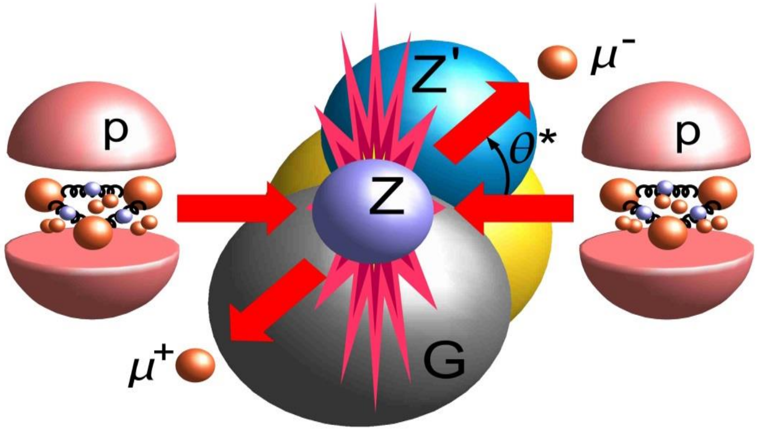
MC Modelling and Production at CMS

Spring-Summer, 2019

Qiang Li (Peking University)

CMS MC Simulation Overview



– **Hard process/Matrix Element generation**:
 Desired process up to parton level
 using perturbative QCD

– **Parton Shower/Hadronization**:
 QCD and QED emissions down to a low scale,
 and produces hadrons from QCD partons

– *Multiple Parton Interaction*

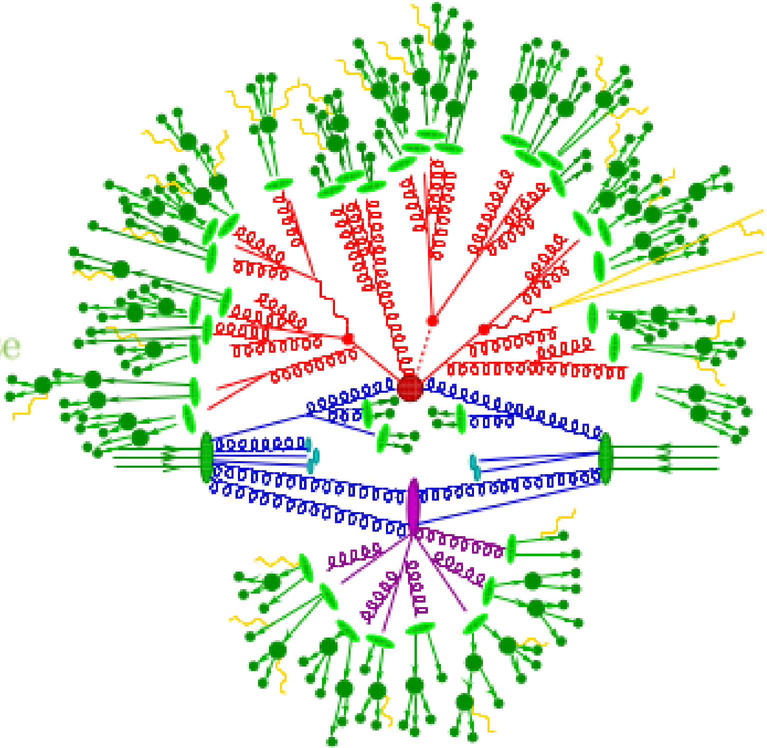
– *Detector Simulation and Digitization*:
 Detailed Geant4 simulation of the interactions of the outgoing particles with the CMS detector, followed by simulation of detector electronics and creation of simulated raw data

– *Reconstruction*:
 Reconstruction of simulated raw data into higher level physics objects
 To a good approximation, identical code as runs on real data

IN CMS Jargon: LHE –> GEN–> SIM–> DIGI–> RECO

MC Generator

- Hard interaction
- Particle decays
- Final state radiation
- Initial state radiation
- Underlying event
- Final-state partons hadronise
- Hadrons decay
- Photon radiation
- Beam remnants



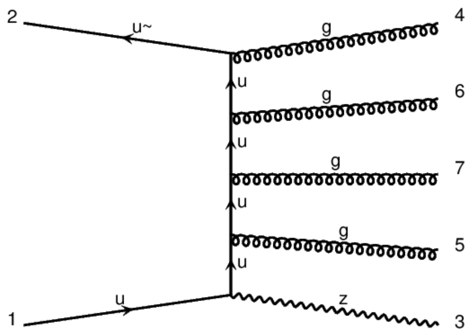
LHC is a QCD factory:

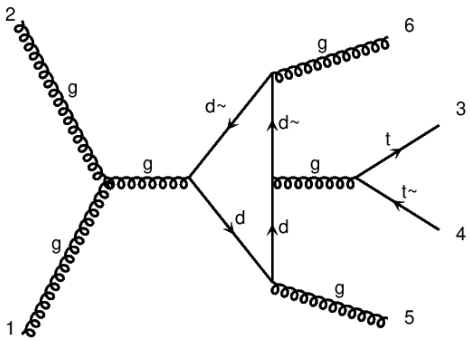
NLO and/or multi-leg/merged-multiplicity

Factorised approach may lead to large uncertainties:

“tune” and PS weights etc needed

State-of-the-art MC Simulation in CMS





- **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**

    Most commonly used in CMS: MG5_aMC+Pythia8 with FxFx merging

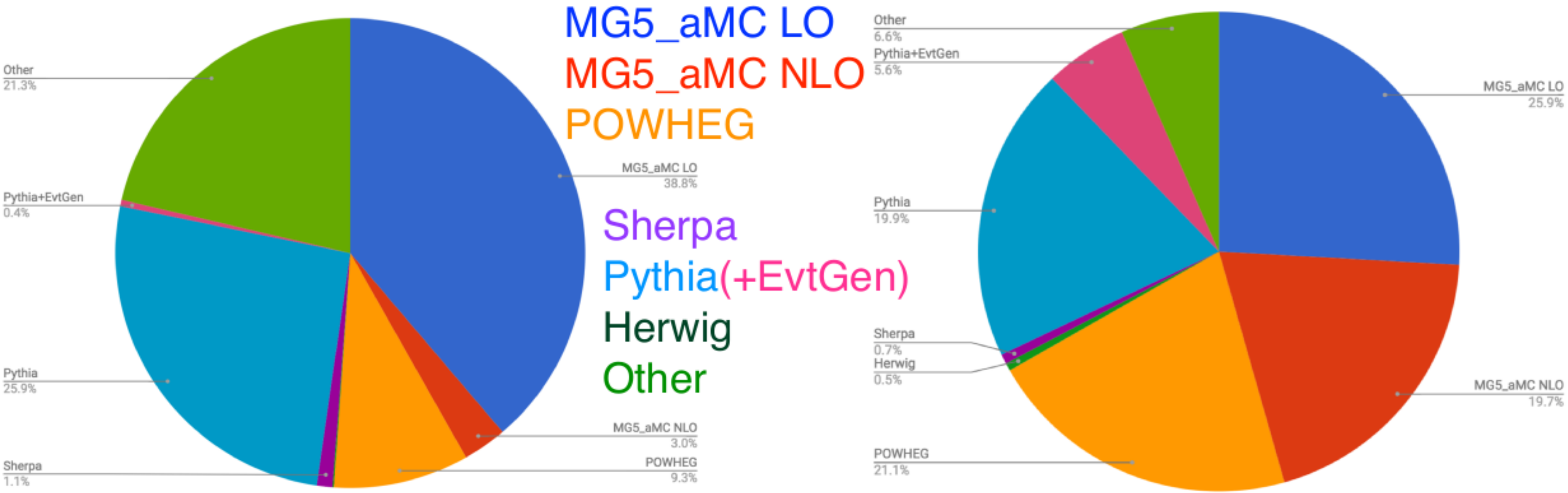
    Most complex process up to 2 additional jets at NLO.
- **For signal, NNLO+PS**

  POWHEG: MINLO-NNLOPS

  CMS HWW reweight the nominal signal to this one

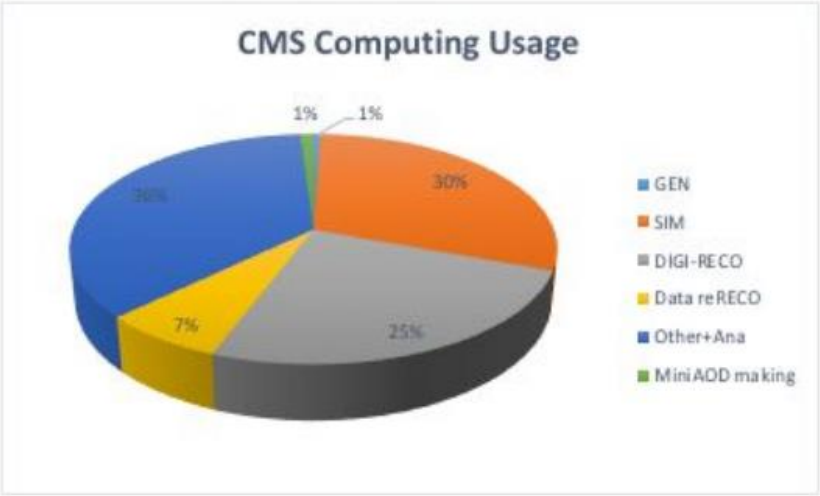
Generator Usage in CMS

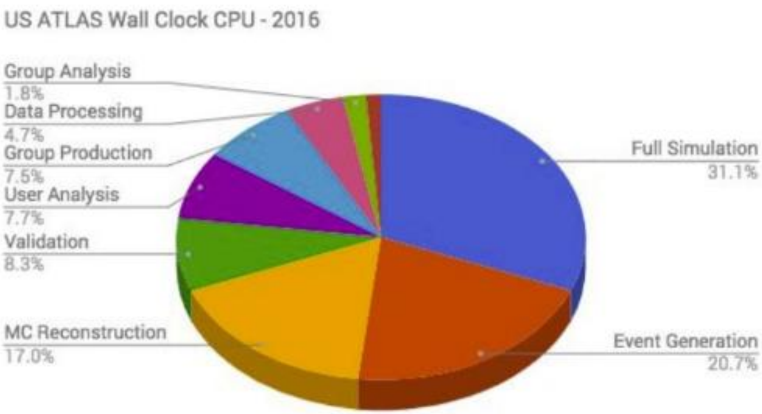
ME generator — by samples | ME generator — by events



