Monte Carlo Simulation in CMS

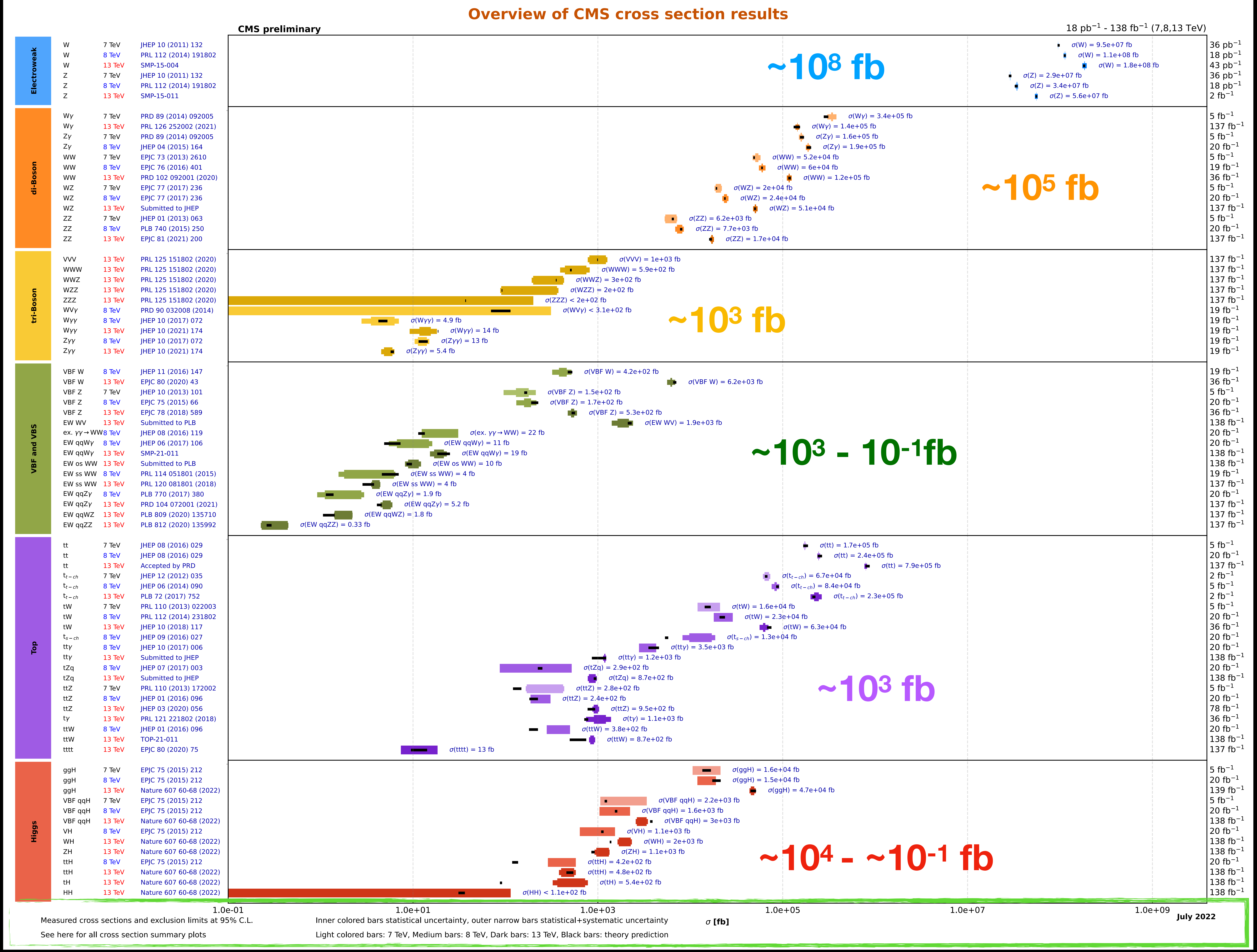
Saptaparna Bhattacharya

Why are we talking about event generation?

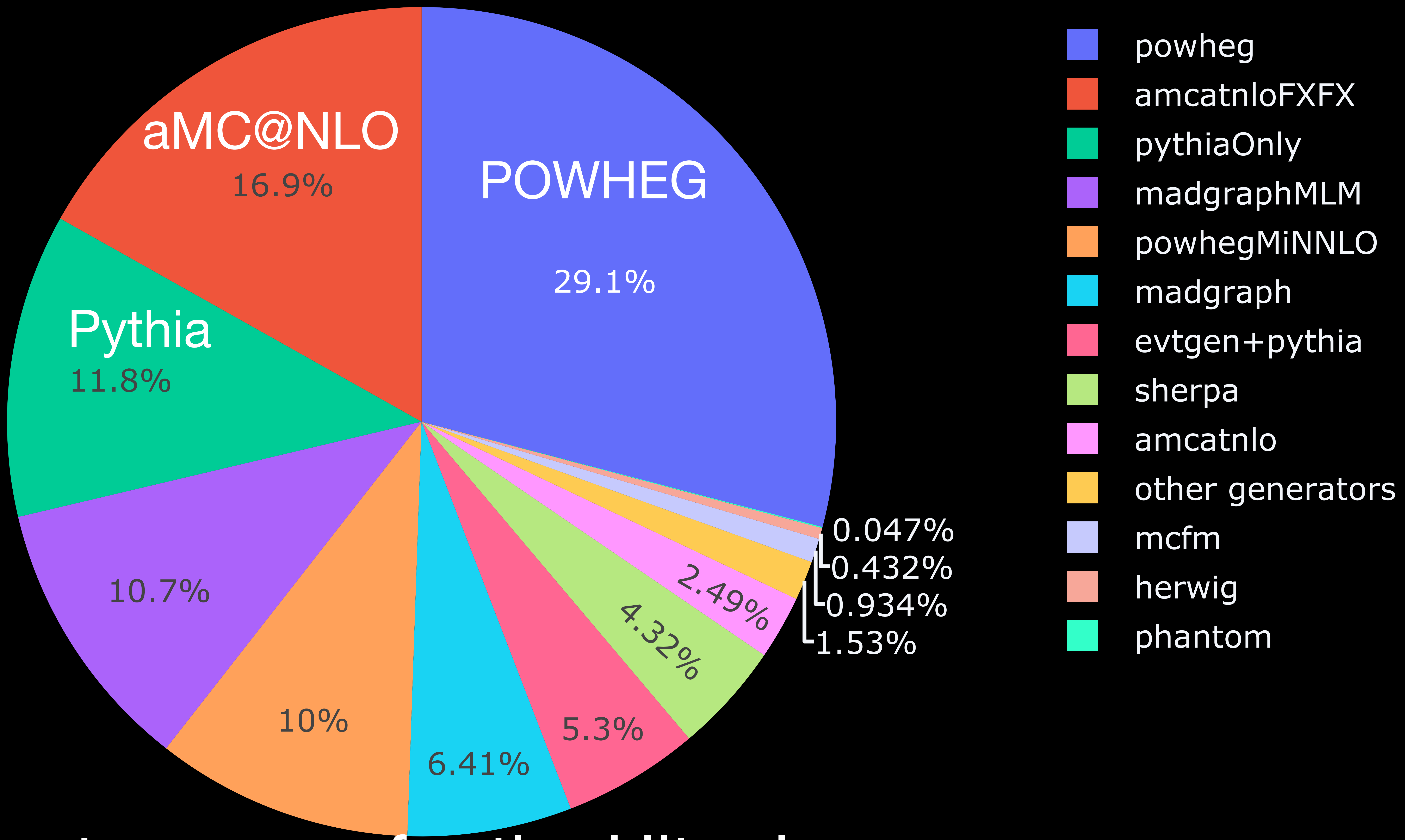
- We can make detailed calculations about the physics processes that we are sensitive to at the LHC
- The calculations are not cheap because of the vast array of processes produced at the LHC: theoretical calculations required span 14 orders of magnitude
- These calculations are an essential part of our field as we keep pushing the precision frontier with hadron colliders
- Which process and at what accuracy to compute is an important choice — need to pay attention to both quality and quantity

Spans 10 orders of magnitude

Spans 14 orders of magnitude when inelastic cross sections are taken into account

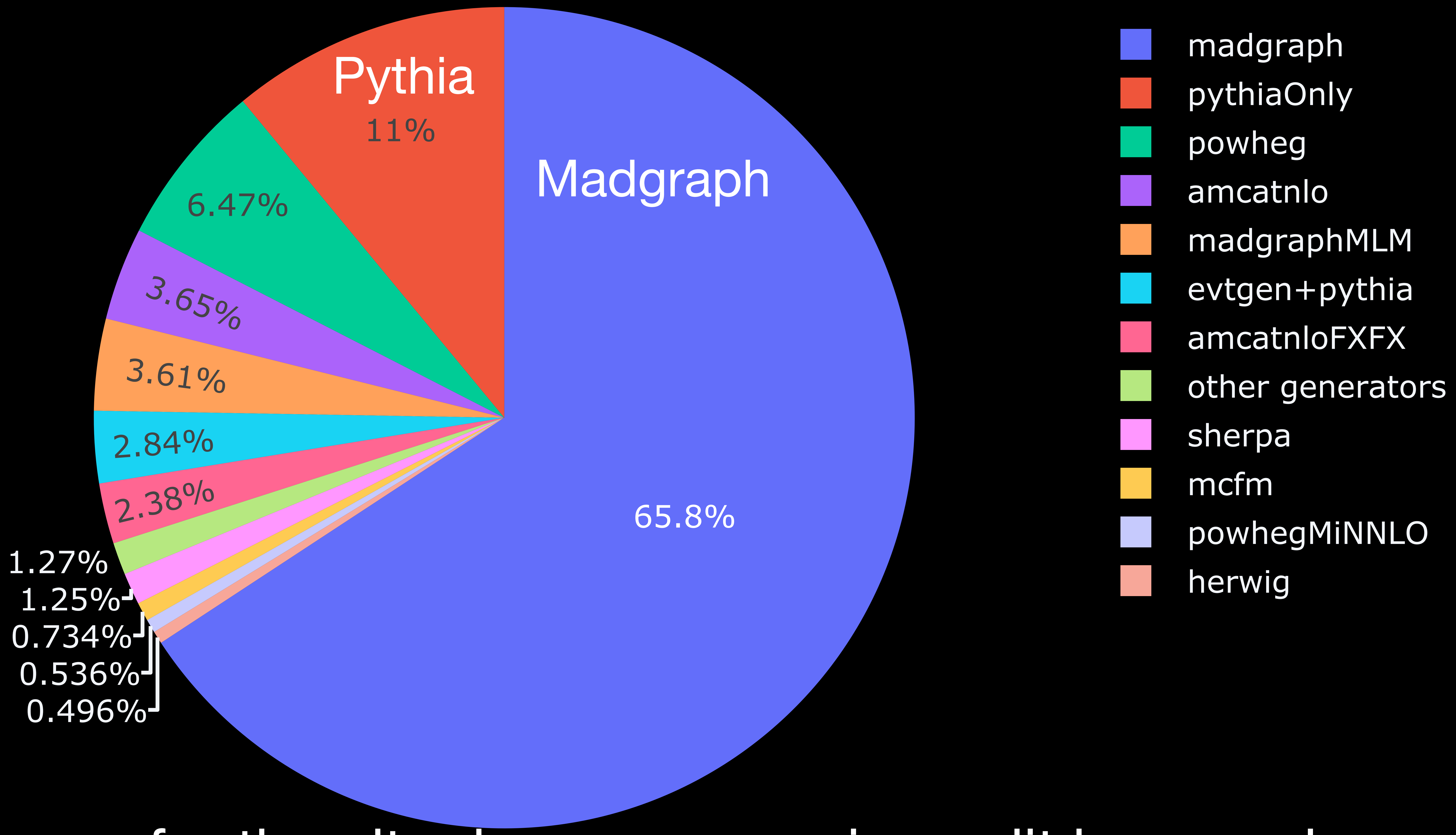


Generator Usage



Generator usage for the Ultra Legacy campaign split by events

Generator Usage

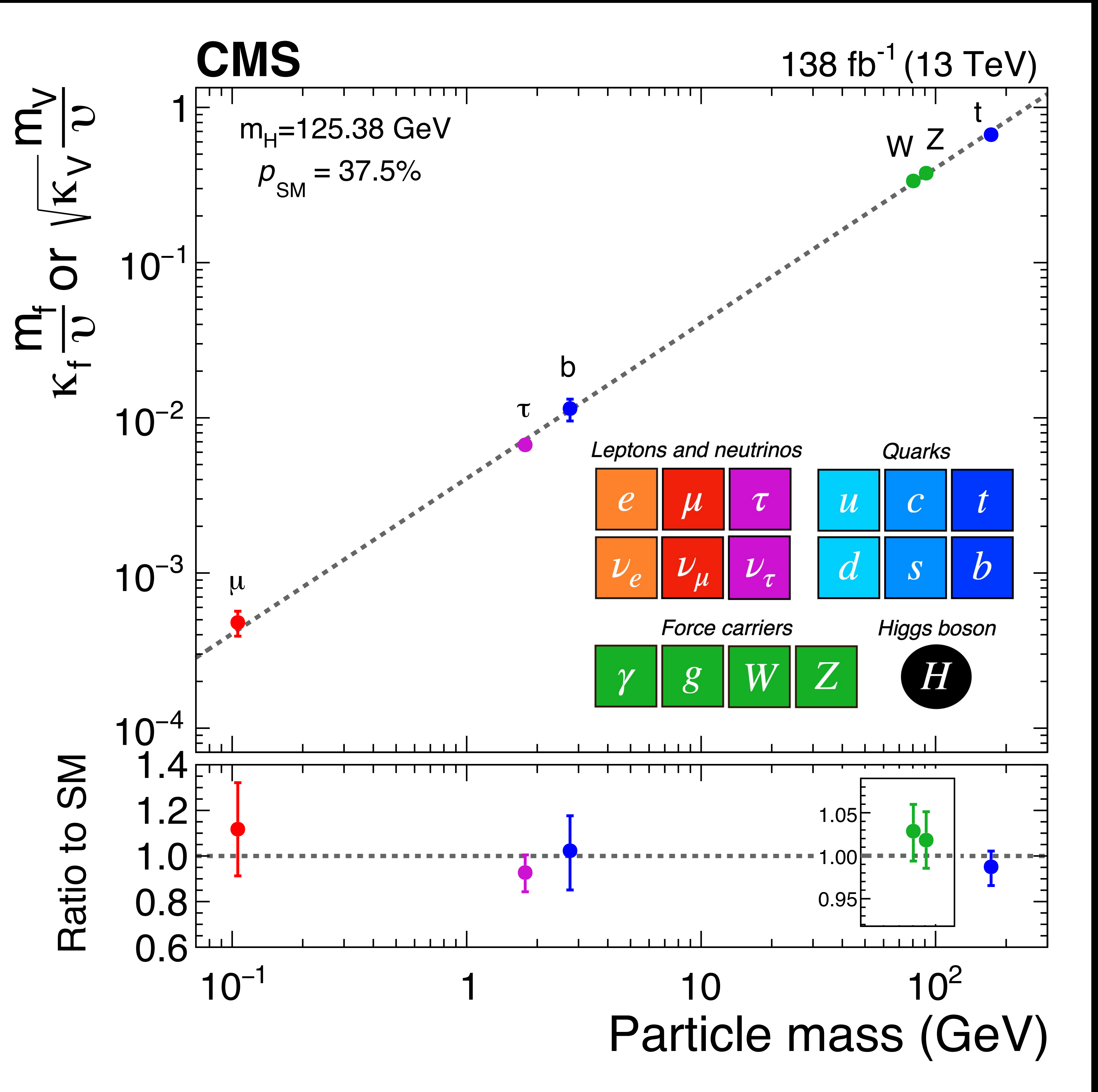


Generator usage for the ultra legacy campaign split by samples (each MC request)

The LHC is a precision machine

- For example, consider Higgs physics

- Higgs turned 10 years old in July 2022
    - 10 years of LHC = 8 million Higgs bosons!
    - From 5 σ to differential measurements in 360 weeks
- Major goal of the HL-LHC: nail down Higgs couplings at percent level

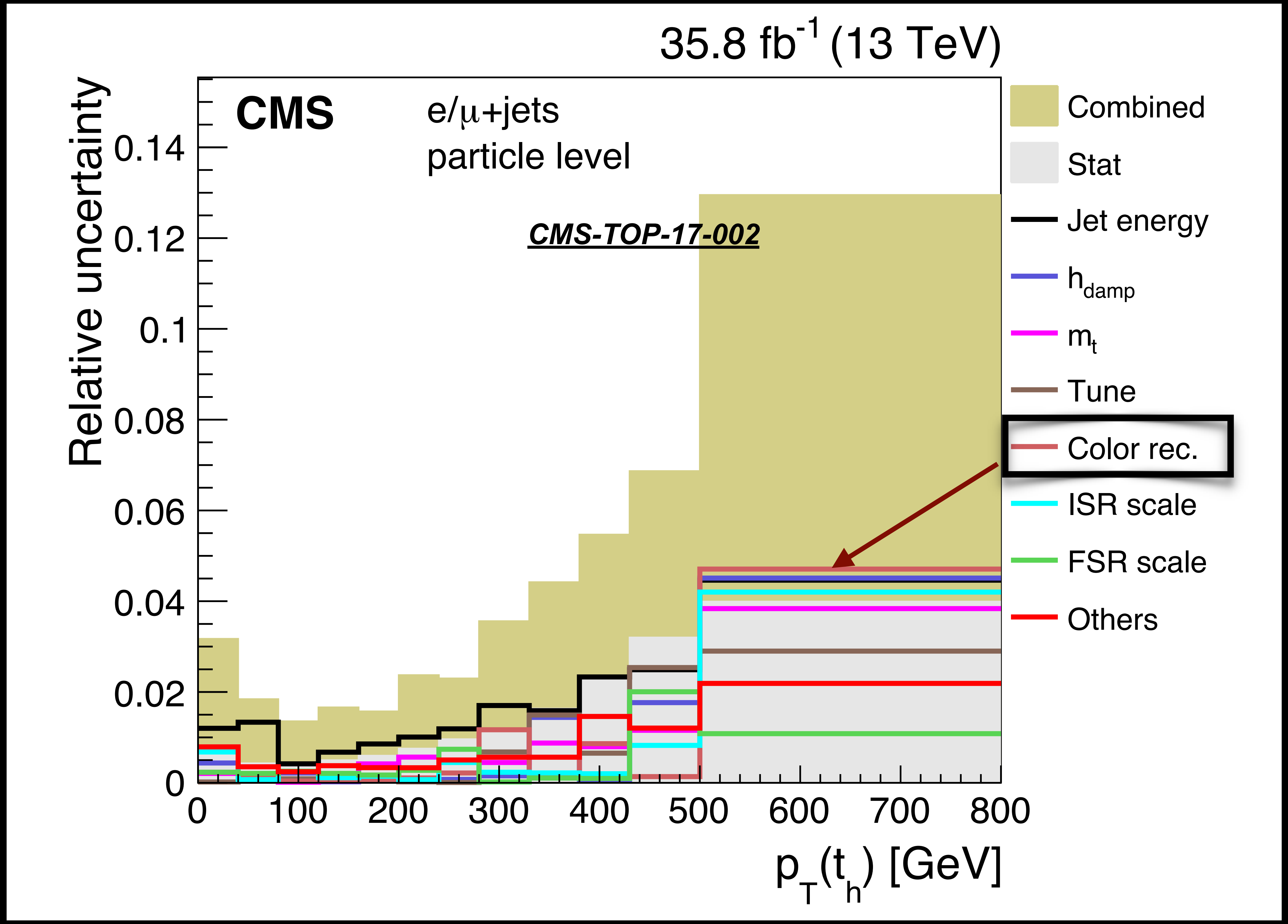


Outline of my talk

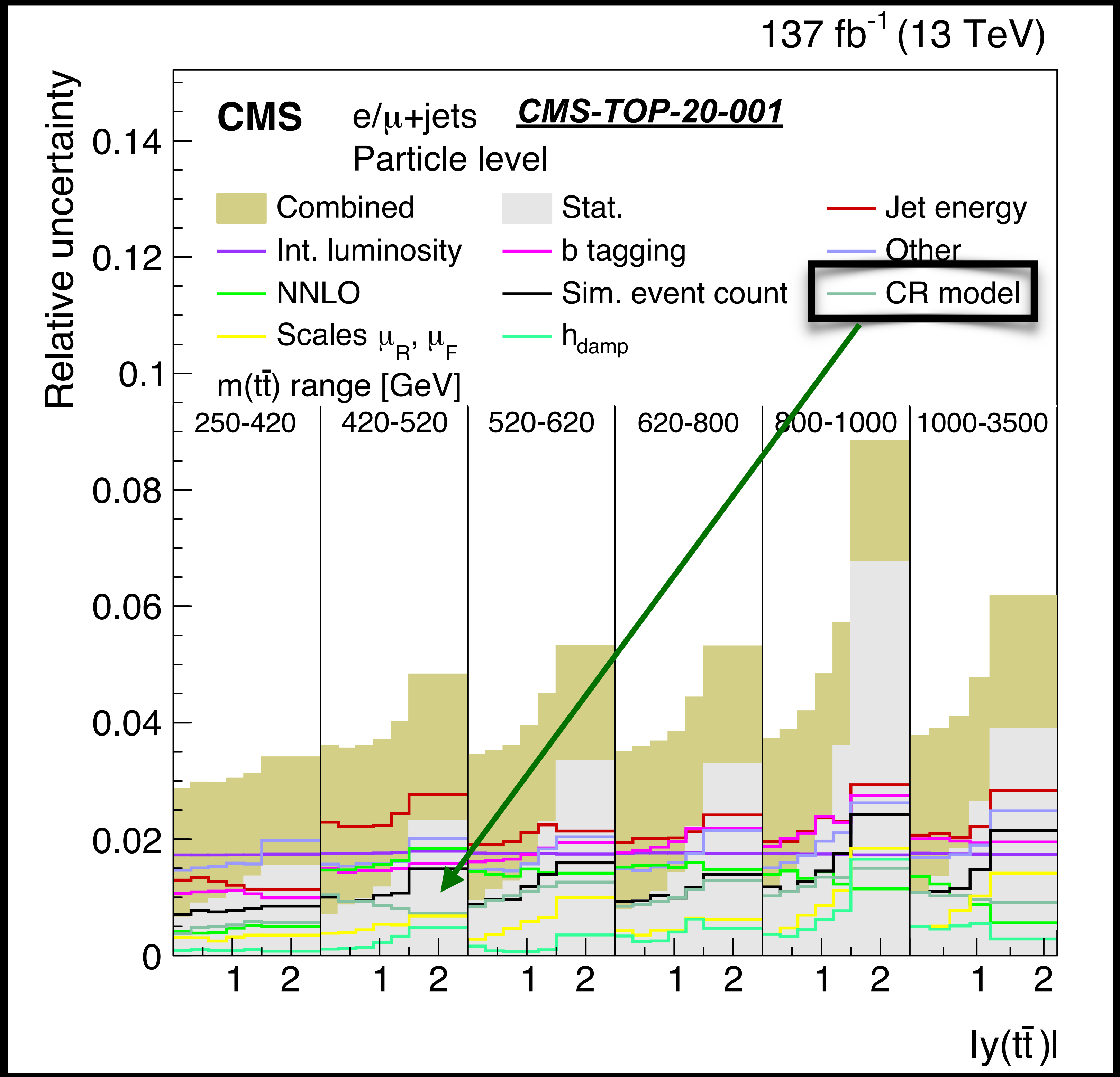
- The need for higher accuracy
  - What the data tells us!
- Improvements
  - V+Jets modeling
  - Third jet in vector boson fusion
  - Tuning studies: GEN-17-002 and what we are currently doing
- What does the generator group do?
  - Run III readiness
  - Common pitfalls

