CMS views on LHC MC WG

LHC MC WG
2024-11-14

Si Hyun Jeon

Overview

BOSTON UNIVERSITY

- How CMS Generator and relevant group is formed and works?
- What CMS expects through LHC MC WG?
  - With theory community
  - With other experiments
  - As a whole group altogether
- Small request from CMS to theory community
- LHC MC WG subgroup considerations

CMS Generator Group : GEN

BOSTON UNIVERSITY

- Subgroup structure for GEN
  - Physics Comparisons and Generator Tunes : Provides and validates underlying event tunes
  - Matrix Element and Future Generators : Maintains matrix element generator infrastructure for gridpacks (precompiled ME calculations)
  - Generator Validation : Maintenance and integration of validation tools, validations of MC samples in new CMSSW or generator releases
  - Generator Integration : Integration of generator related softwares into CMSSW
  - + Other contact persons for specific MC generator experts within CMS
- Liaisons for other groups
  - Top physics modelling, Exotic physics MC & interpretation
  - + MC contacts to facilitate the communication with GEN and physics working groups
  - Common background production team : Formed at the beginning of Run3, group of 4 people in charge of commonly used background processes (DY, W, top pair, VV, … 300 MC samples in total) with a sustainable database that can be extended up to HL-LHC era

CMS MC Production Overview

BOSTON UNIVERSITY

- Matrix element level calculations
  - BSM processes mostly rely on MadGraph5_aMCatNLO
  - SM processes majorly split into two, MadGraph5_aMCatNLO and Powheg, depending on the process
- Parton shower & hadronization
  - Pythia8 taken as default (underlying event simulation is also tuned & generated with Pythia8) for most of cases
- Underlying event tunes
  - Baseline tune developed during Run 2 with Pythia8 link, consistent interest in developing new tunes such as color reconnection link, intrinsic kt link, and also with different shower generator Herwig link
- Computing resources concentrating on priority physics programs : The most intense computation is for Powheg MiNNLO W/Z processes (~1m/event, mainly due to PDF reweighting) which was crucial for recent W mass measurement analysis link
- For validation workflows, we've started employing Rivet since Run 3 to verify our MC production setups

