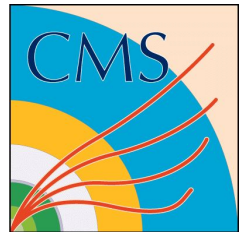


MC Modelling and Production at CMS

April 20th, 2019

[14th workshop on TeV Physics](#), Nanjing

Qiang Li (Peking University)



CMS MC Simulation

(LHE \rightarrow) GEN \rightarrow SIM \rightarrow DIGI \rightarrow RECO

- **Hard process/Matrix Element generation:**

parton level, perturbative QCD

- **Parton Shower/Hadronization:**

QCD/QED emissions to a low scale,
Produces hadrons from QCD partons

- *Multiple Parton Interaction*

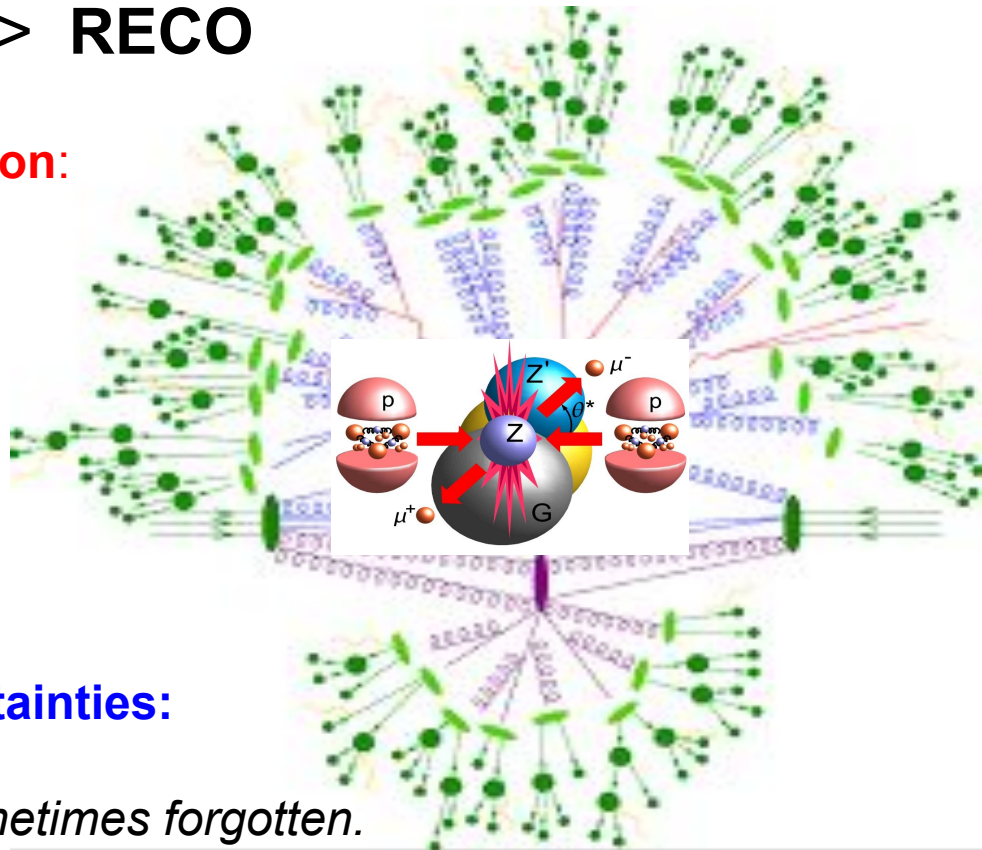
- *Detector Simulation and Digitization:*

- *Reconstruction:*

Factorised approach may lead to uncertainties:

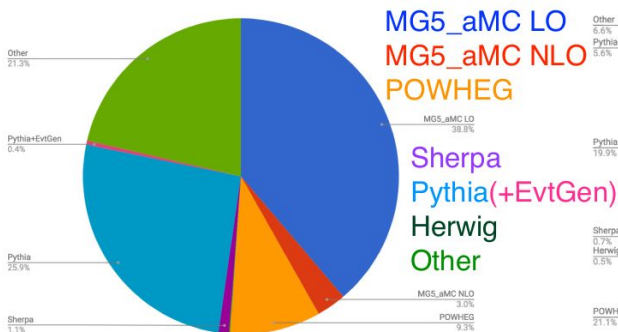
PDF, ME, “Tune” and PS, ...

\rightarrow *In most cases they are relevant but sometimes forgotten.*

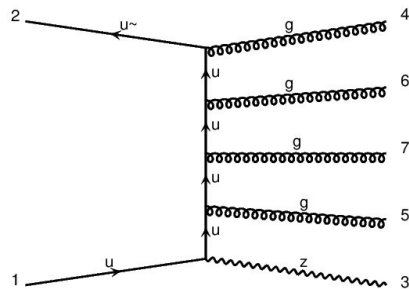
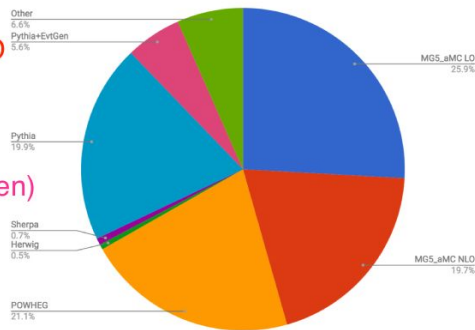


Generator usage in CMS

ME generator — by samples



ME generator — by events



based on 2016 MC campaign

- **Multi-leg LO and NLO consistently matched to the parton shower**

- **LO: Z+0/1/2/3/4 Jets**

Most commonly used in CMS: MG5_aMC@NLO+Pythia8 with MLM matching
 Most complex process up to 4 additional jets

- **NLO TTbar+0/1/2 Jets** **O(10-100)s/event**

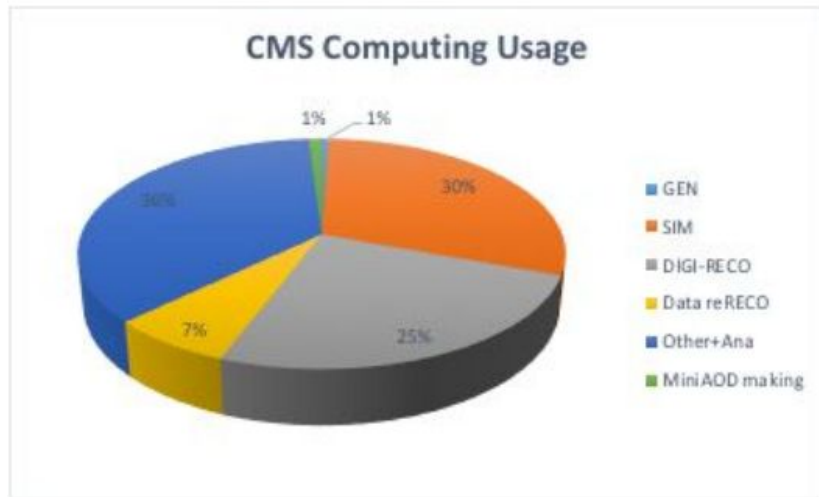
Most commonly used in CMS: MG5_aMC+Pythia8 with FxFx merging
 Most complex process up to 2 additional jets at NLO.

