Generator Tunes & Sherpa in CMS

IP2I Group Meeting
16-July-2025

Jie Xiao

Outline

Activities related to CMS GEN Tuning and Sherpa generator

- GEN Tuning L3 convener: Sep 2023 ~ Sep 2025
- CMS Sherpa contact: Sep 2023 ~

Generator tunes in CMS

- Extraction of HERWIG7 and PYTHIA8 **DPS tunes** from CMS multi-jet measurements
- Tune with **jet substructure** analyses to fix the Data/MC discrepancies observed in some jet-substructure measurements
- Future plans:
  - Tune for **NNPDF4.0**
  - Automation tuning workflow

Sherpa generator in CMS

- Supports: documentations, Q&As
- Usage: physics processes with Sherpa in production
- Upgrade to Sherpa3

Strategy of event generator

Principle: divide et impera

Hard process

- fixed order perturbation theory
  - traditionally: Born-approximation

Bremsstrahlung:

- including ISR/FSR
- resummed perturbation theory

Hadronisation

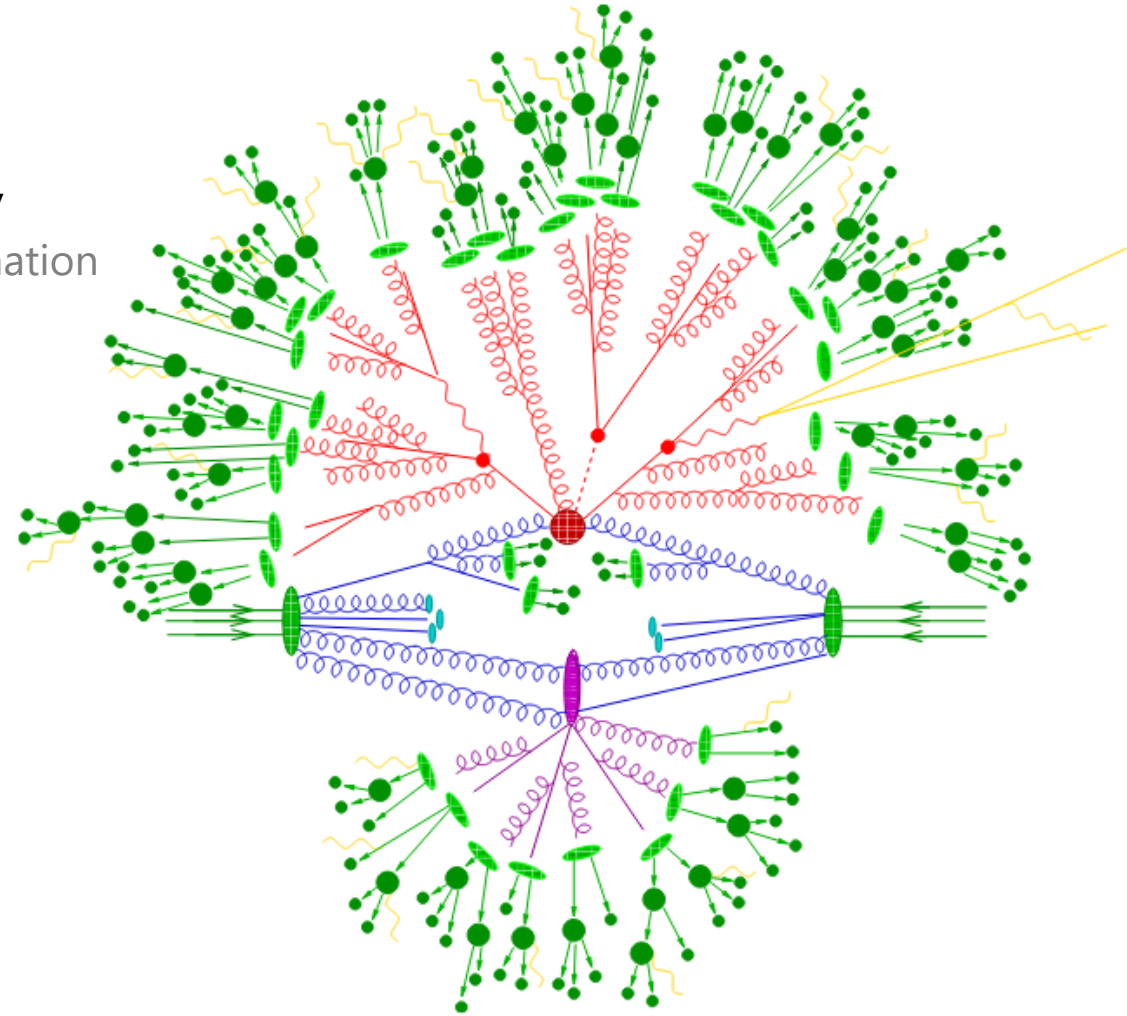
- phenomenological models

Hadron decays

- effective theories, data

"underlying event"

- phenomenological models



Strategy of event generator

Principle: divide et impera

Hard process

- fixed order perturbation theory
  - traditionally: Born-approximation

Bremsstrahlung:

- including ISR/FSR
- resummed perturbation theory

Hadronisation

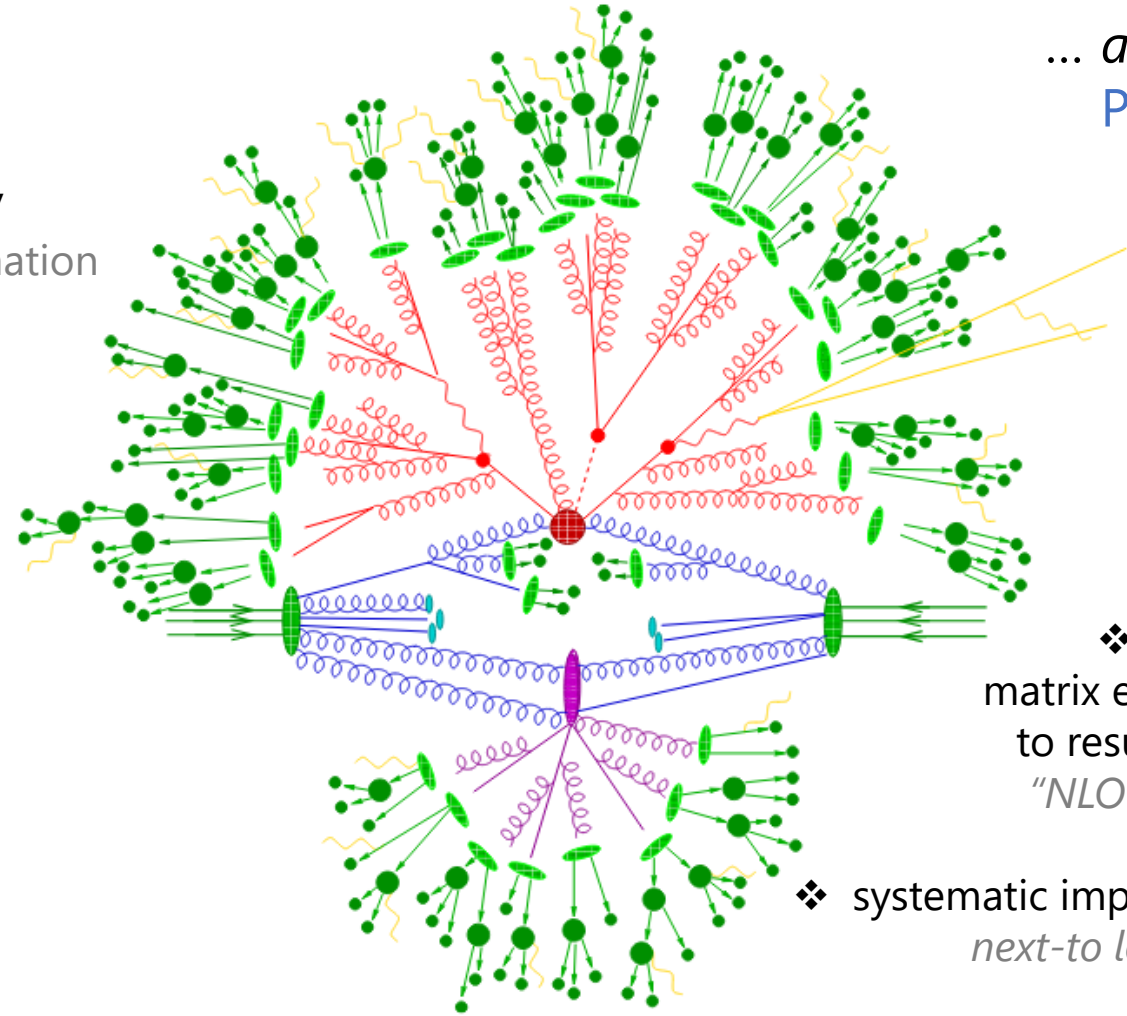
- phenomenological models

Hadron decays

- effective theories, data

"underlying event"

- phenomenological models



... and possible improvements

Phenomenological models

- "tuning" (fitting parameters to data)
- better models on "underlying event" simulation
- ...

