Monte Carlo Production
Energy Frontier Plans

Meenakshi Narain (Brown U)

Laura Reina (FSU)

Alessandro Tricoli (BNL)

May 4, 2020

Background: Snowmass 2013 MC Production (1)

- For Snowmass 2013 EF New Particles group, a large SM MC sample was produced [1] for pp collider studies:
  - A technical group within EF was set up.
  - A common detector simulation was used to generate samples at 14, 33 ,and 100 TeV pp colliders using Delphes.
    - common detector = average of ATLAS+CMS
  - Tools for MC generation developed by members of CMS/LPC in collaboration with theory colleagues from SLAC [2].
  - A central production using OSG[3] was set up to generate and validate the samples [4, 5].
- ILC, CLIC, TLEP, gamma-gamma collider, and Muon Collider studies used their own samples and frameworks.

[1] https://arxiv.org/abs/1309.1057; [2] https://arxiv.org/abs/1308.1636;
[3] https://sciencenode.org/feature/power-sharing-osg.php; [4]https://arxiv.org/abs/1308.0843;
[5] https://lpc.fnal.gov/lpcsnomass/index.shtml

Background: Snowmass 2013 MC Production (2)

- These samples were widely used [also after Snowmass 2013] by initial HL-LHC CMS EOI and other studies.
- Links to some old twikis with info on event generation and common setup, which give the scale of the MC task at Snowmass 2013 and the samples generated.
  - [just FYI and completeness, as the papers have the final information].
  - Instructions for generation of samples (from the common agreement - including common cards with detector parameters). [6]
  - The list of samples. [7]

[6] https://twiki.cern.ch/twiki/bin/view/CMSPublic/NPSnowmass2013SamplesInstructions

[7] https://twiki.cern.ch/twiki/bin/view/CMSPublic/NPSnowmass2013Samples

Task Force - Monte Carlo Production

- Since 2013, the landscape for simulations for future colliders have changed.
- Both future e+e- circular colliders (FCC-ee, CepC) and pp colliders (FCC-hh, CppC) have developed simulation and analysis frameworks, and generated MC samples for their studies.
- Thus, we have a two step plan:

1) **Assess the MC needs** for EF (exp+th) and formulate a plan. We have formed a "**Task force**" for this purpose.
2) **Produce the "needed" MC samples** by the community to carry of the necessary studies for EF.

"needed" ⇒ some reasonable level, as well as prioritized.

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.

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”.

Membership of EF MC Task Force

- Membership includes representatives from EW, QCD and BSM TGs as well as other people with past Snowmass experience & technical expertise.
- Composition of the EF MC task force:
  - member of the Snowmass 2013 team - **John Stupak (Chair)**
  - Energy Frontier (EF) Topical Group conveners:
    - EWK: Isobel Ojalvo
    - QCD: Michael Schmitt
    - BSM: Simone Pagan Griso
  - Theorists with MC expertise: Fabio Maltoni (convener of TF07: Collider Phenomenology) & Stefan Hoeche (convener of EF05: Precision QCD)
  - member of OSG for technical input & resource support: Robert Gardner (or designee)
  - EF conveners

Interactions with Community

- **Input to Task Force from members of the community will be facilitated by each of the EF Topical Groups.**
- **EF Topical group conveners will interact with the community**
  - to gather the needed MC requests - large samples for background and systematic studies
    - (Please note that for Snowmass 2013 signal samples were generated by the proponents of the studies, though this is an open question for the TF!)
  - to ensure that there is a (/are) group(s) which will use the samples!
  - The theory co-convenors in each TG own the responsibility to provide guidance to their TG community on theory predictions and how to assess theory uncertainties.
- **EF Topical Group conveners will interact with the MC TF**
  - to provide guidance on (or negotiate?) priorities (if needed to constrain the insatiable needs of the community):
    - they have the knowledge and expertise to prioritize and advise if the requests are reasonable.

Snowmass 2021 EF Monte Carlo Production Team

- We envision that the **MC production team** will take charge of executing the plan developed by the “MC task force”.
- This group will be formed around July and we will recruit volunteers with technical expertise from the community.
- The main activities of the team are:
  - Set up the technical infrastructure for the central MC production. This will also involve working with the OSG team for compute and storage resources.
  - Recruit members from the community to help with the production.
  - Produce the samples and make them available to the EF community.
  - Provide documentation for use by the EF community for generation of signal and other smaller samples which will not be centrally produced.
  - The chair of this team will help the EF conveners in any future requests for central production, if any, from the community.
  - Document the efforts of this team as “snowmass contributed papers”.

Generalize to other Frontiers?

- We welcome partnership with other Frontiers who may have a similar need.
- The partnership would be focussed on setting up and monitoring production of the MC with the help of the OSG.
- **Discussion today!**

Backup

Some snippets from e-mail exchange with Frank Wuerthwein (OSG Director):

- You should assume that OSG will support this.
- We can do so with both compute and storage.