- **For signal, NNLO+PS**

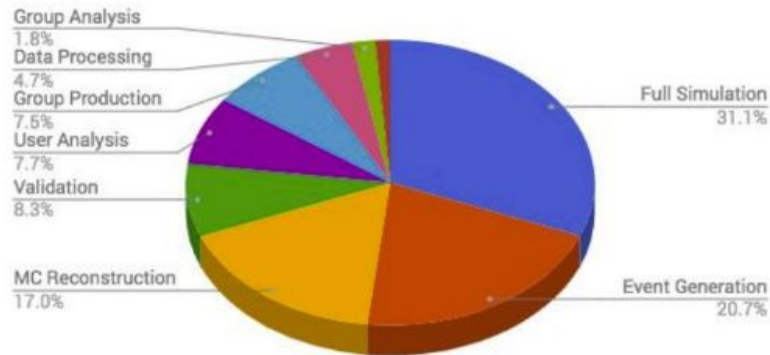
- **POWHEG: MINLO_NNLOPS**

CMS HWW reweight the nominal signal to this one

Generators Can be Computationally Intensive



US ATLAS Wall Clock CPU - 2016



CMS usage from 2017, ATLAS went down to 14.3% in 2017

[Elizabeth Sexton-Kennedy](#)

- These values vary from year to year as analysis needs vary
- CMS uses more LO samples in this year and grid-pack configurations

**Most HEP tools Typically executes 1 instruction at a time (per thread)
Much room for improvement.**

CMS Software

- **Main CMS software application:** [CMSSW](#)
 - Modular C++ application used for event generation, detector simulation, reconstruction, and analysis
- **Configuration of CMSSW runs steered with python configuration files**
- **Input and output with ROOT-based Event Data Model (EDM) files**
- **CMSSW links directly to many externals:** [Pythia](#), [Sherpa](#)

The screenshot shows the GitHub repository for CMS Software. At the top, it displays the repository name 'cms-sw / cmssw' and statistics: 70 Watchers, 594 Stars, and 2,747 Forks. Below this, there are navigation links for Code, Issues (354), Pull requests (95), Projects (0), Wiki, and Insights. A link to 'CMS Offline Software' is provided with the URL 'http://cms-sw.github.io/'. There are also tags for 'hep', 'cern', 'cms-experiment', and 'c-plus-plus'. A progress bar shows 195,338 commits, 76 branches, 1,820 releases, and 756 contributors. Below the progress bar, there are buttons for 'Branch: master', 'New pull request', 'Create new file', 'Upload files', 'Find file', and 'Clone or download'. The main content area shows a list of recent pull requests and commits, including a merge pull request from Dr15Jones/cleanupDQM_L1TMonitor and several other pull requests related to Alignment, AnalysisAlgos, AnalysisDataFormats, BigProducts/Simulation, and CalibCalorimetry.

cms-sw / cmssw

Watch 70 Star 594 Fork 2,747

<> Code Issues 354 Pull requests 95 Projects 0 Wiki Insights

CMS Offline Software <http://cms-sw.github.io/>

hep cern cms-experiment c-plus-plus

195,338 commits 76 branches 1,820 releases 756 contributors

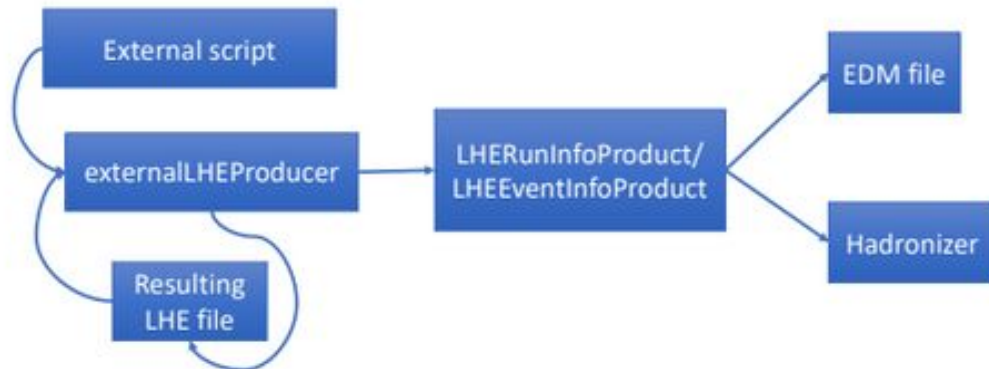
Branch: master New pull request Create new file Upload files Find file Clone or download

cmsbuild Merge pull request #25161 from Dr15Jones/cleanupDQM_L1TMonitor Latest commit ac29e29 21 hours ago

Alignment	Merge pull request #24862 from guitargeek/unary_binary_2	5 days ago
AnalysisAlgos	Code checks	13 days ago
AnalysisDataFormats	AnalysisDataFormats/TopObjects: Fix bug found by clang warning:	8 months ago
BigProducts/Simulation	* Add SimG4Core/PrintGeomInfo to Big products	3 years ago
CalibCalorimetry	Fixed potential memory leak in CastorDbASCIIO	9 days ago

ExternalLHEProducer and Gridpack

- **Compiling code on batch workers discouraged, long init time discouraged**
- **Pre-generated and compiled code** with initial phase space integration results stored in a tarball (with fixed model/run parameters) -> **Gridpack**
- Gridpack **placed in CVMFS** and accessed by remote jobs
- Gridpack location – a parameter of the **externalLHEProducer** module
- Input arguments: number of events, random number seed and **possibly nCPU**



Gridpack Production

Gridpack (LHE) → GEN → SIM → DIGI → RECO

LSF / Condor / CMS-Connect (grid-like condor jobs using CMS Global Pool)

Gridpack size can be an issue (>500MB for the tarball and 5GB decompressed)

