Monte Carlo Task Force: Status Report

John Stupak
on behalf of the Snowmass EF MC Task Force

7/20/20

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	Snowmass2013
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

[Narain]

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

[Narain]

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 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.

Work In Progress

8

collaboration and community surveys distributed on May 31

[Narain]

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

[Narain]

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".

Work In Progress

Community Survey

Work In Progress

9

collaboration and community surveys distributed on May 31

Activities

Collaboration Survey

Full Simulation

Delphes

Which proposed collider are you responding on behalf of?	Have you developed a simulation framework based on GEANT?	Is this framework publicly-available? If so, where?	Have you developed a DELPHES detector card (or other form of fast and/or parameterized simulation)?	Is this card publicly-available? If so, where?
ILC	Yes.	https://github.com/iLCSoft	Work is currently ongoing.	https://github.com/ILDAnaSoft/ILDDelphes (still WIP)
CLIC	Yes, DD4hep+DDG4+ddsim	e.g: /cvmfs/clicdp.cern.ch/iLCSoft/builds/2020-02-07/x86_64-slc6-gcc62-opt/init_ilcsoft.sh github.com/aidasoft/dd4hep	yes, CLICdet cards Stage1/2/3, documentation in https://twiki.cern.ch/twiki/bin/view/CLIC/DelphesMadgraphForBSMReport and https://arxiv.org/abs/1909.12728 (contains concrete instructions)	Yes in the delphes code
CEPC	Yes, CEPC has the Mokka-C, a Geant-4 Full simulation toolkit.	The CEPCsoft are capsuled into a package on cvmfs, the installation can be found at http://cepcsoft.ihep.ac.cn/guides/scratch/docs/cvmfs/	Yes, we have developed a Delphes card, which is also validated on the Full simulation tools. In addition, many small fast simulation tools are developed.	The Delphes card is public available, see https://arxiv.org/abs/1712.09517.
FCC-ee	Yes	Yes, the code is on github: https://github.com/HEP-FCC/ and it is installed centrally on cvmfs (same as FCC-ee). Also for the CDR, full simulation samples where produced with CMSSW.	Yes	Yes, several cards for e+e- detector concepts are available on the delphes GitHub (https://github.com/delphes/delphes/tree/master/cards/).
FCC-hh	yes	Yes, the code is on github: https://github.com/HEP-FCC/ and it is installed centrally on cvmfs	yes	Yes, it is available for FCC-hh on the FCCSW website http://hep-fcc.github.io/FCCSW/
LHeC/FCC-eh	Yes: DD4HEP interface to GEANT	Yes: http://dd4hep.web.cern.ch/dd4hep/	Yes	Not yet
Muon collider	Yes	yes, github	No DELPHES, no parametric simulation	no

Available here:
https://docs.google.com/spreadsheets/d/19KWScsrEgmHRBtqq3tKxHiREEbT0e_IC3DIPcTdEixc/edit?usp=sharing

Collaboration Survey

MC Access

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?	Can they be accessed by external collaborators? If so, how?
ILC	generator-level event samples, stdhep format for √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	If they have access to VO ILC. For full simulation samples an ILD or SiD guest membership will be required.
CLIC	See: https://twiki.cern.ch/twiki/bin/view/CLIC/MonteCarloSamplesForCLICdet	CERN EOS Storage	Access can be granted for CERN account holders, membership to the ILC VO can probably be granted without too much hassle. Files could be transfered to other StorageElements if they allow the ILC vo access, if that would help
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.	Currently, an IHEP account is needed to access these samples. You can contact us and we will be happy to apply an account for you.
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	Possibly on demand, in a way to be defined
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	Yes, either by subscribing to dedicated e-groups at CERN, or could setup web based but less prefered
LHeC/FCC-eh	Signal: several Higgs decay modes plus backgrounds	CERN and University servers	Possibly upon request
Muon collider	Higgs to bb and b backgrounds	University of Padova Cloud	Yes

Available here:
https://docs.google.com/spreadsheets/d/19KWScsrEgmHRBtqq3tKxHiREEbT0e_IC3DIPcTdEixc/edit?usp=sharing

Collaboration Survey

Documentation & Contacts

Which proposed collider are you responding on behalf of?	Is any documentation available to facilitate analysis of these samples?	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	In preparation.	jenny.list@desy.de	https://arxiv.org/abs/2007.03650
CLIC	LCIO API lcio.desy.de, The full simulation is based on the Marlin framework using LCIO for input output. The collections *SelectedPandoraPFOs contain reconstructed particle flow objects that can be used as input to JetClustering, Vertexing or Flavourtagging, depending on the requirements of the analysis. LCFIPlus (https://github.com/lcfiplus/LCFIPlus), FastJet(https://github.com/iLCSoft/MarlinFastJet)	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	Yes, the samples description can be found in http://cepcdoc.ihep.ac.cn/DocDB/0002/000203/002/CEPCNoteCover.pdf. There is also a CEPC software webpage (http://cepcsoft.ihep.ac.cn), describing the overview of all the softwares we use and some illustration. A CEPC notes database (http://cepcdoc.ihep.ac.cn/cgi-bin/DocDB/DocumentDatabase) with plenty of existing analysis can be also used as an indication.	manqi.ruan@ihep.ac.cn, ligang@ihep.ac.cn, yudan@ihep.ac.cn.	
FCC-ee	No	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	Yes, https://cds.cern.ch/record/2717892	Michele.Selvaggi@cern.ch, gerardo.ganis@cern.ch, clement.helsens@cern.ch	Yes, https://cds.cern.ch/record/2717892
LHeC/FCC-eh	Only in talks at workshops	oliver.fischer@liverpool.ac.uk	No
Muon collider	Yes	donatella lucchesi (donatella.lucchesi@pd.infn.it)	https://sites.google.com/site/muoncollider/home

