Collider Simulations for Snowmass 2021

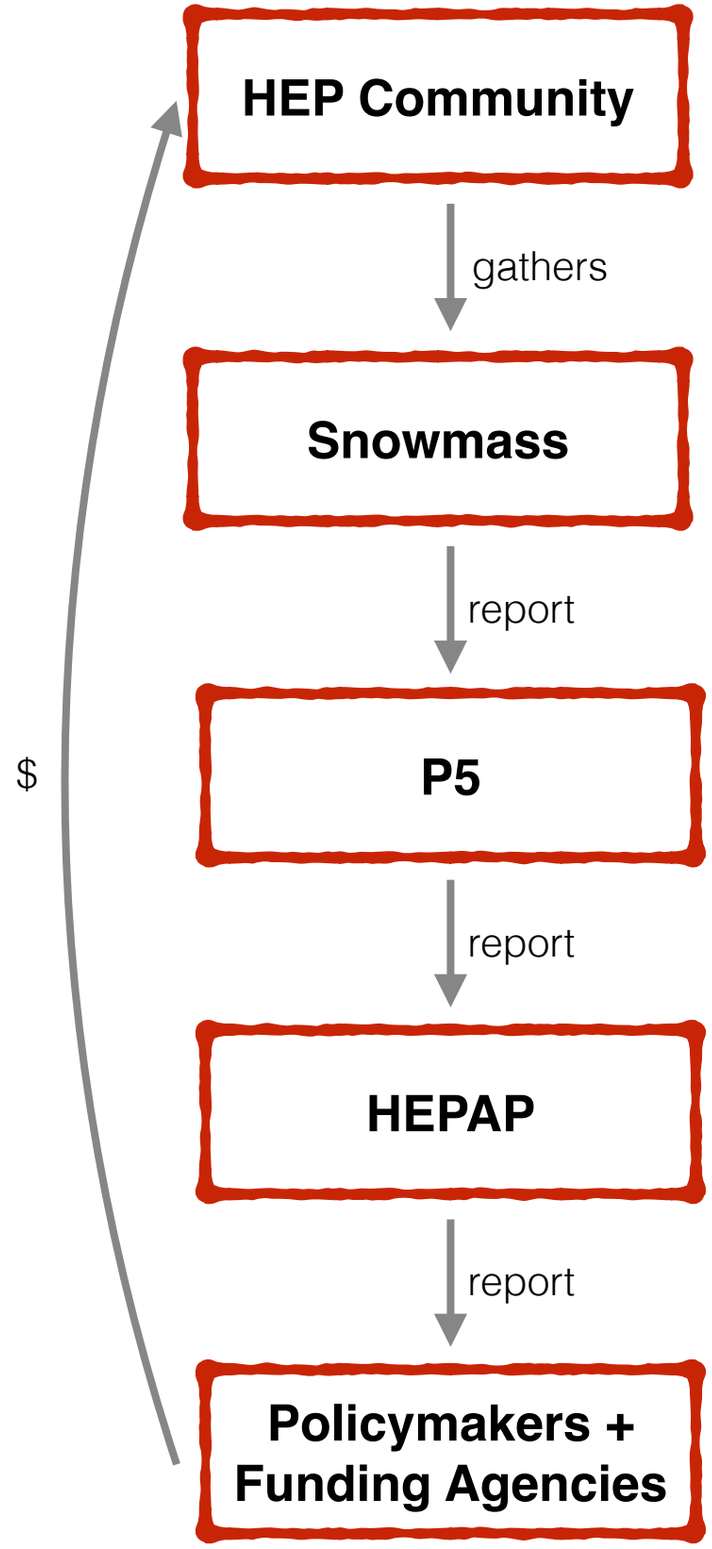
John Stupak
on behalf of the Energy Frontier