We maintain [scripts](#)

for all major generators

to produce gridpack tarballs

– **Madgraph aMC@NLO**

tt012FxFx ~72h@lxplus batch

DY01234MLM ~15h@cms-connect

– **POWHEG**

– **Sherpa, Herwig7** and others

tt01234 MEPS Sherpa

~70h@cms-connect

~O(100s)/event

cms-sw / [genproductions](#) Unwatch 18 Star 23 Fork 390

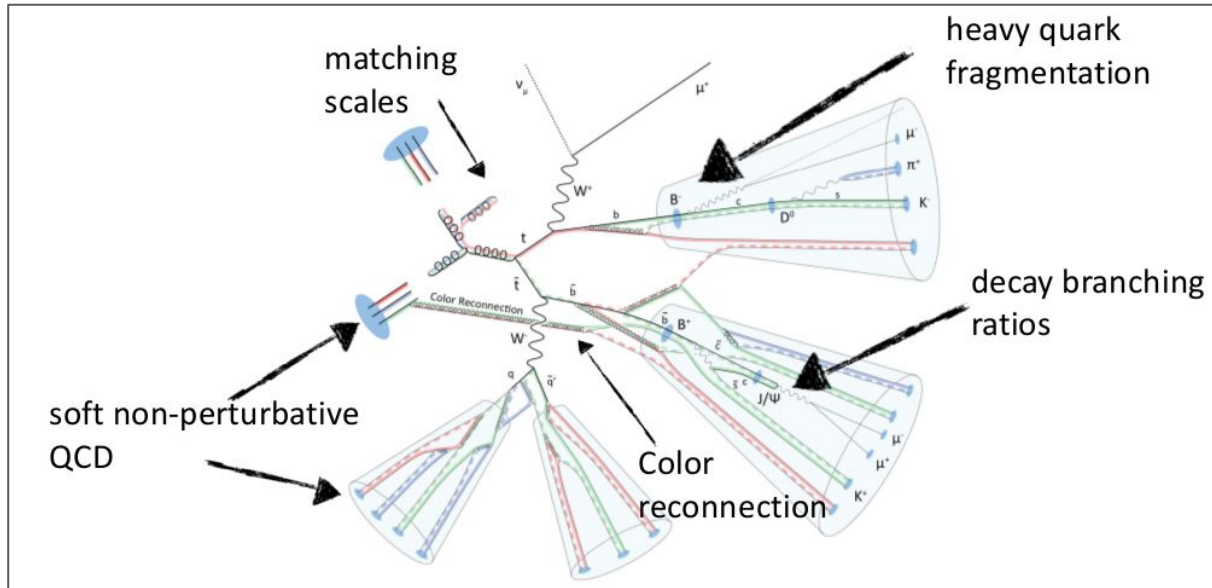
<> Code Issues 13 Pull requests 20 Projects 0 Wiki Insights Settings

Branch: master [genproductions / bin /](#) Create new file Upload files Find file History

File/Folder	Description	Last Commit
..		Latest commit f93c:112 3 days ago
Alpgen/cards/production/13TeV	Cards from Emrah, pLHE request.	3 years ago
BlackMax/cards/production/13TeV	String Ball Cards from the Black-Hole Analysis Group for 2016	3 years ago
CalcHEP/cards/production	Delete random.txt	a month ago
Charybdis/cards/production/13TeV/...	Cards for 2016 BH analysis with significant improvements over the 201...	2 years ago
CompHEP/cards/13TeV/CompHEP...	CompHEP files for Wprime->t+b with mixed chirality	3 years ago
FPMC	more detailed description	9 months ago
GenValidation	added exit command in case job fails	7 months ago
JHUGen	update VBF offshell card	16 days ago
MCFM	use the random seed to choose the events to keep	7 months ago
MadGraph4/cards/production/13Te...	Merge pull request #1126 from Saptaparna/GenFragmentsV5	2 years ago
MadGraph5_aMCatNLO	cards for wide width Tprime to Wb vlq sample	3 days ago
Phantom	adding explicit reference to the top cut	5 months ago
Powheg	update the POWHEG Wgamma folding parameters in order to reduce the nu...	4 days ago

Showering and Hadronization

- Pythia 8.226 default; Herwig++ replaced by Herwig7.
- Fragment settings different depend on Matrix Element
 - LO, NLO; MLM, FxFx matching/merging; POWHEG emission vetoing
- Pythia 8.240 integrated in CMS recently:
 - NLO shower DIRE, VINCIA; [Dipolerecoil](#) option for better description on VBS



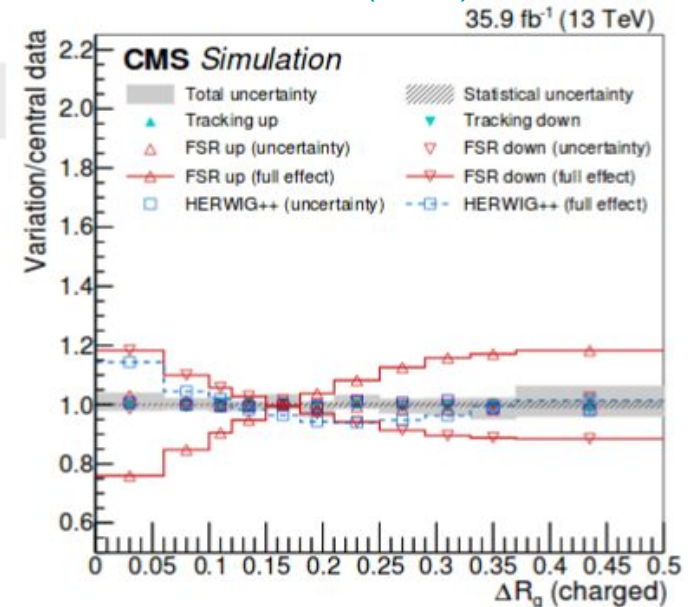
Showering and Hadronization

- Parton Shower Weights, Tune Variations etc included

$(ISR, FSR) \otimes (\mu_R, cNS) \otimes (g \rightarrow gg, g \rightarrow q\bar{q}, q \rightarrow qg, b/t \rightarrow b/t + g) \otimes (\text{up, down})$

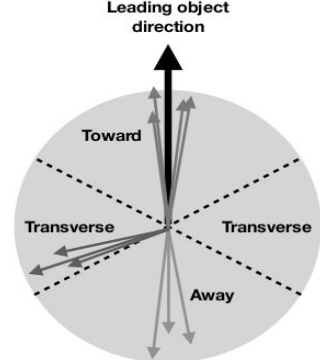
Source	Handle	Weights	Variation	Note/Reference	Dedicated studies
Shower scales	ISR scale (SpaceShower:renormMultFac)	YES	0.5-2.0	FSR variations can be scaled down by $\sqrt{2}$ from LEP	TOP-15-011, TOP-16-021 TOP-17-13, TOP-17-015, ...
	FSR scale (TimeShower:renormMultFac)		0.5-2.0		
ME-PS Matching	hdamp	No	hdamp=1.58 m_t +0.66-0.59 m_t	see TOP-16-021	Starting scale variations for MGS_aMC@NLO
Soft QCD	UE parameters	YES	UE tune up/down	See TOP-16-021 MPI & CR strength doesn't affect resonance decays	TOP-17-015 GEN-17-001
Color reconnection (odd clusters)	MPI based, QCD-inspired, gluon move	No	different models	CR affecting resonance decays	TOP-17-13, TOP-17-015
Fragmentation	momentum transfer from the b-quark to the B hadron: $x_b = p_T(B)/p_T(b\text{-jet})$	YES	Vary Bower-Lund parameter within uncertainties from LEP/SLD fits	see TOP-16-022 (re-weight x_b)	
Flavor response/hadronization	Pythia vs Herwig	No	Vary the JES independently per flavour for light, g, c, b.		
Decay tables	B semi-leptonic BR	YES	vary semileptonic BR +0.77%/-0.45%	re-weight the fraction of semi-leptonic b jets by the PDG values (scale Λ_c to match PDG)	

