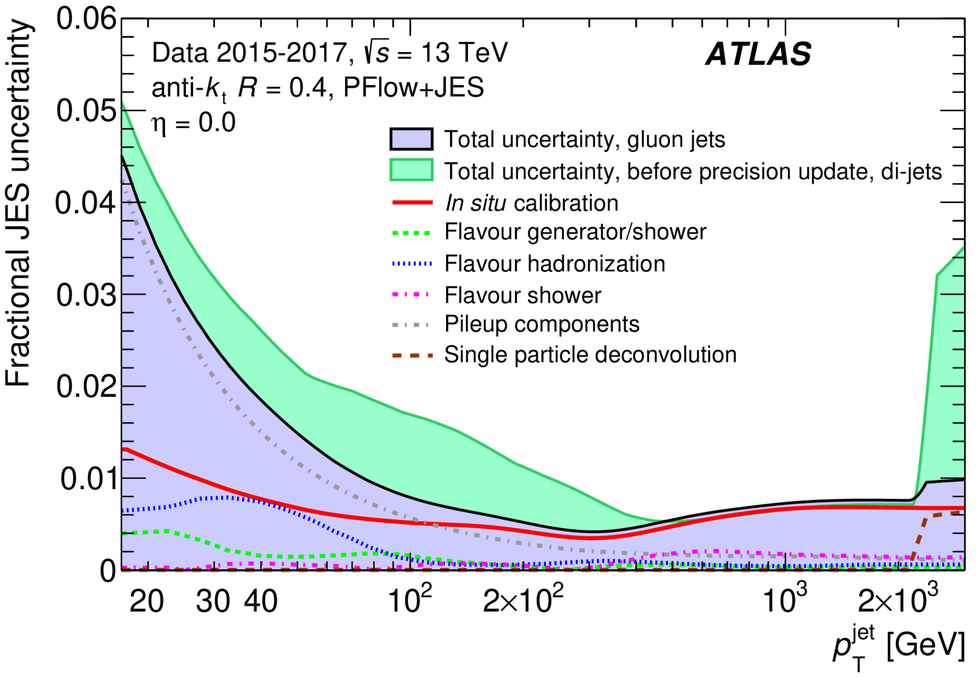
Les Houches MC & Tools experimental overview

Chris Hayes, Alexander Grohsjean, **Jennifer Roloff**
(+Silvia Ferrario Ravasio, Daniel Reichelt, Andrzej Siodmok)

Quantifying uncertainties for parton showers and hadronization

- Modeling uncertainties are often defined in ad-hoc ways, and often rely on 2-point differences between different models
  - Results in both double counting uncertainties, but also in poorly understood results that do not capture all possible effects (and sometimes underestimate uncertainties — no real way to know if results are conservative or not)
- At Les Houches in 2023, many discussions on how to quantify uncertainties in a more meaningful way
  - ATLAS uses an approach that factorizes uncertainties from the parton shower and hadronization
  - Want to understand the interplay between these effects

2405.20206



Quantifying uncertainties for parton showers and hadronization

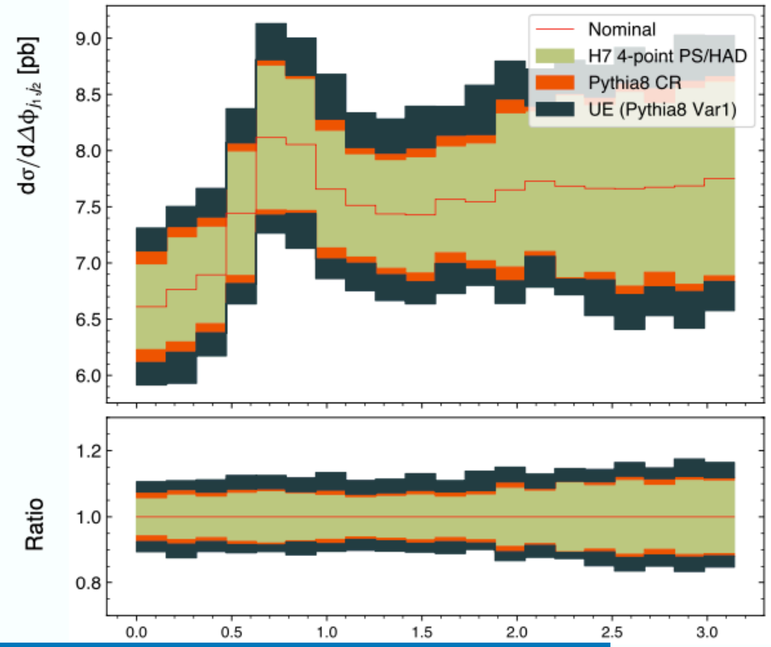
- LH23 studies started conversations on how to improve modeling uncertainties, but this was just the start
- ***LH discussion**: Can we provide stronger tests of these models?*
- ***LH study**: Are we able to start comparing NLL PS predictions?*
- ***LH study**: Can we test these grids?*

LH2023 Report

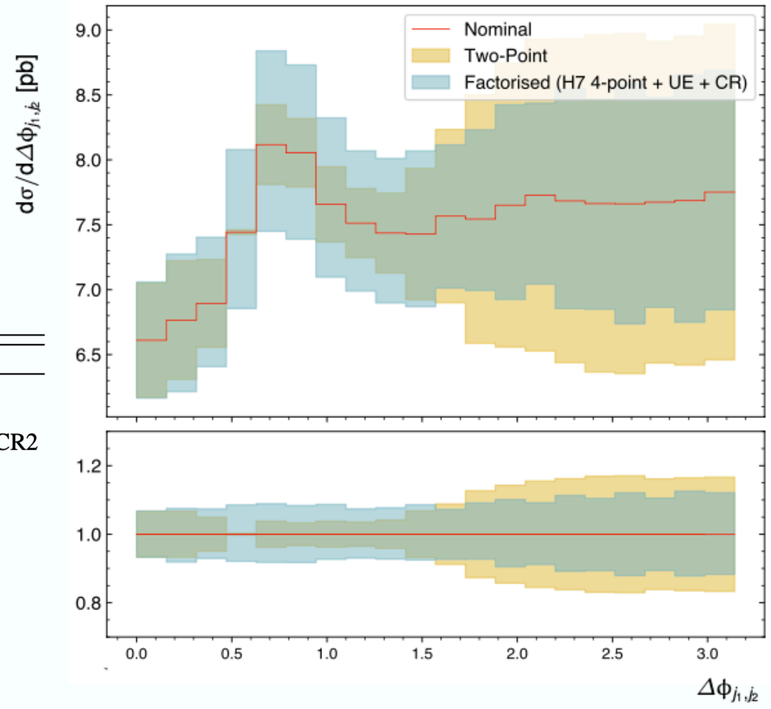
The more factorised two-point comparisons studied by ATLAS were recognised as a step forward; however, some guidance was provided by members of the Les Houches community regarding further improvements to this methodology. In particular, it was pointed out that there can be significant interplay between hadronisation and parton shower algorithms, and that a more reliable way to assess such two-point uncertainties would be to use grids of MC setups that are tuned and configured consistently by the MC authors themselves (e.g., a 2-by-2 grid of HERWIG setups with different parton shower and hadronisation algorithms). Such provisions would be welcomed by the community, and they may also present an opportunity to harmonise the tuning strategies and reference datasets used by the different MC authors.