The LHC is a precision machine

Examples from the TOP sector



Measurement of differential cross sections for the production of top quark pairs and of additional jets in lepton+jets channel



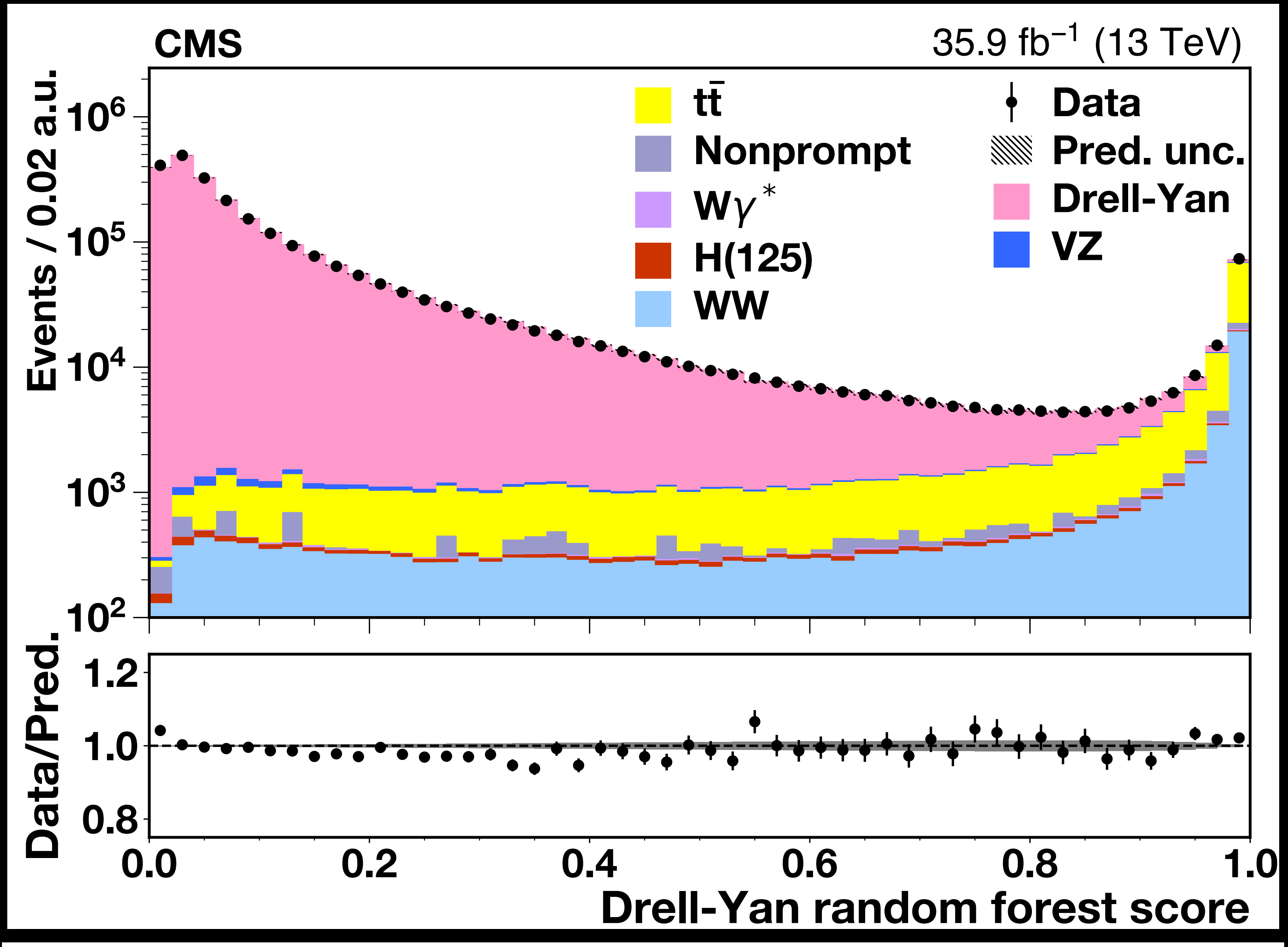
Measurement of differential tt production cross sections in the full kinematic range using lepton+jets events

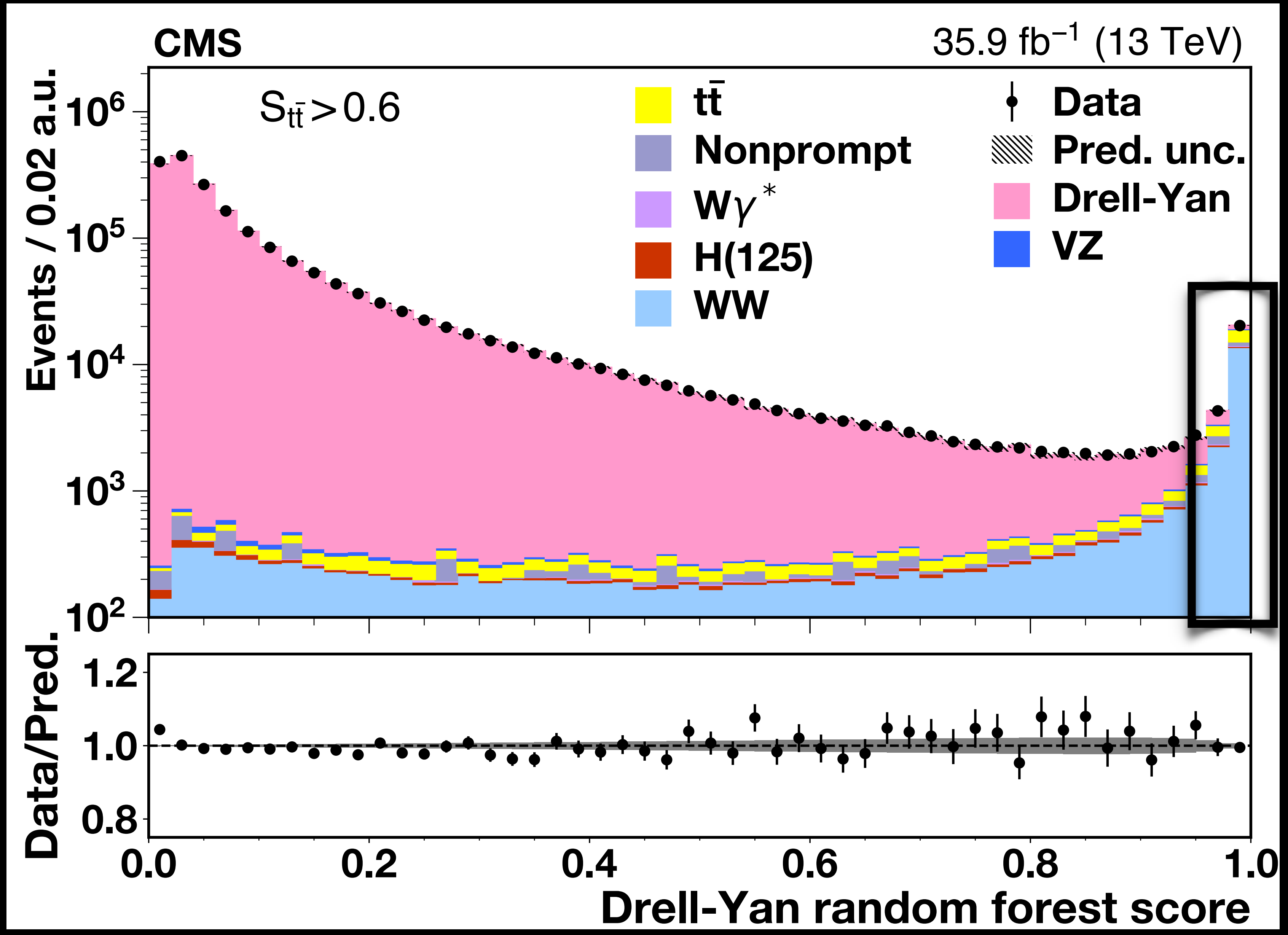
- In many cases theoretical uncertainties (including modeling and cross section uncertainties) are (becoming) dominant
- Experimental uncertainties, such as those associated with the luminosity expected to be lower (~1%) at the HL-LHC (Beam Radiation, Instrumentation, and Luminosity (BRIL-TDR))

The LHC is a precision machine

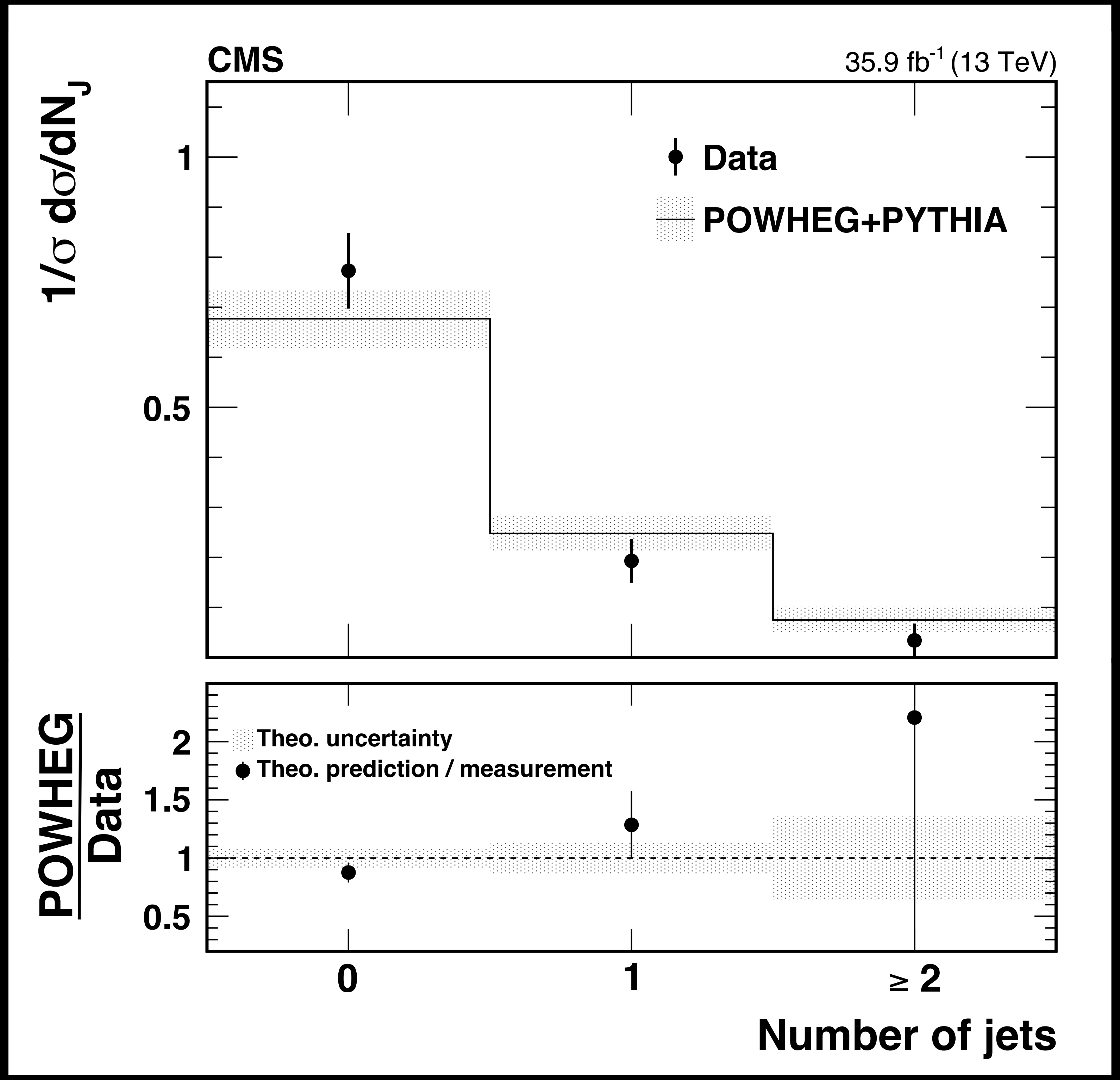
PhysRevD.102.092001

Example from the electroweak sector





- Tension in WW cross section observed at both 8 TeV and early 13 TeV runs at the LHC
- Event selection in jet binned category sensitive to higher order QCD corrections
- Use of random forest discriminators helps mitigate backgrounds without constructing exclusive jet-binned event selection: complements cut-based approach



Total cross section compared to NNLO prediction (with gg → WW incl.)

Using machine learning to probe inclusive regions of phase space: complements cut-based analyses

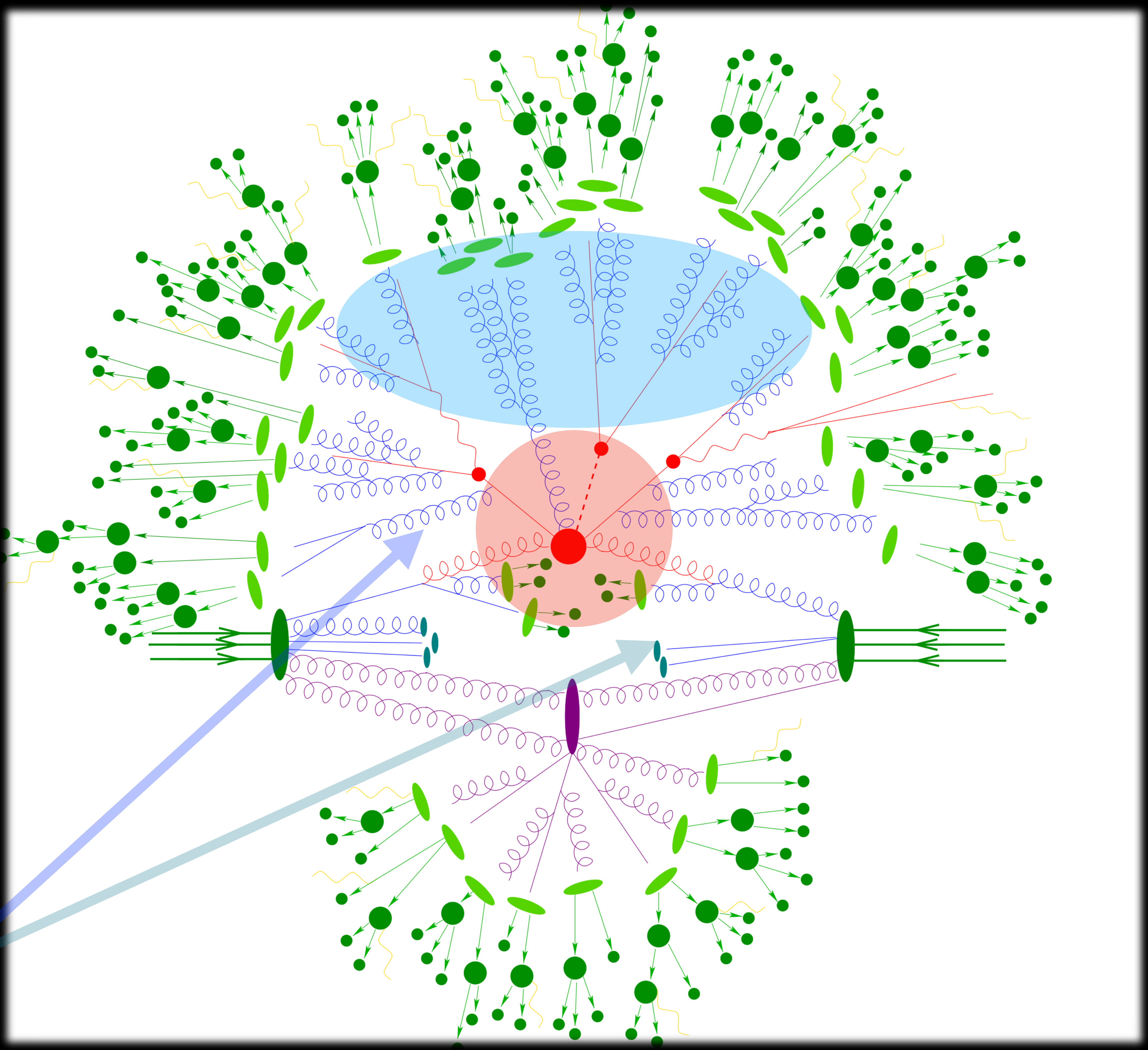
A full event description is possible!

Full event description includes:

Initial state parton shower
Signal process (hard interaction)
Final state parton shower
Fragmentation
Hadron decays
Beam remnants
Underlying event

Components of the underlying event:

Initial state parton shower
Final state parton shower
Beam remnants



$\mathcal{O}$ (1 GeV) $\mathcal{O}$ (1 TeV-10 GeV) $\mathcal{O}$ (1 TeV) $\mathcal{O}$ (1 TeV-10 GeV) $\mathcal{O}$ (1 GeV)

non-perturbative perturbative non-perturbative

Parton shower
$\mathcal{O}$ **(1 TeV - 1 GeV)**

Hard process:
$\mathcal{O}$ **(1 TeV)**

Multiple parton interactions increase with √s (due to increased partonic content)

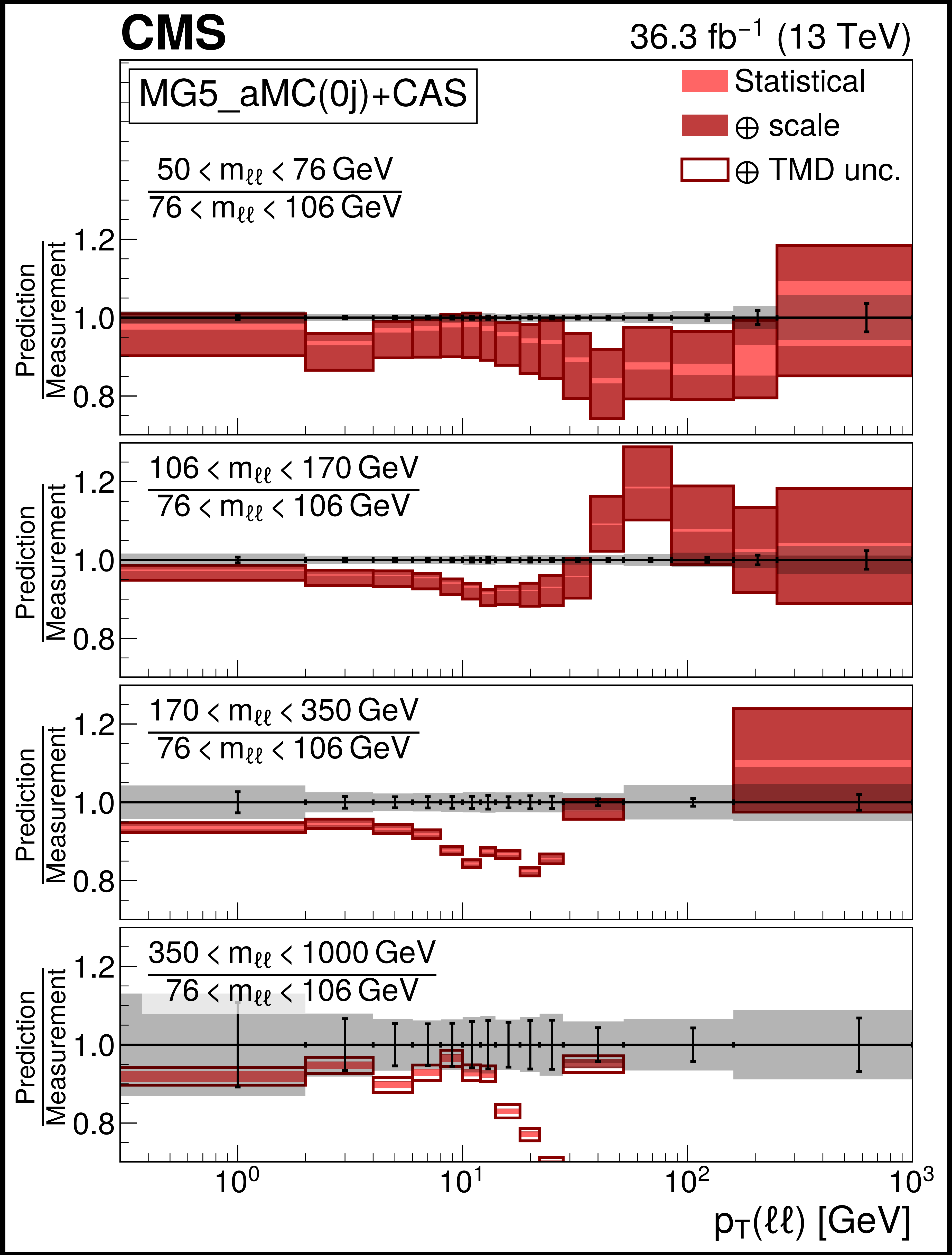
https://arxiv.org/pdf/hep-ph/0311270.pdf

Focusing on the hard scatter:

Probing Standard Model Lagrangian at energy scales far from the confinement scale

