

GSoC 2024

Contributor Welcome Session



May 13th, 2022
Valentin Volkl, Benedikt Hegner, for the
CERN/HSF GSoC admins
CERN

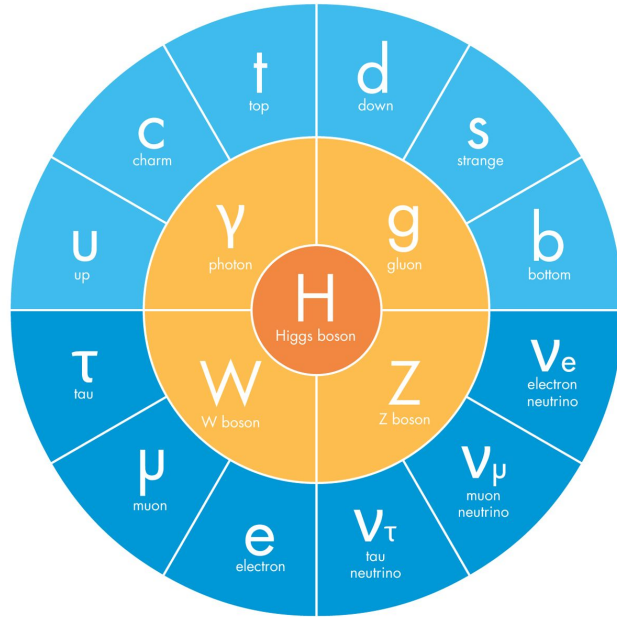
In this session:

- Short introduction by admins
- Reminder of GSoC program rules
- Q&A
- Round table of introductions (if time allows, we have a hard cutoff after 1 hour)

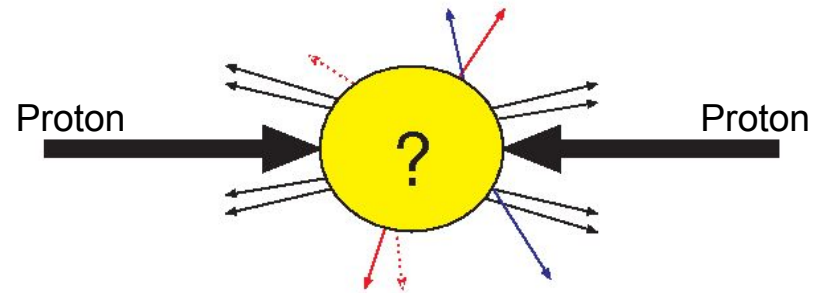
Introduction - Experimental Particle Physics

THE STANDARD MODEL

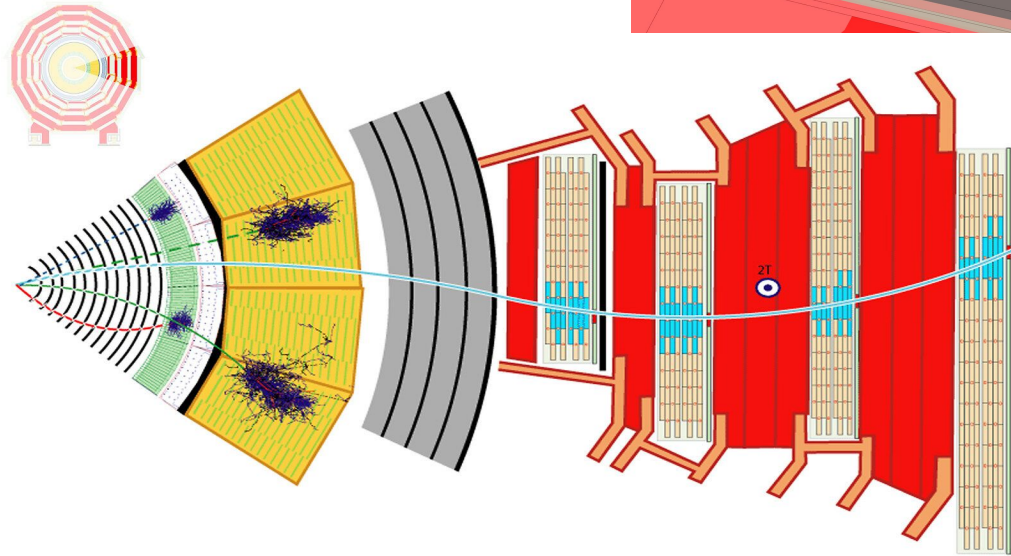
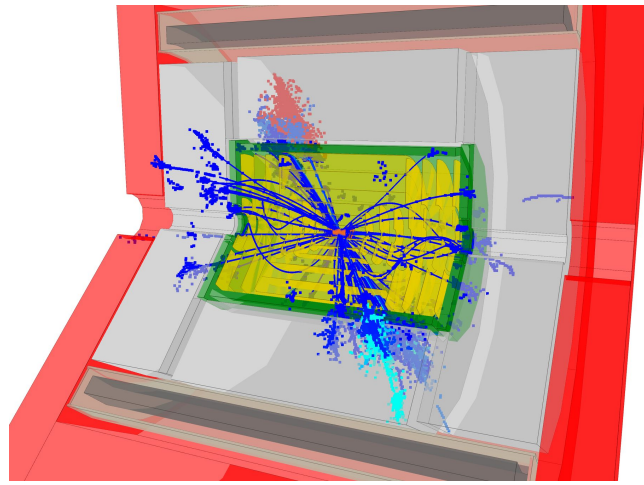
FERMIONS (matter) | BOSONS (force carriers)
● Quarks ● Leptons ● Gauge bosons ● Higgs boson