CMS: [Jet substructure in ttbar events at 13 TeV \(2016\)](#)

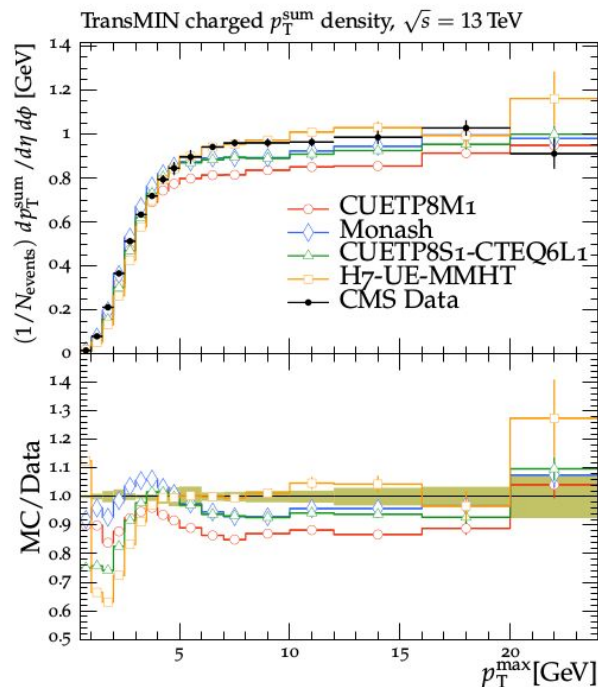


angle between the groomed subjects

MC Tuning: CUETP8M1 Tune



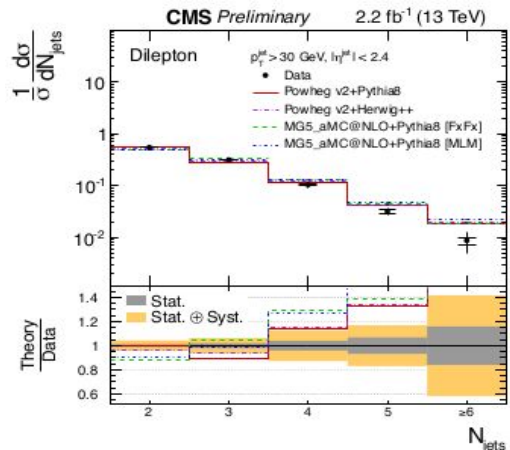
Until 2017 analyses (except 2016 tbar), **Pythia8 CUETP8M1 tune** [EPJC 76 (2016) 155] based on the **Monash** tune was used. Fitting MPI energy dependence parameters to UE data @ \sqrt{s} = 0.9, 1.96 & 7 TeV



- ◆ α_s and shower parameters kept as in Monash $\rightarrow \alpha_s^{\text{ISR/FSR}} = 0.1365$ despite the preferred values of 0.130 in LO and 0.118 in NLO matrix elements/ PDF sets.
 - α_s^{FSR} in Monash \rightarrow by fitting Pythia8 predictions to LEP event shape measurements and α_s^{ISR} is just assumed to be the same as α_s^{FSR} .
 - $\alpha_s^{\text{MPI}} = 0.130$ set to the value preferred in the LO PDF set.
- ◆ Revisited the shower parameters
 - Starting from parton shower in tbar events \rightarrow **CUETP8M2T4 tune**.
 - Using a NNLO PDF set in PS \rightarrow **CP5 (and CP0-4 tunes)**.

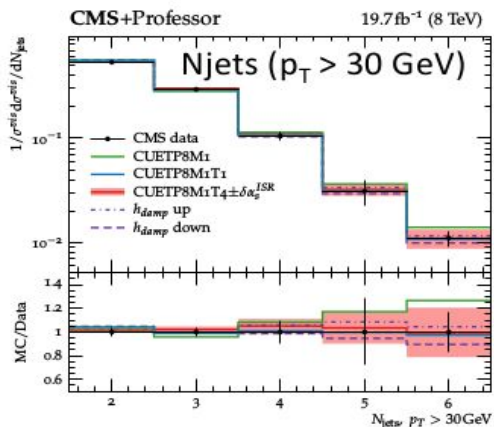
CUETP8M1 does not describe well the central values of 13TeV data

MC Tuning: CUETP8M2T4 Tune [TOP-16-021]



- Predictions overshoot the data for large jet multiplicities when out of the box parameters are used (in Monash-based tunes: $\alpha_s^{ISR}=0.1365$)
- Effect also observed with 8 TeV data.

CMS-PAS-TOP-12-041 (dilepton 8 TeV),
 CMS-PAS-TOP-16-011 (dilepton 13 TeV),
 CMS-PAS-TOP-16-008 (l+jets 13 TeV)



CMS-PAS-TOP-16-021

Tune α_s^{ISR} using 8 TeV ttbar Njets and jet pT data →

$$\alpha_s^{ISR} = 0.1108^{+0.0145}_{-0.0142}$$

$$h_{damp} = 1.581^{+0.658}_{-0.585} \times m_t$$

with `SpaceShowerRapidityOrdering=on` (special care of options needed for the emissions produced by the PS)

- ⇒ Significantly lower shower α_s cures the overshoot of CUETP8M1 at high jet multiplicities.
- ⇒ UE and min-bias are described better
- ⇒ POWHEG+PYTHIA8: generally consistent with data, with residual differences covered by theory uncertainties.

arXiv:1803.0399

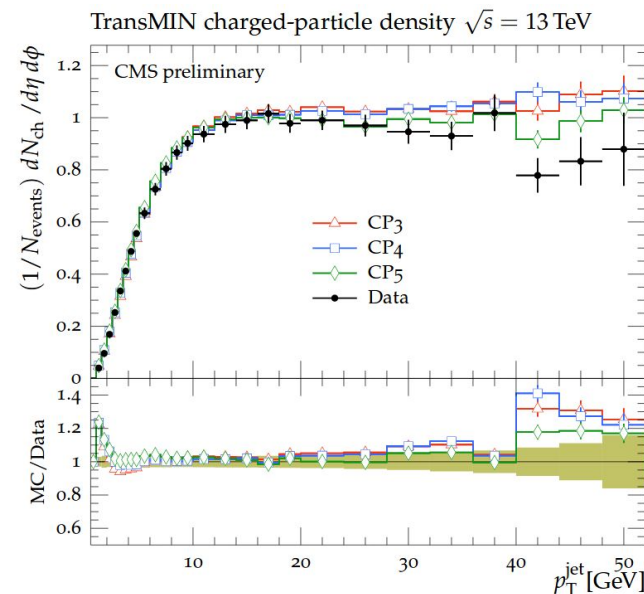
MC Tuning: CPX Tune

[arXiv:1903.12179](https://arxiv.org/abs/1903.12179)