Perturbative description

- inclusion of higher order exact matrix elements and correct connection to resummation in the parton shower: *"NLO-Matching" & "Multijet-Merging"*
- systematic improvement of the parton shower: *next-to leading (or higher) logs & colours*

Current generator tunes in CMS Run3

Generators	❖ Madgraph5_aMC@NLO: **2.9.X** ❖ Other versions of Madgraph5_aMC@NLO are supported ❖ **Pythia: 8.306** ❖ **Herwig: 7.X**
PDF	❖ **NNPDF3.1** (unchanged from Ultra Legacy Run2) ❖ Alternate sets will mostly contain NNLO PDFs including NNPDF4.0
Tune	❖ **CPX** family for **Pythia8** (GEN-17-001) ❖ **CHX** family for **Herwig7** (GEN-19-001) ❖ Intrinsic-k_T tune introduced for Drell-Yan processes
HEPMC	❖ **HEPMC2** (unchanged from Ultra Legacy Run2)

CPX and **CHX** families are recommended for Run3 samples

CP5 and **CH3** tunes are the most commonly used in Run3

Intrinsic-k_T tune is used in Run3 NLO **DYJets** and **WJets** production

Double Parton Scattering (DPS)

SPS and **DPS** show different topologies in the final state

DPS cross section is suppressed w.r.t SPS

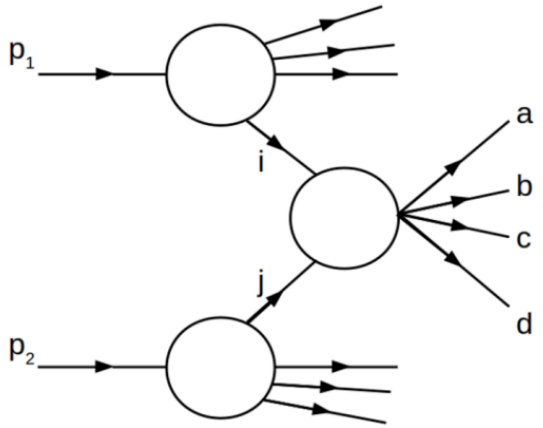
$$\frac{\sigma_{DPS}}{\sigma_{SPS}} \sim \frac{\Lambda^2}{Q^2}$$

(Λ: Hadronic scale ~1GeV; Q: Hard interaction scale)

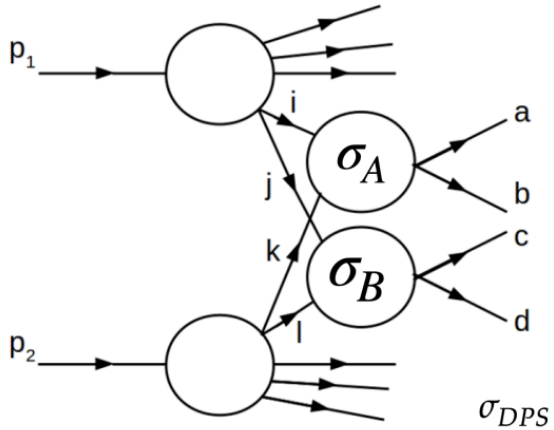
- DPS becomes more important as the collider energy grows. Larger density of partons at small-x values
- DPS can become competitive with SPS when SPS is hindered by small couplings; e.g. same-sign WW production
- …

At the LHC, DPS has been studied in multiple final-states such as

- 4 jets, 4 jets with b-jets, γ+3 jets, W(→*lν*)+dijet, Z(→l^+l^-)+*J/ψ*, *J/ψ*+*J/ψ*, same sign WW, etc



Single Parton scattering (SPS)
One hard scattering in a single pp collision. Final states particles are correlated.



$$\sigma_{DPS} = \frac{m}{2}\frac{\sigma_A \sigma_B}{\sigma_{eff}}$$

Double Parton scattering (DPS)
Two separate hard interactions in a single pp collision. Two pairs of partons from the incoming hadrons interact independently with each other.

DPS sensitive observables: 4 jets example

DPS sensitive observables

- The difference in azimuthal angle between the light jet pair

$$\Delta\phi = |\phi(j_1) - \phi(j_2)|$$

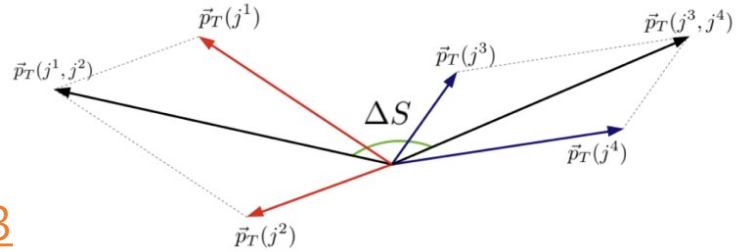
- The balance in p_T of the two light jets

$$\Delta p_T = \frac{|p_T(j_1) + p_T(j_2)|}{(||p_T(j_1)| + |p_T(j_2)||)}$$

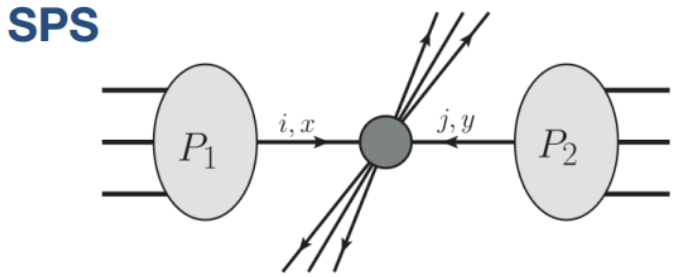
(Soft jets are expected to be produced by a 2nd scattering)

- The azimuthal angle between the two dijet pairs

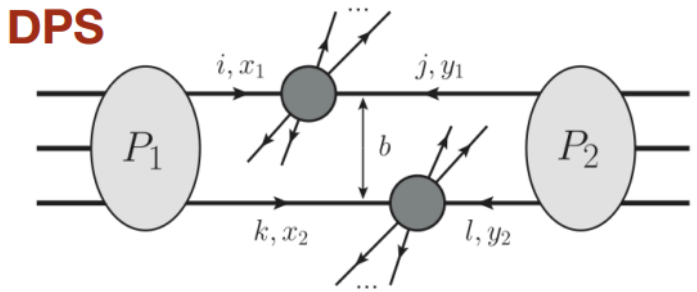
$$\Delta S = \frac{(p_T(j_3, j_4).p_T(j_1, j_2))}{(|p_T(j_3, j_4)| + |p_T(j_1, j_2)|)}$$



Phys. Rev. D **97**, 035013



correlated topologies, back-to-back jets



Uncorrelated topologies, back-to-back jets only for each of the independently produced jet pairs

Description of DPS observables

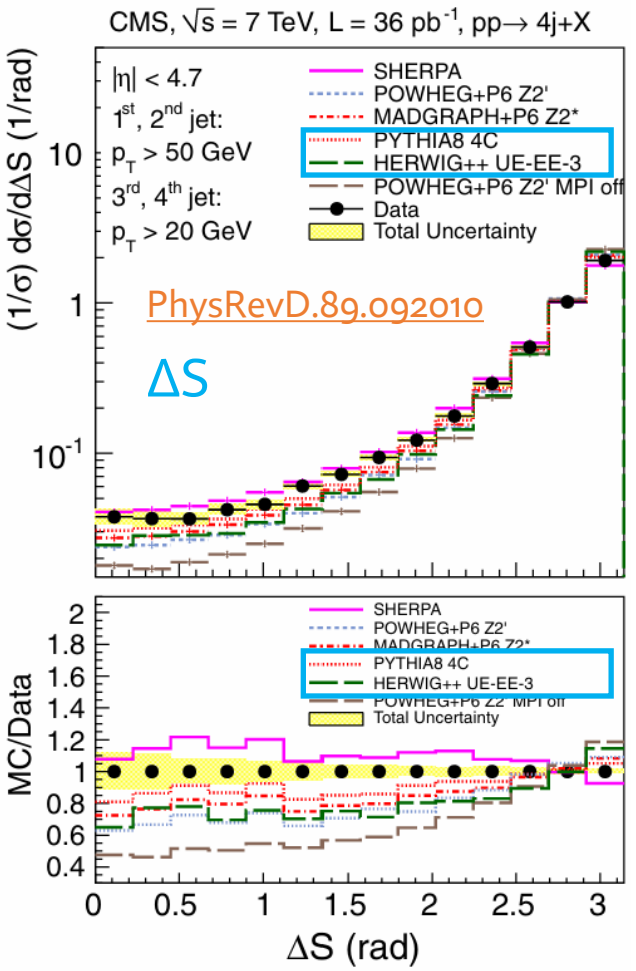
CP5 and CH3 tunes fail at describing DPS observables from CMS multi-jet