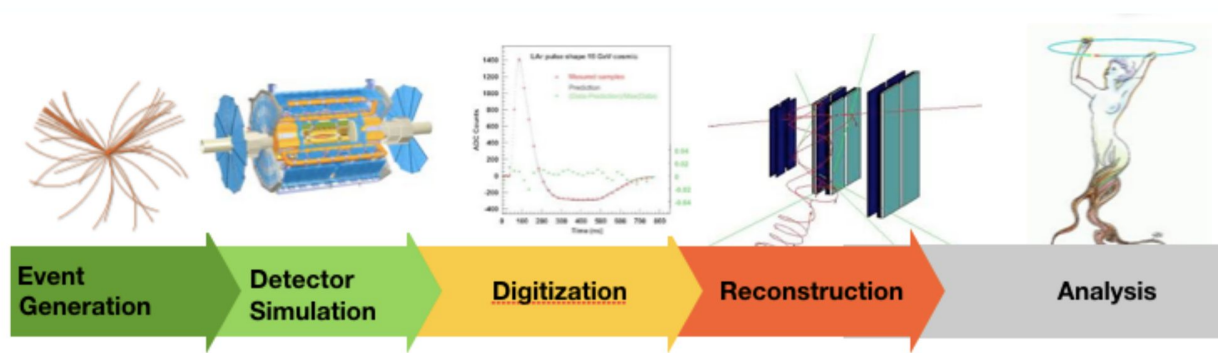
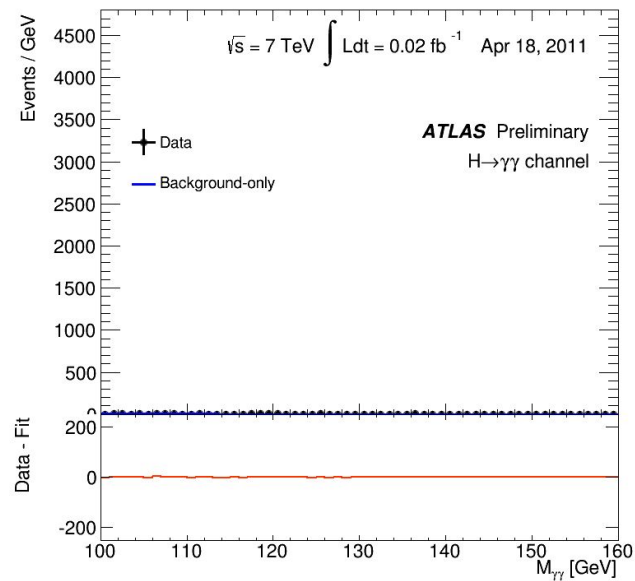
- Study fundamental particles and their interactions ...
 - Most only accessible at high energies
 - And not directly observable
- ... by smashing protons together and measuring what comes out