Available here:
https://docs.google.com/spreadsheets/d/19KWScsrEgmHRBtqq3tKxHiREEbT0e_IC3DIPcTdEixc/edit?usp=sharing

Grossly Oversimplified Results

✅ = yes ⚠️ = somewhat / yes with caveats 🚫 = no / not yet

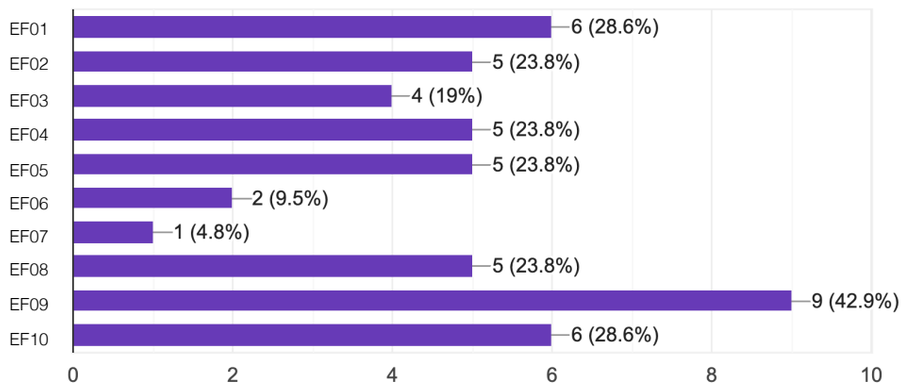
	Full Sim	Delphes Card	Comprehensive SM Bkgd MC	Publicly Available
ILC	✅	⚠️ Not yet, but SGV	✅	⚠️ "Guest" membership required for full sim
CLIC	✅	✅	✅	⚠️ ILC VO only
CEPC	✅	✅	⚠️ Full SM for H run (240 GeV)	⚠️ IHEP account required
FCC-ee	✅	✅	🚫	⚠️ CERN account required
FCC-hh	✅	✅	✅	⚠️ CERN account required
FCC-eh/LHeC	✅	✅	⚠️	⚠️ "Light" membership required
Muon Collider	✅	🚫	🚫	⚠️ INFN Padova account required

Community Survey

- Advertised via EF LISTSERV, Slack, TGCs
- Original deadline was June 9th
  - Still open, please provide your input!
    - If your plans evolve/crystalize later, responses are editable

Which topical groups might your planned studies fall within?

21 responses



More representative sample needed in order to develop concrete plans

https://forms.gle/RMxeXXKcjvPbHicTA

Snowmass Energy Frontier MC Task Force - Community Survey

The success of the Snowmass Energy Frontier effort depends crucially on the analysis of a wide variety of MC samples for various future colliders/detectors. Collaborations for future collider proposals have produced many such samples, using simulation frameworks of varying degrees of sophistication. However, production of additional samples may be necessary to conduct proposed studies for Snowmass 2021.

The purpose of this form is to gather community input regarding the necessity and availability of MC samples for Snowmass 2021. This information will be used to determine if a centralized effort to produce MC samples is required, and if so, how to maximize the utility of the resulting MC (within computing and other constraints).

Before completing this survey, please review the following google doc, which summarizes the current availability of MC samples for various proposed machines:
https://docs.google.com/spreadsheets/d/19KWScsrEgmHRBtqq3tKxHiREEbT0e_lC3DIPcTdEixc/edit?usp=sharing

*Required

link to collaboration survey results

Email address *

Your email address

What is your name? *

Your answer

Which topical groups might your planned studies fall within? *

- [ ] EF01: EW Physics: Higgs Boson properties and couplings
- [ ] EF02: EW Physics: Higgs Boson as a portal to new physics