Quantifying uncertainties for parton showers and hadronization

- Similar approach can be used for modeling uncertainties (not just jet uncertainties)
- ***LH discussion**: How can we handle challenges associated to statistical treatment of these uncertainties (e.g. use of envelopes with profile likelihood)*
- ***LH discussion**: Is factorizing into parton shower and hadronization sufficient?*



From Josh McFayden and Betsy Cunnett



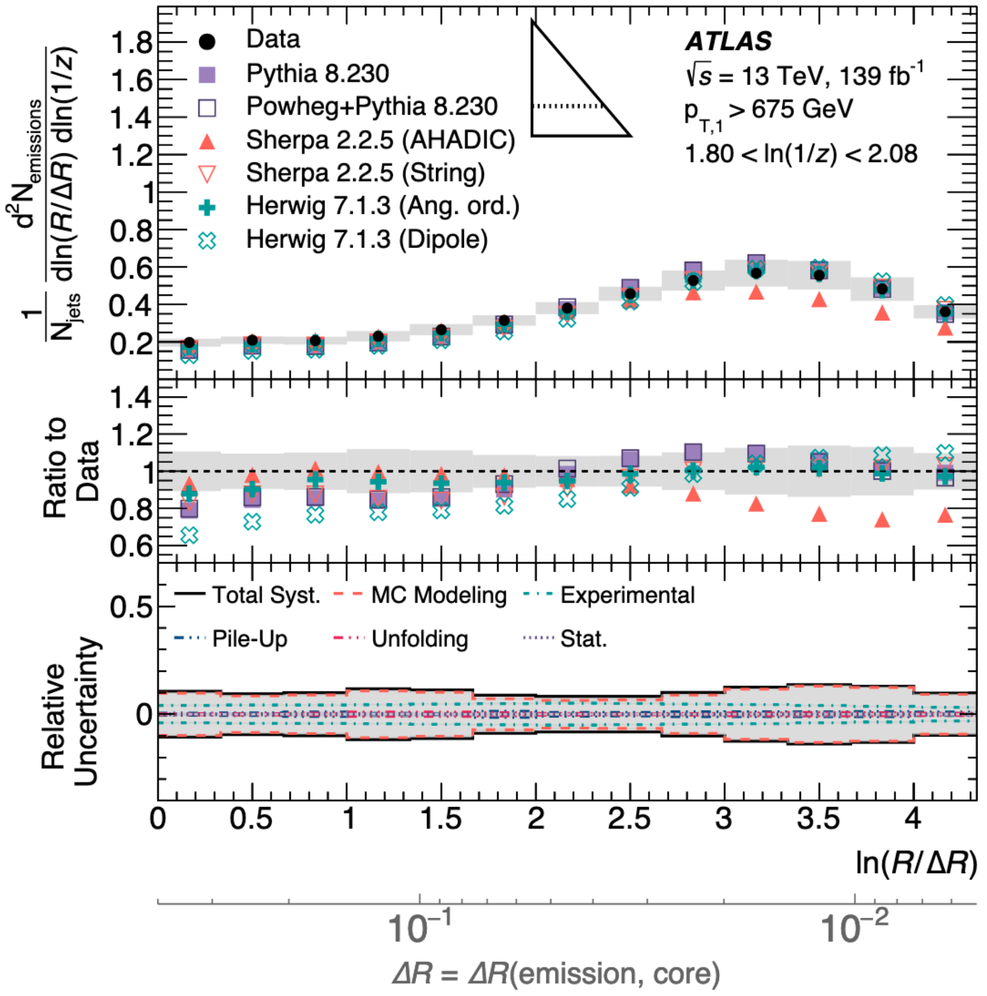
NLO $t\bar{t}$ sample information

	Generator	ME order	Parton Shower	PDF	Tune	Shower Model	Hadronisation Model	Tune Variations	CR Model
Nominal	Powheg Box	NLO	Pythia 8	NNPDF 3.0	A14	Dipole	Lund-String	-	Default
UE Variations	Powheg Box	NLO	Pythia 8	NNPDF 3.0	A14	Dipole	Lund-String	A14 Var1 Up/Down	Default
CR Variations	Powheg Box	NLO	Pythia 8	NNPDF 3.0	A14	Dipole	Lund-String	-	CR0/CR1/CR2
Two-Point Herwig PS Comparison	Powheg Box	NLO	Herwig 7.1	NNPDF 3.0	H7.1-DEFAULT	QTilde	Cluster	-	Default
Herwig Angular+Cluster	Powheg Box	NLO	Herwig 7.3	NNPDF 3.0	H7.3-DEFAULT	QTilde	Cluster	-	Off
Herwig Angular+Lund-string	Powheg Box	NLO	Herwig 7.3	NNPDF 3.0	H7.3-DEFAULT	QTilde	Lund-String	-	Off
Herwig Dipole+Cluster	Powheg Box	NLO	Herwig 7.3	NNPDF 3.0	H7.3-DEFAULT	Dipole	Cluster	-	Off
Herwig Dipole+Lund-string	Powheg Box	NLO	Herwig 7.3	NNPDF 3.0	H7.3-DEFAULT	Dipole	Lund-String	-	Off

Jet substructure exploration

- There are too many JSS observables to remember or measure…
- Do **not** want to measure all of them, but want to know which are most effective at probing different aspects of jet formation
  - ***LH discussion****: what specific PS and hadronization effects are theorists interested in testing?*
- At LH23, started scan over many different MC predictions
  - Compared to existing Rivet routines, and created new routine to study performance of new observables
  - ***LH study****: can we do a more complete analysis of these results?*
  - ***LH study****: can we expand these results to more processes (currently just dijets)*

