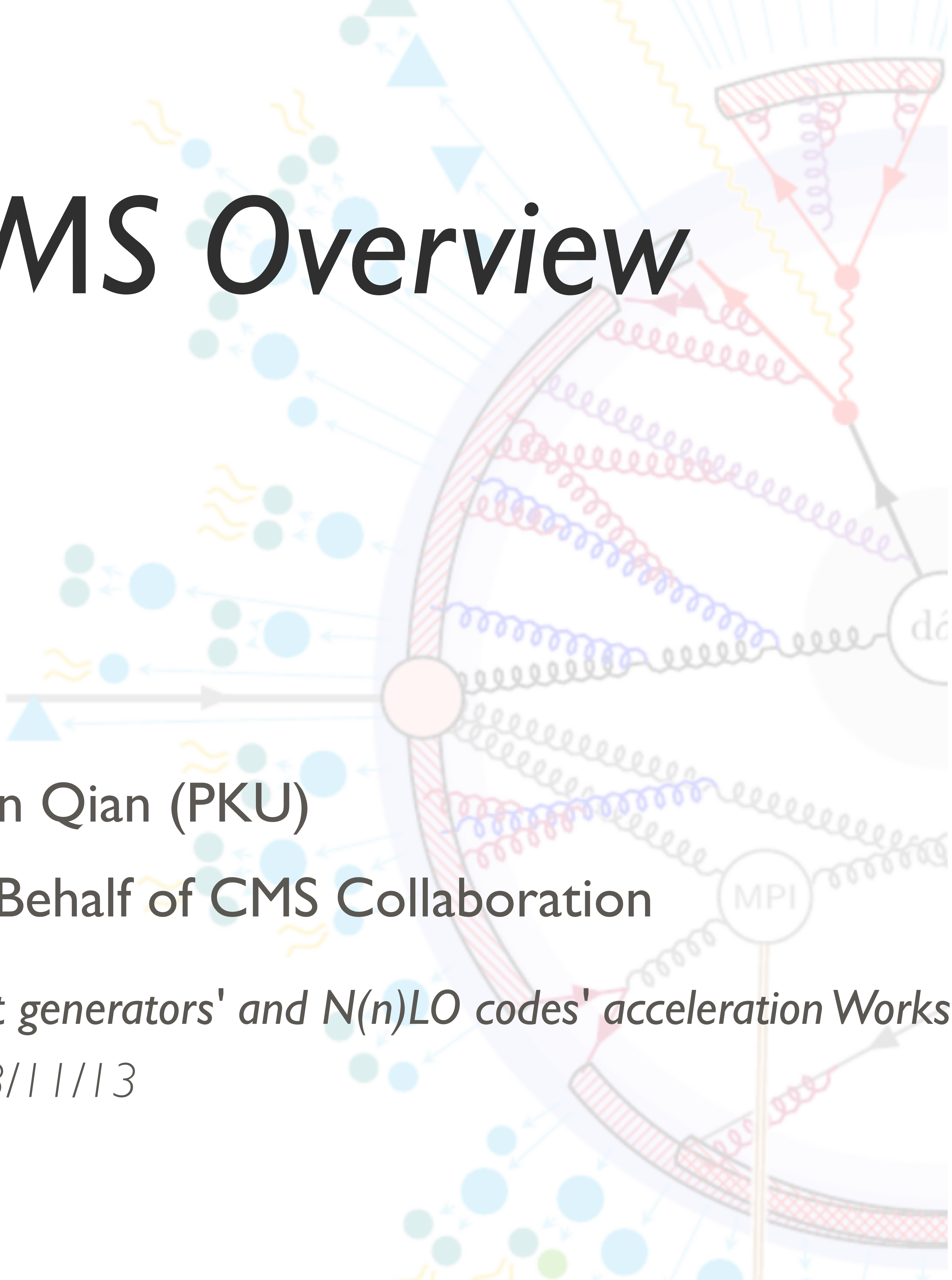
CMS Overview

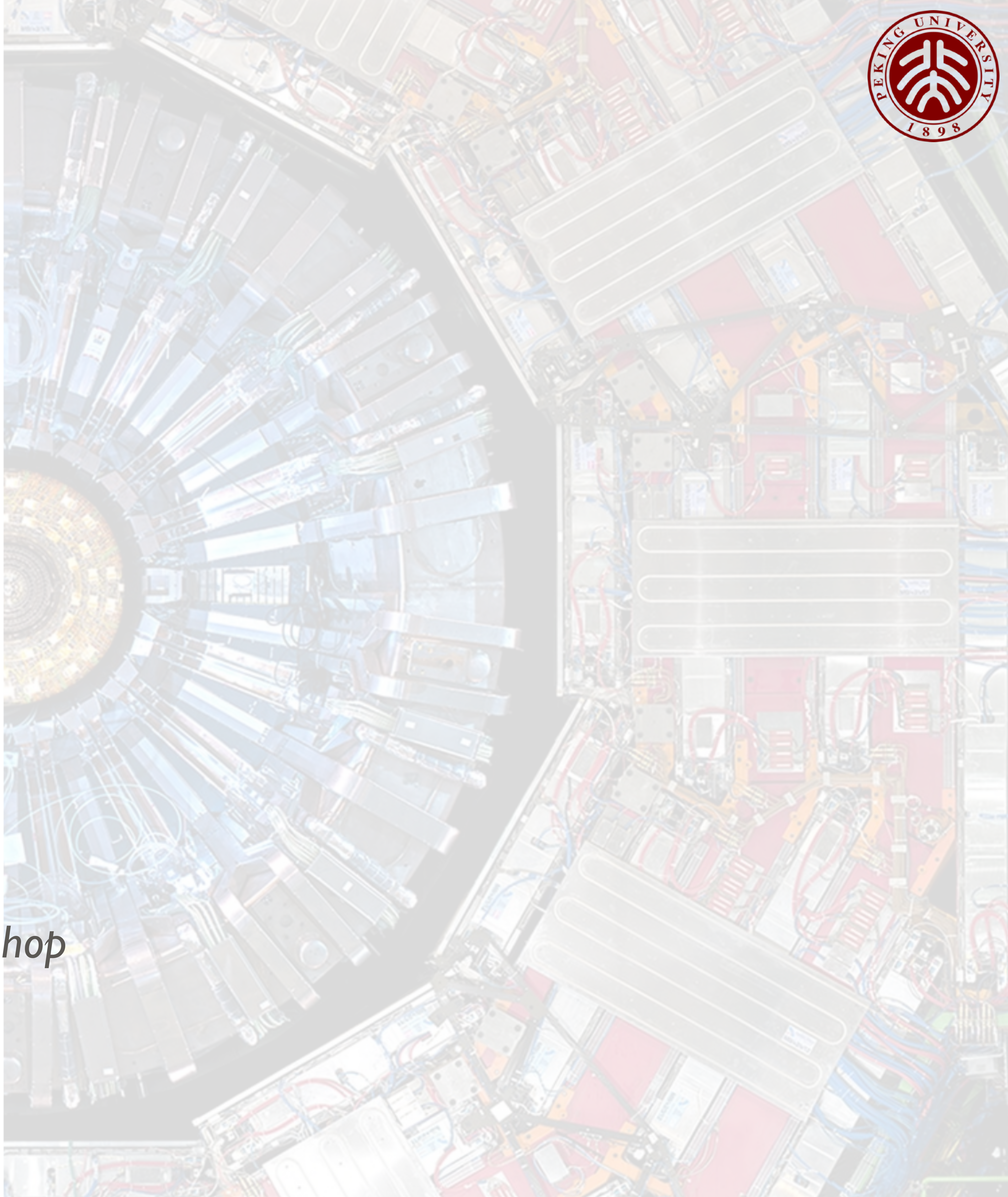
Sitian Qian (PKU)

On Behalf of CMS Collaboration

Event generators' and N(n)LO codes' acceleration Workshop

2023/11/13



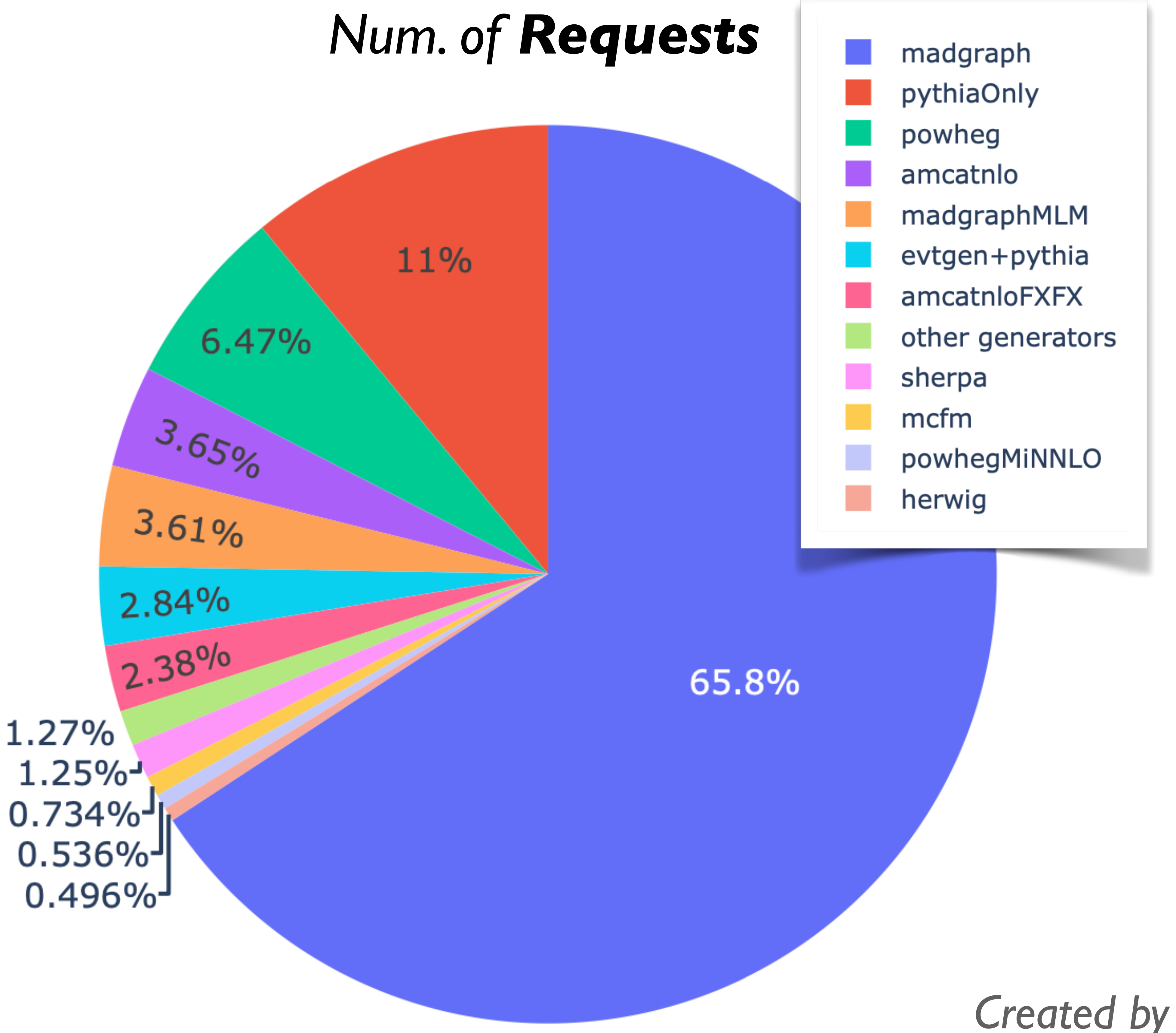


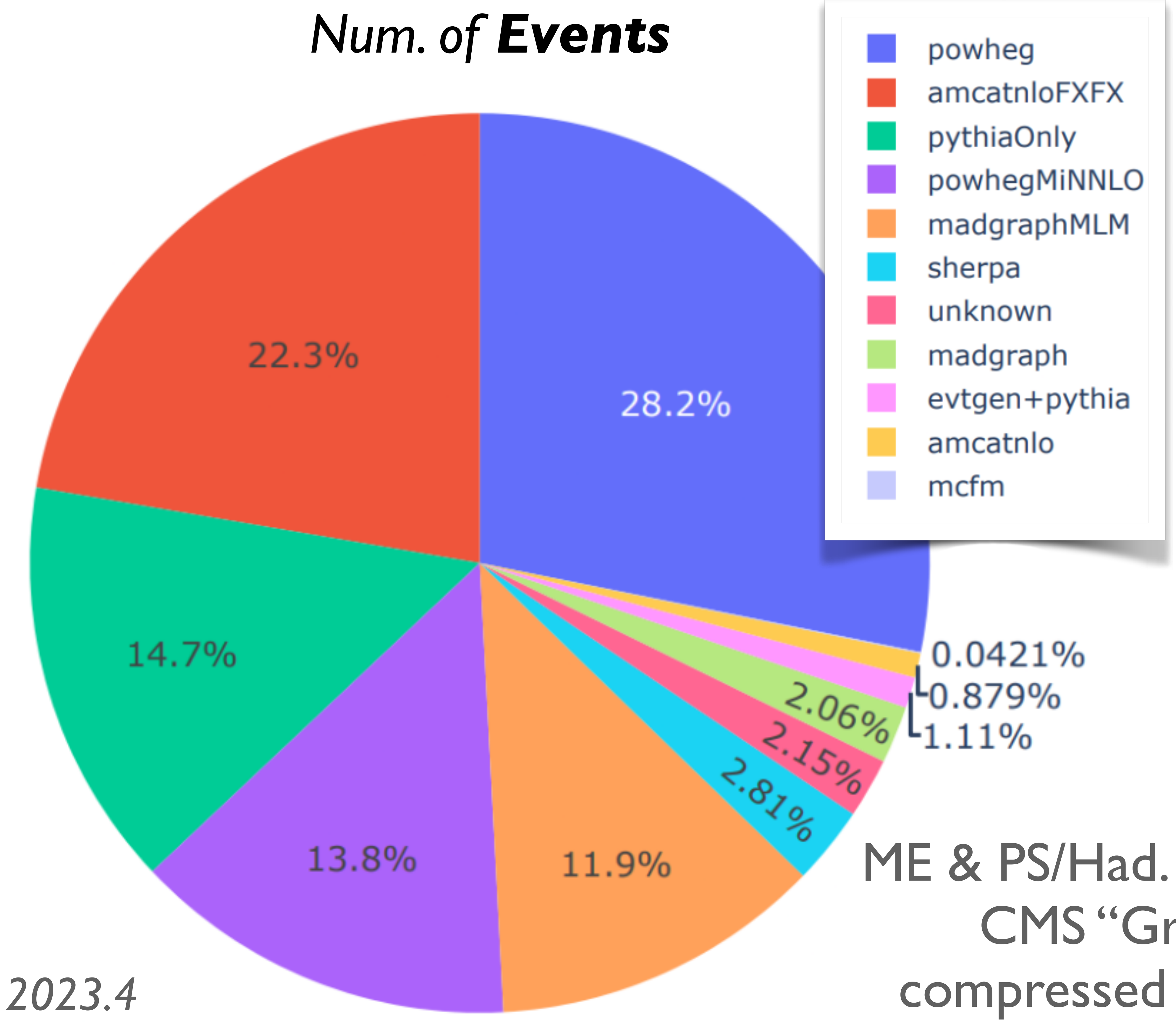
Outline (Plan)

- Current CMS Generator status
- Development in progress
  - Algorithmic improvement
  - Workflow improvement
  - Preparation for new computing infrastructure
- Summary

CMS Generator Usage: A glance from Legacy Run2 dataset

Generator (***Matrix Element*** modeling) usage breakdown based on legacy Run2 dataset
Pythia8 mostly chosen for ***parton shower*** and ***hadronization***


Num. of ***Requests***


Num. of ***Events***

Created by 2023.4

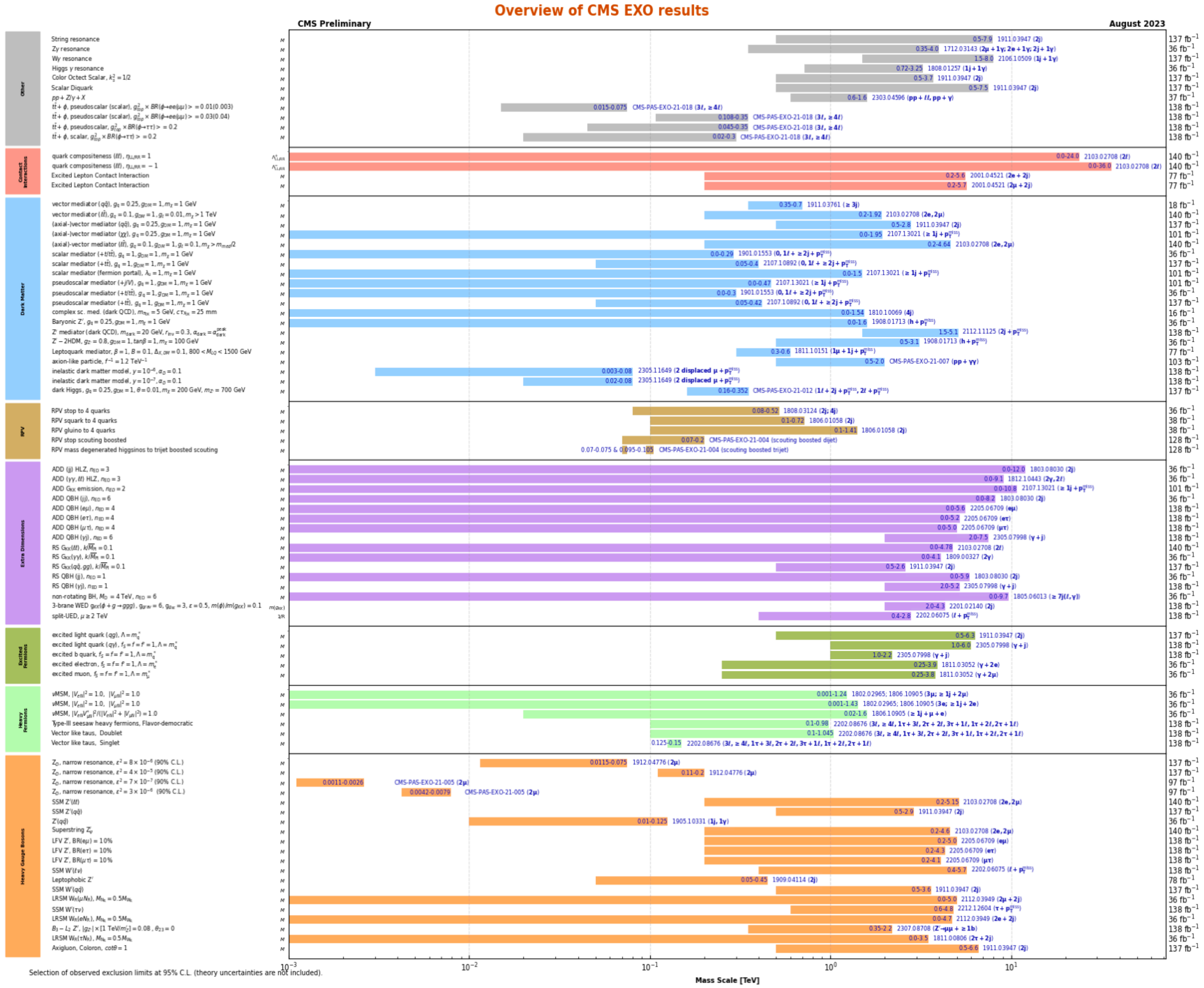
ME & PS/Had. factorization:
CMS "Gridpack":
compressed tarball with
precompiled ME grids

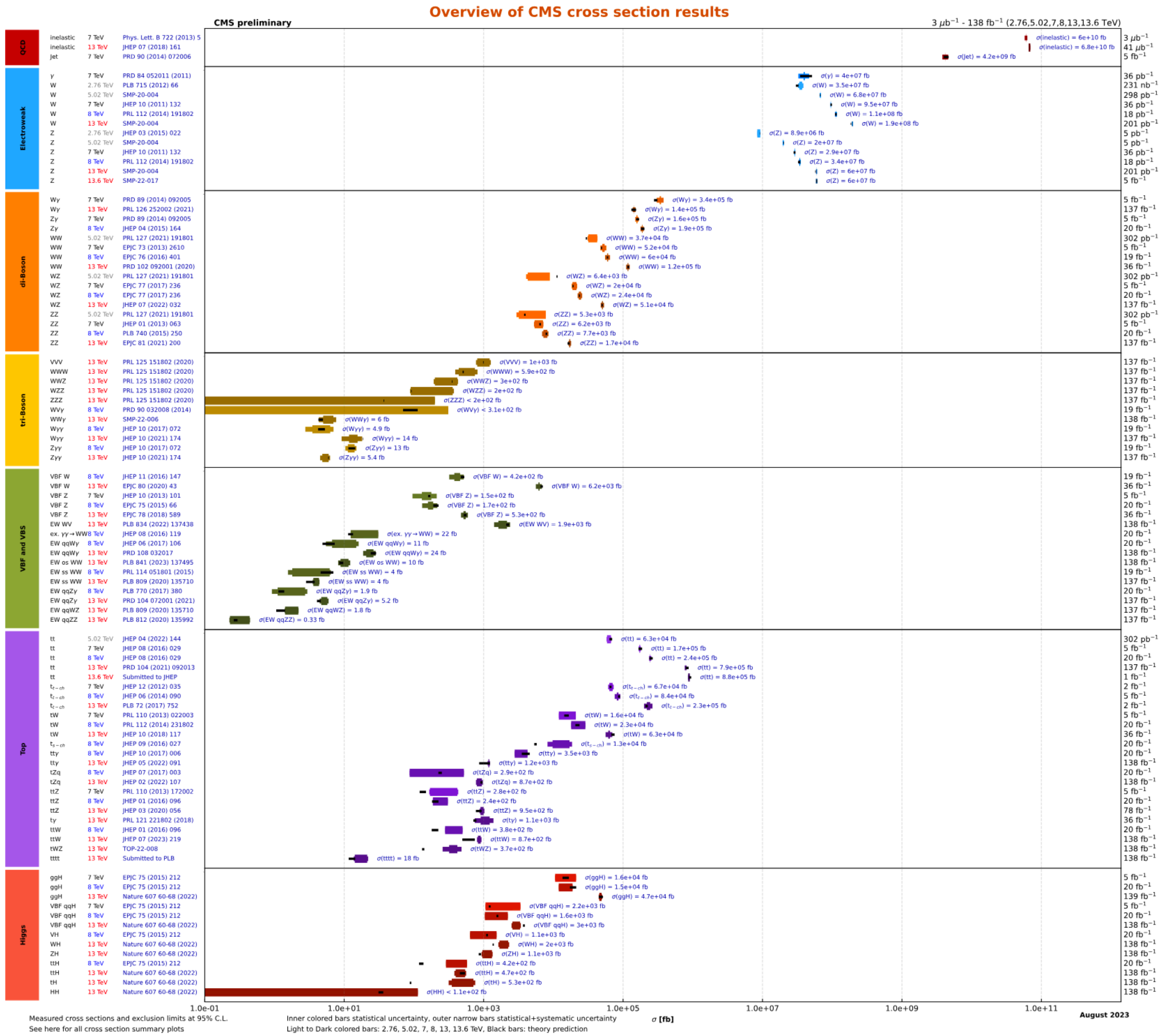
Benefit a lot from the convenience of MadGraph!
More events calls for more sophisticated modeling (high order, jet merging…)!

CMS HAS A **RICH** PHYSICS PROGRAM

More than 10 orders of magnitude SM cross section coverage!