Studying the Z-p_T distribution — sensitive to new electroweak particles

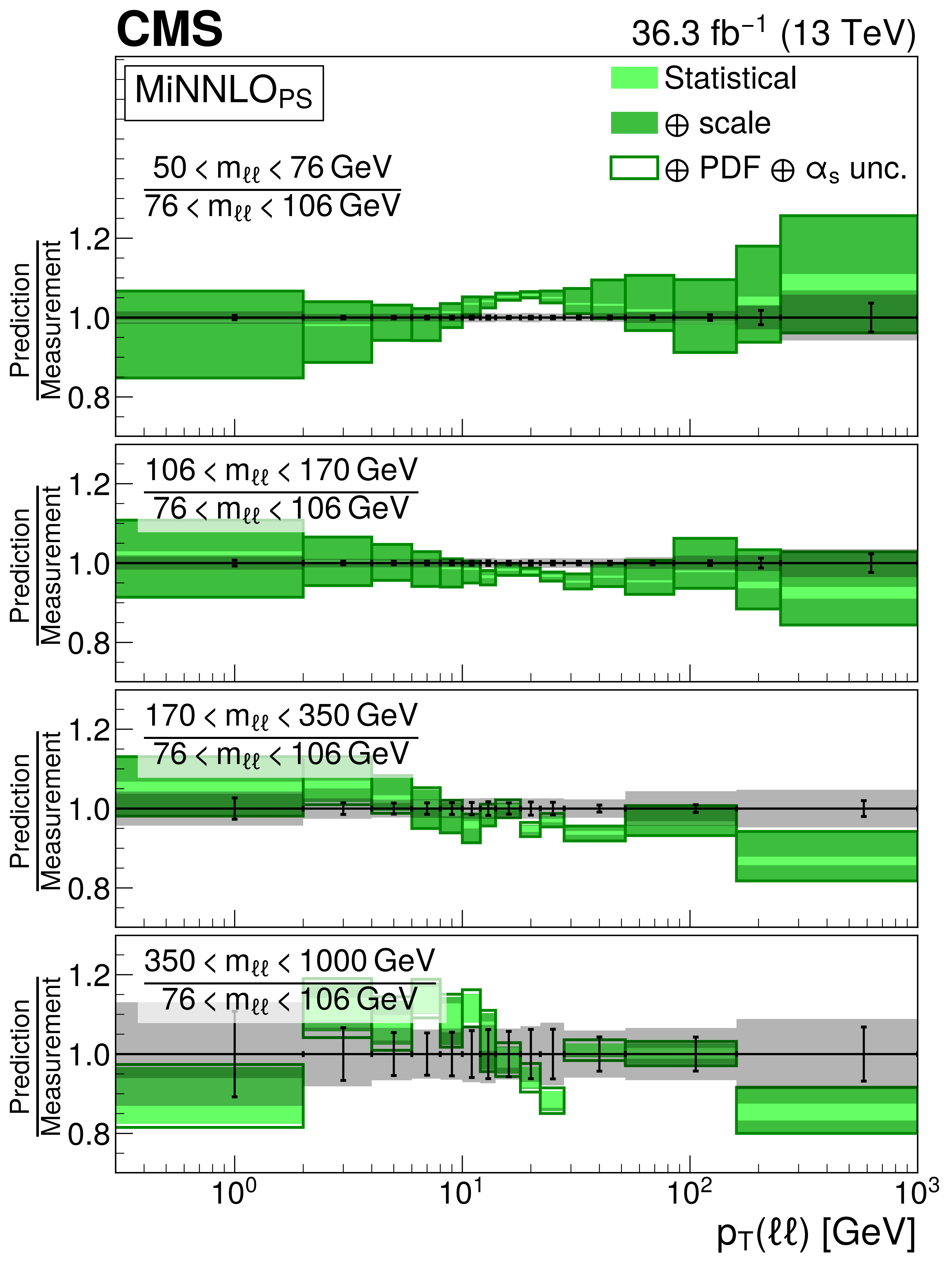
- Measurement of $p_T(\ell\ell)$ spectrum in a wide range of $m_{\ell\ell}$ bins is a test of the validity of resummation approach
- Calculation of inclusive Drell-Yan production as a function of $m_{\ell\ell}$ and $p_T(\ell\ell)$ available up to next-to-next-to-leading order in perturbative QCD
- Possible to calculate soft gluon resummation analytically in transverse momentum dependent parton distributions (TMD) or in parton showers of MC generators



https://arxiv.org/abs/2205.04897

Studying the Z-p_T distribution

- Event sample at NNLO with jet merging generated with MiNNLO$_{PS}$ (arXiv:1908.06987)
- α_s evaluated independently at each vertex (scale determined by kinematic configuration)
- Similar approach adopted for the W-mass analysis
- Low fraction of negative weights

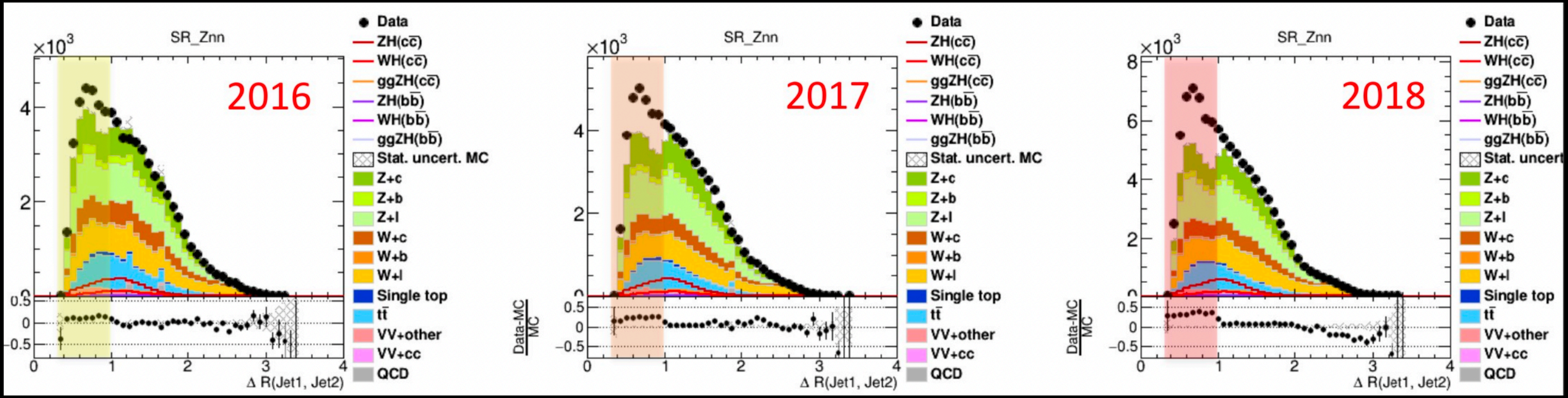


https://arxiv.org/abs/2205.04897

Areas of improvement

V+Jets

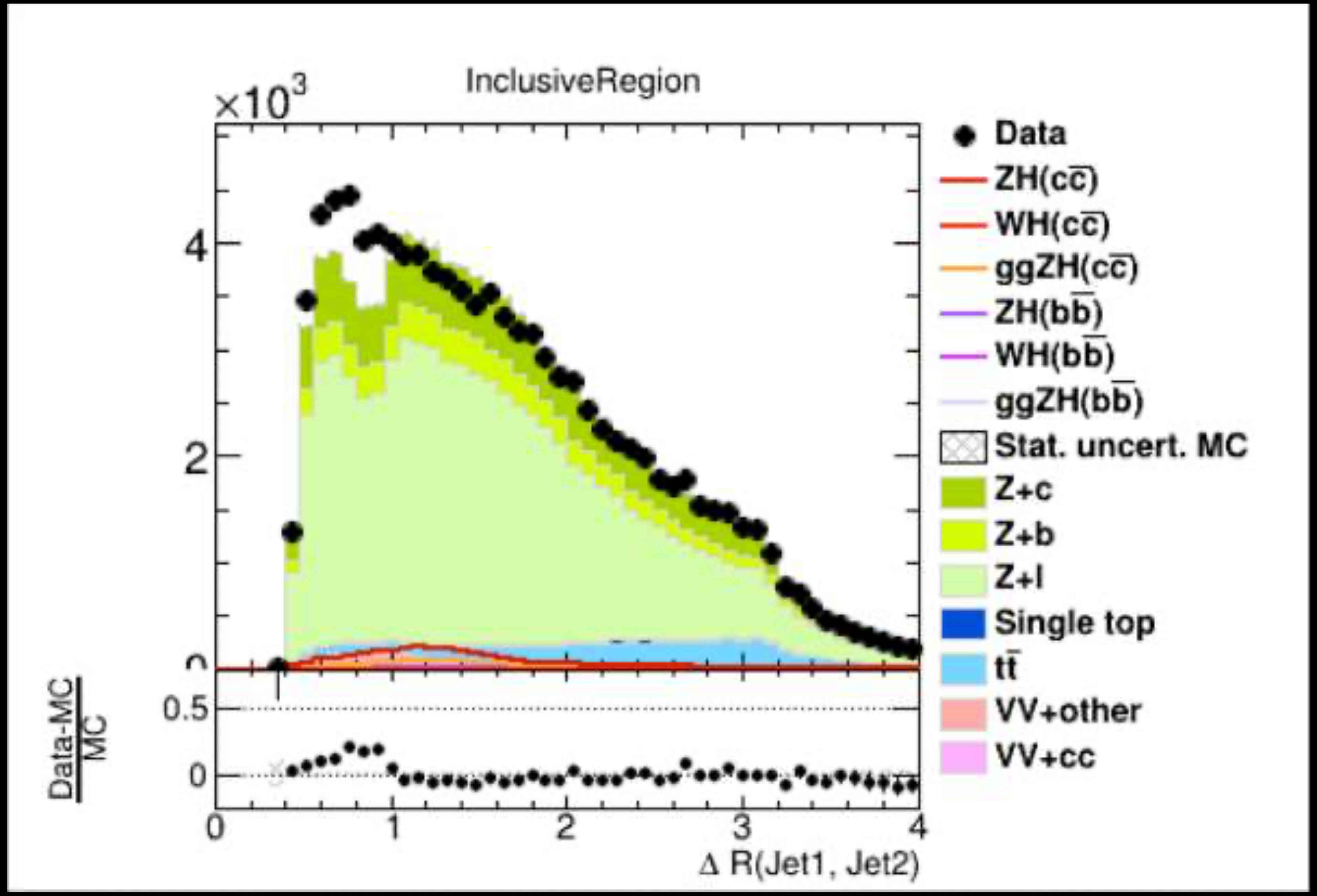
- NLO V+jets generated with Madgraph_aMC@NLO
- Best description of data not achieved in the low $\Delta R(j,j)$ region — boosted region
- Studies of matching and merging parameters between matrix element (ME) and parton shower (PS) models currently underway
  - Are the parameters chosen for FxFx merging and matching ideal? Need to revisit them?
- Disagreement slightly worse for CP5 tune — major effort has taken shape to address these specific regions of phase space



Luca Mastrolorenzo's talk

V+Jets

- Crucial to understand modeling issues in this region of phase space
  - Can lead to significant uncertainty: main source of uncertainty in ATLAS VH(bb)
  - MiNNLO sample generated for W-mass studies, plan to generate Sherpa samples, recent developments on NLO+PS simulations with Herwig Matchbox
  - Discussion of details like 2D binning (p_T and mass) of samples underway



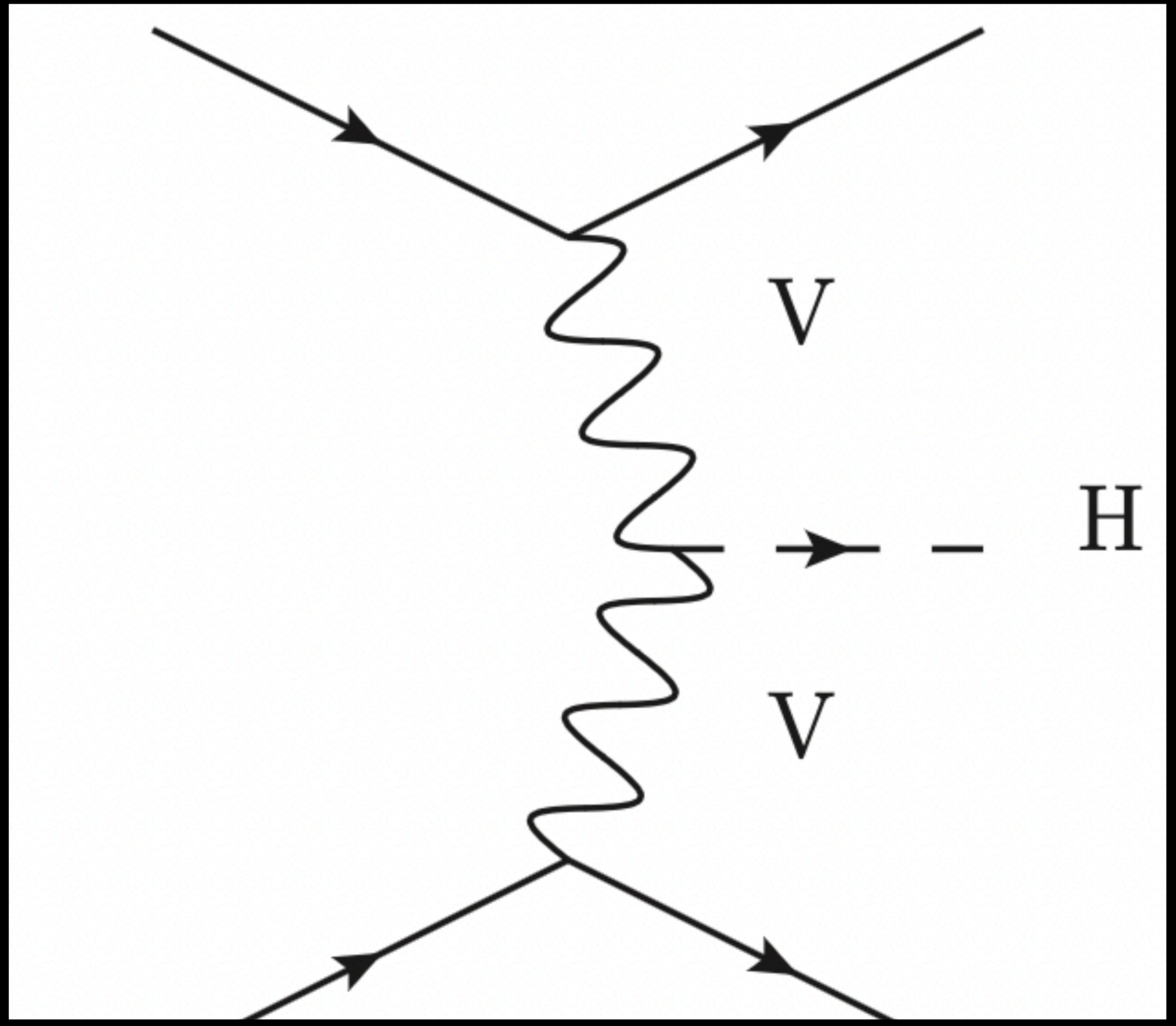
Similar behavior observed for W+Jets

Luca Mastrolorenzo's talk

Vector Boson Fusion (VBF) Modeling

https://arxiv.org/pdf/2003.12435.pdf

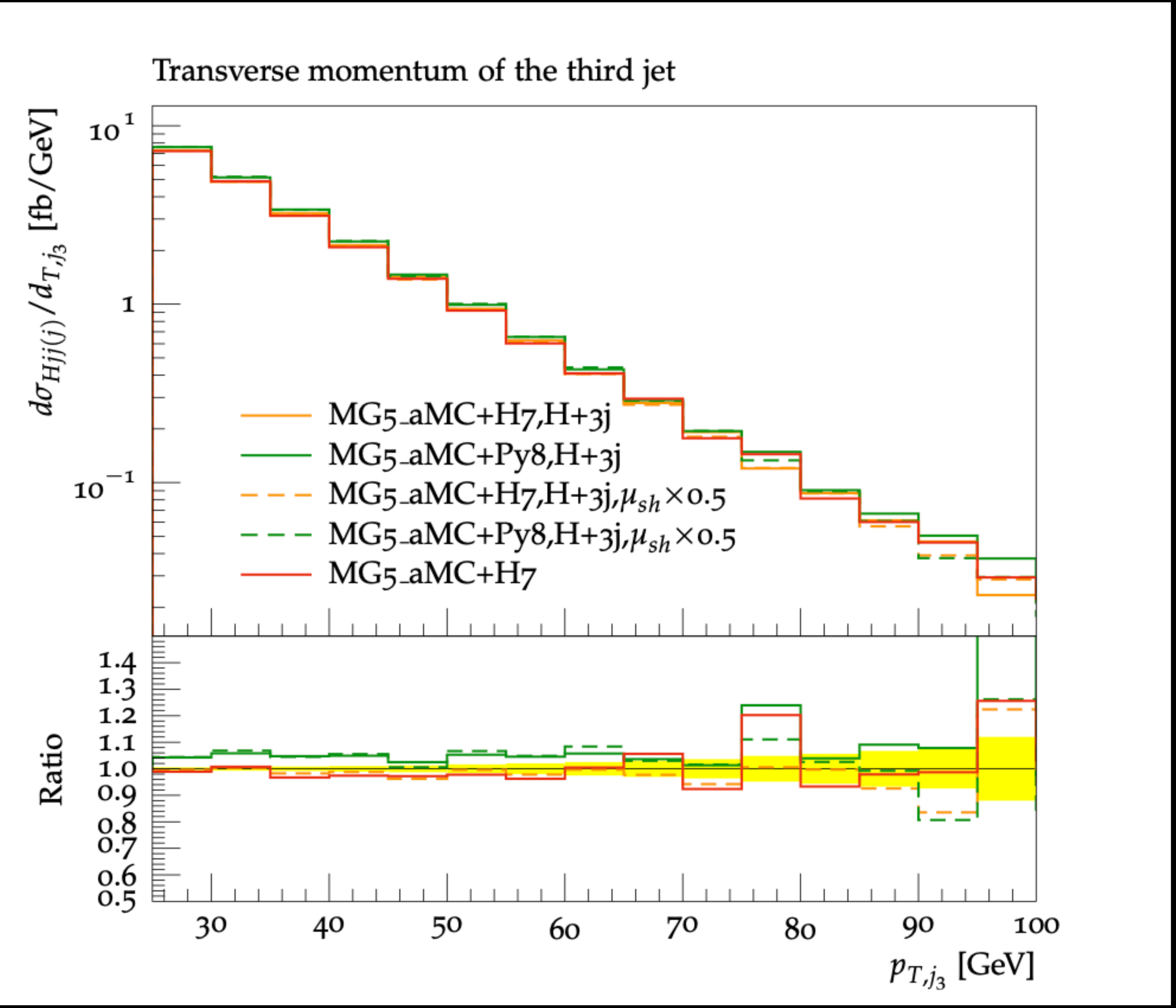
- VBF Higgs production enables probing of HVV coupling: one of the most basic properties of Higgs mechanism
- Large theoretical uncertainties associated with VBF modeling
- Behavior of third jet merits investigation → sensitive to influence of PS modeling
- Sources of possible variations arise from choice of generator, matching scheme, shower MC program, recoil scheme



generator	matching	SMC	shower recoil	used in comparison
VBFNLO+Herwig7/Matchbox	$\oplus$	HERWIG 7.1.5	global ($\tilde{q}$) / local (dipole)	✓($\tilde{q}$)
HJets+Herwig7/Matchbox	$\oplus$	HERWIG 7.1.5	global ($\tilde{q}$) / local (dipole)	
MadGraph5_aMC@NLO 2.6.1	$\oplus$	HERWIG 7.1.2	global	✓
MadGraph5_aMC@NLO 2.6.1	$\oplus$	PYTHIA 8.230	global	
POWHEG-BOX V2	$\otimes$	PYTHIA 8.240	local (dipole)	✓
POWHEG-BOX V2	$\otimes$	PYTHIA 8.240	global	
POWHEG-BOX V2	$\otimes$	HERWIG 7.1.4	global ($\tilde{q}$)	

Vector Boson Fusion (VBF) Modeling

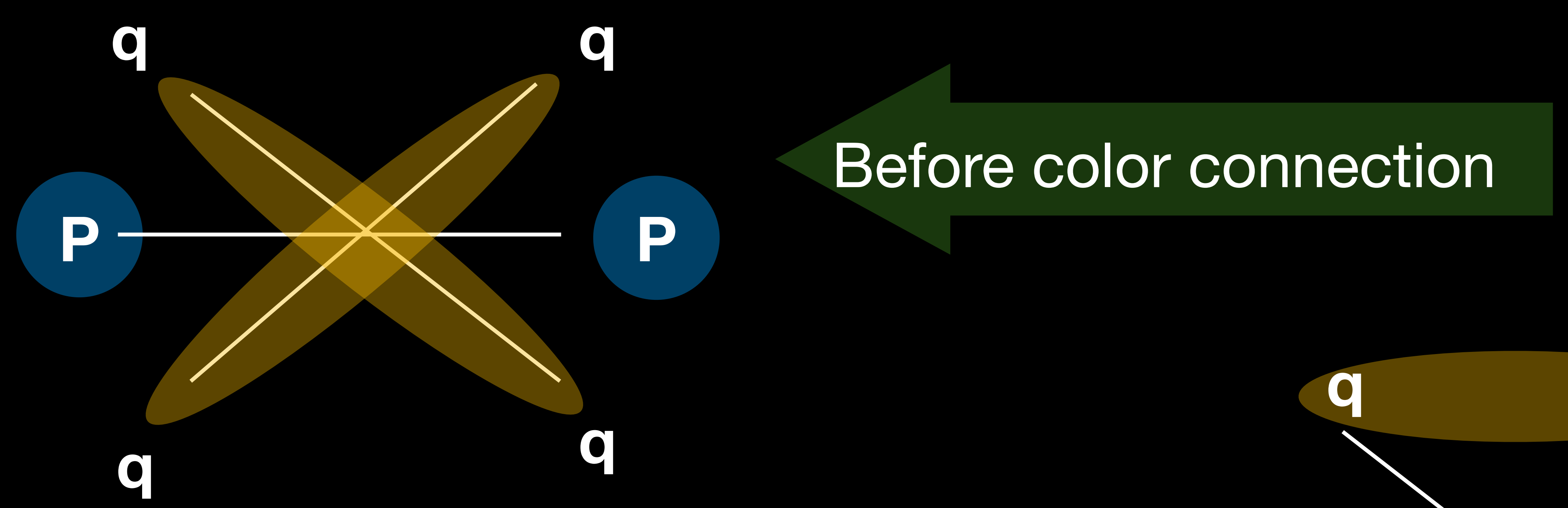
- Inclusive (in the number of jets) observable — p_T of the Higgs modeled well by all generators
- Modeling differences in jet bins almost exactly compensate
- Little to no dependence on choice of renormalization, factorization and shower scales
- Global dipole recoil scheme clearly incorrect
- Study of the impact of the 3rd jet presented at GEN meeting on July 26th, 2021

