Co-designing MPI Runtimes and Deep Learning Frameworks for Scalable Distributed Training on GPU Clusters

Ammar Ahmad Awan and Dhabaleswar K. Panda (Advisor)
awan.10@osu.edu, panda@cse.ohio-state.edu

**MOTIVATION**
- Resurgence of Deep Learning (DL)
  - Availability of Large Datasets like ImageNet and massively-parallel modern hardware like NVIDIA GPUs
  - Emergence of DL frameworks (Caffe, TensorFlow, CNTK, etc.)
  - Competability of Deep Neural Networks (DNNs)
  - Single GPU/node is not enough!
  - Scale-up and Scale-out training: an emerging research area

**RESEARCH CHALLENGES**
- Various Parallelization Strategies for DNNs
  - Model Parallelism / Data Parallelism
  - Alternative Implementation Styles
- Parameter Server approach / Reduction-Tree approach
- Distributed Address-Space Design Constraints
- Parallel Data Reading Mechanisms
- Challenges for Communication Runtimes
  - Very Large GPU-based Buffers
  - Overlap of Computation and Communication

**PROPOSED SOLUTIONS AND PERFORMANCE EVALUATION**

**MPI_Bcast**
- [Design Broadcast for DL Workloads using NCCL](#)
  - NCCL-augmented hybrid design in MVAPICH2-GDR for intra-node communication
  - Tuned inter-node communication using various algorithms like K-nomial Tree, Scatter-Allgather, etc.
  - Combine performance features of NCCL and MPI in a unified communication runtime
- **Performance Benefits**
  - Up to 2X improvement for micro-benchmarks
  - Up to 38% improvement for VGG training with CNTK

**OSU-Caffe**
- Co-Design MVAPICH2-GDR and Caffe
  - Provide design principles to overlap DNN training with MPI communication
  - MPI_Reduce: Efficient GPU-based designs for large-message reductions
  - Delivers better or comparable performance to production-grade DL frameworks
- **Performance Benefits**
  - MPI_Reduce: 130X speedup over OpenMP and 2.5X improvement over MVAPICH2-GDR
  - OSU-Caffe: Better/comparable performance to CNTK for AlexNet training
  - OSU-Caffe: Scale-out to 160 GPUs for GooleLeNet

**SUMMARY OF CONTRIBUTIONS**
- Tackle the challenge of designing a scalable and distributed DL framework
- Efficient Intra-node and Inter-node training
- Proven scale-out for GooleLeNet up to 160 GPUs
- Support for Small (CIFAR10/MNIST) and Large Datasets (ImageNet)
- Optimized Model Propagation and Gradient Aggregation
- Various Design Alternatives to provide Optimal Performance for Small and Large scale training