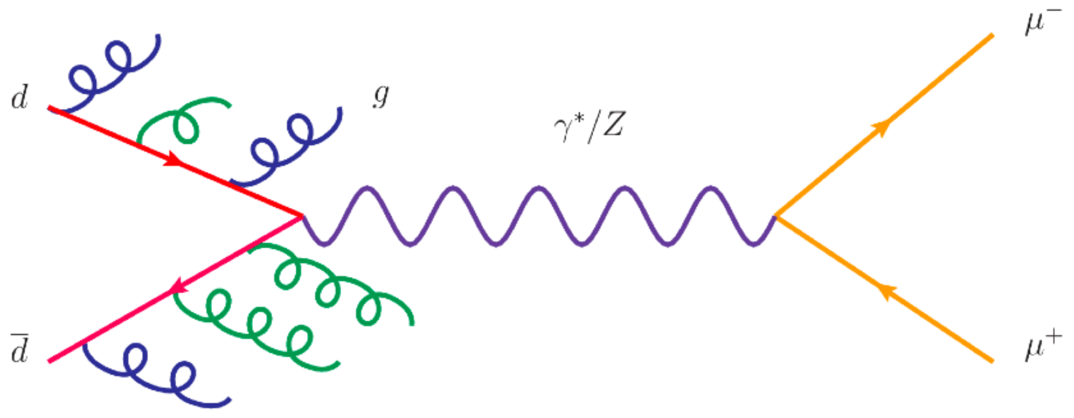
Cherry-picked CMS BSM Results



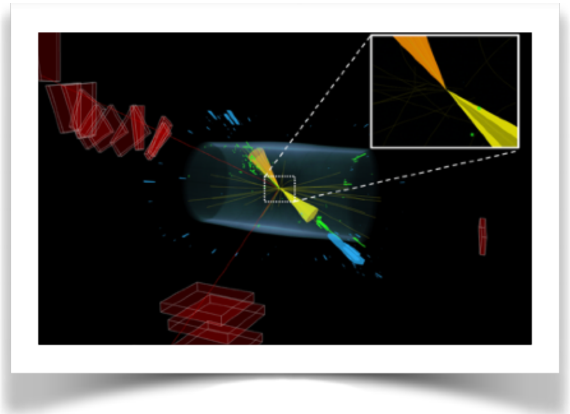


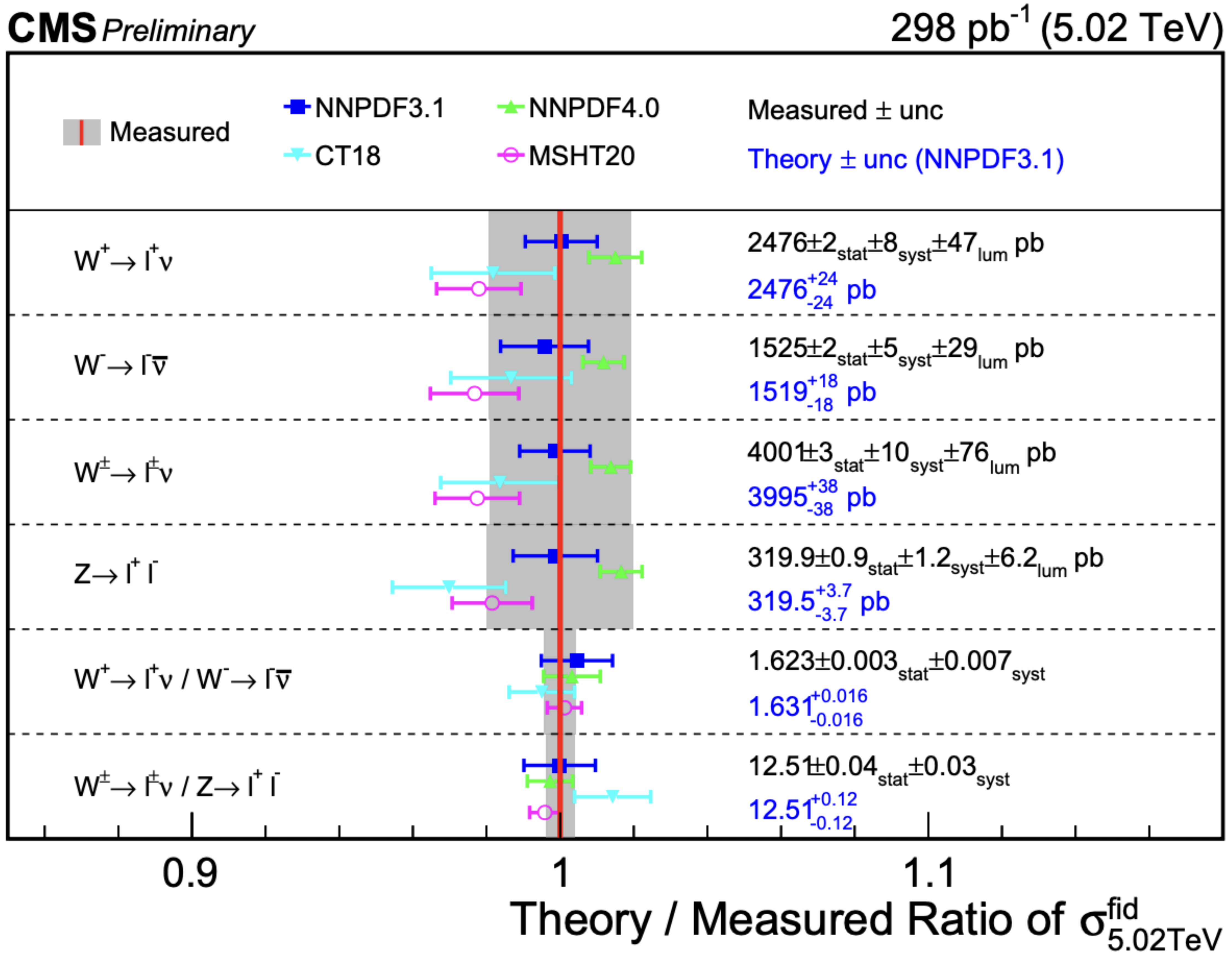
And generator usage is a **crucial** part to all of them!

GENERATOR IS CRUCIAL! TO CMS PHYSICS PROGRAM



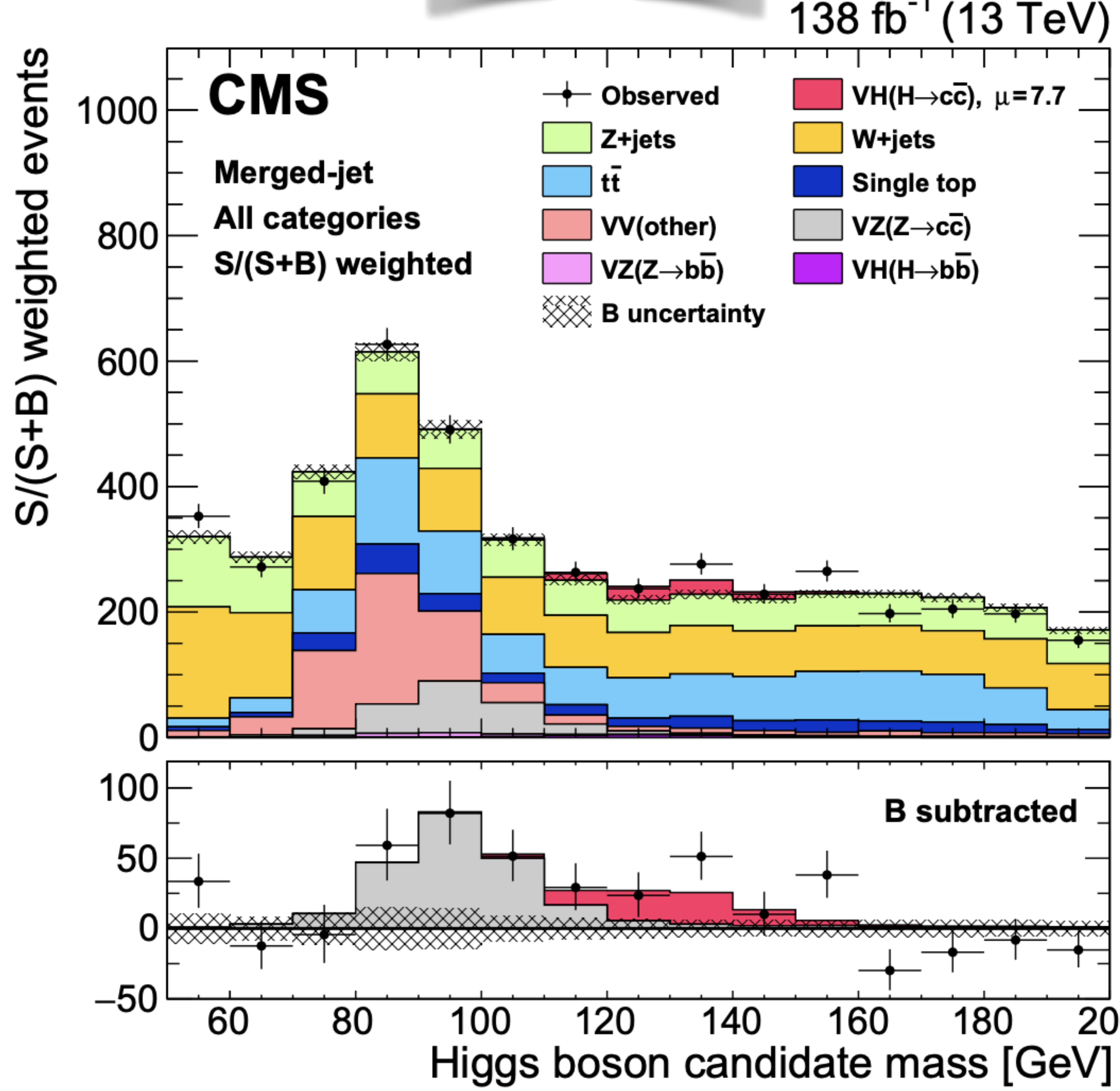
V(+jets) modeling as an example





CMS Preliminary
298 pb^{-1} (5.02 TeV)
Measured
NNPDF3.1
NNPDF4.0
CT18
MSHT20
Measured ± unc
Theory ± unc (NNPDF3.1)
$W^+\to l^+\nu$
$2476\pm2_{stat}\pm8_{syst}\pm47_{lum}$ pb
2476^{+24}_{-24} pb
$W^-\to l^-\bar{\nu}$
$1525\pm2_{stat}\pm5_{syst}\pm29_{lum}$ pb
1519^{+18}_{-18} pb
$W^\pm\to l^\pm\nu$
$4001\pm3_{stat}\pm10_{syst}\pm76_{lum}$ pb
3995^{+38}_{-38} pb
$Z\to l^+l^-$
$319.9\pm0.9_{stat}\pm1.2_{syst}\pm6.2_{lum}$ pb
$319.5^{+3.7}_{-3.7}$ pb
$W^+\to l^+\nu\,/\,W^-\to l^-\bar{\nu}$
$1.623\pm0.003_{stat}\pm0.007_{syst}$
$1.631^{+0.016}_{-0.016}$
$W^\pm\to l^\pm\nu\,/\,Z\to l^+l^-$
$12.51\pm0.04_{stat}\pm0.03_{syst}$
$12.51^{+0.12}_{-0.12}$
0.9
1
1.1
Theory / Measured Ratio of $\sigma^{fid}_{5.02TeV}$

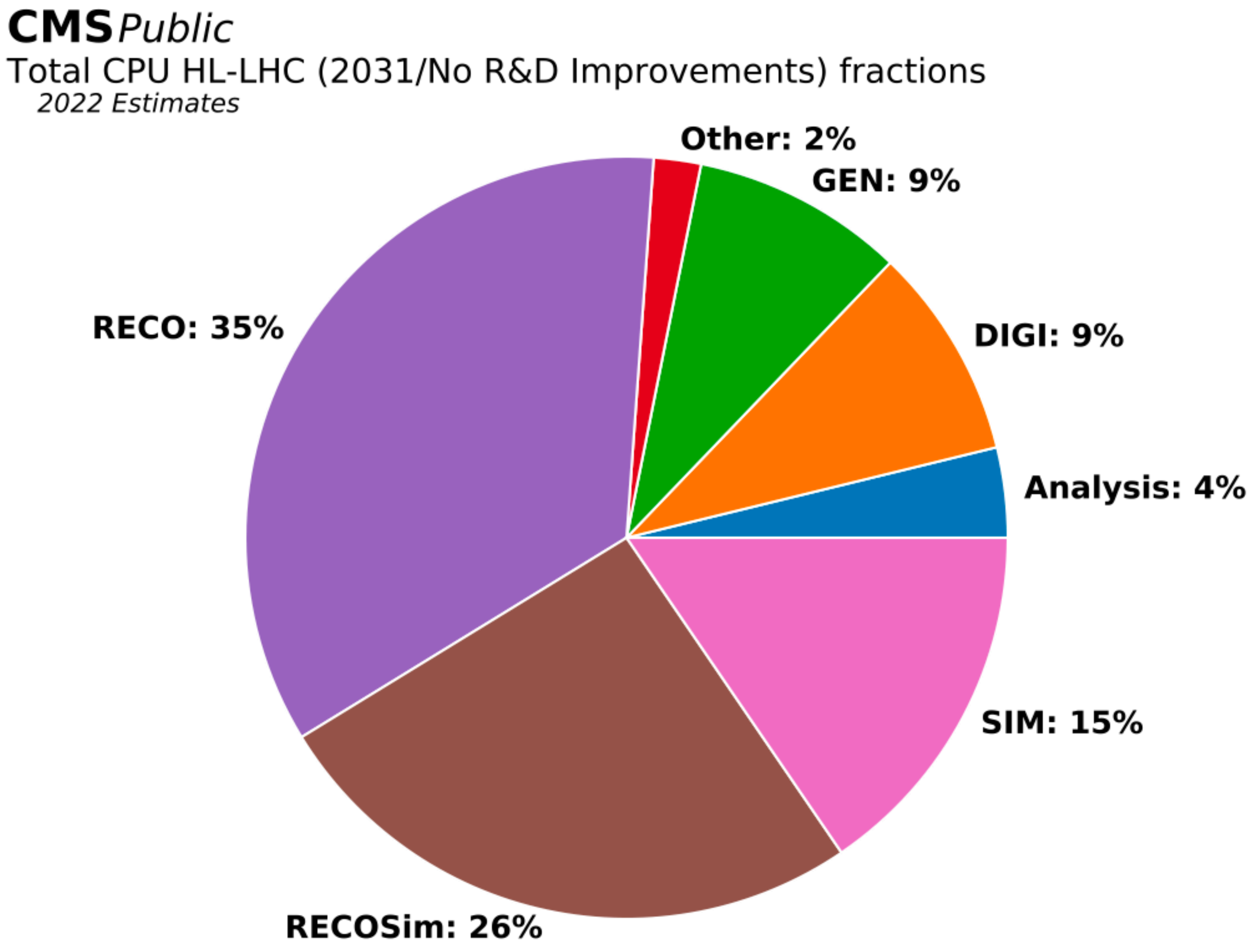
Subpercent precision era of Electroweak measurements



V+Jets production is important background for CMS Higgs program

GENERATOR IS CRUCIAL! AND CHALLENGES AHEAD

HL-LHC: order of magnitude higher integrated luminosity → order of magnitude higher required MC statistics!



Deeper discussion @ Liz's talk!

- Will discuss here our homework for *today*
  - Algorithmic improvement
    - Negative weight elimination
    - Heavy I/O issue for production
    - Phase space biasing and filtering
  - Workflow reorganization and optimization
    - Automized and centralized gridpack production
  - New infrastructure test
    - MG4GPU test

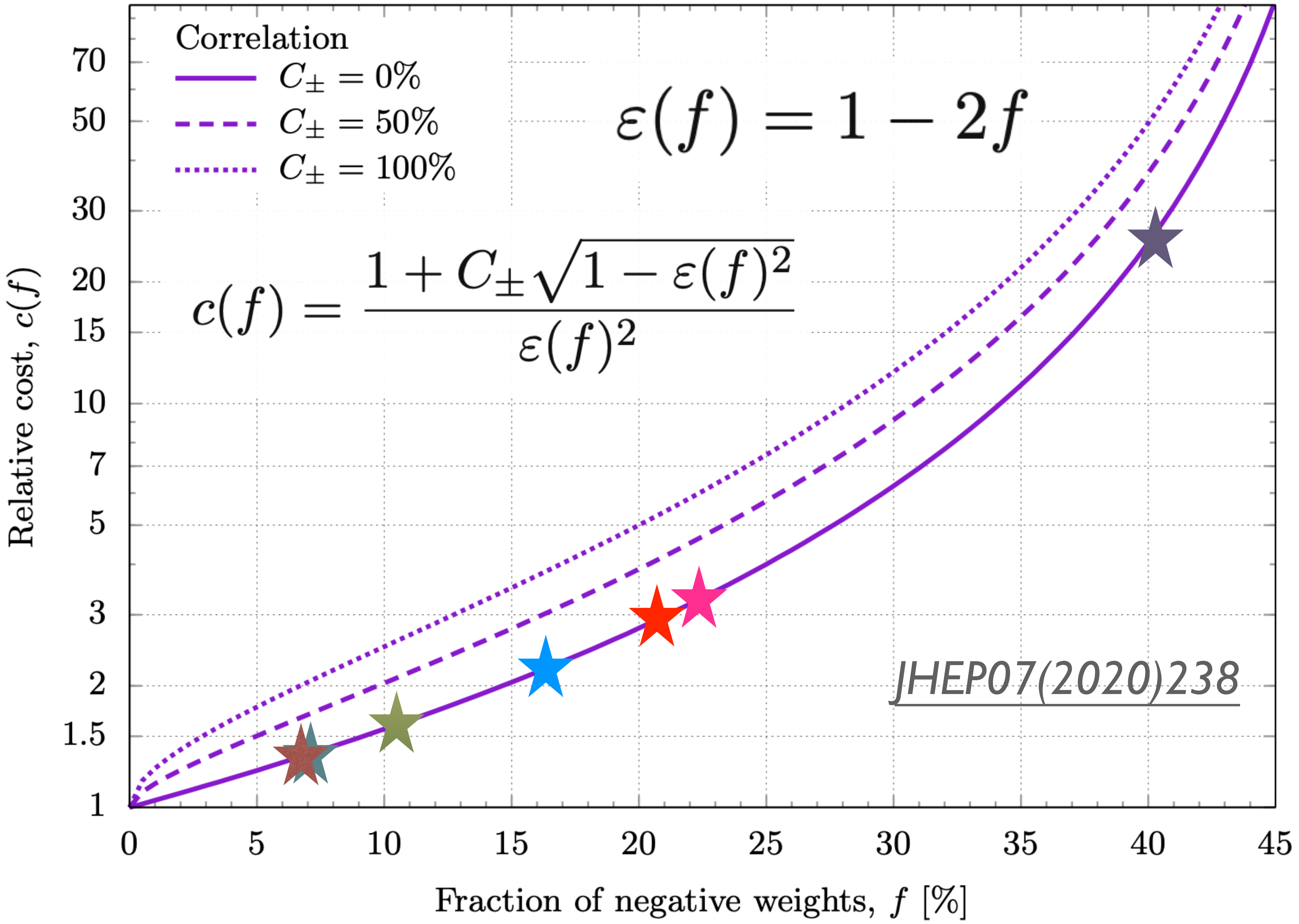
…

NEGATIVE WEIGHT: PRICE TO PAY FOR NLO

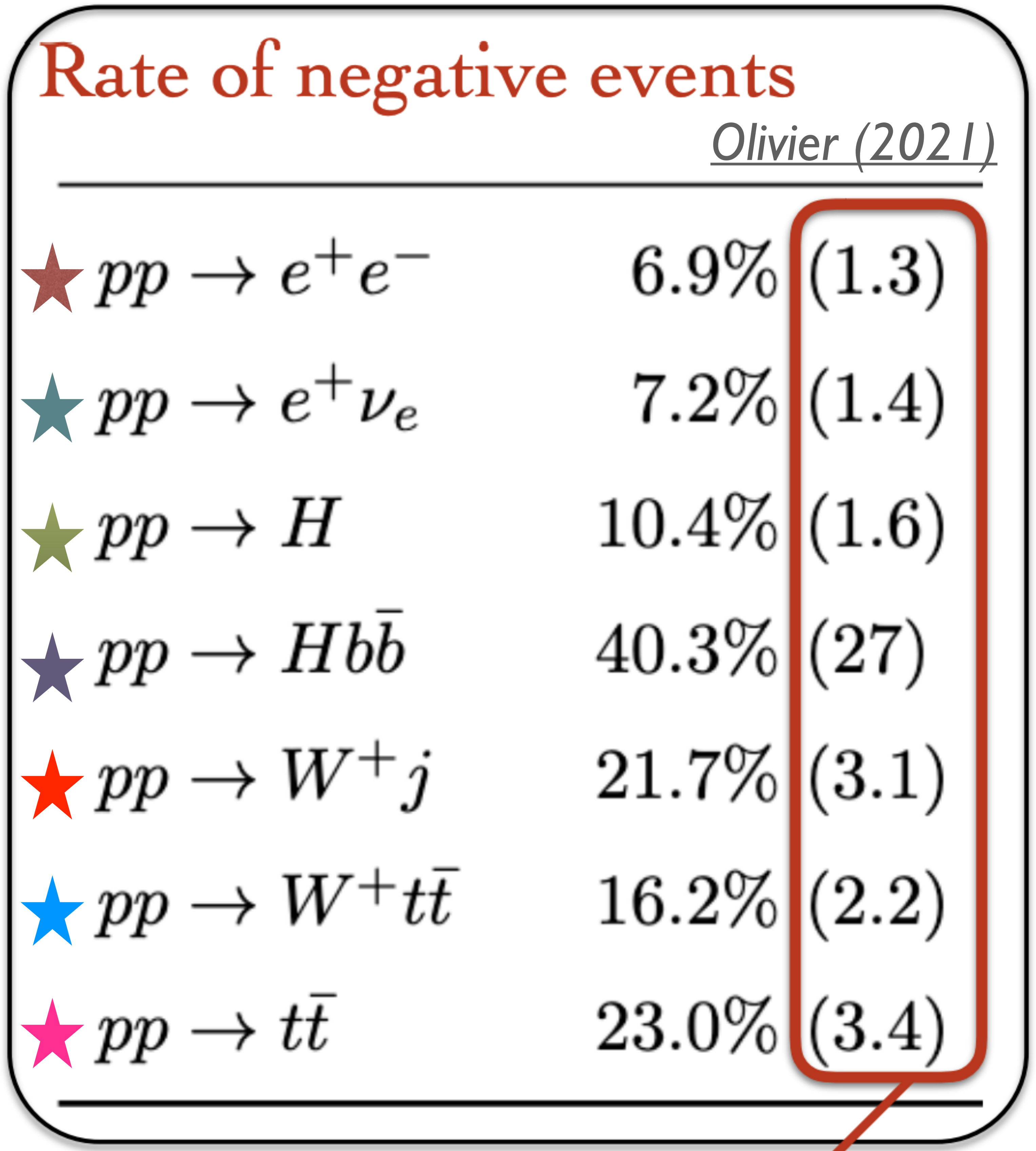
- High order calculation needed for *today*
  - NLO calculation includes real emission and virtual correction
  - Substraction needed for matching to patron shower → Negative weight introduced

Relative cost:
Ratio b/w number
of events with
negative weights
to that from
positive weights
only generation
with the same
statistical power

The Lower
The Better!



