

# Experimental Needs

## MC net Meeting

Saptaparna Bhattacharya

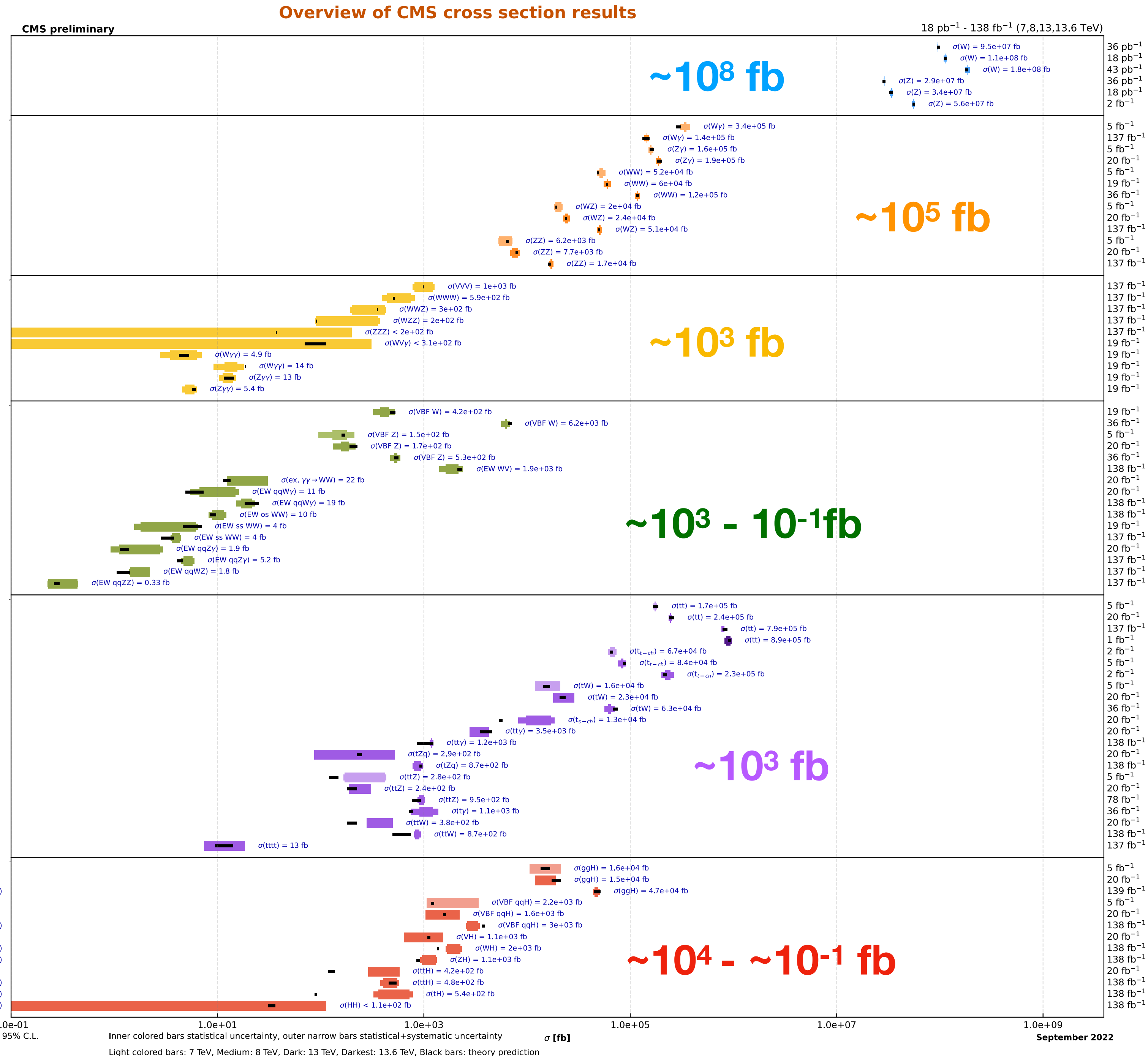
DESY, Northwestern University, Humboldt Fellow