Detecting particles

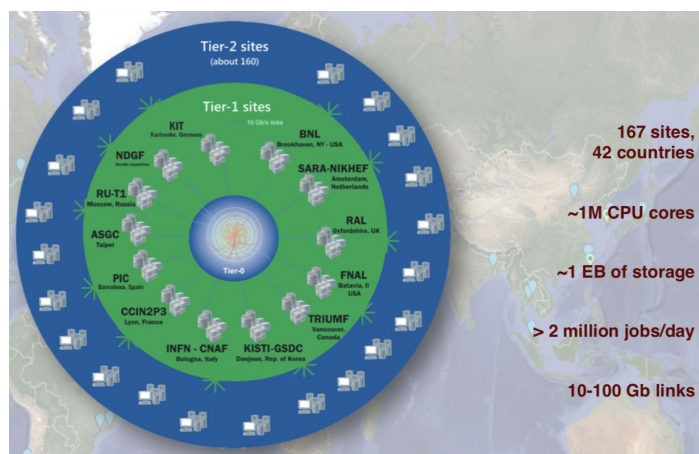


Computing Workloads in HEP



HEP Software and Computing

- High Energy Physics has a vast investment in software
 - Estimated to be around 50M lines of C++
 - Almost all open source!
- It is a critical part of our physics production pipeline, from triggering all the way to analysis and final plots as well as simulation
- LHC experiments use about 1M CPU cores every hour of every day, we have around 1000PB of data with 1000PB of data transfers per year (10-100Gb links)
 - We are in the exabyte era already
- This is a *huge* and *ongoing* cost in hardware and human effort
- With significant challenges ahead of us to support our ongoing physics programme



