Energy Frontier MC Task Force Report

John Stupak
on behalf of the Energy Frontier MC TF

10/7/20

Introduction

- Background, charge, etc
- Wiki: https://snowmass21.org/montecarlo/energy
- Membership:

Name	Institution	email	
John Stupak (chair)	University of Oklahoma	john.stupak[at]cern.ch	Snowmass 2013
Robert Gardner	University of Chicago	rwg[at]uchicago.edu	OSG rep.
Simone Pagan Griso	LBNL	spagangriso[at]lbl.gov	BSM rep.
Stefan Hoeche	FNAL	shoeche[at]fnal.gov	MC expert
Fabio Maltoni	CP3, Catholic University of Louvain	maltoni.fabio[at]gmail.com	MC expert
Meenakshi Narain	Brown University	meenakshi.narain[at]cern.ch	EF convener
Isabel Ojalvo	Princeton University	isabel.rose.ojalvo[at]cern.ch	EWK rep.
Laura Reina	Florida State University	reina[at]hep.fsu.edu	EF convener
Michael Schmitt	Northwestern University	m-schmitt[at]northwestern.edu	QCD rep.
Alessandro Tricoli	Brookhaven National Laboratory	atricoli[at]bnl.gov	EF convener

Charge of the EF MC Task Force (I)

1. Assess the MC needs for studies by each Energy Frontier Topical Group.
    a. This should include the processes, the MC generators, the accelerator configurations (c.o.m, integrated luminosity, pileup scenarios, if any), detector configurations, and number of events for each process type.
2. Survey existing frameworks for MC generation and analysis for future circular colliders (FCC-ee, FCC-hh, CepC, CppC, LHeC, EIC...etc...).
    a. Are the existing samples and framework sufficient for our studies?
    b. Need to request permission to use the existing samples?
3. Check/confirm that ILC, CLIC, Muon collider studies will use their frameworks, and no MC generation by EF group needs to be planned.
4. Finalize the plans and submit the recommendations by the end of June 2020 to the EF conveners.
5. The plan and recommendations will be presented to the EF community and discussed during the July 2020 EF Workshop.
6. The OSG has kindly agreed to support the MC generation for EF, and will provide both compute resources and storage on the OSG Data Federation.

8

Charge of the EF MC Task Force (I)

1. Assess the MC needs for studies by each Energy Frontier Topical Group.
   a. This should include the processes, the MC generators, the accelerator configurations (c.o.m, integrated luminosity, pileup scenarios, if any), detector configurations, and number of events for each process type.
2. Survey existing frameworks for MC generation and analysis for future circular colliders (FCC-ee, FCC-hh, CepC, CppC, LHeC, EIC...etc...).
   a. Are the existing samples and framework sufficient for our studies?
   b. Need to request permission to use the existing samples?
3. Check/confirm that ILC, CLIC, Muon collider studies will use their frameworks, and no MC generation by EF group needs to be planned.
4. Finalize the plans and submit the recommendations by the end of June 2020 to the EF conveners.
5. The plan and recommendations will be presented to the EF community and discussed during the July 2020 EF Workshop.
6. The OSG has kindly agreed to support the MC generation for EF, and will provide both compute resources and storage on the OSG Data Federation.

Community Survey

Collaboration Survey

8

collaboration and community surveys were distributed on May 31

Charge of the EF MC Task Force (I)

1. Assess the MC needs for studies by each Energy Frontier Topical Group.
   a. This should include the processes, the MC generators, the accelerator configurations (c.o.m, integrated luminosity, pileup scenarios, if any), detector configurations, and number of events for each process type.

Community Survey

2. Survey existing frameworks for MC generation and analysis for future circular colliders (FCC-ee, FCC-hh, CepC, CppC, LHeC, EIC...etc...).
   a. Are the existing samples and framework sufficient for our studies?
   b. Need to request permission to use the existing samples?
3. Check/confirm that ILC, CLIC, Muon collider studies will use their frameworks, and no MC generation by EF group needs to be planned.

Collaboration Survey

4. Finalize the plans and submit the recommendations by the end of ~~June~~ September 2020 to the EF conveners.
5. The plan and recommendations will be presented to the EF community and discussed during the ~~July 2020 EF Workshop.~~ CPM
6. The OSG has kindly agreed to support the MC generation for EF, and will provide both compute resources and storage on the OSG Data Federation.

8

collaboration and community surveys were distributed on May 31

Charge of the EF MC Task Force (II)

7. **Develop a plan,** in the event the EF group has to mount a production of a large set of samples for Standard Model backgrounds. The plan should address the following questions:
   a. Shall we adopt a "common framework" both for generation & analysis of the various samples, if so, which one(s)?
   b. Which samples are needed to be produced as a central production?
      i. Include detailed information about the samples (as listed in 1.a above).
      ii. Should signal samples be produced by the proponents and only large SM background samples be produced centrally?
   c. What scale of CPU resources are needed for sample generation?
   d. What projected size of storage is required for production and long term storage of the samples?
   e. Recommendation on the formation and activities of the "EF Monte Carlo Production team".

9

Charge of the EF MC Task Force (II)

7. **Develop a plan,** in the event the EF group has to mount a production of a large set of samples for Standard Model backgrounds. The plan should address the following questions:
   a. Shall we adopt a “common framework” both for generation & analysis of the various samples, if so, which one(s)?
   b. Which samples are needed to be produced as a central production?
      i. Include detailed information about the samples (as listed in 1.a above).
      ii. Should signal samples be produced by the proponents and only large SM background samples be produced centrally?
   c. What scale of CPU resources are needed for sample generation?
   d. What projected size of storage is required for production and long term storage of the samples?
   e. Recommendation on the formation and activities of the “EF Monte Carlo Production team”.