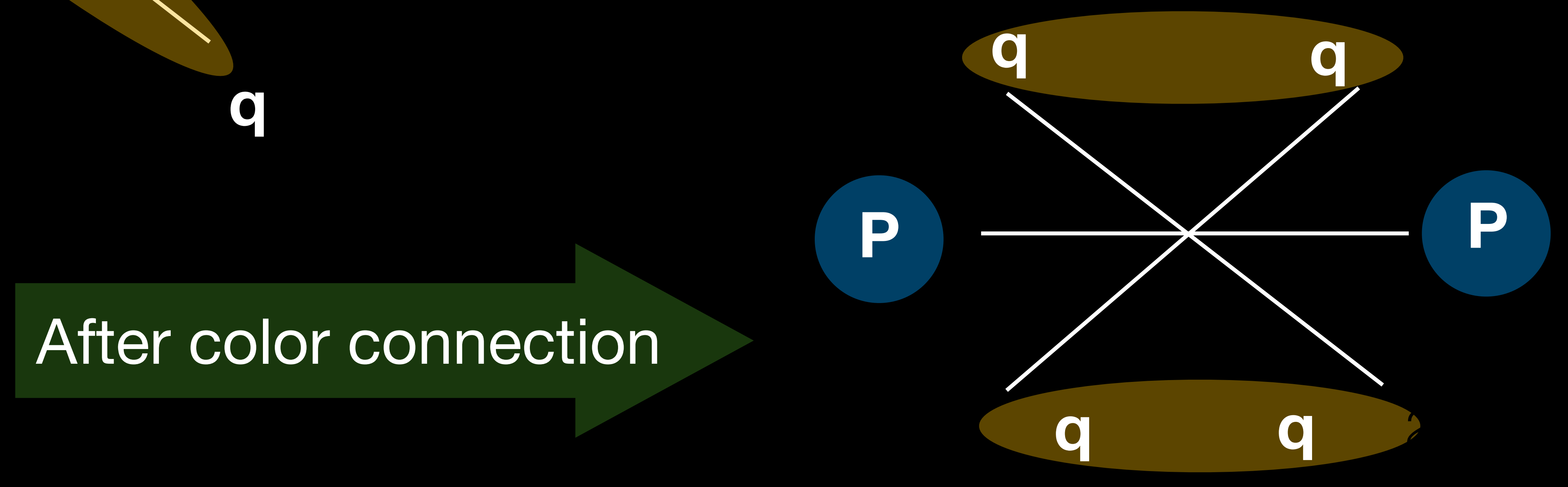


Focusing on the non-perturbative regime

Color reconnection (CR)

- Color reconnection (CR) reconfigures color strings after parton shower
- Performed in the non-perturbative regime
- CR effects dominant in the low p_T region (p_T 2~5 GeV)
- Incorrect color associations can lead to large differences ⇒ unphysical
- Precision at the LHC necessitates an understanding of the color reconnection modeling



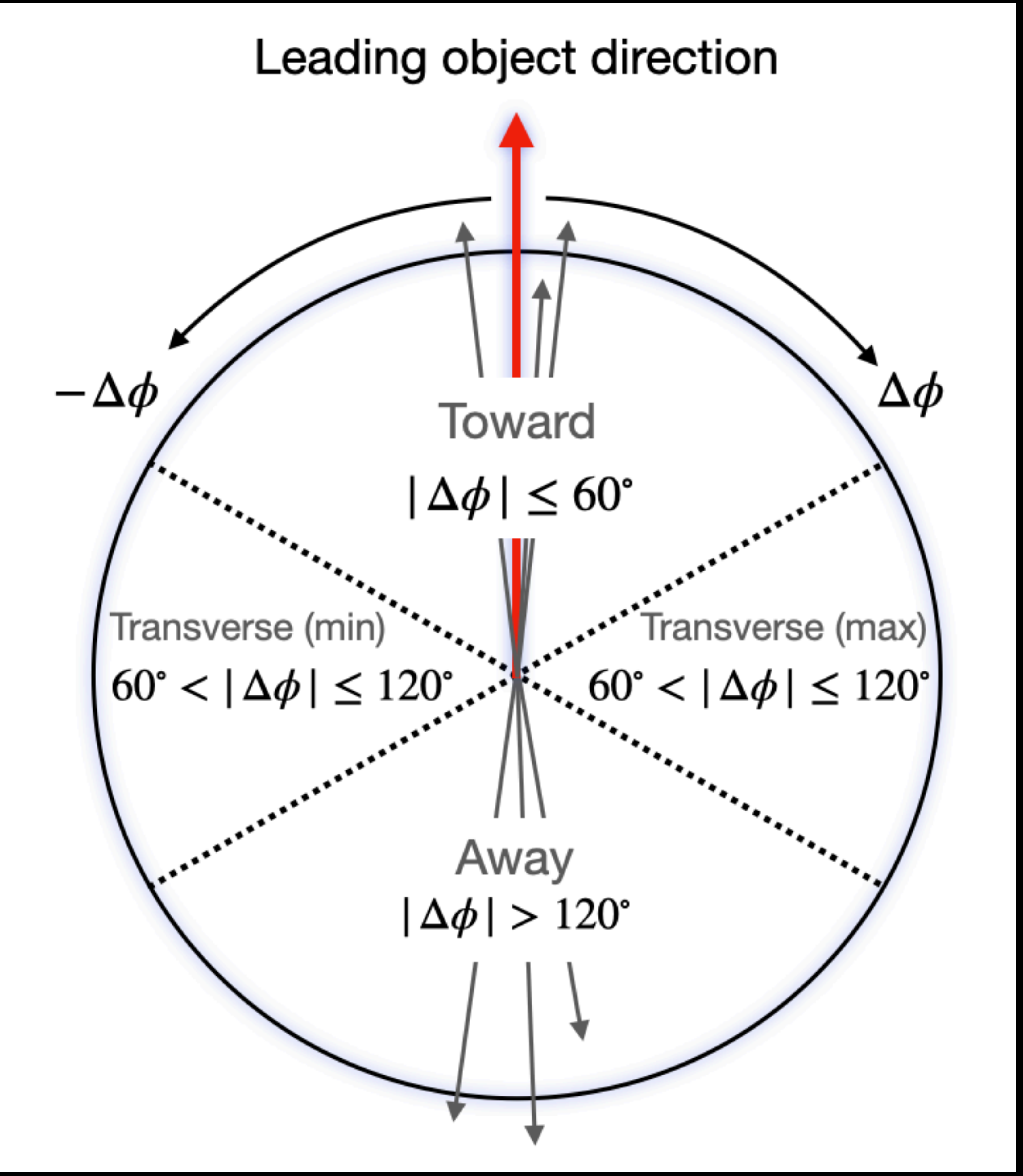


	δm_t^{hyb} [GeV]		
	all-jets	ℓ+jets	combination
Experimental uncertainties			
Method calibration	0.06	0.05	0.03
JEC (quad. sum)	0.15	0.18	0.17
– Intercalibration	−0.04	+0.04	+0.04
– MPFInSitu	+0.08	+0.07	+0.07
– Uncorrelated	+0.12	+0.16	+0.15
Jet energy resolution	−0.04	−0.12	−0.10
b tagging	0.02	0.03	0.02
Pileup	−0.04	−0.05	−0.05
All-jets background	0.07	–	0.01
All-jets trigger	+0.02	–	+0.01
ℓ+jets background	–	+0.02	−0.01
Modeling uncertainties			
JEC flavor (linear sum)	−0.34	−0.39	−0.37
– light quarks (uds)	+0.07	+0.06	+0.07
– charm	+0.02	+0.01	+0.02
– bottom	−0.29	−0.32	−0.31
– gluon	−0.13	−0.15	−0.15
b jet modeling (quad. sum)	0.09	0.12	0.06
– b frag. Bowler–Lund	−0.07	−0.05	−0.05
– b frag. Peterson	−0.05	+0.04	−0.02
– semileptonic b hadron decays	−0.03	+0.10	−0.04
PDF	0.01	0.02	0.01
Ren. and fact. scales	0.04	0.01	0.01
ME/PS matching	+0.24	−0.07	+0.07
ME generator	–	+0.20	+0.21
ISR PS scale	+0.14	+0.07	+0.07
FSR PS scale	+0.18	+0.13	+0.12
Top quark p_T	+0.03	−0.01	−0.01
Underlying event	+0.17	−0.07	−0.06
Early resonance decays	+0.24	−0.07	−0.07
CR modeling (max. shift)	−0.36	+0.31	+0.33
– "gluon move" (ERD on)	+0.32	+0.31	+0.33
– "QCD inspired" (ERD on)	−0.36	−0.13	−0.14
Total systematic	0.70	0.62	0.61
Statistical (expected)	0.20	0.08	0.07
Total (expected)	0.72	0.63	0.61

Brief aside: The underlying event (UE)

GEN-17-002

- UE defined as activity not associated with particles originating from the hard scatter
- Generally studied in events that contain hard scattering with pT ≳ 2 GeV
- Leading object defined on an event-by-event basis
- Φ regions relative to the leading object that are sensitive to the underlying event
- Azimuthal separation between charged particles and leading object $\Delta\Phi = \Phi - \Phi_{max}$ used to define sensitive regions



Description of the tunes

GEN-17-002

- **CR1 stands for the QCD-inspired model** and **CR2 stands for the gluon-move CR model**
  - Complements CP5 with color reconnection variations
- **CP5 default:** CP5 uses NNPDF31_nnlo_as_0118 PDF set, α_S = 0.118, and the MPI-based CR model

			CP5-CR1			CP5-CR2		
RIVET routine	$\sqrt{s}$ (TeV)	Distribution	Fit range (GeV)	N_{bins}	R	Fit range (GeV)	N_{bins}	R
CMS_2015_I1384119	13	N_{ch} vs η		20	1		20	1
CMS_2015_PAS_FSQ_15_007	13	TransMIN charged p_T^{sum}	2–28	15	1	3–36	15	0.5
		TransMAX charged p_T^{sum}	2–28	15	1	3–36	15	0.5
		TransMIN N_{ch}	2–28	15	1	3–36	15	0.1
		TransMAX N_{ch}	2–28	15	1	3–36	15	0.1
CMS_2012_PAS_FSQ_12_020	7	TransMAX N_{ch}	3–20	10	1	3–20	10	0.1
		TransMIN N_{ch}	3–20	10	1	3–20	10	0.1
		TransMAX charged p_T^{sum}	3–20	10	1	3–20	10	0.1
		TransMIN charged p_T^{sum}	3–20	10	1	3–20	10	0.1
CDF_2015_I1388868	2	TransMIN N_{ch}	2–15	11	1	2–15	11	0.1
		TransMAX N_{ch}	2–15	11	1	2–15	11	0.1
		TransMIN charged p_T^{sum}	2–15	11	1	2–15	11	0.1
		TransMAX charged p_T^{sum}	2–15	11	1	2–15	11	0.1

Tuning Parameters

GEN-17-002

- Tuning variables chosen based on modeling of multiple parton interactions **(MPI)**
- Key Pythia8 parameter p_{T0} (function of √s): regularizes primary hard-scattering processes and MPI
- Energy dependence:

$$p_{T0}(\sqrt{s}) = p_{T0}^{\text{ref}} \left(\frac{\sqrt{s}}{\sqrt{s}_0} \right)^{\epsilon}$$

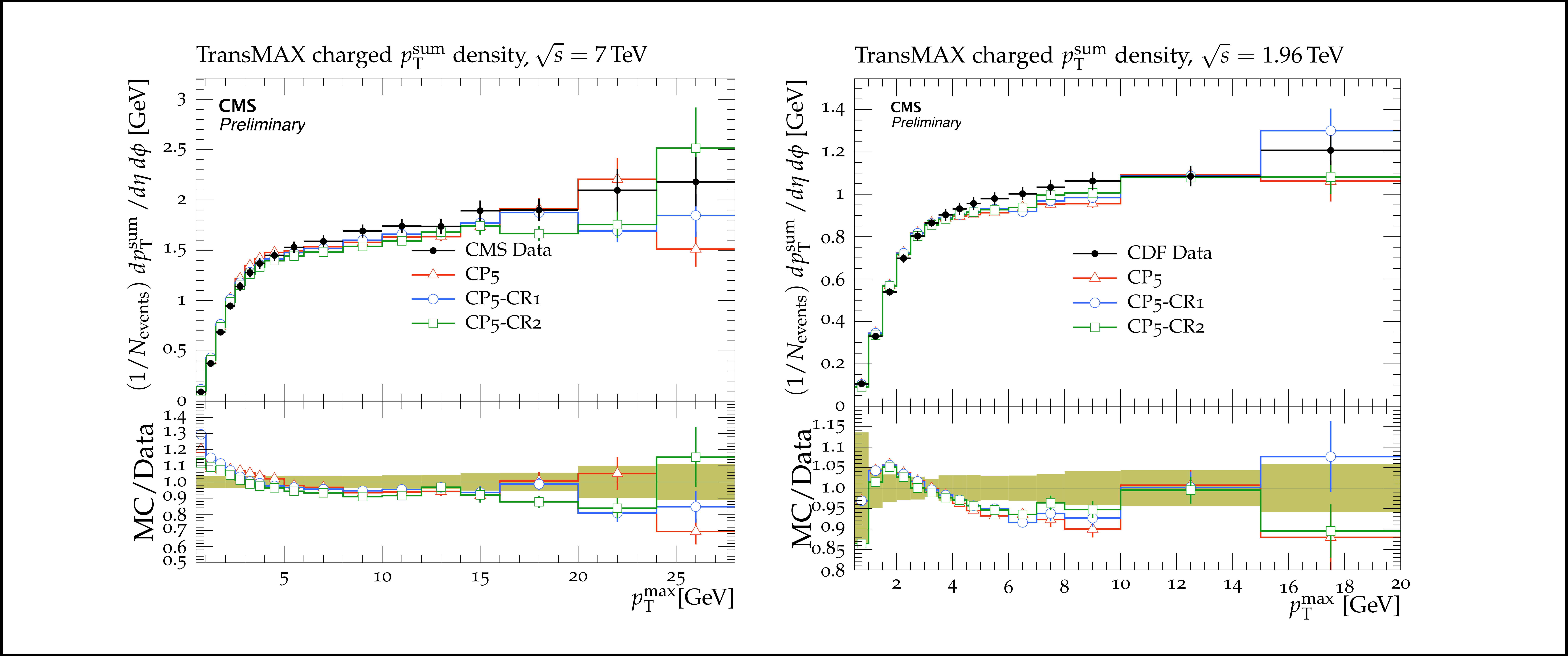
Tunable parameter (ϵ)

Reference energy ($\sqrt{s}_0$)

- coreRadius: width of the core when a double Gaussian profile is assumed for the overlap distribution between the two colliding protons. A double Gaussian core identifies an inner dense part, which is called the core
- coreFraction: the fraction of quark and gluon content enclosed in the core when a double Gaussian matter profile is assumed
- **MPI-based modeling** assigns a probability to each parton pair to reconnect with a harder system high p_T → less likely to be color connected $P = \frac{p_{T\text{Rec}}^2}{p_{T\text{Rec}}^2 + p_T^2}$ where, $p_{T\text{Rec}} = R \cdot p_{T_0}$
- $p_{T\text{Rec}}$ is a regularization term that prevents divergence of partonic cross sections at low p_T, R is a parameter

Performance of the tune

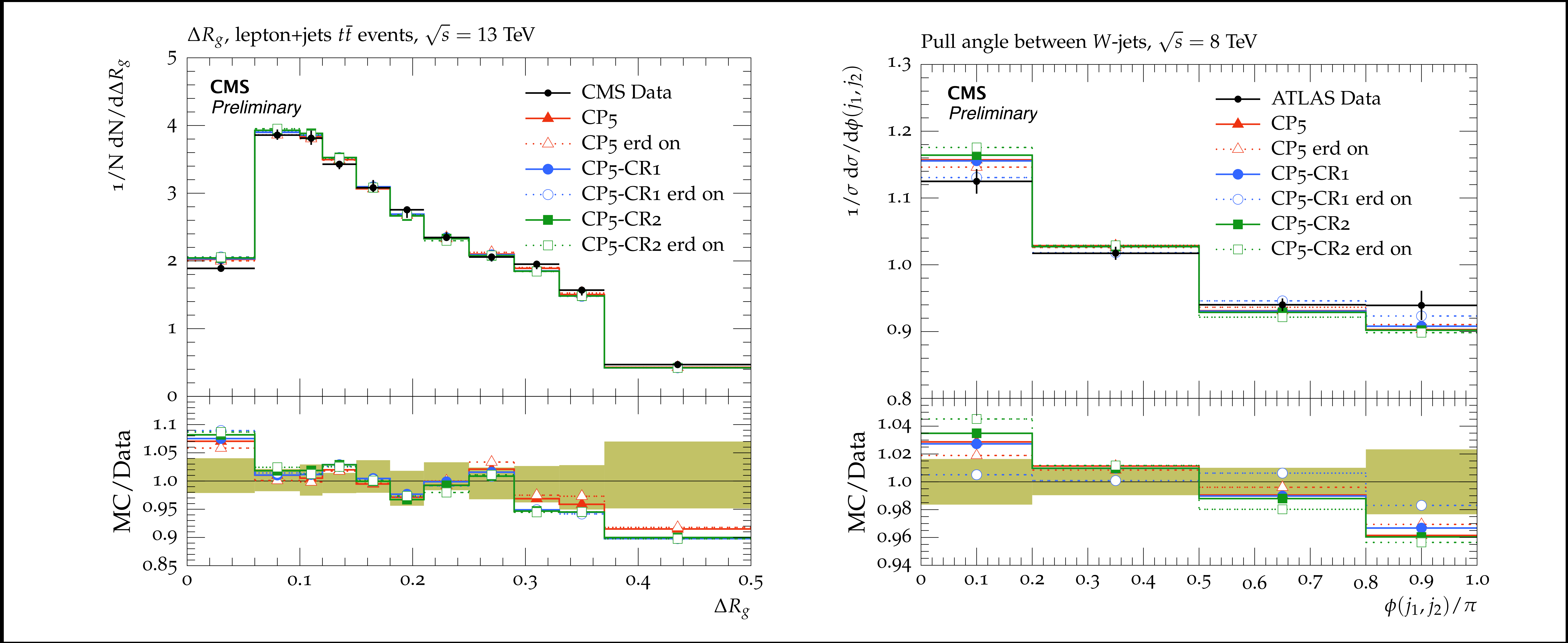
GEN-17-002



- Charged particle density in the transMAX region with CMS (left) and CDF (right) data
- Similar behavior seen at $\sqrt{s}$ = 13 TeV and with ATLAS data
- Tunes perform well, inconsistencies seen in final states with charmed baryons (Λ)

Jet substructure variables in $t\bar{t}$ final states

GEN-17-002



- ΔR_g: angle between two groomed subjets
- $\phi(j_1, j_2)/\pi$: pull angle between jets from the W-boson in top decays (using charged constituents of the jets)
- Early resonance decay (ERD) option: color reconnection before and after the top quark decay

Extraction of the uncertainty on top-mass using tunes

Table 4: The top quark mass, m_{t}, and W mass, m_{W}, extracted by a fit to the predictions of the different PYTHIA8 tunes. The uncertainties in the m_{t} and m_{W} values correspond to the uncertainty in the fitted m_{t} and m_{W}.

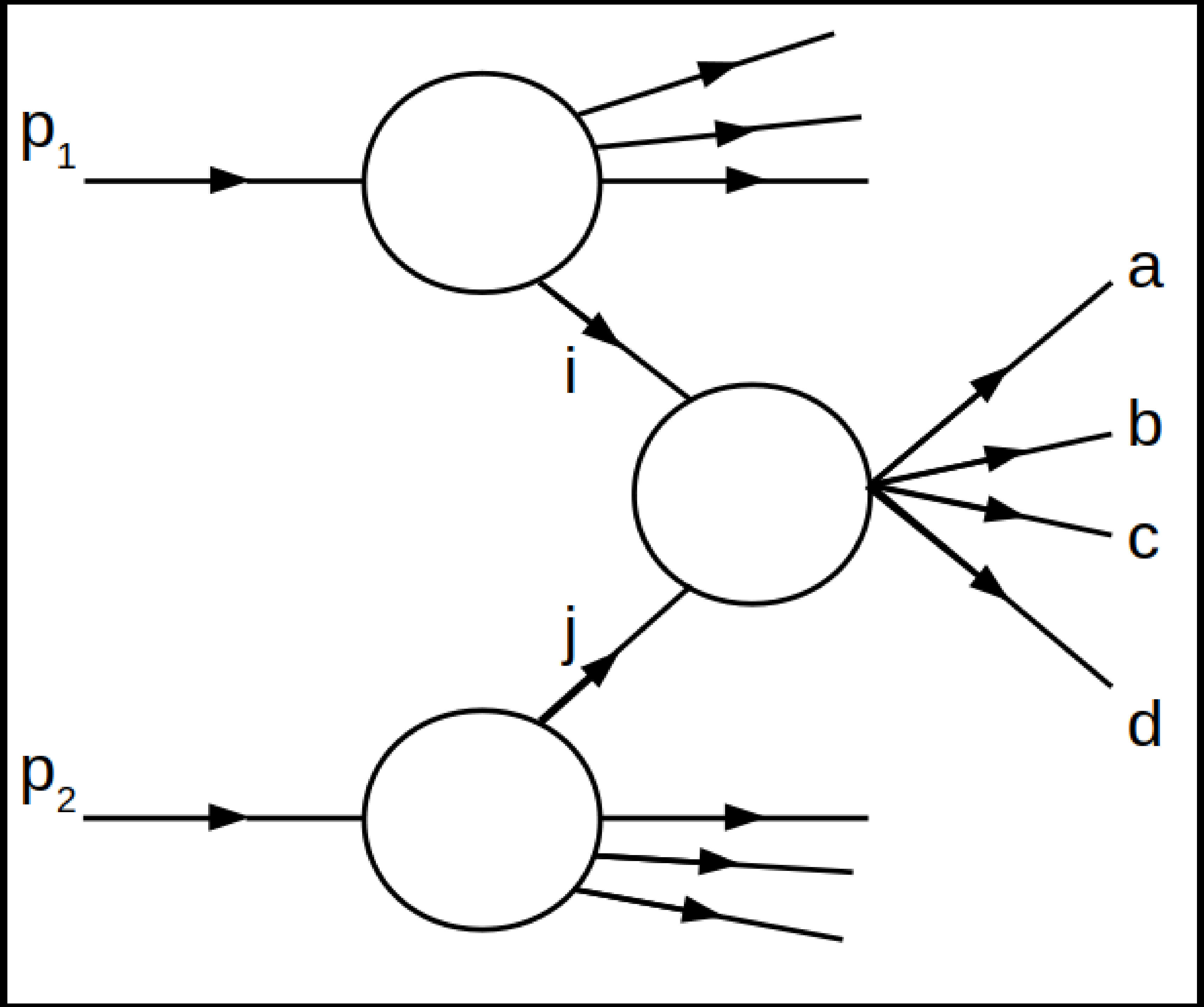
Tune	m_{t} [GeV]	Δm_{t} [GeV]	m_{W} [GeV]	Δm_{W} [GeV]	$\Delta m_{\mathrm{t}} - 0.5 \times \Delta m_{\mathrm{W}}$ [GeV]
CP5	171.93 ± 0.02	0	79.76 ± 0.02	0	0
CP5 erdOn	172.18 ± 0.03	0.25	80.15 ± 0.02	0.40	0.13
CP5-CR1	171.97 ± 0.02	0.04	79.74 ± 0.02	-0.02	0.05
CP5-CR1 erdOn	172.01 ± 0.03	0.08	79.98 ± 0.02	0.23	-0.04
CP5-CR2	171.91 ± 0.02	-0.02	79.85 ± 0.02	0.10	-0.07
CP5-CR2 erdOn	172.32 ± 0.03	0.39	79.90 ± 0.02	0.14	0.32