University of Oklahoma

3/4/21

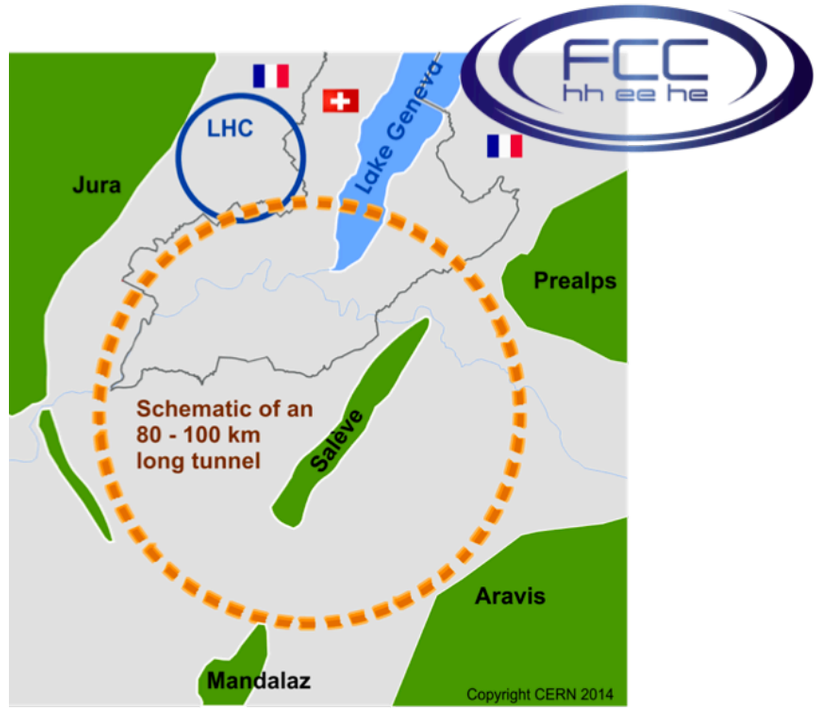
HEP Long-Term Planning

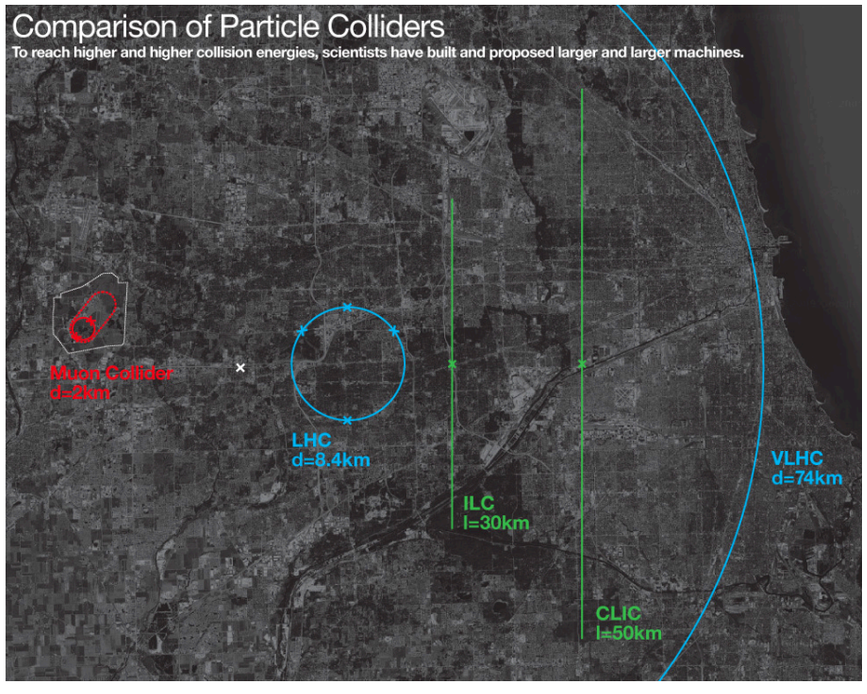
- Snowmass
  - High-energy physics community exercise
    - Every 8-10 years since 1982
    - Details and history
  - Energy frontier → large/expensive particle colliders
    - Identify and compare physics potential of proposed colliders
- Particle Physics Project Prioritization Panel (P5)
  - Digests Snowmass results
  - Prioritizes various projects over next 10-20 years
- High-Energy Physics Advisory Panel (HEPAP)
  - Digests P5 report
  - Provides recommendations to NSF and DOE