Studies ongoing to get a better description of these variables

- Multi-parton interaction (MPI) parameters are obtained through a fit to multi-jet measurements data collected by the CMS experiment at $\sqrt{s}$ = 7 TeV [1,2]

Relevant parameters

- **PYTHIA8**
  - pT0Ref
  - coreFraction
  - coreRadius
- **HERWIG7**
  - ColourReconnector:ReconnectionProbability
  - MPIHandler:InvRadius
  - MPIHandler:Power
  - MPIHandler:pTmin0



The PYTHIA8 and HERWIG7 have similar behavior in CP5 and CH3

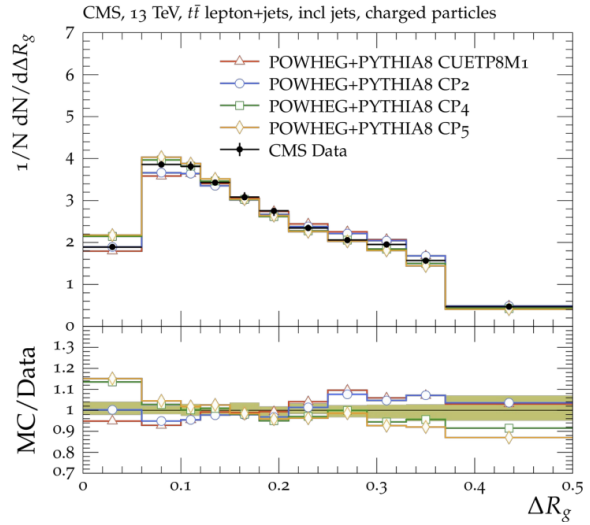
Description of jet substructure

Standard CP5 and CH3 are tunes of 4~5 minimum-bias and color reconnection parameters

- No focus on jet substructure (mostly sensitive to shower development (FSR) & hadronization effects)

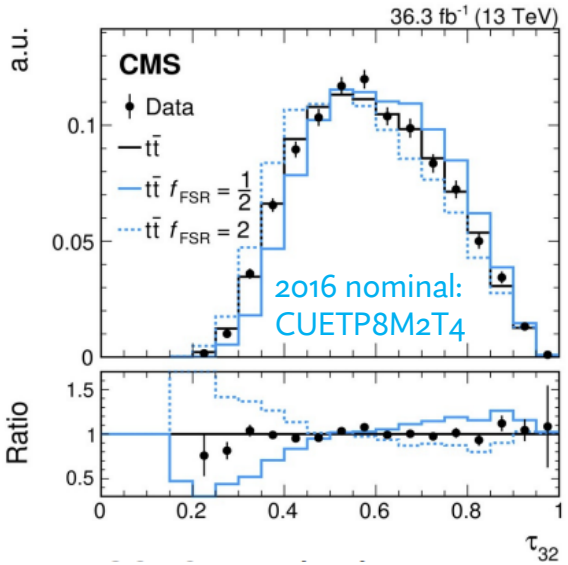
Jet substructure in ttbar TOP-17-013

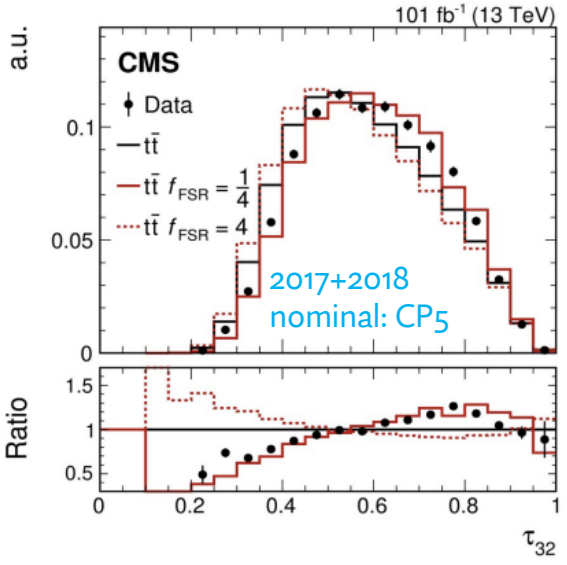
- **Angle between two groomed subjets** GEN-17-001, data from TOP-17-013



Boosted top quark jet substructure TOP-21-012

- **N-subjettiness ratio**



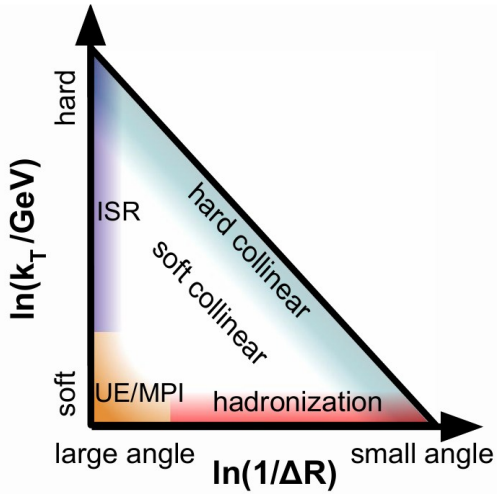


Measurements show jet substructure is not well modeled by simulation

Description of jet substructure

Lund plane: 2D representation of QCD radiation

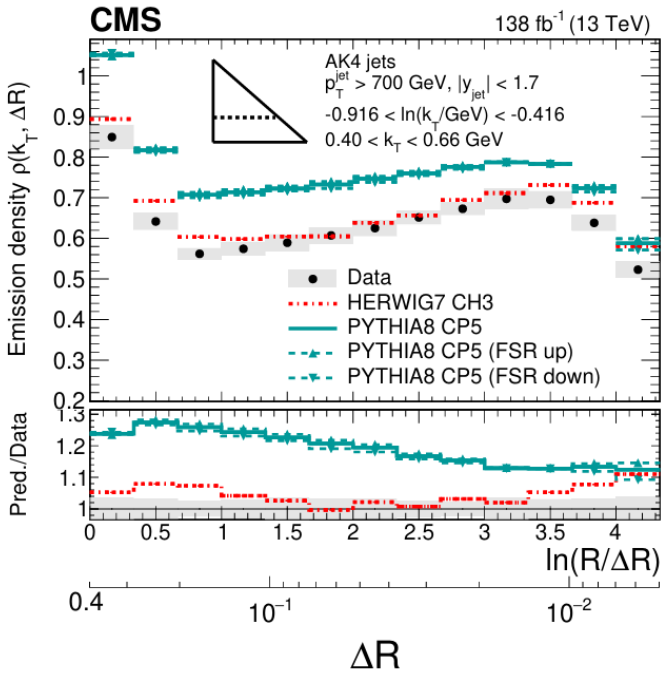
- A given jet is represented as a number of points in the Lund plane

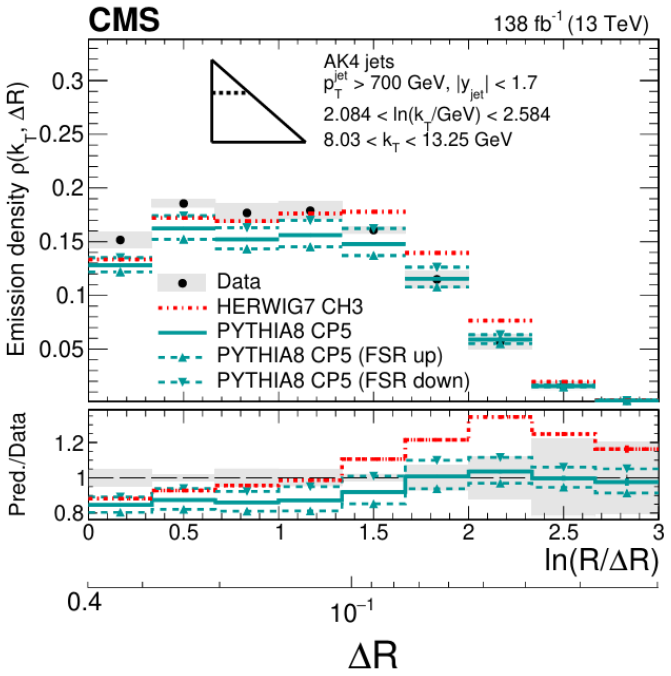


SMP-22-007 Measurement of the primary Lund jet plane density in pp collisions at 13 TeV