Approximate and based on 2016 MC campaign

Generators Can be Computationally Intensive





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

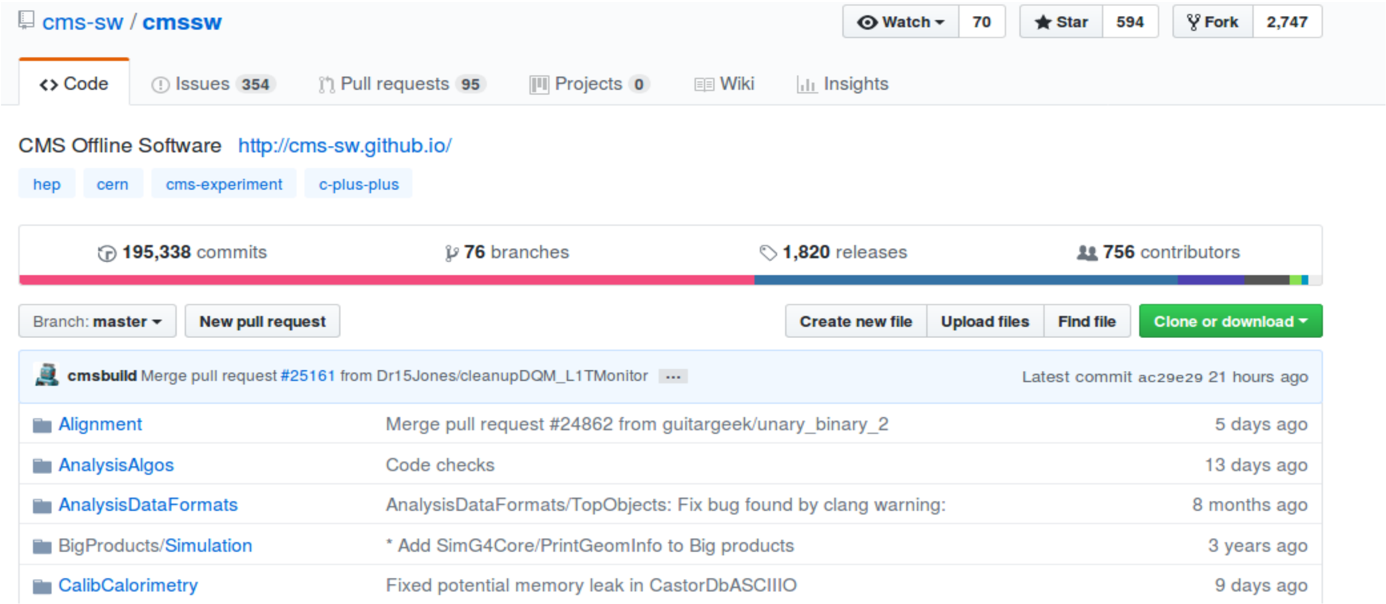
Elizabeth Sexton-Kennedy

- 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, see recent summary talk.

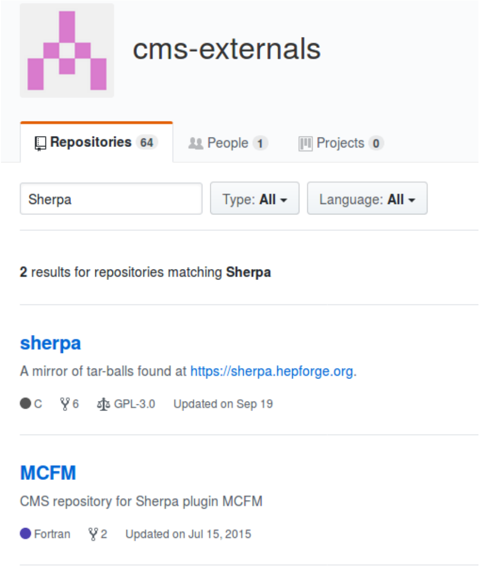
CMS Software

- **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**
  – Storing run-, lumi-section-(23s periods for real data), or event-level data products



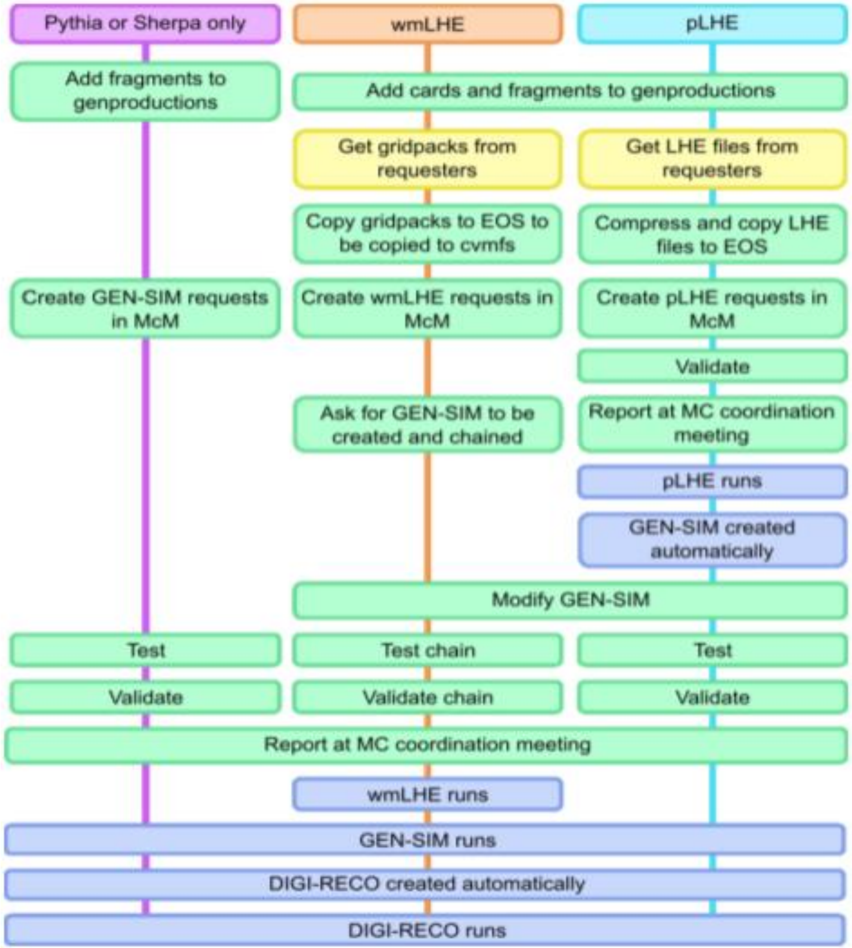
CMS Software

- **CMSSW links directly to many externals**
  - Either an indirect dependency or is directly called from within CMSSW
  - ***Compiled with the same common libraries***, as CMSSW and packaged together with a given release
  - From either a tarball from the author's website, GENSER, or a cms-managed github mirror



CMSSW software and externals are made available on worker nodes through CVMFS – a network file system based on HTTP

CMS MC Workflow



Basic paradigm for event generation

– C++ module making calls to linked external generator code to produce HepMC::GenEvent to be stored as EDM

Matrix element generators which generate LHE files (Madgraph, POWHEG, ...)

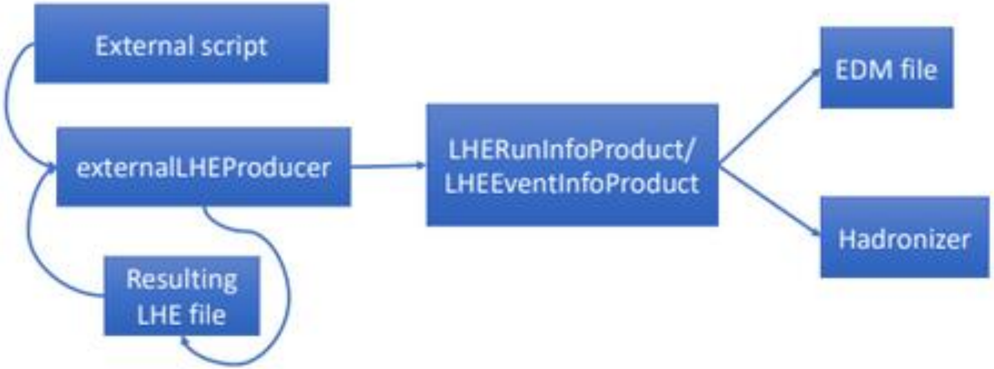
– Loosely coupled to CMSSW, calling of external generation script handled by an integrated CMSSW module **"externalLHEProducer"**

ascii LHE files are transient and immediately packed into binary compressed format

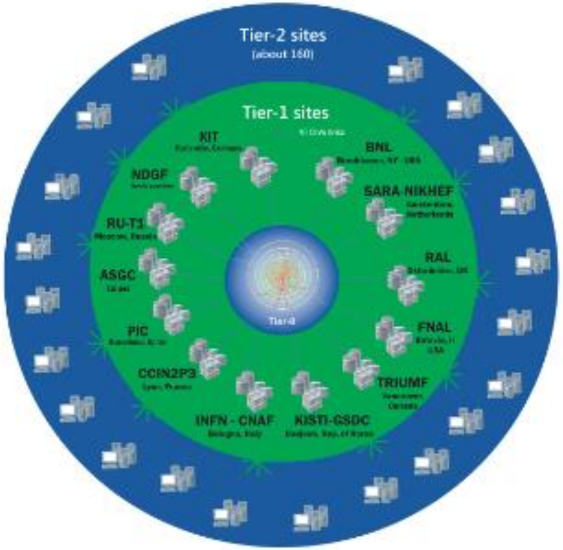
ExternalLHEProducer and Gridpack

- **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**



Worldwide LHC Computing Grid (WLCG)



- WLCG is composed of four levels, or “Tiers”, called 0, 1, 2/3.
- Each tier is made up of several computer centres and provides a specific set of services → tiers process, store and analyse data from LHC
- Tier 0 is the CERN Data Centre provides less than 20% of the Grid’s total capacity, 40% at T1s, and 40% at T2s

CVMFS

- **CernVM File System (CernVM-FS)** is a network file system based on HTTP and optimized to deliver experiment software in a fast, scalable, and reliable way
- Used extensively by High Energy Physics Community
- Developed at CERN (CERN Virtual Machine project)
- Optimized specifically for deploying software
- Read-only filesystem - immutable data model
- Content-addressable files
- HTTPS + FUSE mount