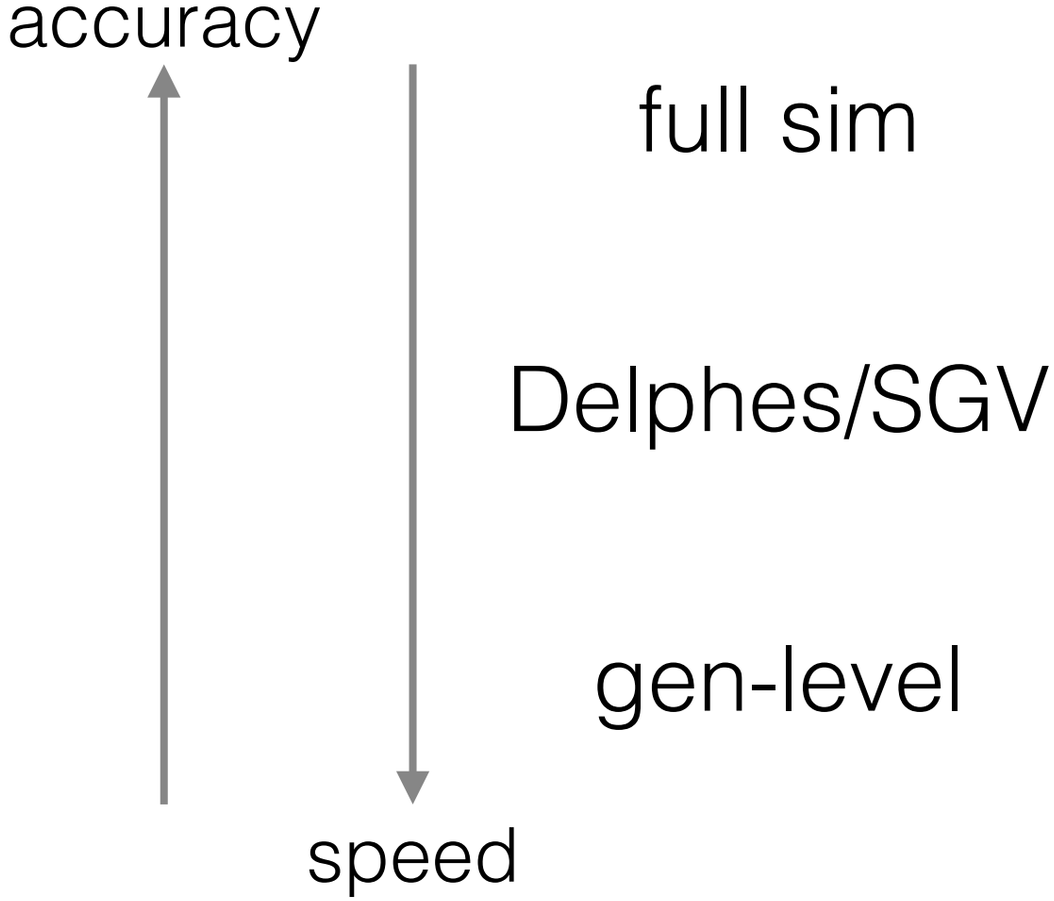
Computing

- Preparing for production of MC on the OSG
  - ≳10 TB storage (~200-300 TB soon)
    - XRootD accessible, no read restrictions
  - Login and job submission server deployed at UChicago
    - Configured to use HTCondor
- Web interface for account management (WIP)
  - Folks will soon be able to request accounts for signal production, analysis, etc.
  - Developing mechanism to run Jupyter notebooks via web interface

Preliminary Plans

MC Simulation

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



Full Sim

- Goal: Facilitate access to, and analysis of, existing MC samples
  - Hosted by various collaborations, with minimal access restrictions
- Access has been achieved
  - Thank you to all collaborations for their cooperation and help to make Snowmass 2021 a success!

	Publicly Available
ILC	⚠ "Guest" membership required for full sim
CLIC	⚠ CERN account required
CEPC	⚠ IHEP account required
FCC-ee	⚠ CERN account required
FCC-hh	⚠ CERN account required
FCC-eh/LHeC	⚠ "Light" membership required
Muon Collider	⚠ INFN Padova account required

MC/Analysis Tutorials

- Planning to hold a series of tutorials for future machines (WIP)
  - Possibly 2-3 hours/machine * 7 machines
  - **Hands on**
  - Led by experts from within the collaborations
  - Covering:
    - Available MC samples
    - Structure of the MC
    - Signal MC generation
    - Analysis examples
  - Sometime mid-August through September (could be repeated, if useful)

Delphes

- Goal: Comprehensive SM background samples with Delphes available to enable studies at each of the energies listed below
  - Hosted on OSG (or elsewhere) with no access restrictions
- Considering using HepSim for organization/metadata storage/distribution
- We plan to provide computing resources and instructions for signal MC production

	Type	Energy						
ILC	linear ee	250	350	500	1000			GeV
CLIC			380		1000	1500	3000	
CEPC	circular ee	m_Z	$2m_W$	240				
FCC-ee		m_Z	$2m_W$	240	$2m_t$			
FCC-hh	hh	100						TeV
FCC-eh/LHeC	eh	1.3	3.5					

NB: we are very supportive of muon collider studies as well, but not clear yet if Delphes is appropriate

Delphes

- Questions:

	Type	Energy						
ILC	linear ee	250	350	500	1000			GeV
CLIC			380		1000	1500	3000	
CEPC	circular ee	m_Z	$2m_W$	240				
FCC-ee		m_Z	$2m_W$	240	$2m_t$			
FCC-hh	hh	100						TeV
FCC-eh/LHeC	eh	1.3	3.5					

Delphes

- Questions:
  - Do we need two sets of MC for different detectors at the same energy?

	Type	Energy						
ILC	linear ee	250	350	500	1000			GeV
CLIC			~~380~~		~~1000~~	1500	3000	
CEPC	circular ee	m_Z	$2m_W$	240				
FCC-ee		~~m_Z~~	$2m_W$	~~240~~	$2m_t$			
FCC-hh	hh	100						TeV
FCC-eh/LHeC	eh	1.3	3.5					

Delphes

- Questions:
  - Do we need two sets of MC for different detectors at the same energy?
  - Can we neglect some energies?