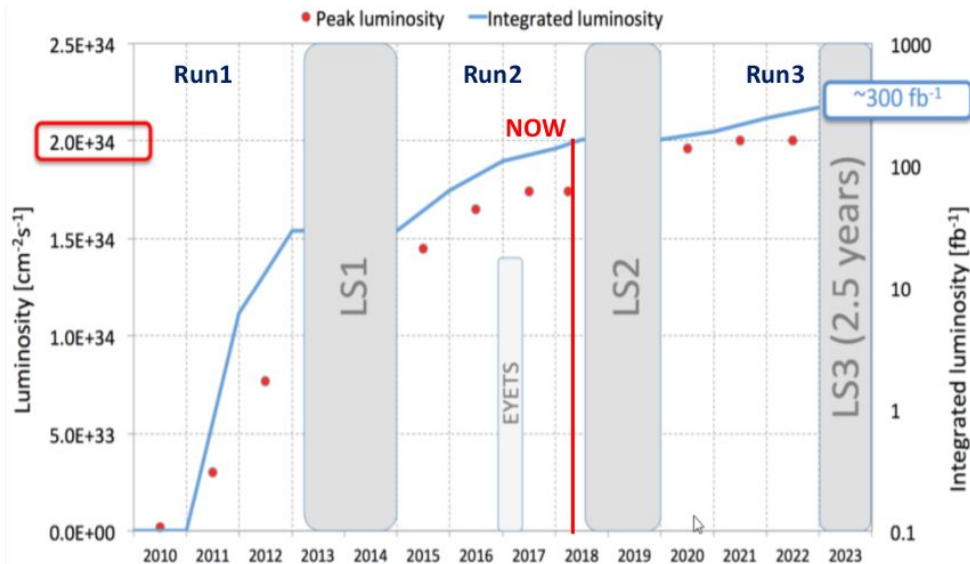
- **First CMS tune with 13 TeV LHC data**
- **Match PDF and α_s in the PS and in the ME.**
 - PYTHIA8 tunes are mostly based on LO PDFs.
 - Sherpa tunes are based on NNLO PDFs.
 - HERWIG7 provide tunes based on NLO PDFs (MPI still based on LO).
- **Test the effect of using different PDF orders of NNPDF sets in PYTHIA8**
- **CP5 with NNPDF3.1NNLO** is the default for 2017 and 2018 MC productions

- CP1: NNPDF3.1 LO ($\alpha_s = 0.130$)
- CP2: NNPDF3.1 LO ($\alpha_s = 0.130$)
- CP3: NNPDF3.1 NLO ($\alpha_s = 0.118$)
- CP4: NNPDF3.1 NLO ($\alpha_s = 0.118$)
- CP5: NNPDF3.1 NNLO ($\alpha_s = 0.118$)

PYTHIA8 parameter	CP1	CP2	
PDF Set	NNPDF3.1 LO	NNPDF3.1 LO	Fixed inputs
$\alpha_s(m_Z)$	0.130	0.130	
SpaceShower:rapidityOrder	off	off	
MultipartonInteractions:EcmRef [GeV]	7000	7000	
$\alpha_s^{\text{ISR}}(m_Z)$ value/order	0.1365/LO	0.130/LO	
$\alpha_s^{\text{FSR}}(m_Z)$ value/order	0.1365/LO	0.130/LO	
$\alpha_s^{\text{MPI}}(m_Z)$ value/order	0.130/LO	0.130/LO	
$\alpha_s^{\text{ME}}(m_Z)$ value/order	0.130/LO	0.130/LO	
MultipartonInteractions:pT0Ref [GeV]	2.4	2.3	Fitted parameters
MultipartonInteractions:ecmPow	0.15	0.14	
MultipartonInteractions:coreRadius	0.54	0.38	
MultipartonInteractions:coreFraction	0.68	0.33	
ColorReconnection:range	2.63	2.32	
χ^2/dof	0.89	0.54	



Beyond Run2/3 and Future



Need to “fight” against conflicting requirements:

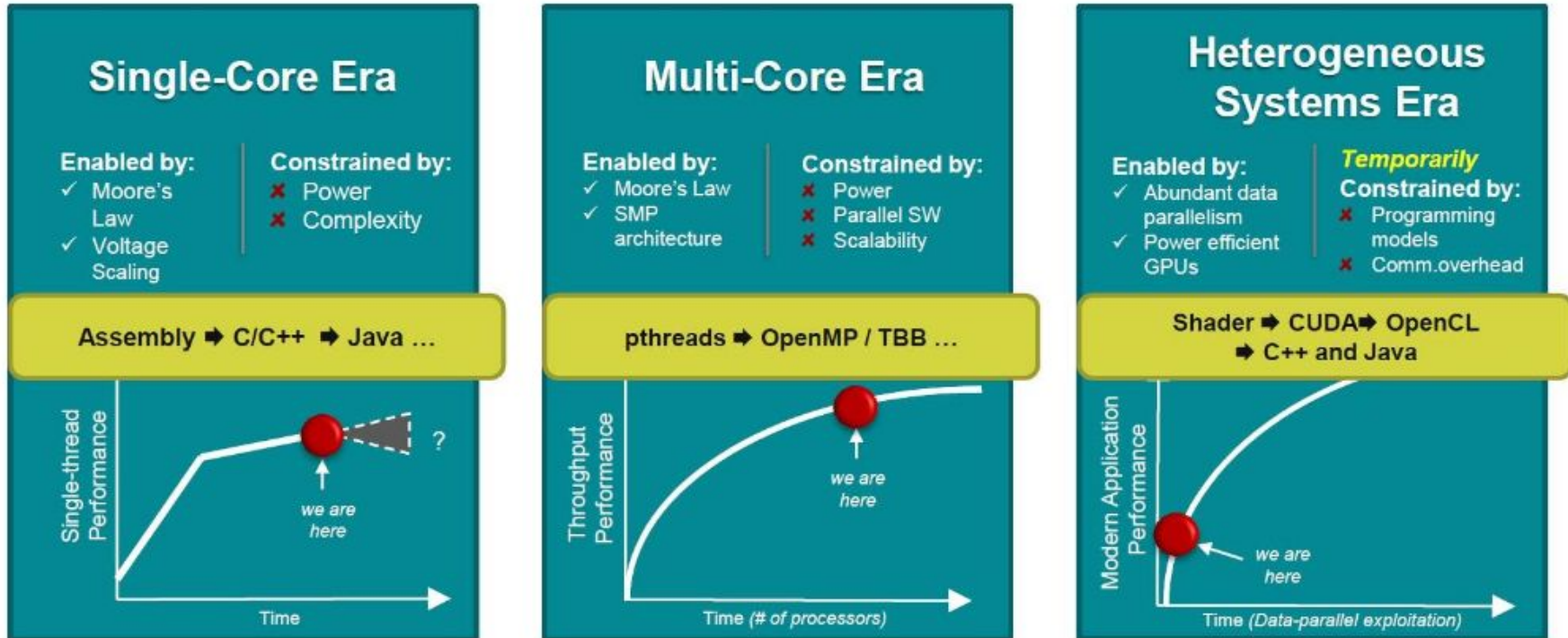
- (Much) larger datasets
- Increased measurement precision
- Need for alternative samples for systematics
- Flattening of computing resources (both cpu and disk space)