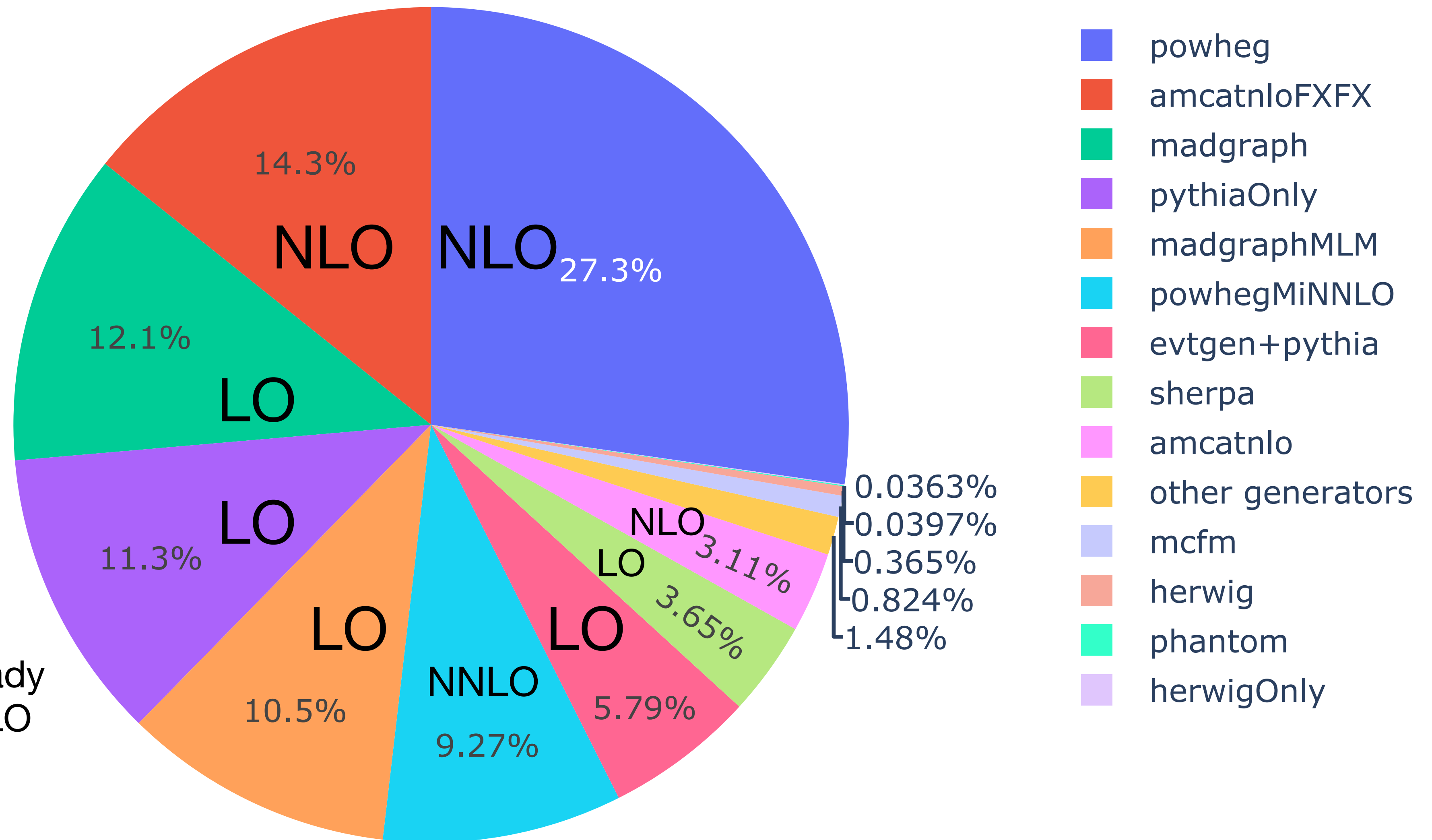


# Spans 10 orders of magnitude Spans 14 orders of magnitude when inelastic cross sections are taken into account

Category	Process	Energy	Reference	
Electroweak	W	7 TeV	JHEP 10 (2011) 132	
	W	8 TeV	PRL 112 (2014) 191802	
	W	13 TeV	SMP-15-004	
	Z	7 TeV	JHEP 10 (2011) 132	
	Z	8 TeV	PRL 112 (2014) 191802	
	Z	13 TeV	SMP-15-011	
di-Boson	W $\gamma$	7 TeV	PRD 89 (2014) 092005	
	W $\gamma$	13 TeV	PRL 126 252002 (2021)	
	Z $\gamma$	7 TeV	PRD 89 (2014) 092005	
	Z $\gamma$	8 TeV	JHEP 04 (2015) 164	
	WW	7 TeV	EPJC 73 (2013) 2610	
	WW	8 TeV	EPJC 76 (2016) 401	
	WW	13 TeV	PRD 102 092001 (2020)	
	WZ	7 TeV	EPJC 77 (2017) 236	
	WZ	8 TeV	EPJC 77 (2017) 236	
	WZ	13 TeV	Submitted to JHEP	
	ZZ	7 TeV	JHEP 01 (2013) 063	
	ZZ	8 TeV	PLB 740 (2015) 250	
ZZ	13 TeV	EPJC 81 (2021) 200		
tri-Boson	VVV	13 TeV	PRL 125 151802 (2020)	
	WWW	13 TeV	PRL 125 151802 (2020)	
	WWZ	13 TeV	PRL 125 151802 (2020)	
	WZZ	13 TeV	PRL 125 151802 (2020)	
	ZZZ	13 TeV	PRL 125 151802 (2020)	
	WV $\gamma$	8 TeV	PRD 90 032008 (2014)	
	W $\gamma\gamma$	8 TeV	JHEP 10 (2017) 072	
	W $\gamma\gamma$	13 TeV	JHEP 10 (2021) 174	
	Z $\gamma\gamma$	8 TeV	JHEP 10 (2017) 072	
	Z $\gamma\gamma$	13 TeV	JHEP 10 (2021) 174	
	VBF and VBS	VBF W	8 TeV	JHEP 11 (2016) 147
		VBF W	13 TeV	EPJC 80 (2020) 43
VBF Z		7 TeV	JHEP 10 (2013) 101	
VBF Z		8 TeV	EPJC 75 (2015) 66	
VBF Z		13 TeV	EPJC 78 (2018) 589	
EW WV		13 TeV	Submitted to PLB	
ex. $\gamma\gamma \rightarrow WW$		8 TeV	JHEP 08 (2016) 119	
EW qqW $\gamma$		8 TeV	JHEP 06 (2017) 106	
EW qqW $\gamma$		13 TeV	SMP-21-011	
EW os WW		13 TeV	Submitted to PLB	
EW ss WW		8 TeV	PRL 114 051801 (2015)	
EW ss WW		13 TeV	PRL 120 081801 (2018)	
EW qqZ $\gamma$		8 TeV	PLB 770 (2017) 380	
EW qqZ $\gamma$		13 TeV	PRD 104 072001 (2021)	
EW qqWZ		13 TeV	PLB 809 (2020) 135710	
EW qqZZ		13 TeV	PLB 812 (2020) 135992	
Top		tt	7 TeV	JHEP 08 (2016) 029
		tt	8 TeV	JHEP 08 (2016) 029
	tt	13 TeV	Accepted by PRD	
	tt	13.6 TeV	TOP-22-012	
	t $\tau$ -ch	7 TeV	JHEP 12 (2012) 035	
	t $\tau$ -ch	8 TeV	JHEP 06 (2014) 090	
	t $\tau$ -ch	13 TeV	PLB 72 (2017) 752	
	tW	7 TeV	PRL 110 (2013) 022003	
	tW	8 TeV	PRL 112 (2014) 231802	
	tW	13 TeV	JHEP 10 (2018) 117	
	t $\tau$ -ch	8 TeV	JHEP 09 (2016) 027	
	tty	8 TeV	JHEP 10 (2017) 006	
	tty	13 TeV	Submitted to JHEP	
	tZq	8 TeV	JHEP 07 (2017) 003	
	tZq	13 TeV	Submitted to JHEP	
	ttZ	7 TeV	PRL 110 (2013) 172002	
	ttZ	8 TeV	JHEP 01 (2016) 096	
	ttZ	13 TeV	JHEP 03 (2020) 056	
	t $\gamma$	13 TeV	PRL 121 221802 (2018)	
	ttW	8 TeV	JHEP 01 (2016) 096	
ttW	13 TeV	TOP-21-011		
tttt	13 TeV	EPJC 80 (2020) 75		
Higgs	ggH	7 TeV	EPJC 75 (2015) 212	
	ggH	8 TeV	EPJC 75 (2015) 212	
	ggH	13 TeV	Nature 607 60-68 (2022)	
	VBF qqH	7 TeV	EPJC 75 (2015) 212	
	VBF qqH	8 TeV	EPJC 75 (2015) 212	
	VBF qqH	13 TeV	Nature 607 60-68 (2022)	
	VH	8 TeV	EPJC 75 (2015) 212	
	WH	13 TeV	Nature 607 60-68 (2022)	
	ZH	13 TeV	Nature 607 60-68 (2022)	
	ttH	8 TeV	EPJC 75 (2015) 212	
	ttH	13 TeV	Nature 607 60-68 (2022)	
	tH	13 TeV	Nature 607 60-68 (2022)	
HH	13 TeV	Nature 607 60-68 (2022)		