Gridpack Production

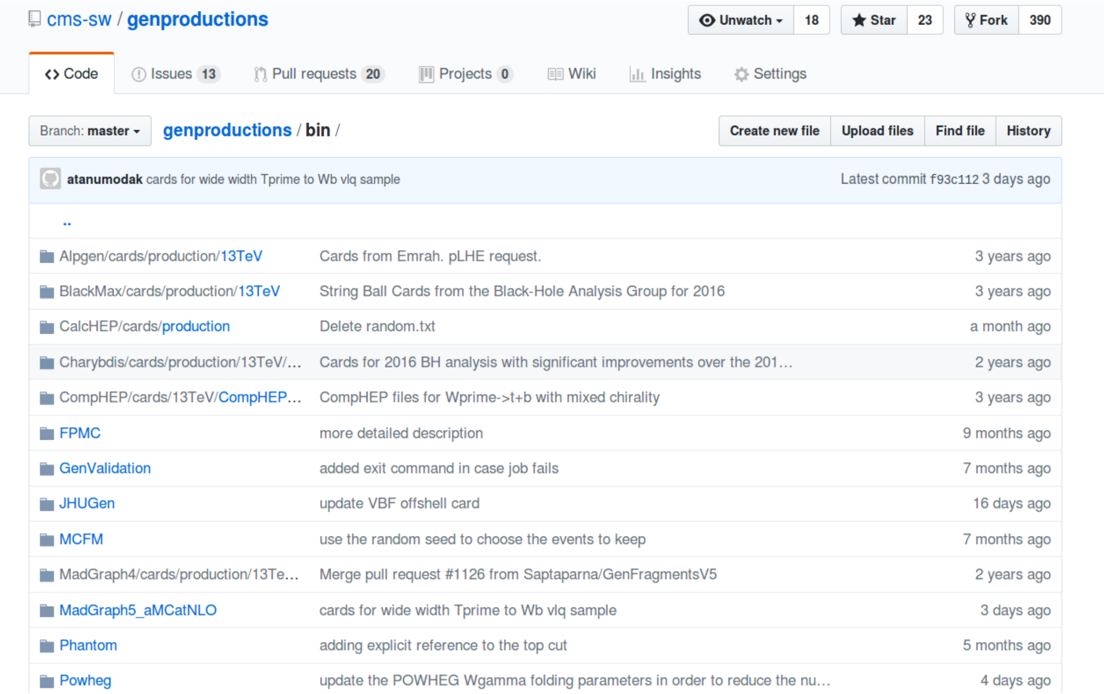
Gridpack (LHE) –> GEN–> SIM–> DIGI–> RECO

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



Pythia8 Showering and Hadronization

We have provided detailed example fragments on GEN twiki.
1) simple LO; 2) MLM matching; 3) simple NLO; 4) FxFx Merging

MadGraph

1) LO with no matching/merging (single jet multiplicity, ickkw=0 in run card)

No special settings are needed in pythia, and a fully generic hadronizer can be used as for TuneCUETP8M1 or TuneCP5.

2) LO with jet matching (multiple jet multiplicities, ickkw=1 in run card)

For kT-MLM matching (the only one extensively used in CMS so far), example fragment is here for TuneCUETP8M1 or TuneCP5. qCut, nQmatch, nJetMax will need to be modified on a per process basis as described in the comments within the fragment.

3) NLO with no matching/merging (single jet multiplicity, ickkw=0 in run card)

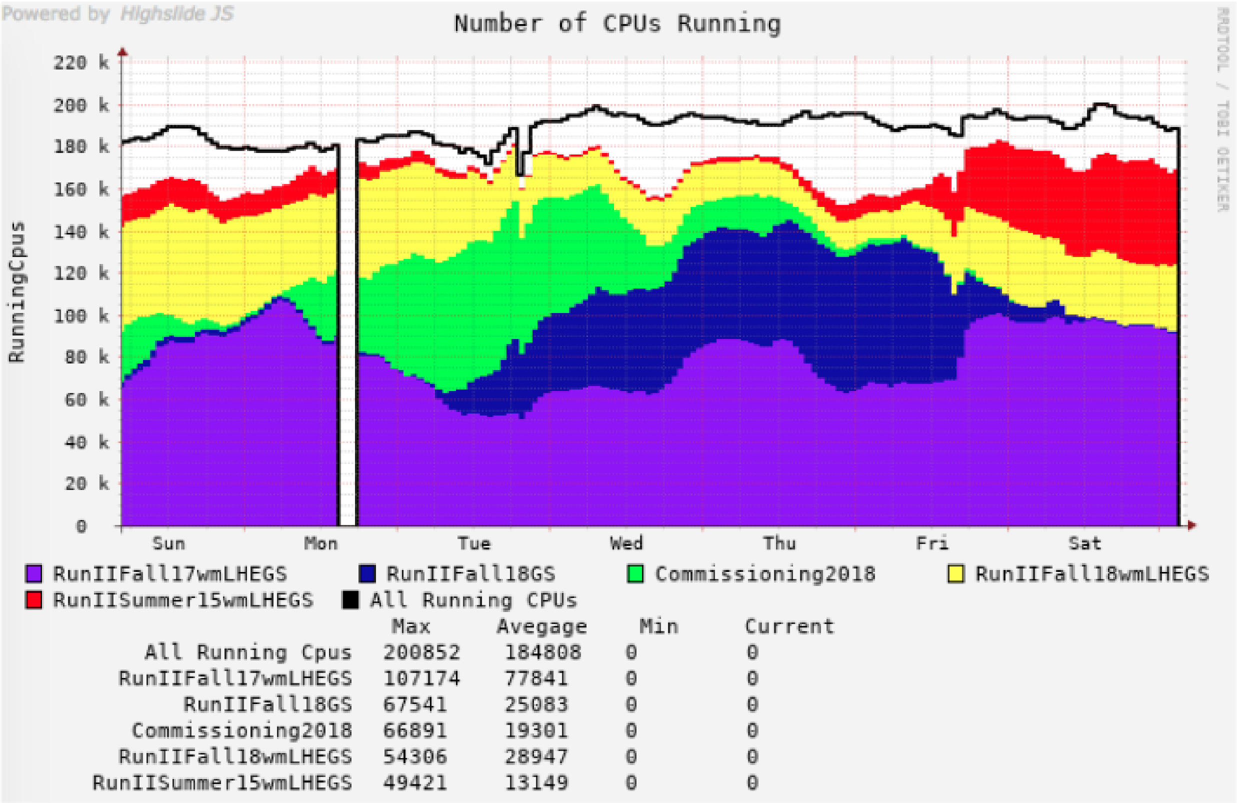
Dedicated hadronizer fragment is needed with special shower settings for aMC@NLO. The setting TimeShower::nPartonsInBorn must be set by hand according to the particular process. An example is here or here (Note for a process with no jets in the born matrix element, set in addition "processParameters = cms.vstring('TimeShower:nPartonsInBorn = 0')", please refer to e.g. SMP-RunIISummer15GS-00061)

4) NLO with jet matching (multiple jet multiplicities, ickkw=3 in run card)

Currently only FXFX merging is supported for this case. An example config is here. The parameters qCut, qCutME, nQmatch, nJetMax will need to modified according to the particular process as described by the comments in the fragment. (note that TimeShower:nPartonsInBorn should NOT be set in this case, as it is now set automatically as of pythia 8.205 when doing FXFX merging)

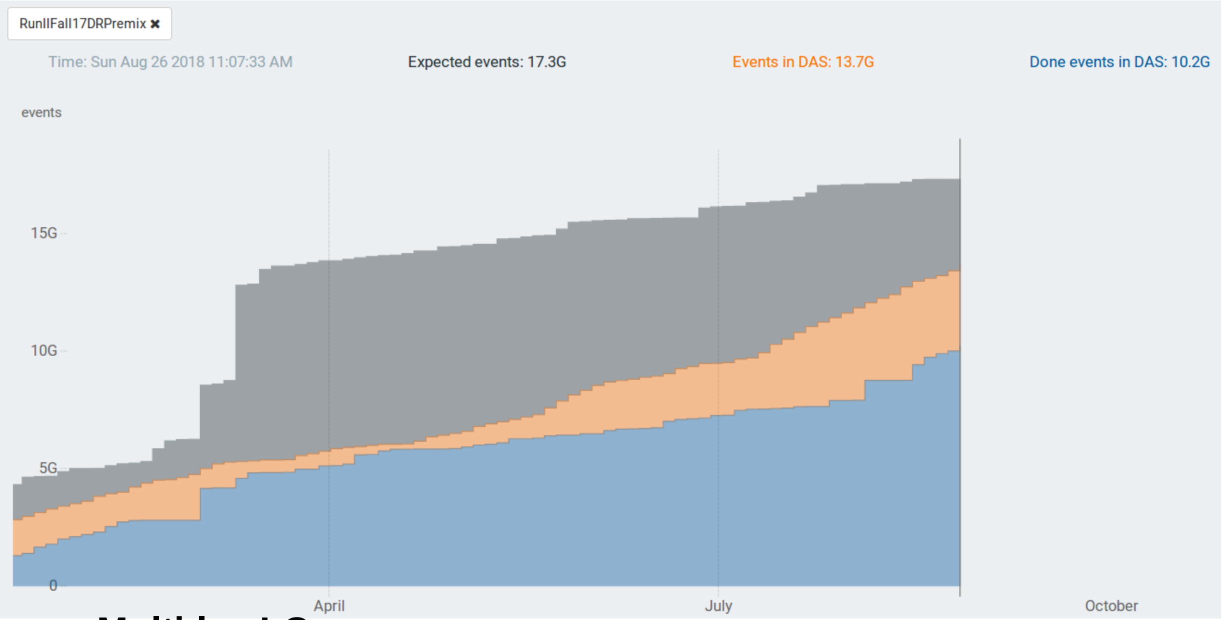
POWHEG has different settings, check e.g. this example

CMS MC Production Status



- Major Campaigns sharing the computing resources

CMS MC Production Status



Multi-leg LO

- up to ~10s/gen-evt
- ~50% matching efficiency -> 20s/full- sim-evt

Multi-leg NLO

- up to ~30s/gen-evt
- ~30% matching efficiency -> 100s/full- sim-evt

Taking 2017 as an example:

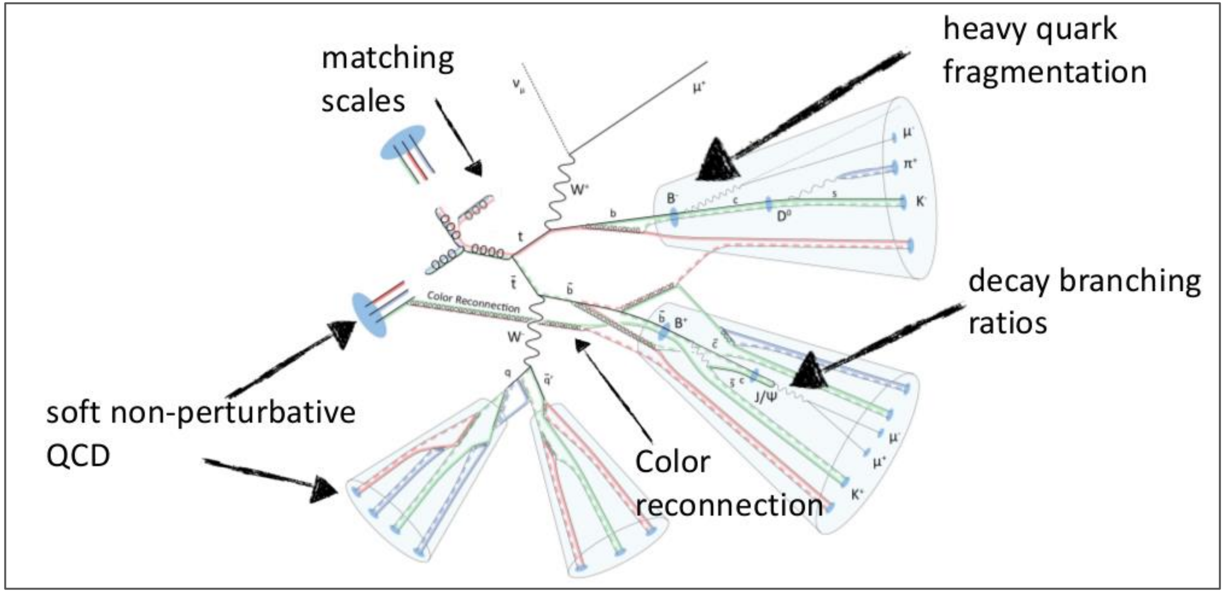
15 B (+ some other production ~ 20 B) in 8 months

- **GEN-SIM-DIGI-RECO** ~85 sec/evt
- **60k cores** (~1/3 of the CMS production power)

- Large fraction of negative weights of up to ~40% -> larger samples!