Community Survey

9

collaboration and community surveys were distributed on May 31

Plan

- Production of SM background MC samples will be carried out by the EF MC Production Team
    - Membership is being worked out now
    - Please contact me if you would like to volunteer
        - All levels of experience welcome
- Goal - support studies for these future detector benchmarks:

Machine	Energy							
CEPC	m_Z	$2m_W$	240					GeV
FCC-ee	m_Z	$2m_W$	240	$2m_t$				
ILC	250	350		500	1000			
CLIC			380			1500	3000	
HL-LHC/FCC-hh	14	75	100	150				TeV
LHeC/FCC-eh	1.3	3.5						
μμ	3	10	14	30				

Simulation Requirements

- Wide variety of studies anticipated within EF
  - Different types of MC needed for different purposes

accuracy ↑ / speed ↓

full sim

detector R&D/design
unconventional signatures
most-critical studies, i.e. Higgs couplings

Delphes/SGV

studies requiring large statistics
comprehensive surveys of machines

truth-level

Full Simulation

- All of the future collider study groups will permit outside collaborators to access to existing full-simulation samples
    - Minor impositions: Generally, abide by publication rules (present and discuss preliminary results at internal study group meetings)
- Generally willing to produce (limited) new MC samples, if needed (and strongly-motivated)
    - Needs/motivation should be discussed with relevant Topical Group Conveners → MC Production Team → future collider study group MC Contact

Full Simulation

- Usage of "foreign" MC/simulation framework involves some degree of a learning curve
  - Organized MC/Simulation Framework Tutorial Series to facilitate this process
    - Recorded for posterity (linked from timetable contributions)

Machine	Date	Link
ILC	Aug 28	https://indico.fnal.gov/event/45031/
CEPC	Sept 8	https://indico.fnal.gov/event/45183/
FCC-ee/hh	Sept 22-23	https://indico.cern.ch/event/945608/
Whizard for e+e-	Sept 28	https://indico.fnal.gov/event/45413/
FCC-ee/hh	Sept 29	https://indico.cern.ch/event/949950/
Muon Collider	Sep 30	https://indico.fnal.gov/event/45187/
LHeC/FCC-eh	Oct 13	https://indico.fnal.gov/event/45185/
ILC Analysis Walkthrough	Oct 14	https://indico.fnal.gov/event/45721/

Today

Fast Simulation

- Delphes is widely-used, well-known, fast, accurate enough for many studies
- Detector cards exist for all proposed machines (some only very recently developed - might be tweaked)
  - Details on "blessed" cards available on TF wiki
- Will use Delphes for production of large SM background MC samples for all collider benchmarks
  - Will re-use existing truth-level events whenever possible

Circular ee Colliders

Machine	Energy							
CEPC	m_Z	$2m_W$	240					GeV
FCC-ee	m_Z	$2m_W$	240	$2m_t$				
ILC	250	350		500	1000			
CLIC			380			1500	3000	
HL-LHC/FCC-hh	14	75	100	150				TeV
LHeC/FCC-eh	1.3	3.5						
μμ	3	10	14	30				

- (Mostly) common energies
- Similar beam characteristics and detector performance expected for CEPC and FCC-ee

Circular ee Colliders

Machine	Energy							
Circular ee	m_Z	$2m_W$	240	$2m_t$				
ILC	250	350		500	1000			GeV
CLIC			380			1500	3000	
HL-LHC/FCC-hh	14	75	100	150				
LHeC/FCC-eh	1.3	3.5						TeV
μμ	3	10	14	30				

truth-level events:
CEPC
FCC-ee

- (Mostly) common energies
- Similar beam characteristics and detector performance expected for CEPC and FCC-ee
- Finalizing common benchmark detector card

Linear ee Colliders

Machine	Energy							
Circular ee	m_Z	$2m_W$	240	$2m_t$				
ILC	250	350		500	1000			GeV
CLIC			380			1500	3000	
HL-LHC/FCC-hh	14	75	100	150				
LHeC/FCC-eh	1.3	3.5						TeV
μμ	3	10	14	30				

truth-level events:
CEPC
FCC-ee

- No common energies
- Beam characteristics and detector performance not as similar

Linear ee Colliders

Machine	Energy							
Circular ee	m_Z	$2m_W$	240	$2m_t$				GeV
ILC	250	350		500	1000			GeV
CLIC			380			1500	3000	GeV
HL-LHC/FCC-hh	14	75	100	150				TeV
LHeC/FCC-eh	1.3	3.5						TeV
μμ	3	10	14	30				TeV

truth-level events:
CEPC
FCC-ee

- No common energies
- Beam characteristics and detector performance not as similar
- Will not adopt common benchmark detector card
- Don't believe it is necessary to generate backgrounds at 380 GeV

Hadron Colliders

Machine	Energy						
Circular ee	m_Z	$2m_W$	240	$2m_t$			
ILC	250	350	500	1000			GeV
CLIC					1500	3000	
HL-LHC/FCC-hh	14	75	100	150			
LHeC/FCC-eh	1.3	3.5					TeV
μμ	3	10	14	30			

truth-level events:
CEPC
FCC-ee

- In addition to the 100 TeV benchmark, we will produce background samples at 75 and 150 TeV using the FCC-hh detector card
- For phenomenological studies, we will also generate background samples at 14 TeV using the same card
  - ATLAS and CMS folks interested in doing HL-LHC studies should go through their respective collaborations