	Type	Energy						
ILC	linear ee	250	350	500	1000			GeV
CLIC			~~380~~		~~1000~~	1500	3000	
CEPC	circular ee	m_Z	~~$2m_W$~~	240				
FCC-ee		~~m_Z~~	~~$2m_W$~~	~~240~~	$2m_t$			
FCC-hh	hh	100						TeV
FCC-eh/LHeC	eh	1.3	3.5					

Delphes

- Questions:
  - Do we need two sets of MC for different detectors at the same energy?
  - Can we neglect some energies?
  - To factorize detector performance and collider parameters, should we use a generic parameterization of detector performance for ee colliders?
    - Or perhaps separate circular and linear benchmarks?

	Energy								
generic ee detector	m_Z	$2m_W$	240/250	350/380	500	1000	1500	3000	GeV
FCC-hh	100								TeV
FCC-eh/LHeC	1.3	3.5							

Summary

- TF is working hard to fulfill its mandate
  - Full simulation: samples are available with minimal hurdles
  - Fast simulation:
    - Preliminary (broad) plan in place
    - Need more input from community to finalize details
      - Survey: https://forms.gle/RMxeXXKcjvPbHicTA
    - Preparing for MC production on OSG

Discussion

- Delphes
  - Do we need two sets of MC for different detectors at the same energy?
  - Can we neglect some energies?
  - To factorize detector performance and collider parameters, should we use a generic parameterization of detector performance for ee colliders?
- OSG data access: what is your expected use case?
  - Stream via XRootD?
  - Copy to local cluster?
  - Run condor jobs on OSG?
- Is there a need for an open meeting to discuss these issues in more detail?

Backup

https://arxiv.org/abs/1905.03764

Collider	Type	$\sqrt{s}$	$\mathscr{P}$ [%] $[e^-/e^+]$	N(Det.)	$\mathscr{L}_{\text{inst}}$ $[10^{34}]$ cm^{-2}s^{-1}	$\mathscr{L}$ [ab^{-1}]	Time [years]	Refs.	Abbreviation
HL-LHC	pp	14 TeV	-	2	5	6.0	12	[13]	HL-LHC
HE-LHC	pp	27 TeV	-	2	16	15.0	20	[13]	HE-LHC
FCC-hh(*)	pp	100 TeV	-	2	30	30.0	25	[1]	FCC-hh
FCC-ee	ee	M_Z	0/0	2	100/200	150	4	[1]	
		$2M_W$	0/0	2	25	10	1-2		
		240 GeV	0/0	2	7	5	3		FCC-ee$_{240}$
		$2m_{top}$	0/0	2	0.8/1.4	1.5	5		FCC-ee$_{365}$
							(+1)	(1y SD before $2m_{top}$ run)	
ILC	ee	250 GeV	±80/±30	1	1.35/2.7	2.0	11.5	[3,14]	ILC$_{250}$
		350 GeV	±80/±30	1	1.6	0.2	1		ILC$_{350}$
		500 GeV	±80/±30	1	1.8/3.6	4.0	8.5		ILC$_{500}$
							(+1)	(1y SD after 250 GeV run)	
		1000 GeV	±80/±20	1	3.6/7.2	8.0	8.5	[4]	ILC$_{1000}$
							(+1-2)	(1-2y SD after 500 GeV run)	
CEPC	ee	M_Z	0/0	2	17/32	16	2	[2]	CEPC
		$2M_W$	0/0	2	10	2.6	1		
		240 GeV	0/0	2	3	5.6	7		
CLIC	ee	380 GeV	±80/0	1	1.5	1.0	8	[15]	CLIC$_{380}$
		1.5 TeV	±80/0	1	3.7	2.5	7		CLIC$_{1500}$
		3.0 TeV	±80/0	1	6.0	5.0	8		CLIC$_{3000}$
							(+4)	(2y SDs between energy stages)	
LHeC	ep	1.3 TeV	-	1	0.8	1.0	15	[12]	LHeC
HE-LHeC	ep	1.8 TeV	-	1	1.5	2.0	20	[1]	HE-LHeC
FCC-eh	ep	3.5 TeV	-	1	1.5	2.0	25	[1]	FCC-eh

Inputs from Collaborations: ILC

We envision 4 levels of participation:

Fast simulation: A fast-simulation framework and large samples of SM events at ECM = 250, 350, 500, and 1000 GeV. We are compiling a list of study questions covering all aspects of ILC physics.

Full simulation: Corresponding full simulation data in a high-level analysis format that can be used directly in root.

Detailed detector: For detailed detector design, one should learn the ILCSoft simulation framework. The SiD and ILD groups will offer no-cost guest memberships.

Detector R&D: A comprehensive list of current R&D projects is available; scan this list for potential collaborators. The R&D collaborations invite new members.