- σ_{var} Negative weights strongly reduce statistical power
- For weighted events w_i , effective events N_{eff} for fraction of negative weights f :

$$N_{eff} = \frac{(\sum_i w_i)^2}{\sum w_i^2} = N(1 - 2f)^2$$
- for 35% negative weights (common at for high jet-multiplicity/ high pt)

⇒ 9% effective events compared to $w_i = 1$

The Future : Heterogeneous Architecture



<https://opensourceforu.com/2016/12/how-heterogeneous-systems-evolved-and-the-challenges-going-forward/>

The computing available in 2026 will be heterogeneous and highly concurrent.
different types of compute units and interconnects

HSF and Generator Workshop

The HEP Software Foundation facilitates cooperation and [common efforts](#) in High Energy Physics software and computing internationally.

[Community White Paper](#): summarising R&D in a variety of technical areas for HEP Software and Computing



Physics Event Generator Computing Workshop

📅 26 Nov 2018, 09:00 → 28 Nov 2018, 18:00 Europe/Zurich

📍 4-3-006 - TH Conference Room (CERN)

Goals of this workshop:

- Identify the most crucial areas for **technical improvements to the generators** used by the experiments.
- Define a programme of work that can be used to **attract investment** in these technical areas, aiming to **have software engineers who can work together with the generator authors**.
- Identify ways of making new **theoretical advances** easier to implement in a **computationally efficient way**.

Thanks



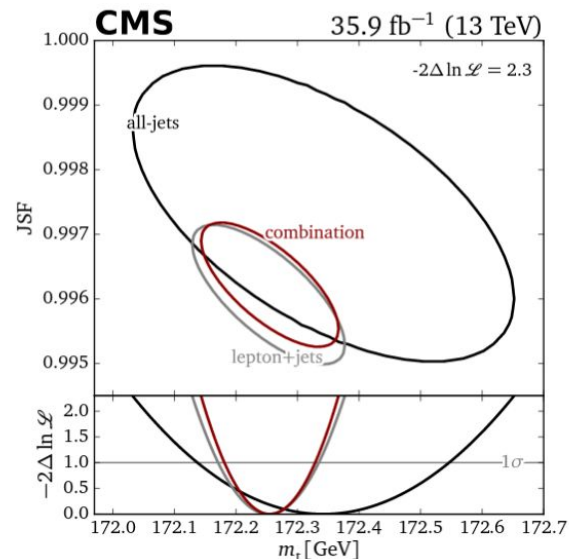
MC Modelling: b jet and Color-Connection ...

- The Bowler–Lund fragmentation function varied within uncertainties determined by the ALEPH and DELPHI
- Alternatively, the Peterson fragmentation function is used to derive additional uncertainty.
- Semileptonic b hadron branching fraction, varied by -0.45% and $+0.77\%$, motivated by measurements of B^0/B^+ decays and their corresponding uncertainties

	2D		1D	hybrid	
	δm_t^{2D} [GeV]	δJSF^{2D} [%]	δm_t^{1D} [GeV]	δm_t^{hyb} [GeV]	$\delta \text{JSF}^{\text{hyb}}$ [%]

Top quark mass from $t\bar{t}$ fully hadronic (2016)

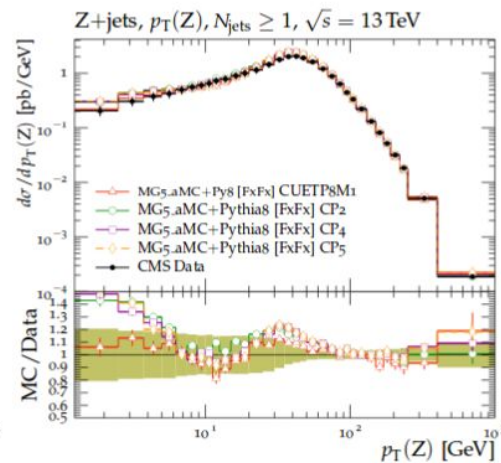
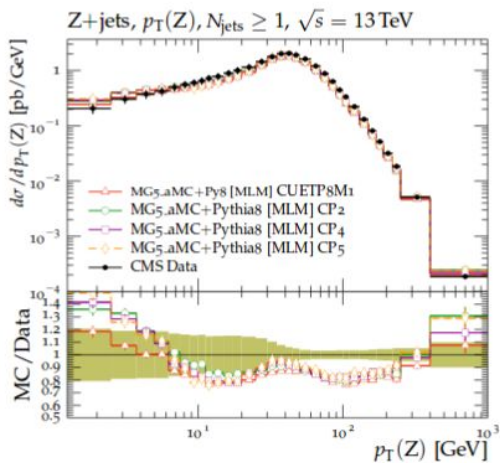
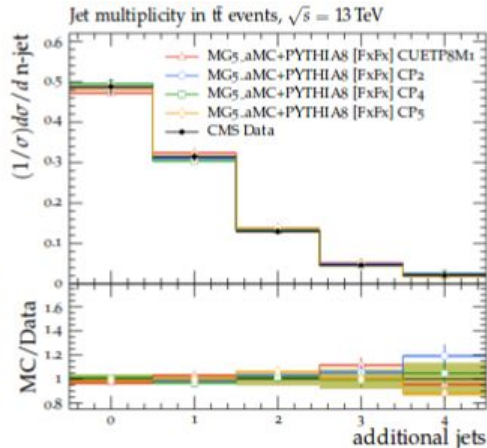
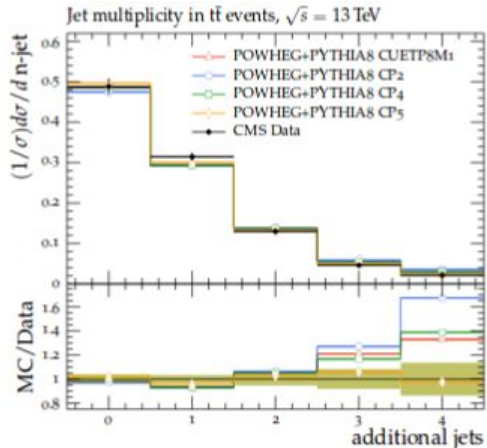
b jet modeling (quad. sum)	0.09	0.0	0.09	0.09	0.0
- b frag. Bowler–Lund	-0.07	0.0	-0.07	-0.07	0.0
- b frag. Peterson	-0.05	0.0	-0.04	-0.05	0.0
- semileptonic b hadron decays	-0.03	0.0	-0.03	-0.03	0.0
PDF	0.01	0.0	0.01	0.01	0.0
Ren. and fact. scales	0.05	0.0	0.04	0.04	0.0
ME/PS matching	$+0.32 \pm 0.20$	-0.3	-0.05 ± 0.14	$+0.24 \pm 0.18$	-0.2
ISR PS scale	$+0.17 \pm 0.17$	-0.2	$+0.13 \pm 0.12$	$+0.12 \pm 0.14$	-0.1
FSR PS scale	$+0.22 \pm 0.12$	-0.2	$+0.11 \pm 0.08$	$+0.18 \pm 0.11$	-0.1
Top quark p_T	+0.03	0.0	+0.02	+0.03	0.0
Underlying event	$+0.16 \pm 0.19$	-0.3	-0.07 ± 0.14	$+0.10 \pm 0.17$	-0.2
Early resonance decays	$+0.02 \pm 0.28$	+0.4	$+0.38 \pm 0.19$	$+0.13 \pm 0.24$	+0.3
CR modeling (max. shift)	$+0.41 \pm 0.29$	-0.4	-0.43 ± 0.20	-0.36 ± 0.25	-0.3
- "gluon move" (ERD on)	$+0.41 \pm 0.29$	-0.4	$+0.10 \pm 0.20$	$+0.32 \pm 0.25$	-0.3
- "QCD inspired" (ERD on)	-0.32 ± 0.29	-0.1	-0.43 ± 0.20	-0.36 ± 0.25	-0.1
Total systematic	0.81	0.9	1.03	0.70	0.7
Statistical (expected)	0.21	0.2	0.16	0.20	0.1
Total (expected)	0.83	0.9	1.04	0.72	0.7