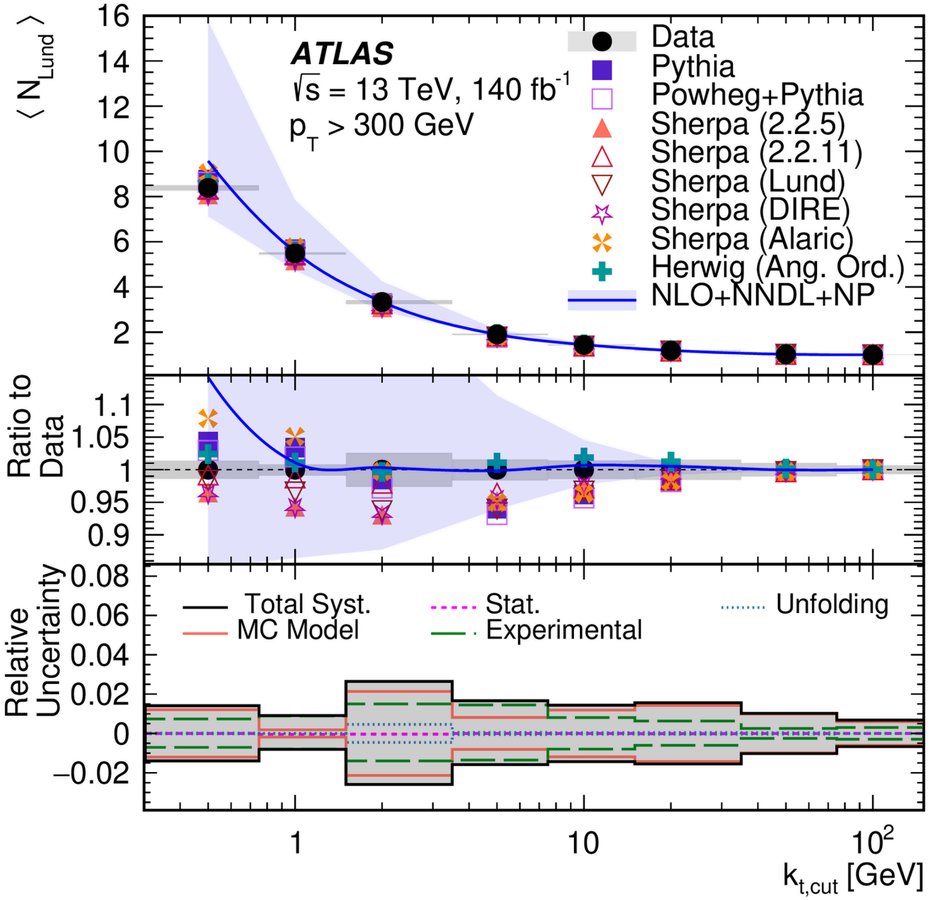
2004.03540



Advanced parton shower models: experimental connections

- Many recent developments in improving the accuracy of parton shower models
- Need to *test* these models, and to understand experimental impacts of missing effects
- Complementary to understanding which observables are most important — need to know which observables and event topologies give us good insight into these new effects
  - Also need to discuss experimental feasibility of any strategies
- ***LH discussion**: what observables should we study to test NLL and NNLL parton showers?*

