LPCC Fast Sim Workshop

Anna Zaborowska

CERN EP-SFT

October 7, 2021

LPCC Fast Simulation Workshop

- 22-23 November
- 15:00-18:00 CET
- an extra hour (till 19:00) in case of need for further discussions
- zoom-only meeting

LPCC Fast Simulation Workshop

- 22-23 November
- 15:00-18:00 CET
- an extra hour (till 19:00) in case of need for further discussions
- zoom-only meeting
- proposal to have several mini-presentations on the specific topics to facilitate the discussions (instead of usual per-experiment status presentations)
- speakers within one topic would be asked to cover same questions (as much as possible)

Topics

1. Fast Sim applicability and accuracy
2. Validation
3. Machine Learning: Data preprocessing
4. Machine Learning: Tuning
5. Machine Learning: Inference Integration
6. Reusability and Generalisation

Topics

1. Fast Sim applicability and accuracy
   - How is fast simulation used: applied to which detectors?
   - Are there specific “sensitive” parts of the detector where it is difficult to get fast simulation?
   - Is it used widely in production?
   - Which physics analyses can use fast simulation, which cannot?
   - What are the demands on the accuracy of the (fast) simulation?
   - Do you/can you tune fast simulation to collision data to get better results than from the full simulation?
   - How does it look now, but also how it changes for Run3 & HL-LHC?
2. Validation
3. Machine Learning: Data preprocessing
4. Machine Learning: Tuning
5. Machine Learning: Inference Integration
6. Reusability and Generalisation

Topics

1. Fast Sim applicability and accuracy
2. Validation
   - How do you validate fast simulation, whether it's "classical" or ML approach.
   - Is it on e.g. particle shower characteristics (longitudinal/transverse profiles, etc.), on high-level results of physics analyses, ...?
   - For ML, what is the metric used to validate the model and what is the objective function you are optimizing during the training?
3. Machine Learning: Data preprocessing
4. Machine Learning: Tuning
5. Machine Learning: Inference Integration
6. Reusability and Generalisation

Topics

1. Fast Sim applicability and accuracy
2. Validation
3. Machine Learning: Data preprocessing
   - Which particle types and features (energy, angle) are considered?
   - How large is the training dataset?
   - How to define the structure of the input data (hits, cells, clusters, custom voxels,..)?
   - Which data structure is used for the training (1D vector, images, graphs..)?
   - Which scaling is used? Is the dataset balanced (is the number of events for the different particle properties almost the same?).
   - How do you store the preprocessed data? How are the condition values (energy of the particle, angle,..) encoded?
4. Machine Learning: Tuning
5. Machine Learning: Inference Integration
6. Reusability and Generalisation

Topics

1. Fast Sim applicability and accuracy
2. Validation
3. Machine Learning: Data preprocessing
4. Machine Learning: Tuning
   - What tuning approach is adopted?
   - What metric is used to compare between the performance of the model during the tuning process?
5. Machine Learning: Inference Integration
6. Reusability and Generalisation

Topics

1. Fast Sim applicability and accuracy
2. Validation
3. Machine Learning: Data preprocessing
4. Machine Learning: Tuning
5. Machine Learning: Inference Integration
   - How is the ML trained model stored (file format)?
   - Are there any specific files to store in addition to the ML model?
   - Which inference library is used? How much time does it take to generate one event?
   - What is the memory footprint?
   - Are there any optimization approaches considered to reduce the memory?
   - Has the ML simulation been tested on GPU?
6. Reusability and Generalisation

Topics

1. Fast Sim applicability and accuracy
2. Validation
3. Machine Learning: Data preprocessing
4. Machine Learning: Tuning
5. Machine Learning: Inference Integration
6. Reusability and Generalisation
   - Are there any valuable lessons to be learnt by others?
   - Tools which can be reused? Maybe some ideas have failed for your detectors (or parts of it) and it can be useful to others?
   - What are your plans, maybe there is room for collaboration?

Topics

1. Fast Sim applicability and accuracy
2. Validation
3. Machine Learning: Data preprocessing
4. Machine Learning: Tuning
5. Machine Learning: Inference Integration
6. Reusability and Generalisation

6 hours (minus breaks) per 6 topics = on avg. 1h per topic, so ~4 10 min presentations