Neither CP5 nor CH3 describe data well everywhere

- 10–20% differences across phase space





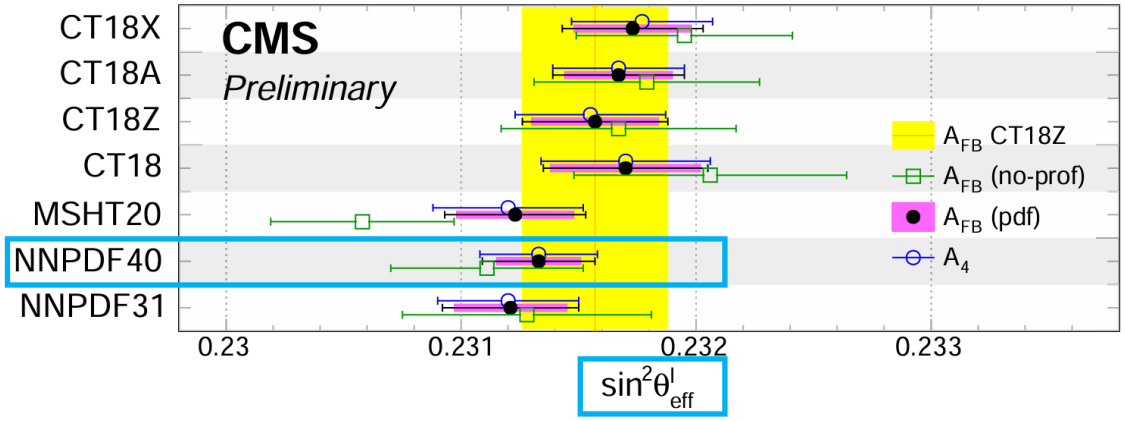
NNPDF4.0 in CMS

NNPDF4.0 was published with outstanding uncertainties (around 1% at the most x)

- NNPDF4.0 was already included in Run3 sample production
- Corresponding tune has not been developed

CMS GEN group got several queries about how to use it

- Measurement of the weak mixing angle with DY: SMP-22-010 tested NNPDF4.0 (details in Rhys Taus' talk)



NNPDF4.0 shows the best performance

NNPDF4.0 in CMS

NNPDF4.0 was published with outstanding uncertainties (around 1% at the most x)

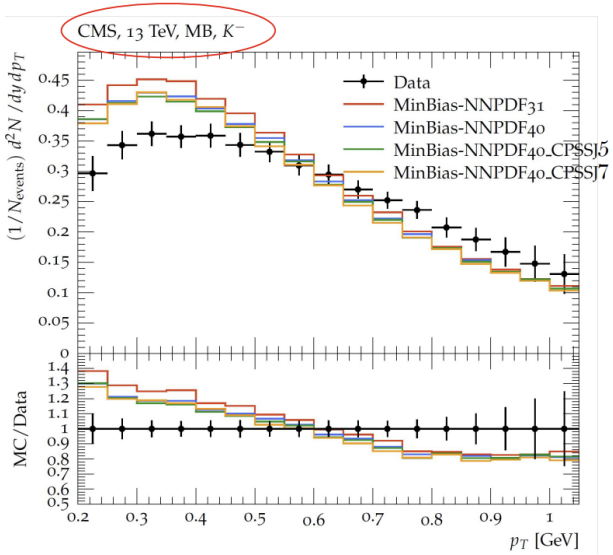
- NNPDF4.0 was already included in Run3 sample production
- Corresponding tune is in development

CMS GEN group got several queries about how to use it

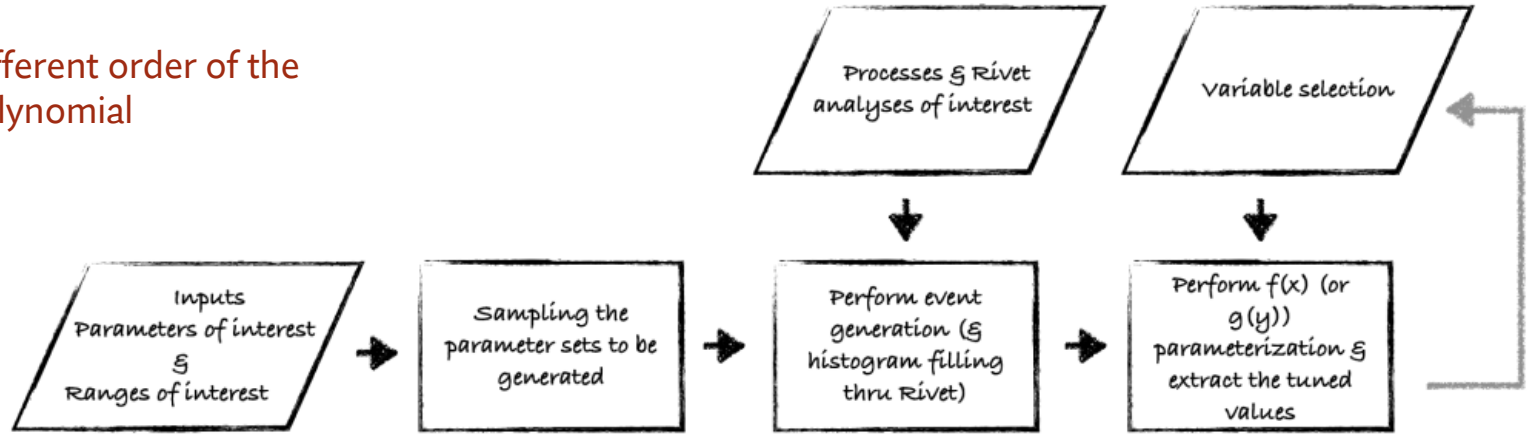
- Measurement of the weak mixing angle with DY: SMP-22-010 tested NNPDF4.0 (details in Rhys Taus' talk)

Need a new tune for NNPDF4.0: chance to have a common automated workflow for tunning

- Tuning tool itself is applicable to different event generators, use NNPDF4.0 as a demonstration