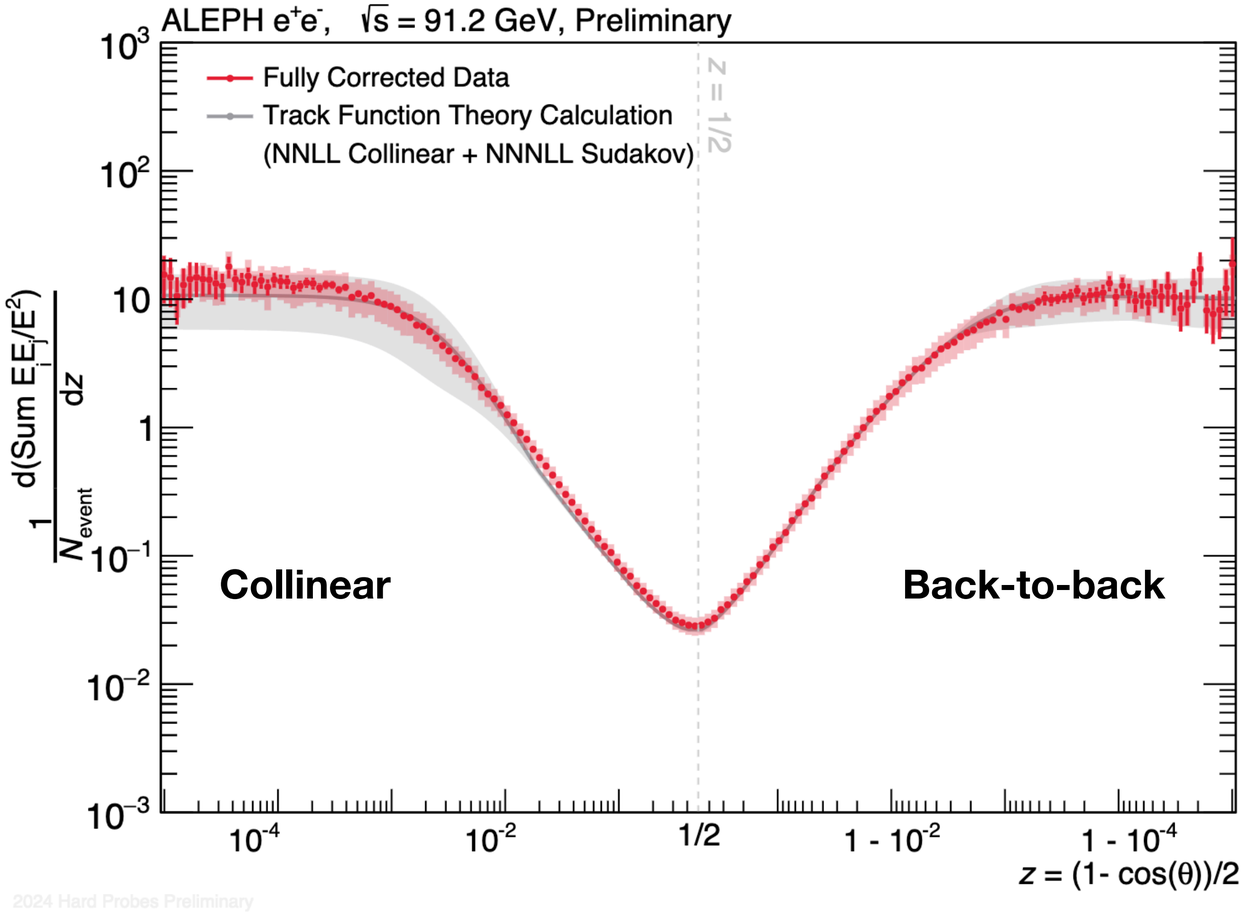
2402.13052



Jet substructure in ee events

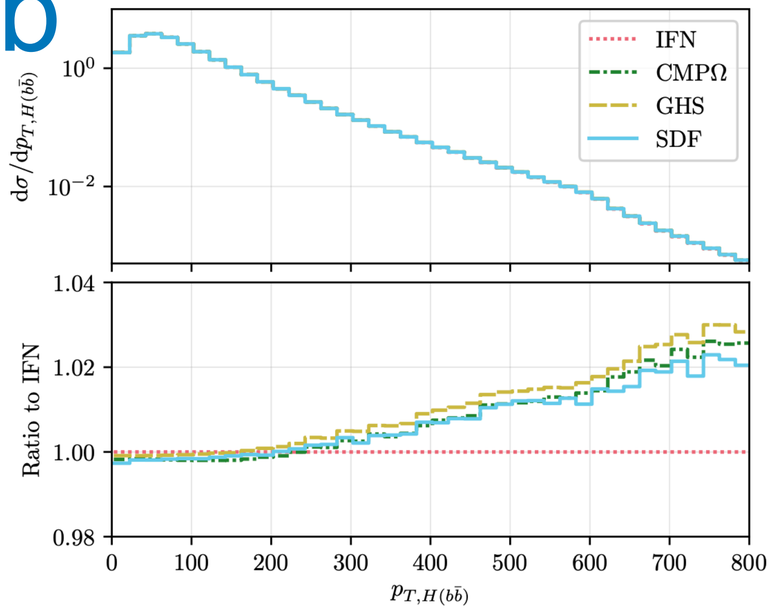
2111.09914 2211.13519 2501.01968

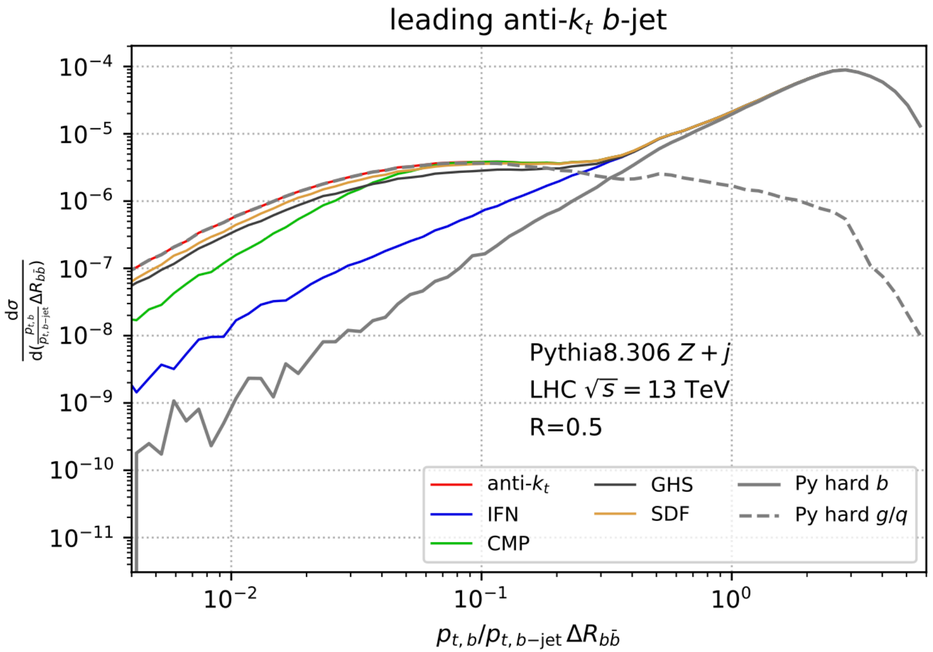
- Resurgence of JSS measurements in e^+e^- events using reanalyzed LEP data
- Opportunity to test out **effectiveness of new techniques** in preparation for the possible future e^+e^- colliders colliders
- ***LH discussion**: Are there any new measurements we could do with the existing data that would complement the LHC measurements?*
  - Cleaner environment, but also smaller energy range, less access to perturbative QCD
- ***LH discussion**: What detector and reconstruction advances are needed to enable new measurements at future colliders?*
  - *e.g.* better SV reconstruction, particle ID, etc.



Heavy flavor jet tagging and g→bb

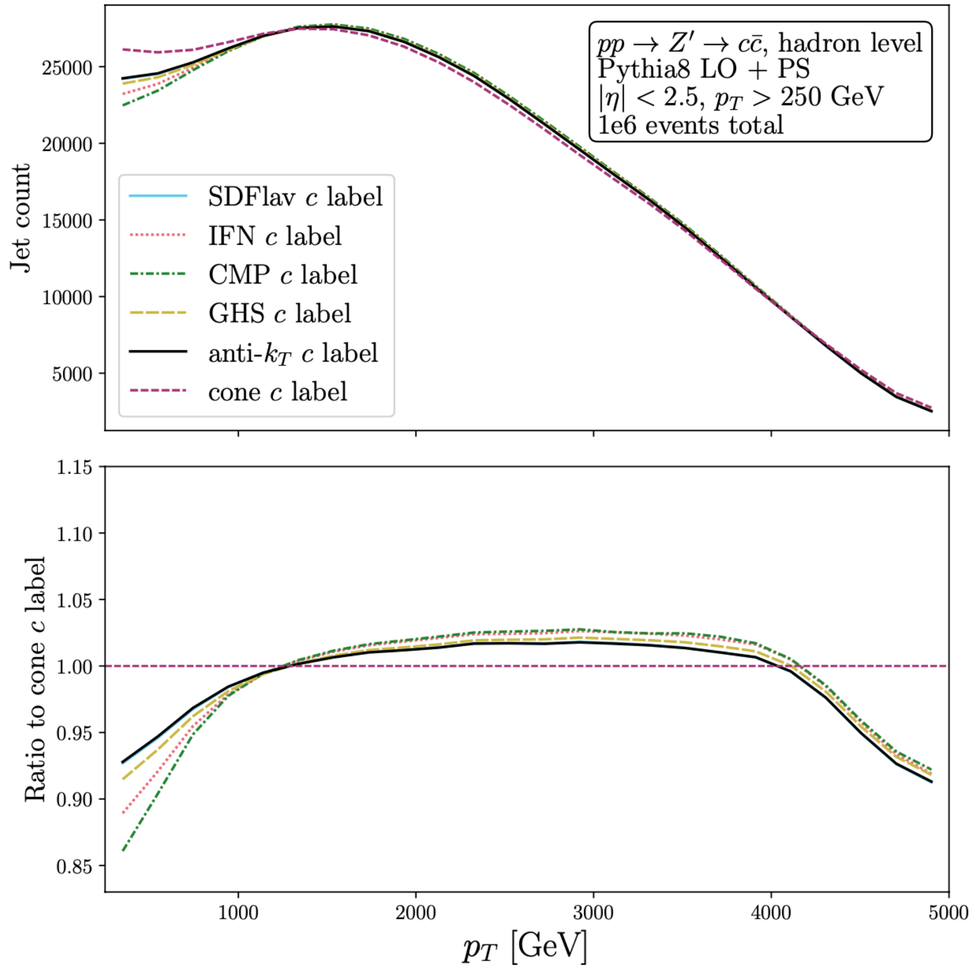
- Recent theoretical advances in definitions of heavy flavor jets that can be experimentally reconstructed
  - Many discussions at LH23, with an upcoming paper describing the conclusions
  - Rough agreement between models in many cases, but still some mislabeling in many cases
  - ***LH study**: how do these algorithms interact with JSS observables?*





Heavy flavor jet tagging and g→bb

- Interaction between choice of label and experimental flavor tagging training
  - Can result in inclusion of gluon splitting, or can aggressively remove training jets
  - ***LH discussion**: Should we have a unified definition across experiments?*
  - ***LH study**: What are the experimental impacts of these choices?*