JHEP07(2020)238



Rate of negative events

Olivier (2021)

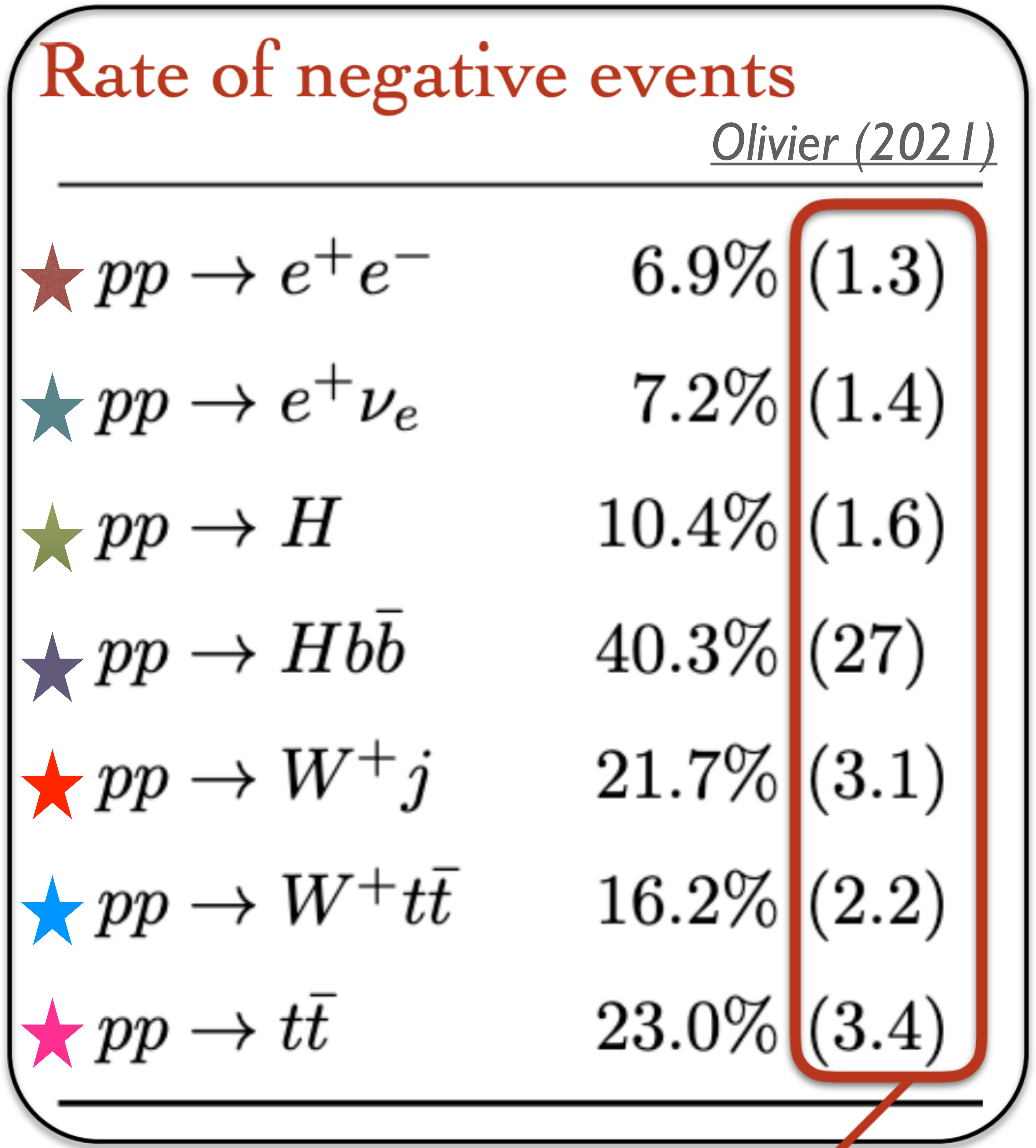
★ $pp \to e^+e^-$	6.9%	(1.3)
★ $pp \to e^+\nu_e$	7.2%	(1.4)
★ $pp \to H$	10.4%	(1.6)
★ $pp \to Hb\bar{b}$	40.3%	(27)
★ $pp \to W^+j$	21.7%	(3.1)
★ $pp \to W^+t\bar{t}$	16.2%	(2.2)
★ $pp \to t\bar{t}$	23.0%	(3.4)

Cost In sample size

$$c(f) = \frac{1}{(1-2f)^2}$$

NEGATIVE WEIGHT: PRICE TO PAY FOR NLO

- High order calculation needed for *today*
  - NLO calculation includes real emission and virtual correction
  - Substraction needed for matching to Parton shower → Negative weight introduced
  - POWHEG has its way of eliminating negative weights!
  - But we still want to leverage the flexible generation from aMC@NLO (and FxFx)



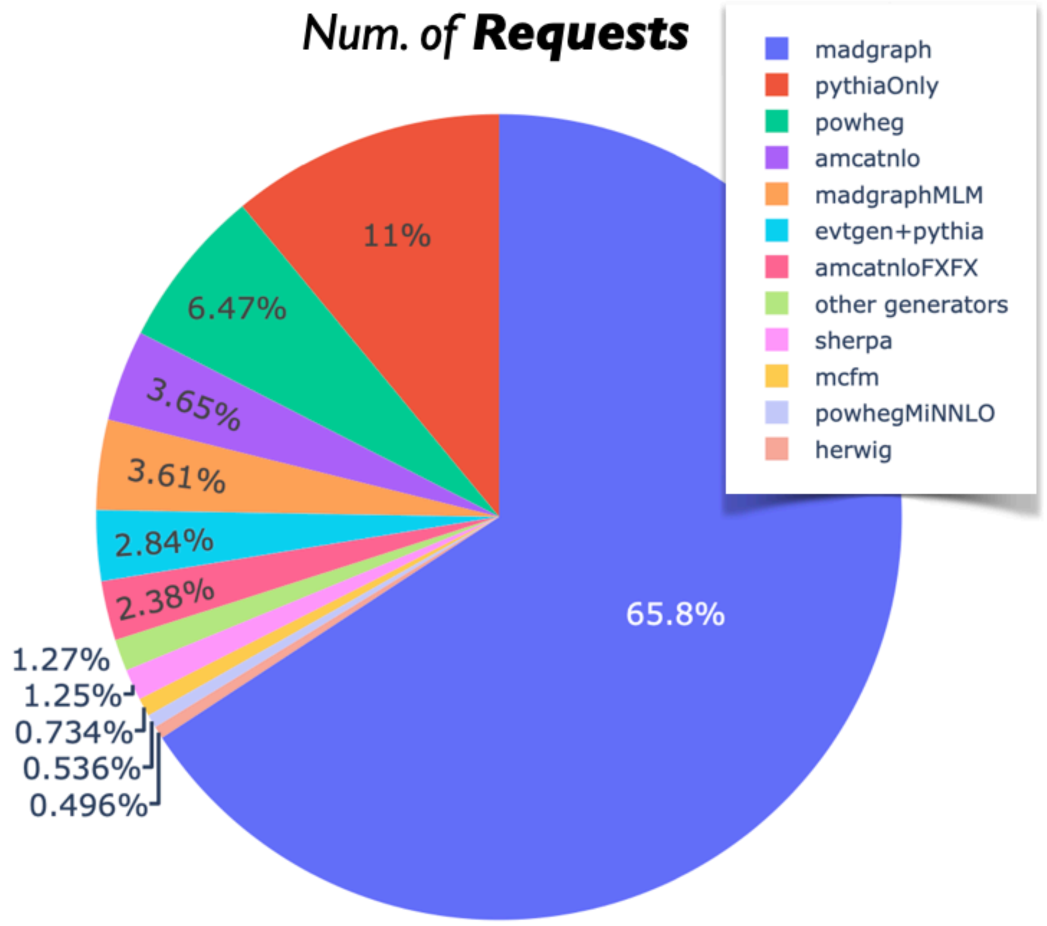
Rate of negative events

Olivier (2021)

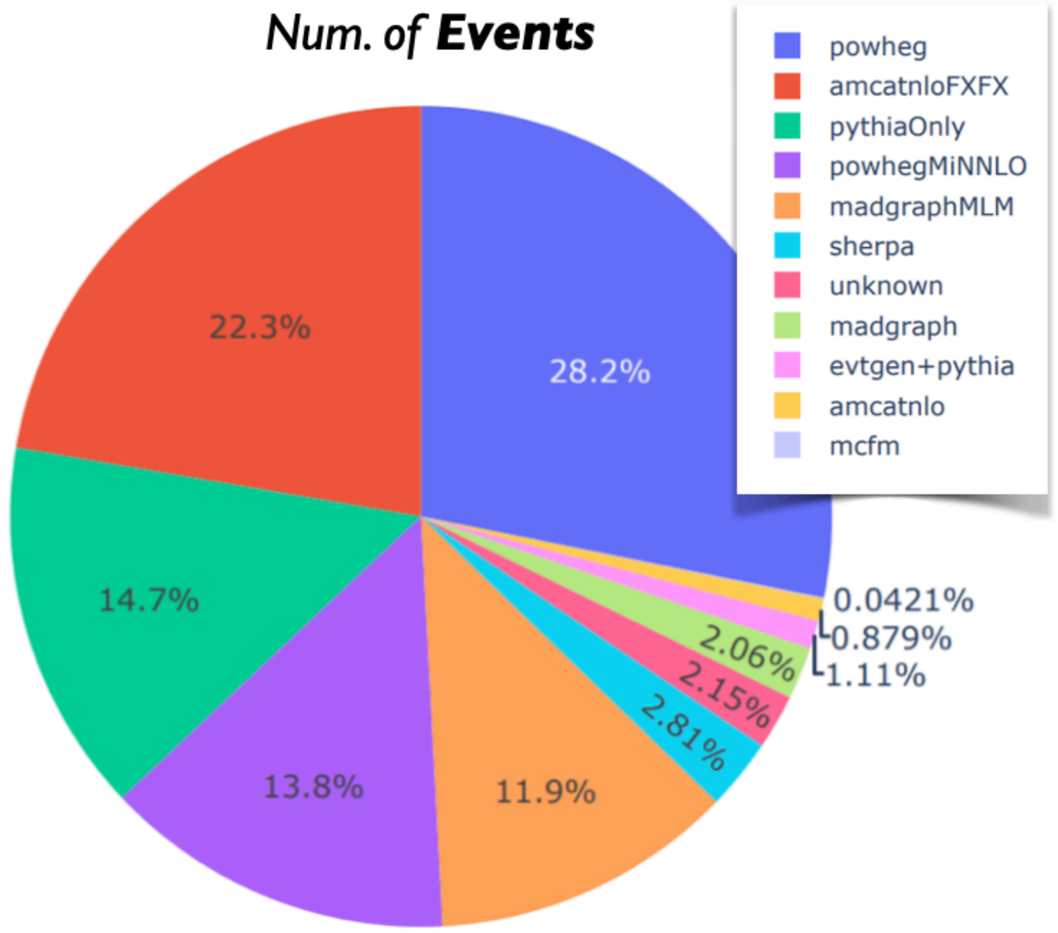
★ $pp \to e^+e^-$	6.9%	(1.3)
★ $pp \to e^+\nu_e$	7.2%	(1.4)
★ $pp \to H$	10.4%	(1.6)
★ $pp \to Hb\bar{b}$	40.3%	(27)
★ $pp \to W^+j$	21.7%	(3.1)
★ $pp \to W^+t\bar{t}$	16.2%	(2.2)
★ $pp \to t\bar{t}$	23.0%	(3.4)

Cost In sample size

$$c(f) = \frac{1}{(1-2f)^2}$$



Num. of **Requests**



Num. of **Events**

NEGATIVE WEIGHT: ELIMINATION WITH PHYSICS CONSIDERATION

aMC@NLO follows MC@NLO matching prescription

$$d\sigma^{(\mathbb{H})} = d\sigma^{(\mathrm{NLO},E)} - d\sigma^{(\mathrm{MC})}\,,$$

$$d\sigma^{(\mathbb{S})} = d\sigma^{(\mathrm{MC})} + \sum_{\alpha=S,C,SC} d\sigma^{(\mathrm{NLO},\alpha)}\,.$$

Negative weights originate from both H and S terms

Introducing "Delta" factor to suppress negative weights

$$d\sigma^{(\Delta,\mathbb{H})} = \left(d\sigma^{(\mathrm{NLO},E)} - d\sigma^{(\mathrm{MC})}\right)\Delta\,,$$

$$d\sigma^{(\Delta,\mathbb{S})} = d\sigma^{(\mathrm{MC})}\Delta + \sum_{\alpha=S,C,SC} d\sigma^{(\mathrm{NLO},\alpha)} + d\sigma^{(\mathrm{NLO},E)}(1-\Delta)\,.$$

JHEP07(2020)238

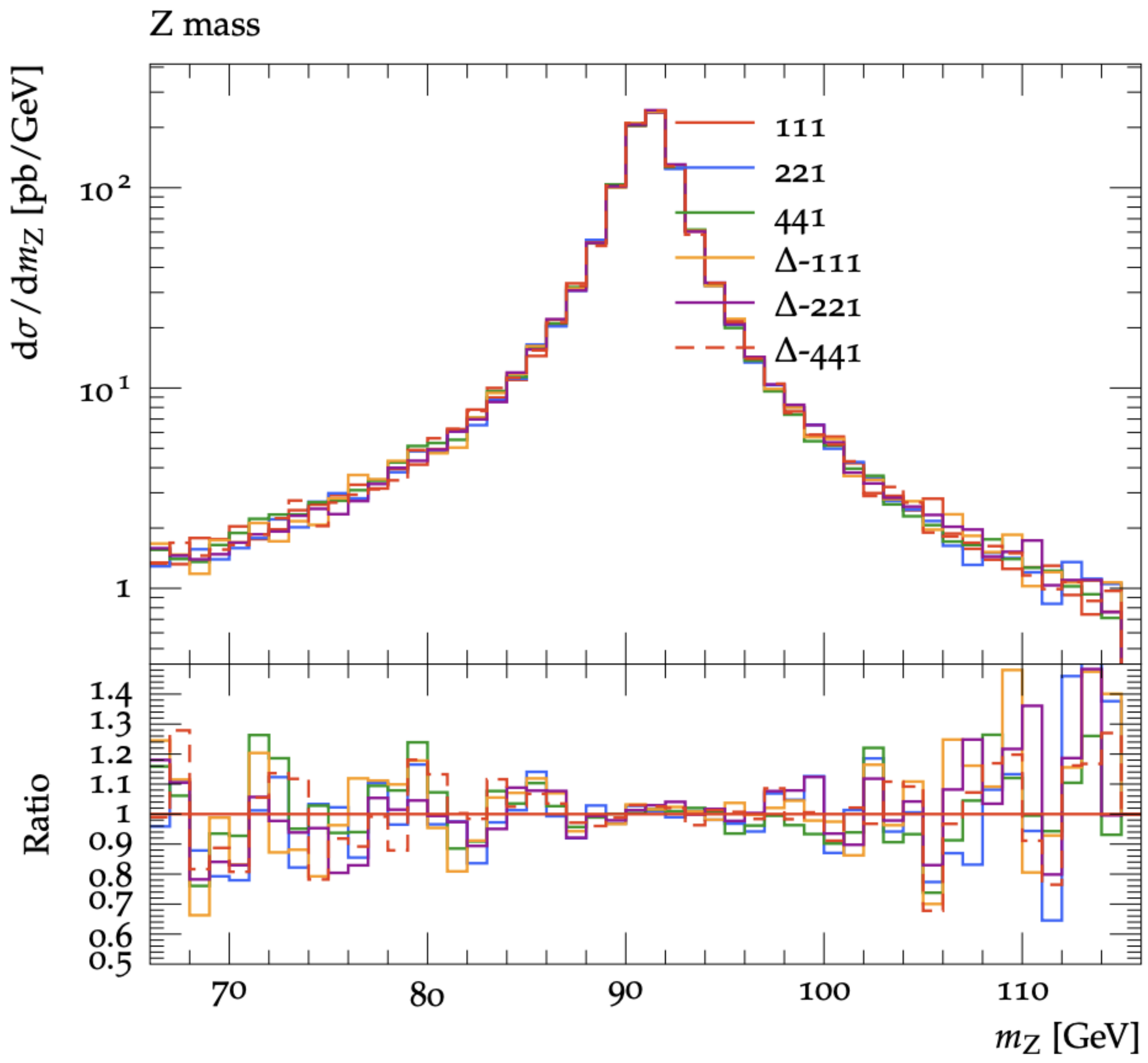
MC@NLO-Delta prescription

Folding needed for further suppression

Ongoing validation with DY(ee)

111/221/441 → folding prescription

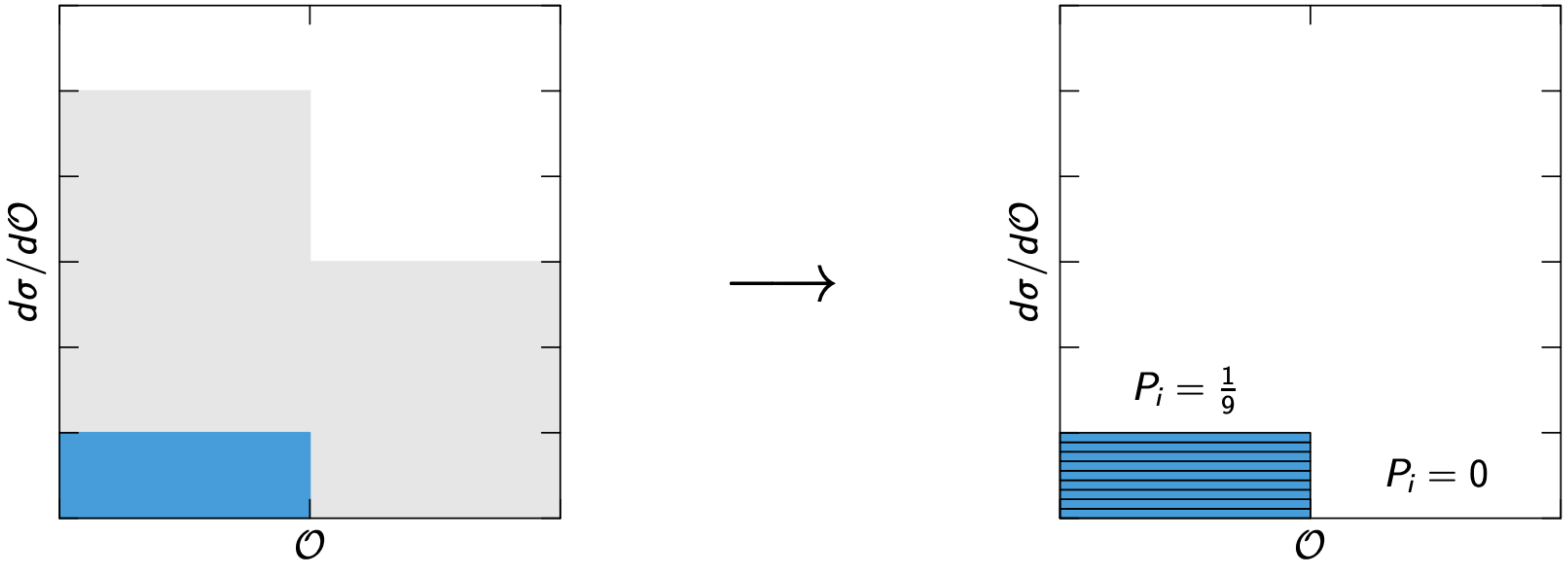
w/o(w/) Delta-: aMC@NLO(-Delta)



Next step: CMS integration and further test!

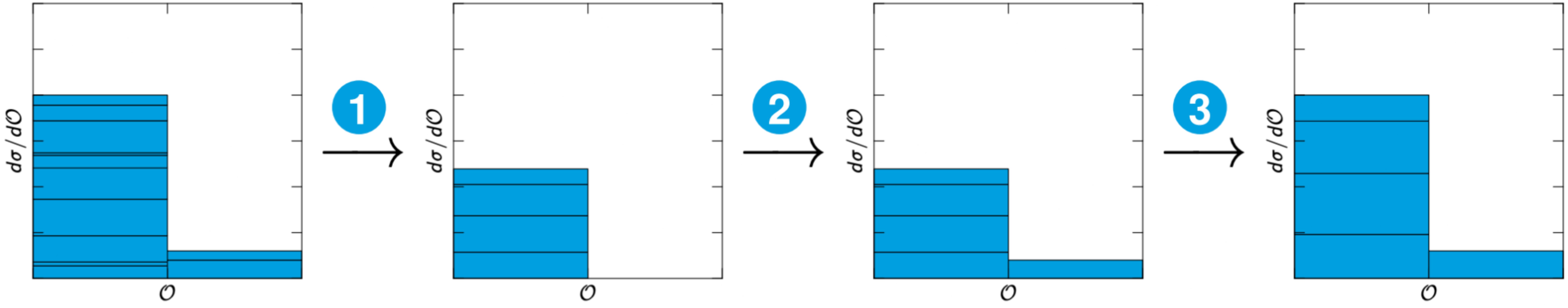
NEGATIVE WEIGHT: ***“DISTRIBUTION-DRIVEN”*** PROPOSALS

- Negative weights are introduced for correct predictions
    - Trustful distribution when considering negative weights
    - Reweighting of events possible!



NEGATIVE WEIGHT: ***“DISTRIBUTION-DRIVEN”*** PROPOSALS

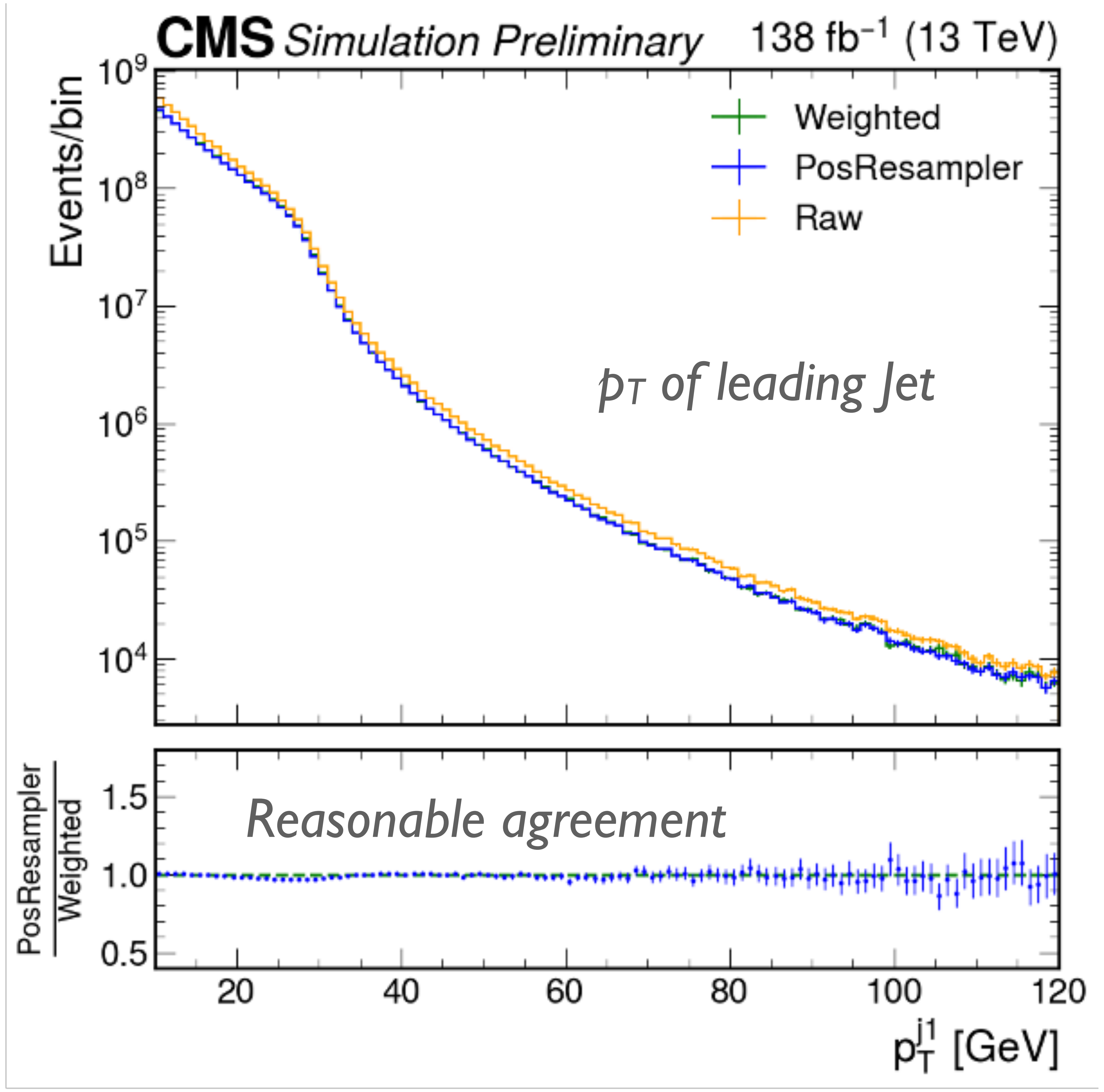
- Negative weights are introduced for correct predictions
  - Trustful distribution when considering negative weights
  - Reweighting of events possible!
  - #events (for SIM) **reduced** then from unweighting (resampling)