# Generator usage split by events

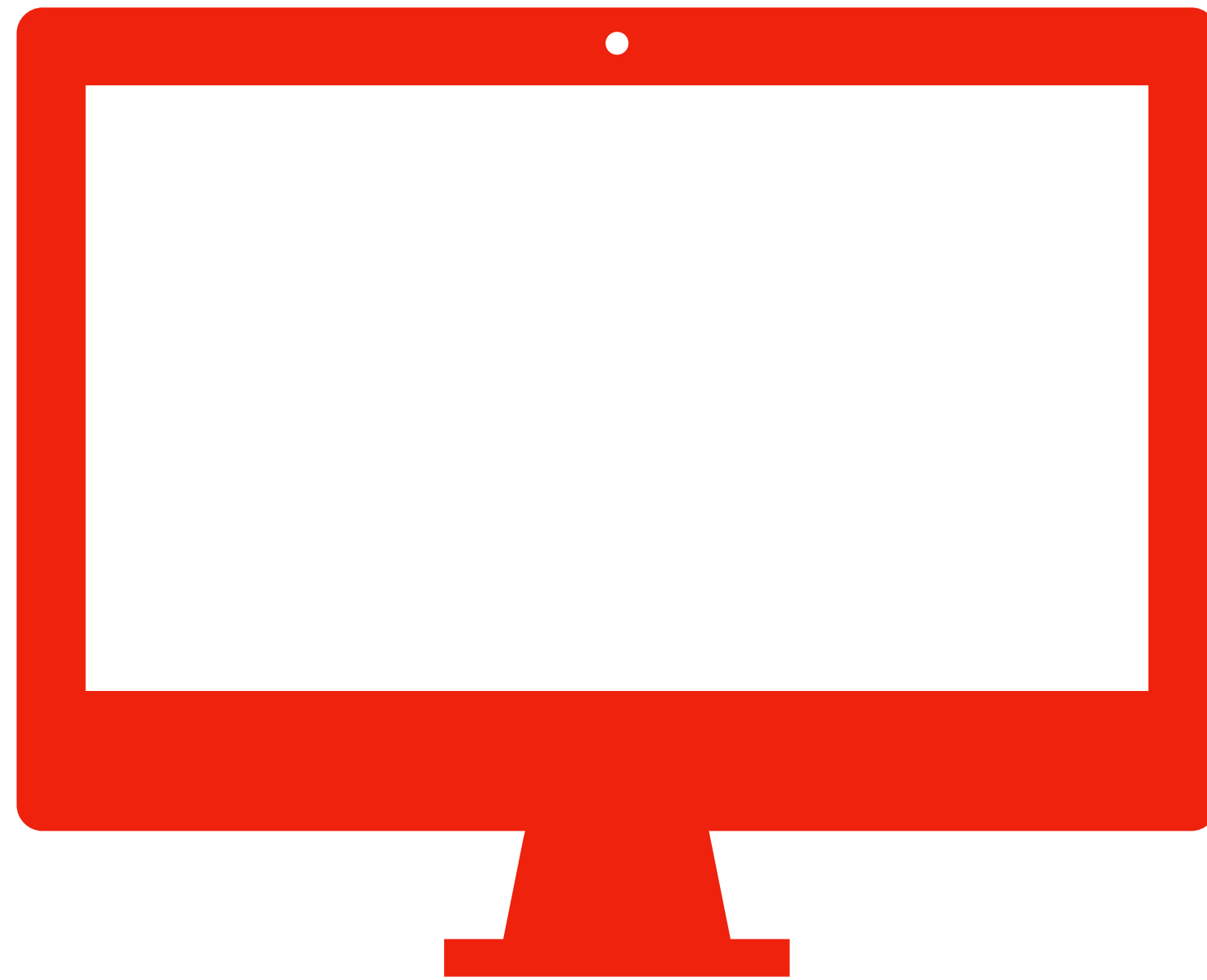


- NNLO: 9.27%
- NLO: 44.71%
- LO: 46.03%

- Most of the events already generated at least at NLO accuracy

- Major consumers of Madgraph\_aMC@NLO

# Computing challenges ahead...



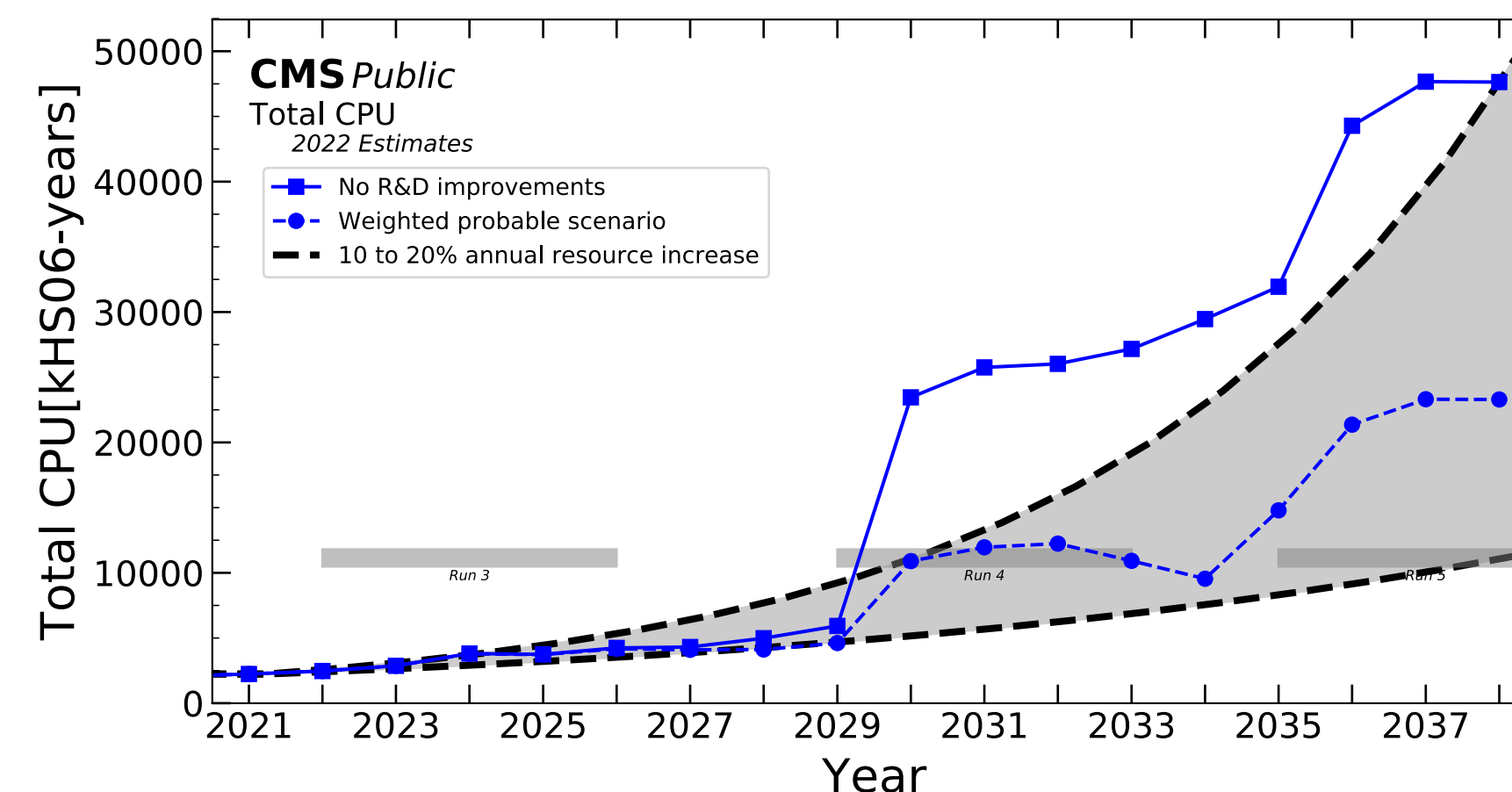
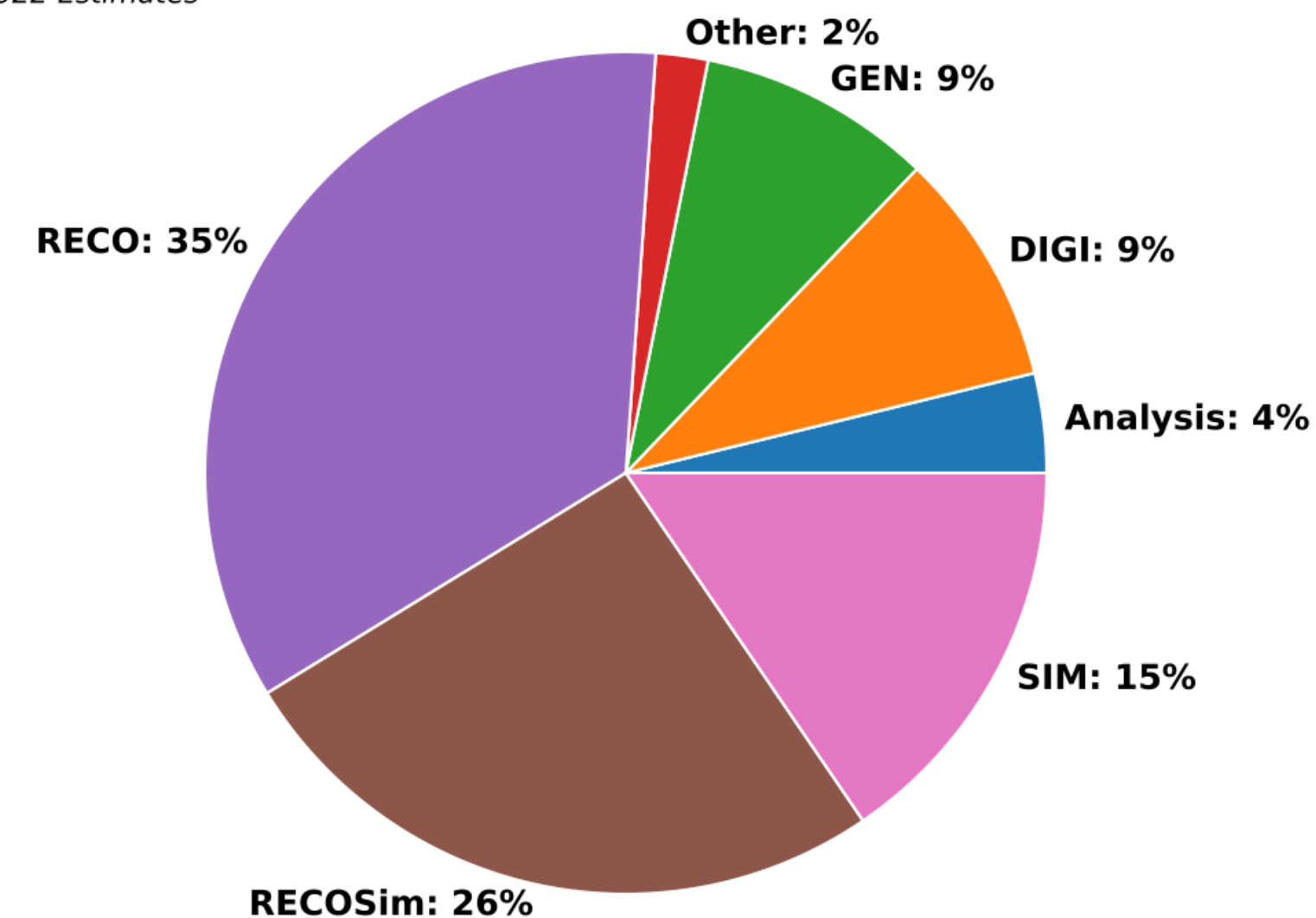