Energy Frontier

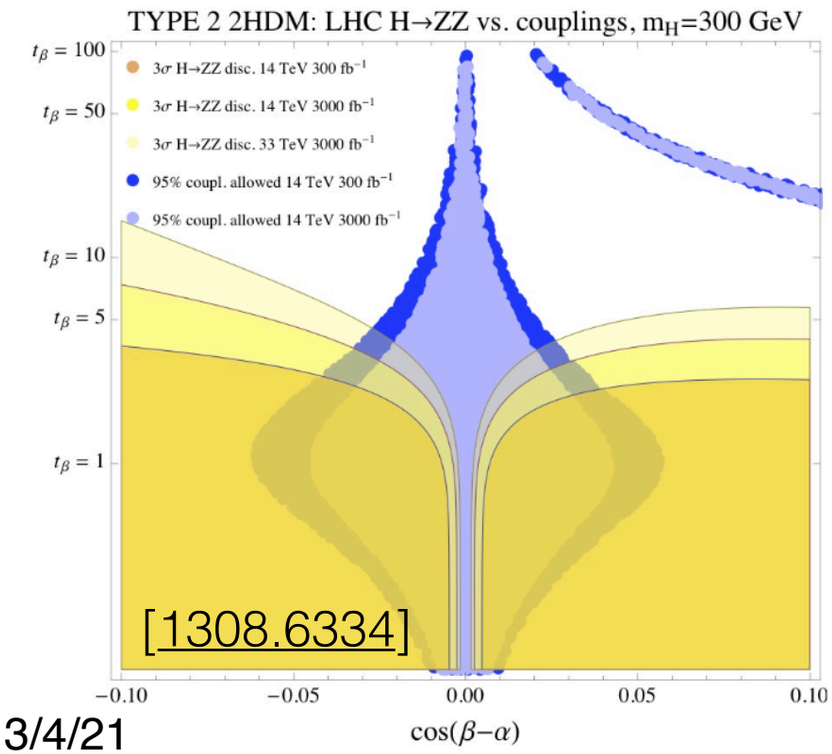
- Oversimplification: Searching for heavy particles and new forces
  - $E = mc^2$ → need high-energy collider (~$10B) to produce massive particles
- Following the LHC, what should we build next?
  - ee, ep, pp, μμ?
  - Collision energy?
  - Linear or circular?