NEGATIVE WEIGHT: ***“DISTRIBUTION-DRIVEN”*** PROPOSALS

- Negative weights are introduced for correct predictions
  - Trustful distribution when considering negative weights
  - Reweighting of events possible!
  - #events (for SIM) **reduced** then from unweighting (resampling) *epjc/s10052-020-08548-w*
  - Positive resampler: simply using histograms!
- Implemented with CMS workflow, now under validation

NLO W production
Reweighted with GEN W boson p_T



p_T of leading Jet

Reasonable agreement

NEGATIVE WEIGHT: *“DISTRIBUTION-DRIVEN”* PROPOSALS

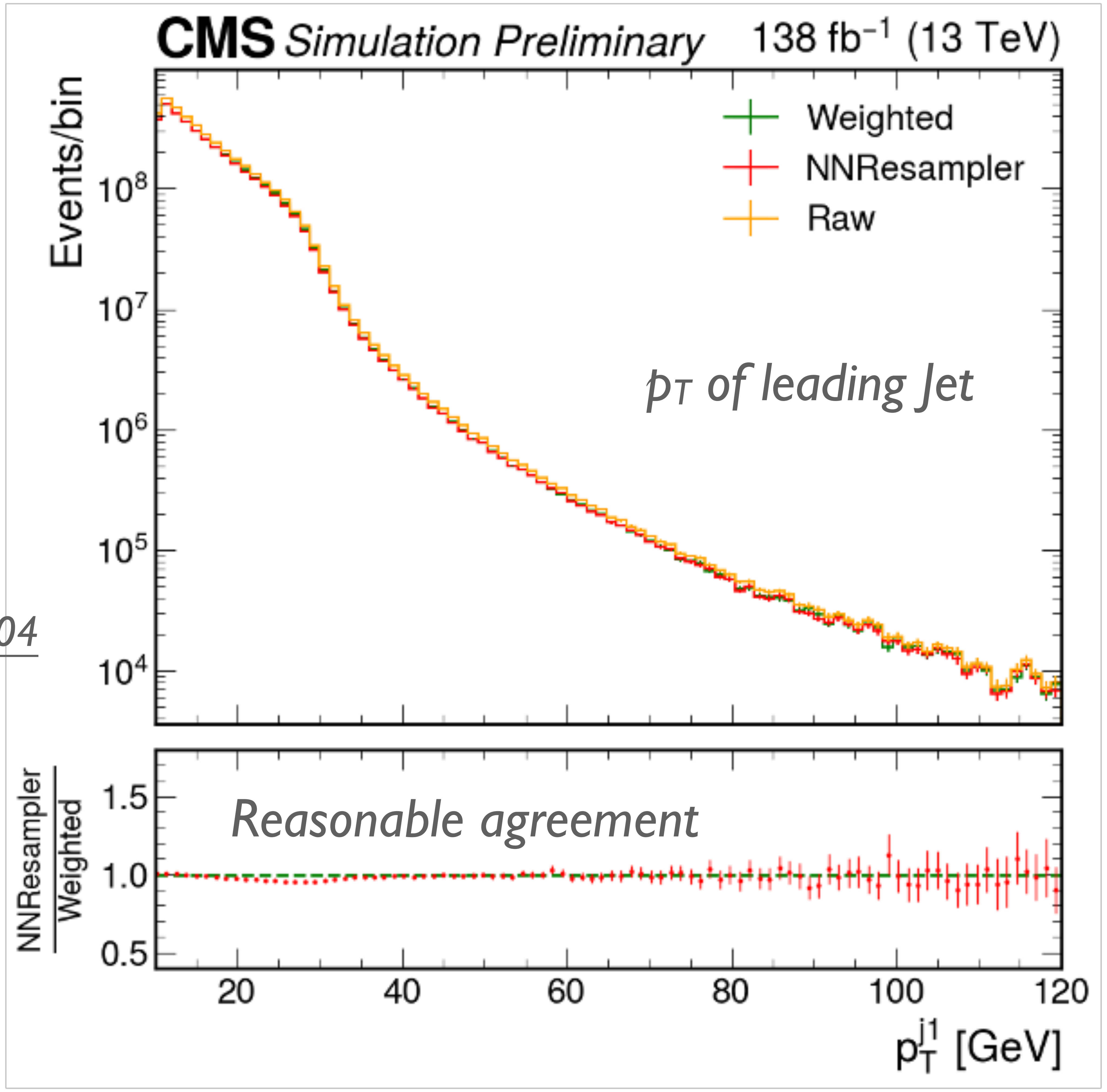
- Negative weights are introduced for correct predictions
  - Trustful distribution when considering negative weights
  - Reweighting of events possible!
  - #events (for SIM) **reduced** then from unweighting (resampling)
  - A **neural network** can be used for reweighting → predicting per-event weight
  - Done via a special loss function!

PhysRevD.102.076004

$$\mathcal{L}[g] = -\sum_{i=1}^{N} w_i \log g(x_i) - \sum_{i=1}^{N} \log\left(1 - g(x_i)\right)$$

- Now under validation within CMS!

NLO W production

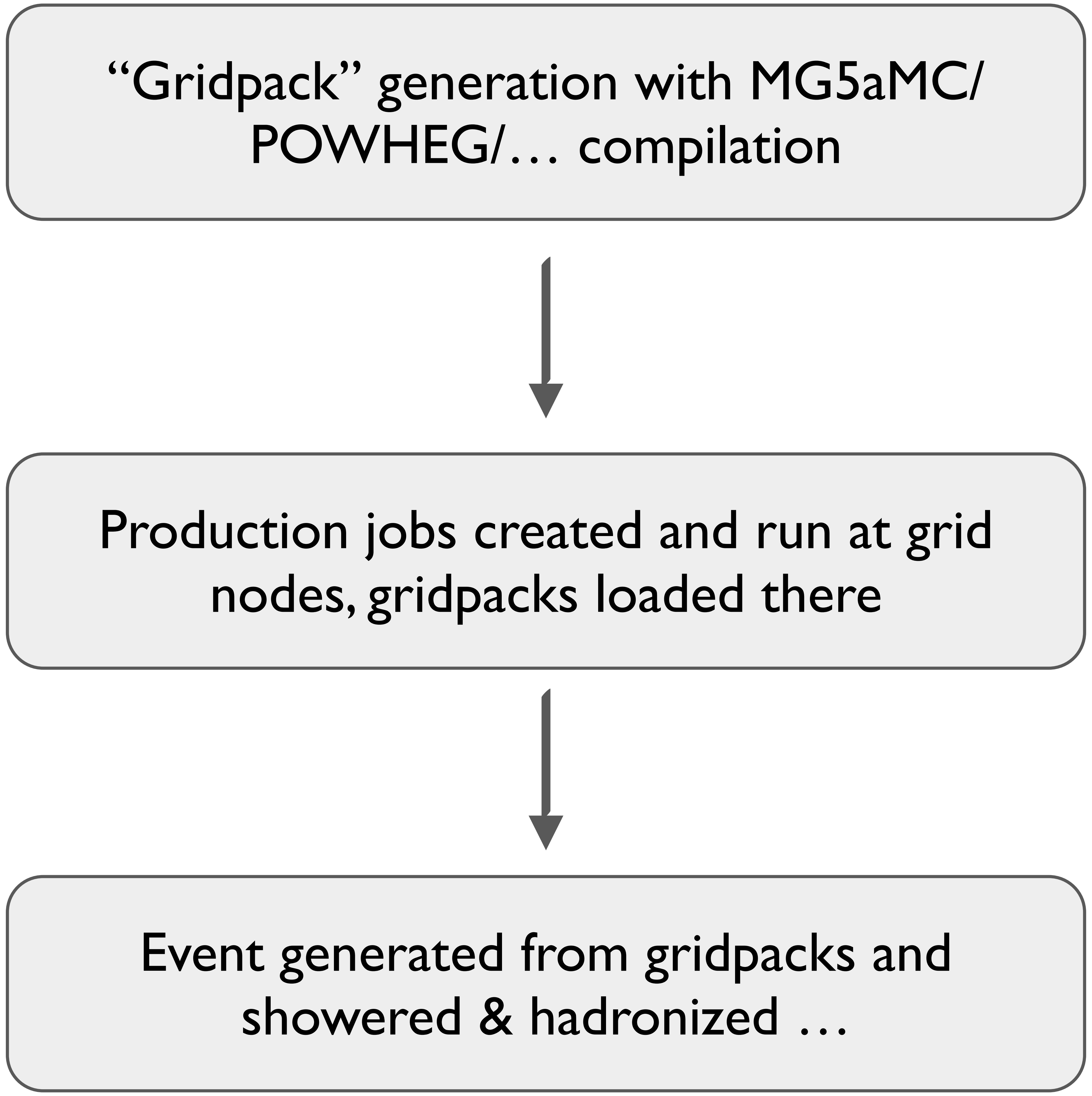


NN Backbone: 1D PCNN (DeepAK8-like)

HEAVY GRIDPACK: PRICE FOR SOPHISTICATED MODELING

CMS MC Production Workflow

highly simplified version

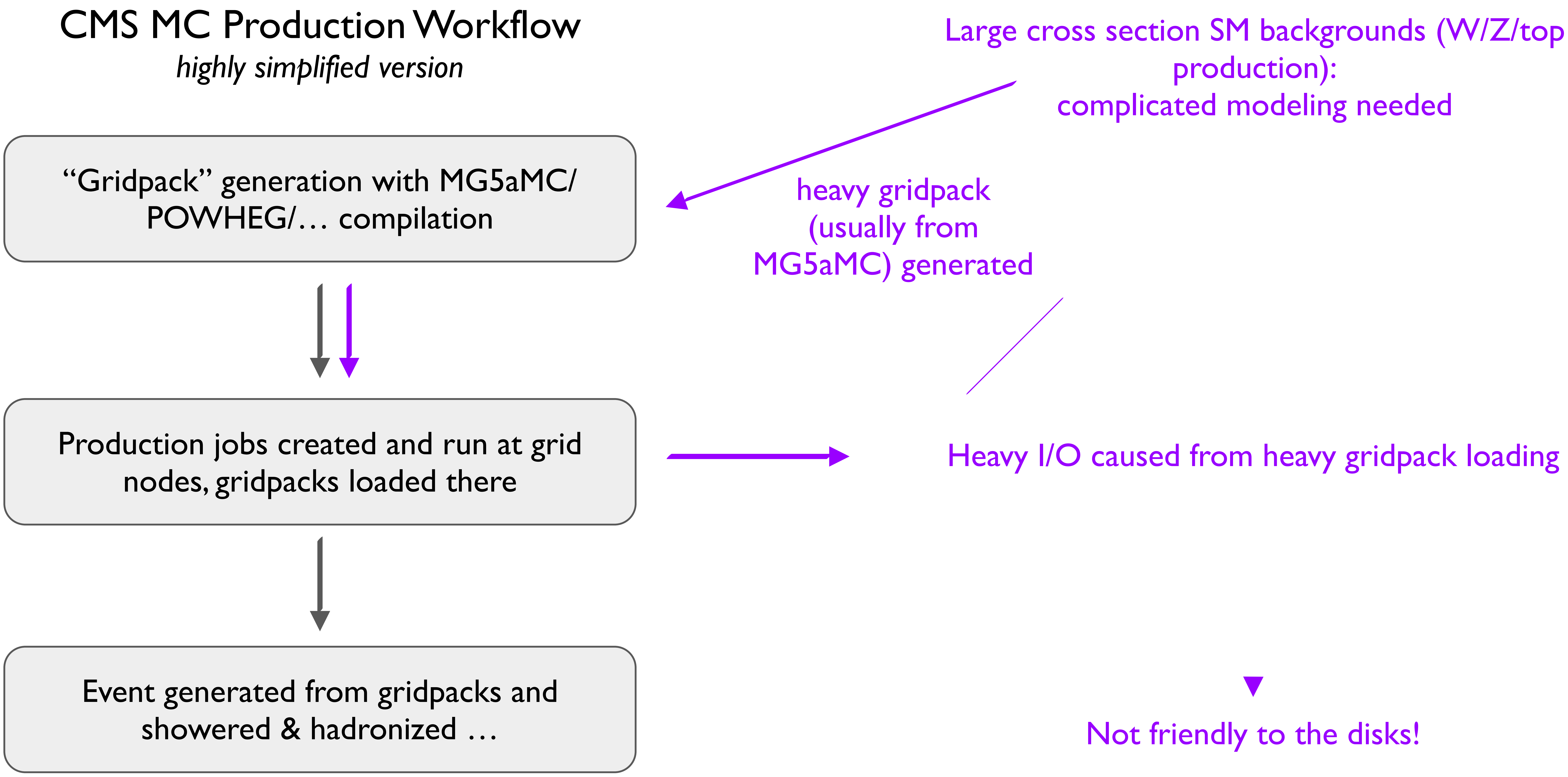


"Gridpack" generation with MG5aMC/POWHEG/… compilation

Production jobs created and run at grid nodes, gridpacks loaded there

Event generated from gridpacks and showered & hadronized …

HEAVY GRIDPACK: PRICE FOR SOPHISTICATED MODELING



CMS MC Production Workflow

highly simplified version

"Gridpack" generation with MG5aMC/ POWHEG/… compilation

Production jobs created and run at grid nodes, gridpacks loaded there

Event generated from gridpacks and showered & hadronized …

Large cross section SM backgrounds (W/Z/top production): complicated modeling needed

heavy gridpack (usually from MG5aMC) generated

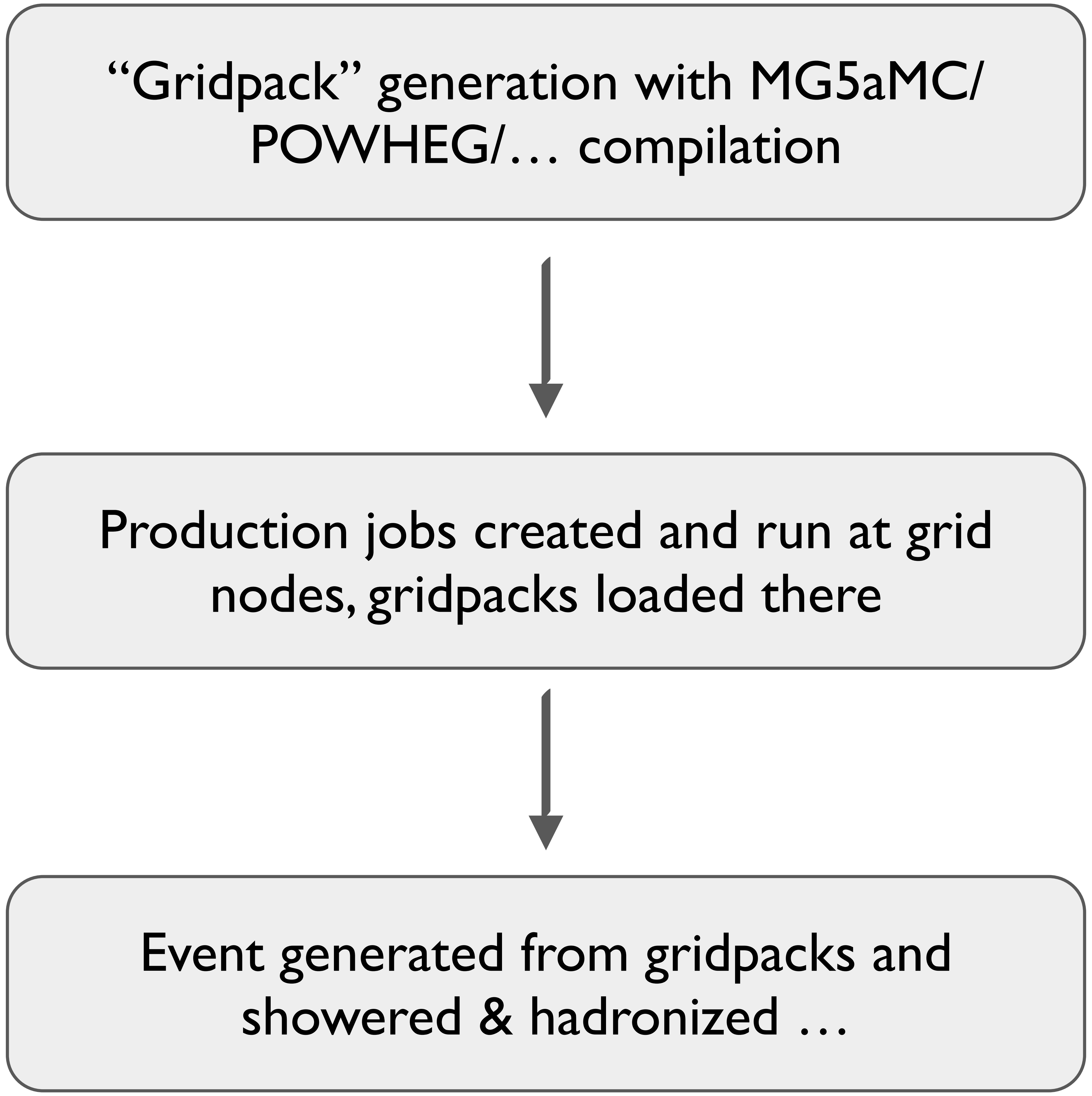
Heavy I/O caused from heavy gridpack loading

Not friendly to the disks!

HEAVY GRIDPACK: PRICE FOR SOPHISTICATED MODELING

CMS MC Production Workflow

highly simplified version



"Gridpack" generation with MG5aMC/ POWHEG/… compilation

Production jobs created and run at grid nodes, gridpacks loaded there

Event generated from gridpacks and showered & hadronized …

Example from CMS W+012j FxFx Modeling

Condition	Run3	Run2 Legacy
UFO model	loop_sm-ckm_no_b_mass	loop_sm-ckm_no_b_mas s
Size (compressed)	774M	762M
Size (uncomp.)	**14G** subprocesses **104M** MG source **Negligible** **Negligible** auxiliary files **Negligible** **Negligible**	**16G** ***x nThreads!*** **177M** **Negligible** **Negligible** **Negligible** **Negligible**

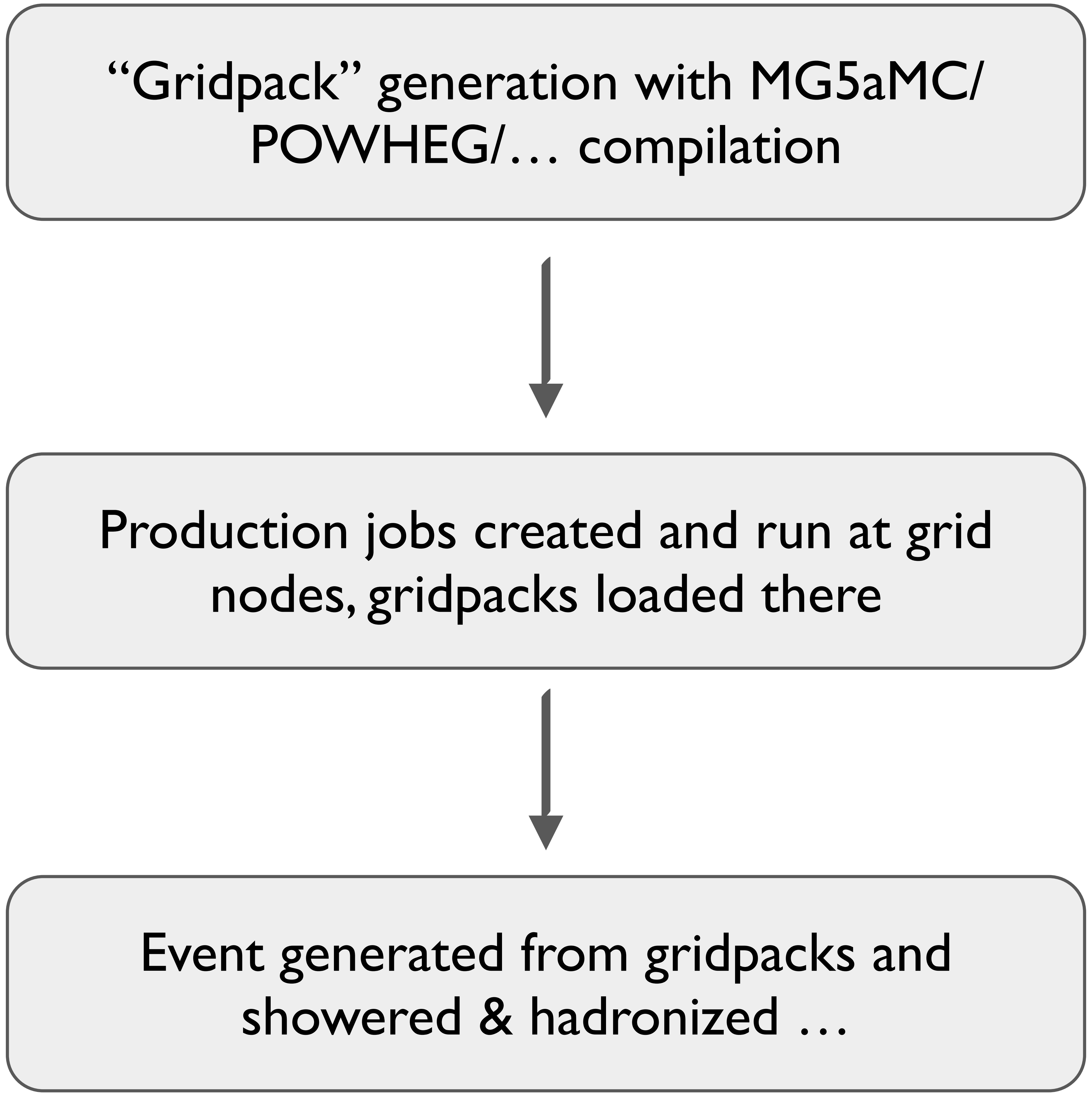
O(10 GB) I/O per thread

Not friendly to disks!