# Computing Budgets

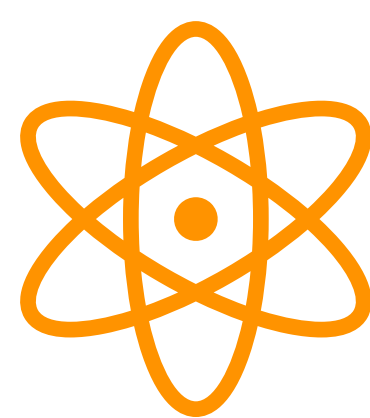
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/CMSOfflineComputingResults>

- The precision program comes at a cost of computing resources
- Current event generation step accounts for 9% of the total CPU footprint
- Expected to increase as physics processes are generated at high precision
- Current area of focus:
  - Implement multithreading for all workflows
  - Making large gridpacks read-only
  - Reduction of negative weights
  - GPU off-loading

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



CPU Efficiency



Physics based improvement

Complete or nearing completion indicated with:

# Need to focus on computing improvements now...

- As the workflows become more complicated and higher order corrections are implemented, it is foreseeable that event generators will start costing more in terms of CPU usage
- To this end, over the last few years, several improvements have been explored and implemented
  - Multithreading at the event generator level (talk given in [HSF meeting](#))
  - Mitigation of negative weights
  - Synergy with Madgraph4GPU group



# LO vs. NLO — the problem of negative weights

- While next-to-leading order (NLO) often provides best description of data
- Samples suffer from large fraction of negative weights
  - Lead to statistical dilution and computational costs for downstream processing
- For tails of distributions, often easier to generate samples with additional partons at leading order (LO)
- Recommend analysts to exercise caution and decide what is needed for their analysis

Rate of negative events

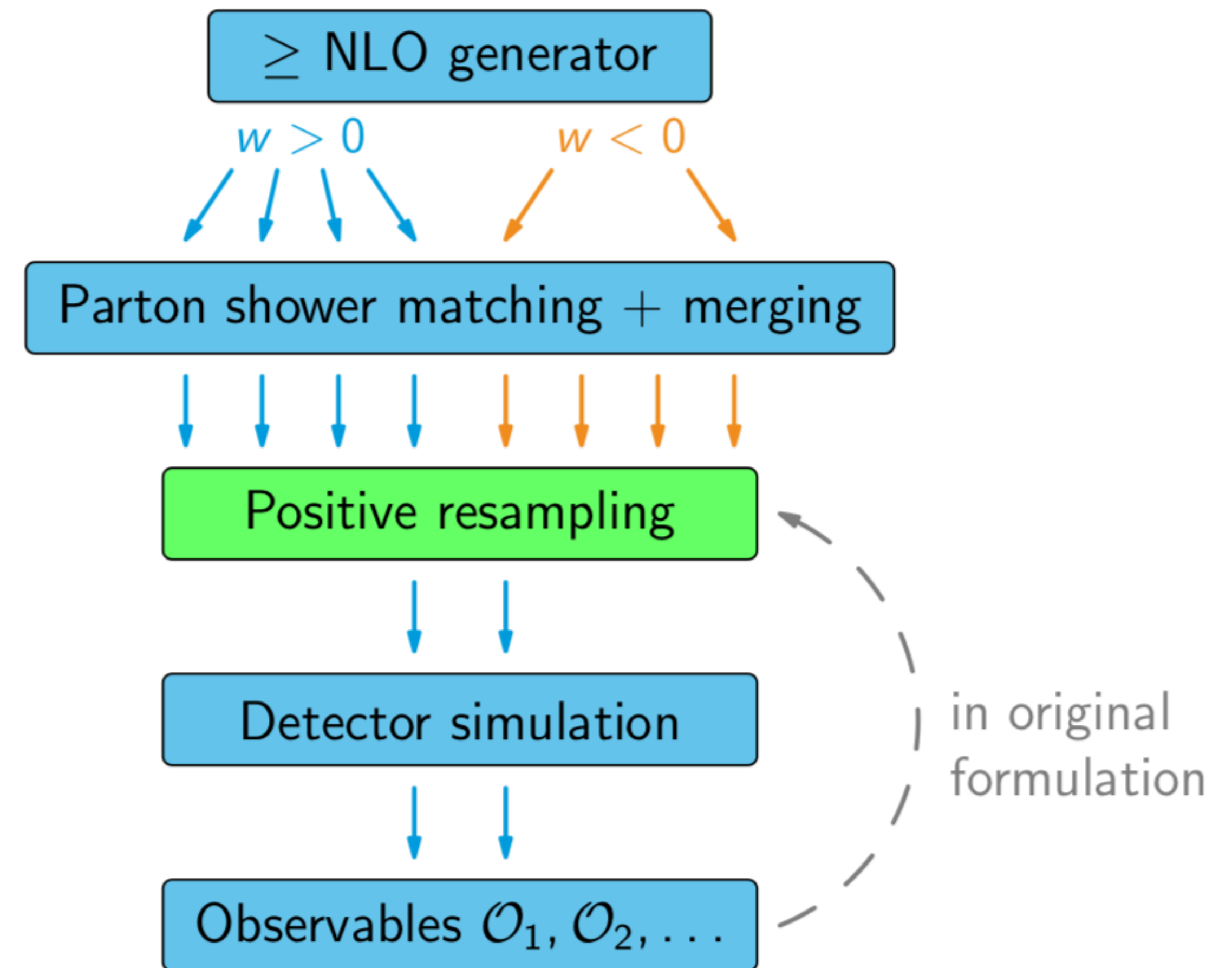
$pp \rightarrow e^+e^-$	6.9%	(1.3)
$pp \rightarrow e^+\nu_e$	7.2%	(1.4)
$pp \rightarrow H$	10.4%	(1.6)
$pp \rightarrow Hb\bar{b}$	40.3%	(27)
$pp \rightarrow W^+j$	21.7%	(3.1)
$pp \rightarrow W^+t\bar{t}$	16.2%	(2.2)
$pp \rightarrow t\bar{t}$	23.0%	(3.4)