Hadron Colliders

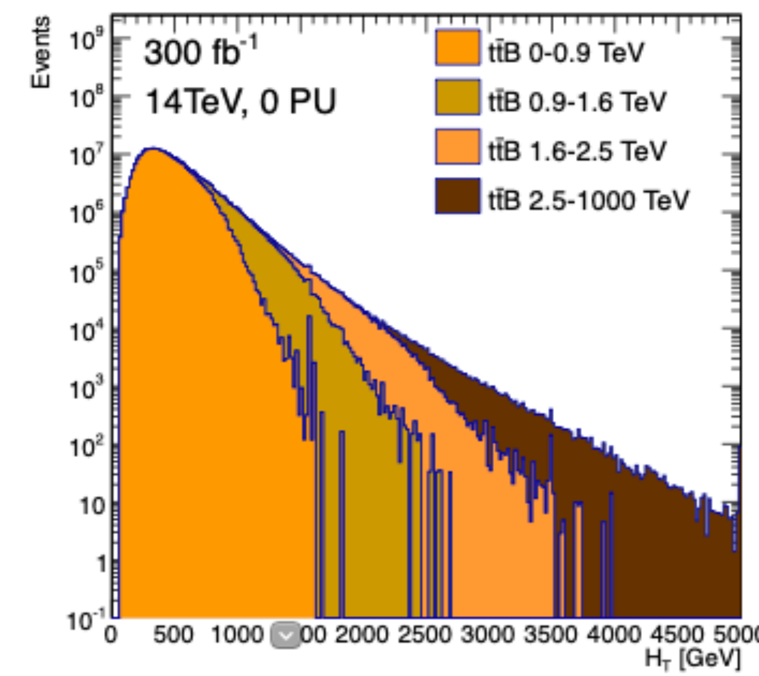
- Need a coherent strategy to produce wide set of representative SM backgrounds with sufficient statistics in tails of many distributions, with minimal complexity/book keeping
  - Will use similar strategy to that adopted for Snowmass 2013
    - Combine processes with similar cross sections in single MC sample
    - On-shell internal propagators excluded → fully orthogonal
    - On-shell heavy resonances treated as stable
      - Subsequently decayed democratically to leptons+hadrons using BRIDGE
    - Store event weight: σ_{LO} * NLO k-factor * w_{BR}

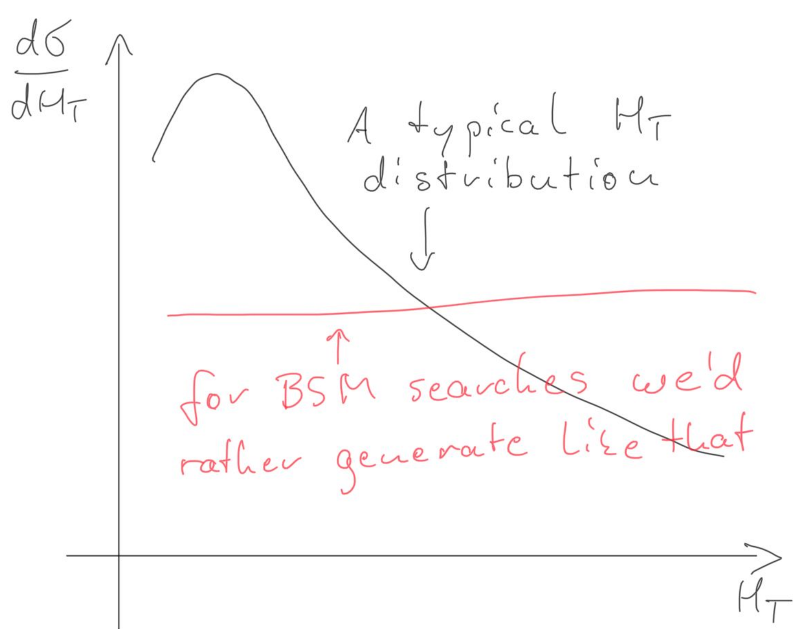
Dataset name	Physics process	Number of recoil jets
B-4p	γ or on-shell W, Z	0
Bj-4p	γ or on-shell W, Z	1-3
Bjj-vbf-4p	γ or off-shell W, Z, H in VBF topology	2-3
BB-4p	Diboson (γ, W, Z) processes	0-2
BBB-4p	Tri-boson (γ, W, Z) processes including BH	0-1
LL-4p	Non-resonant dileptons (including neutrinos) with $m_{ll} > 20$ GeV	0-2
LLB-4p	Non-resonant dileptons with an on-shell boson, $m_{ll} > 20$ GeV	0-1
H-4p	Higgs	0-3
tj-4p	Single top (s- and t-channel)	0-2
tB-4p	Single top associated with a boson	0-2
tt-4p	$t\bar{t}$ pair production	0-2
ttB-4p	$t\bar{t}$ associated with γ, W, Z, H	0-1

Table 1-2. *Table of background processes. All processes include the particles in the dataset name plus additional recoil jets up to four generated particles. On-shell vector bosons, off-shell dileptons, Higgs bosons, top quarks, and jets are denoted B, LL, H, t, and j, respectively.*