HEAVY GRIDPACK: PRICE FOR SOPHISTICATED MODELING

CMS MC Production Workflow

highly simplified version



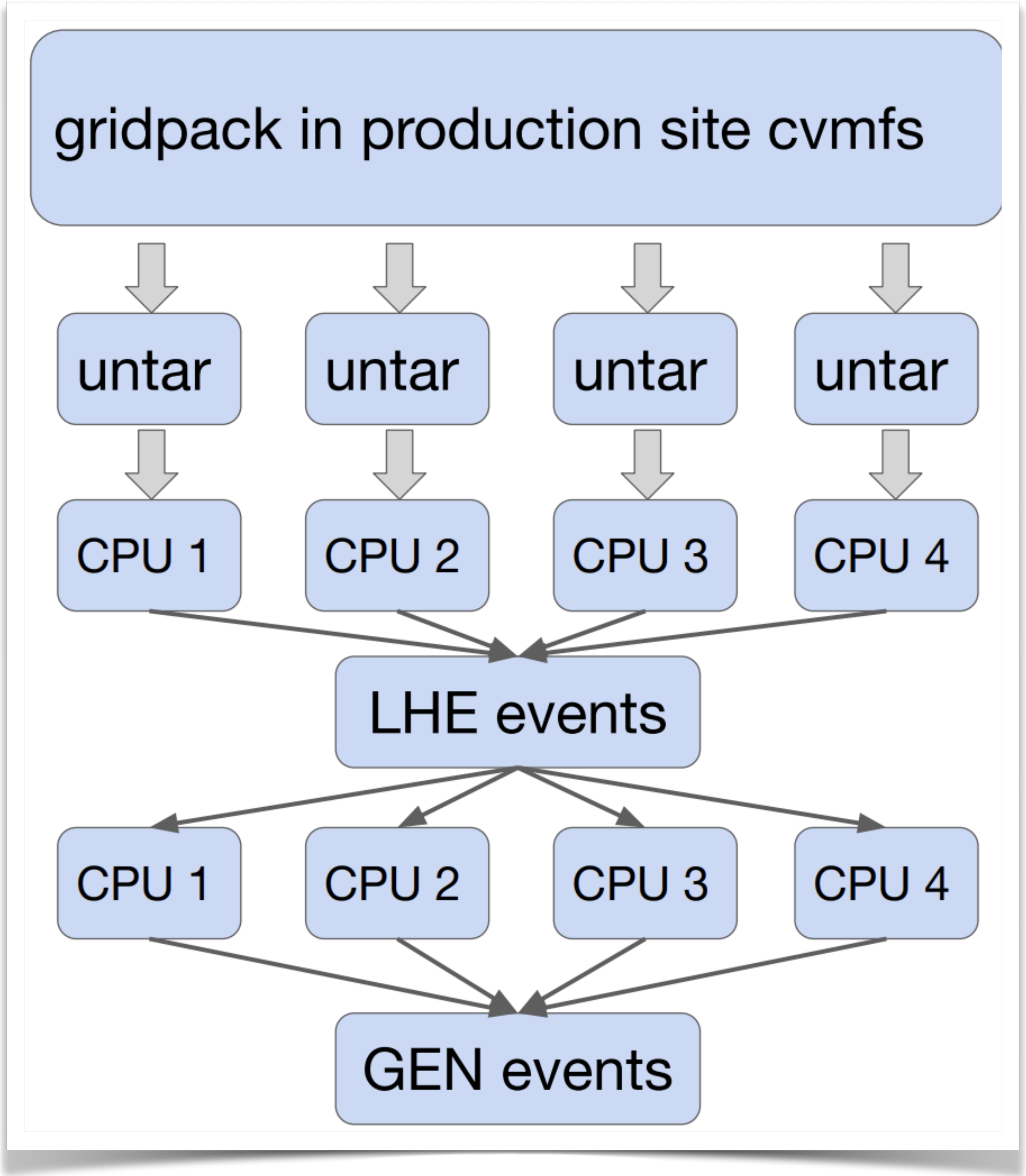
"Gridpack" generation with MG5aMC/ POWHEG/… compilation

Production jobs created and run at grid nodes, gridpacks loaded there

Event generated from gridpacks and showered & hadronized …

More practical workflow:

Multithreading via concurrent jobs

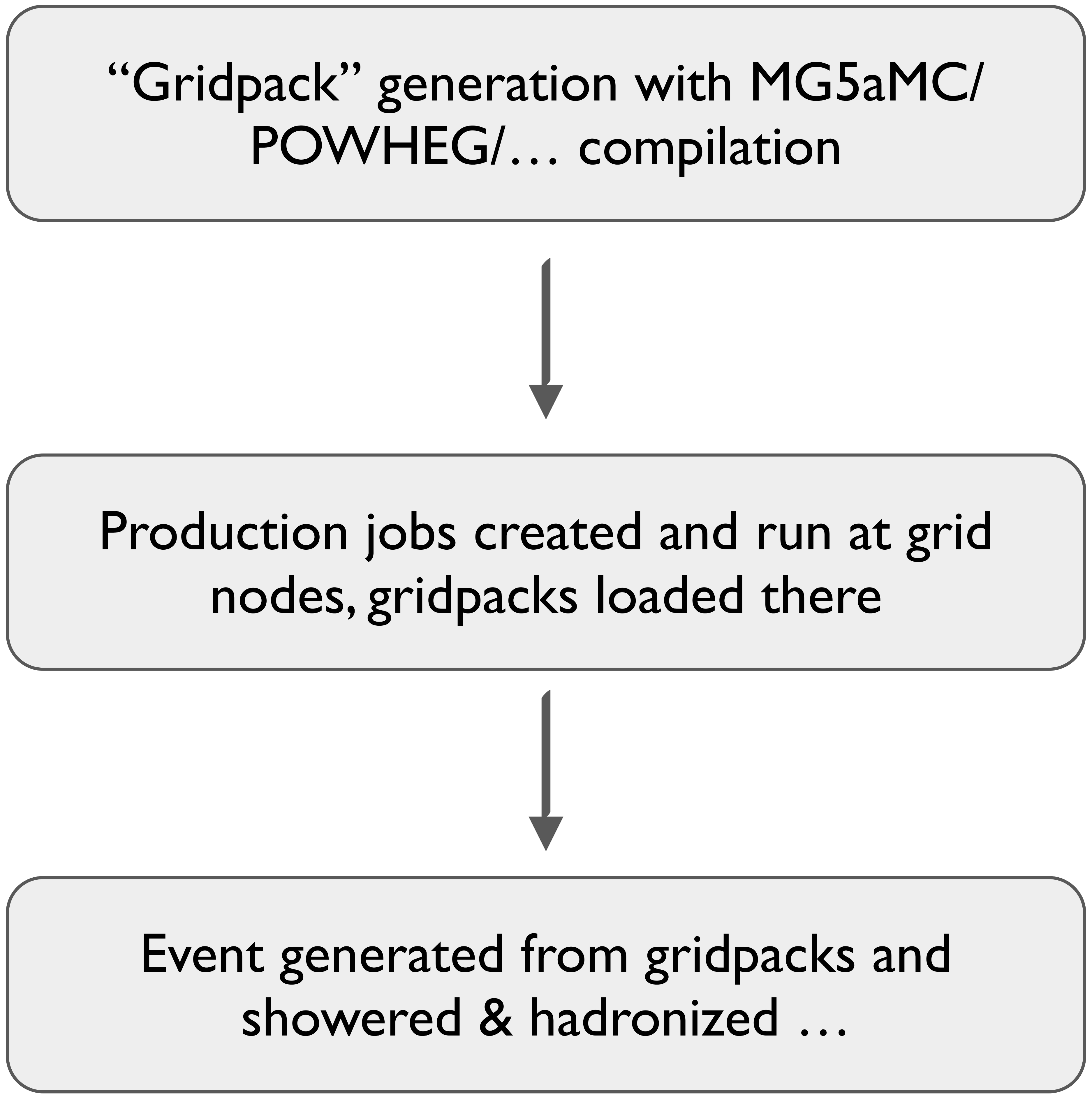


Straightforward multicore utilization!

HEAVY GRIDPACK: PRICE FOR SOPHISTICATED MODELING

CMS MC Production Workflow

highly simplified version



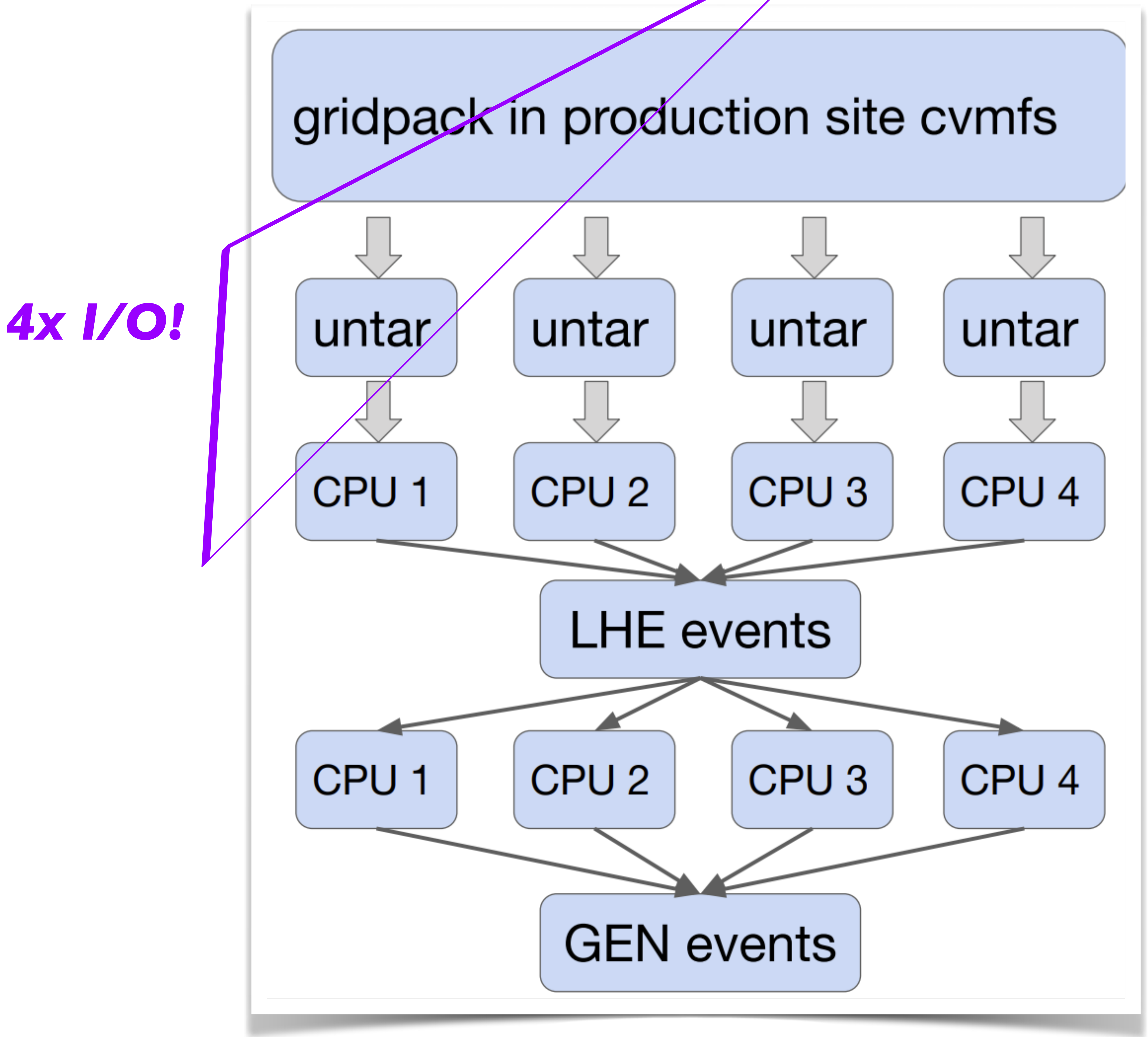
"Gridpack" generation with MG5aMC/ POWHEG/… compilation

Production jobs created and run at grid nodes, gridpacks loaded there

Event generated from gridpacks and showered & hadronized …

More practical workflow:

Multithreading via concurrent jobs



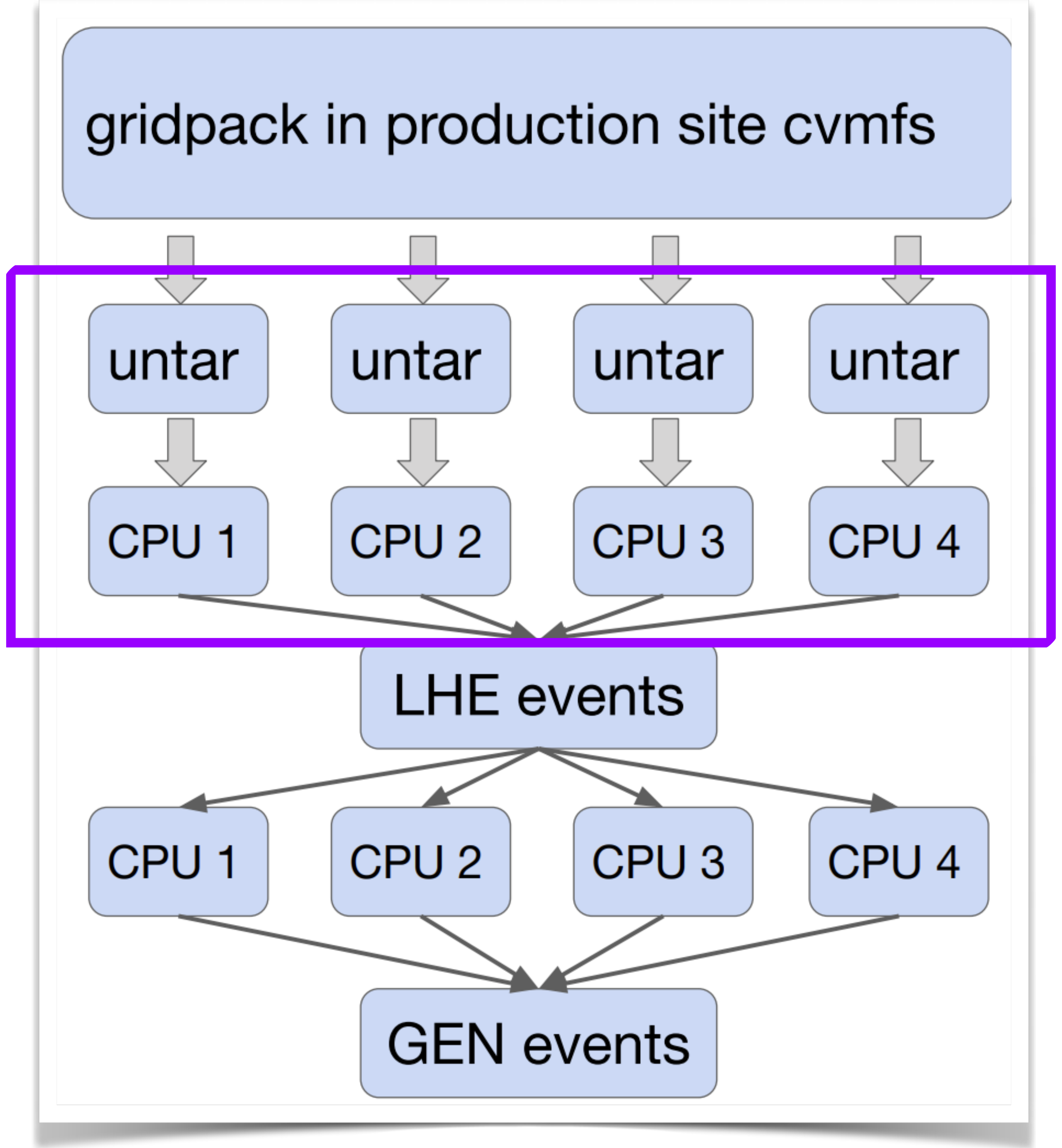
Straightforward multicore utilization!

Not I/O friendly

HEAVY GRIDPACK: PRICE FOR SOPHISTICATED MODELING

Before *MG5aMC*

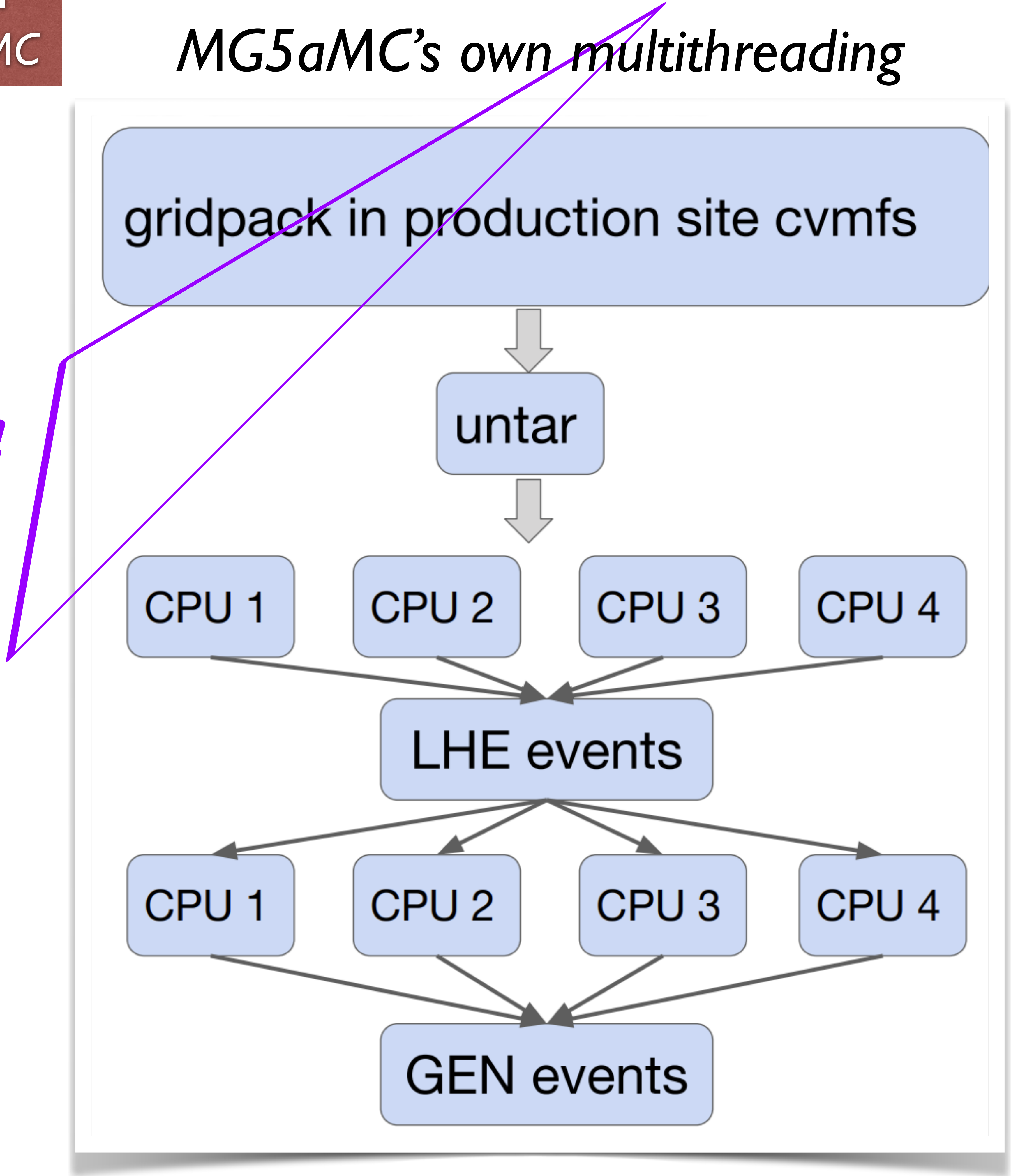
More practical workflow:
Multithreading via concurrent jobs



Straightforward multicore utilization!
Not I/O friendly

After *MG5aMC*

Current workaround:
MG5aMC's own multithreading

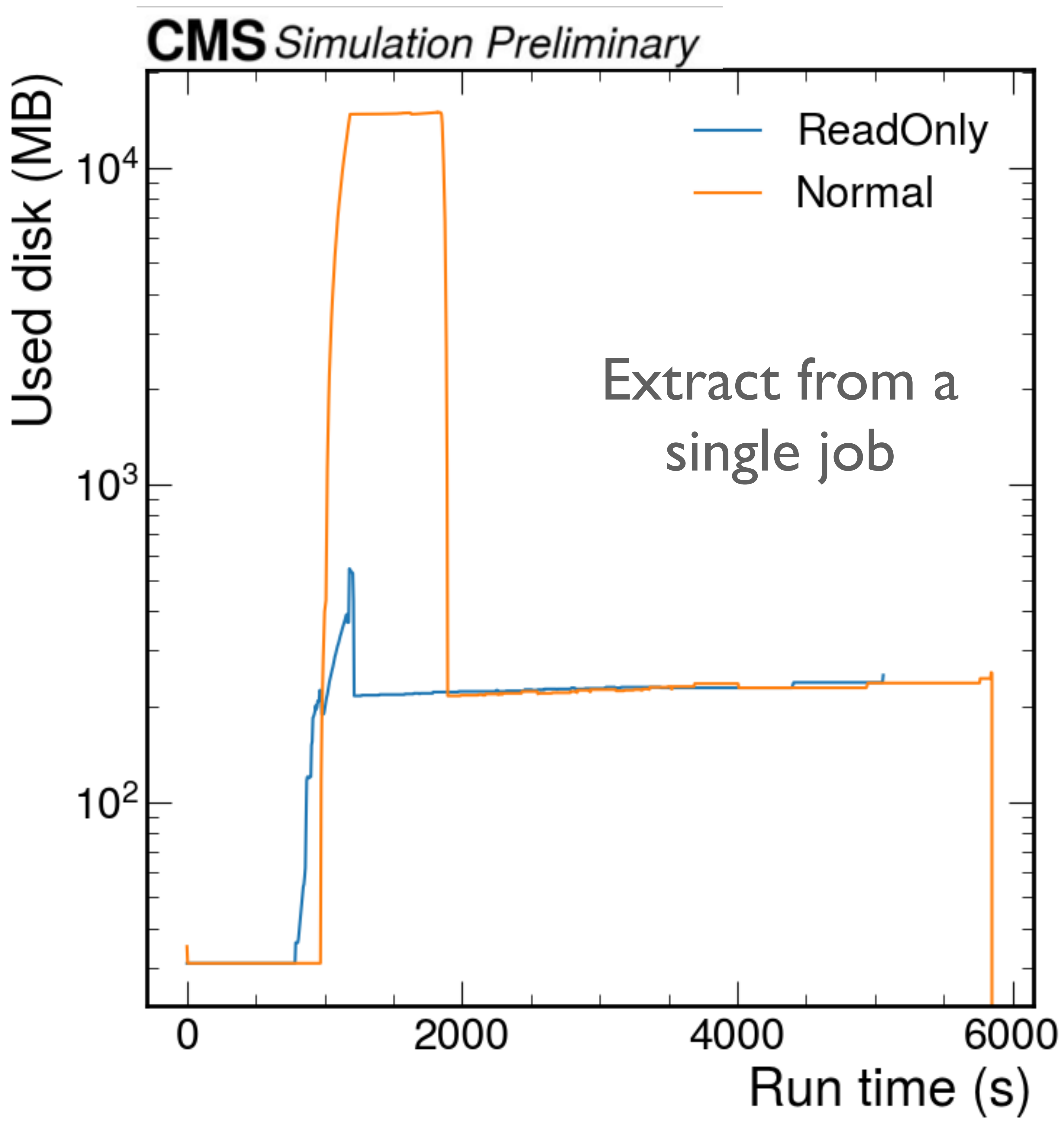


Lesson taken: more jobs per untarred gridpack I/O → better!