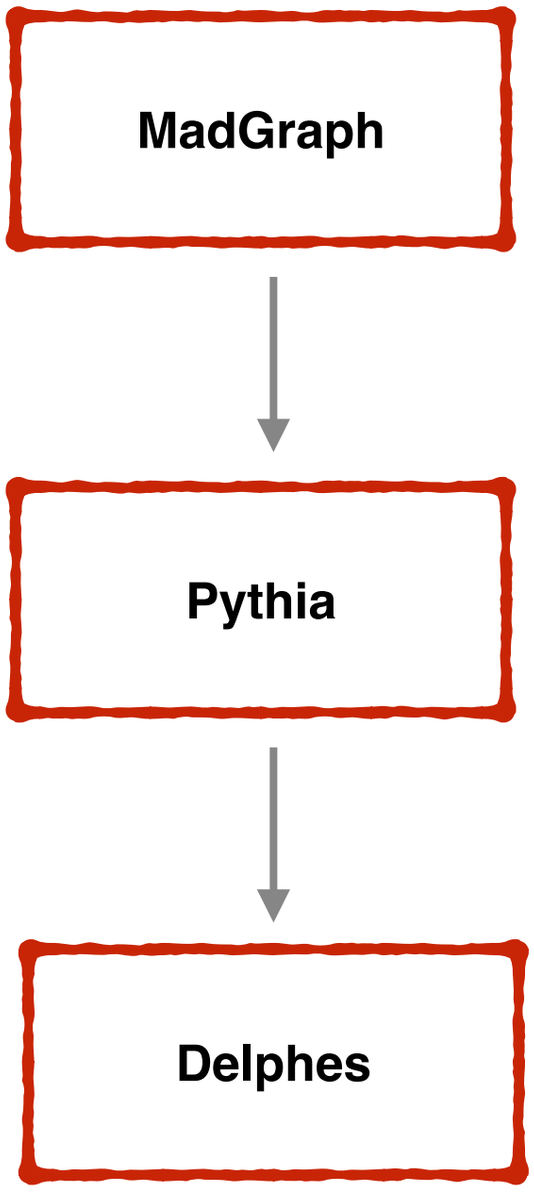
Simulation

- Problem: Need large samples of simulated collisions to assess sensitivity of proposed colliders for variety of new physics scenarios
  - Production is disk and CPU intensive
- Energy Frontier MC Task Force established to develop plan for production
  - Formulated plan and presented to community



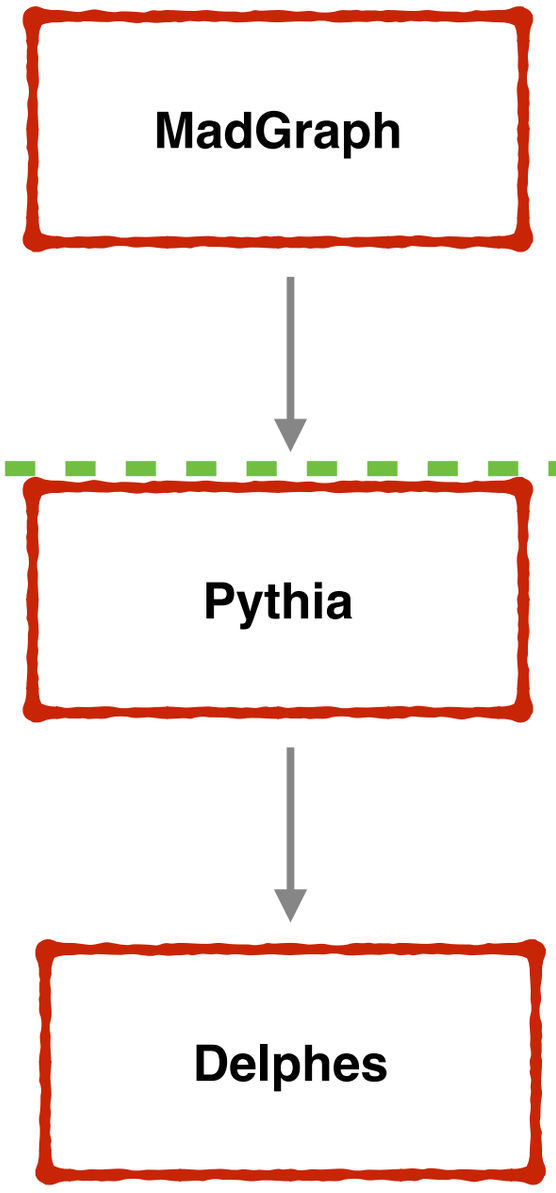
Name	Institution	
John Stupak (chair)	University of Oklahoma	Snowmass2013
Robert Gardner	University of Chicago	OSG rep.
Simone Pagan Griso	LBNL	BSM rep.
Stefan Hoeche	FNAL	MC expert
Fabio Maltoni	CP3, Catholic University of Louvain	MC expert
Meenakshi Narain	Brown University	EF convener
Isabel Ojalvo	Princeton University	EWK rep.
Laura Reina	Florida State University	EF convener
Michael Schmitt	Northwestern University	QCD rep.
Alessandro Tricoli	Brookhaven National Laboratory	EF convener

Workflow



- MadGraph
  - Simulates “hard collision” process
    - Long runtime
    - Negligible output size
- Pythia
  - Simulates fragmentation and hadronization of colored particles
    - Short runtime
    - Non-negligible output size
- Delphes
  - Simulates detector response and event reconstruction
    - Short runtime
    - Non-negligible output size