ILC

https://arxiv.org/abs/2007.03650

Summary

- **available already:**
  - **DATA: generator-level event samples 250 GeV, 350 GeV, 500 GeV, 1 TeV**
  - **TOOL: SGV fast simulation**
- **in preparation, hopefully in place by ~mid July:**
  - **DATA: SGV-miniDST of the above generator-level samples**
  - **TOOL: Delphes card for a “generic ILC detector”**
- **in preparation, coming during ~fall:**
  - **DATA: ILD-miniDST of new 250 GeV samples, other energies tbd**
  - **TOOL: miniDST output for Delphes ?**
- **data with “ILD” in the name will require at least “guest membership”**

CEPC

https://drive.google.com/file/d/1jIlxKS4c_cwOcRMn-tKfW7yloUtxNEIF/view?usp=sharing

CDR Samples & consumed CPU time

	ZH	ZZ	SM bkg (2f+4f)	Zpole
event number	1.14m	6.39m	801m+101m	10^12
generated number	1.32m+5m(with different final states)	5.9m	26m+91m	54.6m
time	2.5w	2.36w	5.2w+36.4w	10.9w
Simu	12T	6.9T	22T+104T	38T
	12T	6.9T	22T+104T	———————
Reco	No data (if need to be stored, ~130% simu size)			
Dst	1.1T	767G	4.3T+11.2T	4.4T
	854G	740G	2.9T+5.1T	———————

In total: ~ o(1P) data generated, using several months with up to 1k CPUs

FCC-hh

http://cds.cern.ch/record/2717892

CERN-FCC-PHYS-2020-0005

Future Circular Collider

PUBLICATION

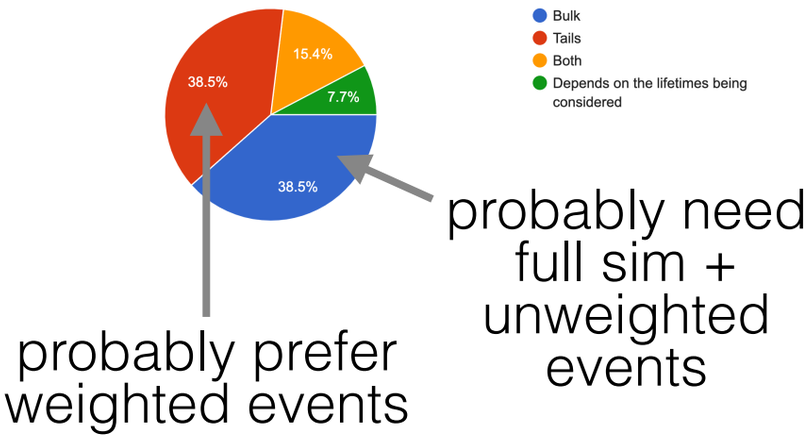
A framework and goals for FCC-hh physics studies at Snowmass 2021