Allows comparison with TOP-17-007, factoring in shifts in Mw

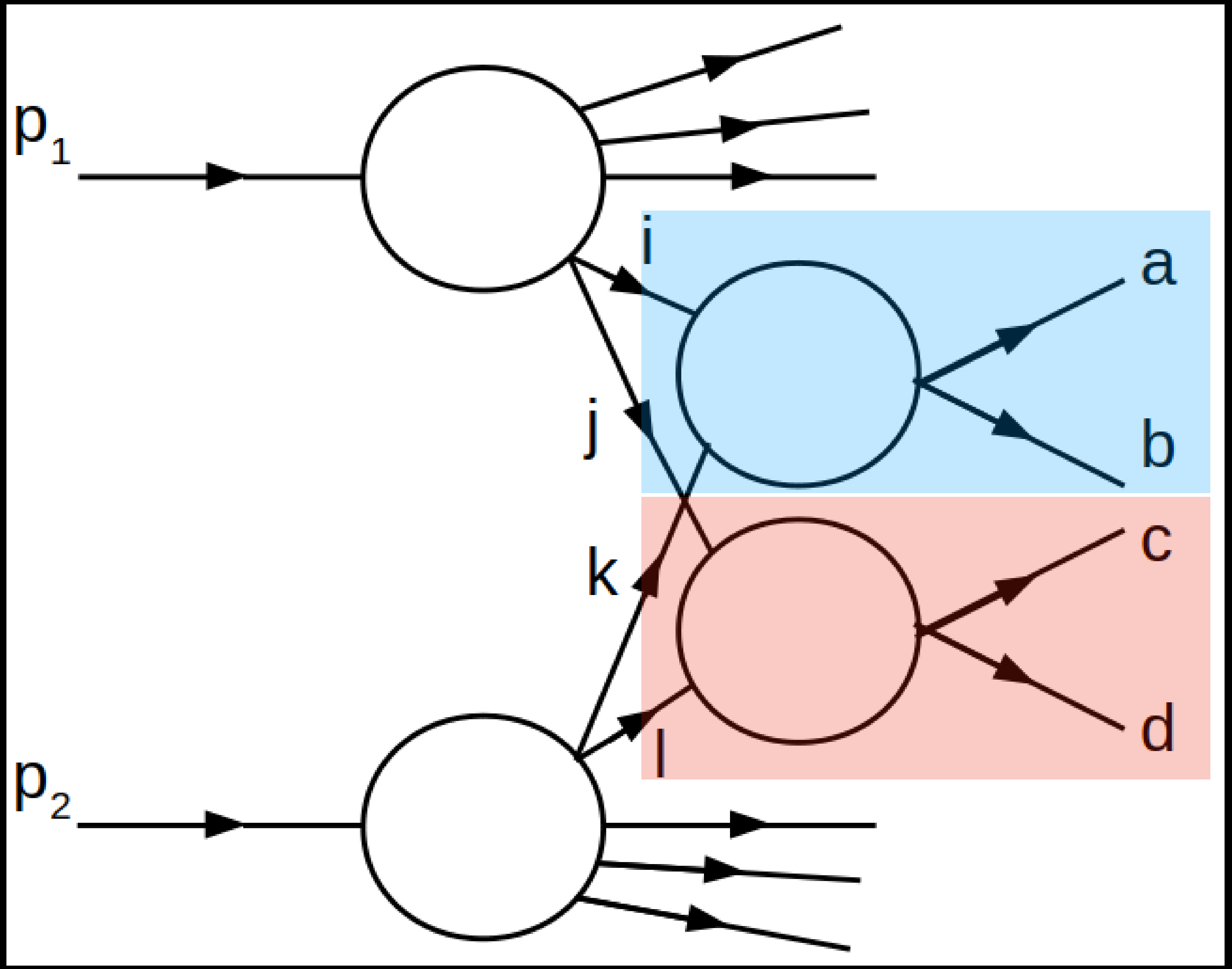
- Top-quark candidates constructed by a RIVET routine (kinematic requirements imposed and “dressed” leptons used)
- Largest deviation (from default CP5) found for CP5-CR2 erdon (0.32 GeV) ~ similar to what is observed in TOP-17-007

https://arxiv.org/abs/2109.13822

Double Parton Scattering of 4 jet events



Single parton scattering (SPS)



Double parton scattering (DPS)

- In the case of SPS, one hard scattering produces the jets *a* through *d*
- Two jet pairs are created independently in a DPS event → different kinematic correlations than an SPS event
- Can allow for studies of p_T ordered and angular ordered showers

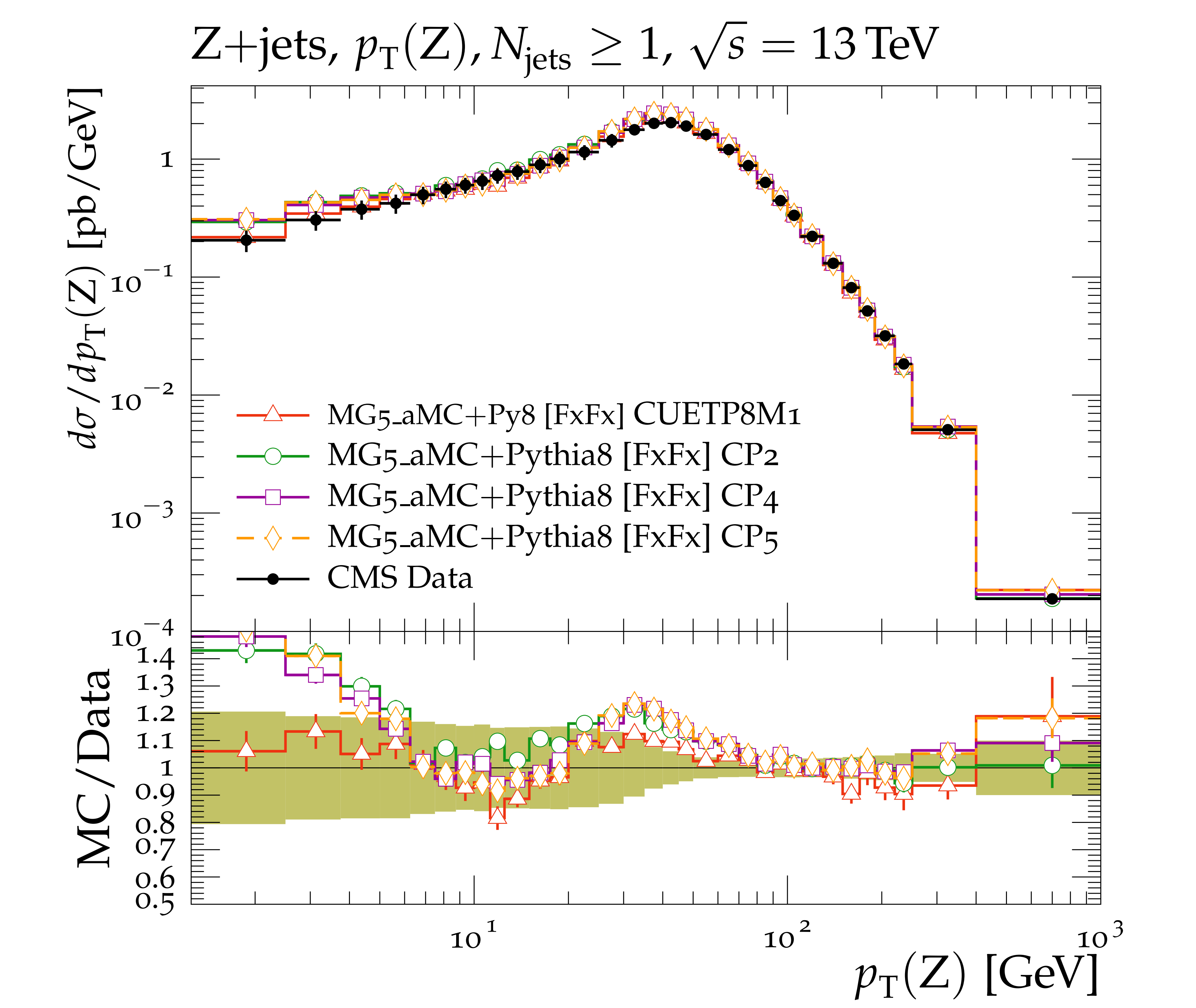
Current tuning studies

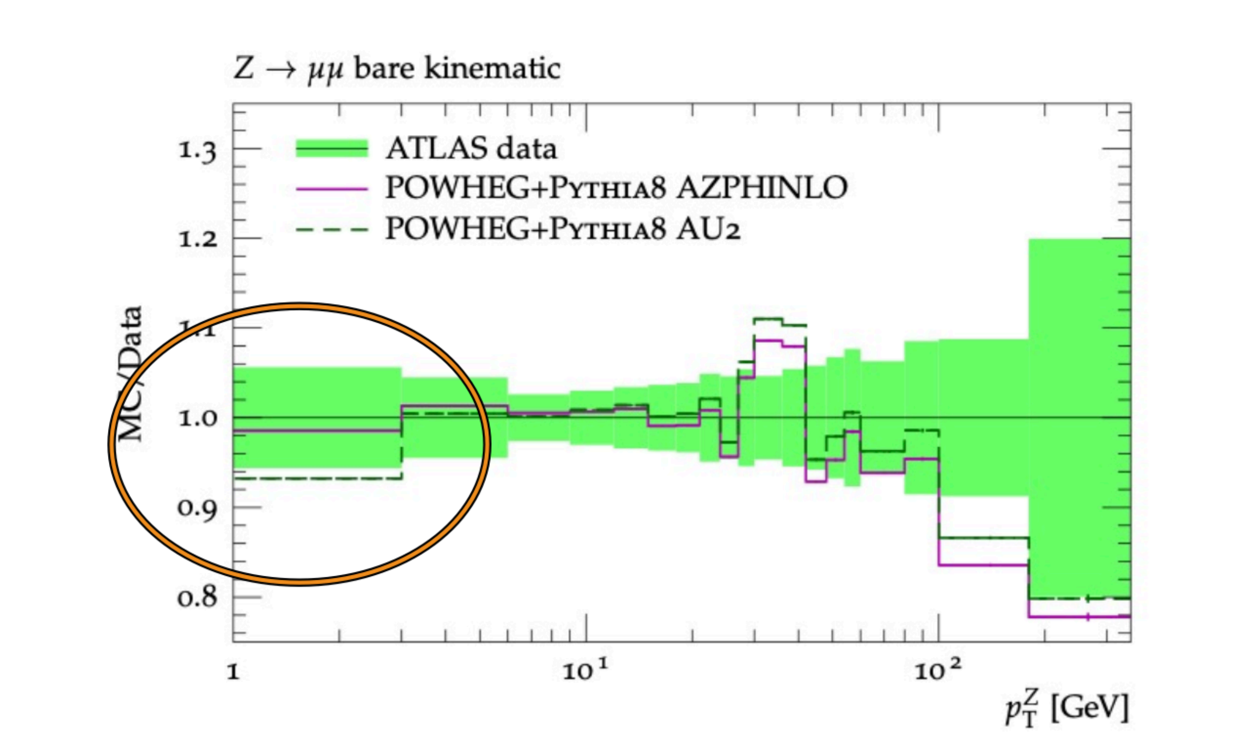
- We perform this task as discrepancies seen in data might not be public or can be easily shared for tuning purposes
- Often Generator authors do not have computing resources for performing global tunes
- Urgency in the experiments to make the best simulations available

Current tuning studies — work in progress

[GEN-17-002](#)

- Challenges
  - Underlying event description needs to remain compatible with CP5 description
  - Choose variables to carefully tune only a certain region of phase space where CP5 does not provide the best description of data
- Variable of choice and target region:
  - Intrinsic k_T: intrinsic transverse momentum of the initial state partons (100 MeV/s)
  - Tune intrinsic k_T at low mass DY
- Exercise caution:
  - Make sure no other additional change is introduced





[ATL-PHYS-PUB-2013-017](#)

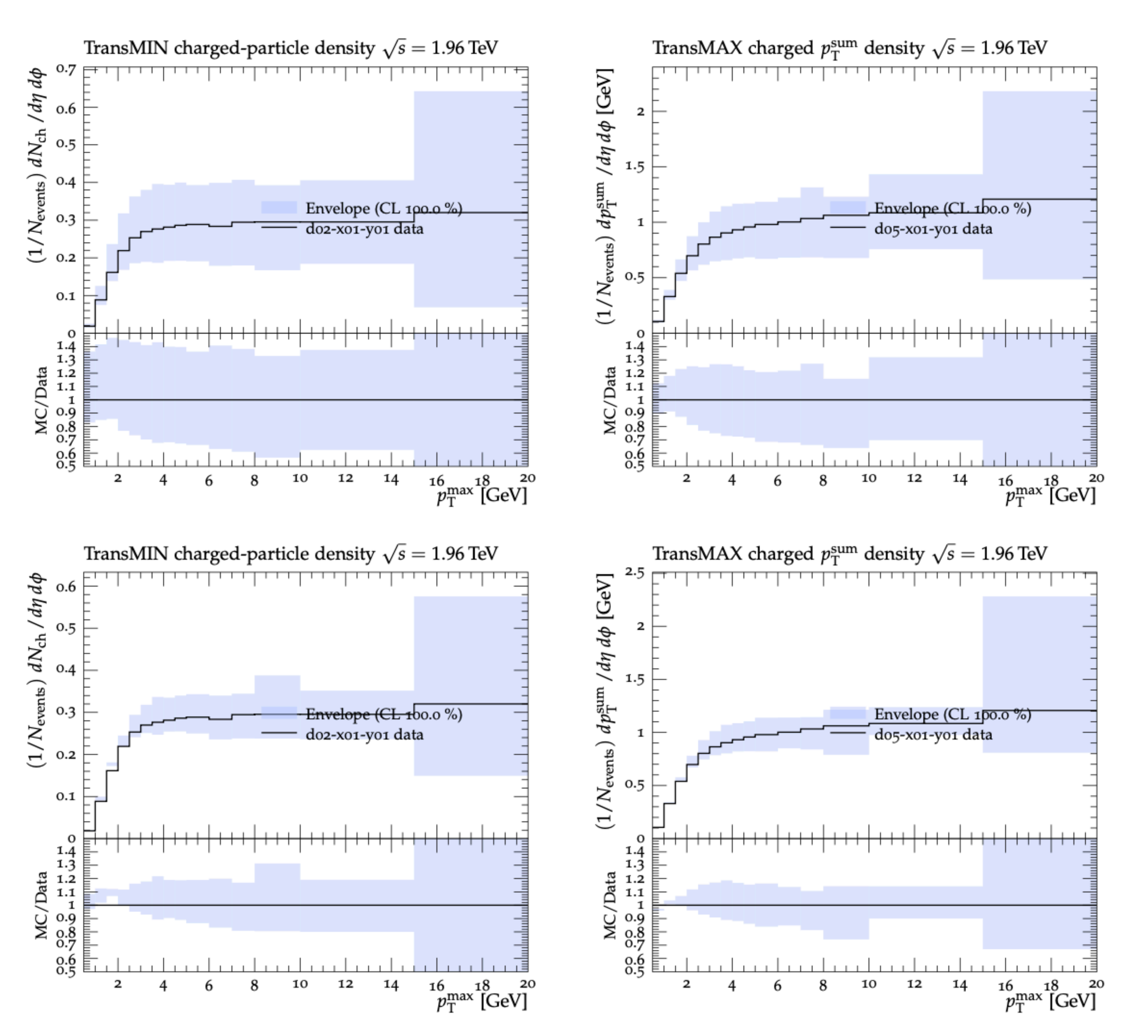
Current tuning studies — work in progress

- Immediate goal:
  - Tuning studies underway with an aim to make minor modifications to CP5
    - Target parameter: intrinsic kT (set in the Monash tune)
      - Intrinsic k_T → partons have small *transverse* momenta in the hadron (100 MeV/s)
      - Introduced as a non-perturbative parameter
      - Effect can be seen in low p_T regime in Drell-Yan processes
    - Explore any benefits associated with moving to NNPDF4.0
- Medium-term goal:
  - Study dependence of intrinsic k_T on ISR/FSR parameters (use the same value of α_s for ISR and FSR)
- Compare results with Professor and MCNNTunes
- First results presented at GEN meeting on May 2nd: focus on reproducing CP5

Current tuning studies — work in progress

- Exercise caution:
  - Make sure no other additional change is introduced
  - Check predictions for the underlying event observables
- CP5 + intrinsic k_T parameter introduces minimal changes to the envelope
- Similar effort ongoing in the LHC Jets and EW boson subgroup

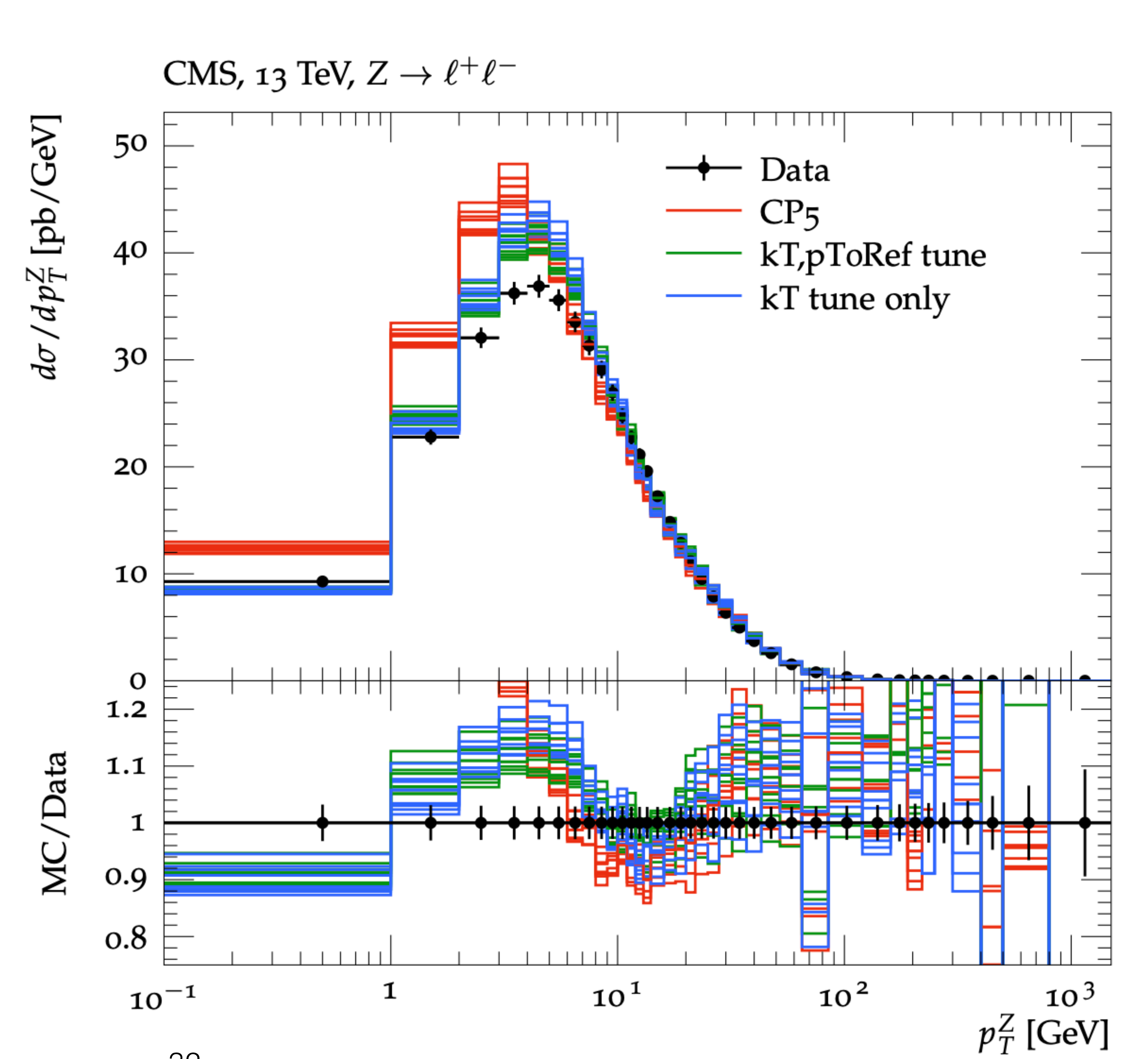
2 parameters



Only intrinsic k_T

Current tuning studies — work in progress

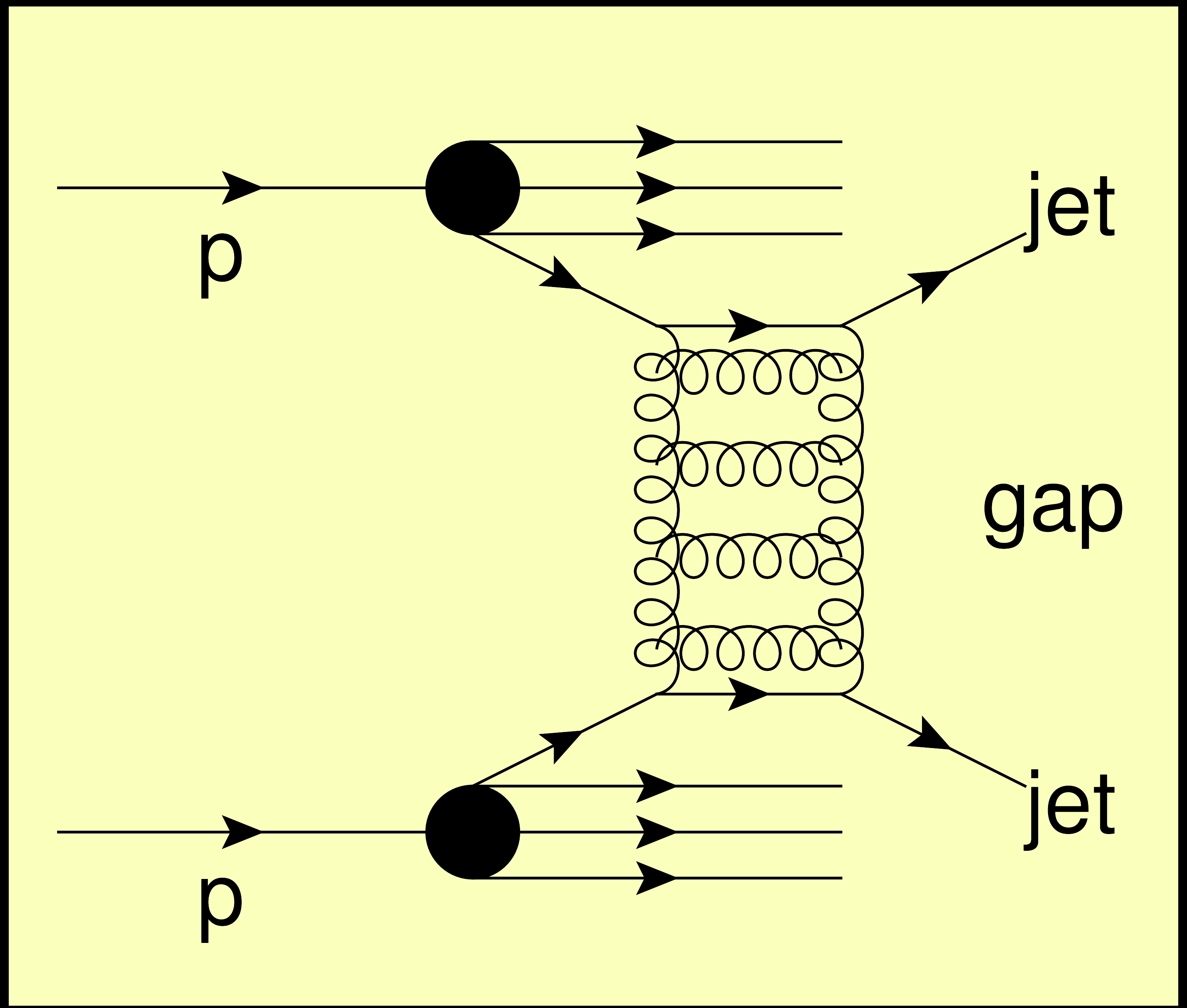
- Only tuning intrinsic k_T preserves underlying event description, while providing better description of low p_T^Z
- Finalizing the tune for dedicated sample production in a few months
- Aiming for pre-approval in the coming weeks
- Stay tuned!

