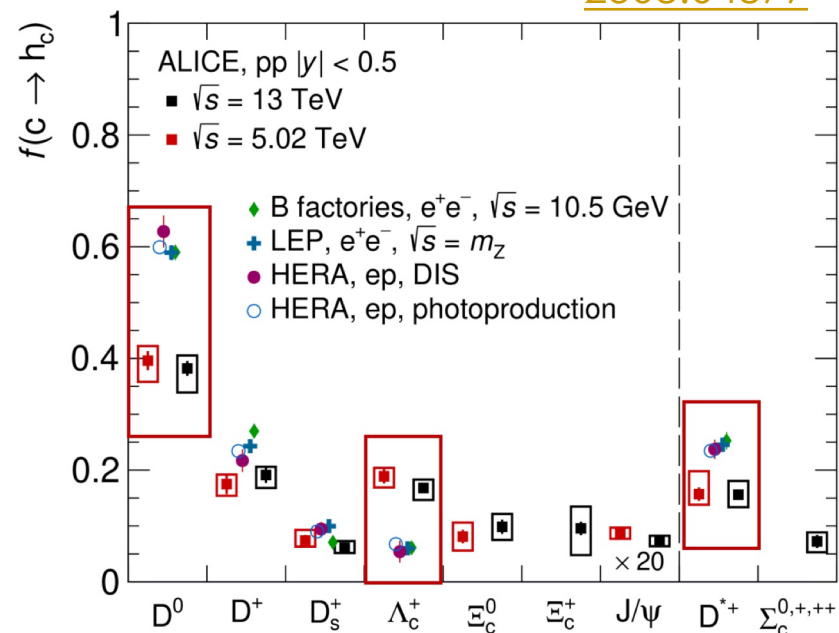
2308.04877

➤ Pythia8 QCD CR in HF sector

- Description of $\Sigma_c^{0,+,++}$ (2455) spectra + D mesons production (HF identified hadrons)

➤ Tuning of the **Rope Hadronization** model

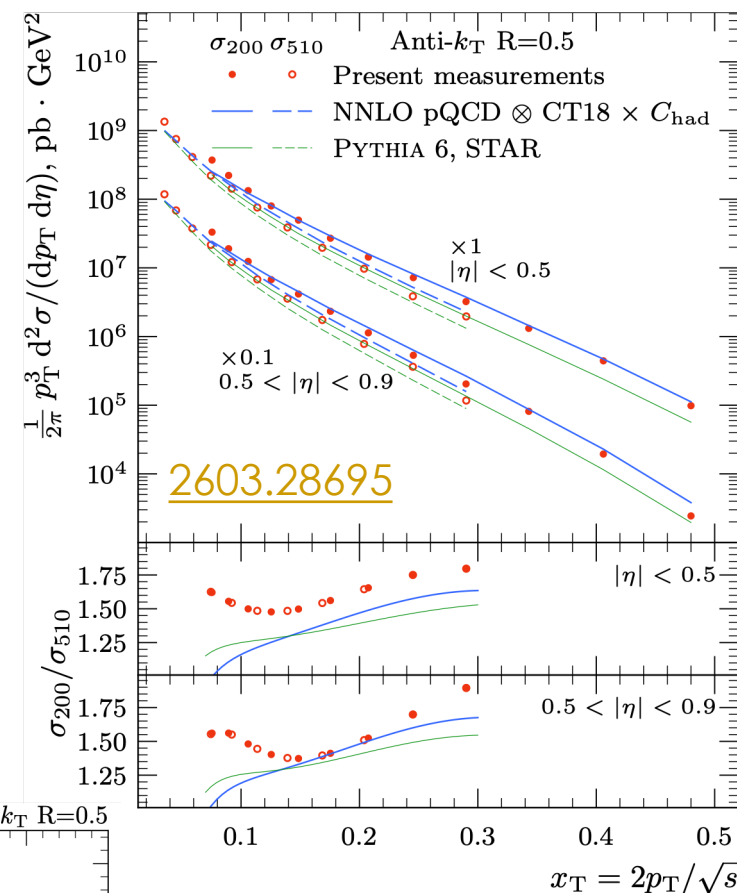
- Find a correct description for p/π vs $\langle dN_{ch}/d\eta \rangle_{|\eta|<0.5}$
- Describe strangeness enhancement from ALICE