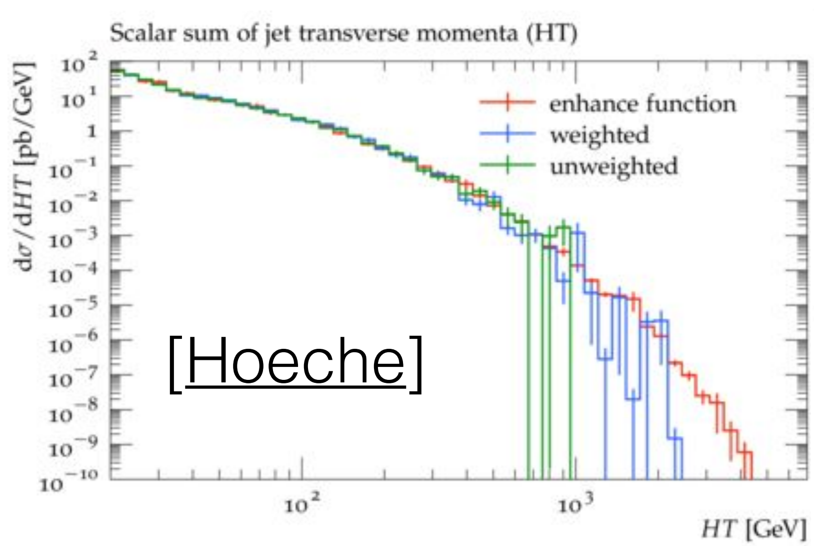
Hadron Colliders

- One significant improvement since 2013
  - Rather than binning samples in H_T, use weighted events or generate with some functional bias (pending validation)
    - Will also correct for this with event weight

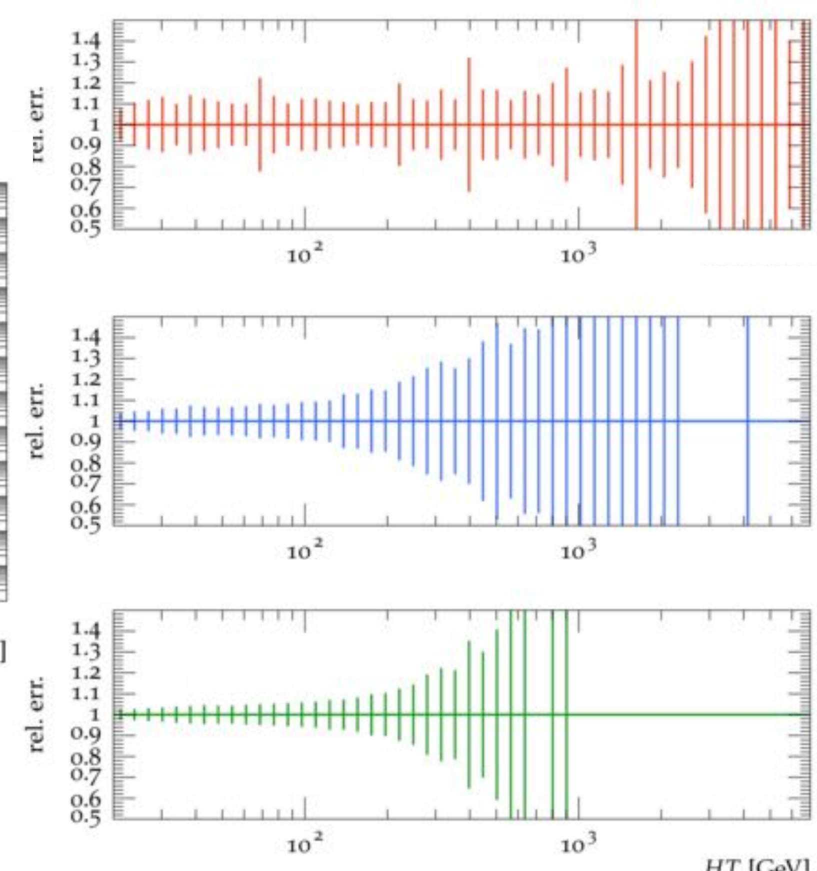




Example: Jet-HT



- Each distribution generated with 100k events
- Timing: EF 1168s, W 572s, UW 944s



Other Colliders

Machine	Energy						
Circular ee	m_Z	$2m_W$	240	$2m_t$			GeV
ILC	250	350	500	1000			
CLIC					1500	3000	
HL-LHC/FCC-hh	14	75	100	150			TeV
LHeC/FCC-eh	1.3	3.5					
μμ	3	10	14	30			

truth-level events:
CEPC
FCC-ee

- Also plan to support studies for the eh and μμ collider benchmarks
    - Energies are not set in stone

Production Procedure

- hh samples will be generated proactively
  - From event generation through Delphes
- ee, eh, and μμ samples will be generated on demand
  - Requests can be made here
    - Requests will be accepted on a rolling basis, but those received by the end of August will receive priority
  - With the exception of requests for very large statistics, users will be expected to provide truth-level events (new or existing), MC production team will process with Delphes

Signal Production

- Signal production is specialized and not very resource intensive → no need to centralize
  - MC Production Team will provide instructions to generate signal MC samples with the same parameters as the background samples, but will not directly produce signal MC
- If available computing resources are not sufficient, users can utilize the Open Science Grid (OSG)

Computing

- OSG has kindly offered to support Snowmass and supply computing resources
  - Interactive login node + batch (opportunistic) job submission + storage
- Background MC will be produced opportunistically, and hosted, on the OSG
- These resources are also available for other Snowmass use cases
  - Signal MC production, analysis jobs, etc.
  - We only ask that before submitting a large batch of jobs, discuss workflow with OSG computing folks to ensure workflow is okay
  - Request an account here
  - Documentation is available here
  - See Slack for support and discussion

Validation

- Authors of Delphes cards (or MC Production Team) will perform validation to ensure performance matches that of full simulation at the few percent level
- Analyzers should further validate by comparing some differential distributions in “nearby” region of phase space, for which full simulation results are publicly available

Additional Info

- Further details will be provided in a written document (forthcoming)
- The production plan is briefly summarized on the MC Task wiki
  - Also includes some additional useful info (will be expanded soon)
    - Summary of existing MC samples, details of full simulation frameworks, and contacts for each proposed collider:

