

Ammar Ahmad Awan

2015 Neil Ave. • Columbus • OH 43210 • USA
+1 614 360 8349 • ammar.ahmad.awan@gmail.com
<http://cse.osu.edu/~awan.10>

Publications

Select

1. **A. A. Awan**, K. V. Manian, C-H Chu, H. Subramoni, and DK Panda, *Optimized Large-Message Broadcast for Deep Learning Workloads: MPI, MPI+NCCL, or NCCL2?*, Parallel Computing (PARCO '19), Vol. 85, Pages 141-152, Jul 2019.
2. **A. A. Awan**, C-H Chu, X. Lu, H. Subramoni, and D. K. Panda, *OC-DNN: Exploiting Advanced Unified Memory Capabilities in CUDA 9 and Volta GPUs for Out-of-Core DNN Training*, 25th IEEE International Conference on High-Performance Computing, Data, Analytics, and Data Science (HiPC '18) '18, Dec 2018.
3. **A. A. Awan**, C-H Chu, X. Lu, H. Subramoni, and DK Panda, *Can Unified-Memory support on Pascal and Volta GPUs enable Out-of-Core DNN Training?*, ISC High-Performance (ISC '18), June 2018. **Best Student Poster Award**.
4. **A. A. Awan**, K. Hamidouche, J. Hashmi, and D. K. Panda, *S-Caffe: Co-designing MPI Runtimes and Caffe for Scalable Deep Learning on Modern GPU Clusters*, 22nd ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP '17), Feb 2017.
5. **A. A. Awan**, K. Hamidouche, A. Venkatesh, and D. K. Panda, *Efficient Large Message Broadcast using NCCL and CUDA-Aware MPI for Deep Learning*, 23rd European MPI Users' Group Meeting (EuroMPI '16), Sep 2016. **Best Paper Runner-Up**.

All Publications

Most updated list of publications is available from my [Google Scholar](#) page.

Journal Articles

1. **A. A. Awan**, A. Jain, C-H Chu, H. Subramoni, and DK Panda, *Communication Profiling and Characterization of Deep Learning Workloads on Clusters with High-Performance Interconnects*, IEEE Micro (Early Access: doi: 10.1109/MM.2019.2949986).
2. **A. A. Awan**, K. V. Manian, C-H Chu, H. Subramoni, and DK Panda, *Optimized Large-Message Broadcast for Deep Learning Workloads: MPI, MPI+NCCL, or NCCL2?*, Parallel Computing (PARCO '19), Vol. 85, Pages 141-152, July 2019.
3. C-H Chu, X. Lu, **A. A. Awan**, H. Subramoni, Bracy Elton, and DK Panda, *Exploiting Hardware Multicast and GPUDirect RDMA for Efficient Broadcast*, IEEE Transactions on Parallel and Distributed Systems (TPDS '19), Vol. 30, No. 3, Pages 575-588, Mar 2019.
4. K. Hamidouche, A. Venkatesh, **A. A. Awan**, H. Subramoni, and D. K. Panda, *CUDA-Aware OpenSHMEM: Extensions and Designs for High Performance OpenSHMEM on GPU Clusters*, Parallel Computing (PARCO '16), Vol. 58, Pages 27-36, Oct 2016.
5. Z. Pervez, **A. A. Awan**, A. M. Khattak, S. Y. Lee, and Eui-Nam Huh, *Privacy-aware searching with oblivious term matching for cloud storage*, Journal of Supercomputing, Vol. 63, Issue 2, Pages 538-560, Feb 2013.

Refereed Conference/Workshop Papers

1. A. Jain, **A. A. Awan**, H. Subramoni, and DK Panda, *Scaling TensorFlow, PyTorch, and MXNet using MVAPICH2 for High-Performance Deep Learning on Frontera*, 3rd Deep Learning on

Supercomputers Workshop, held in conjunction with SC '19, Nov 2019.