Showering and Hadronization

- Pythia 8.226 default; Herwig++ replaced by Herwig7.
- Fragment settings different depent 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



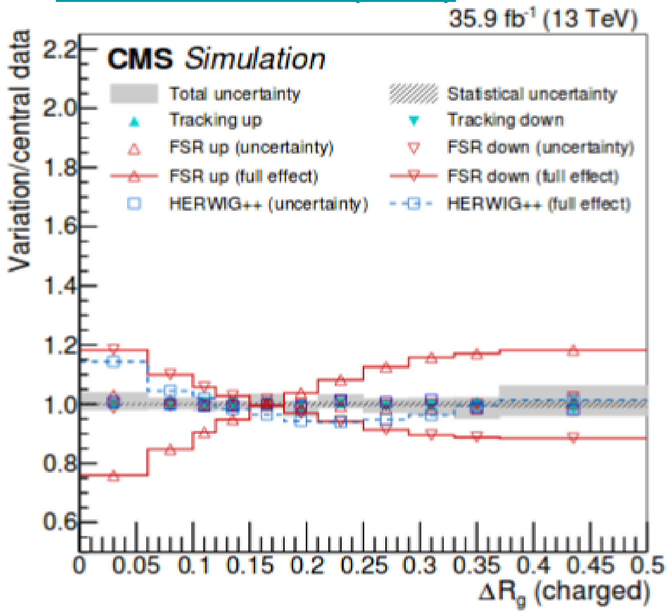
Showering and Hadronization

- Parton Shower Weights, Tune Variations etc included

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

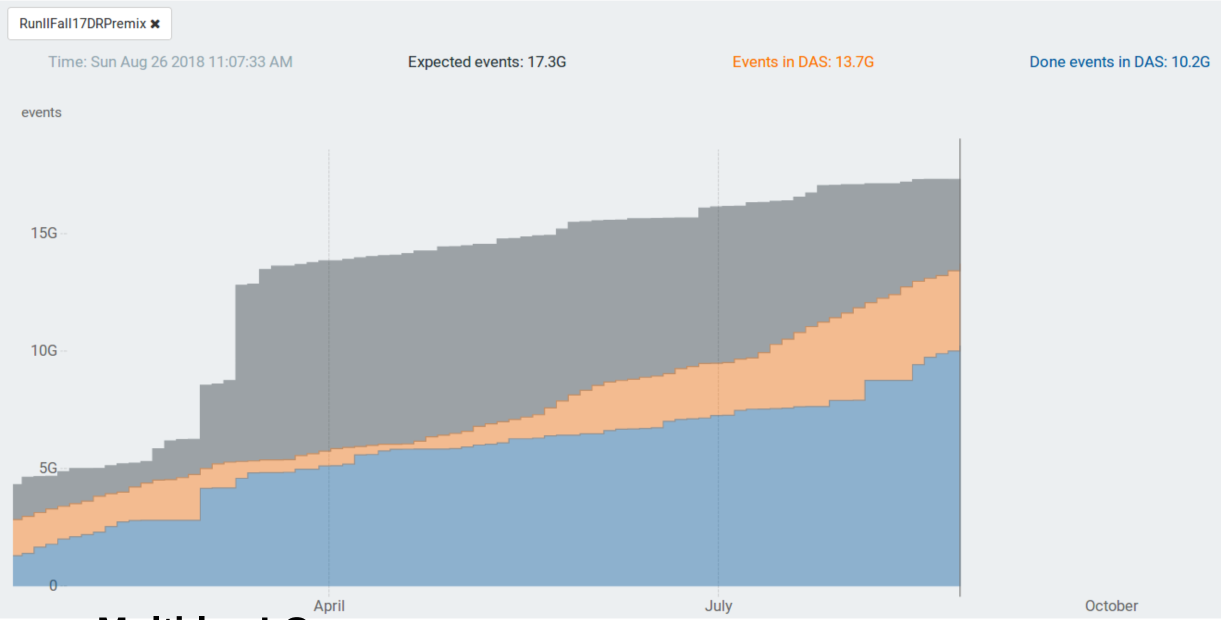
Source	Handle	Weights	Variation	Note/Reference	Dedicated studies
Shower scales	ISR scale (SpaceShower:renormMultFac) FSR scale (TimeShower:renormMultFac)	YES	0.5-2.0 0.5-2.0	FSR variations can be scaled down by √2 from LEP	TOP-15-011, TOP-16-021 TOP-17-13, TOP-17-015, ...
ME-PS Matching	hdamp	No	hdamp=1.58m_t +0.66-0.59 m_t	see TOP-16-021	Starting scale variations for MG5_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-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_b to match PDG)	

CMS: Jet substructure in ttbar events at 13 TeV (2016)



angle between the groomed subjets

CMS MC Production



Multi-leg LO

- up to ~10s/gen-evt
- ~50% matching efficiency -> 20s/full- sim-evt

Multi-leg NLO

- up to ~30s/gen-evt
- ~30% matching efficiency -> 100s/full- sim-evt

Taking 2017 as an example:

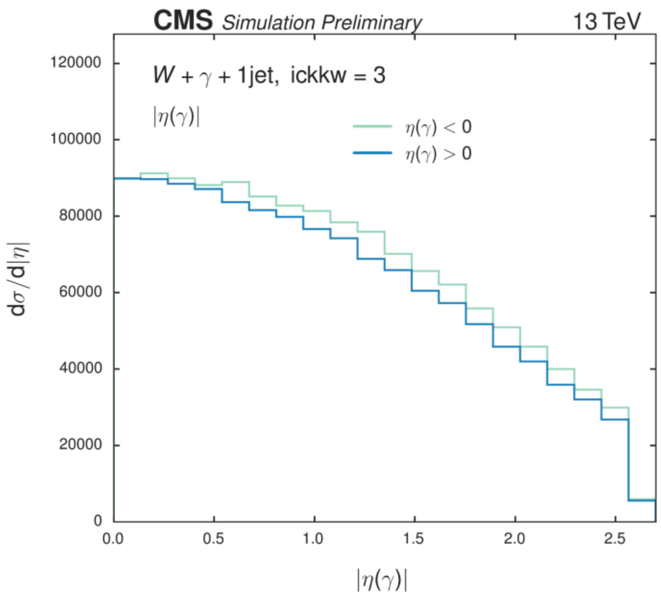
15 B (+ some other production ~ 20 B) in 8 months

- **GEN-SIM-DIGI-RECO** ~85 sec/evt
- **60k cores** (~1/3 of the CMS production power)

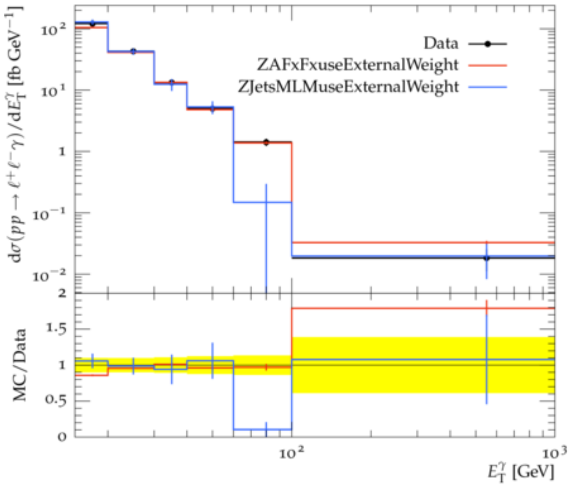
- Large fraction of negative weights of up to ~40% -> larger samples!

MC Validation

- Automated validation workflow
- Validation across different Generators, Generator Versions and CMSSW releases
- Using Rivet to check with data
- Optimizing some parameters: pdfwgt, hdamp, …..

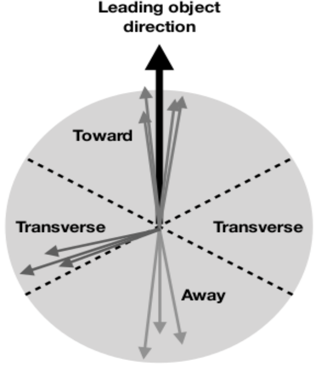


https://answers.launchpad.net/mg5amcnlo/+question/631090
WGamma FxFx bug (photon eta assymetry) reported by CMS and acknowledged by the MG authors.   Fixed since MG version 2.5.5

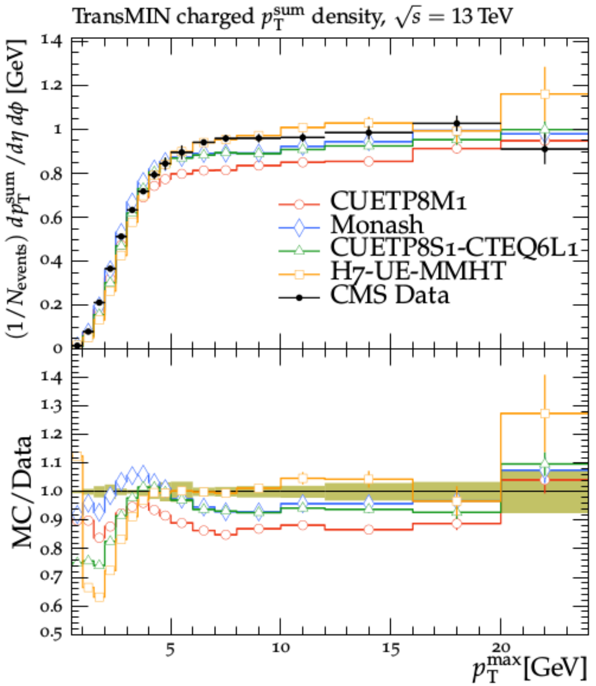


ZGamma vs ZJets after photon selection with Rivet from 7TeV ATLAS_2013_I1217863_Z

MC Tuning: CUETP8M1 Tune



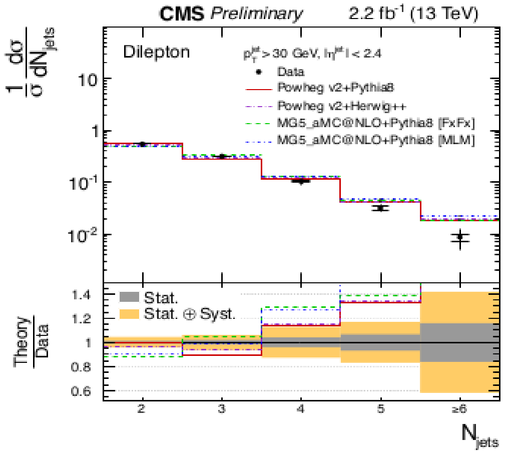
Until 2017 analyses (except 2016 ttbar), **Pythia8 CUETP8M1 tune** [EPJC 76 (2016) 155] based on the **Monash** tune was used. Fitting MPI energy dependence parameters to UE data @sqrt(s) = 0.9, 1.96 & 7 TeV