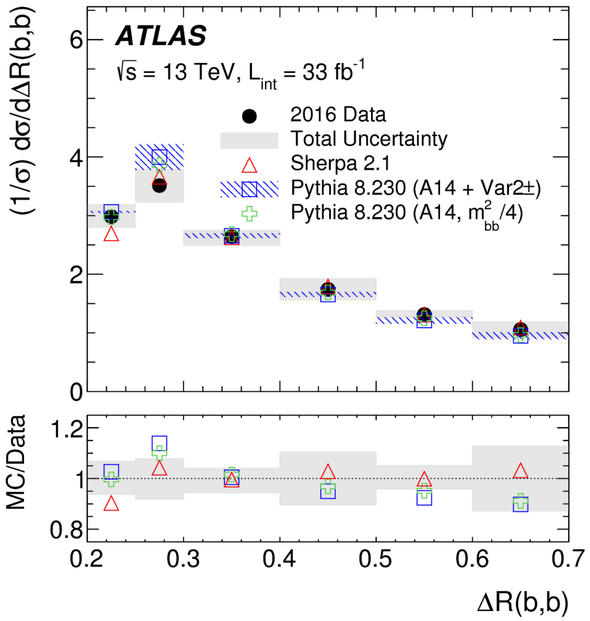
Heavy flavor jet tagging and g→bb

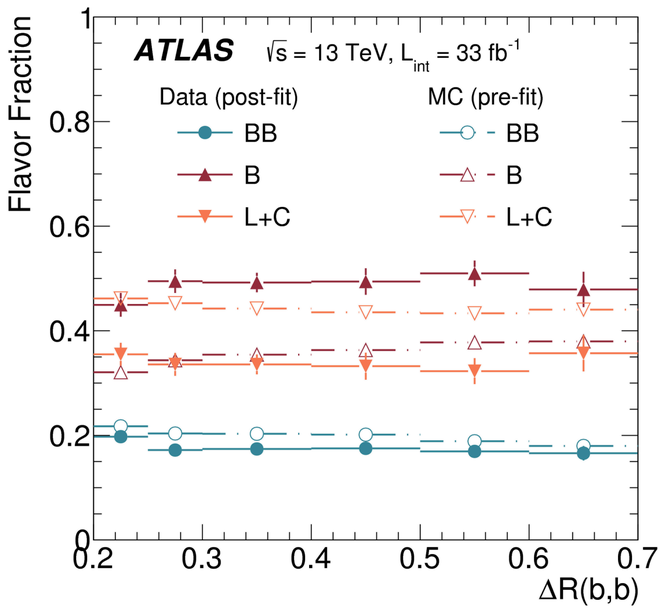
- A few measurements of g→bb exist, but more precision is needed
  - Modeling is poorly constrained, but needed for better understanding and comparison of the different flavor algorithms
  - ***LH discussion**: Are there any particular aspects that are important to measure? (e.g. rates, angles, p_T dependence, etc)*
  - ***LH discussion**: How experimentally feasible are these measurements?*
- Relevant for better measurements of b-quark-initiated jets, but also for H→bb measurements
  - ***LH study**: Does double-b-tagging bias the selection?*
  - ***LH study**: Can we build on techniques for single b-jet reconstruction to create less biased measurements?*
  - Many of these studies could be done at particle-level, with more detailed detector simulation later

LH2023 Report

Our findings suggest the best approach to a successful implementation may be gradual: using these algorithms first at truth level, to improve Monte Carlo-based corrections, and progressively moving toward direct application to reconstructed data as experimental techniques mature. Several open directions remain. Firstly, deeper studies are needed on the modelling of gluon splitting into $b\bar{b}$-pairs, possibly through double b-tagged jets. Secondly, validations of the algorithms' performance under different hadronisation and parton shower models (e.g., string vs. cluster models in HERWIG and SHERPA) must continue. These variations are typically small, but still non-negligible in some regimes, especially for charm.