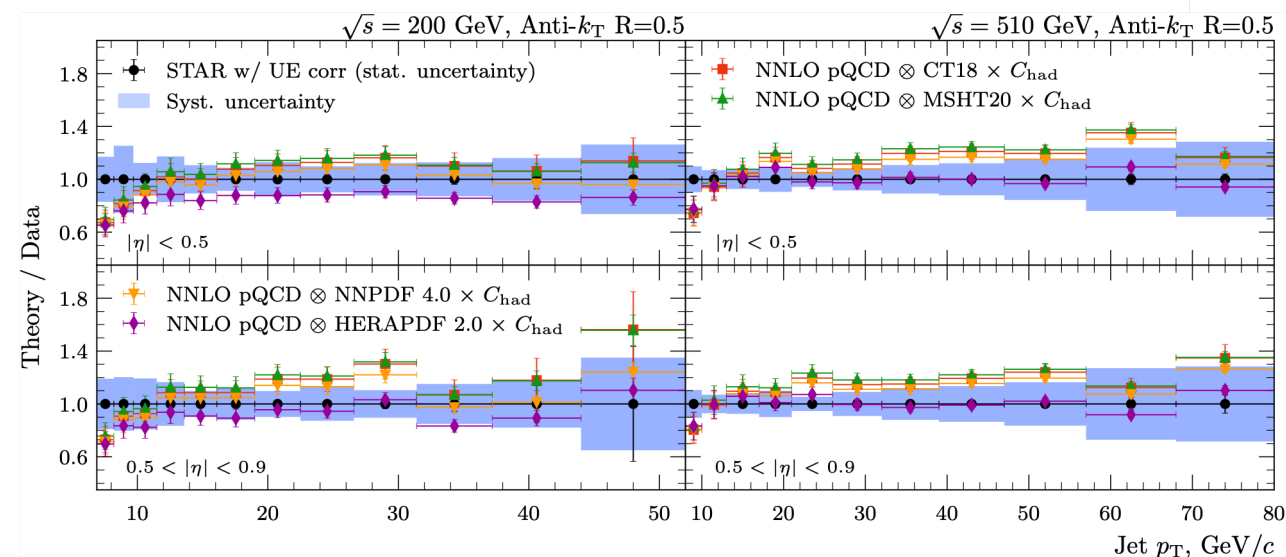
- Attempts at tuning have investigated CR, MPI and strong flavour options
- **QCD-inspired CR** options show beneficial behaviour for baryons, but are a factor of ~ 2 slower
- Bespoke tunes could be made for different physics purposes i.e. with weights to shift to another tune

Inclusive jets at RHIC

- **Double differential cross sections** in η and p_T at $\sqrt{s} = 200$ and 510 GeV
- Pythia describes the shape well, but slight offset $\rightarrow$ **valuable input for tuning**
- NNLO pQCD with recent PDF sets tends to overestimate data, which align most closely with HERAPDF
- Data could constrain xg for $x > 0.1$



**Hadronisation
corrections
are crucial!**

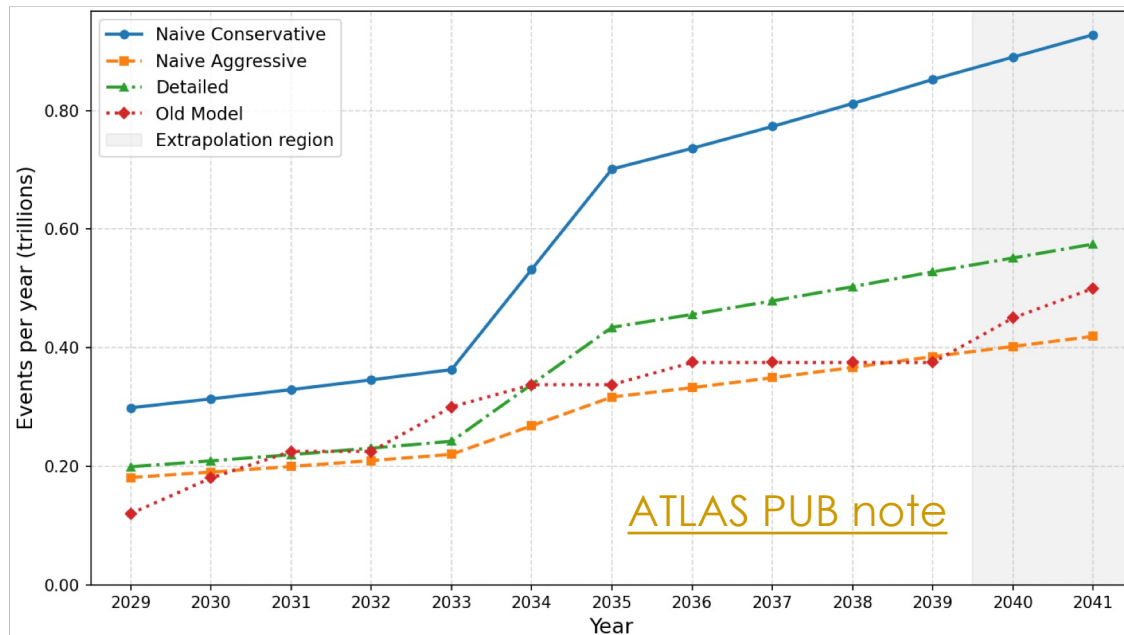


HL-LHC Event Projections

- Some significant work in 2025, and now a rather detail analysis based on:
 - Existing production over the last 5-10 years
 - Rucio records of ~all existing EVGEN
 - Classifications of datasets into several major categories using AML
 - Intuitive extrapolation rules for those categories based on their contents
- A couple of **very important projects** towards HL-LHC:
 - Negative event weights in vector boson samples
 - Slicing/biasing of top, single-top and diboson samples (by default)