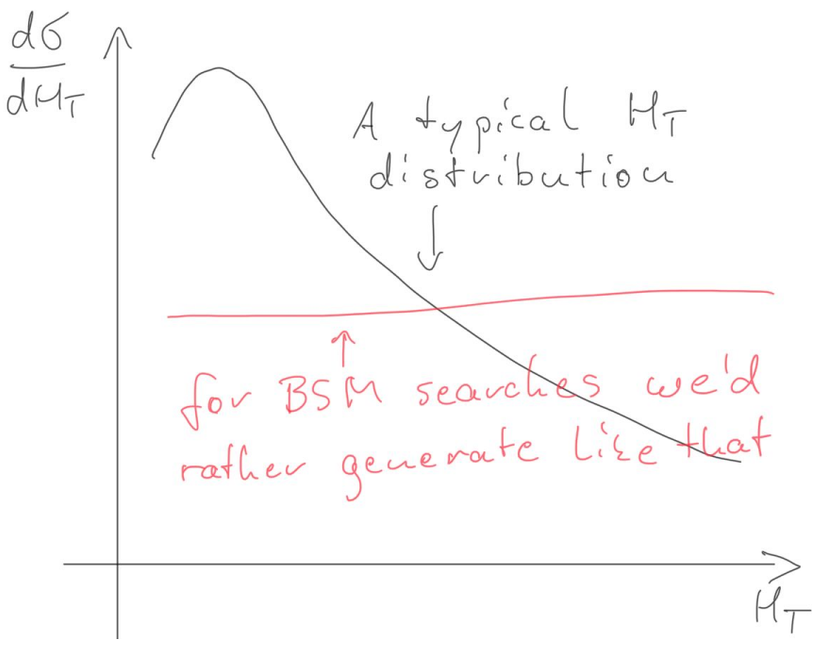
Helsens, Clement (CERN) *et al.*

12 May 2020

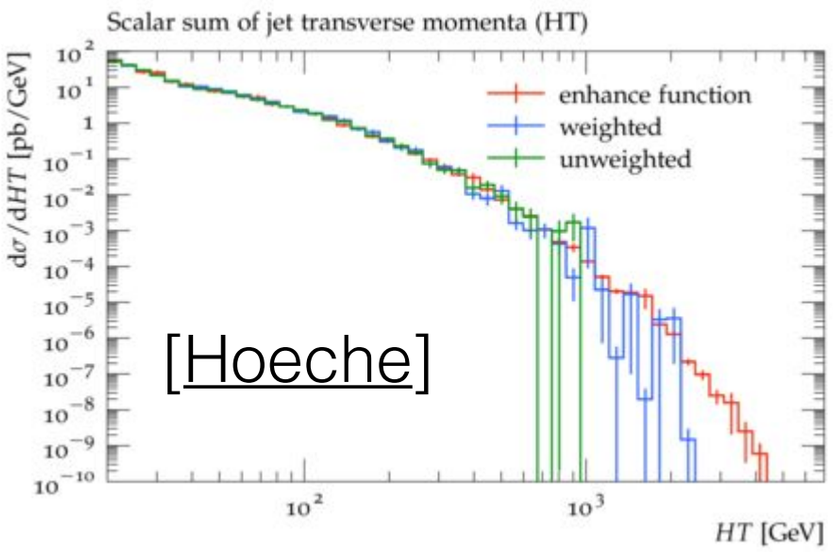
(Un)Weighted MC

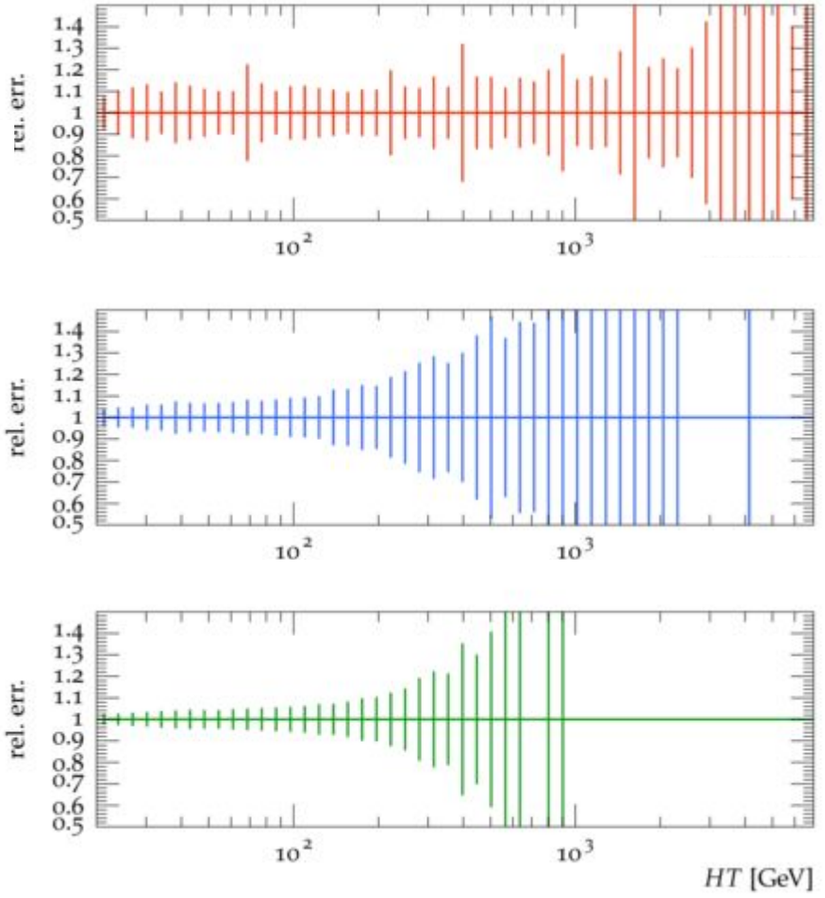
Broadly speaking, are your planned studies more sensitive to MC statistics in the bulk or tails of distributions?
13 responses





Example: Jet-HT





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

Snowmass 2013

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

Challenge

- σ_{ttbar} @ 14 TeV = 1 nb
  - For L_{eff} = 10 * 3/ab: N = 3E10
    - 1 kB/event → 30 TB
    - 1 minute/event → 50k CPU·years
- Generated weighted events to reduce statistics required by ~2000x
- Used parameterized detector simulation to reduce CPU consumption to manageable level
- Utilized opportunistic OSG resources

Event Generation

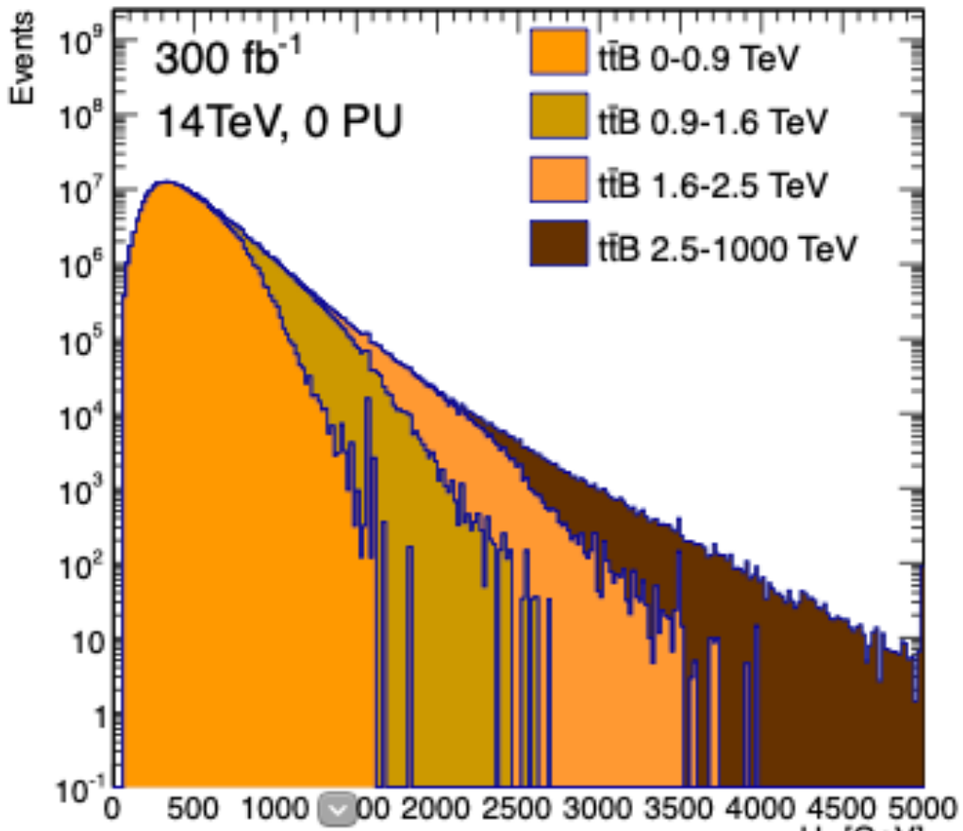
- With many background processes and E/PU combinations, adopted "container" scheme to simplify organization/book-keeping
  - Combined processes with similar cross sections in single MC sample
    - On-shell internal propagators excluded → fully orthogonal
  - On-shell heavy resonances treated as stable (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

