Optimizing Distributed Deep Learning Training by Tuning NCCL

Majid Salimi Beni[a], Ruben Laso[b], Biagio Cosenza[c], Siegfried Benkner[b], and Sascha Hunold[a]

[a] *Faculty of Informatics, TU Wien, Austria*
[b] *Faculty of Computer Science, University of Vienna, Austria*
[c] *Department of Computer Science, University of Salerno, Italy*

Distributed Deep earning is essential for training large-scale neural networks when the entire data set or model cannot fit into a single machine. The communication layer of such a deep learning framework is responsible for synchronizing model updates and exchanging gradients between nodes, and the communication operations in that layer must be efficient. The NVIDIA Collective Communications Library (NCCL) is a widely used back-end for communication in GPU-accelerated clusters. Similar to the Message Passing Interface (MPI), NCCL's efficiency depends on its parameter configuration [1, 2], including the choice of communication algorithms, buffer sizes, and network types.

NCCL Parameter Tuning: We propose a two-step *offline tuner* to optimize the NCCL parameter configuration for multi-GPU clusters. First, we *profile* the training of the models to determine the most relevant message sizes. Second, we employ a Bayesian optimizer to find an *efficient parameter configuration.*

Experimental Results: Figure 1 compares the performance of two deep learning models (`Bert` and `NasNetMobile`) on 2 nodes of the Leonardo supercomputer, using TensorFlow and Horovod. On top, we compare the bandwidth obtained after tuning the collectives for the most frequently used message size of each model. The tuned configurations improved the bandwidths of the respective NCCL operations in the microbenchmarks by 2.26 and 21.02 times. On the bottom, we show an improvement in training performance of 12% and 13% for `Bert` and `NasNetMobile`, respectively, when using the tuned configuration. Our experiments highlight the significant performance gains achievable through optimizing NCCL in distributed deep learning training.

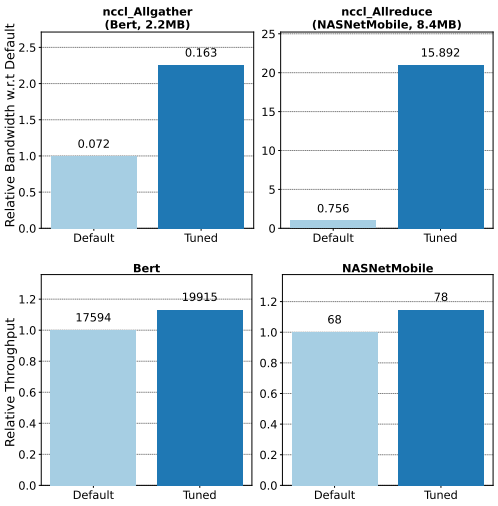

Figure 1: Default vs. Tuned NCCL collectives on 2×4 NVIDIA A100 GPUs. The raw values are shown on top of the bars: Bandwidth in GB/s, and Throughput in samples/s. Higher is better.

References

[1] De Sensi, D., *et al.* "Exploring GPU-to-GPU Communication: Insights into Supercomputer Interconnects," SC (2024).

[2] Salimi Beni, M., Cosenza, B., and Hunold, S., "MPI Collective Algorithm Selection in the Presence of Process Arrival Patterns," CLUSTER (2024).