HEAVY GRIDPACK: PRICE FOR SOPHISTICATED MODELING

- How to maximize #jobs per untarred gridpack I/O?
  - Have it host on sites and **uncompressed**, then directly load it without I/O on disks
  - Broadcasting through cvmfs: one time cache, multiple times of utilization!
- Prerequisite: Read-Only gridpack
  - No NLO gridpacks from MG5aMC in the past!
  - CMS has worked with MG5aMC authors (many thanks!) for a solution

Source codes implemented.
Ongoing validation with W+012j FxFx

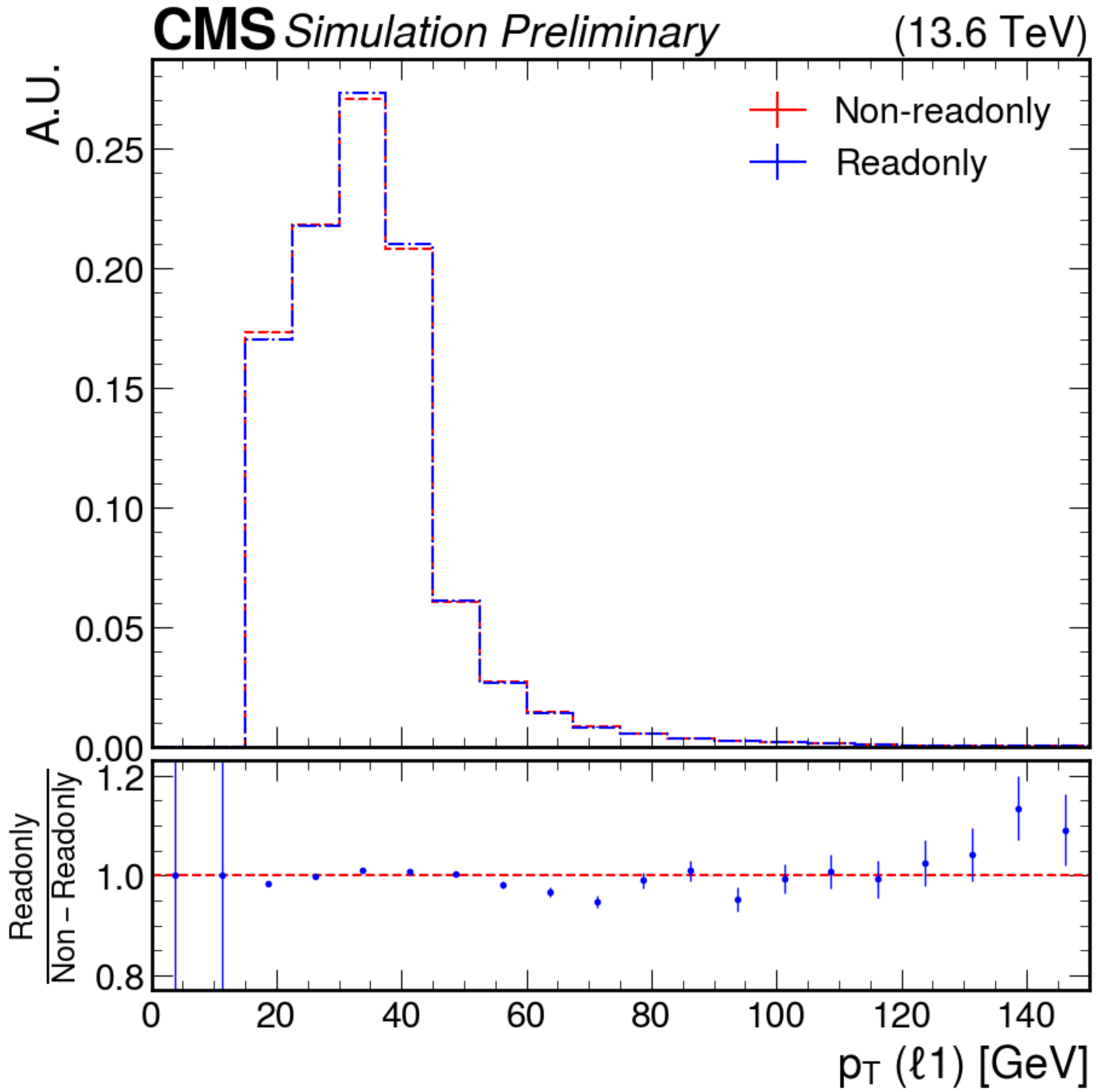


Dramatic improvement on disk usages!

HEAVY GRIDPACK: PRICE FOR SOPHISTICATED MODELING

- How to maximize #jobs per untarred gridpack I/O?
  - Have it host on sites and **uncompressed**, then directly load it without I/O on disks
  - Broadcasting through cvmfs: one time cache, multiple times of utilization!
- Prerequisite: Read-Only gridpack
  - No NLO gridpacks from MG5aMC in the past!
  - CMS has worked with MG5aMC authors (many thanks!) for a solution

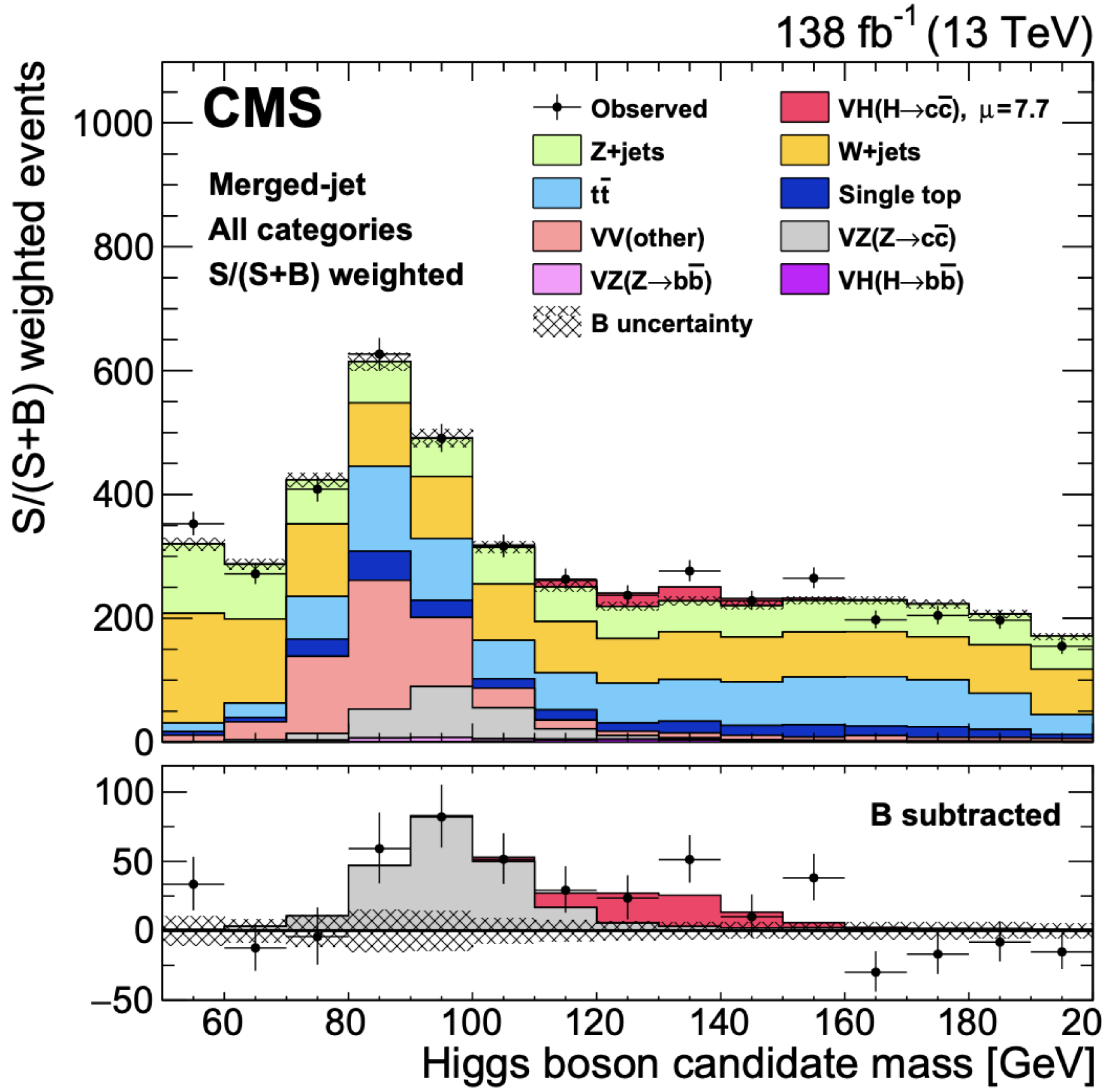
Source codes implemented.
Ongoing validation with W+012j FxFx



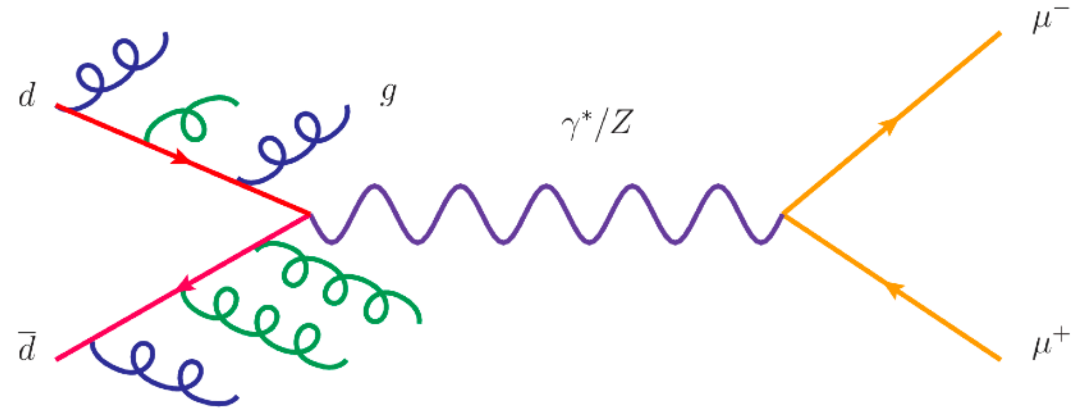
Discrepancies under investigation

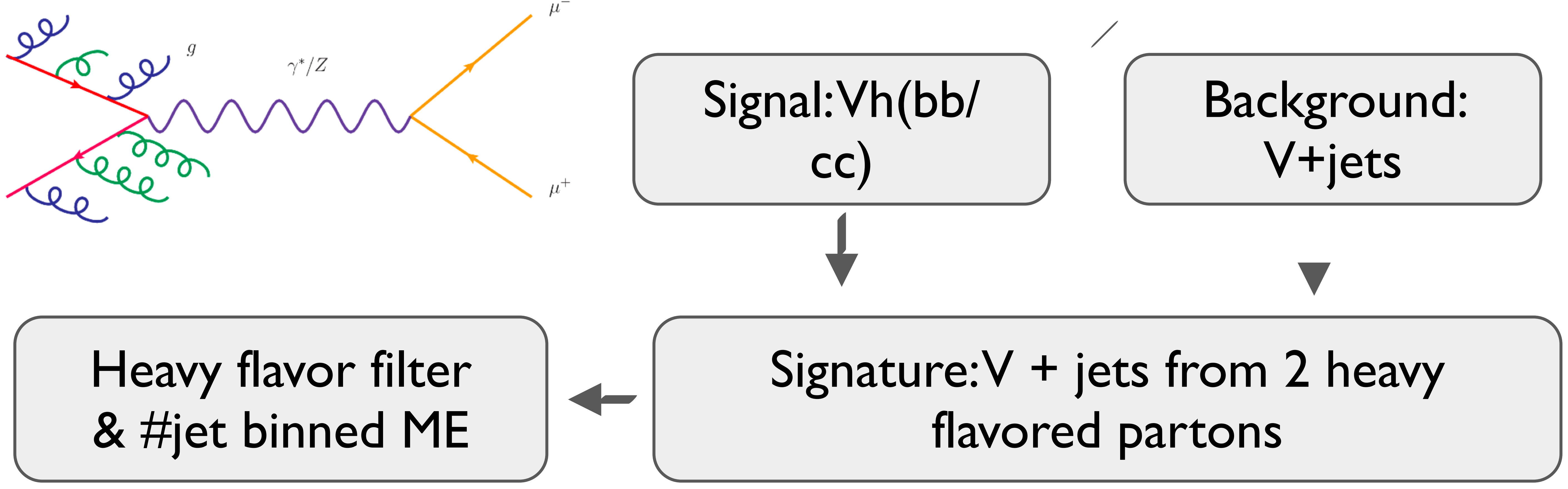
Selecting phase-space of interest: biasing & filtering

Again V+jets as an example



V+jets are important backgrounds of Vh(bb/cc) analyses!



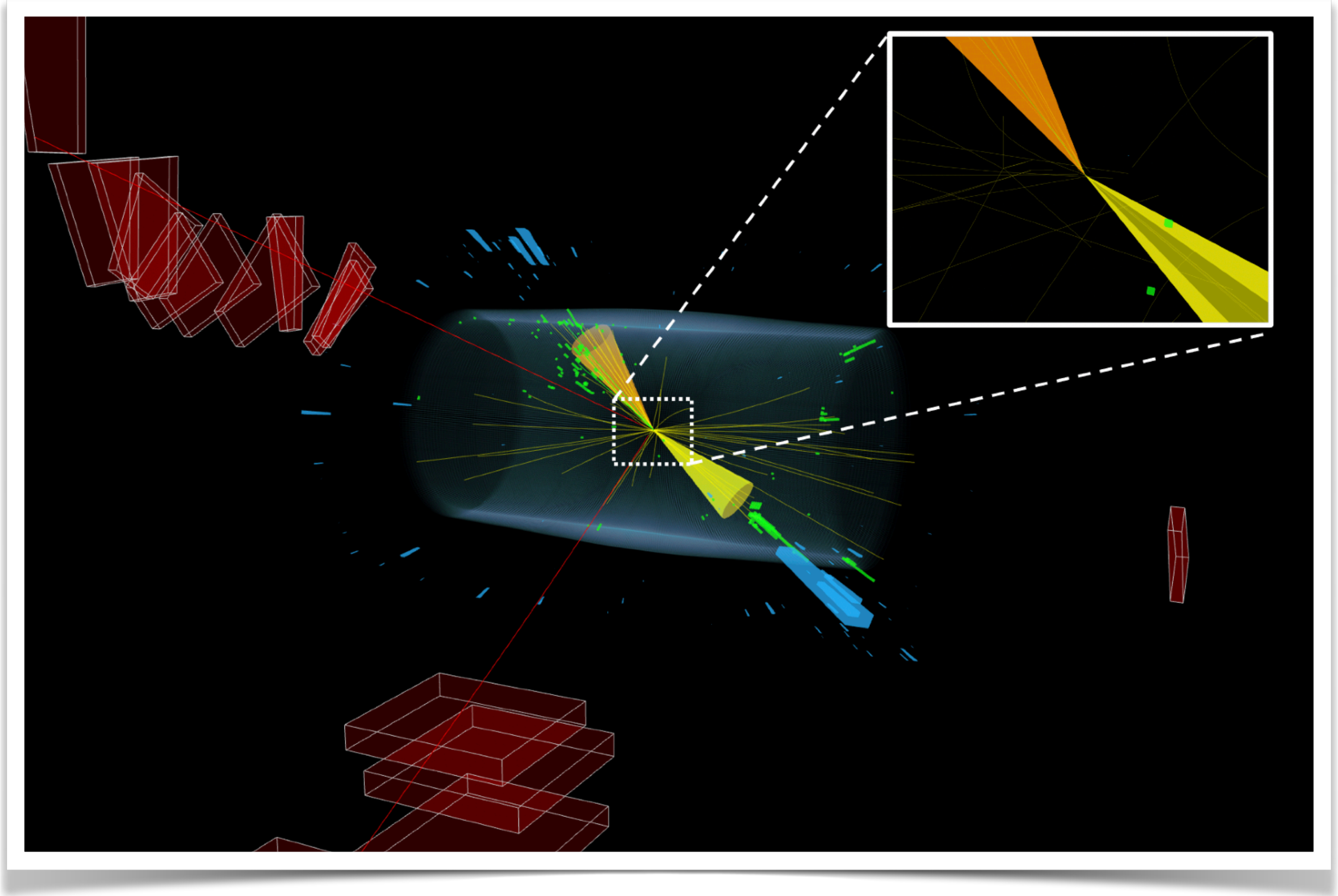


Signal: Vh(bb/cc)

Background: V+jets

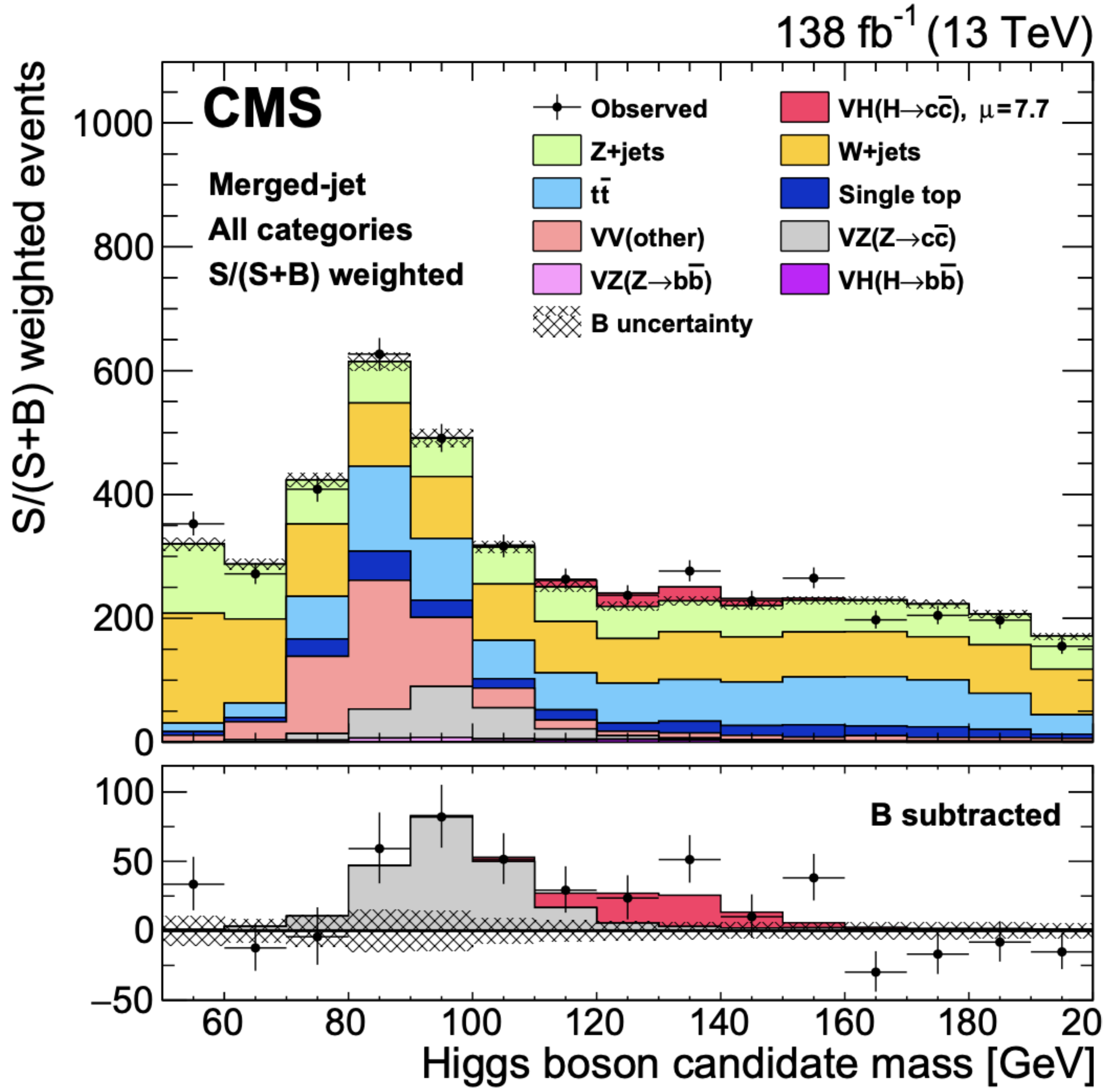
Signature: V + jets from 2 heavy flavored partons

Heavy flavor filter & #jet binned ME

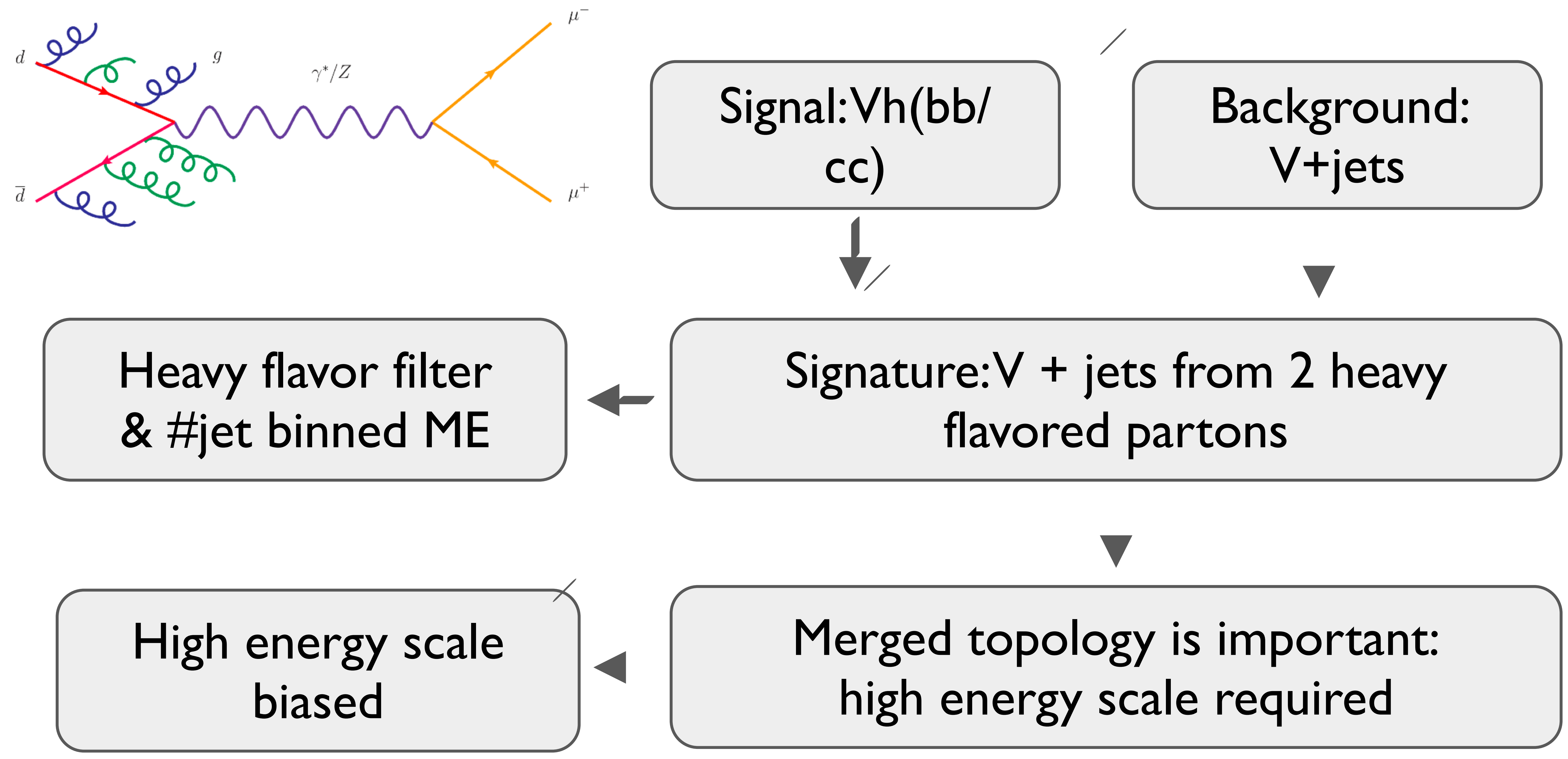


SELECTING PHASE-SPACE OF INTEREST: BIASING & FILTERING

Again V+jets as an example



V+jets are important backgrounds of Vh(bb/cc) analyses!