$$m_t^{\text{hyb}} = 172.26 \pm 0.07 (\text{stat+JSF}) \pm 0.61 (\text{syst}) \text{ GeV}$$

MC Tuning: CPX Tune

[arXiv:1903.12179](https://arxiv.org/abs/1903.12179)



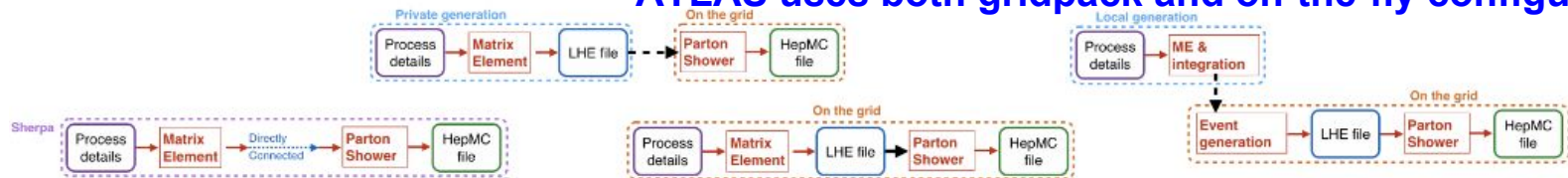
Predictions from the **new Tunes** based on higher-order PDF sets, **interfacing with higher-order and multileg matrix element generators**, such as POWHEG and MG5_aMC, are shown to give a reliable description of observables measured in multijet final states, Drell–Yan, and top quark production processes

ATLAS Generator Usage

Josh McFayden

- ▶ Various possible configurations result in many different running modes
- ▶ Also requires flexibility in the software integration and production system configuration.

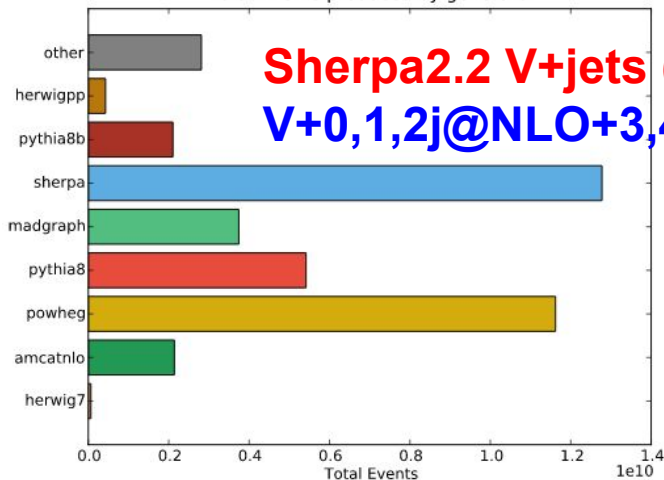
ATLAS uses both gridpack and on-the-fly configurations