Cost In sample size  
 $c(f) = \frac{1}{(1 - 2f)^2}$

From Olivier Mattelaer's talk

# Handling negatively weighted events

- Handling of negatively weighted events currently pursued with the generator theory community
- New and improved positive resampling and neural network resampling introduced by Andreas Maier and Jeppe Andersen
  - Implements redistribution of weights without incurring any bias and affecting an observable: leads to dramatic reduction in negative weight fraction
- First implementation in place → validation in progress





# Madgraph4GPU

Stephan Hageboeck, Stefan Roiser,  
Andrea Valassi, Olivier Mattelaer

- The CMS generators group is working closely with CERN computing and theorists to take advantage of developments toward a GPU based version of Madgraph
- Matrix element calculations are offloaded to GPUs
- Overall execution is still dominated by the Fortran part of the computation
- Currently in discussions on implementing a GPU-based event generation in central workflows

CUDA grid size		madevent		
		8192		
$gg \rightarrow t\bar{t}gg$	MEs precision	$t_{TOT} = t_{Mad} + t_{MEs}$ [sec]	$N_{events}/t_{TOT}$ [events/sec]	$N_{events}/t_{MEs}$ [MEs/sec]
Fortran	double	55.4 = 2.4 + 53.0	1.63E3 (=1.0)	1.70E3 (=1.0)
CUDA	double	2.9 = 2.6 + 0.35	3.06E4 (x18.8)	2.60E5 (x152)
CUDA	float	2.8 = 2.6 + 0.24	3.24E4 (x19.9)	3.83E5 (x225)

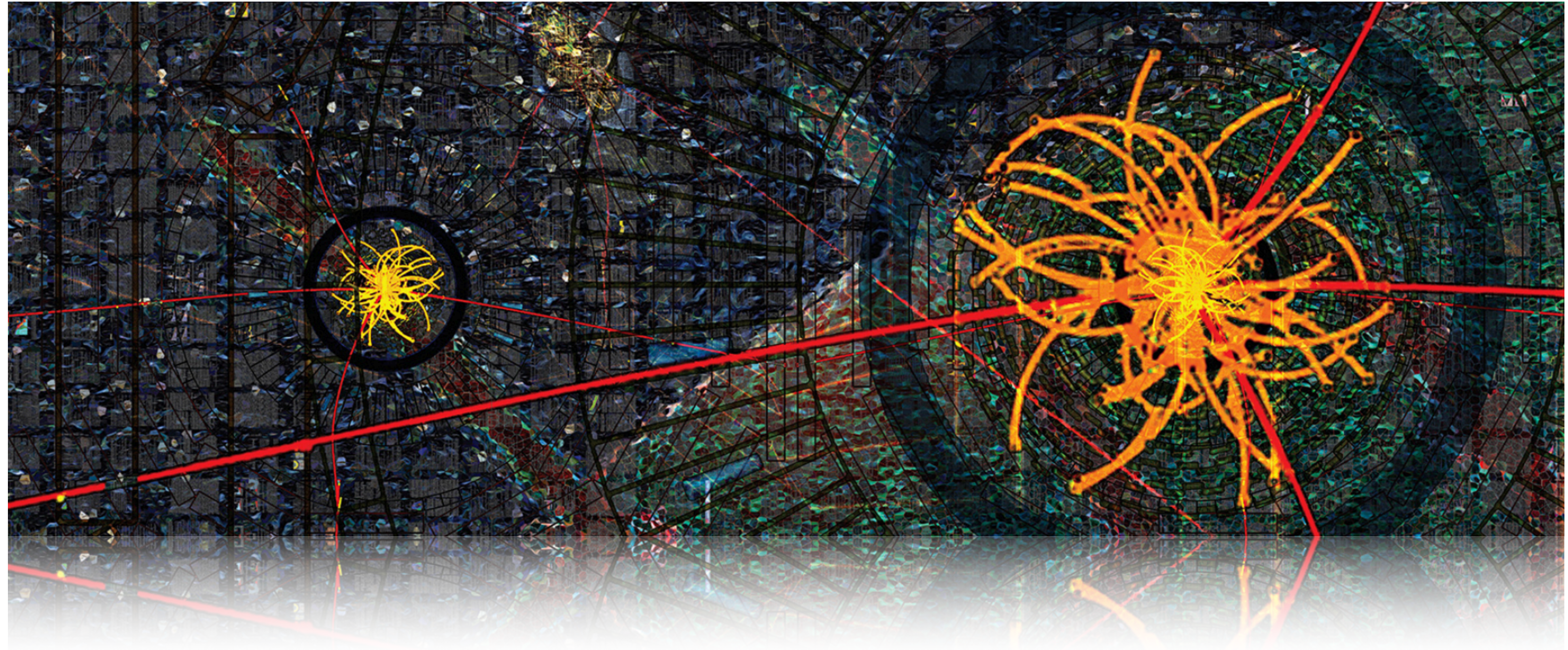
NVidia V100, Cuda 11.7, gcc 11.2

CUDA grid size		madevent		
		8192		
$gg \rightarrow t\bar{t}ggg$	MEs precision	$t_{TOT} = t_{Mad} + t_{MEs}$ [sec]	$N_{events}/t_{TOT}$ [events/sec]	$N_{events}/t_{MEs}$ [MEs/sec]
Fortran	double	1228.2 = 5.0 + 1223.2	7.34E1 (=1.0)	7.37E1 (=1.0)
CUDA	double	19.6 = 7.4 + 12.1	4.61E3 (x63)	7.44E3 (x100)
CUDA	float	11.7 = 6.2 + 5.4	7.73E3 (x105)	1.66E4 (x224)
CUDA	mixed	16.5 = 7.0 + 9.6	5.45E3 (x74)	9.43E3 (x128)