Signal: Vh(bb/cc)

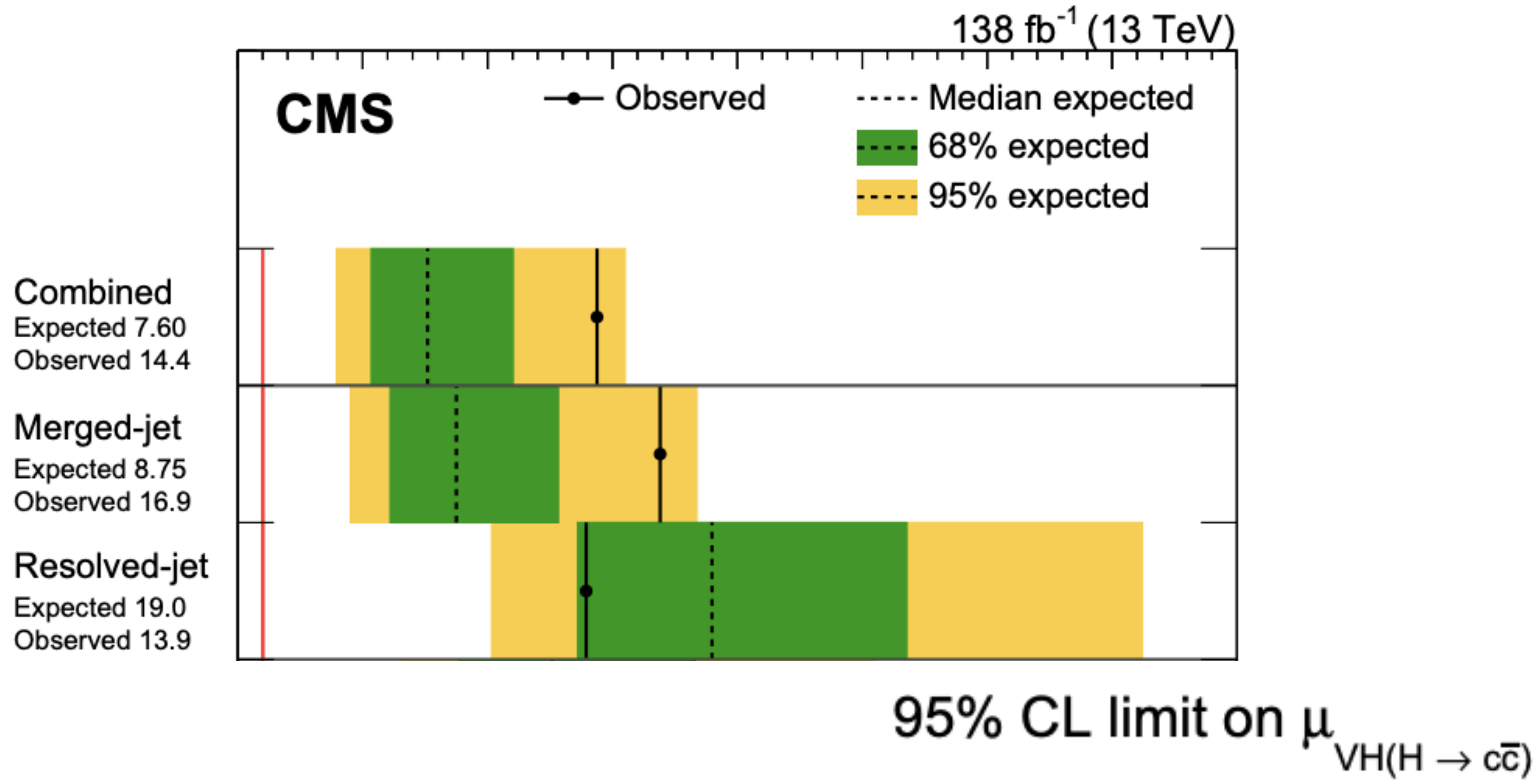
Background: V+jets

Signature: V + jets from 2 heavy flavored partons

Heavy flavor filter & #jet binned ME

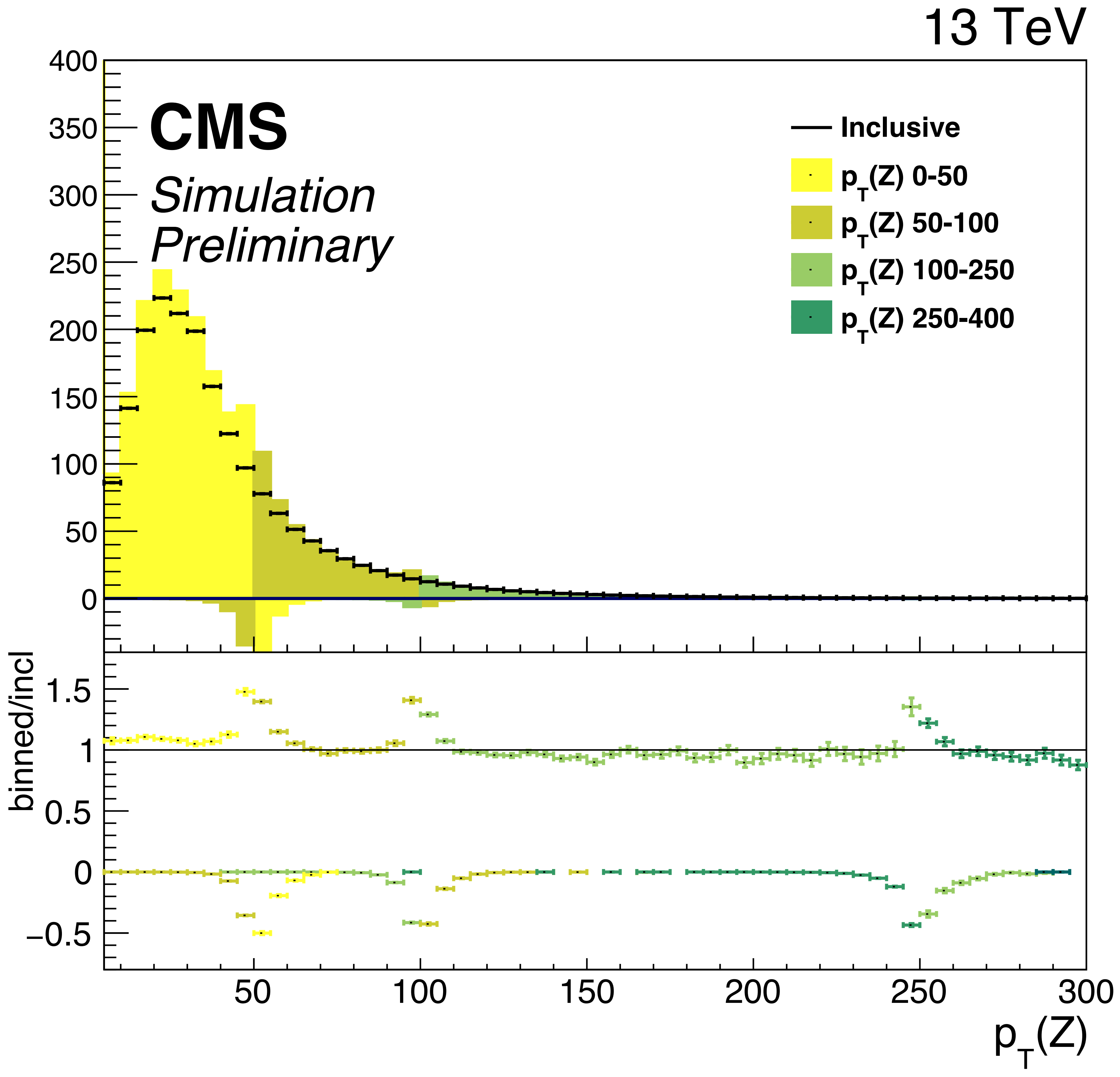
Merged topology is important: high energy scale required

High energy scale biased



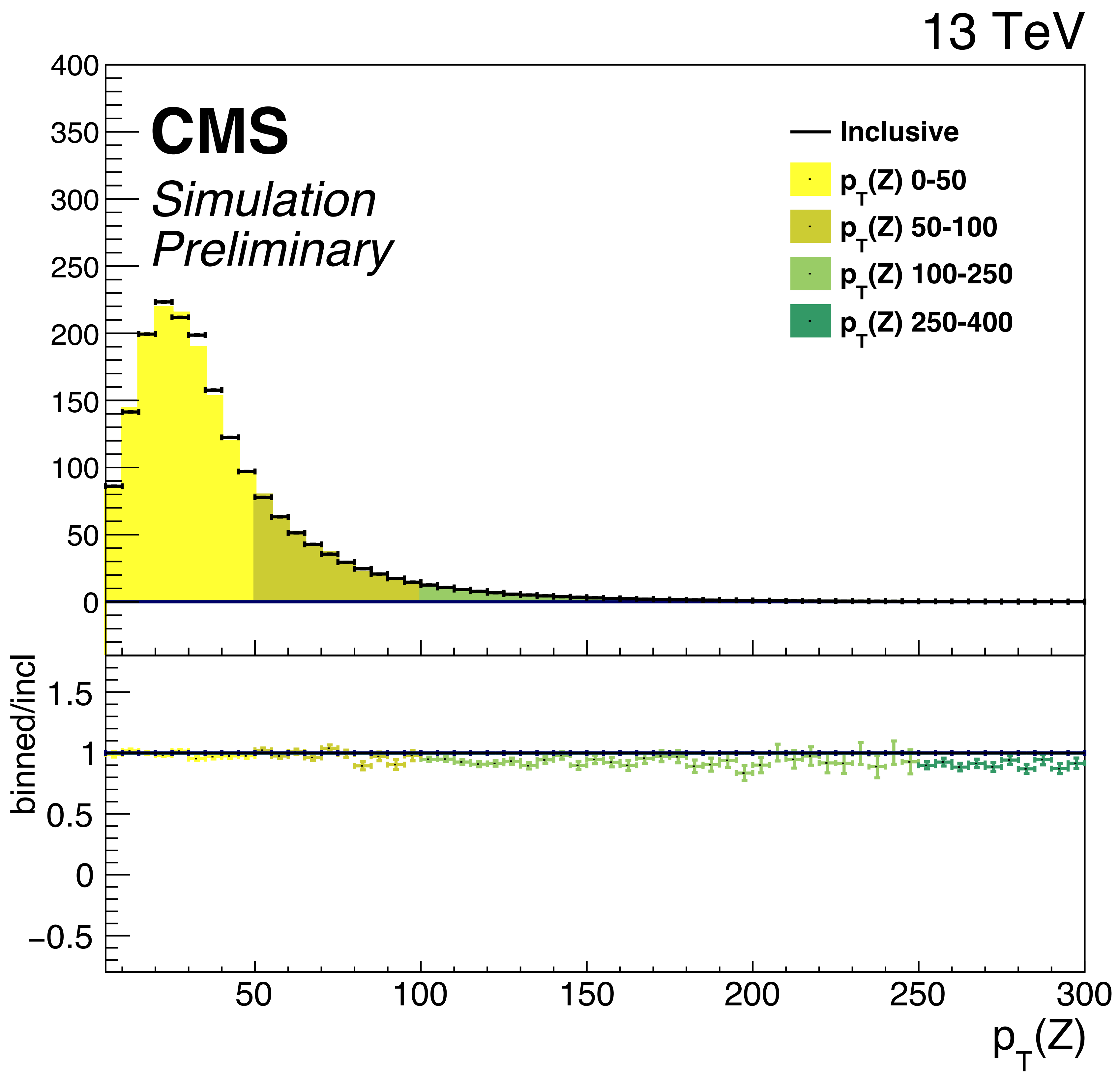
SELECTING PHASE-SPACE OF INTEREST: BIASING & FILTERING

- Conventional approach for high energy scale biasing → binned production (pT(V), HT, etc)
  - Actively iterating with generator authors modeling improvements



Customized NLO pT(V) binned MG5aMC

Spikes around bin boundaries spotted and resolved



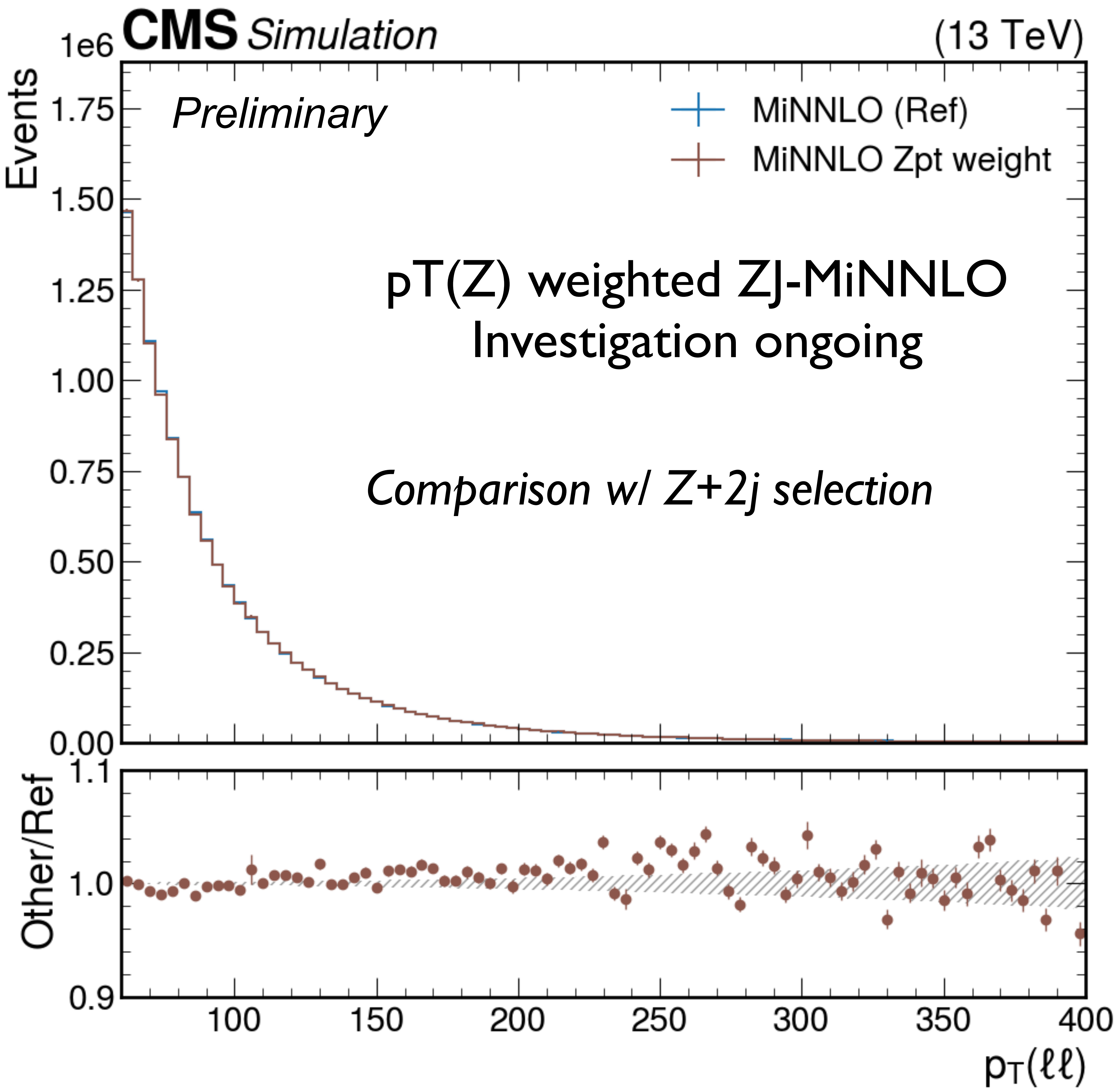
***Left**: Initial modeling with **only cuts** on N-body kinematics for generation*

***Right**: **Two set of cuts**, one on N-body kinematics for generation,, the other on recorded LHE events*

SELECTING PHASE-SPACE OF INTEREST: BIASING & FILTERING

- Conventional approach for high energy scale biasing → binned production (pT(V), HT, etc)

- Alternative approach: produce weighted events with generator bias module
  - Smooth distribution by construction
  - Only one gridpack needed



pT(Z) weighted ZJ-MiNNLO
Investigation ongoing

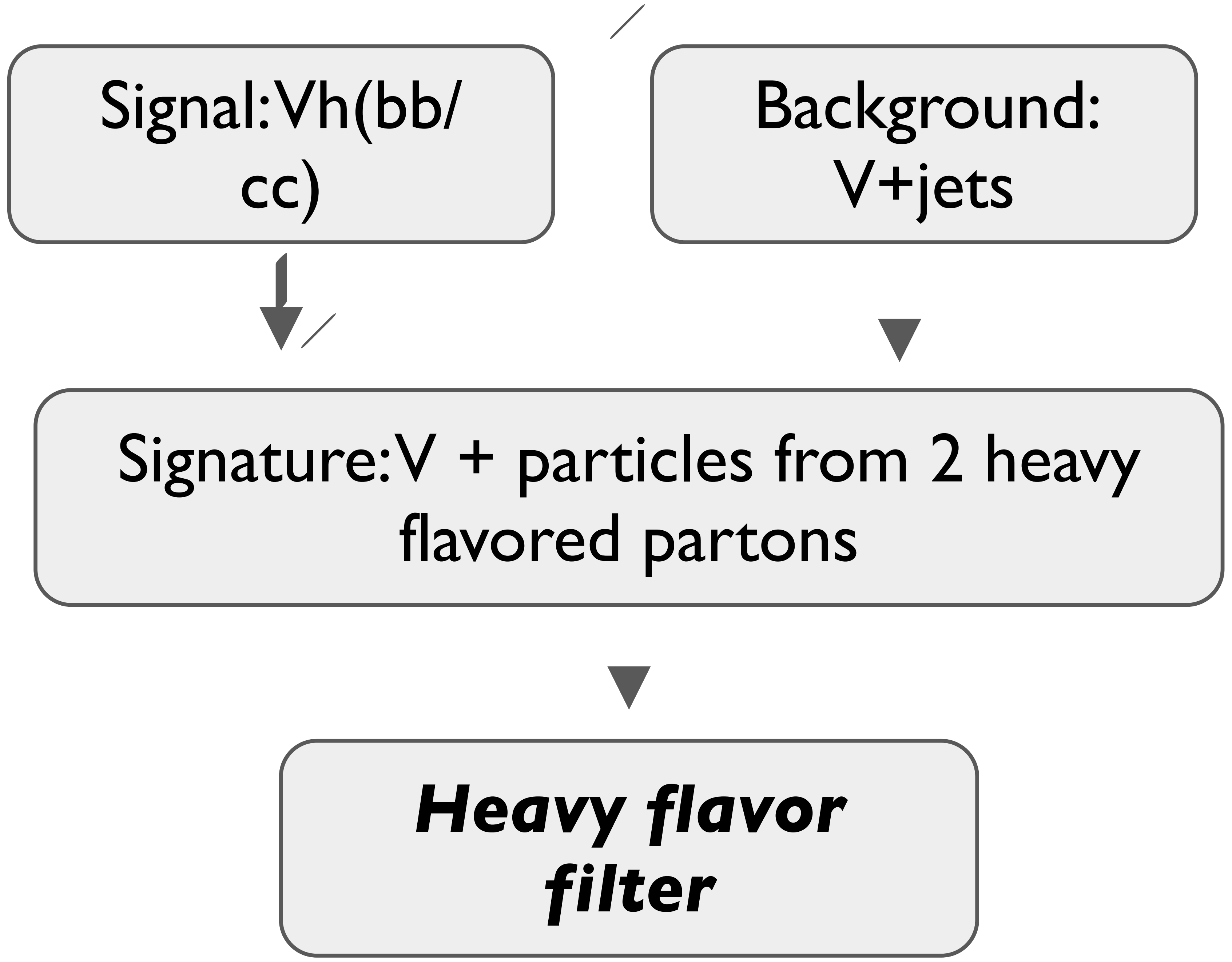
Comparison w/ Z+2j selection

Similar attempts exist for other processes: e.g. QCD dijet production

AUTOMIZED & CENTRALIZED GRIDPACK PRODUCTION

- CMS moves to centralized and automized gridpack production for robust common SM background modeling
- Automized and centralized production helps improving computing efficiency as well:
  - Minimizing human intervention helps reducing computing via ***avoiding***
    - **reproduction** for correcting mistakes
    - **repeated production** from miscommunication

e.g. heavy flavor ***filter*** for V+jets



Signal: Vh(bb/cc)

Background: V+jets

Signature: V + particles from 2 heavy flavored partons

Heavy flavor filter

The heavy flavor ***filter*** could be implemented on top of a normal gridpack by ***filtering*** events from it

Automized & centralized gridpack production

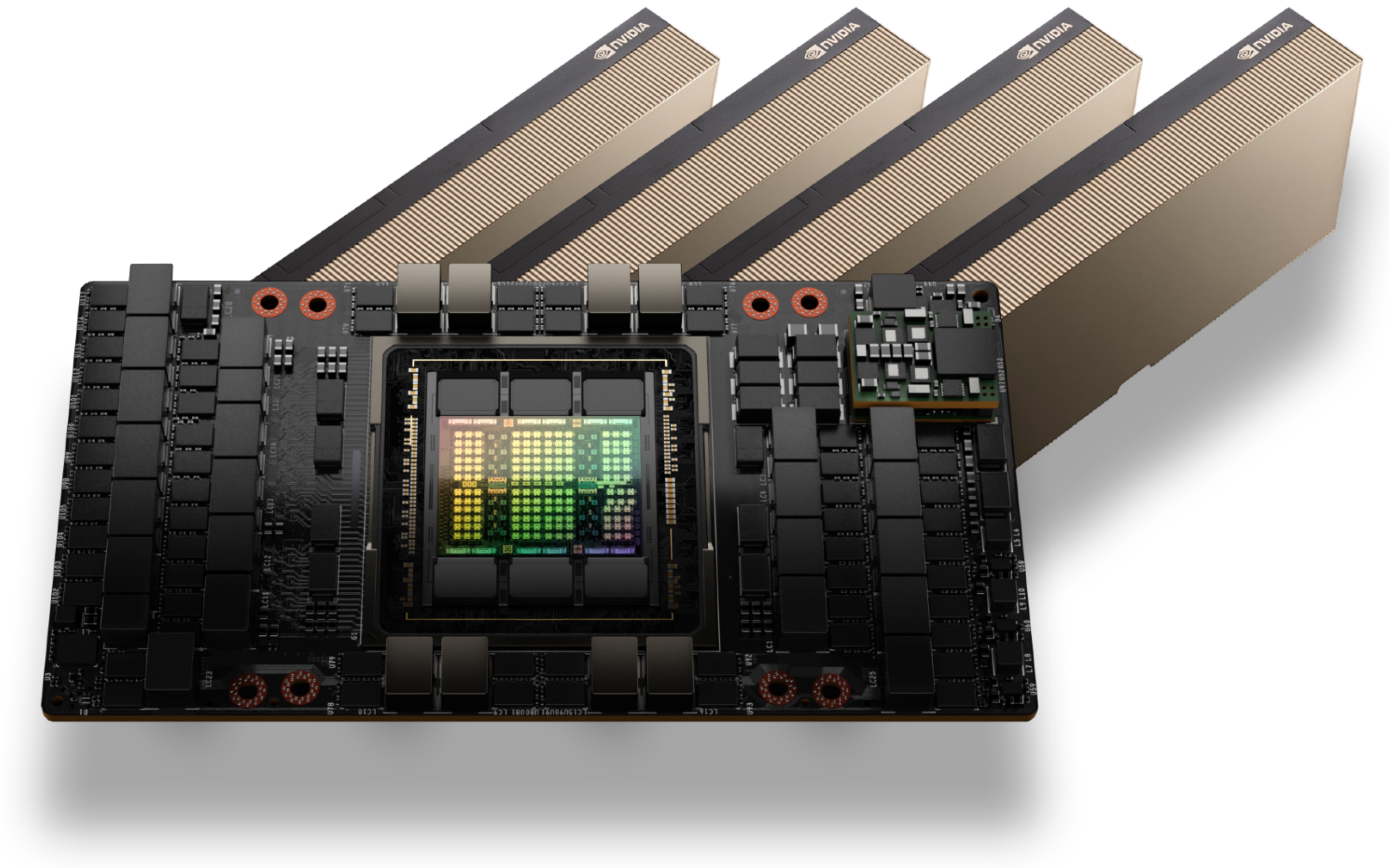
- CMS moves to centralized and automized gridpack production for robust common SM background modeling
- Automized and centralized production helps improving computing efficiency as well:
  - Minimizing human intervention helps reducing computing via ***avoiding***
    - **reproduction** for correcting mistakes
    - **repeated** production from miscommunication
  - Dedicated production platform with a huge bunch of jobs :)
    - Maximized CPU occupancy
    - Dedicated optimization could be carried out
      - ✳ E.g. process specific CPU consumption