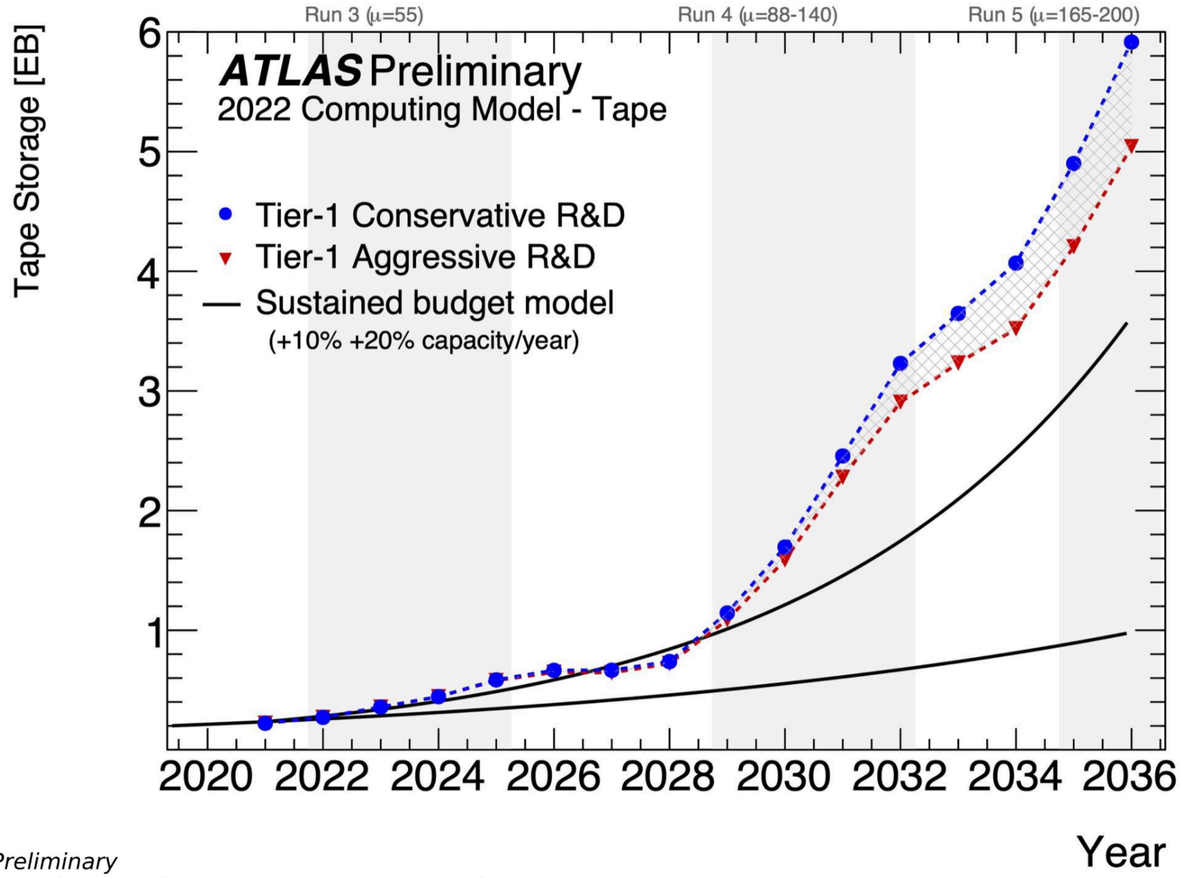
1812.09283

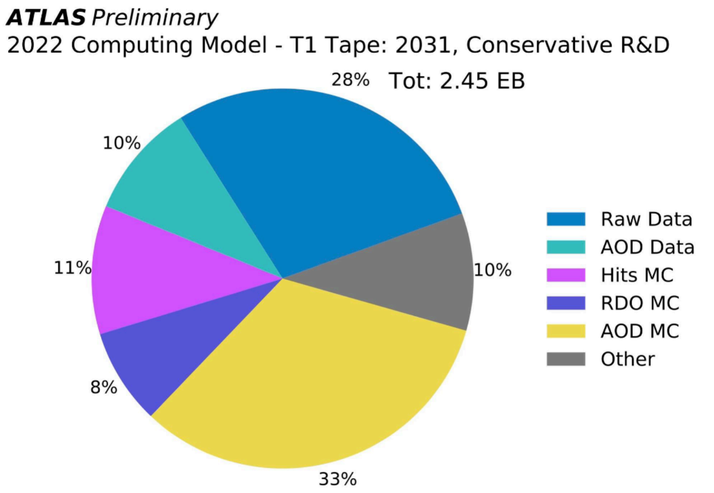




Negative weights and matching

- The expected amount of tape and disk space that we will need to store samples is not sufficient
- Monte Carlo is a significant portion of this (and the part that we have the most control of)
  - 15% negative weights already means we need to generate double the number of events!
- *How can we be more efficient in our event generation to reduce amount of MC that we need to generate and store?*

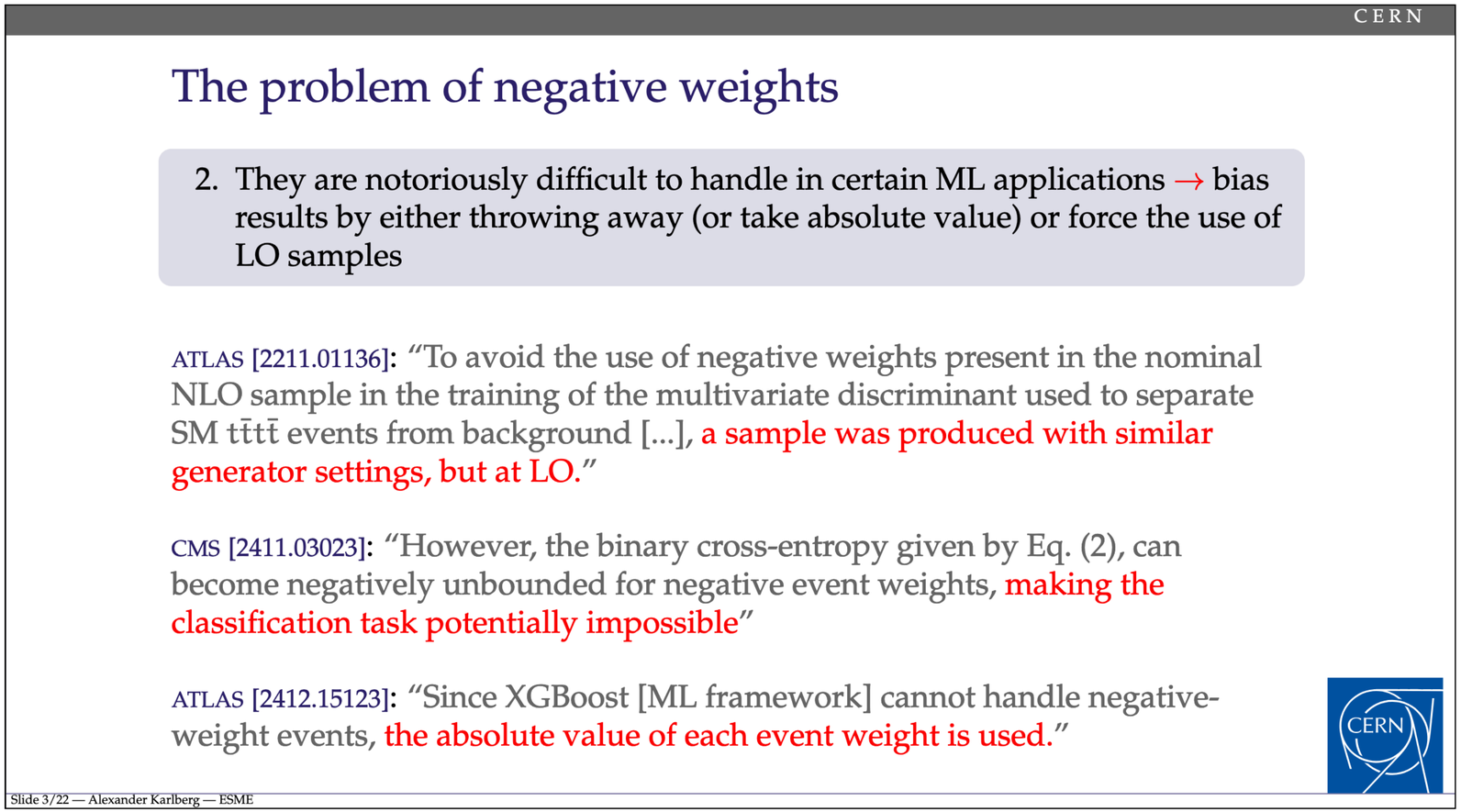




Negative weights and matching

A. Karlberg's slides

- Negative weights also cause difficulties in ML training
  - Most models are not set up to handle negative weights
- ***LH study/ discussion**: Are there better ways of handling these cases for ML training?*



CERN

The problem of negative weights

2. They are notoriously difficult to handle in certain ML applications → bias results by either throwing away (or take absolute value) or force the use of LO samples

ATLAS [2211.01136]: "To avoid the use of negative weights present in the nominal NLO sample in the training of the multivariate discriminant used to separate SM $t\bar{t}t\bar{t}$ events from background [...], a sample was produced with similar generator settings, but at LO."

CMS [2411.03023]: "However, the binary cross-entropy given by Eq. (2), can become negatively unbounded for negative event weights, making the classification task potentially impossible"

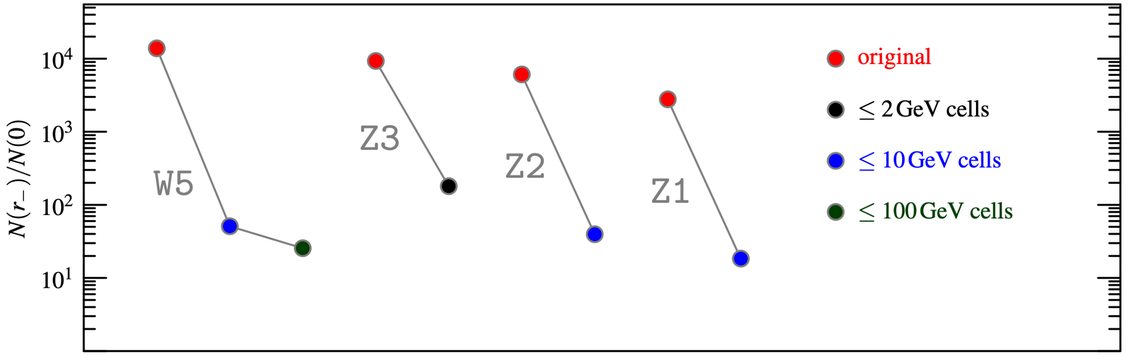
ATLAS [2412.15123]: "Since XGBoost [ML framework] cannot handle negative-weight events, the absolute value of each event weight is used."