NVidia V100, Cuda 11.7, gcc 11.2



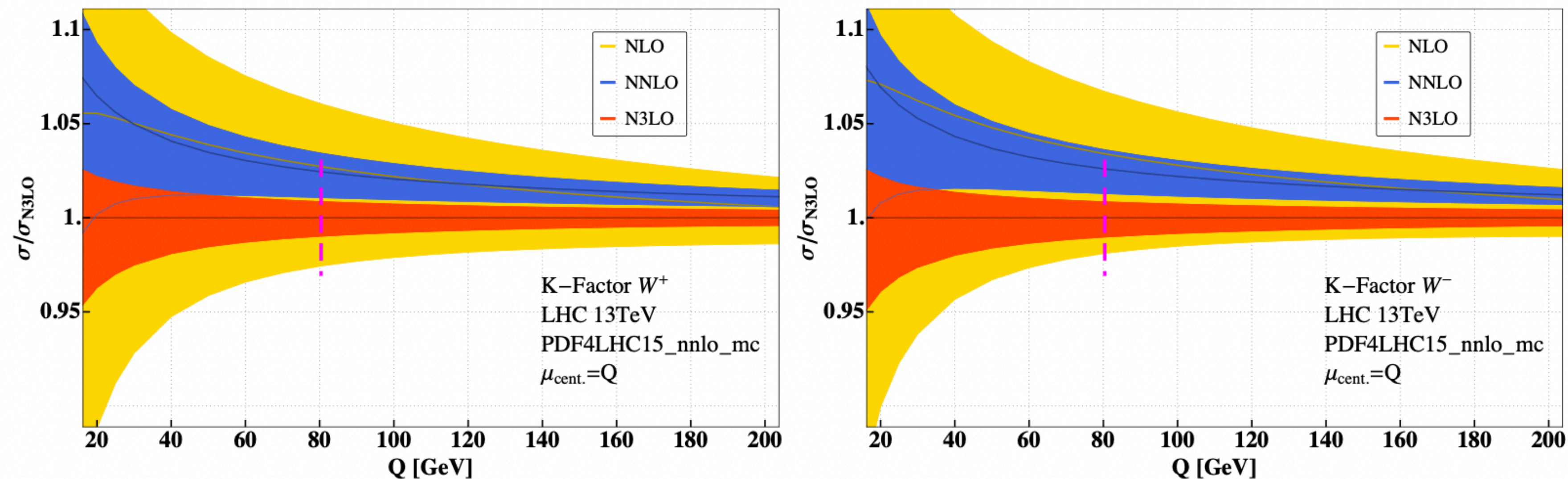
# Modeling of physics processes





# Uncertainties for Matrix Element predictions

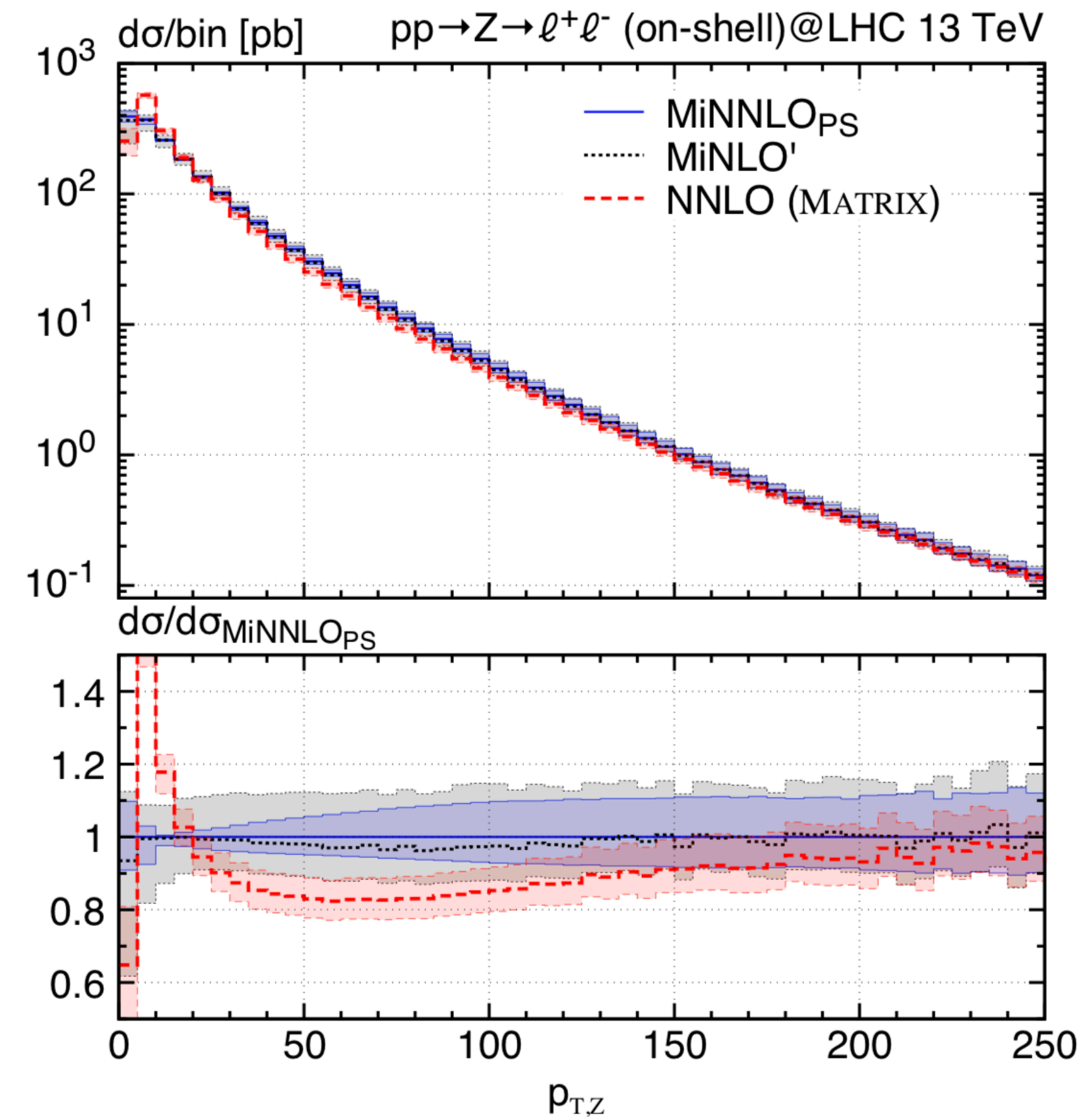
- Conventional method of estimating scale uncertainties (obtaining NNLO if we calculate up to NLO) exist
- Unfortunately uncertainty band has no statistical interpretation



N3LO computation lies outside NNLO band obtained by varying scales

# Sample generation at the precision frontier

- MiNNLO ([arXiv:1908.06987](https://arxiv.org/abs/1908.06987)) sample generated at next-to-next-leading-order (NNLO) accuracy + parton shower (NNLO+PS)
- 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
- However, MiNNLO is computationally expensive