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

<u>CUETP8M1 does not describe well the central values of 13TeV data</u>

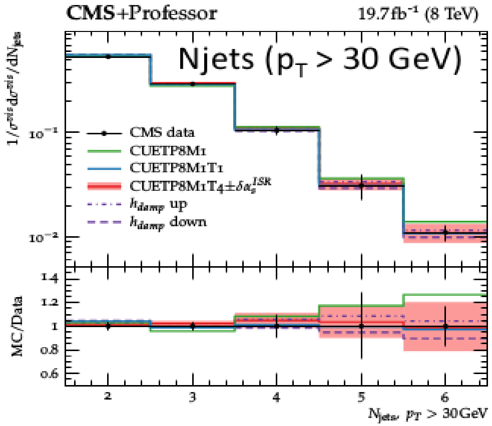
MC Tuning: CUETP8M2T4 Tune [TOP-16-021]



- Predictions overshoot the data for large jet multiplicities when out of the box parameters are used (in Monash-based tunes: α_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)*

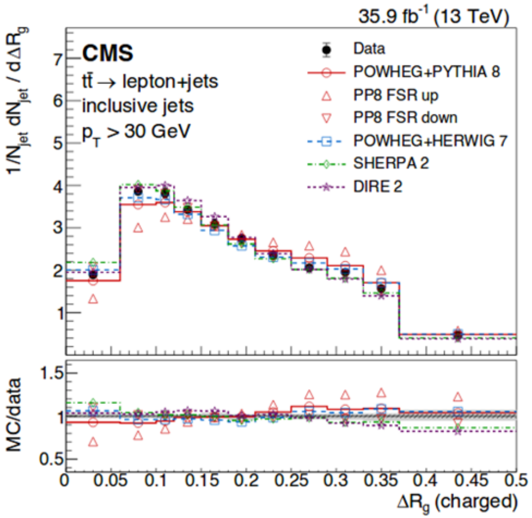
==> Significantly lower shower α_s cures the overshoot of CUETP8M1 at high jet multiplicities.
==> UE and min-bias are described better
==> POWHEG+PYTHIA8: generally consistent with data, with residual differences covered by theory uncertainties. arXiv:1803.0399

MC Tuning: jet substructure

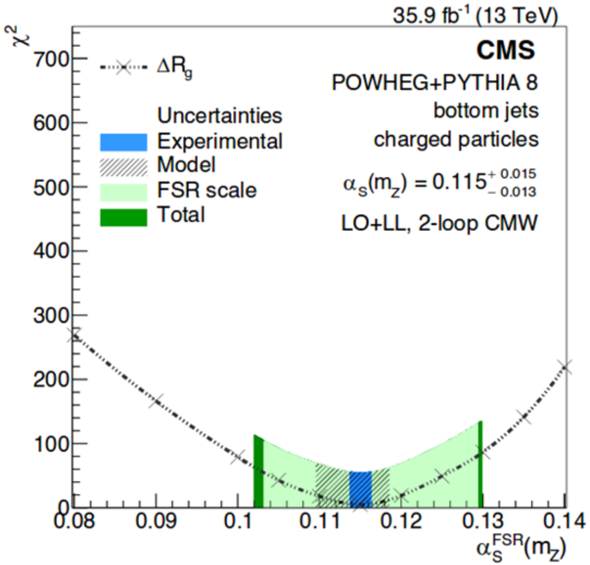
α_s^{FSR} from jet substructure in ttbar l+jets events

- Measured using charged+neutral and with only charged jet constituents (particle pT > 1 GeV).
- b, light, or gluon jet enriched samples.

PRD.98.092014



Angle between groomed subjets
at particle level (correlated to jet width)



Angle between groomed subjets
at particle level.

Pythia8:
CUETP8M2T4 for ttbar
CUETP8M1 for the rest.

None of the default tunes yield a good overall description of the data.

Powheg+Pythia8 with LO+LL, 2-loop :

$$\alpha_s^{FSR}(M_Z) = 0.115^{+0.015}_{-0.013}$$

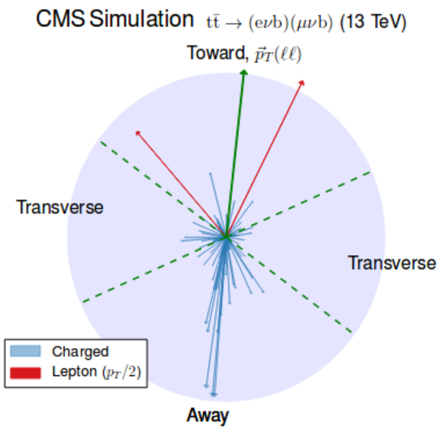
Efe Yazgan DIS2019

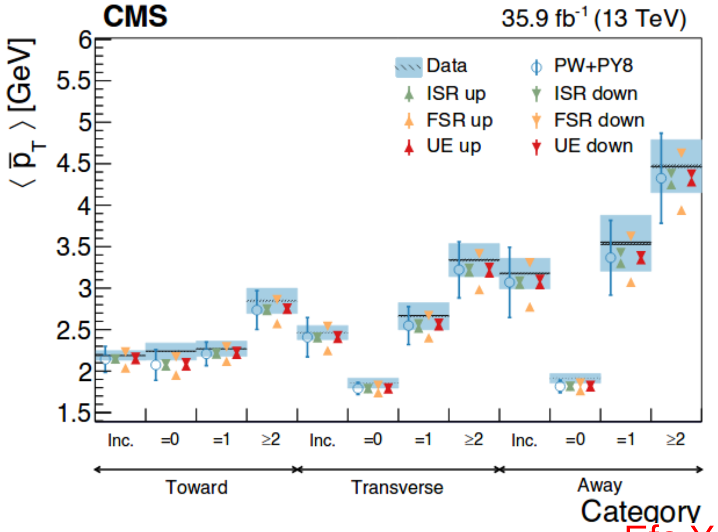
MC Tuning: UE in TTbar

α_s^{FSR} from Underlying Event in ttbar

EPJC 79 (2019) 123

- Measurement of the UE for the first time at a scale of > $2m_t$.
- > 200 distributions investigated in different categories to enhance sensitivity to the modeling of MPI, color reconnection, $\alpha_s^{FSR}(M_Z)$ in Pythia8.
- Measurement unfolded to particle level.
- Good agreement of POWHEG+PYTHIA8 with CUETP8M2T4 in UE event regions.





MC Tuning: CPX Tune

arXiv:1903.12179

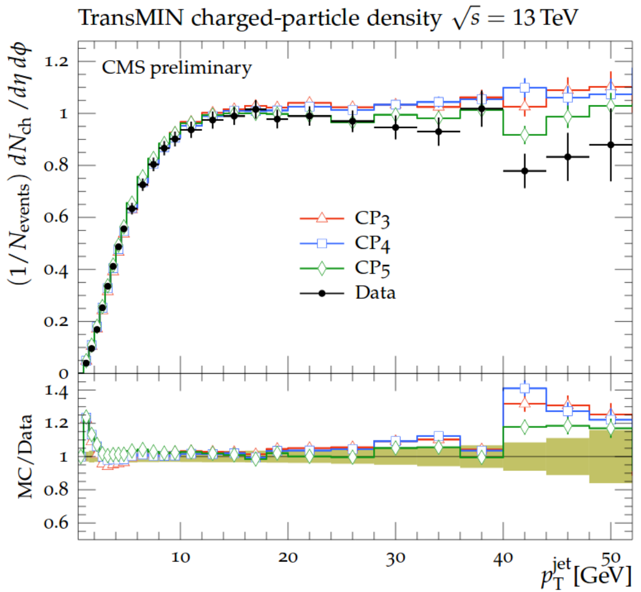
- **First CMS tune with 13 TeV LHC data**
- **Match PDF and αs in the PS and in the ME.**
  - PYTHIA tunes are mostly based on LO PDFs.
  - Sherpa tunes are based on NNLO PDFs.
  - HERWIG7 provide tunes based on NLO PDFs (MPI still based on LO).
- **Test the effect of using different PDF orders of NNPDF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

However, NNPDF3.1 NNLO is not positive definitely

For high mass tail, CP2 is the alternative.

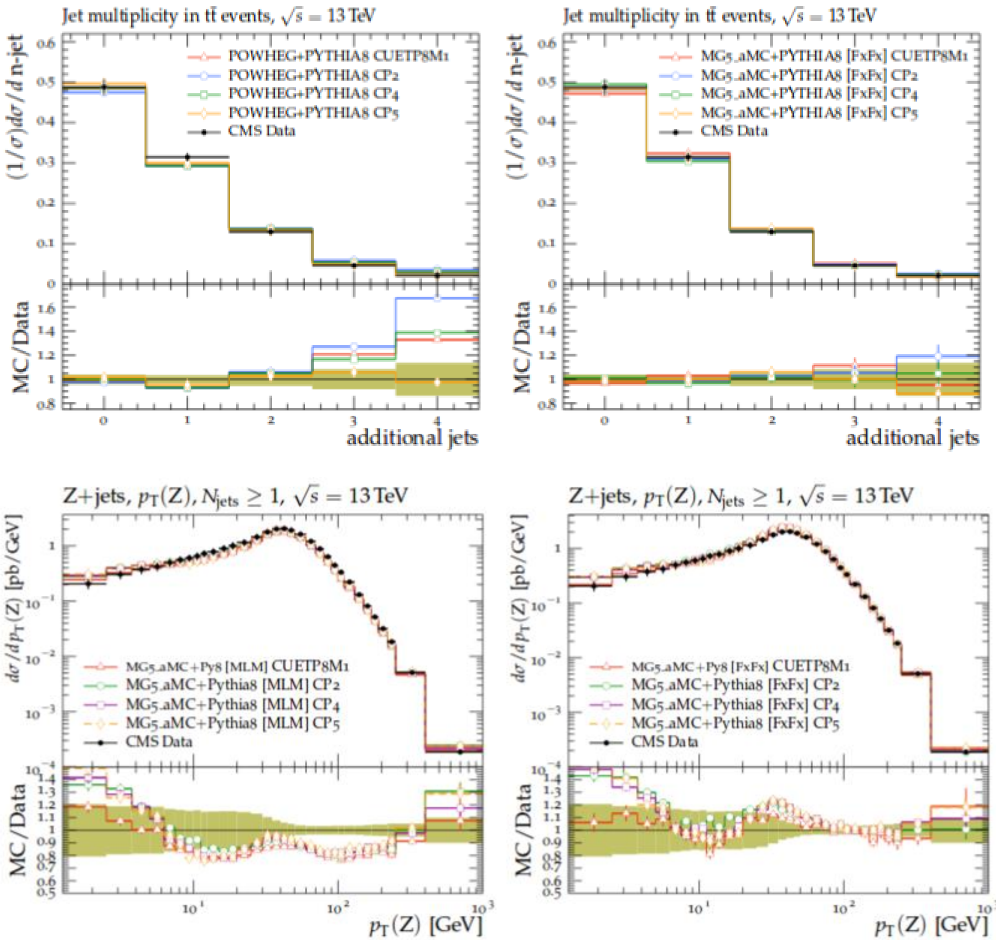
- CP1: NNPDF3.1 LO (α s =0.130)
- CP2: NNPDF3.1 LO (α s =0.130)
- CP3: NNPDF3.1 NLO (α s =0.118)
- CP4: NNPDF3.1 NLO (α s =0.118)
- CP5: NNPDF3.1 NNLO (α s =0.118)

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	
$\alpha_S^{ISR}(m_Z)$ value/order	0.1365/LO	0.130/LO	
$\alpha_S^{FSR}(m_Z)$ value/order	0.1365/LO	0.130/LO	
$\alpha_S^{MPI}(m_Z)$ value/order	0.130/LO	0.130/LO	
$\alpha_S^{ME}(m_Z)$ 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	



MC Tuning: CPX Tune

arXiv: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

MC Modelling: Data-Driven Estimation

For BSM searches, relevant phase space are in tails of N(jets), VpT..., challenging to model.