4

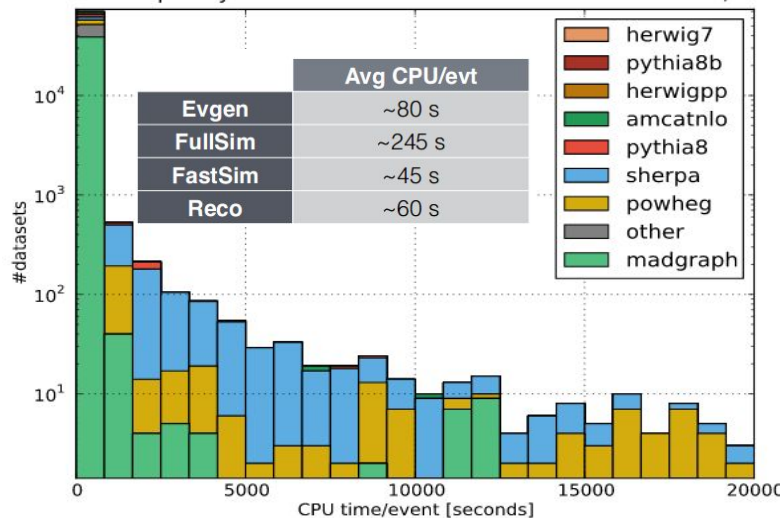
Josh McFayden | Gen HSF WS | 26/11/2018

Total Events produced by generator

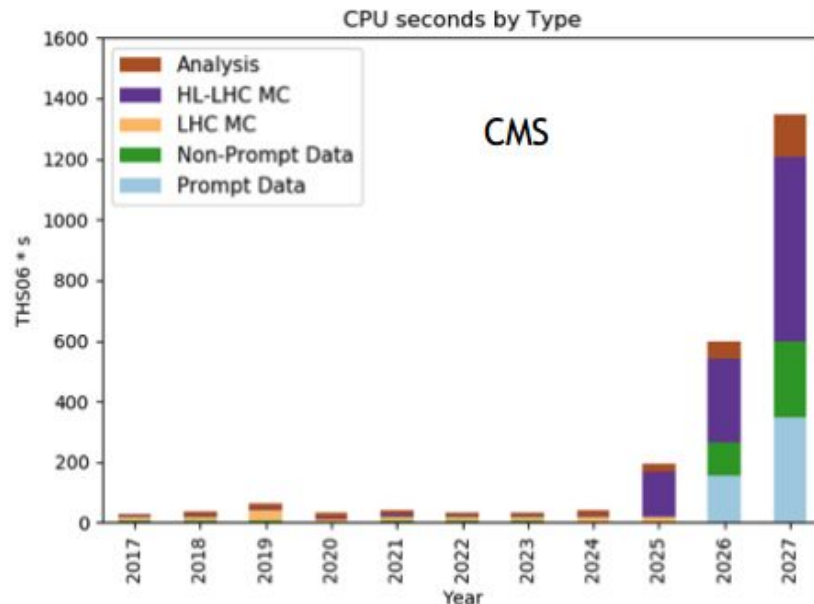
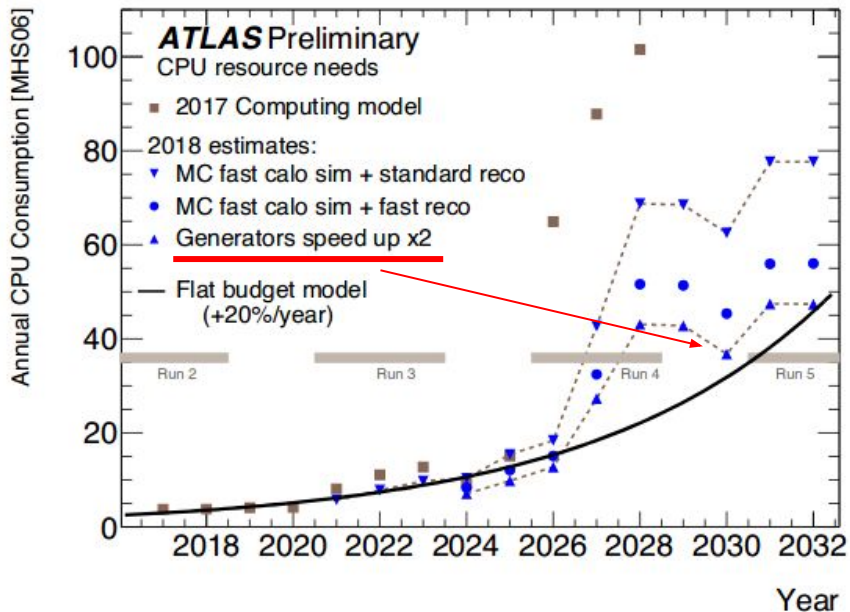


Sherpa2.2 V+jets (3.2B)
V+0,1,2j@NLO+3,4j@LO

CPU time/event for 2015 MC event generation at $\sqrt{s} = 13$ TeV
 (All physics processes included, correlations with process complexity and filter efficiencies not taken into account)



Motivation for Generator upgrade

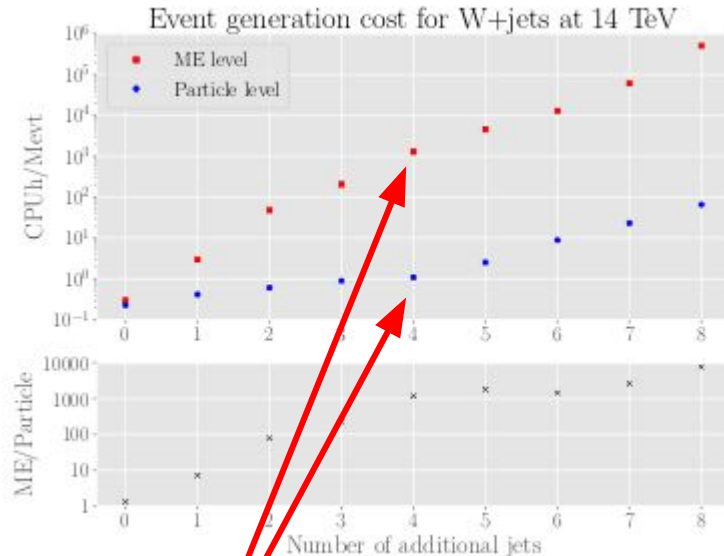
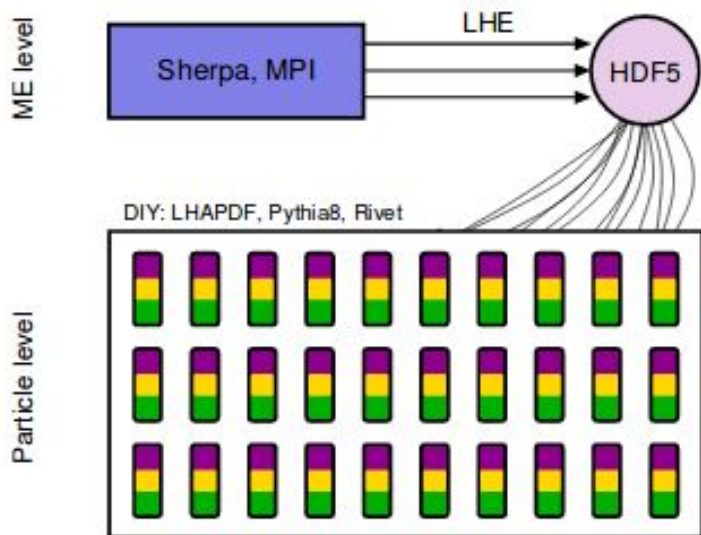


- We are OFF by ~5x on CPU power when considering Moore's law
- HL-LHC salvation will come from software improvements, not from hardware

Elizabeth Sexton-Kennedy

High-multiplicity multi-jet merging with HPC technology **W+up to 8 Jets**

1. **HDF5** (Hierarchical Data Format) storage for ME events: The CPU expensive part of the simulation is stored in a parton-shower independent format.
2. Particle level and merging with Pythia with ASCR's (Advanced Scientific Computing Research) **DIY**, which does all the low-level MPI communication. **Particle level run-time up to 4 orders of magnitude faster than ME.**



For comparison, [CMS WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8](#) (W+0/1/2/3/4Jets) time/evnet~16s, then **1M events->4000hr (ME+PS)**