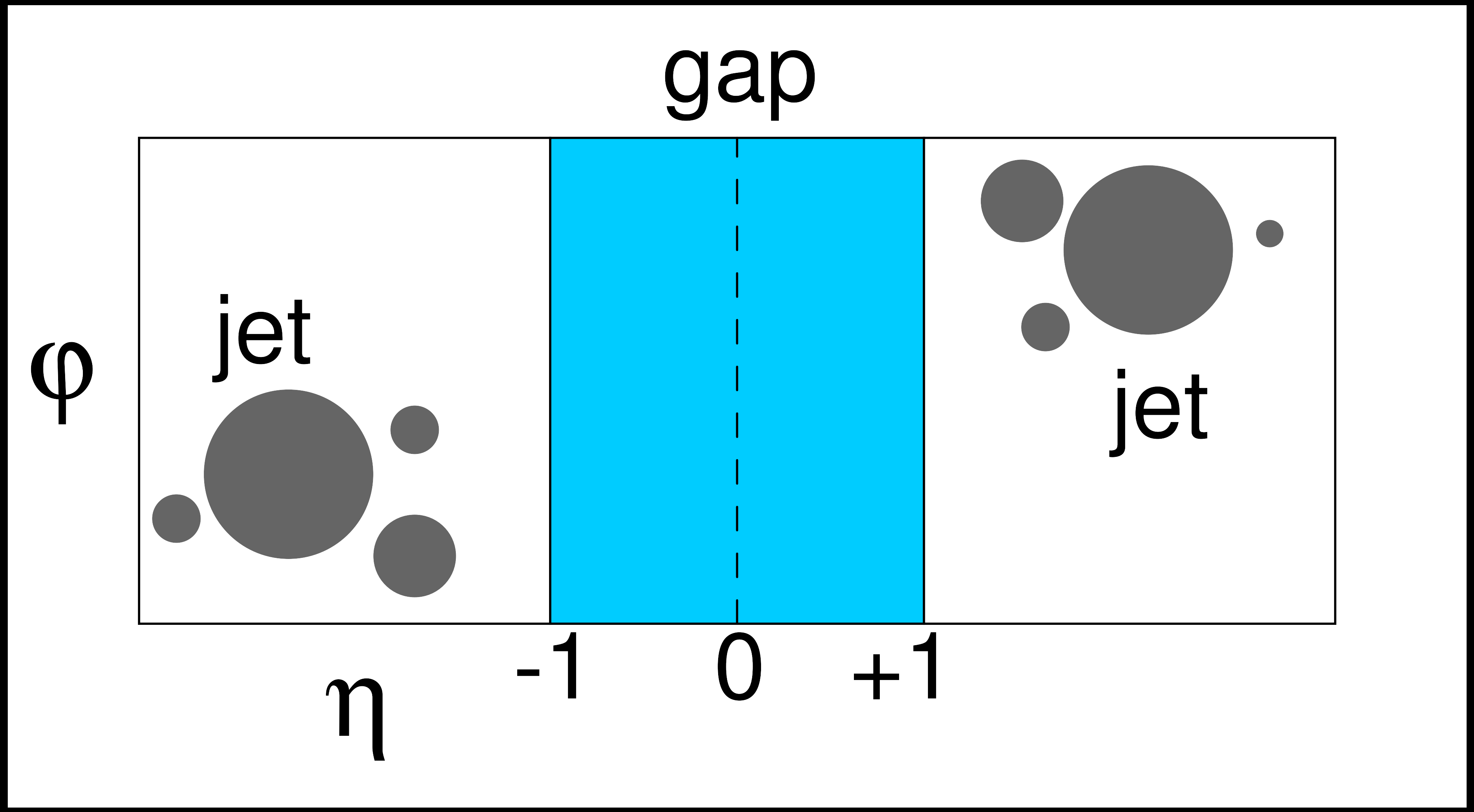


A novel analysis
Parton showers beyond DGLAP

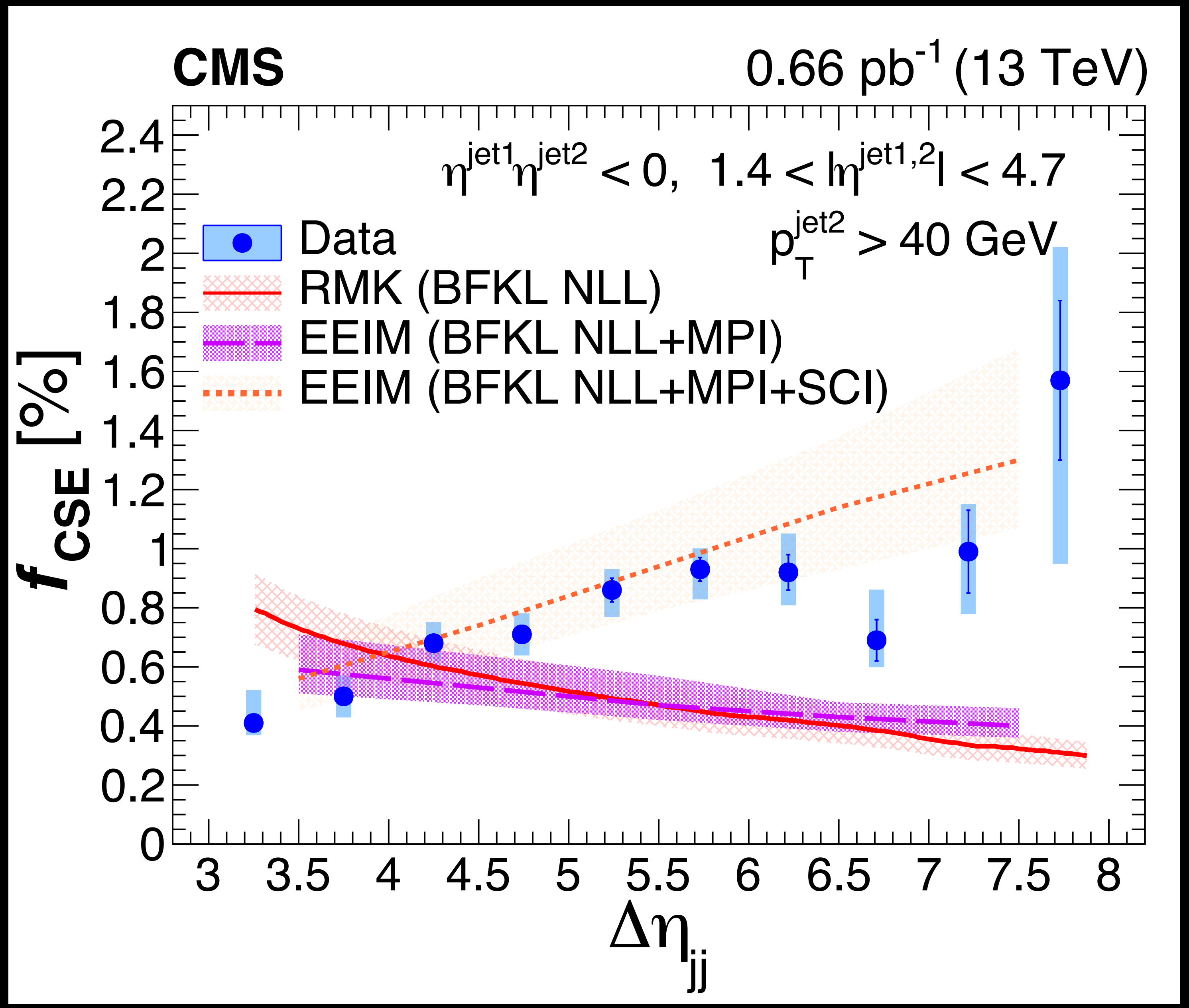
Phys. Rev. D 104 (2021) 032009

Parton Shower description beyond DGLAP





- 2 → 2 scattering: for values of Λ_{QCD} ($\hat{s} \gg \Lambda^2_{QCD}$), where QCD is no longer strongly coupled, fixed order perturbation theory no longer valid
- Balitsky–Fadin–Kuraev–Lipatov (BFKL) evolution equation resums logarithmic terms to all orders in α_S → NLL accuracy
- In dijet production, BKFL dynamics expected to manifest when two jets are separated by a large rapidity interval



f_{CSE}: fraction of events produced via color singlet exchange

The CMS Generator Group

CMS Generator Group: who are we?

- L2 conveners: Gurpreet Singh Chahal, Meng Lu, Saptaparna Bhattacharya
- Summary of our activities: https://twiki.cern.ch/twiki/ bin/viewauth/CMS/GeneratorMain
- Meetings held every Monday at 14:00 CERN

Generator Integration **Mikhail Kirsanov** **Siew Yan Hoh**	**Matrix Element and Future Generators** **Mattia Lizzo**
Generator Validation **Jie Xiao** **Sanghyun Ko**	**Physics Comparisons and Generator Tunes** **Armando B. Martinez** **Arthur Moraes**

- **Generator Contacts:**
  - MG5_aMC@NLO : Sihyun Jeon, Alexander Grohsjean
  - Powheg: Roberto Covarelli, Alexander Grohsjean
  - Sherpa: Gurpreet Singh Chahal
  - Pythia: Steve Mrenna
  - Herwig7: Dominic Stafford, Kostas Theofilatos
  - MCFM/ MATRIX: Bingran Wang, vacant starting September 2021
  - Pyquen: Andrey Belyaev
- **General purpose tool contacts:**
  - **Rivet framework** : Tahys Janssen, Markus Seidel
  - Professor: Sercan Vieri
  - XSDB: Sen Deng, Qilong Guo, Yuzhe Zhao
- **Contacts to other groups:**
  - ML: Andre Sznajder
- **Technical contacts:**
  - NanoAOD /NanoGEN: Dylan Teague
  - cmsconnect: Hyon San Seo
  - lxplus@CERN, HTCondor: Rajat Gupta
  - request checking script: Efe Yazgan
- **Generator Operators**: Duncan Leggat, Youn Roh
- **Release validation**: Dongwoon Kim
- **HSF contact**: Efe Yazgan

Run III readiness — outline

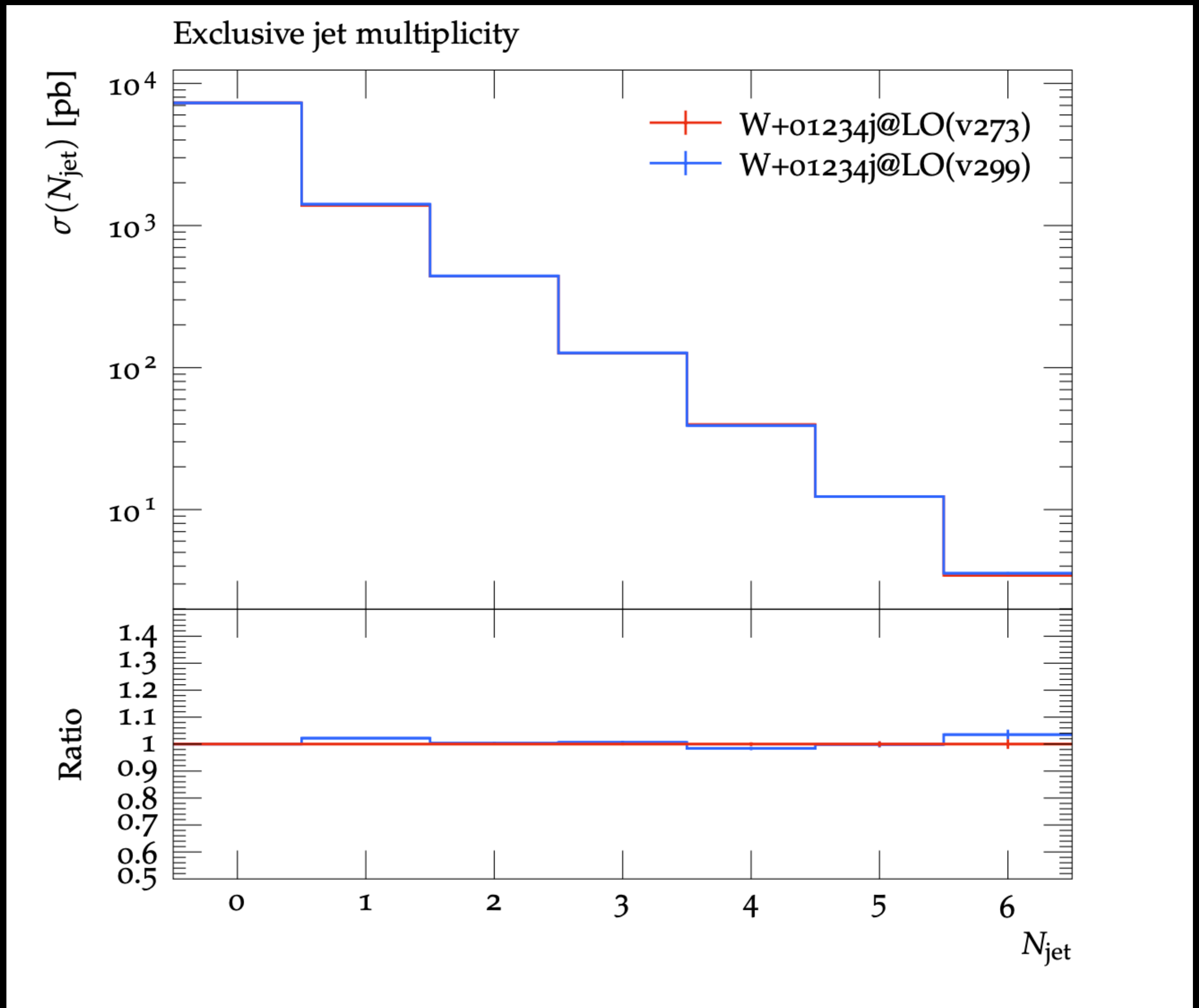
- Major areas of focus (right now):
  - Core generator development work
    - Exploring new version of generators
      - It is imperative to upgrade to a new version of Madgraph_aMCatNLO as the Ultra Legacy version is no longer supported by the authors
      - Identifying bugs/inconsistencies with previous settings
    - Monitoring and/or updating CMSSW for major updates (e.g. HEPMC3)
    - Facilitating the automated generation of common backgrounds

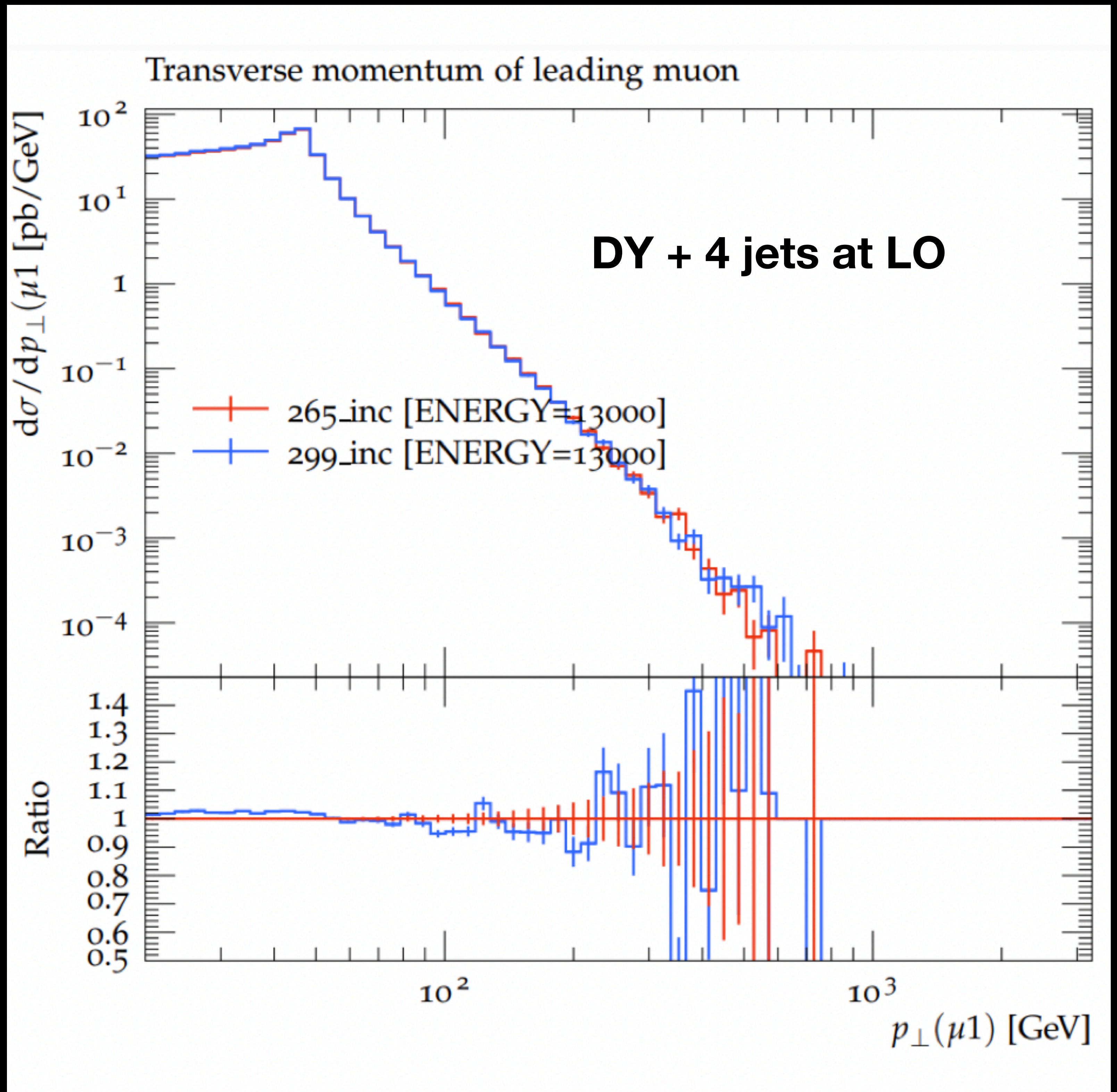
Core generator development - I

- Current plan is to update to Madgraph5_aMC@NLO 2.9X (UL branch: 2.6.5)
  - Version 2.9X includes helicity recycling for optimizing production of VBF processes
  - Expect speedup for LO processes
  - Choice of version dictated by:
    - several issues with 3.3X (received support from authors but reporting and fixing bugs is likely to make progress slow)
    - The generator validation Level-3 conveners (Sanghyun Ko and Jie Xiao) have validated 2.9X

Generator development for Run III

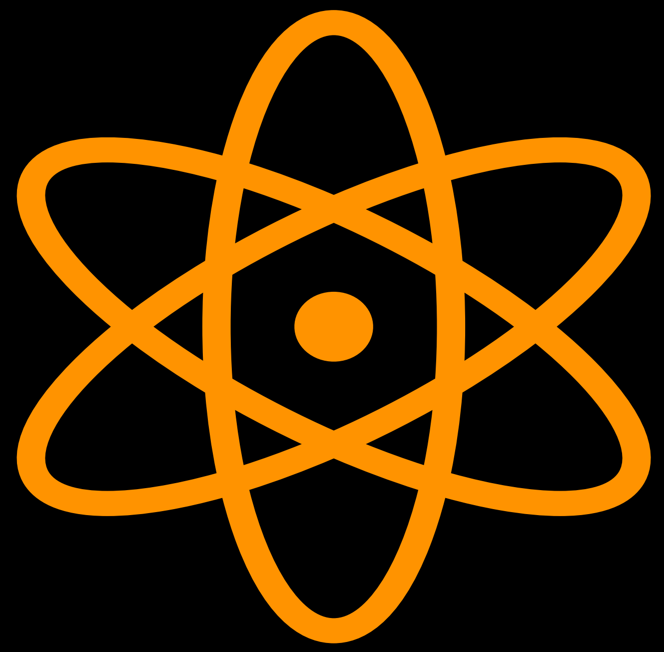
- Validation of 🍞 and 🧈 processes being run with MG5_299
  - binned and inclusive: W+Jets, Drell-Yan and $t\bar{t}$
  - Good agreement for V+jets





Run III and beyond

- Major areas of focus:
  - Tuning studies
    - Goal: Updating CP5 for cases where known issues exist
    - Twiki in place: https://twiki.cern.ch/twiki/bin/viewauth/CMS/Rivet_Professor
    - Meetings held every week (Fridays at 14:00 CERN)
  - Negative weight reduction
    - Deleterious effect on sample generation as more samples are being generated at NLO
    - Possible areas of improvement include: MC@NLO-Δ scheme and positive resampling
    - Detailed discussion at O&C week on multi-threading



Physics based improvement

CPU Efficiency

LO vs. NLO

- While next-to-leading order (NLO) often provides best description of data
  - Samples suffer from large fraction of negative weights
  - For tails of distributions, often easier to generate samples with additional partons at leading order (LO)
- Exercise caution and decide what is needed for YOUR analysis

Rate of negative events

$pp \to e^+e^-$	6.9%	(1.3)
$pp \to e^+\nu_e$	7.2%	(1.4)
$pp \to H$	10.4%	(1.6)
$pp \to Hb\bar{b}$	40.3%	(27)
$pp \to W^+j$	21.7%	(3.1)
$pp \to W^+t\bar{t}$	16.2%	(2.2)
$pp \to t\bar{t}$	23.0%	(3.4)

Cost In sample size

$$c(f) = \frac{1}{(1-2f)^2}$$

From Olivier Mattelaer's talk

Automating background production

- Automation of background production
  - Discussed in several GEN meetings
  - Current implementation includes a full Madgraph@LO production
  - First successful prototype for DY+4jets successfully produced
  - Extended it to other generators (e.g. POWHEG)
  - Synergistic with existing efforts in GEN regarding validation of the new Madgraph_aMCatNLO