- Preference for storage would be to export it into the OSG Data Federation.
- Slide 3 is the current use for a recent 6 month period.
- Slide 4 is the deployed caches globally.

- Fermilab already supports the main endpoint that is connected into this.
- We should just use that for snowmass as well.

- We can then do something similar to the covid19 work we do right now, and CMS can provide extra access for Snowmass related work. Just like they did for covid19 work last weekend. We peaked at 40,000 cores. ATLAS can do the same, and voila! we have large amount of resource access from the submit hosts OSG supports.
- We have submit hosts at Chicago, and will have some in Wisconsin by this fall. People ask for accounts and off they go. We could provide a submit option also from Fermilab, or from cms-connect or atlas-connect.

- Bottom line, this is all pretty straightforward to support.

Current Use in OSG

Collaboration	Working Set	Data Read	Reread Multiplier
DUNE	14.2GB	1.8PB	30k
LIGO (private)	18.2TB	596TB	30
LIGO (public)	7.2TB	96TB	14
MINERVA	435GB	305TB	700
DES	272GB	127TB	500
NOVA	4.8TB	8.5TB	2

Depending on the expected size of working sets, a given science use case may need larger caches than we have deployed.

LIGO has driven international deployment, and DUNE is the major beneficiary of it.

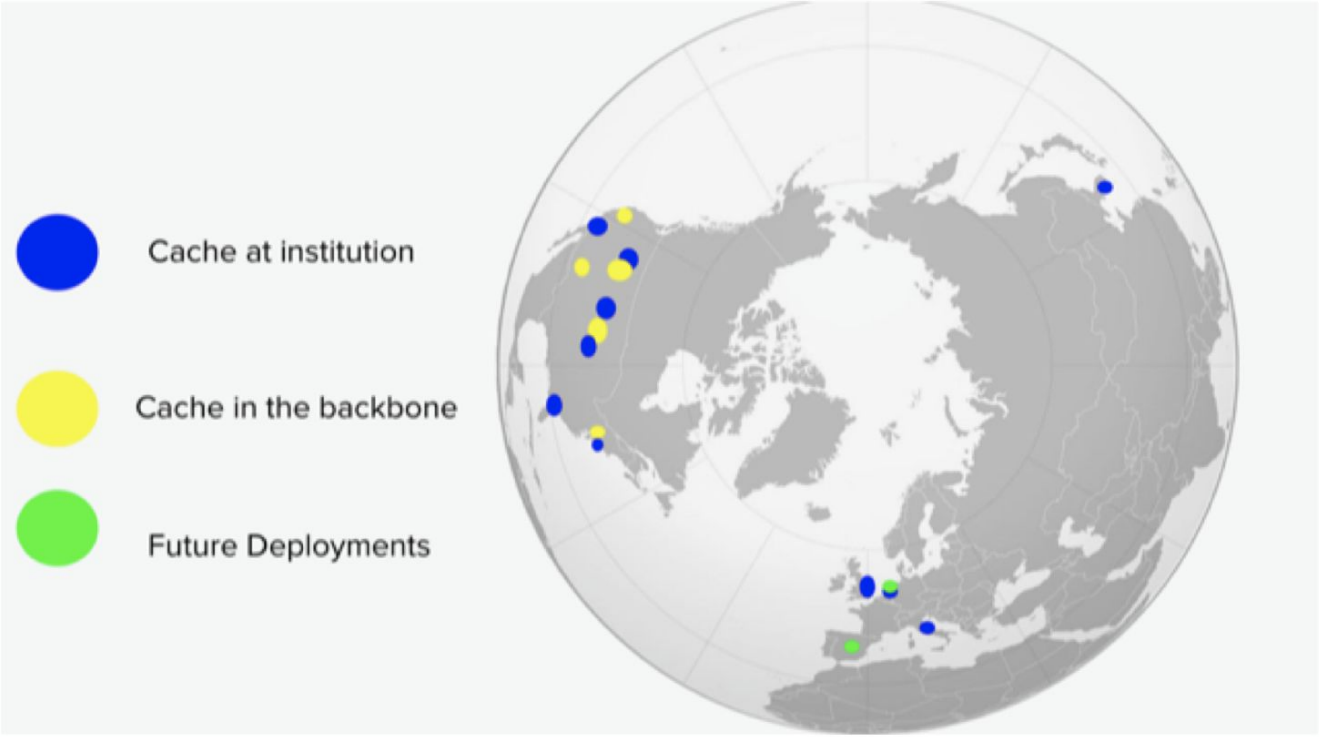
Typical cache sizes are large enough to fit the sum of all working sets at this point in time.

Current Deployment in OSG

Open Science Grid

UCSD

More than a dozen caches deployed across 3 continents



Including several at POPs of Internet2 and GPN.