Which proposed collider are you responding on behalf of?	What signal/background MC samples have been produced thus far (using either fast or full simulation)?	Where are they located?	Who within the collaboration should members of the Energy Frontier contact with further questions regarding MC/simulation frameworks?	Is there some place where additional information related to MC/simulation frameworks can be found?
ILC	generator-level event samples, stdhep format for $\sqrt{s}$ = 250GeV, 350 GeV, 500 GeV, 1 TeV, further samples based on fast simultion (SGV) and full simulation (iLCSoft) are in preparation	On the GRID under VO ILC	jenny.list@desy.de	https://arxiv.org/abs/2007.03650
CLIC	See: https://twiki.cern.ch/twiki/bin/view/CLIC/MonteCarloSamplesForCLICdet	CERN EOS Storage	Please contact us at clicdp-snowmass-samples-contacts@cern.ch	https://arxiv.org/abs/1812.07337 contains a section on the software, github.com/ilcsoft
CEPC	We have full simulation of CEPC ZH and SM background at 240GeV, 350GeV, and Z pole events. See CEPC Note http://cepcdoc.ihep.ac.cn/DocDB/0002/000203/002/CEPCNoteCover.pdf	With the support of the computing center of Institute of High Energy Physics, the CEPC samples are stored on the IHEP clusters.	manqi.ruan@ihep.ac.cn, ligang@ihep.ac.cn, yudan@ihep.ac.cn.	
FCC-ee	A limited number of useful e+e- event samples, processed through full CMS simulation and reconstruction, still exist (though producing again these events won't take very long).	On private areas	gerardo.ganis@cern.ch, clement.helsens@cern.ch, patrick.janot@cern.ch, patrizia.azzi@cern.ch	Same as FCC-hh (https://cds.cern.ch/record/2717892)
FCC-hh	Full and Delphes samples are listed here http://fcc-physics-events.web.cern.ch/fcc-physics-events/Delphesevents_fcc_v02.php http://fcc-physics-events.web.cern.ch/fcc-physics-events/FCCsim_v03.php	There are located on eos at CERN	Michele.Selvaggi@cern.ch, gerardo.ganis@cern.ch, clement.helsens@cern.ch	Yes, https://cds.cern.ch/record/2717892
LHeC/FCC-eh	Signal: several Higgs decay modes plus backgrounds	CERN and University servers	oliver.fischer@liverpool.ac.uk	No
Muon collider	Higgs to bb and b backgrounds	University of Padova Cloud	donatella lucchesi (donatella.lucchesi@pd.infn.it)	https://sites.google.com/site/muoncollider/home

Conclusion

- MC Task Force plan is nearly finalized
  - Please let us know if you have any comments/complaints/concerns, or if you would like to volunteer for the MC Production Team
- Background MC production should begin very soon
  - As in 2013, this may be an iterative process
  - Early feedback from analyzers will be very helpful

Backup

Snowmass 2013

- Ran MadGraph5 + Pythia6 + Delphes3 opportunistically on the OSG to produce large-statistic SM background MC samples for future pp colliders

Parameter	LHC	HL-LHC	HE-LHC	VLHC
Energy [TeV]	14	14	33	100
Mean additional interactions per crossing ($< \mu >$)	50	140	140	140
Integrated Luminosity [fb^{-1}]	300	3000	3000	3000

- Documentation:
  - MC Simulation: https://arxiv.org/abs/1308.1636
  - "Snowmass" detector: https://arxiv.org/abs/1309.1057
  - OSG production: https://arxiv.org/abs/1308.0843

Snowmass 2013

- Signal MC production isn't resource intensive
  - Provided analysts with recipe for production from LHE, as well as analysis pointers
- Studies for ILC, CLIC, etc. used their own frameworks/samples
- Common data format for all future pp machines facilitated easy analysis/comparison
  - Tune a few cuts and turn the crank
- These samples were useful well beyond Snowmass itself
  - I am still occasionally asked if they are accessible