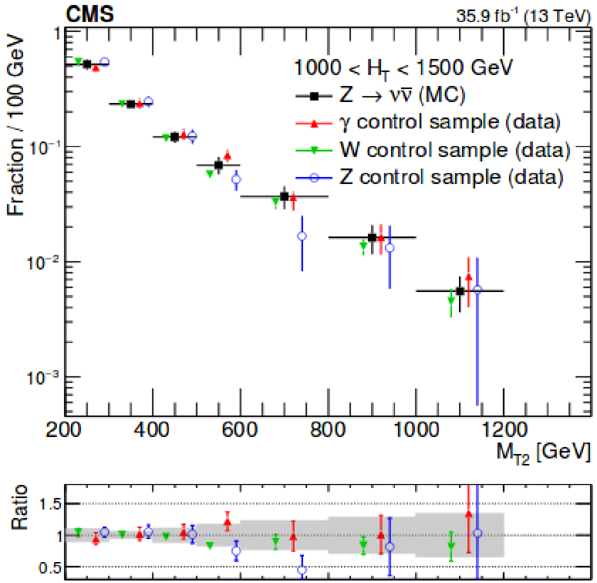
Since NLO samples can be quite statistically limited, some searches rely on LO samples and correction factors from data to describe these corners of phase space.

Control region choice represents trade-off between statistical and systematic uncertainty

V(DATA-others, CR)*V(MC, SR)/V(MC, CR)

Examples of CR —> SR translation

Final state	Background	Control sample	Features of extrapolation
0(1)-lepton + jets + MET	W+jets/ttbar with a "lost" lepton	1(2)-lepton	Correct lepton ID inefficiencies in MC to data, generally rely on MC to describe lepton acceptance
0-lepton + jets + MET	Z(νν)+jets	Z(ll)+jets or γ+jets	Hadronic recoil (MET excluding leptons or photon) in CR proxy for MET in SR, differences between Z and γ



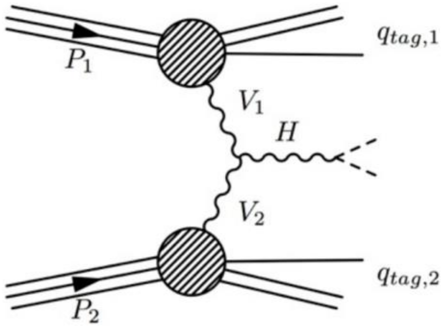
SUSY MT2

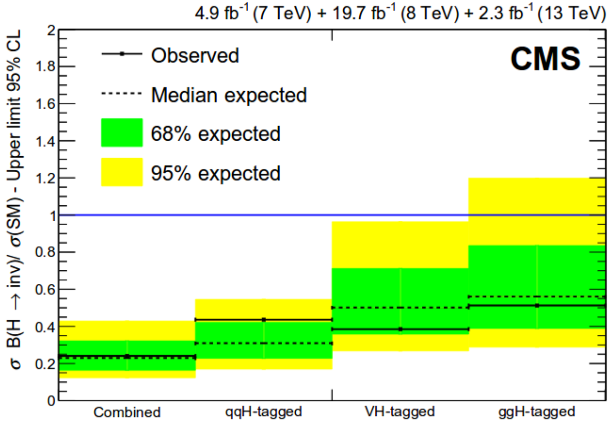
https://arxiv.org/abs/1705.04650

MC Modelling: Theo. Sys.

VBF Invisible Higgs https://arxiv.org/pdf/1610.09218.pdf

Systematic uncertainty	Impact
Common	
W to Z ratio in QCD produced V+jets	13%
W to Z ratio in EW produced V+jets	6.3%
Jet energy scale and resolution	6.0%
QCD multijet normalisation	4.3%
Pileup mismodelling	4.2%
Lepton efficiencies	2.5%
Integrated luminosity	2.2%
Signal specific	
ggH acceptance	3.8%
Renorm. and fact. scales and PDF (qqH)	1.8%
Renorm. and fact. scales and PDF (ggH)	$<0.2\%$
Total systematic	$^{+15}_{-19}\%$
Total statistical only	$^{+28}_{-27}\%$
Total uncertainty	$^{+32}_{-33}\%$





Dominant systematic uncertainty from theory. Need higher order calculations in MC generators to increase sensitivity of our searches in big data era of LHC

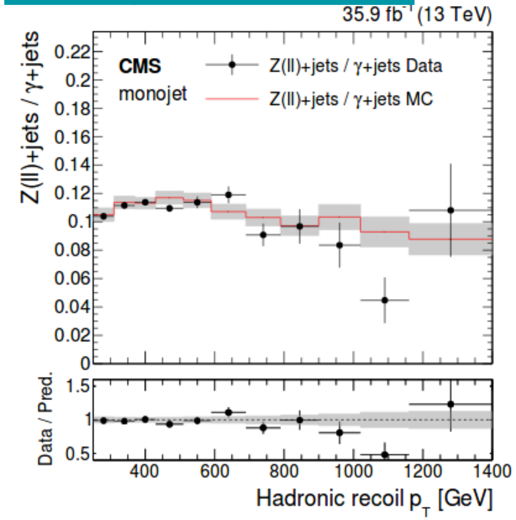
MC Modelling: V+Jets high tail, NLO EWK

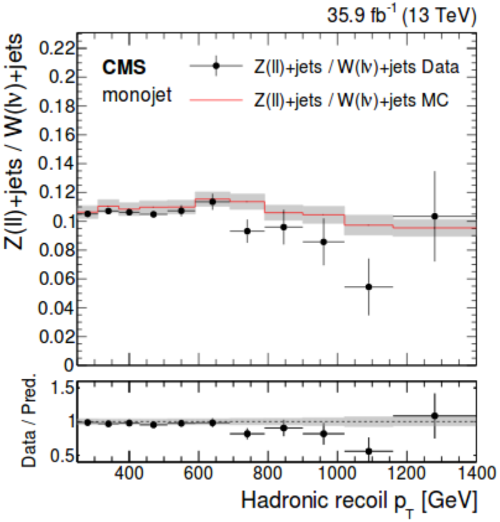
PTv up to 3TeV:

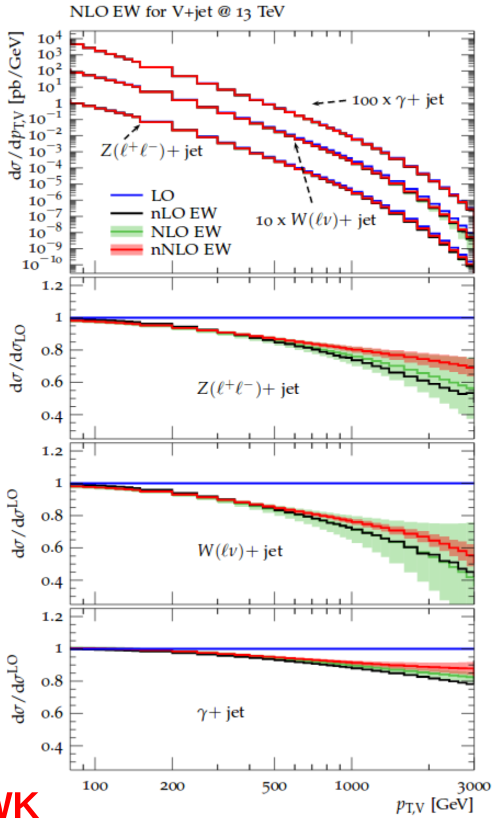
https://arxiv.org/abs/1705.04664

LO simulations for these processes are corrected using boson pT-dependent NLO QCD K-factors derived using MADGRAPH5_aMC@NLO. They are also corrected using pT-dependent higher-order EW corrections .

Monojet Searches:





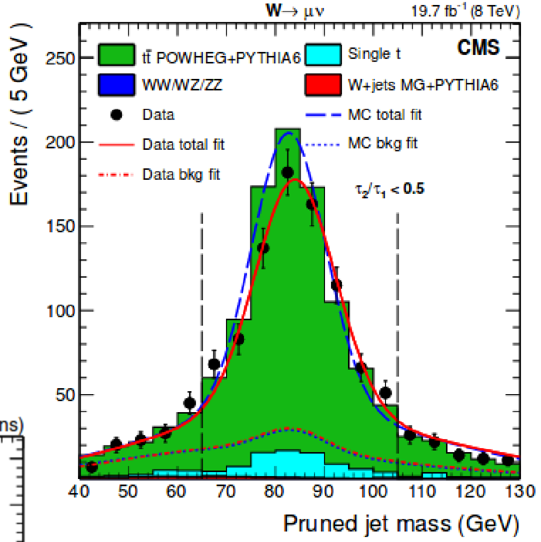


The theory uncertainty would have been very large without NLO QCD/EWK

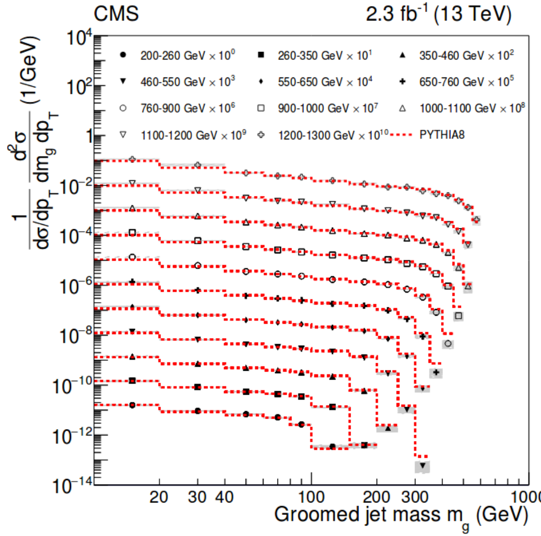
MC Modelling: Boosted Jets

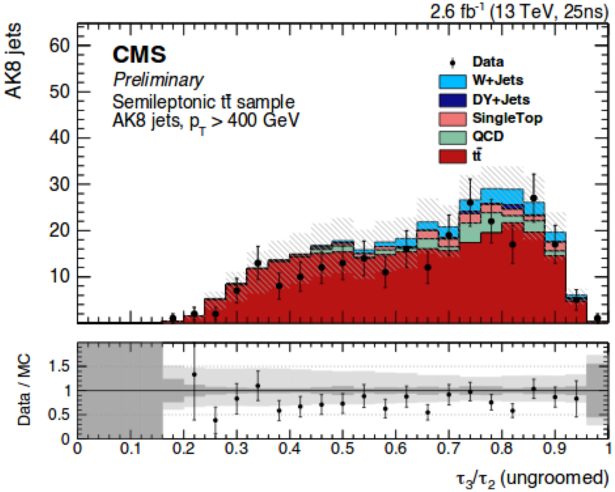
W-tagging can be calibrated with TTbar-Control region.