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

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



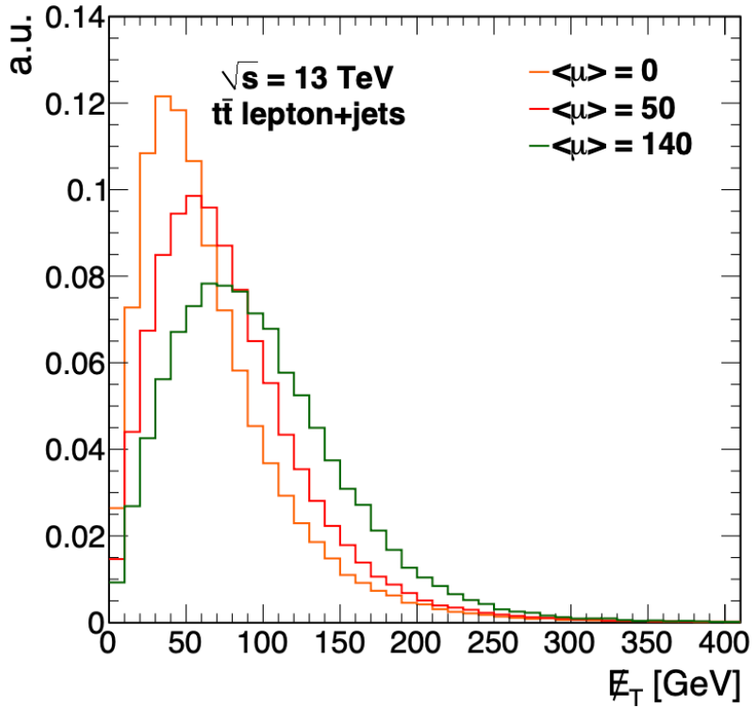
Pythia

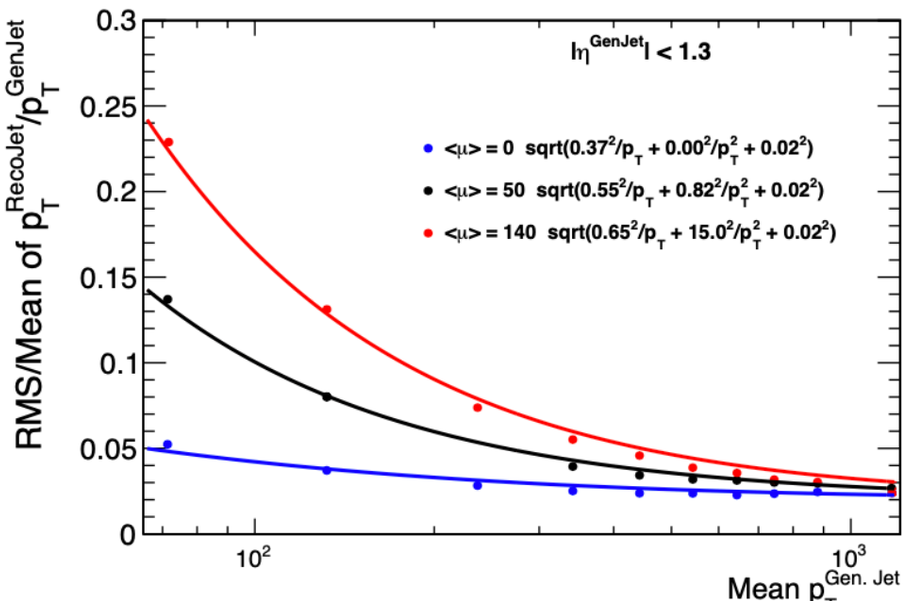
- Five-flavor MLM matching in the shower-kt scheme

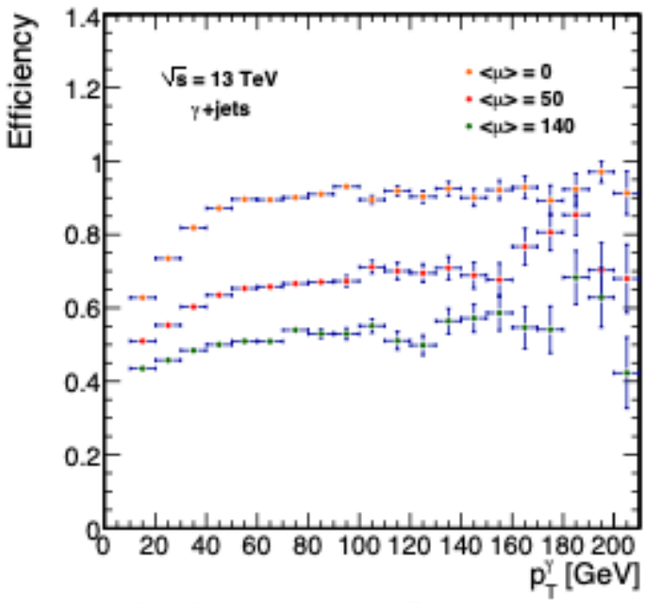
Dataset Names	QCUT
B-4p, BJ-4p, BJJ-vbf-4p, BB-4p, BBB-4p, LL-4p, LLB-4p, H-4p	40 GeV
TJ-4p, TB-4p	60 GeV
TT-4p, TTB-4p	80 GeV