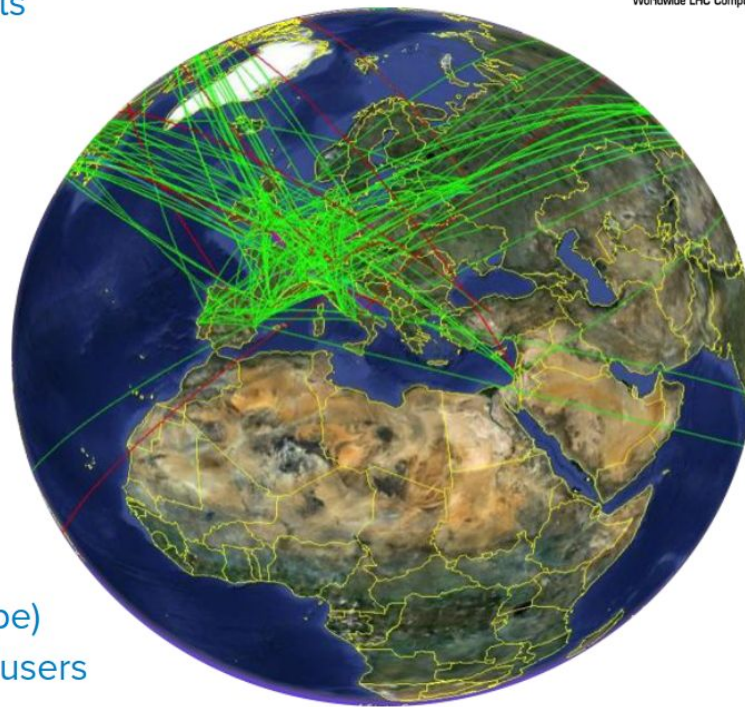
athena ○

ATLAS Experiment main repository for Athena ○

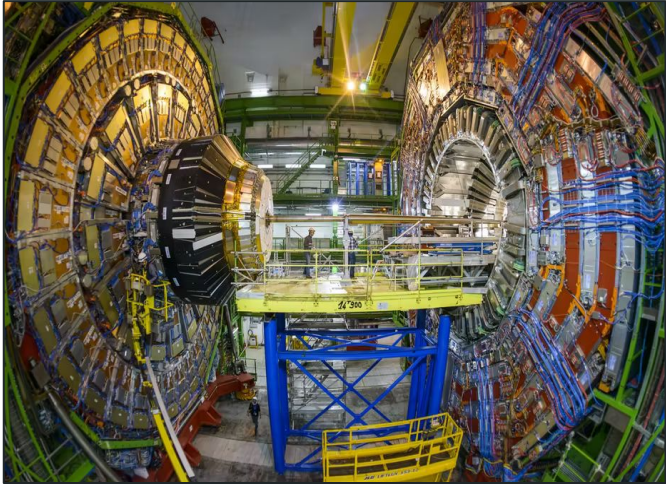


The Worldwide LHC Computing Grid - WLCG

- International collaboration to distribute and analyse LHC data
- Integrates computing centres worldwide that provide **computing** and **storage** resource into a single infrastructure accessible by all LHC physicists
- Global **network** connectivity
 - **Tier-0 (at CERN):** Large resource, recording and custodial archival of collision data, prompt reconstruction
 - **Tier-1s:** Memory and CPU intensive tasks second tape copy of detector data, occasional Tier-0 overspilling
 - **Tier-2s:** Processing centres, nowadays many are similar to the Tier-1s
- Around 1.5 million CPU cores fully occupied 24/7
- Around 1.5 EB data (~600 PB on disk and >800 PB on tape)
- More than 100 PB moved every month, accessed by 10k users



HEP Software and Computing



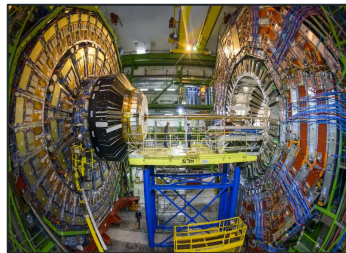
```
Model &desc = fSource->ConnectModel(desc, mSourceModel);  
Model.GetFieldZero()->SetOnDiskId(desc.GetFieldZeroId());  
// If the model has been created from the descriptor, the on-disk ID is already set.  
// User-provided models instead need to find their corresponding ID in the descriptor  
if (field.GetOnDiskId() == kInvalidDescriptorId) {  
    field.SetOnDiskId(desc.FindFieldId(field.GetName(), field.GetParent()));  
}  
field.ConnectPageSource(*fSource);
```



Questions?

We won't be able to answer all here ...

- Learn more from the many openly available resources!
 - Summer Student lectures
 - <https://indico.cern.ch/category/17540/>
 - Openlab student lectures
 - <https://indico.cern.ch/category/16988/>
- But most importantly, take advantage of this programme to become an expert on your project!



```
Model.GetFieldZero() = fSource->GetDescriptor();
// If the model has been created from the descriptor, the model has been created
// User-provided models instead need to find their corresponding field in the model
if (field.GetOnDiskId() == kInvalidDescriptorId) {
    field.SetOnDiskId(desc.FieldId(field.GetName(), field.GetName()));
}
field.ConnectPageSource(*fSource);
```



CERN/HSF GSoC program rules

- We don't have any special rules*, but as a reminder, the GSoC rules apply
 - <https://summerofcode.withgoogle.com/rules>
 - <https://google.github.io/gsocguides/>
- You'll need to fill in a midterm and final evaluation (students and mentors)
- By default, this is due July 12th (<https://developers.google.com/open-source/gsoc/timeline>)
 - But it can be extended up to mid-August (in agreement with mentor/student!). This should not affect total workload, but give more flexibility.
 - PLEASE NOTE: Let us know early if you want to extend these deadlines, they cannot be changed once the evaluation period starts!
- * One exception: We do require a blogpost from you summarizing the project for the HSF website. This is obligatory, and we won't pass you if we don't get one before the final examination.
 - No formal requirements in terms of length, style...
 - Template <https://github.com/HSF/hsf.github.io/pull/1449/files> (s/2023/2024/)

CERN/HSF GSoC program rules 2

We'll organize one or two meetings like this towards the end of the program where you have the opportunity to give a lightning talk on your project.

Google also organizes lightning talks for GSoC contributors.

If you want to connect with your peers, you can join

<https://groups.google.com/g/hsf-gsoc-2024-students> or organize yourselves

The admins are here to help - let us know early if there is an issue or just give feedback! hsf-gsoc-admin@googlegroups.com

Q&A