✳ E.g. process specific CPU consumption

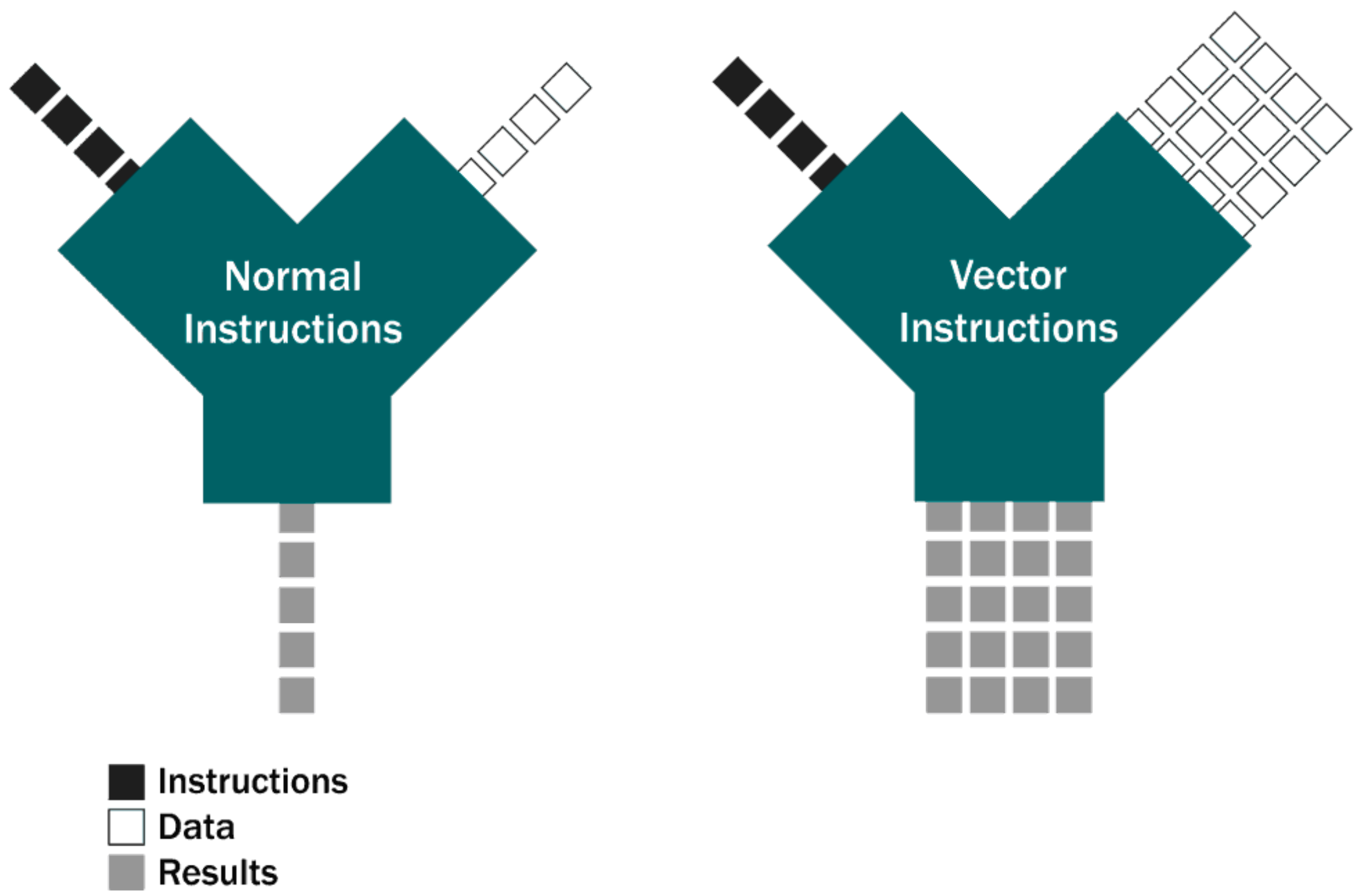
- ✓ No suitable CPU configuration suits all process
  - ➡ *Quite empirical task!*
  - ➡ *e.g.* ***V+0j*** *should consume much less than* ***V+4j****!*
- ✓ Now #CPU cores to use is configurable to maximize CPU efficiency

Preparation for new infrastructure: modern parallelization

GPU Computing



CPU Vectorization



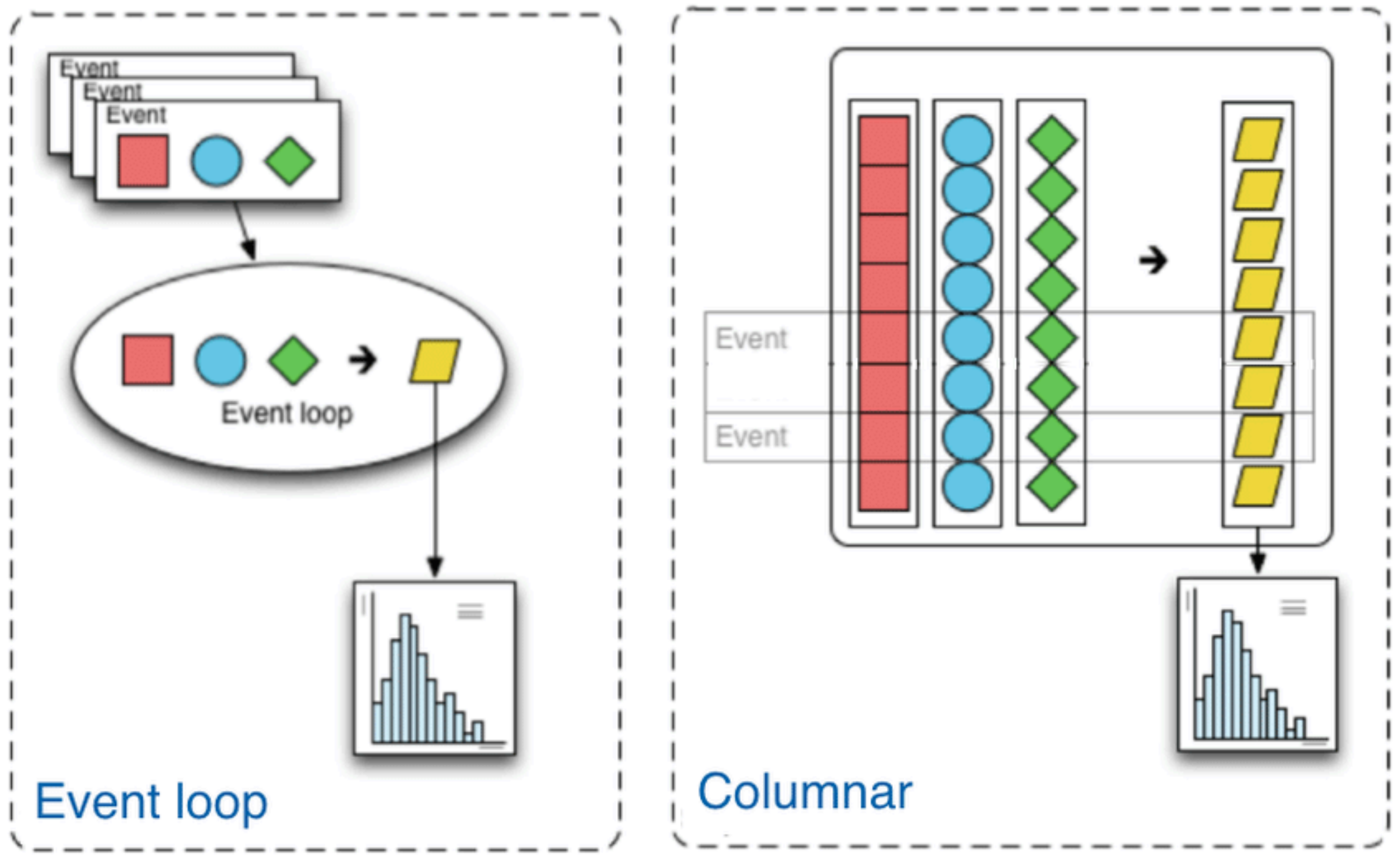
Parallelizaiton is a major topic of modern high performance computing

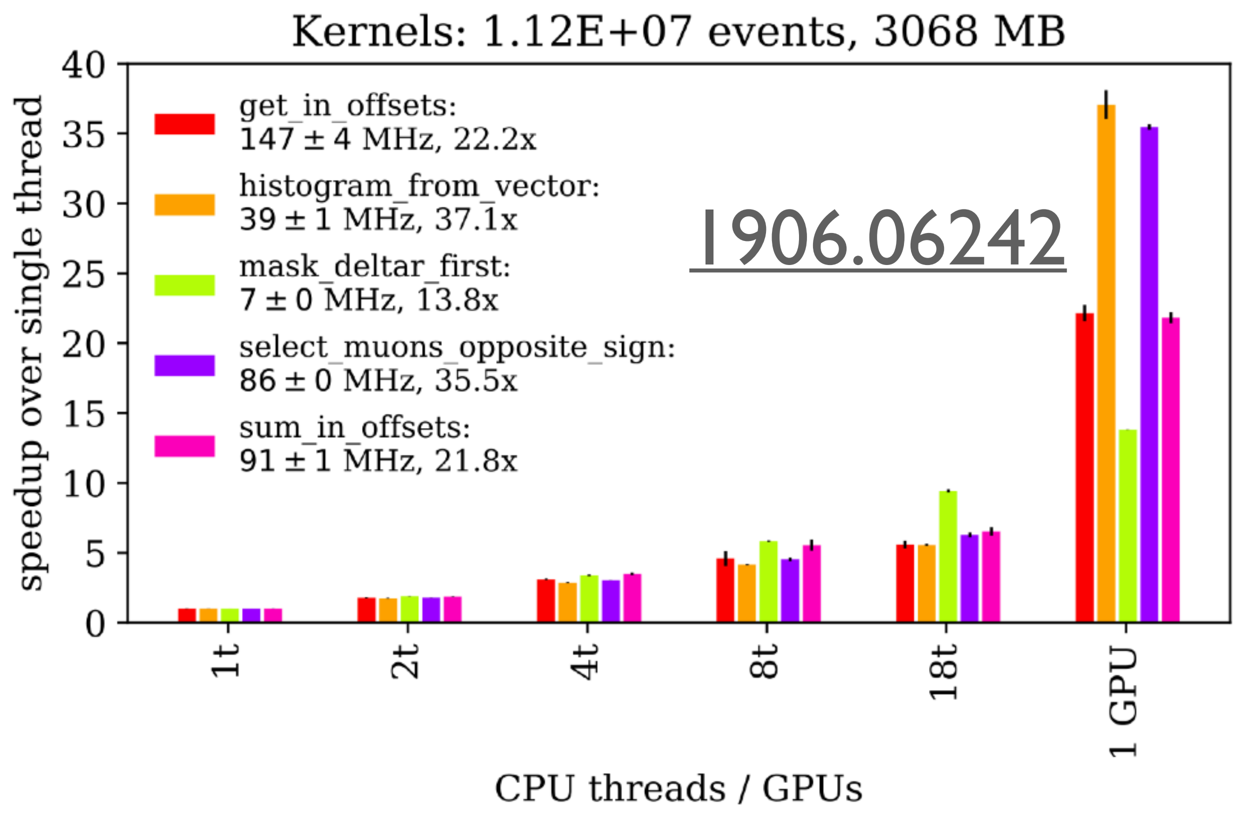
Preparation for new infrastructure: modern parallelization

CMS HLT GPU Farm

https://cms.cern/news/first-collisions-reconstructed-gpus-cms

Columnar analysis



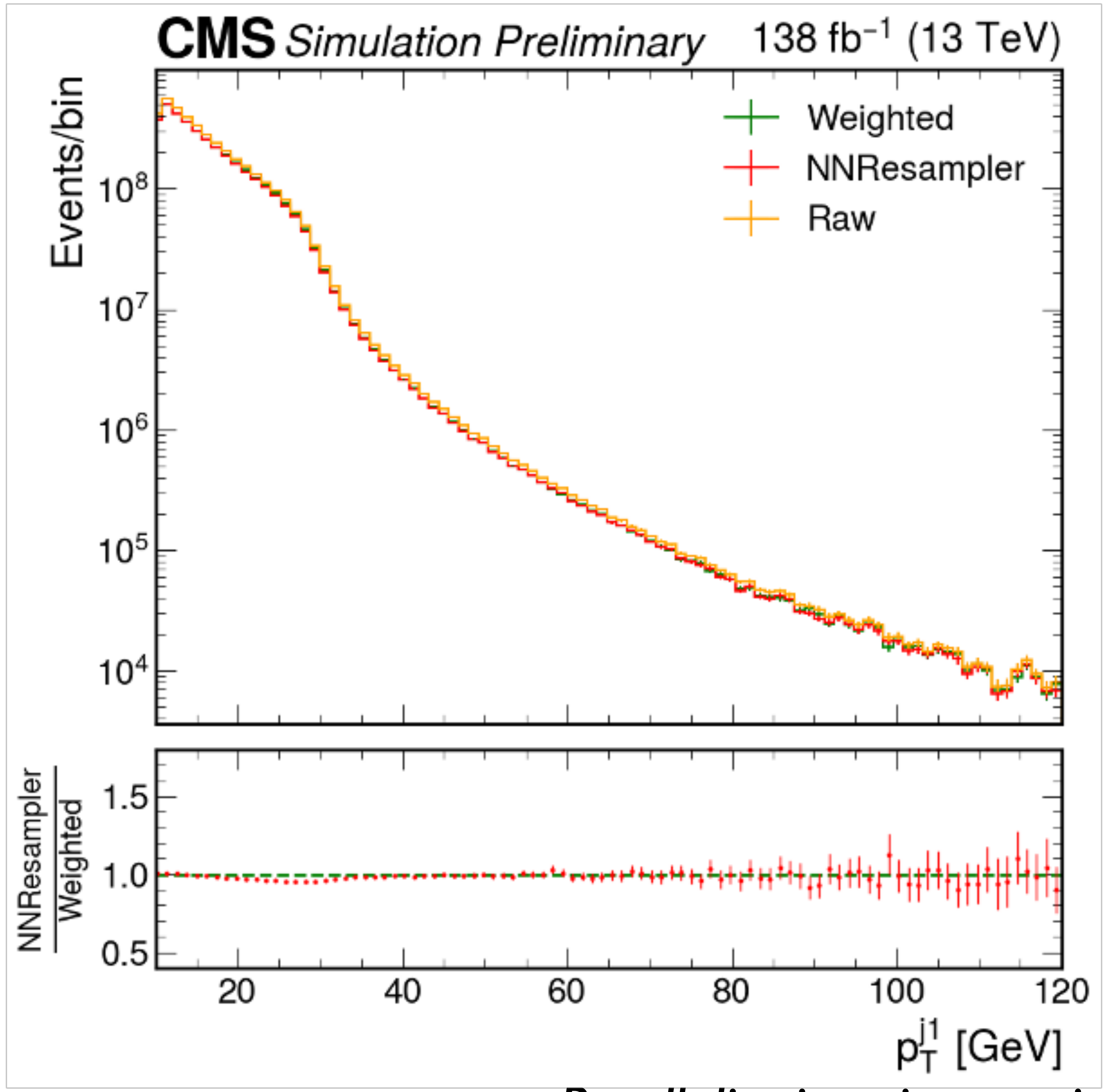


1906.06242

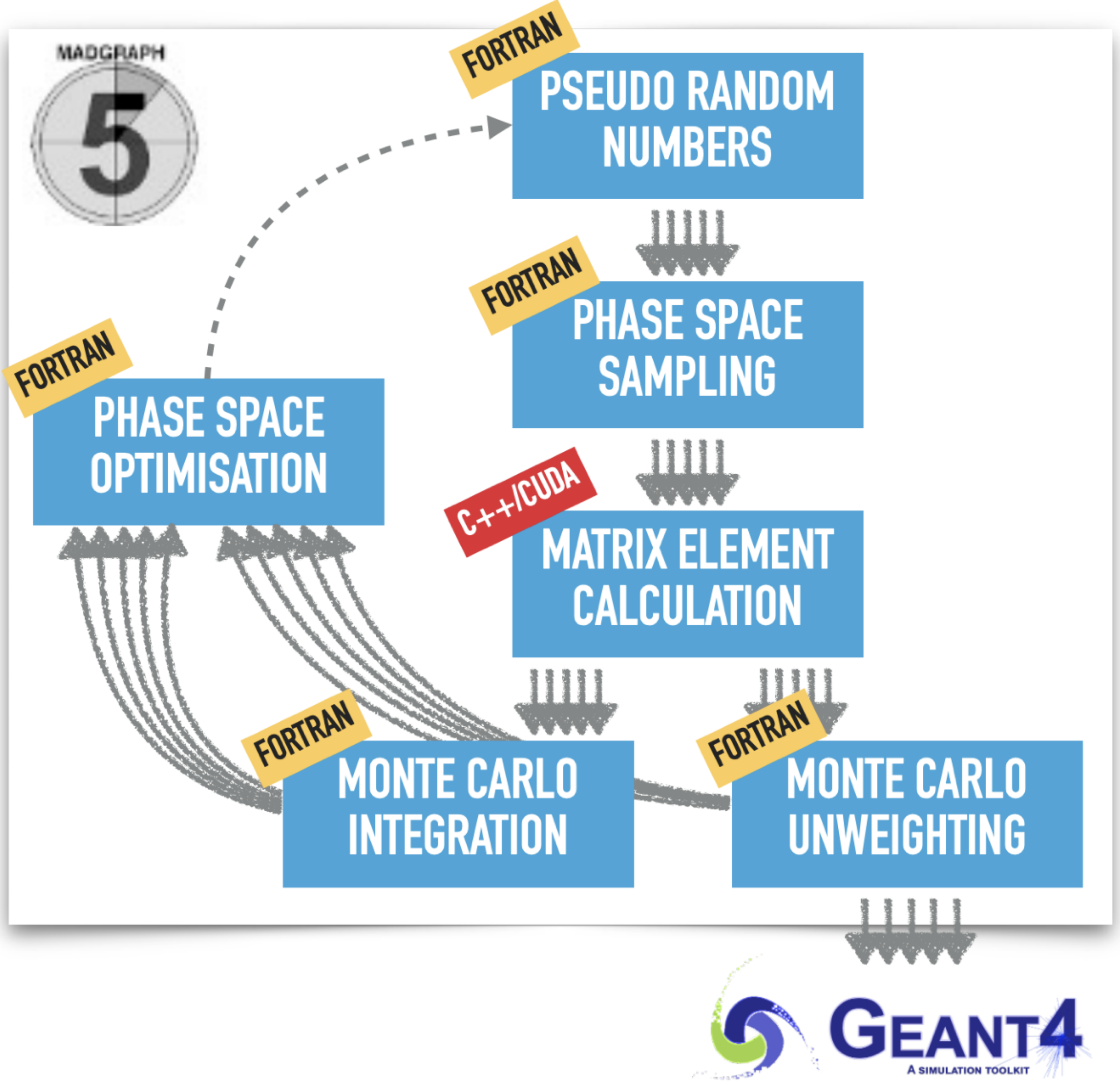
Parallelizaiton is a major topic of modern high performance computing
And extensively employed in HEP!

Preparation for new infrastructure: modern parallelization

Machine learning assists generators and needs GPU



Generators can benefit from GPU/CPU vectorization as well!



Parallelizaiton is a major topic of modern high performance computing
And extensively employed in HEP!
Generator should not be absent!

PREPARATION FOR NEW INFRASTRUCTURE: MODERN PARALLELIZATION

Significant improvement with GPU for gridpack generation!

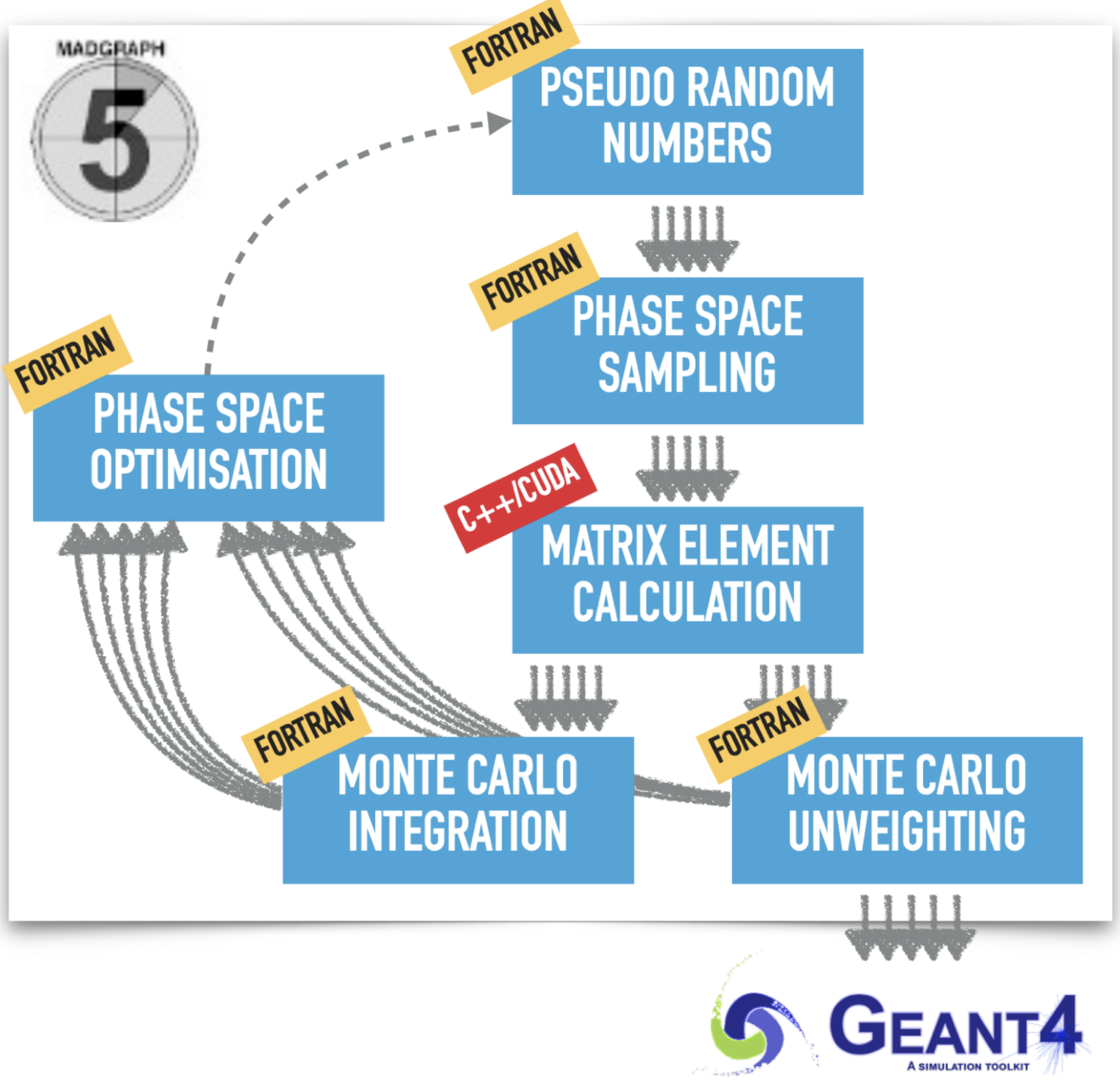
process	Cross section [pb]	Error [pb]	#diagrams (#processes)	Time(FORTRAN)/ Time(CUDA)
TT+0j	504.4	12	8(6)	3.6x
TT+1j	575.7	0.25	9(16)	10.6x
TT+2j	426	0.16	1473(96)	16.1x

CPU Vectorization also helps!

process	Cross section [pb]	Error [pb]	#diagrams (#processes)	Time(FORTRAN)/ Time(CPP)
DY+0j	5711	1.054	30(15)	5.4x
DY+1j	3535	1.263	180(45)	4.7x
DY+2j	2236	0.5005	3120(285)	4.1x

Time value reported here include uncompressing

Generators can benefit from GPU/CPU vectorization as well!



Further checks ongoing
CMS workflow integration & test in future

SUMMARY:

- Overview of CMS efforts in generator acceleration and optimization

- Progresses includes various levels/aspects
  - Algorithmically improving generators: negative weight elimination, efficient phase space biasing & filtering, tackling the heavy gridpack I/O issue
  - Systematically improving the workflow: automized and centralized gridpack production
  - Preparing for new computing infrastructures: testing MG4GPU

- Look forward to a fruitful discussion here!

THANKS!

BackUp: Automized & centralized gridpack production

	SM backgrounds with large cross section	Signal Processes
Number of events to produce	Large	Not too large per process
Phase space coverage	Large	Signal specified region
Sophisticated modeling (e.g. jet merging, high order simulation)	Yes	Not necessary for going too far
Complexity for gridpack production	High	Not quite
Production workflow	Standardized	Might be novel and/or flexible
CMS policy	From Run3: automatically and centrally produced	Not centrally produced