Different order of the polynomial



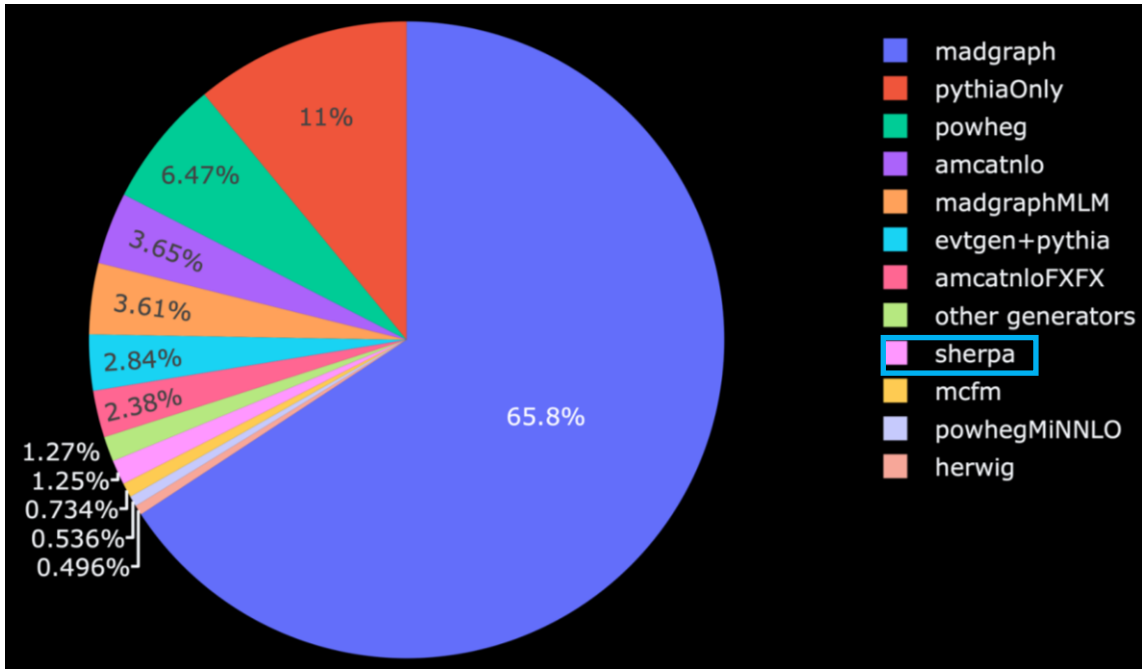
Plots from Sitian Qian

Usage of Sherpa in the CMS

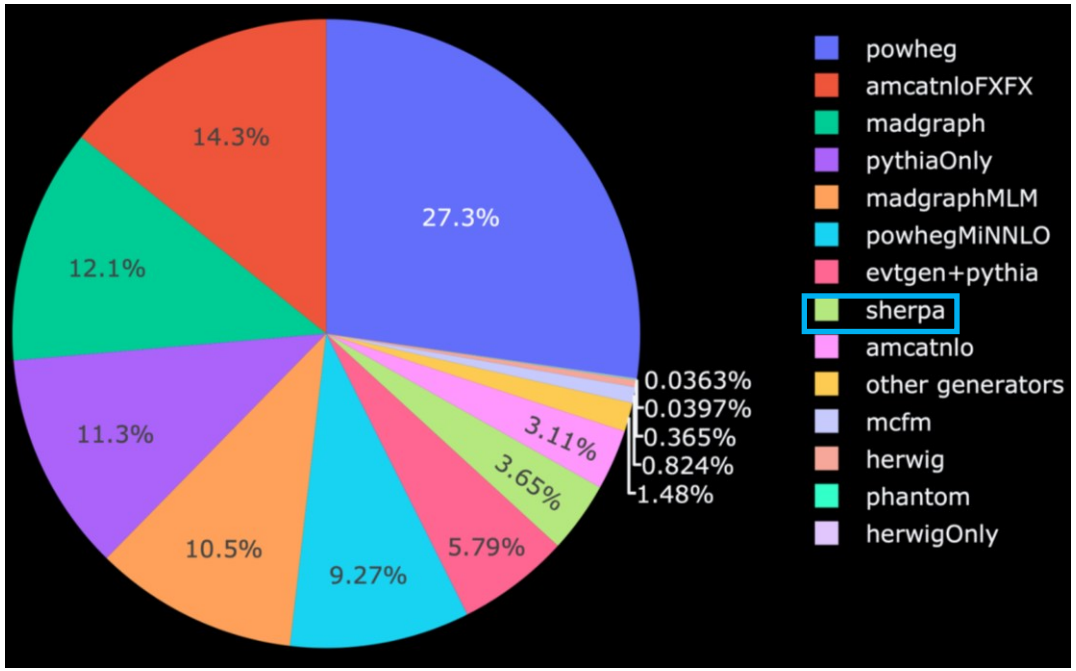
Sherpa haven't been widely used in CMS

- A lot of room for improvement

Generator usage split by samples



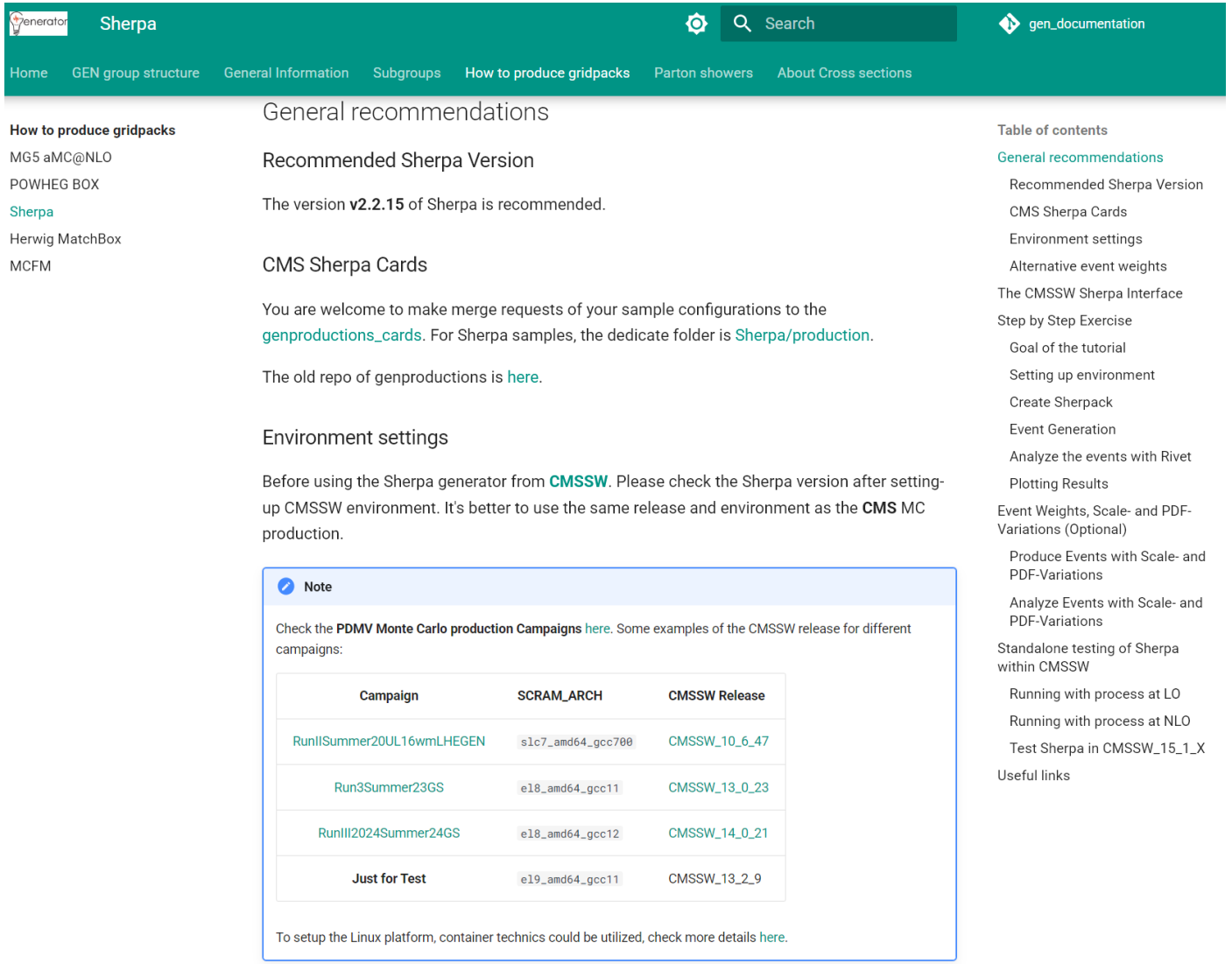
Generator usage split by events



From Sapta created in Apr. 2023

What we have for CMS Sherpa?

Starting point: documentation



Sherpa | Search | gen_documentation

Home | GEN group structure | General Information | Subgroups | How to produce gridpacks | Parton showers | About Cross sections

How to produce gridpacks
MG5 aMC@NLO
POWHEG BOX
Sherpa
Herwig MatchBox
MCFM

General recommendations

Recommended Sherpa Version

The version **v2.2.15** of Sherpa is recommended.

CMS Sherpa Cards

You are welcome to make merge requests of your sample configurations to the genproductions_cards. For Sherpa samples, the dedicate folder is Sherpa/production.

The old repo of genproductions is here.

Environment settings

Before using the Sherpa generator from **CMSSW**. Please check the Sherpa version after setting-up CMSSW environment. It's better to use the same release and environment as the **CMS** MC production.

Note

Check the **PDMV Monte Carlo production Campaigns** here. Some examples of the CMSSW release for different campaigns:

Campaign	SCRAM_ARCH	CMSSW Release
RunIISummer20UL16wmLHEGEN	slc7_amd64_gcc700	CMSSW_10_6_47
Run3Summer23GS	el8_amd64_gcc11	CMSSW_13_0_23
RunIII2024Summer24GS	el8_amd64_gcc12	CMSSW_14_0_21
Just for Test	el9_amd64_gcc11	CMSSW_13_2_9

