

High Performance Computers as a solution for Machine Learning training and hyperparameter optimisation

Jean-Roch Vlimant ^a, Torre Wenaus ^b, Rui Zhang ^{c,1}

- a. California Institute of Technology
- b. Brookhaven National Laboratory
- c. University of Wisconsin–Madison

The High-Luminosity LHC (HL-LHC) is planned to reach an instantaneous luminosity of $5 \times 10^{34} \text{cm}^{-2} \text{s}^{-1}$. This will lead to the integrated luminosity of 3000fb^{-1} at the end of the HL-LHC era, 20 times of what we have now as standing at the end of Run 2. One of the potential solutions is to leverage huge resources provided by High Performance Computers (HPCs). This can be done in many aspects. We would like to focus on Machine Learning (ML) and hyperparameter optimisation (HPO) in our contribution.

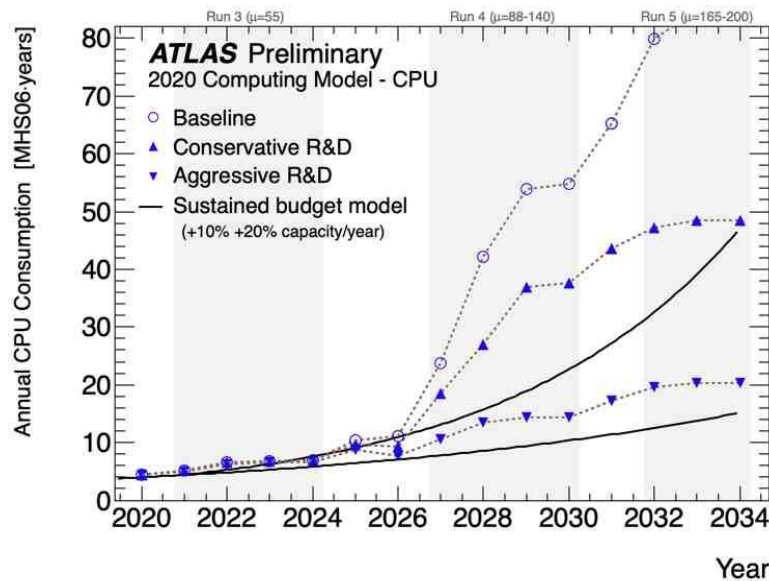


Fig 1. Estimation and projection of CPU resources necessary for the ATLAS experiment from 2018 to 2034 [1].

The HPCs are notoriously more difficult to use than the distributed grid computing for High Energy Physics, where the latter computing model has been deployed by LHC since then. Despite the hurdles to access an HPC as individual users, HPCs have quite different batch queue policies and network connections. The most notable ones that will have big impacts on our applications are:

- HPCs encourage jobs to use a large number of nodes at the same time by granting them longer walltime. For instance, on Summit at ORNL, the walltime is 2h if a job

¹ rui.zhang@cern.ch

requests less than 45 nodes, where one node has two 16-core-processors and six GPUs.

- Due to security considerations, worker nodes on most HPCs have no Internet connection to the outside world. Communications with outside need to be done on front-end nodes before the job starts.

Therefore, it is crucial and challenging to seek a solution for the marriage of HPC and ML/HPO. On the other hand, the nature of embarrassingly parallelisation of ML/HPO make itself an outstanding candidate to utilise the “world leading facilities”. It may as well shed light on how to off-load our other applications to HPCs.

References:

[1] <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ComputingandSoftwarePublicResults>