Slide 3/22 — Alexander Karlberg — ESME

Negative weights and matching

2303.15246

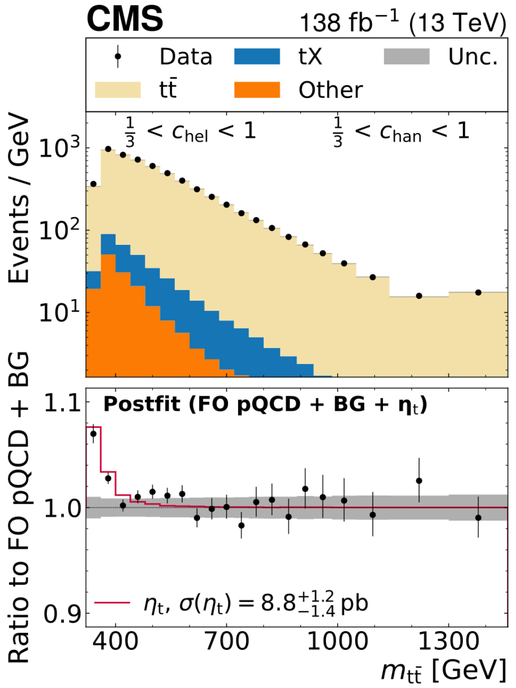
- Which samples are the most important to improve in the coming years?
  - Interplay with precision measurements: 2→3 NNLO jet production (and α_s) is currently limited by negative weights…
- Recent workshop explored many of the different options for mitigating these effects
  - Wide variety of methods currently being explored, including better strategies to avoid negative weights, ML/AI methods to redistribute weights
- ***LH discussion**: What benchmarks should we use to compare their performance?*
  - Multi-dimensional problem, relevant for multiple processes → cannot just use one metric!

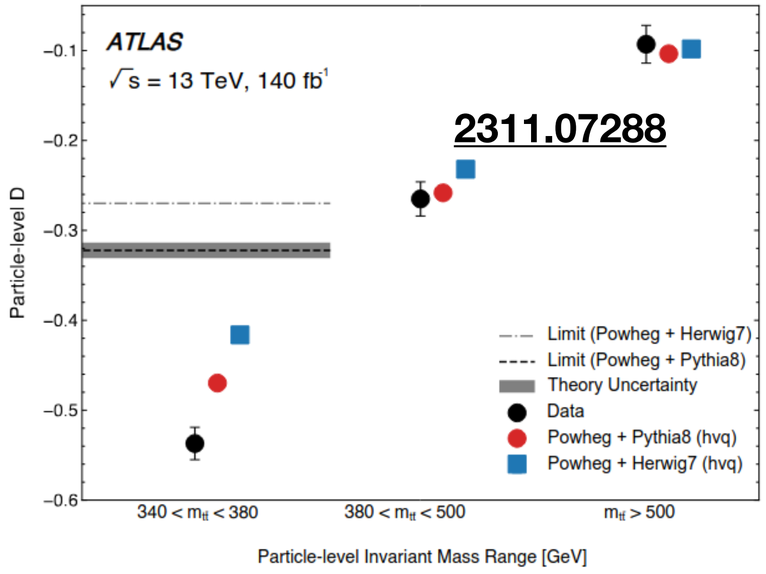


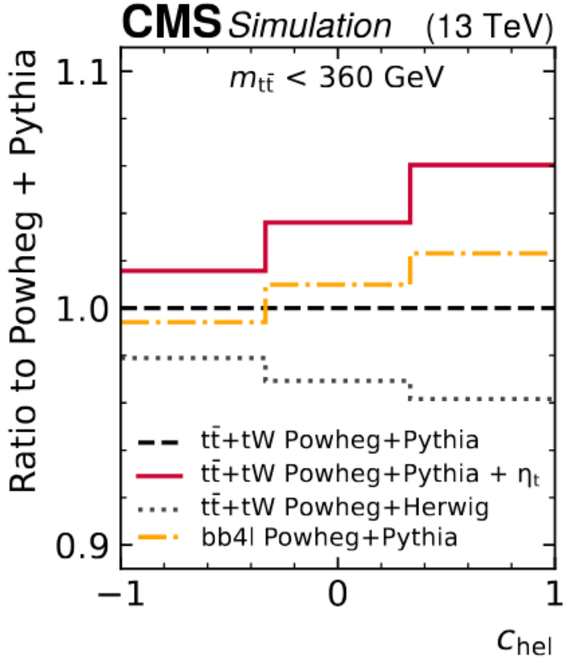
Modelling of the $t\bar{t}$ threshold

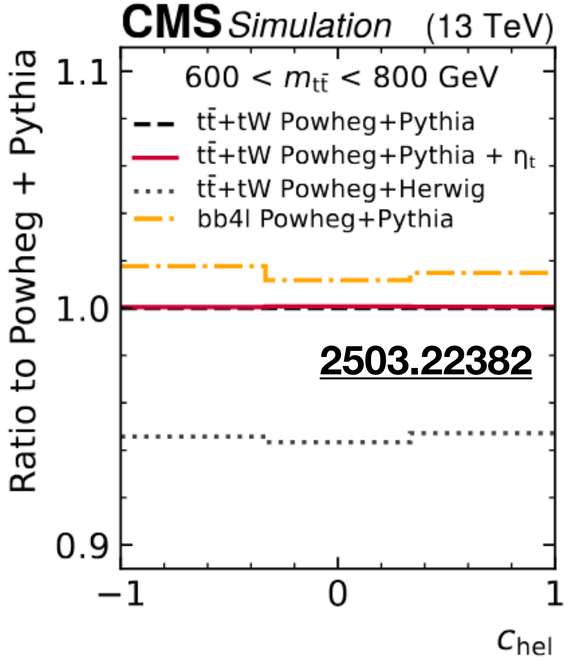
- $t\bar{t}$ threshold received significant attention in the last two years, particularly with the observation of quantum entanglement at the LHC
- Excess compatible with $t\bar{t}$ quasi-bound states
- Crucial to understand:
  - Invariant $t\bar{t}$ mass spectrum
  - Spin correlations
- Issue:
  - Large discrepancies between ME generators
  - Significant impact of parton shower on predicted spin correlation