Cross sections at 13.6 TeV

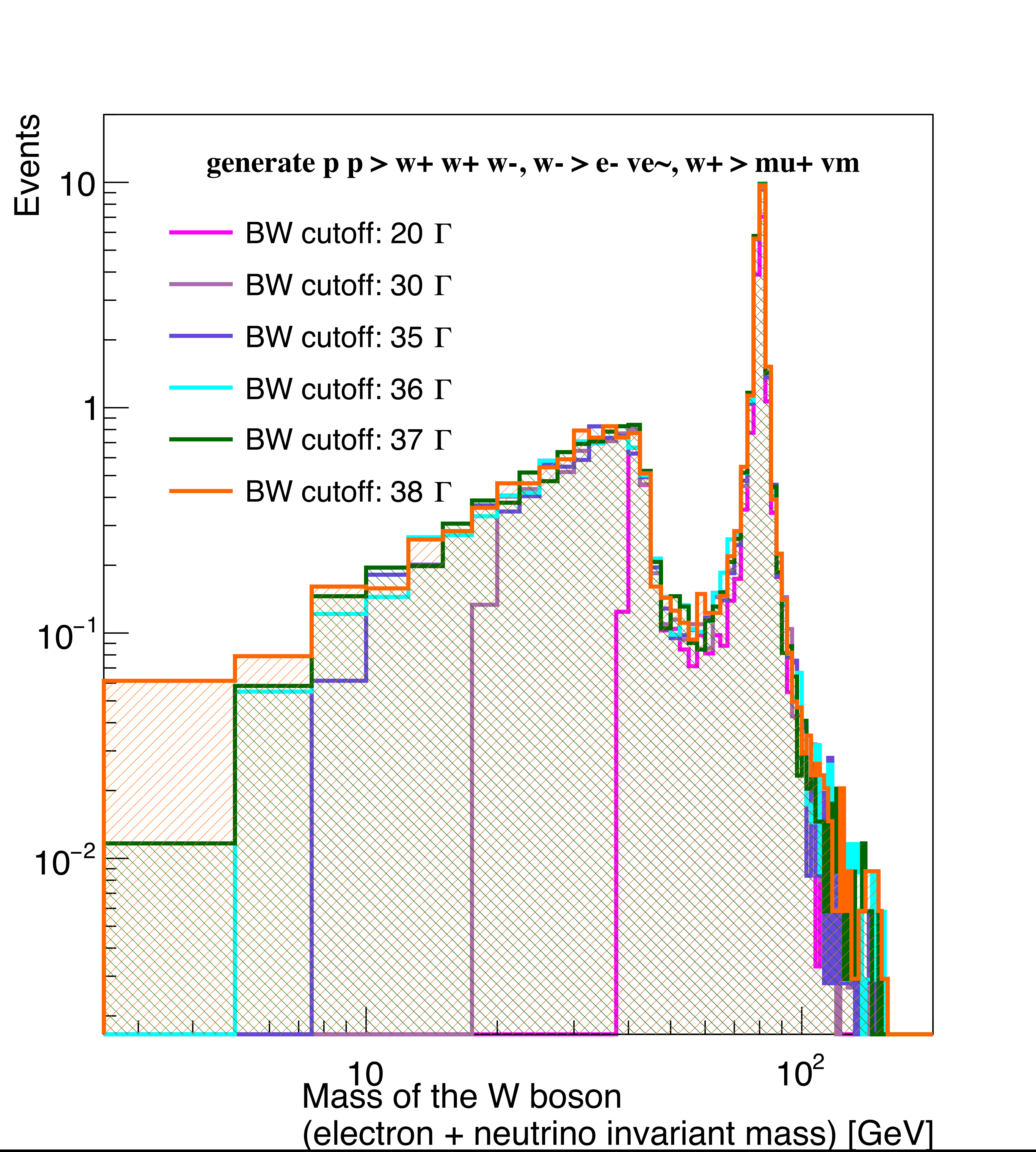
- Higher order cross sections available for TOP and HIG processes available through the various LHC-wide working groups
- Similar effort on-going in conjunction with SMP for multiboson processes
  - Several discussions with ATLAS colleagues
  - MATRIX used for computation (NNLO QCD and NLO EW)
- Cross section database automatically updated with sample cross sections (driven by the accuracy of the sample)
  - For higher order cross sections, need input from PAGs and entered into the database
  - We have provided detailed instructions to run FEWZ on lxplus and recent update given at GEN meeting

Baseline Recommendations

Generators	• Madgraph5_aMC@NLO: 2.9.9 • Pythia: 8.306 (other generators expected to remain the same as the Ultra Legacy campaign)
PDF	NNPDF3.1 (unchanged from Ultra Legacy) Alternate sets will mostly contain NNLO PDFs
Tune	CPX family (CP5 is the most commonly used)
HEPMC version	HEPMC2 (unchanged from Ultra Legacy)
Total number of events to be generated	Assuming 40 fb^{-1} of data will be collected this year (only an assumption!)

Common pitfalls when committing cards to genproductions repository

- Breit-Wigner cut off
- 4 or 5 flavor scheme?
- Extra models?
- Applying cuts?
- Are you setting widths correctly?
- Using madspin?



Common pitfalls when committing cards to genproductions repository

- Breit-Wigner cut off:
  - Do not set it to any arbitrary number, Madgraph authors recommend 15 Γ
  - Remember that Madgraph still operates within the Narrow Width Approximation
- 4 or 5 flavor scheme?
  - 4 flavor scheme, b-quark is massive
- Extra models?
  - Please ping us when they need to be uploaded
- Applying cuts?
  - Please don't apply too restrictive cuts
- Are you setting widths correctly?
  - Are you setting them by hand?
- Using madspin?
  - Please use consistent value of Briet Wigner cut off in the madspin card

- Generator L2s check settings twice, once during card review and then during sample approval
- Please commit your cards to the repository as early in the analysis timeline as possible
- We are just two people that are responsible for card review and sample approval

The Future

What’s next for us?

- Lots of useful discussions at Snowmass on event generators helmed by Gherardo Vita, Joshua Issacson, Tobias Neumann, Andreas Kronfeld
- Cross frontier discussions between theory and energy frontier
- Featured panel discussions on nightmare scenarios, novel observables and new approaches
  - Discussion on Machine Learning for event generation by Tilman Plehn

Comprehensive White Paper submitted to the Snowmass process

High Energy Physics – Phenomenology

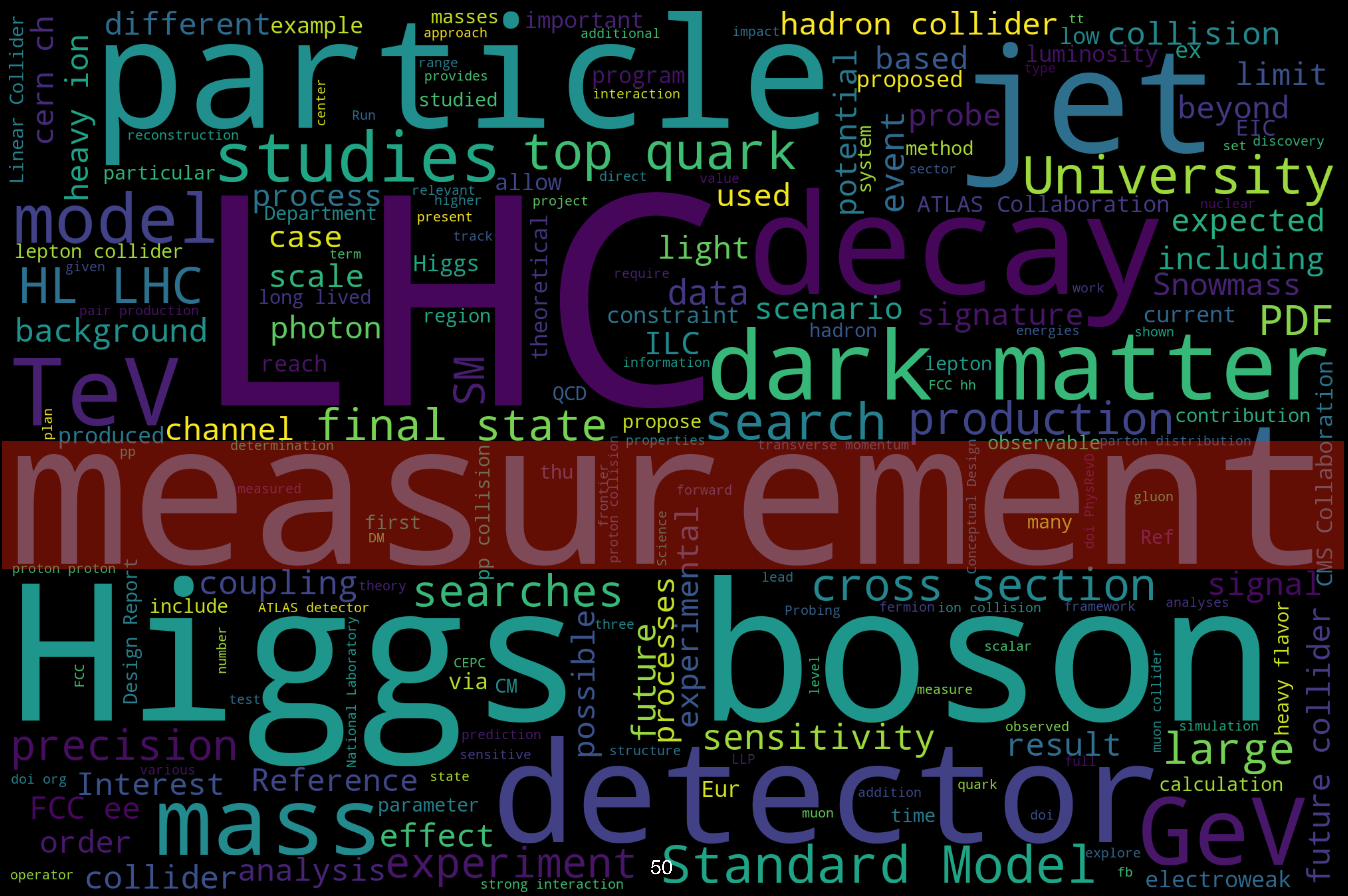
[Submitted on 21 Mar 2022]

Event Generators for High-Energy Physics Experiments

J. M. Campbell, M. Diefenthaler, T. J. Hobbs, S. Höche, J. Isaacson, F. Kling, S. Mrenna, J. Reuter, S. Alioli, J. R. Andersen, C. Andreopoulos, A. M. Ankowski, E. C. Aschenauer, A. Ashkenazi, M. D. Baker, J. L. Barrow, M. van Beekveld, G. Bewick, S. Bhattacharya, C. Bierlich, E. Bothmann, P. Bredt, A. Broggio, A. Buckley, A. Butter, J. M. Butterworth, E. P. Byrne, S. Chakraborty, X. Chen, M. Chiesa, J. T. Childers, J. Cruz-Martinez, J. Currie, N. Darvishi, M. Dasgupta, A. Denner, F. A. Dreyer, S. Dytman, B. K. El-Menoufi, T. Engel, S. Ferrario Ravasio, L. Flower, J. R. Forshaw, R. Frederix, A. Friedland, S. Frixione, H. Gallagher, K. Gallmeister, S. Gardiner, R. Gauld, A. Gavardi, T. Gehrmann, A. Gehrmann-De Ridder, L. Gellersen, S. Gieseke, F. Giuli, E. W. N. Glover, M. Grazzini, A. Grohsjean, C. Gütschow, K. Hamilton, R. Hatcher, I. Helenius, T. Han, O. Hen, V. Hirschi, M. Höfer, J. Holguin, A. Huss, P. Ilten, S. Jadach, A. Jentsch, W. Ju, S. Kallweit, A. Karlberg, T. Katori, W. Kilian, M. M. Kirchgaeßer, M. Knobbe, C. Krause, F. Krauss, J.-N. Lang, G. Lee, S. W. Li, M. A. Lim, J. M. Lindert, D. Lombardi, L. Lönnblad, M. Löschner, N. Lurkin, Y. Ma, P. Machado, A. Maier, I. Majer, M. Marcoli, G. Marinelli, M. R. Masouminia, O. Mattelaer, J. Mazzitelli, J. McFayden et al. (87 additional authors not shown)

We provide an overview of the status of Monte-Carlo event generators for high-energy particle physics. Guided by the experimental needs and requirements, we highlight areas of active development, and opportunities for future improvements. Particular emphasis is given to physics models and algorithms that are employed across a variety of experiments. These common themes in event generator development lead to a more comprehensive understanding of physics at the highest energies and intensities, and allow models to be tested against a wealth of data that have been accumulated over the past decades. A cohesive approach to event generator development will allow these models to be further improved and systematic uncertainties to be reduced, directly contributing to future experimental success. Event generators are part of a much larger ecosystem of computational tools. They typically involve a number of unknown model parameters that must be tuned to experimental data, while maintaining the integrity of the underlying physics models. Making both these data, and the analyses with which they have been obtained accessible to future users is an essential aspect of open science and data preservation. It ensures the consistency of physics models across a variety of experiments.

Word cloud of letter of interest submissions in the energy frontier



Where we are and where we are headed

- The LHC is a precision machine
- Huge interest and push towards measurements (as opposed to searches)
- Wide plethora of modeling studies performed in CMS
- High accuracy matrix element descriptors in use
- Tuning studies on-going to achieve best description of data in the soft, non-perturbative regime
- Successful synergy with event generator community (through the McNET program among other avenues)
- My message today: Join the activities of the Generator group! We give out EPR credits!

Additional Material

Vast array of modeling studies (indicative)

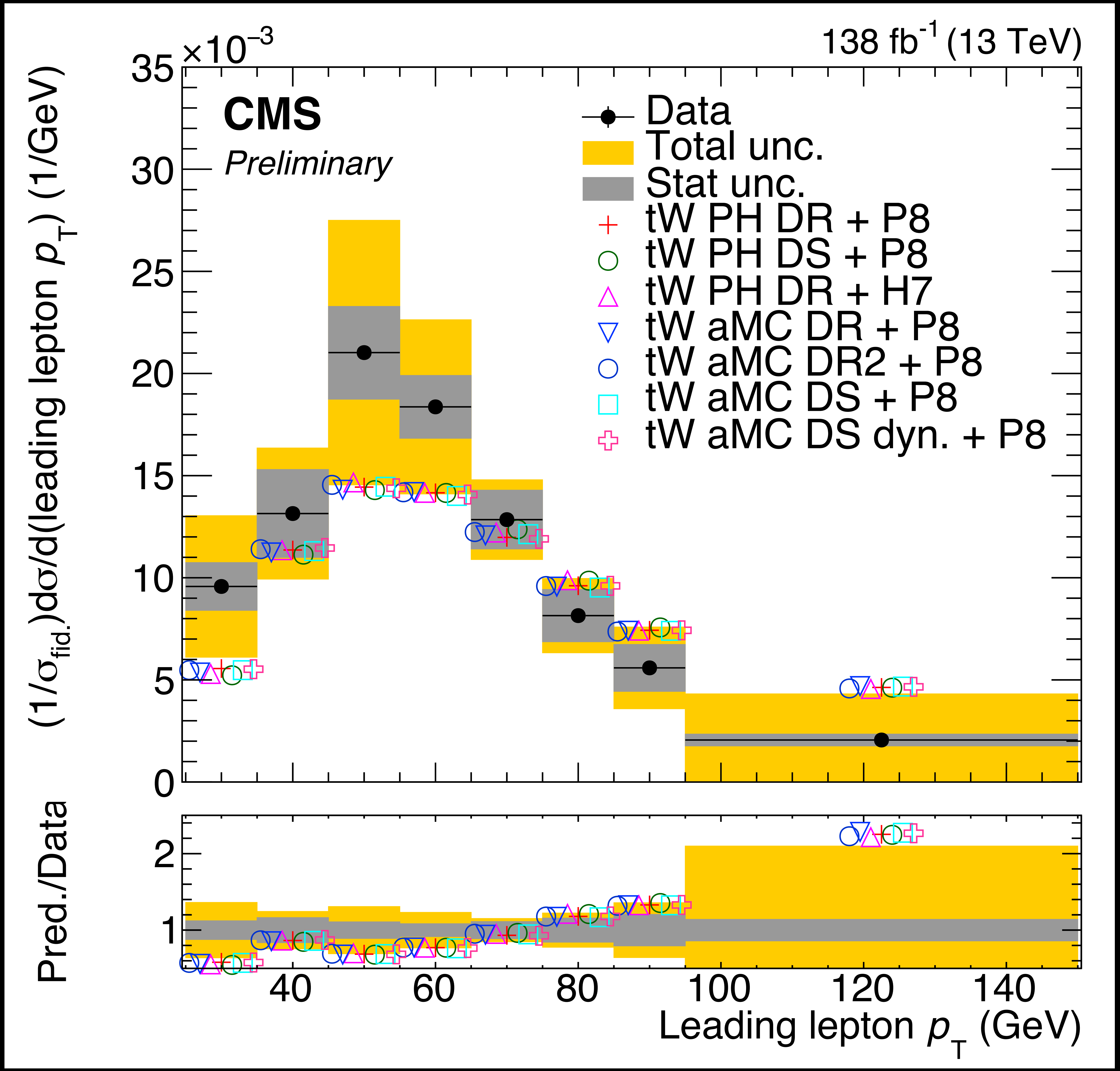
Measurement of W+c production cross section (SMP-21-005)	Includes comparisons with of differential distributions with MCFM
Measurement of Z+b jets cross section (SMP-20-015)	Includes comparisons of Z+heavy flavor with various matrix element generators
Measurement of double parton scattering in four jets and probing distributions sensitive to double parton scattering in Z+jets events (SMP-20-007 and SMP-20-009)	Includes comparison with HERWIG 7 and VINCIA
Measurement of mass dependence of the transverse momentum of Drell Yan lepton pairs (SMP-20-003)	Includes comparisons of Z with various matrix element generators (including comparisons with GENEVA)
First evidence of off-shell production of Higgs bosons (HIG-21-013)	Requires N^3LO description of $gg \to H \to ZZ$

Vast array of modeling studies (indicative)

Measurement of inclusive and differential cross sections for single top quark production in association with a W boson (TOP-21-010)	Features comparisons with POWHEG+Pythia, HERWIG with various diagram removal schemes
Measurement of differential cross sections for the production of top quark pairs and of additional jets (TOP-20-006)	Features comparisons with POWHEG+Pythia and HERWIG
Measurement of the top quark pole mass using pair produced top quarks (TOP-21-008)	The top quark pole mass is extracted using the theoretical predictions at next-to-leading order with the ABMP16NLO PDF.
Measurement of the inclusive and differential $t\bar{t}\gamma$ cross section (TOP-21-004)	Features comparisons with aMC@NLO+Pythia and HERWIG
Measurement of the shape of the b quark fragmentation function using charmed mesons produced inside b jets from (TOP-18-012)	A determination of the shape parameter of the Lund-Bowler fragmentation function for b quarks is presented

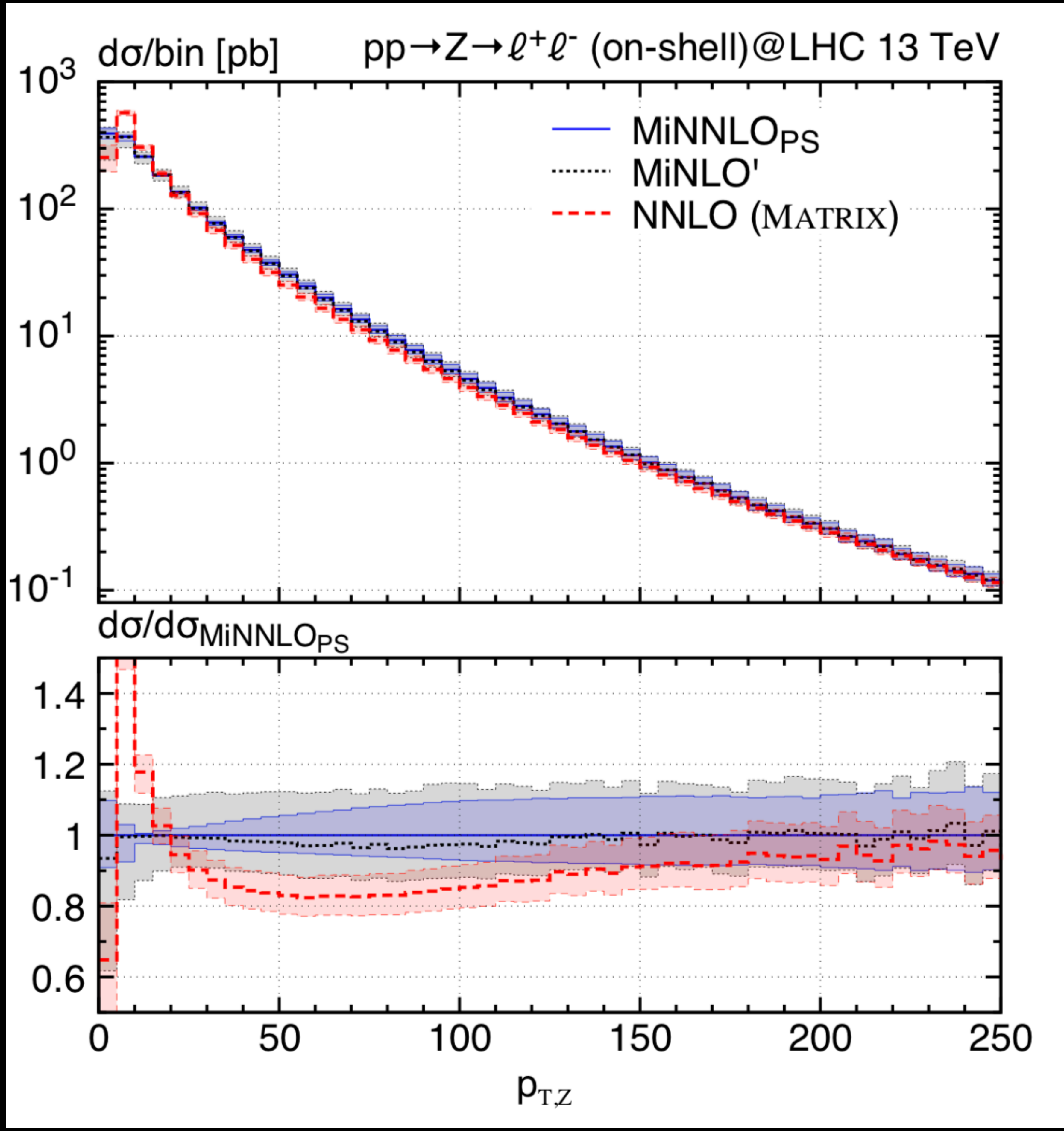
Single top in tW channel

- Single top samples simulated with POWHEG using
  - diagram removal (DR)
    - NLO diagrams that are doubly resonant removed
  - diagram subtraction (DS)
    - differential cross section modified by a gauge invariant subtraction term
  - Dyn → used to model top-quark resonance



