Proposal for a MC samples database for FCC studies

Benjamin Fuks, Clement Helsens, Carlos Solans

Why a database for MC samples?

- The FCC effort counts with very little resources for the time being
  - Synergy between different communities (experimentalist and theorists) interested in FCC studies is mandatory in order to maximize efficiency
  - A very important aspect of the FCC study to share samples
- We want to start producing samples using MadGraph for FCC-hh
  - Produce HepMC (after showering/hadronization) and/or LesHouches files (before showering), store the samples and share the information
  - Would FCC-ee be interested in such a tool? What is the generator being used for FCC-ee?

Proposal

- We propose to build a sample database in order to have a centralized archive of generated samples and provide a platform for generation of new samples
  - The possibility to trace similar samples for one's analysis is a key to success
  - Control over the generation of samples avoids production of statistically irrelevant samples
- Start with the possibility to select the background process type and random seed
  - Control of the chosen random seed is mandatory to produce uncorrelated samples

Requirements

- MC sample database should
  - Provide full description of samples
  - Be flexible in the addition and removal of indexes (p_T cut on jets, leptons, etc…)
  - Rely on a lightweight framework infrastructure
  - Simple to query and fast to reply
  - Have a web interface for public availability

- This MC database could a first step towards a more general tool to registering the different datasets (generation, simulation…) with different versions (alike the ATLAS AMI)

Considerations for the implementation

- Implementation proposal considers
  - Central web services (AFS, virtual host)
  - Central storage space (?)
  - Document oriented database (MongoDB, Redis, Oracle)
  - Simple front-end interface (PHP, Python, javascript)

- If the CPU is not a problem
  - Drop the storage of HepMC events or LesHouches files, and re-produce them on the fly from the input cards upon request
- If the HDD is not a problem
  - Store HepMC events along with the samples

Simple example: pp→ttbar

- Cross section
  - At 13 TeV : 465 pb
  - At 100 TeV: 2.66 10^4 pb
- Multiplicity goes up with √s and so does the size.
  - Sample size for 10k events

	HepMC Size /10k ev (MB)	**LesHouches Size /10k ev (MB)**
13 TeV	214 (482)	5.3 (22)
100 TeV	264 (598)	5.5 (22)

- Aiming for 3ab^{-1}
  - We need 10^{11} events to have the same number of data events
  - Which is 10^7 more events → leading to sample size ~TB