CMS entanglement: 2406.03976





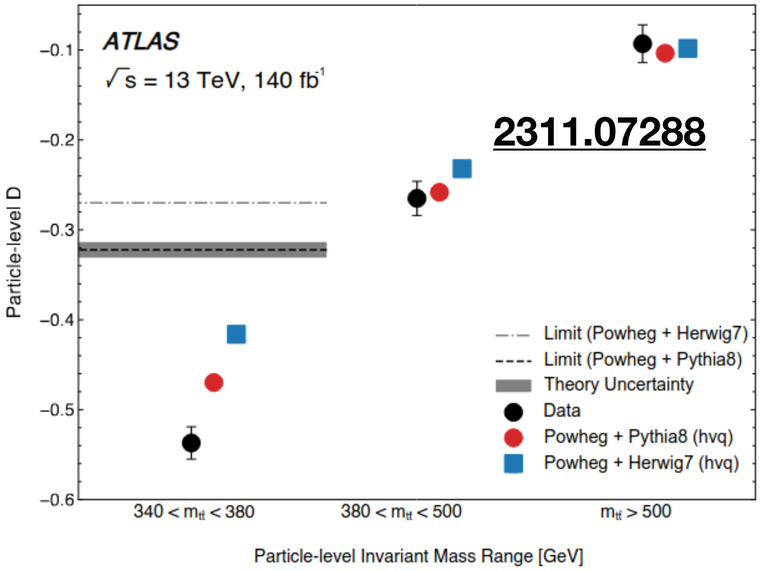


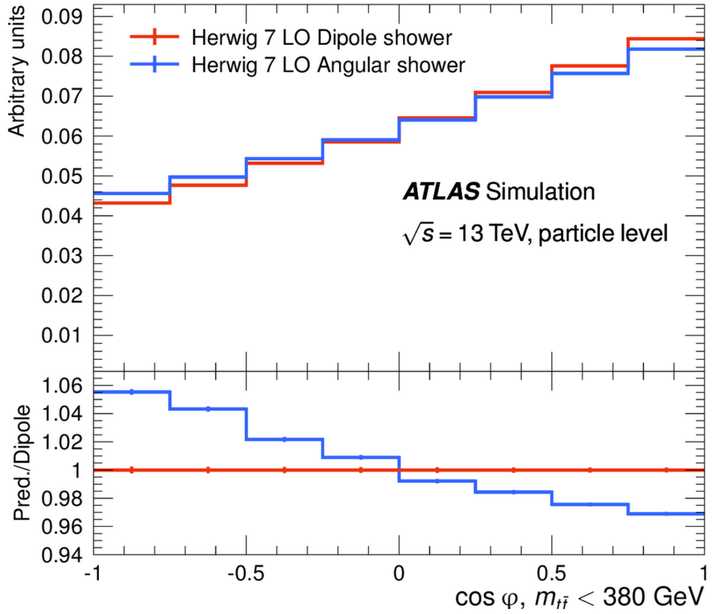


Modelling of the $t\bar{t}$ threshold

- $t\bar{t}$ threshold received significant attention in the last two years, particularly with the observation of quantum entanglement at the LHC
- Excess compatible with $t\bar{t}$ quasi-bound states
- Crucial to understand:
  - Invariant $t\bar{t}$ mass spectrum
  - Spin correlations
- Issue:
  - Large discrepancies between ME generators
  - Significant impact of parton shower on predicted spin correlation

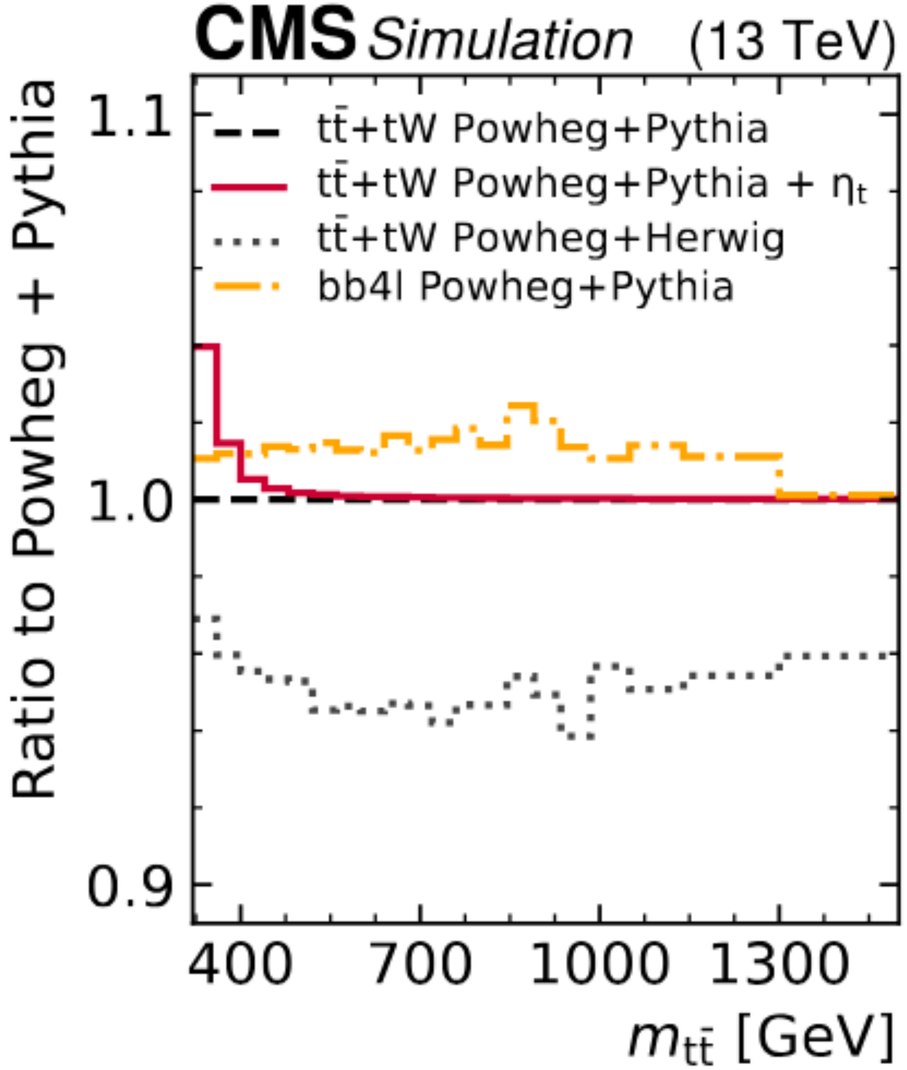
CMS entanglement: 2406.03976





Modelling of the $t\bar{t}$ threshold

- Open questions:
  - Why do angular ordered showers change spin correlation (only) close to the threshold?
  - Are these changes correlated with the change in the predicted invariant mass spectrum ?



Modelling of the $t\bar{t}$ threshold

- CMS uses NNLO+NLO EW tt as baseline and adds a toponium signal inspired by NRQCD
- arXiv:2505.00096 by Emanuele, Paolo and Luca
  - proposal to consider only the first few powers of α_s/v
- part of bound state effect might already be included in MiNNLO+PS
- Open questions for LH:
  - *How compatible are the theoretical approaches ?*
- How well do they fit to the data ?
- What is a good model for quasi bound states ?