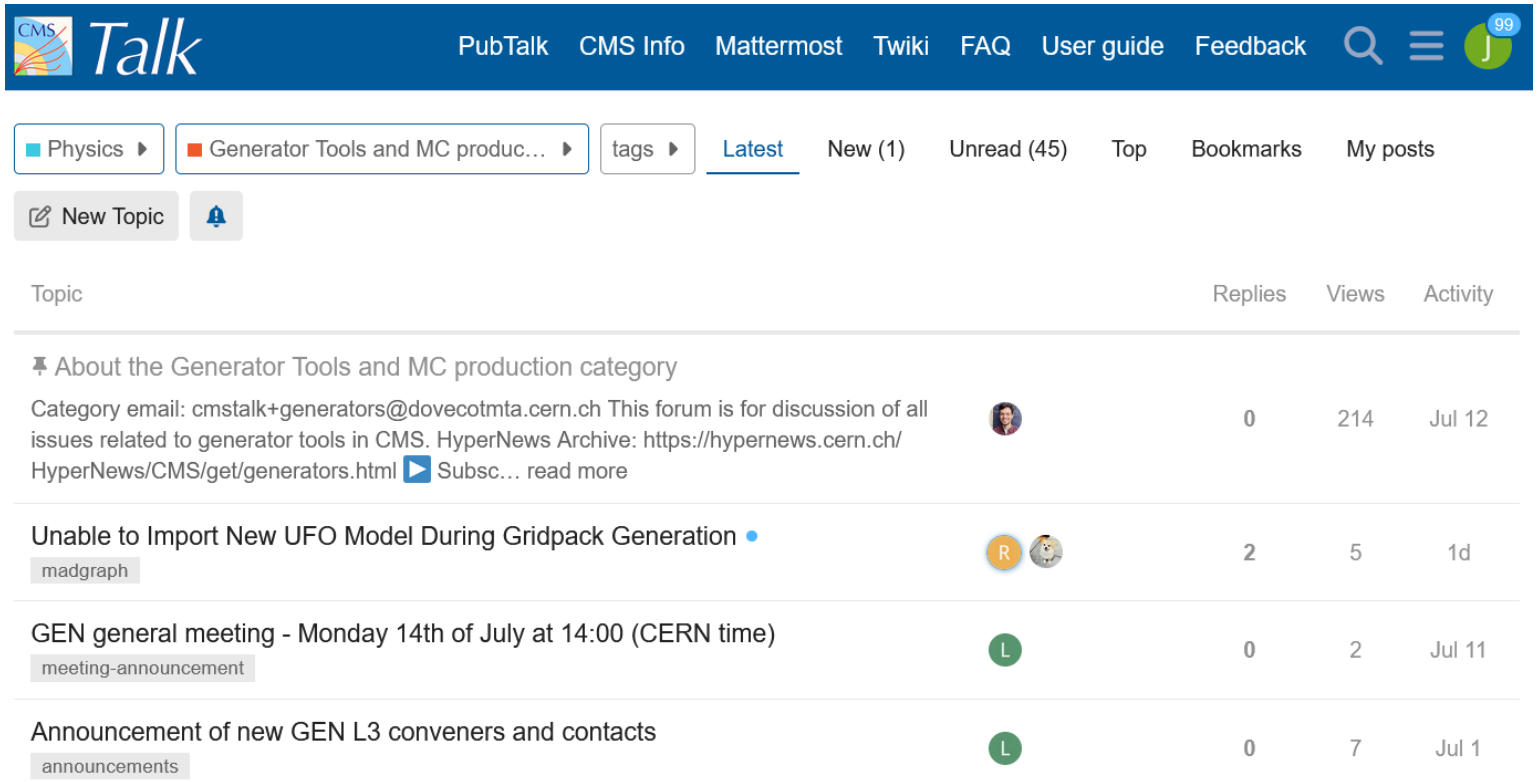
To setup the Linux platform, container technics could be utilized, check more details here.

Table of contents
General recommendations
Recommended Sherpa Version
CMS Sherpa Cards
Environment settings
Alternative event weights
The CMSSW Sherpa Interface
Step by Step Exercise
Goal of the tutorial
Setting up environment
Create Sherpack
Event Generation
Analyze the events with Rivet
Plotting Results
Event Weights, Scale- and PDF-Variations (Optional)
Produce Events with Scale- and PDF-Variations
Analyze Events with Scale- and PDF-Variations
Standalone testing of Sherpa within CMSSW
Running with process at LO
Running with process at NLO
Test Sherpa in CMSSW_15_1_X
Useful links

What we have for CMS Sherpa?

Starting point: documentation

Q&As CMS Talk Generator Tools



What we have for CMS Sherpa?

Starting point: documentation

Q&As CMS Talk Generator Tools

Infrastructure

- Sherpa is installed as an external package in CMS Software components (CMSSW)
  - Current recommended version: **v2.2.15**
- The Sherpa Interface of the CMSSW
  - Creation of the **Sherpack**: tarball file of phase-space integration and created libraries
  - Event generation
    - can be parallelized on grid
    - turn events into compliant format
    - feed output into CMSSW for detector simulation

What we have for CMS Sherpa?

Starting point: documentation

Q&As CMS Talk Generator Tools

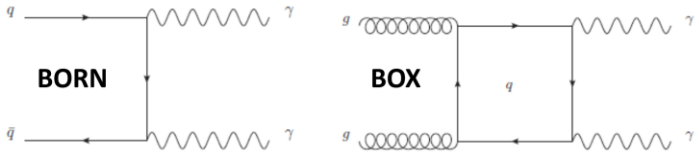
Infrastructure

- Sherpa is installed as an external package in CMS Software components (CMSSW)
  - Current recommended version: **v2.2.15**
- The Sherpa Interface of the CMSSW
  - Creation of the **Sherpack**: tarball file of phase-space integration and created libraries
  - Event generation
    - can be parallelized on grid
    - turn events into compliant format
    - feed output into CMSSW for detector simulation
- Workflow for sample production
  - Submit configuration file, e.g., Sherpa Run.dat, to Git repo for CMS MC configuration: genproductions_cards
  - Configuration review: Generator contact, GEN conveners
  - Prepare gridpack/Sherpack
  - Move to Physics Data And Monte Carlo Validation(PDMV) group for production

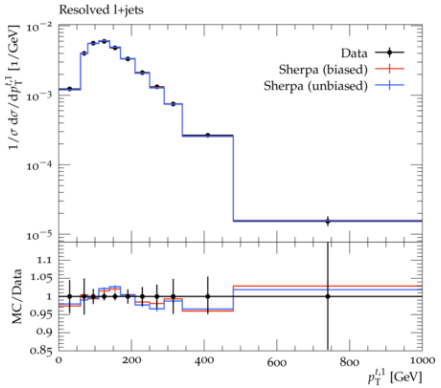
CMS Sherpa samples

Diphoton sample: one of the main backgrounds of H(γγ) related analysis

- **Diphoton box diagram** + diphoton jets at Leading Order (LO)
- Working on NLO sample production, which could give better description

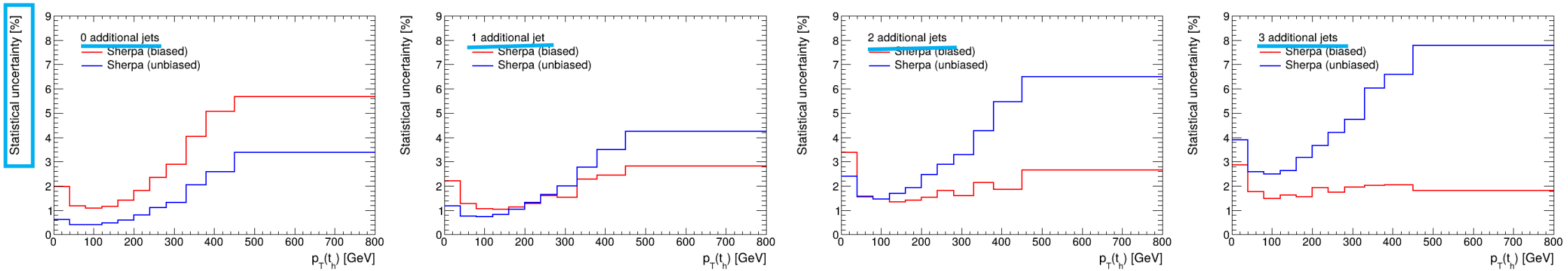


Plots from J. Llorente



TTbar semi-leptonically decaying sample

- The bias module `Enhance_Function` is used
  - No difference on the cross section between the biased and the nominal sample
  - Lower the statistical uncertainty in high jet multiplicity region, important for high HT region



Towards Sherpa3

From Enrico Bothmann

The Sherpa collaboration presents Sherpa 3 in July 2024

Comprises work of last five years

- new physics features
- more intuitive and flexible user interface
- more efficient CPU footprint
- modern build system
- comprehensive validation suite