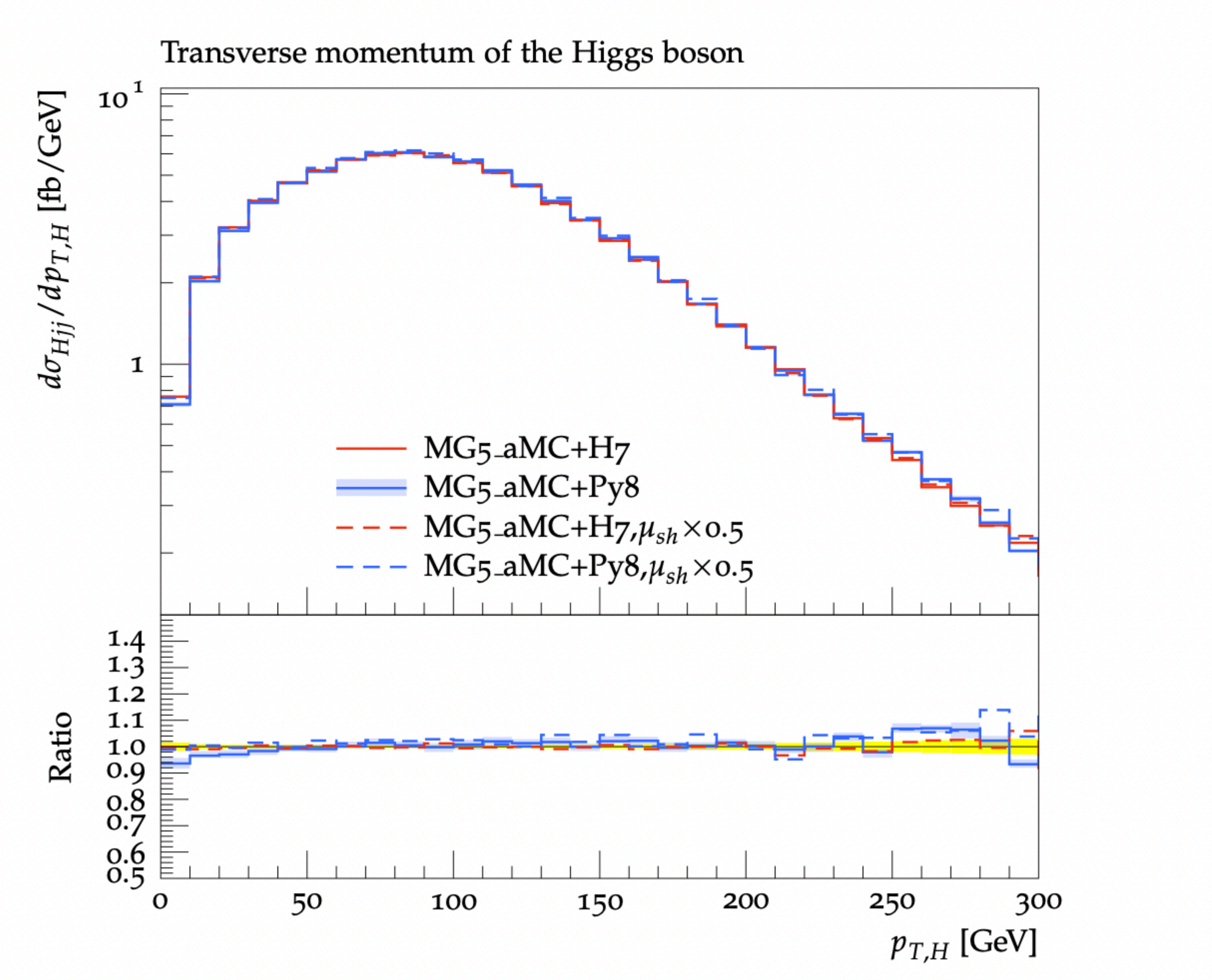
Sample generation at the precision frontier: W-mass

- MiNNLO (arXiv:1908.06987) sample generated at next-to-next-leading-order (NNLO) accuracy
- NNLO generation enabled by merging jet multiplicities with custom scale choice
- Large sample size required to match sample size of data
- Low fraction of negative weights
- Extensive validation performed
- Details in Kenneth Long's talk



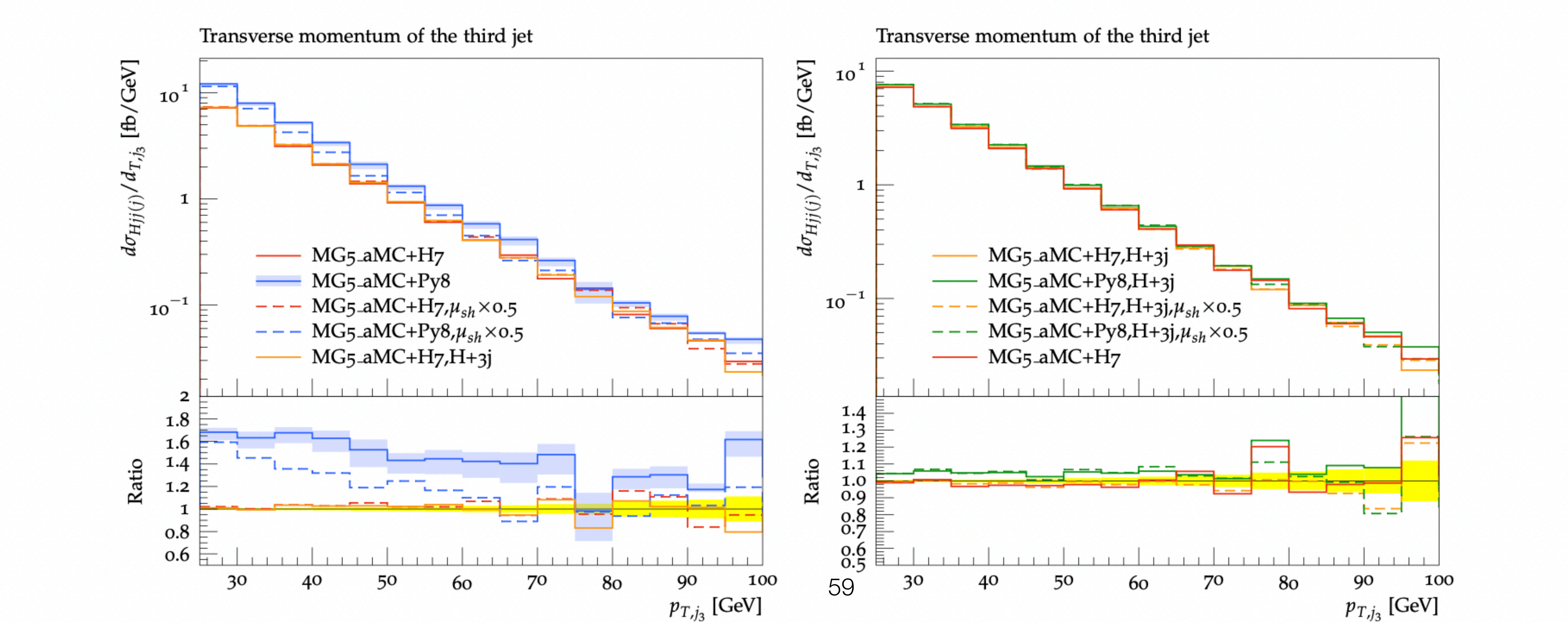
Vector Boson Fusion (VBF) Modeling

- Inclusive (in the number of jets) observable — p_T of the Higgs modeled well by all generators
- Modeling differences in jet bins almost exactly compensate
- Little to no dependence on choice of renormalization, factorization and shower scales
- Global dipole recoil scheme clearly incorrect
- Study of the impact of the 3rd jet presented at GEN meeting on July 26th, 2021



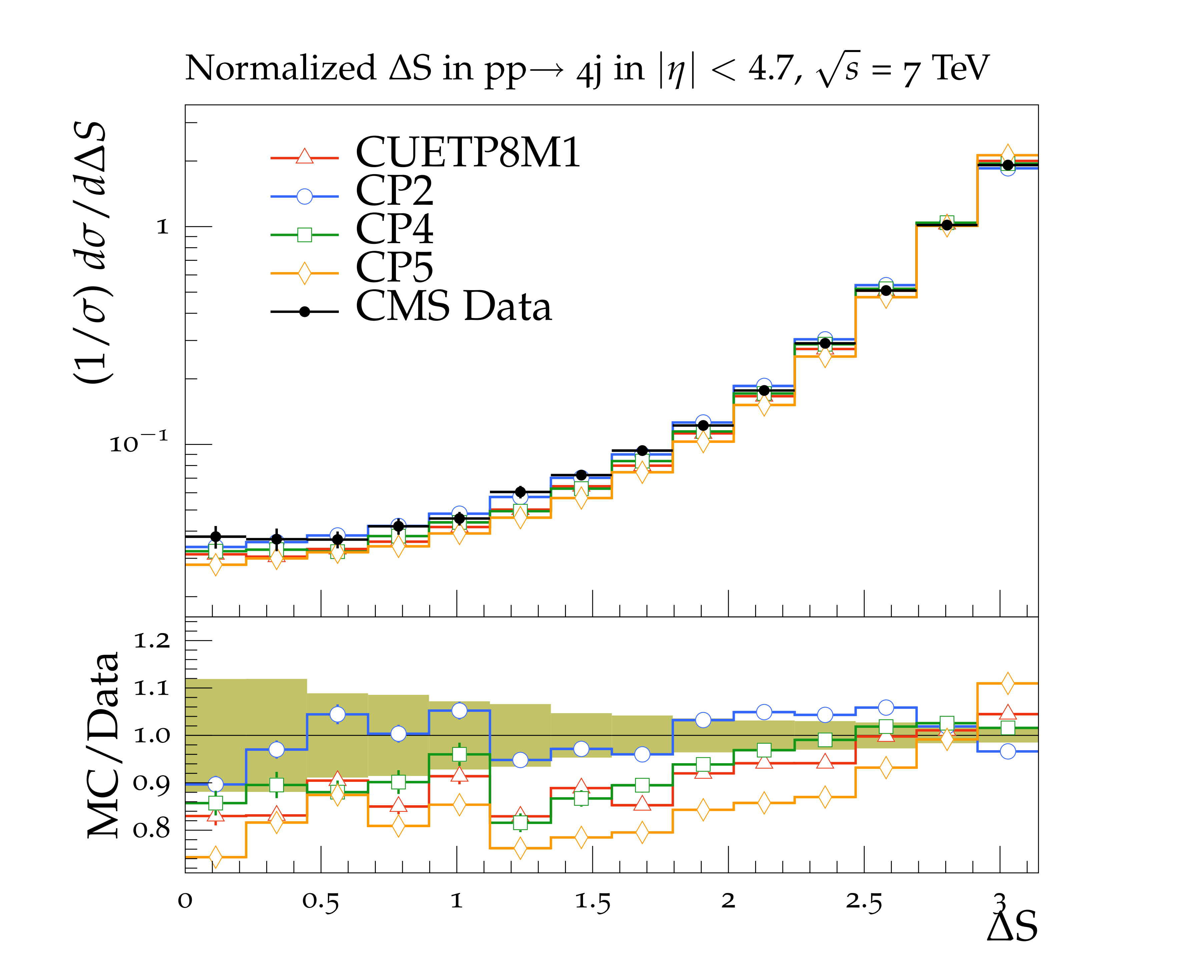
Vector Boson Fusion (VBF) Modeling

- Global dipole recoil scheme clearly incorrect
  - In CMS, samples generated with local dipole recoil scheme
- Study of the impact of the 3rd jet presented at GEN meeting on July 26th, 2021



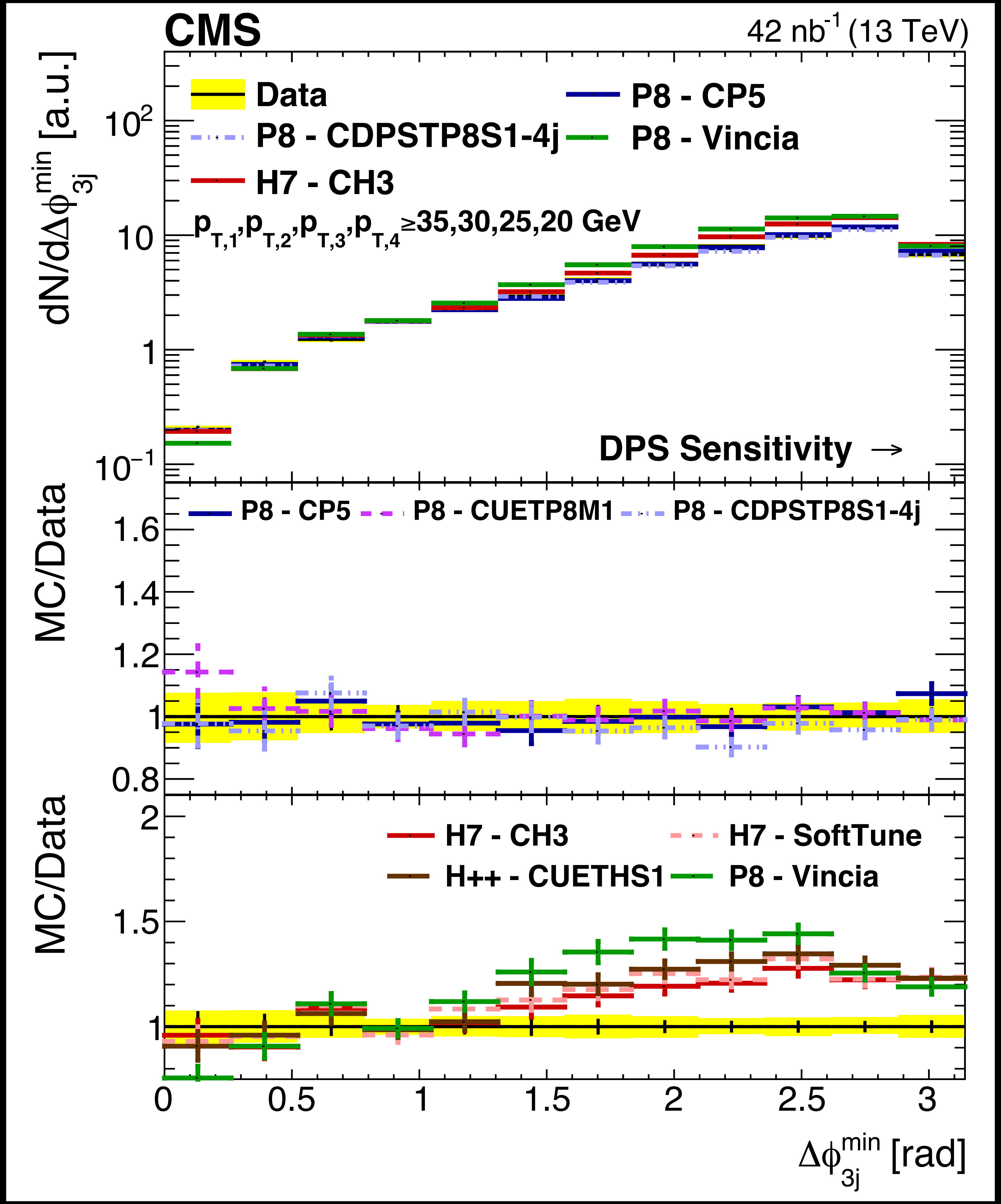
Double Parton Scattering of 4 jet events

Data shows preference for p_T ordered parton shower description (studies performed in GEN-17-001 in agreement)



Azimuthal angular difference between the hard and the soft jet pairs

$$\Delta S = \arccos\left(\frac{(\vec{p}_{T,1}+\vec{p}_{T,2})\cdot(\vec{p}_{T,3}+\vec{p}_{T,4})}{|\vec{p}_{T,1}+\vec{p}_{T,2}||\vec{p}_{T,3}+\vec{p}_{T,4}|}\right)$$



Minimal combined azimuthal angular range of three jets (2 out of 3 jets likely to be back-to-back)

$$\Delta\phi_{3j}^{\min} = \min\left\{ |\phi_i - \phi_j| + |\phi_j - \phi_k| \,\middle|\, i,j,k \in [1,2,3,4], i \neq j \neq k \right\}$$

Core generator development - II

- Emphasis on fixing old issues while the 2.9X validation is ongoing
  - Example: DY@NLO with LHE level filtering
- Sought detailed feedback from PAGs in a series of meetings
  - B2G, EXO, SMP, SUS, TOP
  - Focused on modeling problems in specific regions of phase space
  - Discussed usage of PDFs and tunes
  - Sample size needs for Run III

Tuning studies - overview

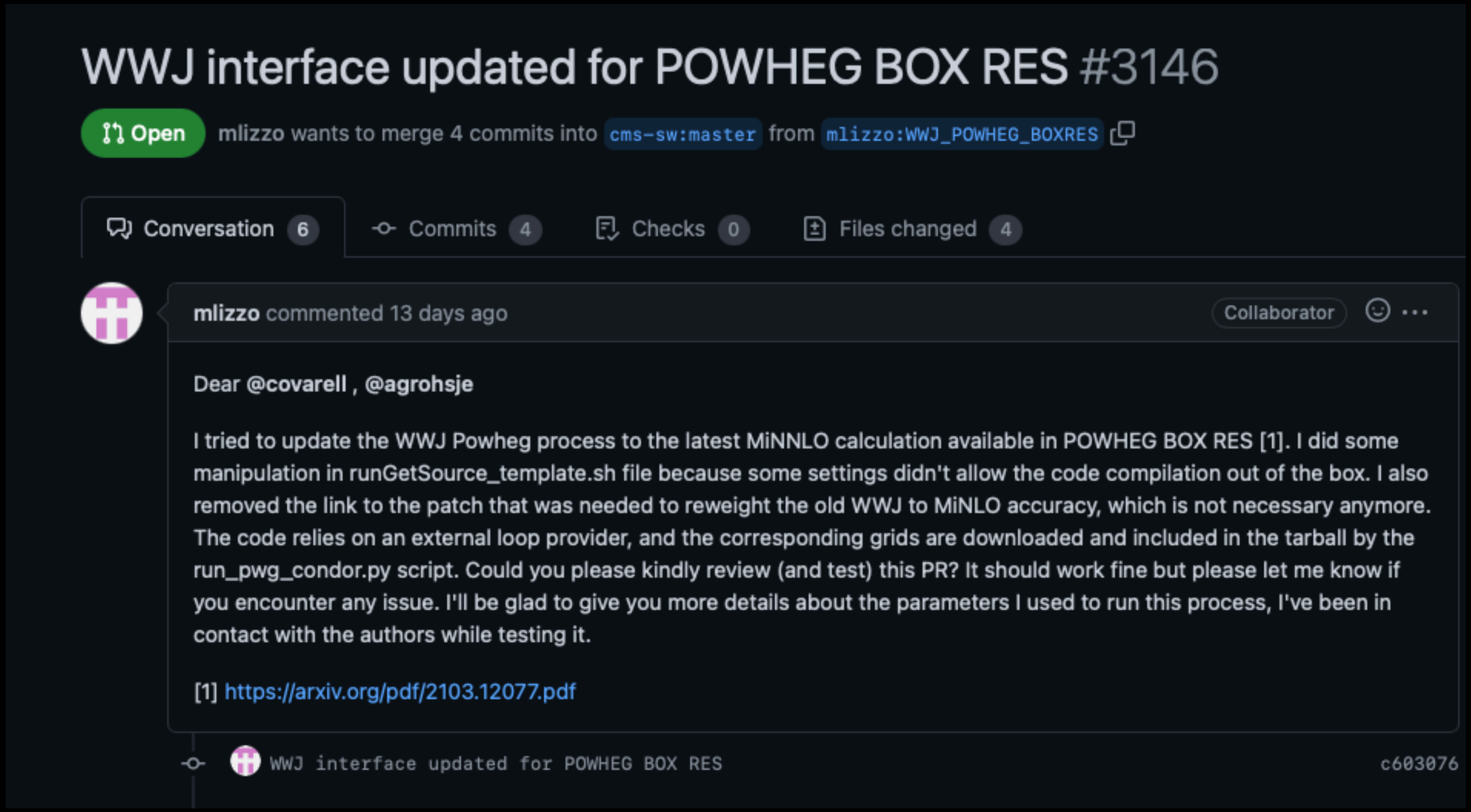
- Immediate goal:
  - Tuning studies underway with an aim to make minor modifications to CP5
    - Target parameter: intrinsic kT (set in the Monash tune)
      - Intrinsic k_T → partons have small *transverse* momenta in the hadron (100 MeV/s)
      - Introduced as a non-perturbative parameter
      - Effect can be seen in low p_T regime in Drell-Yan processes
    - Explore any benefits associated with moving to NNPDF4.0
- Medium-term goal:
  - Study dependence of intrinsic k_T on ISR/FSR parameters (use the same value of α_s for ISR and FSR)
- Compare results with Professor and MCNNTunes
- First results presented at GEN meeting on May 2nd: focus on reproducing CP5
- Detailed talk given at SMP-V workshop

- Check SMP-21-011 on the jet part, SMP-21-005 on W+c, SMP-20-015 for Z+b(b), SMP-20-007 and SMP-20-009 for DPS. Then, there is a large list of VV analyses (18-004, 20-014, 19-001, 22-001/not public yet). There are really many analyses. Let me know if you want more

- There are also VBS analyses, but I think we are not there yet to check generators in detail, except SMP-19-012.

Technical Updates

- Update in progress to incorporate latest MiNNLO calculation available in POWHEG BOX RES (W^+W^- production at NNLO+PS with MiNNLOPS: arXiv:2103.12077)



Generator development for Run III

- Current plan is to update to Madgraph5_aMC@NLO 2.9X
  (UL branch: 2.6.5)
  - Version 2.9X includes helicity recycling for optimizing production of VBF processes
  - Expect speedup for LO processes
  - Choice of version dictated by:
    - several issues with 3.3X (received support from authors but reporting and fixing bugs is likely to make progress slow)
    - The generator validation Level-3 conveners
      (Sanghyun Ko and Jie Xiao) have proposed to validate 2.9X
    - Currently gridpack production in progress
  - However, if your process needs 3.3X, branch exists in the genproductions repository

Interesting modeling studies in the HIG group

- GEN group relies heavily on PAG feedback for modeling studies
  - A number of MC modeling studies and improvements covered by Ulascan Sarica in the talk given recently at a PPD meeting
    - Features discussion of N^3LO description of $gg \to H \to ZZ$
    - Studies of several variables of interest modeled with JHUGEN+POWHEG
  - New updates in JHU GEN with regard to EFT simulations presented by Jeffrey Davis in GEN meeting
    - Corresponding pull request merged into official genproductions repository → available for usage by wider collaboration

Tuning Studies