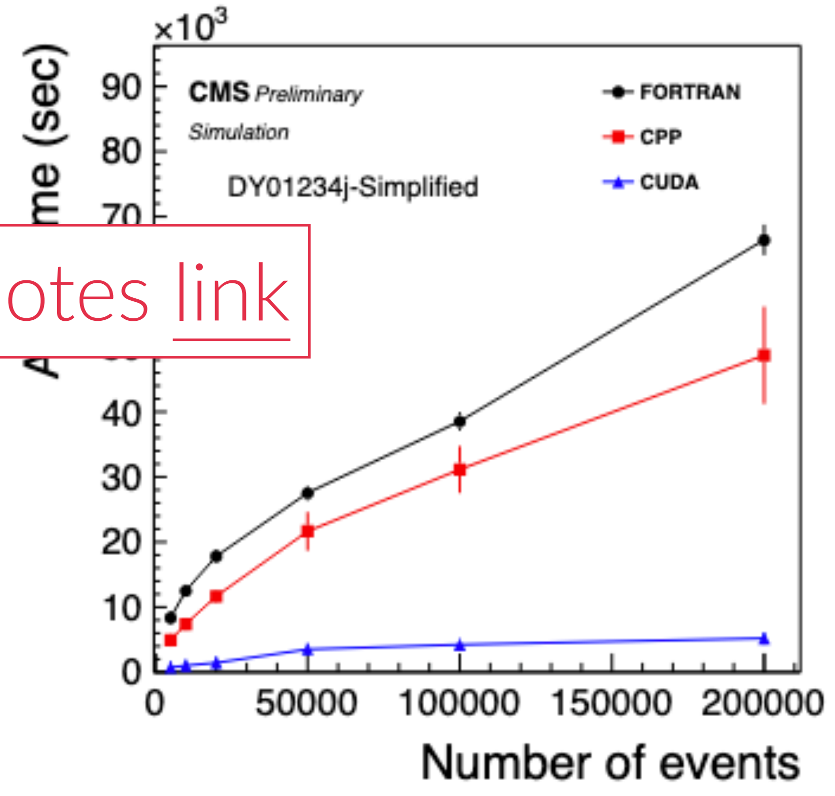
CMS ↔ Theory Community

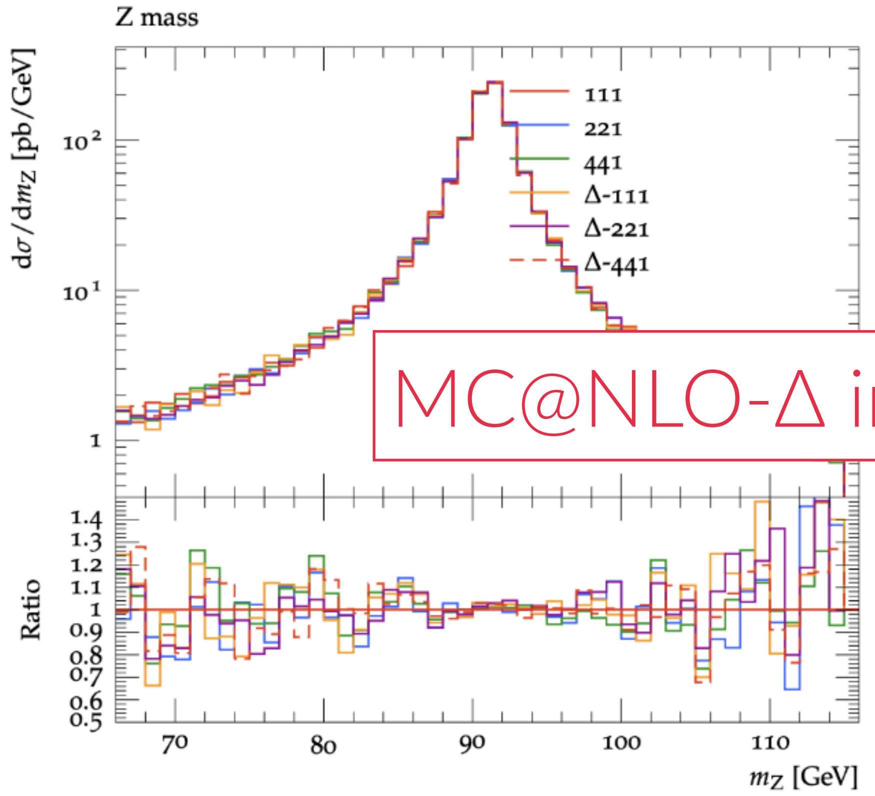
BOSTON UNIVERSITY

- Leveraging computing powers from experiments for MC generator developments and validations
- Benefits for experiment
  - Swiftly responding to theoretical advancements : Proactively plan for usage, avoiding unnecessary validation steps
  - Providing practical guidance tailored to MC production workflows in experiments : Predefined workflows for large scale production in experiments can limit the integration of generator advancements. Early stage collaboration/discussion could help with exploiting the full potential
- In turn, benefits for theory community
  - Access to larger computing resources and validating with real use case scenarios : e.g. CMS uses up to 4 parton multiplicity for DY process from MadGraph5_aMCatNLO, validations on advancements with lesser parton multiplicity is yet questionable for experiments

production time	FORTRAN	CPP - AVX2	CUDA
DY+0j	7m	6m	[illegible]
DY+1j	10m	10m	[illegible]
DY+2j	1h 12m	1h 10m	51m
DY+3j	**22h 40m**	**9h 4m**	**4h 18m**
DY+4j (Simplified)	**440h 46m**	**141h 20m**	**46h 26m**