Event Generation

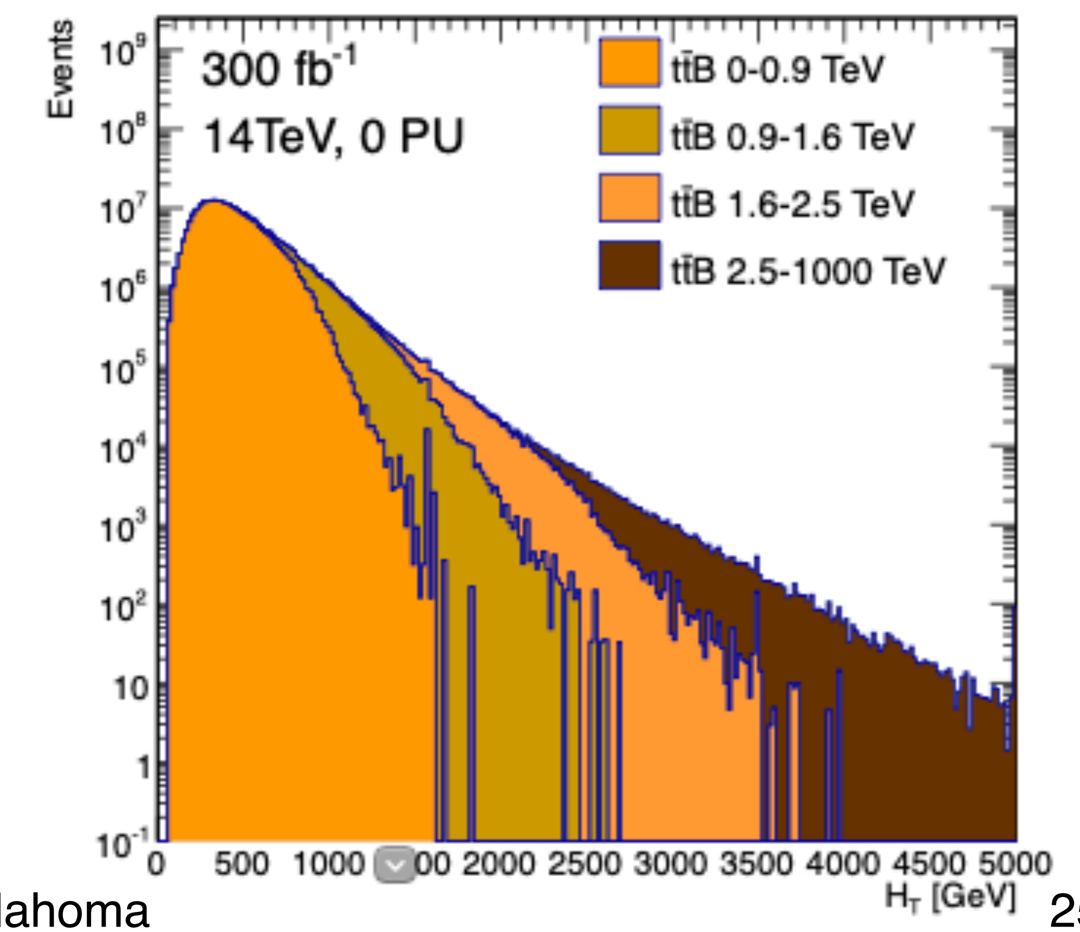
- With many background processes and E/PU combinations, adopted “container” scheme to simplify organization/book-keeping
  - Combined processes with similar cross sections in single MC sample
    - On-shell internal propagators excluded → fully orthogonal
  - On-shell heavy resonances treated as stable (decayed later w/ BRIDGE)
  - Up to 4 final state partons
- Each sample was binned in S_T^*: scalar p_T sum of all final state partons
  - One decade of cross section per bin (up to 7)

particle containers

$$J = \{g, u, \bar{u}, d, \bar{d}, s, \bar{s}, c, \bar{c}, b, \bar{b}\}$$
$$L = \{e^+, e^-, \mu^+, \mu^-, \tau^+, \tau^-, \nu_e, \nu_\mu, \nu_\tau\}$$
$$V = \{W^+, W^-, Z^0, \gamma\}$$
$$T = \{t, \bar{t}\}$$
$$H = \{h^0\}$$