Disclaimer

- v3.0.x are production releases ...
- but params of UE model not fully tuned

Towards Sherpa3

From Enrico Bothmann

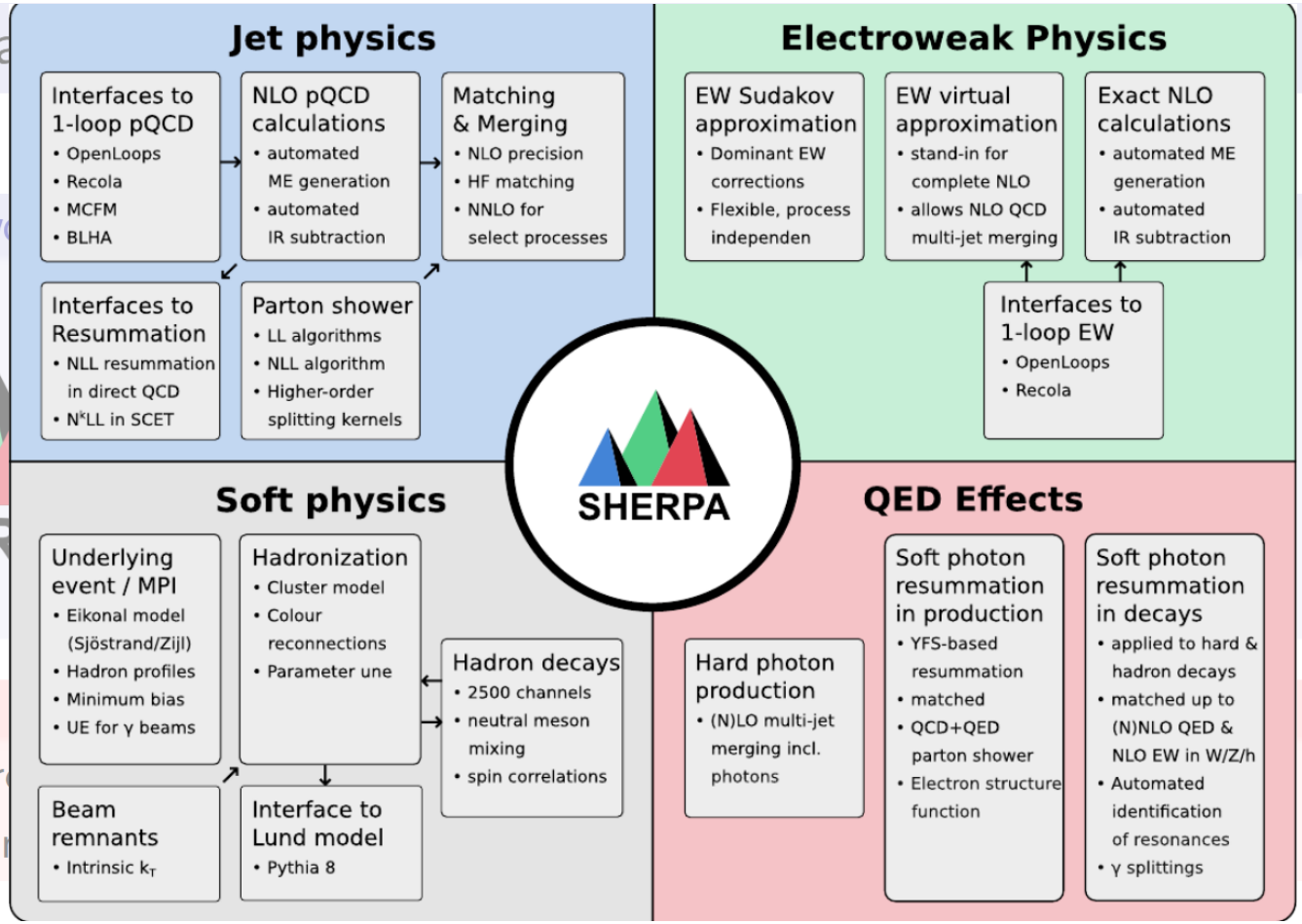
The Sherpa

Comprises w

SHER

Disclaimer

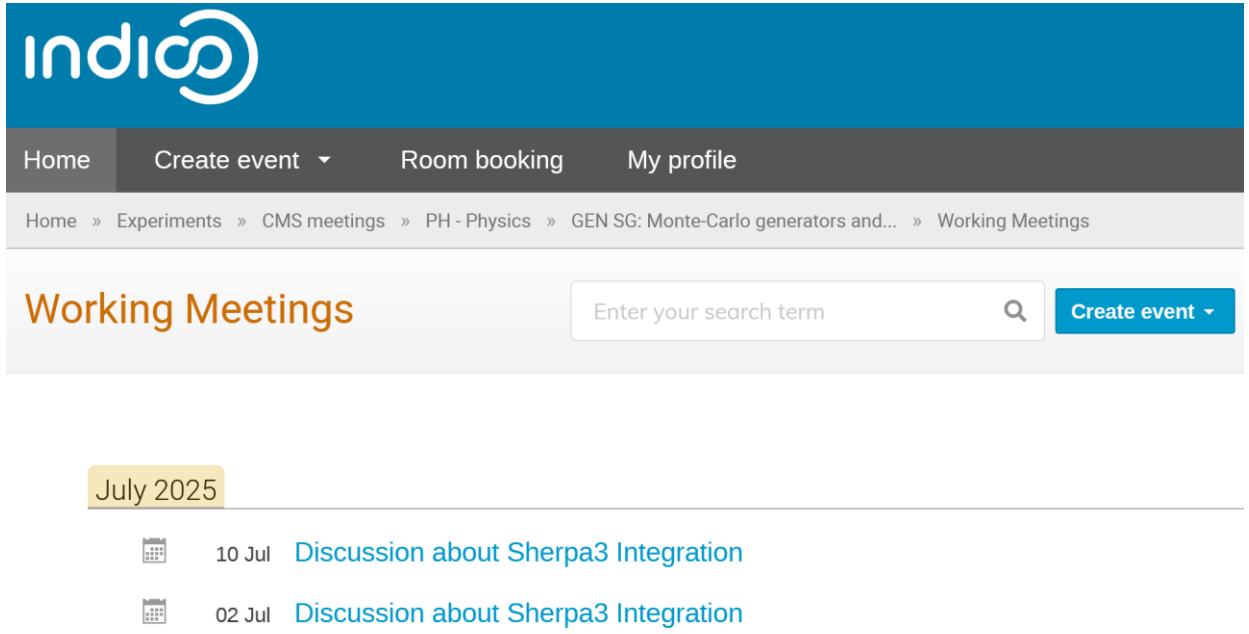
- v3.0.x ar
- but para



Towards Sherpa3: CMS efforts

Early stage

- Dedicated discussion on Sherpa3 integration recently



Towards Sherpa3: CMS efforts

Early stage

- Dedicated discussion on Sherpa3 integration recently

Tasks on-going

- Sherpa3 integration, details in gen_tasklist #48

cms-gen / Gen Tasklist / Issues / #48

Sherpa3 Integration

Open · Issue created 1 week ago by **Jie Xiao**

GOAL: Monitor the progress of `Sherpa3` Integration in `cmsdist`. `cmsdist` is used to build external packages of the `cmssw`.

Version of the Sherpa and related packages

- [x] `Sherpa` v3.0.1.
  - [ ] PS: `Sherpa2` could be upgraded to `v2.2.16`.
- [x] `Recola` v1.4.4: one-loop matrix-element generator could be interfaced to Sherpa for SM process at both NLO QCD and EW accuracy.
- [ ] `MCFM` v10.3: Sherpa MCFM interface can be used for analytic one-loop corrections which would be faster than OpenLoops cmsdist issue#8112.
- [ ] `OpenLoops` v2.1.4: current version in CMS is `v2.1.2`, maybe it's better to decouple the OpenLoops upgration, just keep it here as a reminder.

`Sherpa3` integration example:

- Example based on `CMSSW_14X`: https://github.com/shimashimarin/cmsdist/pull/1/files
- Example based on cmsdist `IB/CMSSW_15_1_X/master`: PR#2 of personal branch `IB/CMSSW_15_1_X/master_test_sherpa3`

Starting point:

- [ ] Get familiar with the documentation of Building external packages.
- [ ] Please make pull requests to the `IB/CMSSW_15_1_X/master` branch of the `cmsdist` fork under CMS-Sherpa. Avoid creating branch directly from the official repo.
  - [ ] Once `Sherpa3` integrated, check if need to backport to the production release. E.g., `RunIII2024Summer24GS` now is based on `CMSSW_14_0_21`.
    - Start from `IB/CMSSW_15_1_X/master` of `cmsdist`, `HepMC3` is fully supported.

2 of 8 checklist items completed · Edited 4 days ago by Jie Xiao

Towards Sherpa3: CMS efforts

Early stage

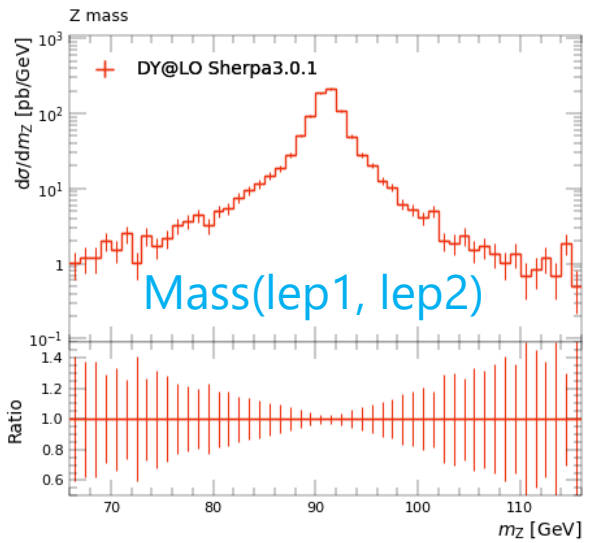
- Dedicated discussion on Sherpa3 integration recently