mg4gpu CMS DP notes link





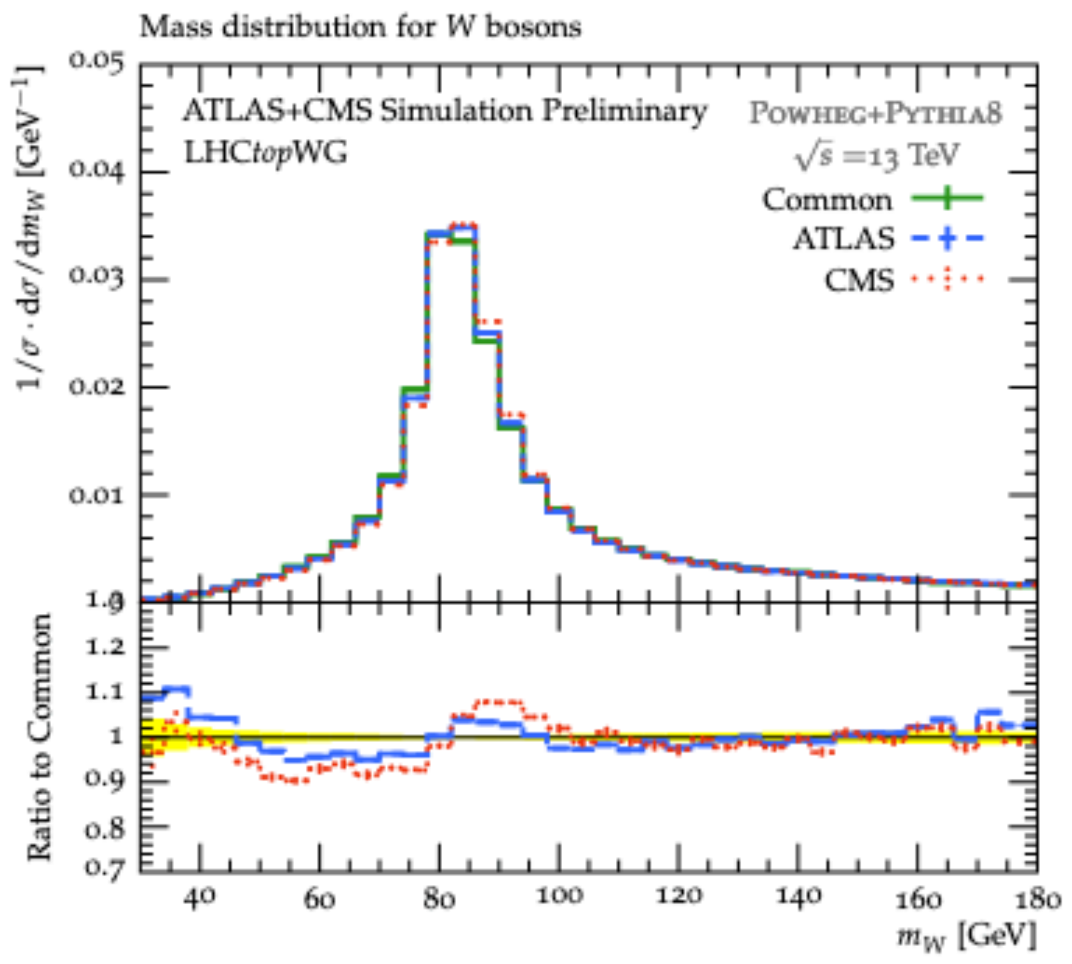
MC@NLO-Δ in CMS

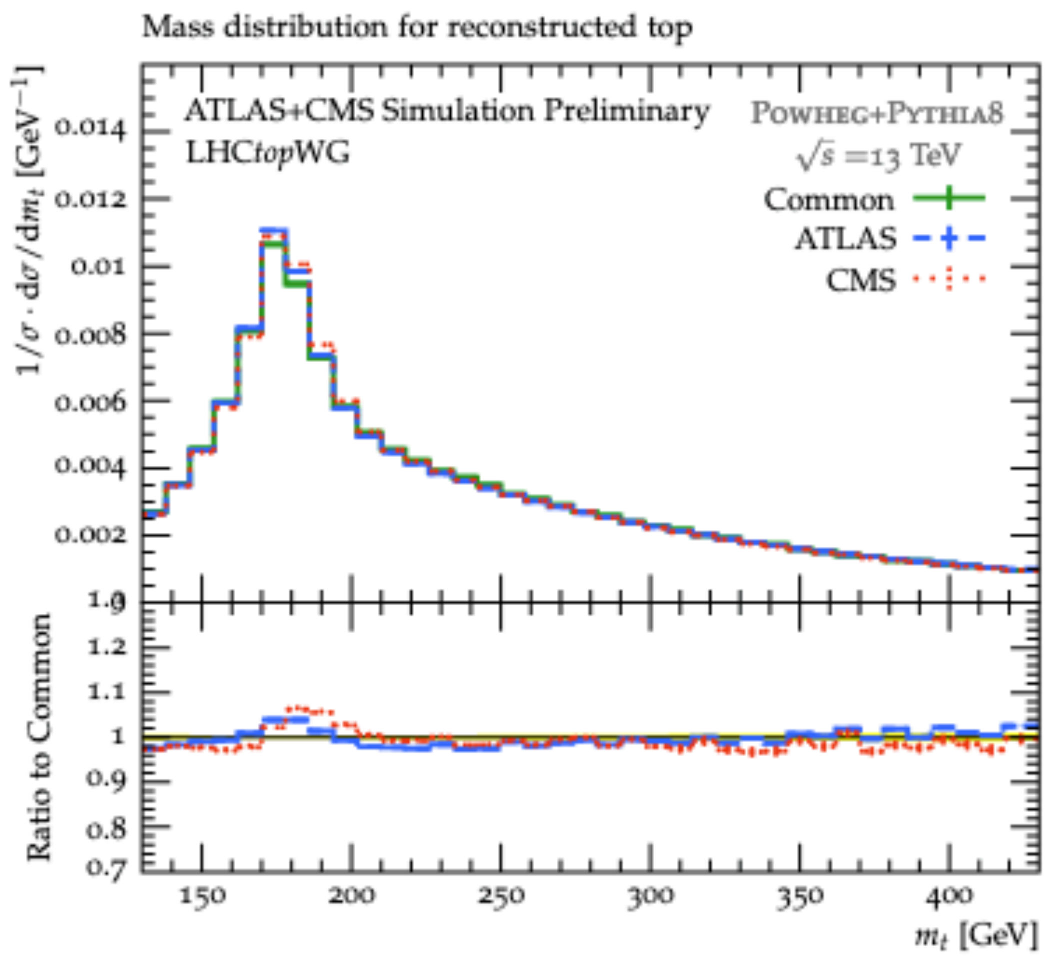
CMS ↔ Other Experiments

BOSTON UNIVERSITY

- Knowledge exchange with other experiments
  - Standard model physics process benchmarking : Already exists effort for top physics modelling link

Setting name	Setting description	CMS default	ATLAS default	Common Proposal
Powheg				
qmass	top-quark mass [GeV]	172.5	172.5	172.5
twidth	top-quark width [GeV]	1.31	1.32	1.315
hdamp	first emission damping parameter [GeV]	237.8775	258.75	250
wmass	$W^{\pm}$ mass [GeV]	80.4	80.3999	80.4
wwidth	$W^{\pm}$ width [GeV]	2.141	2.085	2.11
bmass	b-quark mass [GeV]	4.8	4.95	4.875
Pythia 8				
	Pythia 8 version	v240	v230	v240 (CMS) v244 (ATLAS)
	Tune	CP5	A14	Monash
PDF:pSet	LHAPDF6 parton densities to be used for proton beams	NNPDF31_nnlo _as_0118	NNPDF23_lo _as_0130_qed	NNPDF23_lo _as_0130_qed
TimeShower:alphaSvalue	Value of α_s at Z mass scale for Final State Radiation	0.118	0.127	0.1365
SpaceShower:alphaSvalue	Value of α_s at Z mass scale for Initial State Radiation	0.118	0.127	0.1365
MPI:alphaSvalue	Value of α_s at Z mass scale for Multi-Parton Interaction	0.118	0.126	0.130
MPI:pT0ref	Reference p_T scale for regularizing soft QCD emissions	1.41	2.09	2.28
ColourReconnection:range	Parameter controlling colour reconnection probability	5.176	1.71	1.80





