

# Large-Scale Distributed Training of Deep Neural Net Models

*Thursday 12 July 2018 11:00 (15 minutes)*

In the recent years, several studies have demonstrated the benefit of using deep learning to solve typical tasks related to high energy physics data taking and analysis. Building on these proofs of principle, many HEP experiments are now working on integrating Deep Learning into their workflows. The computation need for inference of a model once trained is rather modest and does not usually need specific treatment. On the other hand, the training of neural net models requires a lot of data, especially for deep models with numerous parameters. The amount of data scales with the many parameters of the models which can be in billions or more. The more categories present in classification or the more wide a range of regression is performed the more data is required. Training of such models has been made tractable with the improvement of optimization methods and the advent of GP-GPU well adapted to tackle the highly-parallelizable task of training neural nets. Despite these advancement, training of large models over large dataset can take days to weeks. To take the best out of this new technology, it would be important to scale up the available network-training resources and, consequently, to provide tools for optimal large-scale trainings. Neural nets are typically trained using various stochastic methods based on gradient descent. One of the avenue to further accelerate the training is via data parallelism, where the computation of the gradients is computed on multiple subset of the data in parallel and used collectively to update the model toward the optimum parameters. Several frameworks exists for performing such distributed training, including framework already developed by the authors, all with their strengths and limitations. In this context, our development of a new training workflow, which scales on multi-node/multi-GPU architectures with an eye to deployment on high performance computing machines is described. Old and new frameworks at put on a benchmark test on a few HEP-specific examples and results are presented.

**Authors:** PIERINI, Maurizio (CERN); PANTALEO, Felice (CERN); VLIMANT, Jean-Roch (California Institute of Technology (US))

**Presenter:** PANTALEO, Felice (CERN)

**Session Classification:** T6 - Machine learning and physics analysis

**Track Classification:** Track 6 –Machine learning and physics analysis