However, **H-tagging/tau32 validation** can be limited at high pT; Various sources of systematic uncertainty considered, e.g. choice of generator (Madgraph vs Powheg), parton shower description (Pythia8 vs Herwig++), pT dependence, purity of control sample.
http://cds.cern.ch/record/2256875?ln=en



Jet Grooming also tested to large extent:



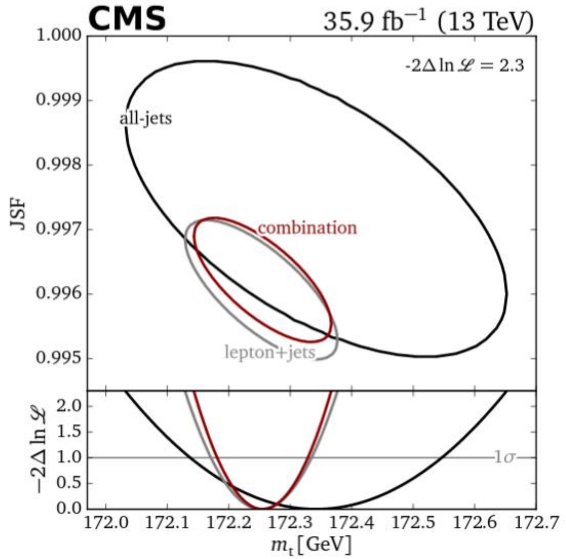


MC Modelling: b jet and Color-Connection ...

- 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 B0/B+decays and their corresponding uncertainties

[Top quark mass from ttbar fully hadronic (2016)]

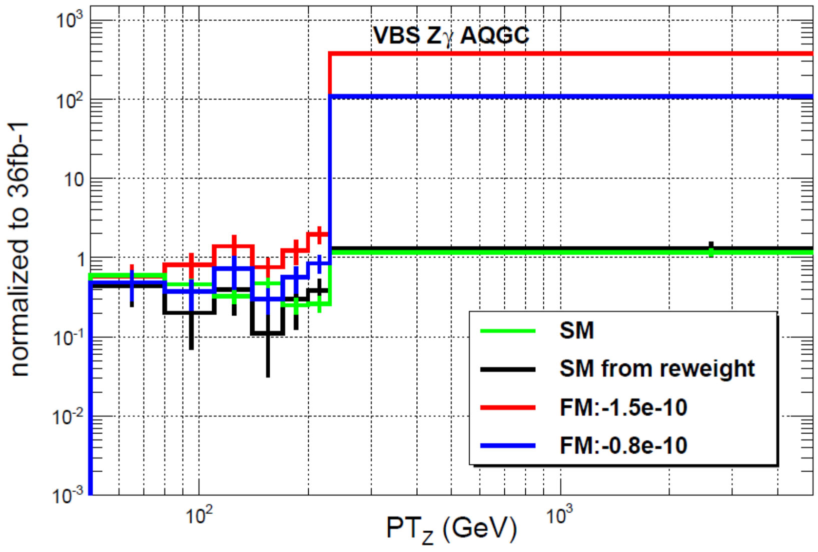
	2D δm_t^{2D} [GeV]	2D δJSF^{2D} [%]	1D δm_t^{1D} [GeV]	hybrid δm_t^{hyb} [GeV]	hybrid δJSF^{hyb} [%]
b jet modeling (quad. sum)	0.09	0.0	0.09	0.09	0.0
– b frag. Bowler–Lund	−0.07	0.0	−0.07	−0.07	0.0
– b frag. Peterson	−0.05	0.0	−0.04	−0.05	0.0
– semileptonic b hadron decays	−0.03	0.0	−0.03	−0.03	0.0
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



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

MC Modelling: PDF/Scale/PS/aGC Weights

- Improved treatment of theory systematics in searches/measurements (**scale/PDF**)
- Support for parameter reweighting in **new physics models** in some generators
- Reduce simulation time and storagment by orders of magnitude
- Recently, **Parton Shower Weights** are also available



See more on PS weights in TOP-17-013/015

Automated Variations of Shower Parameters for Uncertainty Bands in Pythia8:

'UncertaintyBands:doVariations= on',

isr:muRfac=0.5/2.0, fsr:muRfac=0.5/2.0

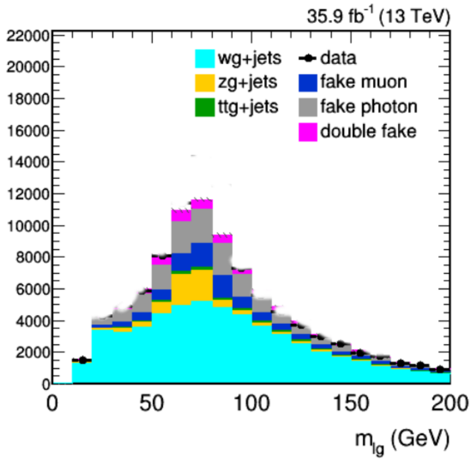
….

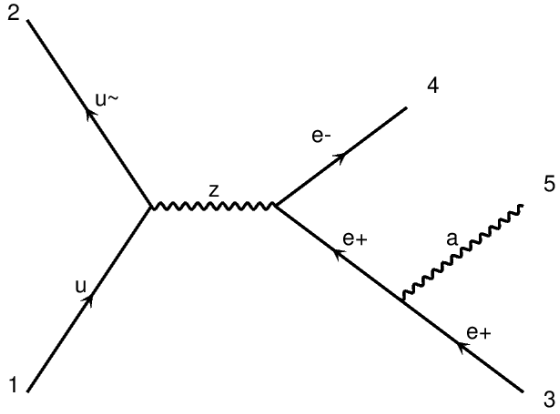
MC Modelling: W/Z Gamma

- Need to separate W/ZGamma and W/Z+Jets, using [particles status code]()

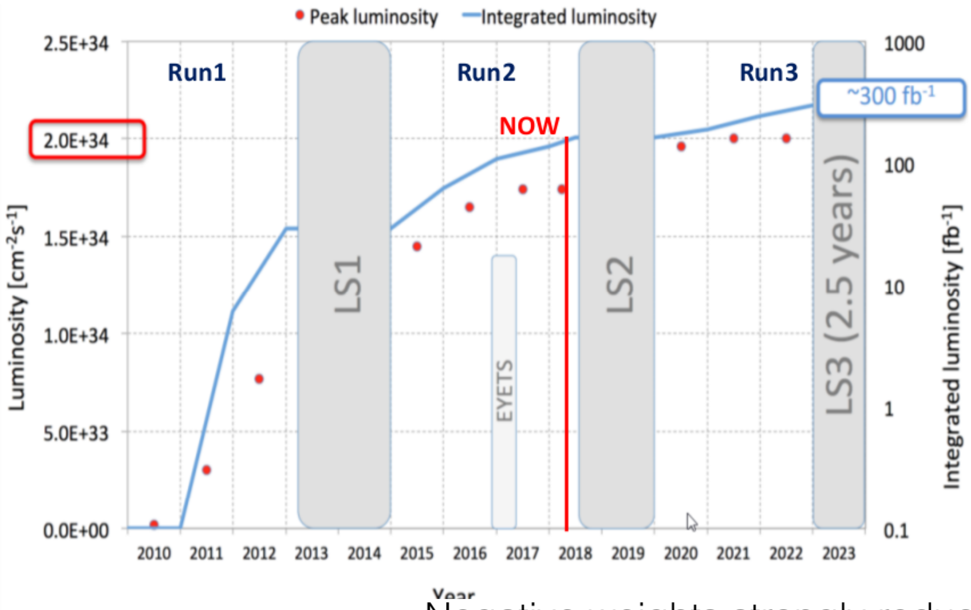
- TTGamma FxFx: may miss large FSR contributions
  generate p p > t t~ a [QCD] @0
  add process p p > t t~ a j [QCD] @1

- ZGamma can contribute largely to WGamma selections: pay attention to Z mass cut! Below ZGamms sample from MadGraph with Mll>30GeV.





Beyond Run2/3 and Future



Need to "fight" against conflicting requirements:

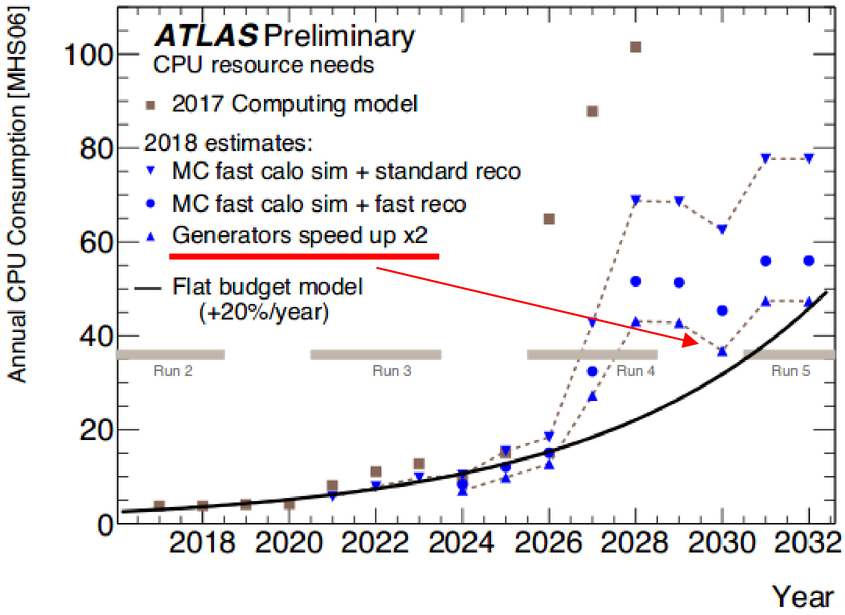
– (Much) larger datasets
– Increased measurement precision
– Need for alternative samples for systematics
– Flattening of computing resources (both cpu and disk space)