- Sharing the computing burden by producing common MC formatted LHE or HepMC files
  - e.g. The time required to produce LHE files for Powheg MiNNLO is on par with the time needed for GEANT4 detector simulation per event. It would be nice if we can divide and conquer as a whole LHC collaboration
  - We can take W+9j study from S. Hoche et al. link as a great example of doing so, sharing HDF5 converted LHE files in Fermilab with (semi-)public access

CMS ↔ LHC MC WG

BOSTON UNIVERSITY

- LHC MC WG serving as a common platform to share issues that each one of us is facing
- There are several cases where if it's a problem for CMS, it's going to be a problem for other experiments as well
  - Small examples that CMS faced would be …
  - GoSam compilation with newer fortran releases failing to treat real and imaginary numbers properly leading to wrong physics results in Powheg higgs processes
  - High mass offshell W process sampling issues with large mlnu cut in MadGraph5_aMCatNLO (fixed by authors 1y ago)
  - Tau polarisation segmentation fault in Pythia8 when plugged in with certain type of LHE files (fixed by authors 2y ago) link
  - All these can be considered as common issues across experiments but often hard to fully notice before facing the same issue yourself from your experiment, even after such fixes are introduced
- Centralised documentations on such issues lets us to keep track of the hard works that theorists are handling and deliver the information back to our own collaborations