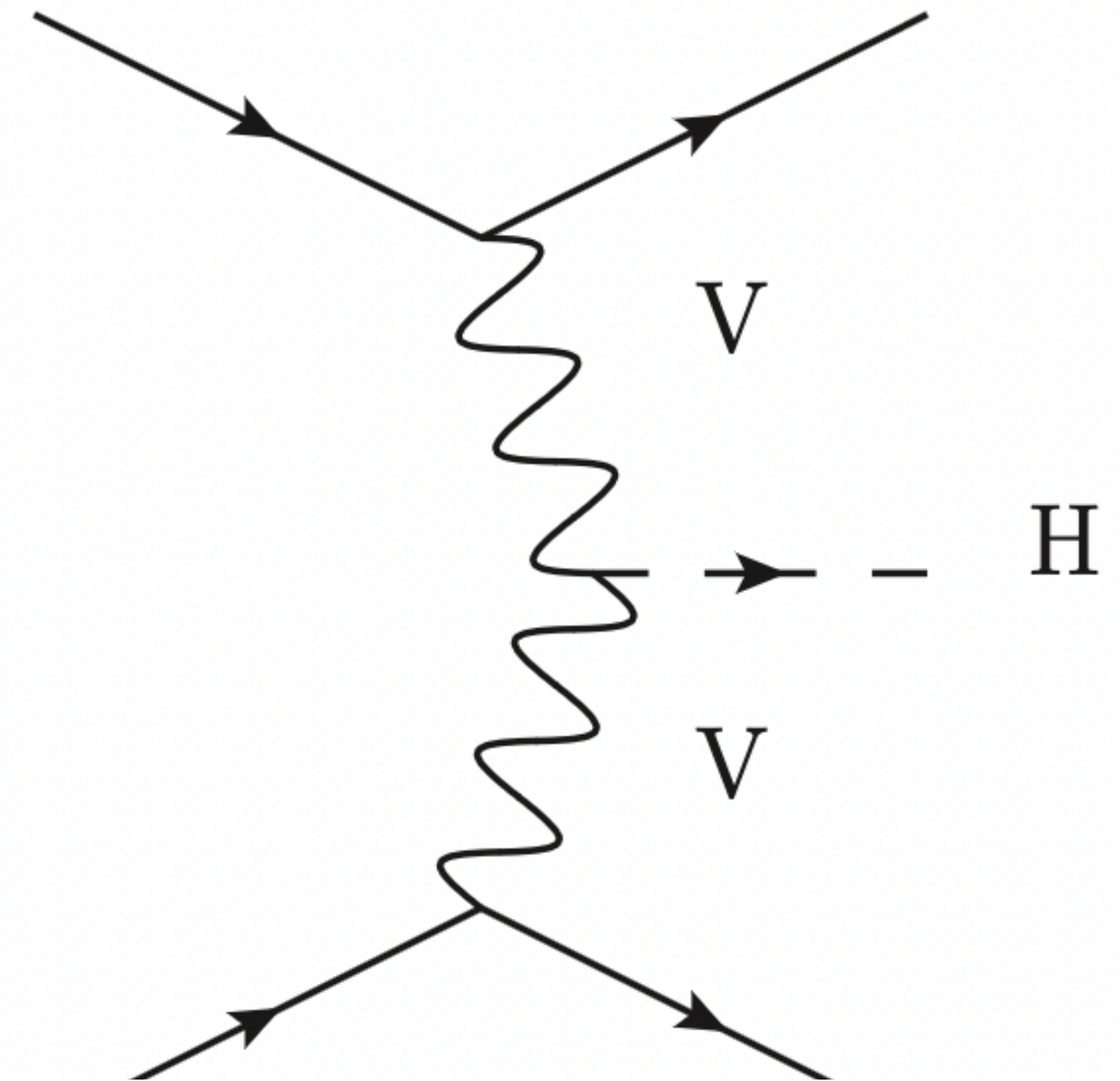




# Vector Boson Fusion modeling

- 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

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

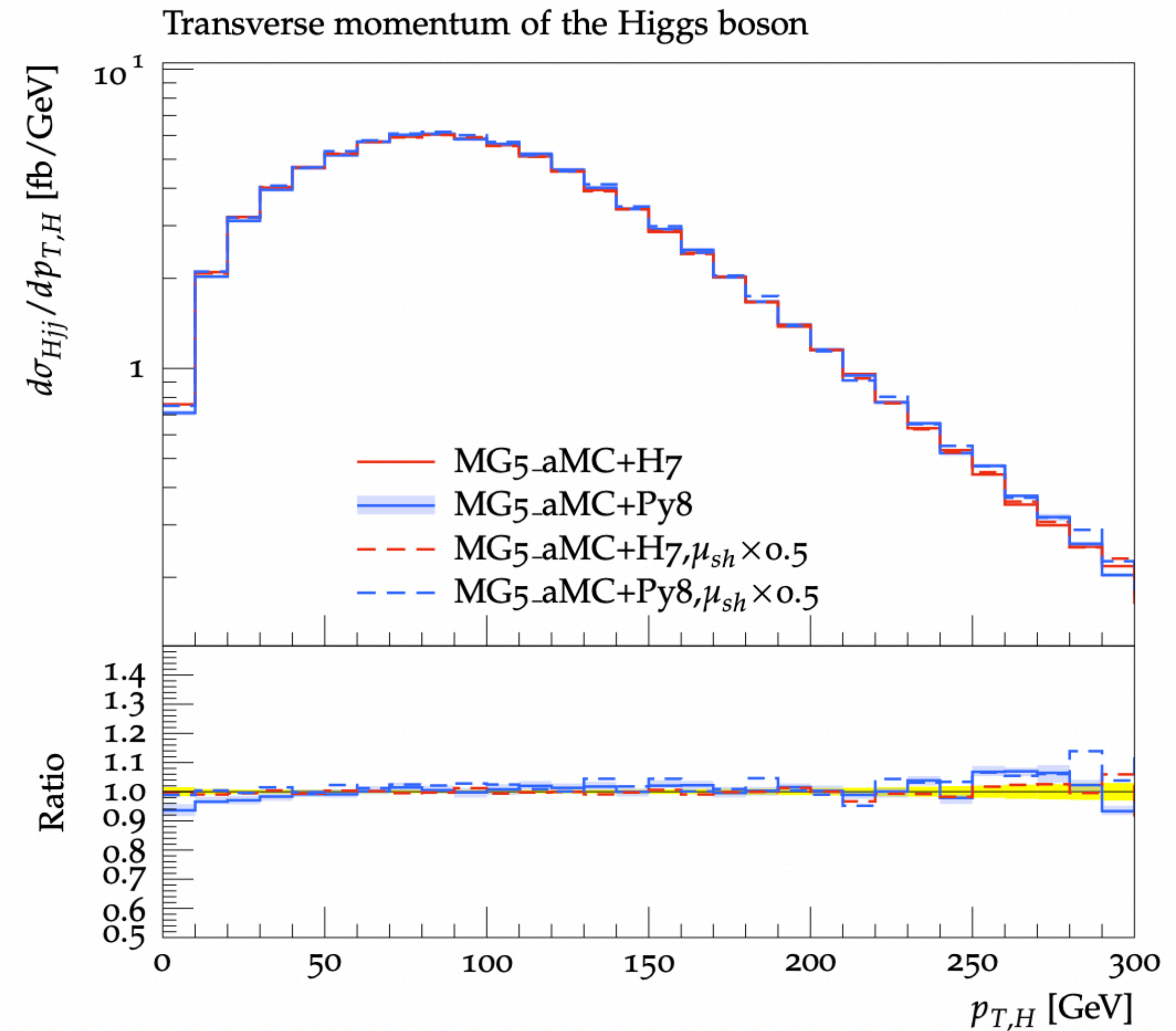


generator	matching	SMC	shower recoil	used in comparison
VBFNLO+Herwig7/Matchbox	⊕	HERWIG 7.1.5	global ( $\tilde{q}$ ) / local (dipole)	✓ ( $\tilde{q}$ )
HJets+Herwig7/Matchbox	⊕	HERWIG 7.1.5	global ( $\tilde{q}$ ) / local (dipole)	
MadGraph5_aMC@NLO 2.6.1	⊕	HERWIG 7.1.2	global	✓
MadGraph5_aMC@NLO 2.6.1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