Workflow



- MadGraph
  - Simulates “hard collision” process
    - Long runtime
    - Negligible output size
- Pythia
  - Simulates fragmentation and hadronization of colored particles
    - Short runtime
    - Non-negligible output size
- Delphes
  - Simulates detector response and event reconstruction
    - Short runtime
    - Non-negligible output size

/software/gridpacks/

gridpack 1 | gridpack 2 | . . . | gridpack 10

gridpack = precompiled MadGraph tarball (~100MB) for each physics processes

ship via condor

worker node 1

MadGraph

long runtime
small output

worker node 2

Pythia

Delphes

short runtime
large output

/collab

/collab/gridpack1

LHE 1 | LHE 2 | LHE 3 | . . . | LHE 10k

ROOT 1 | ROOT 2 | ROOT 3 | . . . | ROOT 10k

/collab/gridpack2

LHE 1 | LHE 2 | LHE 3 | . . . | LHE 10k

ROOT 1 | ROOT 2 | ROOT 3 | . . . | ROOT 10k

. . .

/collab/gridpack10

LHE 1 | LHE 2 | LHE 3 | . . . | LHE 10k

ROOT 1 | ROOT 2 | ROOT 3 | . . . | ROOT 10k

Technical Details (pp)

- 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 sample
    - On-shell internal propagators excluded → fully orthogonal
    - On-shell heavy resonances treated as stable
    - Subsequently decayed democratically to leptons+hadrons using MadSpin
  - Store event weight: σ_{LO} * NLO k-factor * w_{BR}

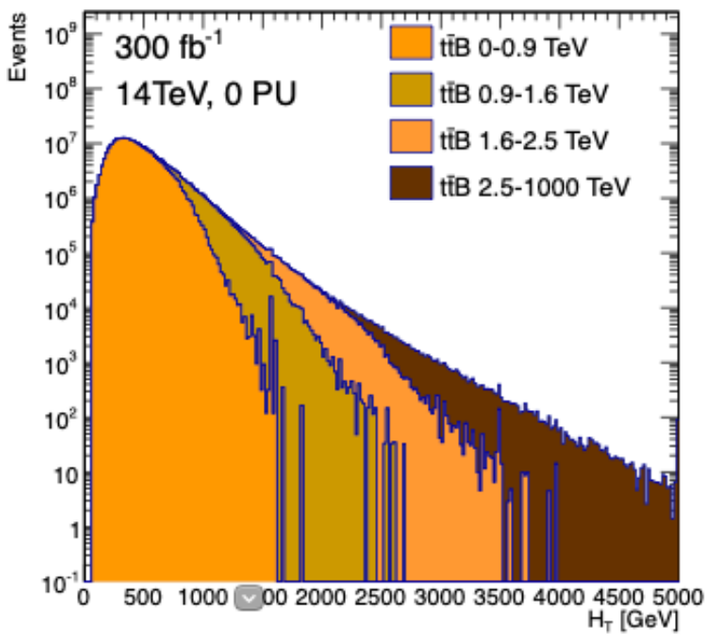
Dataset name	Physics process	Number of recoil jets
B-4p	γ or on-shell W, Z	0
Bj-4p	γ or on-shell W, Z	1-3
Bjj-vbf-4p	γ or off-shell W, Z, H in VBF topology	2-3
BB-4p	Diboson (γ, W, Z) processes	0-2
BBB-4p	Tri-boson (γ, W, Z) processes including BH	0-1
LL-4p	Non-resonant dileptons (including neutrinos) 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