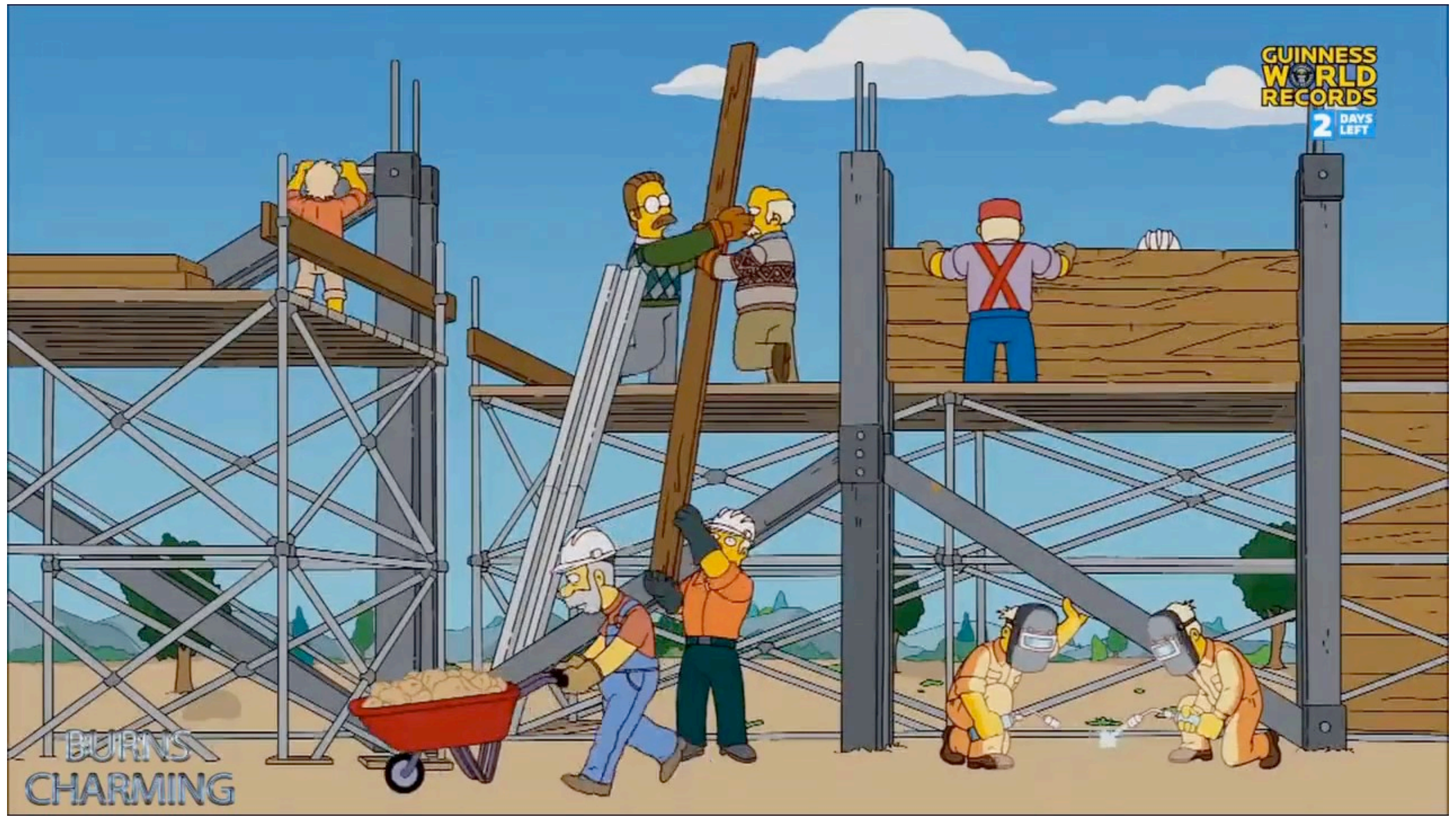
Request from CMS

BOSTON UNIVERSITY

- CMS fixes specific generator release before the data taking and tends to stick to the same version for few years and then move up to newer release
  - Main reason is due to computational stability and to avoid unnecessary surprises
  - For experiments, it's not that easy to update the releases as all possible sides (computing, physics, production management) are relevant to such issues and need to be guaranteed with stability and performance
- We were able to discuss such issues before Run 3 for MadGraph5_aMCatNLO
  - Asked for long-term-support (LTS) version v2.9.x that continuously gets updated (until 2025) with bug fixes but not adding new features that might break the stability of the tool
  - So MadGraph5_aMCatNLO currently supports two releases, v2.9.x (LTS) and v3.x.x which is being upgraded with new features everyday
  - Is this an approach that can be pursued by other generators as well? Obviously it will be a huge burden on theorists side if every different experiments ask for their own version for LTS
  - Would like to hear how much stability other experiments seek and wondering if such solution could help them as well if we can decide LTS version altogether to minimise the burden on theorists and share bugs/issues among experiments

Constructing LHC MC WG

- LHC MC WG subgroup suggestions from CMS
  - Generator performance : GPU/vectorisation or negative weight handling developments with computing advancements
  - Generator validation : Common workflow setup to validate MC generators utilising Rivet, MCPLOTS, or other tools across whole LHC physics enthusiasts
  - Tuning studies : Continuing the former LHC MB & UE WG, joint discussions between theorists and experimentalists on "what are the parameters to tune" and "what will be useful measurements"
- Any other ideas?



Summary

BOSTON UNIVERSITY

- Shared (biased) views from CMS towards LHC MC WG
- CMS would like to build a strong bridge with theory community and other experiments through LHC MC WG and expects fruitful outcomes from this collaboration
- Some considerations brought up
  - What could CMS exchange with theory community?
  - What could CMS exchange with other experiments?
  - Can we come up with a common platform sharing MC issues through LHC MC WG?
  - Moving towards HL-LHC where we expect longer run period, could we consider LTS (sustainable) releases for MC generators?

Backup Slides

CMS Interests

- CMS is in particular interested in
  - Negative weight handling : e.g. Cell resampler link
  - GPU developments : e.g. MadGraph5_aMCatNLO devel link
  - MC data formats : Yet using HepMC2 in CMS, possible to migrate LHE to HDF5 in near future?
  - Sustainable support for MC : e.g. GoSam not fully compatible with newer fortran releases