# Vector Boson Fusion modeling

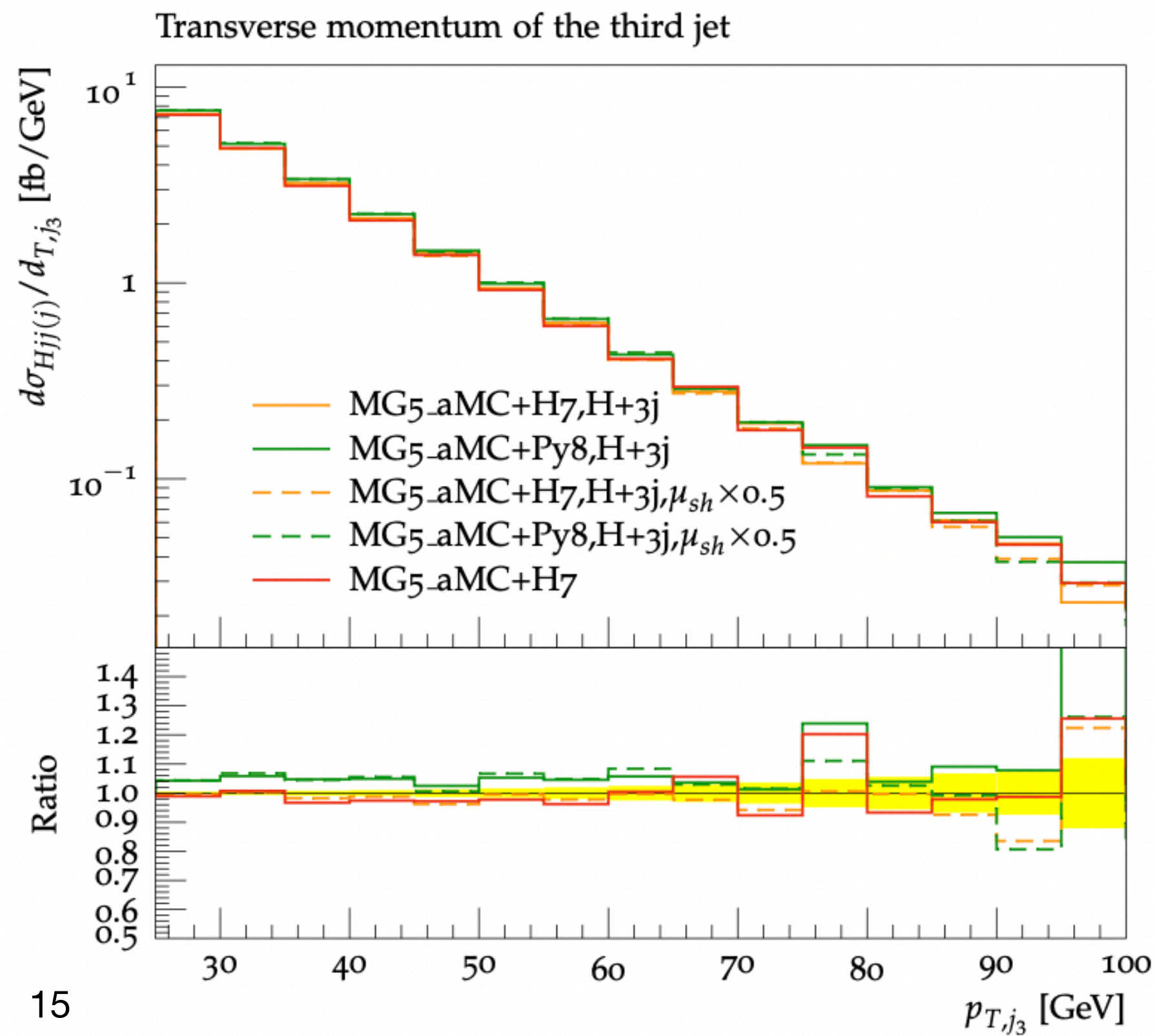
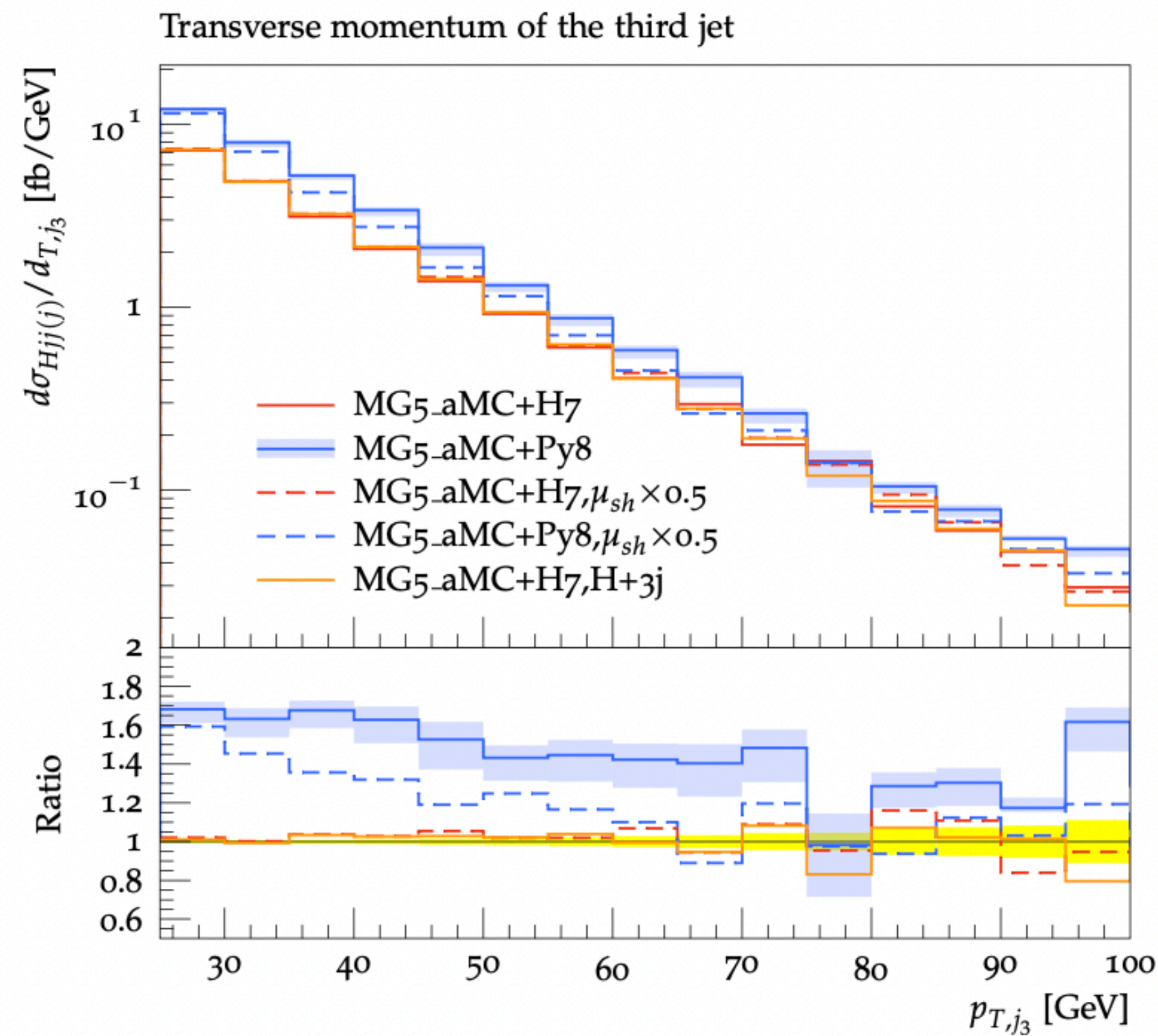
- Inclusive (in the number of jets) observable —  $p_T$  of the Higgs modeled well by all generators
- Modeling differences in jet bins almost exactly compensate
- Little to no dependence on choice of renormalization, factorization and shower scales





# Vector Boson Fusion modeling

- Looking at the third jet:
  - Global dipole recoil scheme clearly incorrect
    - In CMS, samples generated with local dipole recoil scheme
  - Can we do better?





# What we need in CMS

- Improvements in generator workflows *crucial* to usher in the precision era of the LHC
- Specific issues:
  - Handling large gridpacks efficiently — currently I/O is a bottleneck and the read-only functionality is the solution
  - Multithreading
  - GPU offloading
  - Reduction of negatively weighted events in NLO workflows
- Improved physics modeling in specific areas of phase space
  - In many cases improved physics modeling comes at the cost of computing resources
- With the challenges of the High Luminosity LHC in mind, it would be good to structure generator workflows to be compliant with modern software engineering and design standards, so they can be scaled on  $\mathcal{O}(100k)$  cores
  - Challenges arise from the vast number of external packages that are run → could benefit from further optimization
- Reduction of theoretical uncertainties in precision measurements → perhaps the topic of a future meeting?