Category	MC23 Events	Share	Scaling
Higgs	6.95	3.1%	Lumi
JetPhoton	40.10	17.9%	20%
Top	22.03	9.8%	50%
WeakBoson	151.48	67.7%	Lumi
Exotics	1.69	0.8%	Constant
Other	1.57	0.7%	Lumi

Pretty close to what was
estimated in the Roadmap



Storage and CPU needs for Run 4

LHCC March '26

350 x 10⁶ MC events / fb⁻¹ of data

90 x 10⁹ MC events / year (260 fb⁻¹ / year)

↓ with negative weights MCs

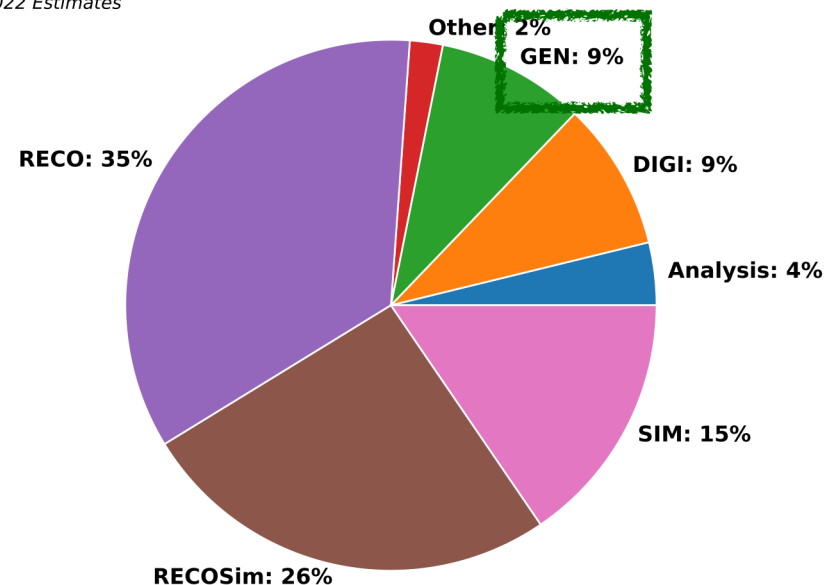
125 x 10⁹ (+40%)

Run 4

Data Formats Size (kB/evt.)	PU=62	PU=150
RAW	1 200	4 600
RECO*transient	4 300	15 000
AOD(SIM)*saved on demand (disc)	565 (605)	1 500
MiniAOD(SIM)	64 (86)	190
NanoAOD(SIM) kB/evt	1.5 (1.5)	4
HLTScout	12	25
L1Scout	360	450

Precision	CPU/evt. HS23-s
LO	150
LO (EvtGen)	1 500
NLO	200
NNLO	1 300
Total CPU/evt.	

CMSPublic
Total CPU HL-LHC (2031/No R&D Improvements) fractions
2022 Estimates



➤ 24 Petabytes/year

➤ **Game changers:** offloading ME part in GPUs and vectorized CPUs

➤ HepMC3 – Next Generation Event Record

➤ 2-4x smaller than HepMC2

➤ Native support for ROOT IO ~ 3x faster

LHC MC WG

- WG main web page: <https://lpcc.web.cern.ch/content/monte-carlo-wg>
- **Mandate:**
 - Address and improve MC generators of pp processes for the LHC experiments
 - MC tuning including simulation of minimum bias and underlying event
 - Computational and technical aspects
- The LHC MC WG operates by holding public meetings where all relevant topics are discussed
- Subgroups with a specific focus i.e. tuning are also formed
 - Very active Tuning Forum - <https://indico.cern.ch/event/1713297/>
- Subgroups report on their activities in the plenary meetings
- WG Indico category: <https://indico.cern.ch/category/18585>
- Next meeting at CERN from 05th till 06th October 2026 – **consider to join!**

Summary and outlook

- MC precision is entering **a new phase**
 - **Higher perturbative accuracy**: NNLO+PS becoming increasingly mature, next generation showers moving towards NLL accuracy
 - **Production must become more efficient**: GPUs/vectorization, reduction of NW and on-the-fly reweighting are key to make HL-LHC MC production feasible
 - **Generator modelling remains a limiting uncertainty**: HF production, UE/MPI and JSS can have sizeable impacts on precision measurements
 - **Towards principled uncertainties**: new tuning methodologies can provide quantitative parameter uncertainties (no ad-hoc variations)
 - **Theory ↔ experiments need to be tighter**: several current uncertainties cannot be solved by either side alone
- Looking towards Run 4!
 - Establish *common benchmarks* and *validation strategies* for PS and generator developments
 - Develop *data/theory-driven constraints* on dominant modelling uncertainties
 - Exploit *AI/ML and heterogeneous computing* without compromising reproducibility
 - Continue coordinated work through the *LHC MC WG and its Tuning Forum*
- The goal: **MC predictions** that are **simultaneously more accurate, more efficient**, and equipped with **reliable uncertainty estimates**!