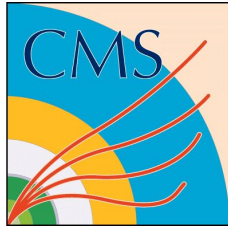


CMS event generation accounting update

Efe Yazgan, Qiang Li
14 March 2019

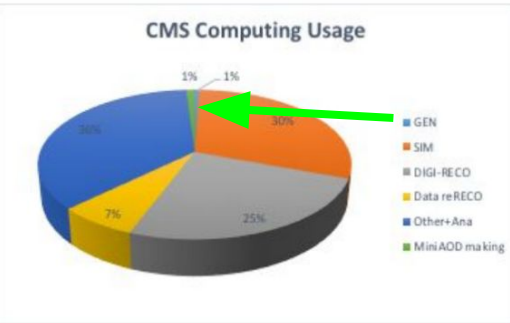


中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

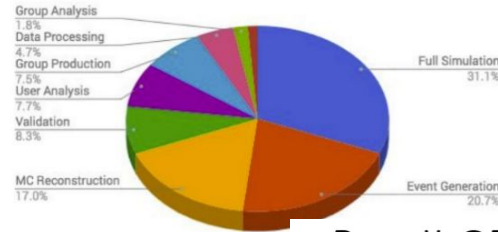


Reminds: Physics Event Generator Computing Workshop, 26-28 November 2018

Generators CAN be Computationally Intensive



US ATLAS Wall Clock CPU - 2016



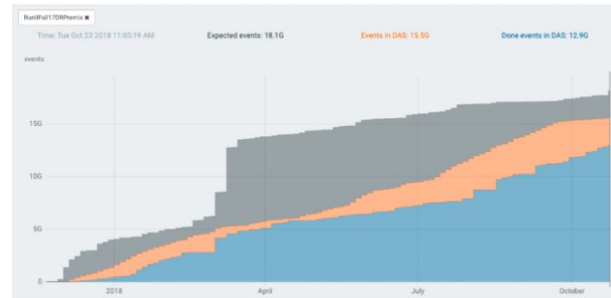
Run II GEN Production

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

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

15B, 8months, 60k cores -> 85sec/evt
For GEN+SIM+DIGI+RECO

CMS by default treats GEN & SIM together.



- 15 B (+ some other production campaigns ~ 20 B) in 8 months
 - GEN-SIM-DIGI-RECO ~85 sec/evt
 - 60k cores (~1/3 of the CMS production power)
- Multi-leg LO
 - up to ~10s/gen-evt
 - ~10% matching efficiency → 100s/full-sim-evt
- Multi-leg NLO
 - up to ~30s/gen-evt
 - ~30% matching efficiency → 100s/full-sim-evt
 - Large fraction of negative weights of up to ~40% → larger samples!

Run II:

- GEN not stored for physics samples in disk. GEN-SIM re-produced whenever needed.
- Generators ~1-10% of the total CPU
 - Variation due to LO, NLO, NNLO, complexity of the process, or different methods of calculation.
 - Most BSM samples at this point are simulated at LO.

A bit more accurate estimation based on a GENOnly campaign:

“GEN is about 10% of GENSIM,
i.e. for a figure of 30% for GENSIM fraction of total
CPU resources, a number like 3% for GEN is
accurate to within a factor of two.” Nicholas Smith

2017 DY and ttbar (Fall17 Campaign)

	DY+WJets	ttbar
LO	1,407,390,000	426,758,223
NLO	1,957,990,000	1,937,376,864
Total	3,365,380,000	2,364,135,087

Samples and number of events are in

https://docs.google.com/spreadsheets/d/1vbNb4fWG5N5y9zqrpFgqFVH29DgkUzpR_jfVFH44Jxs/edit?usp=sharing

- (1) **DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8 (176M for the inclusive sample)**
 - DY1/2/3/4Jets,
 - HT-70to100, 100to200, 200to400, 400to600, 600to800, 800to1200, 1200to2500, 2500toInf
 - Pt-0to50, 50to100, 100to250, 250to400, 400to650, 650toInf
 - M-50, 10to50, 100to200, 200to400, 400to500
- (2) **DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8 (228M for the inclusive sample)**
 - 0/1/2Jets, LHEZpT, MII

- **For both DY (1) LO and (2) NLO Samples listed in last page:
GEN costs 5-17% of the total processing time.**
Note in CMS we have many LO samples for SUSY and Exotica searches, which can decrease the GEN fraction of total CPU usage quite much!
- **1071 Scale and PDF weights for DY LO, corresponding to
~9% of GEN processing time
~0.4% of total processing time**
- **Madgraph (especially LO) in CMS for the moment rely on internal multithreading which is not perfect.**
In one of the test cases with 8 threads, it forced down the CPU efficiency to 60%.
Note in production, the low CPU efficiency will not pass validation, and would run with fewer threads.

GEN CPU fraction = GEN / (GEN+SIM+DIGI+RECO+MINIAOD)

TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8

For Powheg NLO, GEN CPU fraction ~4%

BulkGravToWW_narrow_M-1000_13TeV-madgraph

For Exotica LO sample, GEN CPU fraction <~ 1%

QCD_Pt_50to80_TuneCP5_13TeV_pythia8

For Pure Pythia Samples, GEN CPU fraction <~ 1%

Exp		ATLAS				CMS			
Process	Prec	Gen	nEvts	CPU[s]	s/evt	Gen	nEvts	CPU[s]	s/evt
W/Z/ γ			6600M						
DY+Wjets	LO					MG5_aMC	1410 M		
DY+Wjets	NLO					MG5_aMC	1960 M		
ttbar	LO		900M			MG5_aMC	430 M		
ttbar	NLO		?			PW ¹	1940 M		
Diboson			1700M			PY/PW/MG/HW	712 M		
ttV			75M			MG5_aMC	154 M		
single top			80M			PW ²	876 M		
multijet			175M			PY/MG_aMC	1950 M		
TOTAL <small>including OTHERS</small>	—	—	40B?			—	20B?		

Table 1: Summary of number of events and CPU consumption for large background samples for ATLAS and CMS.

Process definitions and CPU and s/evt still to be added.

Use samples similar in ATLAS and CMS:

- Ttbar powheg+pythia8 (inclusive NLO)
- W+jets LO HT-binned
- W+jets NLO (N-jets)
- ...