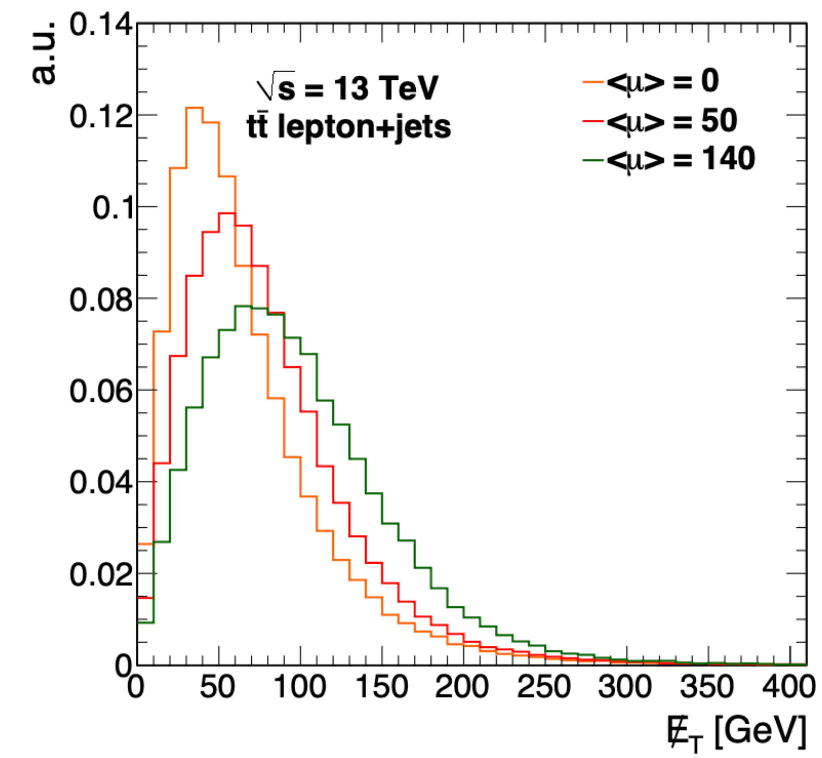
MC samples

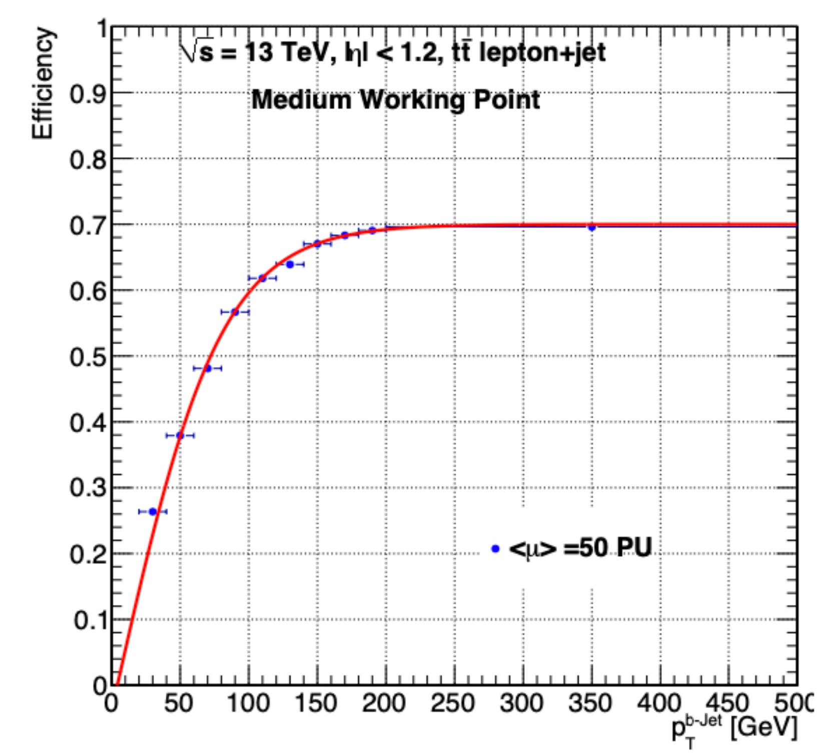
Dataset name	Physics process	Number of recoil jets
B-4p	γ or on-shell W, Z	0
Bj-4p	γ or on-shell W, Z	1-3
Bjj-vbf-4p	γ or off-shell W, Z, H in VBF topology	2-3
BB-4p	Diboson (γ, W, Z) processes	0-2
BBB-4p	Tri-boson (γ, W, Z) processes including BH	0-1
LL-4p	Non-resonant dileptons (including neutrinos) with $m_{ll} > 20$ GeV	0-2
LLB-4p	Non-resonant dileptons with an on-shell boson, $m_{ll} > 20$ GeV	0-1
H-4p	Higgs	0-3
tj-4p	Single top (s- and t-channel)	0-2
tB-4p	Single top associated with a boson	0-2
tt-4p	$t\bar{t}$ pair production	0-2
ttB-4p	$t\bar{t}$ associated with γ, W, Z, H	0-1

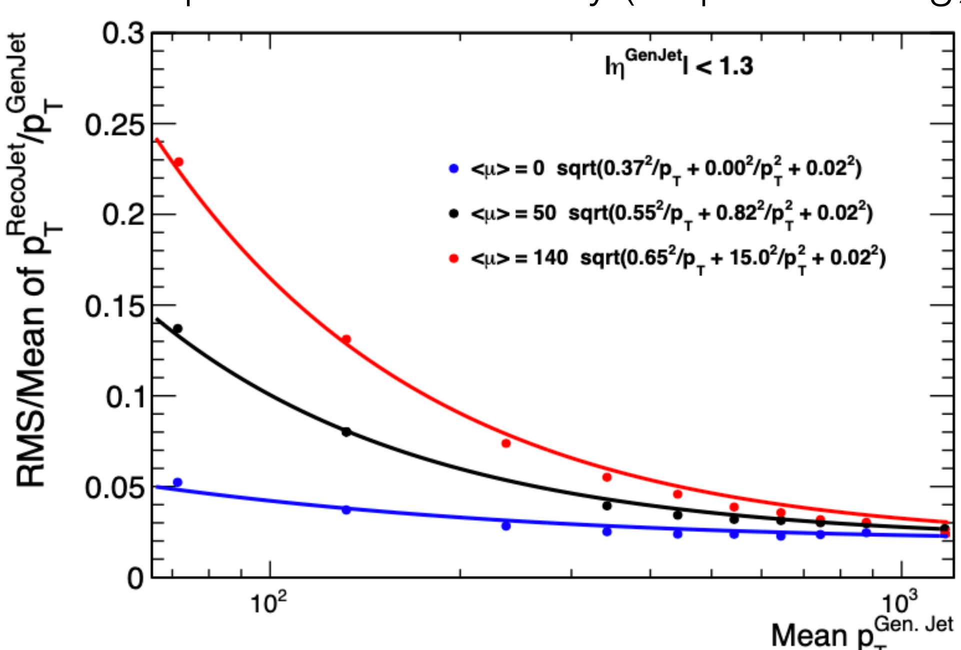


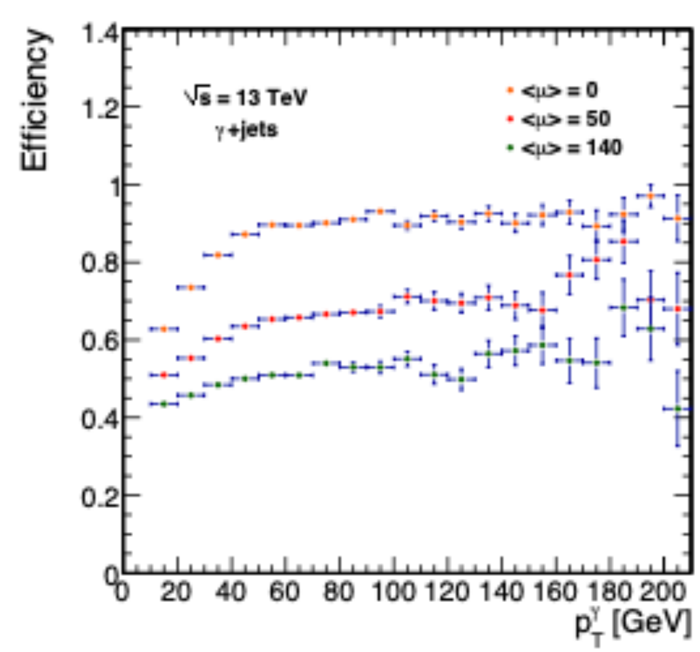
Detector Simulation/Reconstruction

- “Snowmass detector” implemented in Delphes
  - The best of both ATLAS and CMS
    - Performance taken from public documents, reflecting expected future upgrades
- Main simulation parameters (generally specified as p_T- and η-dependent functions):
  - Tracking efficiency (charged hadrons, e, μ)
  - Momentum resolution (charged hadrons, e, μ)
  - Calorimeter resolution (EM, hadronic clusters)
  - Reconstruction/tagging efficiency (e, μ, γ, b-jet, τ_h)
- Isolation determined by simulation
- PU suppression: charged hadron subtraction and area-based correction
- Developed new functionality (output slimming, jet grooming/substructure, t/V/H-tagging)