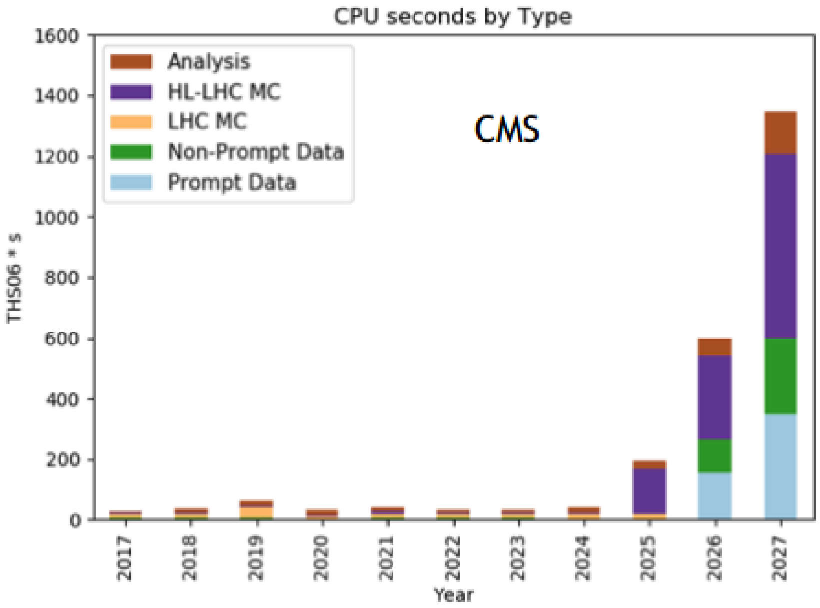
- Negative weights strongly reduce statistical power
- For weighted events w_i, effective events N_{eff} for fraction of negative weights f:

$$N_{eff} = \frac{(\sum_i w_i)^2}{\sum w_i^2} = N(1-2f)^2$$

- for 35% negative weights (common at for high jet-mulitplicity/ high pt)
  $\Longrightarrow$ 9% effective events compared to $w_i = 1$

Motivation for Generator upgrade

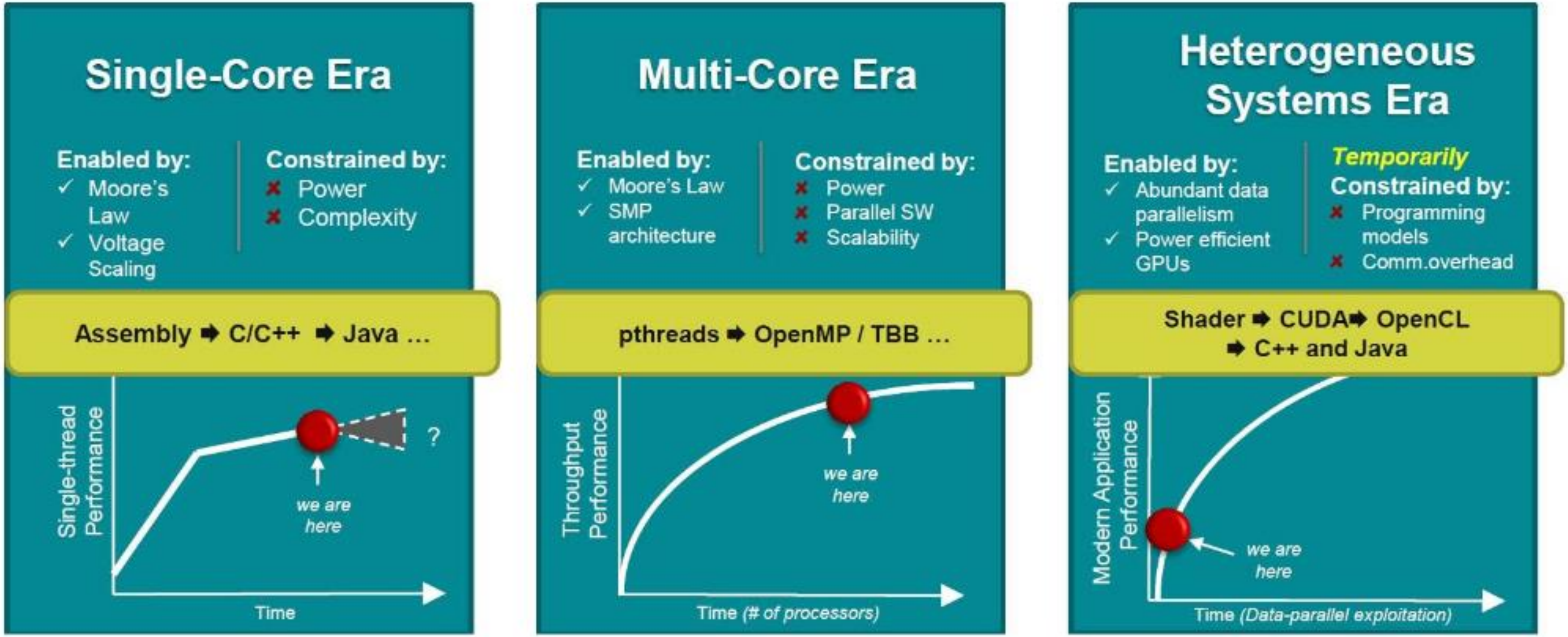




- We are OFF by ~5x on CPU power when considering Moore's law
- HL-LHC salvation will come from software improvements, not from hardware

Elizabeth Sexton-Kennedy

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.

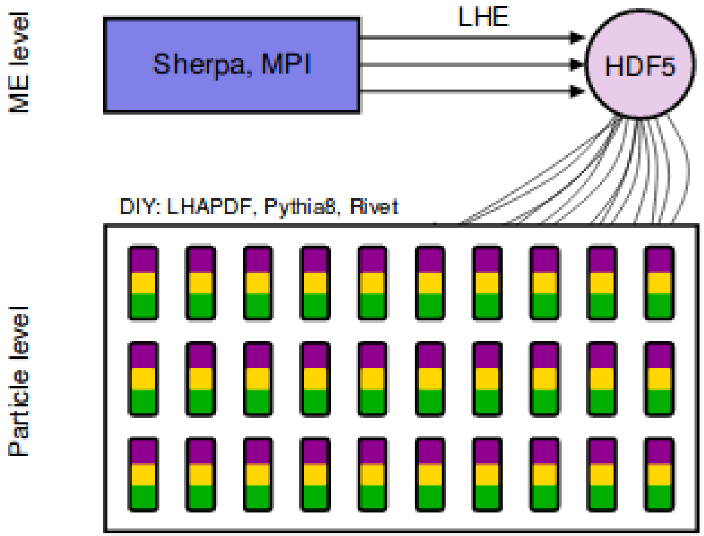
different types of compute units and interconnects

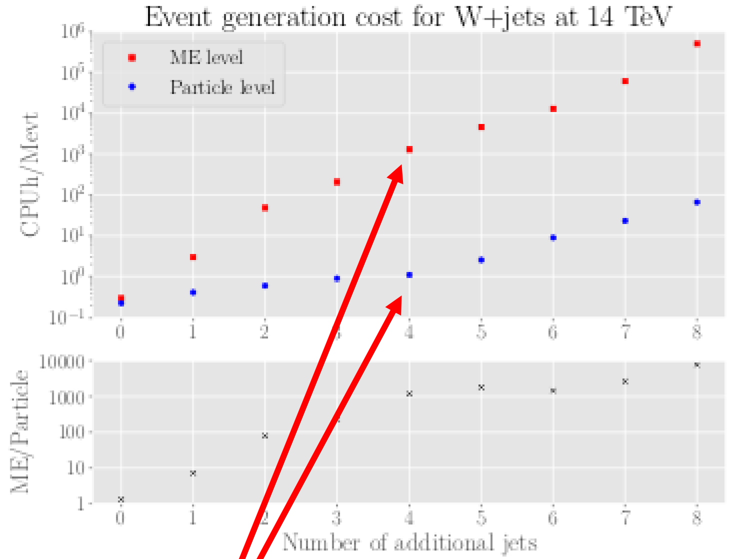
Highlight of recent effort towards HPC

[Holger Schulz](#)

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)

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

HEP Software Foundation

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.

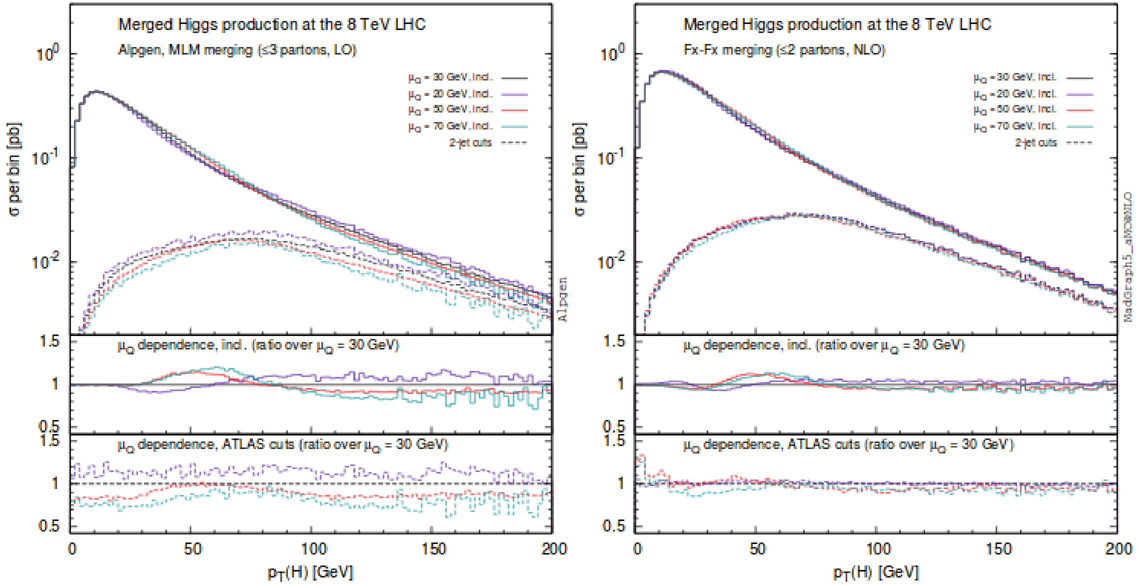
○ Define a programme of work that can be used to attract investment in these technical areas, aiming to have software engineers who can work together with the generator authors.

○ Identify ways of making new theoretical advances easier to implement in a computationally efficient way.

- **Better be right than fast**
- (N)NLO simulations allow one to treat theoretical systematics seriously. This has a very direct impact on both SM and BSM physics
- There is no way to guess the local (in phase-space) behaviour of systematics without performing simulations
- One typical example: merging scale variations for a given PTJ analysis cut. NLO and LO have typically different behaviours.
- Unweighting events means spending time earlier, but saving It (plus disk space) later

Merging: LO ⟶ NLO

Left: LO (Alpgen). Right: FxFx (MadGraph5_aMC@NLO)



Reweighting non-QCD corrections is a tricky business

Thanks