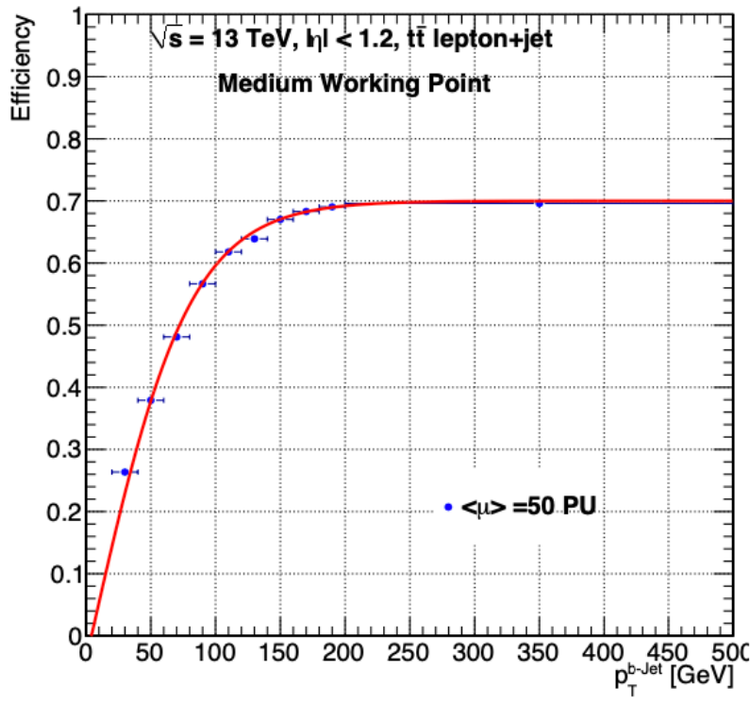
Detector Simulation/Reconstruction

- "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

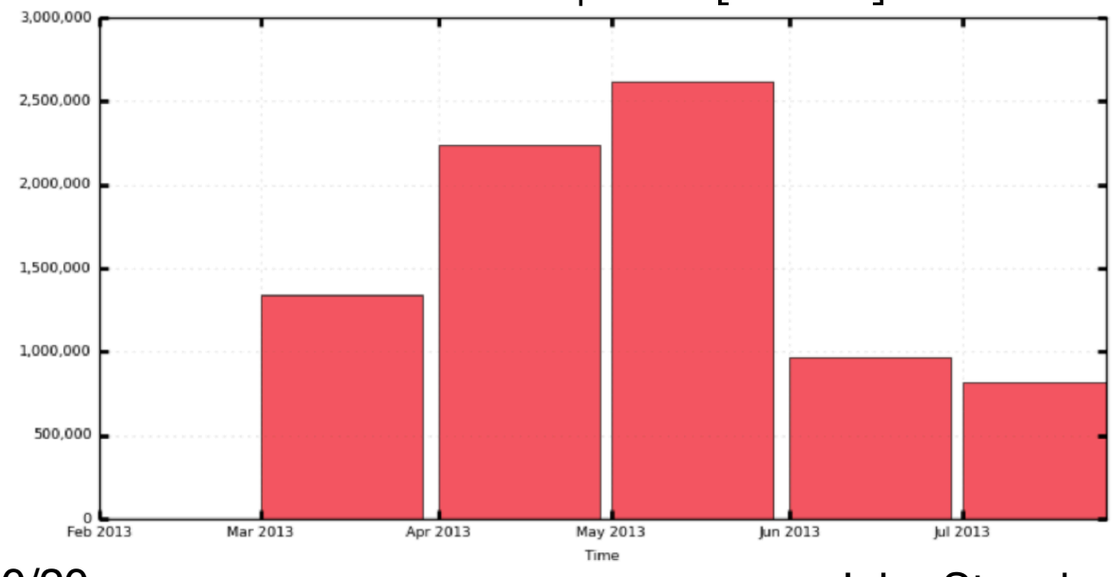
Process	$\sqrt{s} = 14$ TeV	$\sqrt{s} = 33$ TeV	$\sqrt{s} = 100$ TeV
$t\bar{t}$	1.24	1.10	0.96
$W^+ j$	1.17	0.85	0.74
$W^- j$	1.20	0.89	0.75
$Z^0 j$	1.17	0.87	0.76
γj	1.54	1.04	0.89
$W^+ W^-$	1.25	1.08	1.0
$W^+ Z^0$	1.24	1.06	0.95
$W^- Z^0$	1.26	1.09	0.97
$Z^0 Z^0$	1.37	1.29	1.21
$W^+ \gamma$	1.22	0.80	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
tW^-	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

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