Event Weight

- Generator-level events produced at LO
  - NLO k-factor calculated from ratio of MCFM and MadGraph (inclusive) cross sections
- Used BRIDGE to decay heavy resonances democratically
  - Enhances statistics for rare decay modes
- σ_{LO} * k-factor * w_{BR} stored as event weight

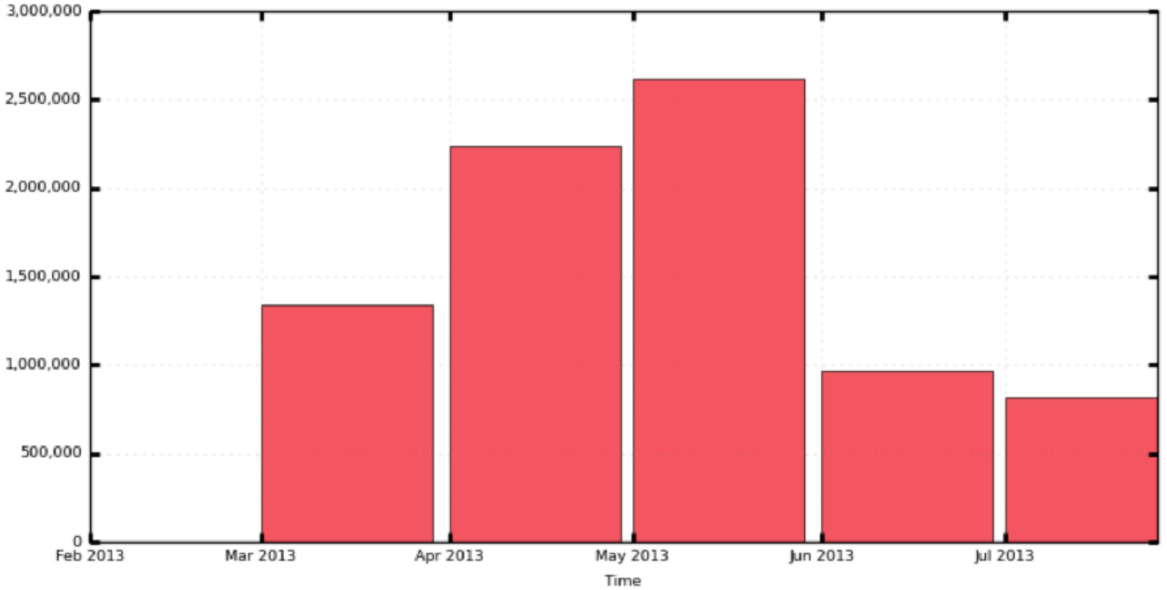
Process	$\sqrt{s} = 14$ TeV	$\sqrt{s} = 33$ TeV	$\sqrt{s} = 100$ TeV
$t\bar{t}$	1.24	1.10	0.96
$W^+ j$	1.17	0.85	0.74
$W^- j$	1.20	0.89	0.75
$Z^0 j$	1.17	0.87	0.76
γj	1.54	1.04	0.89
$W^+ W^-$	1.25	1.08	1.0
$W^+ Z^0$	1.24	1.06	0.95
$W^- Z^0$	1.26	1.09	0.97
$Z^0 Z^0$	1.37	1.29	1.21
$W^+ \gamma$	1.22	0.80	0.67
$W^- \gamma$	1.33	0.83	0.67
$Z^0 \gamma$	1.24	0.95	0.76
$\gamma \gamma$	1.34	1.08	0.98
$t W^-$	1.0	0.77	0.78
$\bar{t} W^+$	1.0	0.77	0.78
$t \bar{b}$	1.76	1.72	1.94
$\bar{t} b$	1.88	1.73	1.78
$\ell^+ \ell^-$	1.20	1.16	1.20

$t\bar{t} \to$	weight	in sample	change
hadronic	44%	25%	0.56
semi-leptonic	44%	50%	1.13
di-leptonic	11%	25%	2.25
$W^{\pm} Z^0 \to$	weight	in sample	change
1ℓ	30%	44%	1.4
2ℓ	6.7%	11%	1.6
3ℓ	3.3%	16%	4.8

Computing (OSG)

- Utilized opportunistic resources to produce ~0.5 billion events
    - ~14k jobs/day totaling ~890 CPU·years
    - Peak usage ≈ 100 kCPU·hours/day
- Job submission via GlideinWMS
- Software dependencies from CvmFS

CPU consumption [hours]



- MadGraph and Pythia/Delphes performed in 2 separate jobs
    - MadGraph
        - ~10 MB input gridpack (output LHE) transferred via HTCondor
        - Responsible for most of the CPU usage
    - Pythia/Delphes
        - 1 GB minimum bias file pre-staged to storage nodes at 10 grid sites
        - Outputs (5-20 kB/event) transferred to FNAL, BNL, UNL
            - UNL was accessible (web and XRootD) without grid certificate (theorists)

data transfer

Month	Fermi dCache (TB)	UNL (TB)
June	65.0	46.4
May	12.4	5.2
April	189.7	10.8
March	1.1	0.0
Total	268.3	62.5