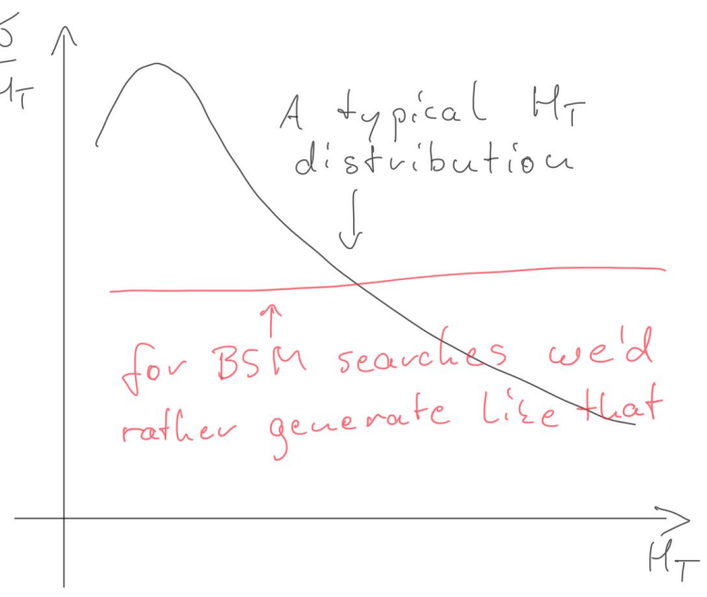
Table 1-2. *Table of background processes. All processes include the particles in the dataset name plus additional recoil jets up to four generated particles. On-shell vector bosons, off-shell dileptons, Higgs bosons, top quarks, and jets are denoted B, LL, H, t, and j, respectively.*

$t\bar{t} \rightarrow$	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

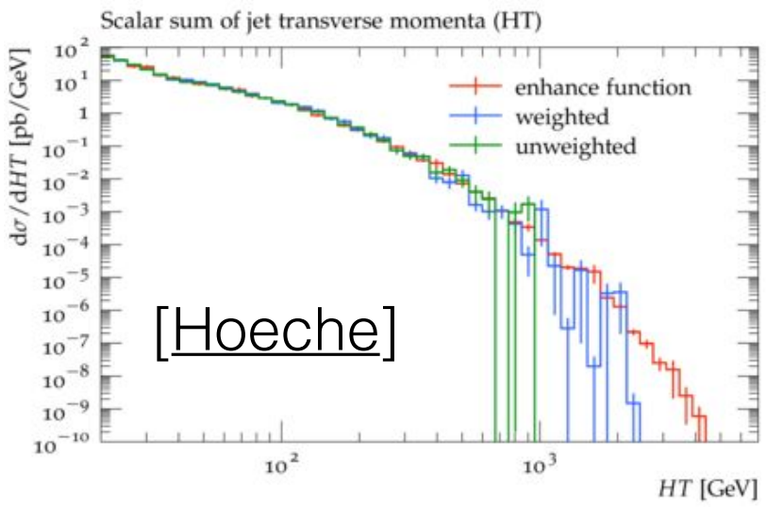
Technical Details (pp)

- One proposed change relative to 2013
  - Rather than binning samples in H_T, use weighted events or some functional bias (pending validation)
    - Correct to physical distribution with event weight