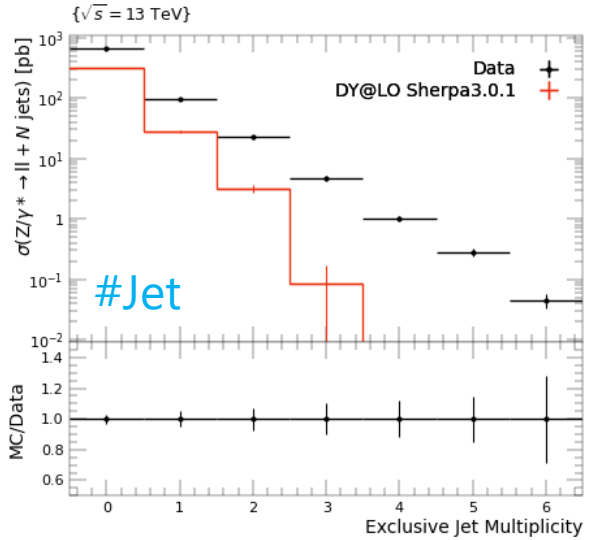
Tasks on-going

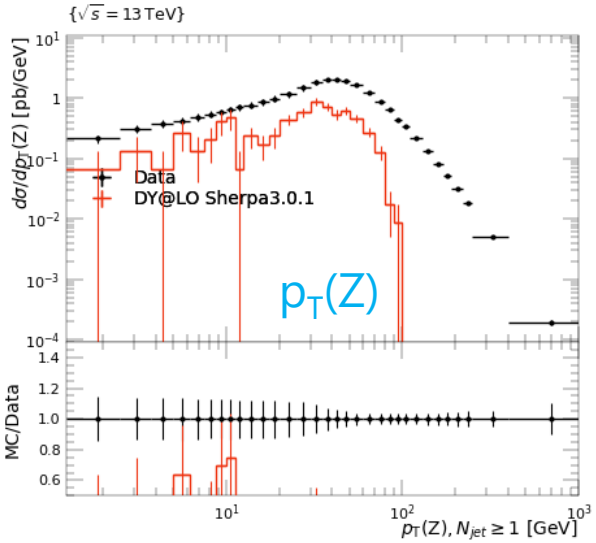
- Sherpa3 integration, details in gen_tasklist #48
  - Preliminary installation done

Quick test with DY@LO configuration

standalone run with the CMSSW installed Sherpa3







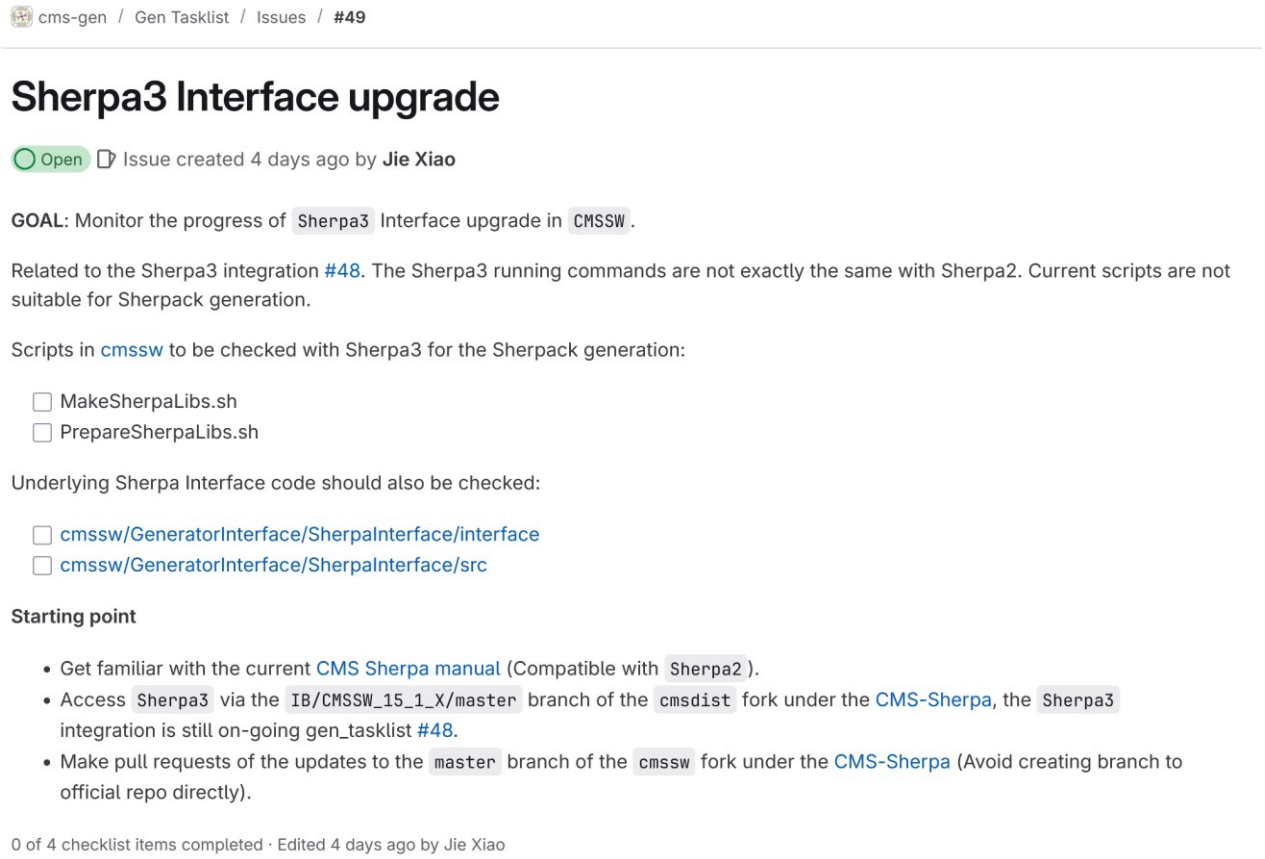
Towards Sherpa3: CMS efforts

Early stage

- Dedicated discussion on Sherpa3 integration recently

Tasks on-going

- Sherpa3 integration, details in gen_tasklist #48
  - Preliminary installation done
- Sherpa3 interface upgrade, details in gen_tasklist #49
  - CMDs changed since Sherpa3, current SherpaInterface no longer valid



cms-gen / Gen Tasklist / Issues / #49

Sherpa3 Interface upgrade

Open · Issue created 4 days ago by **Jie Xiao**

GOAL: Monitor the progress of `Sherpa3` Interface upgrade in `CMSSW`.

Related to the Sherpa3 integration #48. The Sherpa3 running commands are not exactly the same with Sherpa2. Current scripts are not suitable for Sherpack generation.

Scripts in cmssw to be checked with Sherpa3 for the Sherpack generation:

- [ ] MakeSherpaLibs.sh
- [ ] PrepareSherpaLibs.sh

Underlying Sherpa Interface code should also be checked:

- [ ] cmssw/GeneratorInterface/SherpaInterface/interface
- [ ] cmssw/GeneratorInterface/SherpaInterface/src

Starting point

- Get familiar with the current CMS Sherpa manual (Compatible with `Sherpa2`).
- Access `Sherpa3` via the `IB/CMSSW_15_1_X/master` branch of the `cmsdist` fork under the CMS-Sherpa, the `Sherpa3` integration is still on-going gen_tasklist #48.
- Make pull requests of the updates to the `master` branch of the `cmssw` fork under the CMS-Sherpa (Avoid creating branch to official repo directly).

0 of 4 checklist items completed · Edited 4 days ago by Jie Xiao

Towards Sherpa3: CMS efforts

Early stage

- Dedicated discussion on Sherpa3 integration recently

Tasks on-going

- Sherpa3 integration, details in gen_tasklist #48
  - Preliminary installation done
- Sherpa3 interface upgrade, details in gen_tasklist #49
  - CMDs changed since Sherpa3, current SherpaInterface not longer valid
- Sherpa3 Validation Survey, details in gen_tasklist #50
  - Propose processes for validation
  - Need personpower

cms-gen / Gen Tasklist / Issues / **#50**

Sherpa3 Validation Survey

Open · Issue created 4 days ago by **Jie Xiao**

Following the `Sherpa3` integration #48 and interface upgrade #49, detailed validation should be carried out.

This issue is used to collect the physics processes and ideas for `Sherpa3` validation.

List of the validation process (preliminary):

- WJets
- ZJets
- ttbar
- ttV
- GJets
- Diphoton
- Diboson
- QCD
- VBS/VBF

Ideas of validation:

- Automation validation framework with `CI/CD`?

New features of the Sherpa:

- EW correction via Sudakov approximation
- For VBS/VBF process: Consistent treatment of special colour flow via new VBF scale setter
- Polarization

Feel free to add things related, e.g., Run cards, new processes, new ideas ...

Summary

CPX and **CHX** families are recommended for Run3 in CMS

- **CP5** and **CH3** tunes are the most commonly used

Ongoing studies to improve current tunes

- New tune sets to improve the discrepancies shown (such as: DPS variables, jet substructure, …)

New opportunities

- Common automated workflow development for tunning

Sherpa is not widely used in CMS, plenty of room for improvement

- Diphoton related analyses rely on Sherpa for **diphoton background simulation**
- **ttbar** production was studied and in production
- Try to have **higher order** simulation of some backgrounds, e.g., **γ+Jets and QCD**

Challenge of upgrading to **Sherpa3**

- Sherpa3 integration in CMSSW, CMSSW interface upgrade, Sherpa3 validation