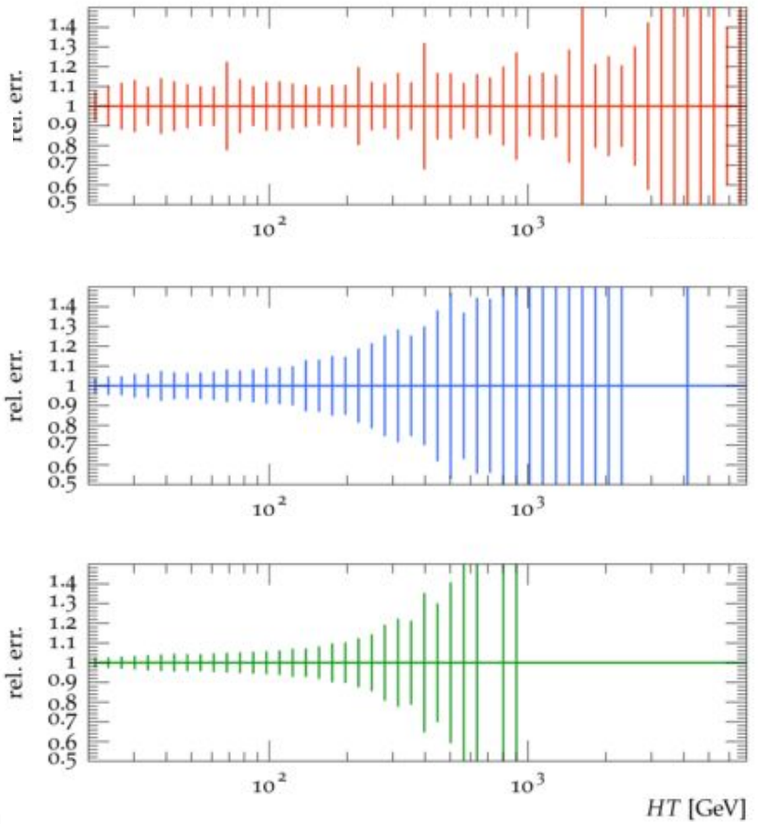




Example: Jet-HT



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



Computing (2021)

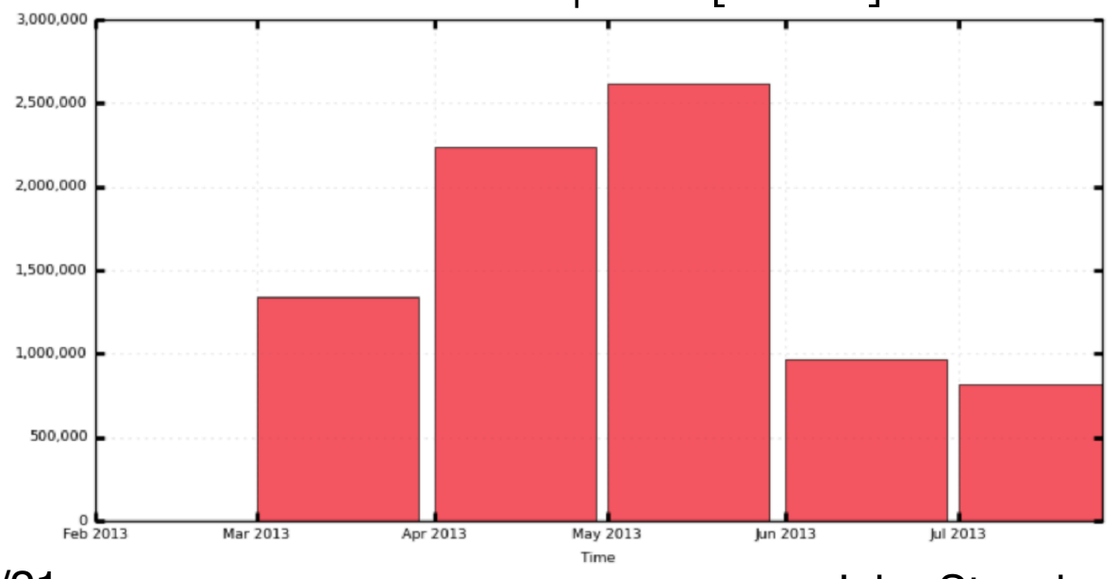
- OSG has kindly offered to support Snowmass effort and supply computing resources
  - Interactive login node + batch job submission + storage (“Snowmass21 Connect”)
- Documentation is available here
- Using Slack for discussion and user support

Computing (2013)

- 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

- 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)

CPU consumption [hours]



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

Summary

- Snowmass process absolutely vital to sustain HEP in the US
  - Requires significant computational resources/support
- Plan for production of simulated collisions on the OSG has been developed
  - Few Technical details are being ironed out and production will begin soon
- References (Snowmass2013):
  - Grid production: https://arxiv.org/abs/1308.0843
  - Simulation (1): https://arxiv.org/abs/1308.1636
  - Simulation (2): https://arxiv.org/abs/1309.1057

Many, many thanks to OSG for resources and support!

Backup