2. A. Jain, **A. A. Awan**, Q. Anthony, H. Subramoni, and DK Panda, *Performance Characterization of DNN Training using TensorFlow and PyTorch on Modern Clusters*, 21st IEEE International Conference on Cluster Computing, (Cluster '19), Sep 2019.
3. **A. A. Awan**, A. Jain, C-H Chu, H. Subramoni, and D. K. Panda, *Communication Profiling and Characterization of Deep Learning Workloads on Clusters with High-Performance Interconnects*, 26th Symposium on High-Performance Interconnects (HotI '19), Aug 2019.
4. **A. A. Awan**, J. Bedorf, C-H Chu, H. Subramoni, and D. K. Panda, *Scalable Distributed DNN Training using TensorFlow and CUDA-Aware MPI: Characterization, Designs, and Performance Evaluation*, 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid '19), May 2019.
5. K. Vadambacheri Manian, **A. A. Awan**, A. Ruhela, C. Chu, and D. K. Panda, *Characterizing CUDA Unified Memory (UM)-Aware MPI Designs on Modern GPU Architectures*, 12th Workshop on General Purpose Processing Using GPU (GPGPU '19), held in conjunction with ASPLOS '19, Apr 2019.
6. **A. A. Awan**, C-H Chu, X. Lu, H. Subramoni, and D. K. Panda, *OC-DNN: Exploiting Advanced Unified Memory Capabilities in CUDA 9 and Volta GPUs for Out-of-Core DNN Training*, IEEE 25th International Conference on High Performance Computing (HiPC '18), Dec 2018.
7. **A. A. Awan**, C-H Chu, H. Subramoni, D. K. Panda, *Optimized Broadcast for Deep Learning Workloads on Dense-GPU InfiniBand Clusters: MPI or NCCL?*, 25th European MPI Users' Group Meeting (EuroMPI '18), Sep 2018.
8. **A. A. Awan**, H. Subramoni, D. K. Panda, *An In-depth Performance Characterization of CPU- and GPU-based DNN Training on Modern Architectures*, 3rd Workshop on Machine Learning in HPC Environments (MLHPC '17), held in conjunction with SC '17, Nov 2017.
9. C-H Chu, X. Lu, **A. A. Awan**, H. Subramoni, J. Hashmi, Bracy Elton, and DK Panda, *Efficient and Scalable Multi-Source Streaming Broadcast on GPU Clusters for Deep Learning*, 46th International Conference on Parallel Processing (ICPP '17), Aug 2017.
10. **A. A. Awan**, K. Hamidouche, J. Hashmi, and D. K. Panda, *S-Caffe: Co-designing MPI Runtimes and Caffe for Scalable Deep Learning on Modern GPU Clusters*, 22nd ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming (PPoPP '17), Feb 2017.
11. K. Hamidouche, **A. A. Awan**, A. Venkatesh, and D. K. Panda, *CUDA M3: Designing Efficient CUDA Managed Memory-aware MPI by Exploiting GDR and IPC*, 23rd IEEE International Conference on High Performance Computing, Data, and Analytics, Dec 2016.
12. **A. A. Awan**, K. Hamidouche, A. Venkatesh, and D. K. Panda, *Efficient Large Message Broadcast using NCCL and CUDA-Aware MPI for Deep Learning*, 23rd European MPI Users' Group Meeting (EuroMPI '16), Sep 2016. **Best Paper Runner-Up**.
13. C. Chu, K. Hamidouche, A. Venkatesh, **A. A. Awan**, and D. K. Panda, *CUDA Kernel based Collective Reduction Operations on Large-scale GPU Clusters*, 16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid '16), May 2016.
14. **A. A. Awan**, K. Hamidouche, A. Venkatesh, J. Perkins, H. Subramoni, and D. K. Panda, *GPU-Aware Design, Implementation, and Evaluation of Non-blocking Collective Benchmark*, 22nd European MPI Users' Group Meeting (EuroMPI '15), Sep 2015.
15. K. Hamidouche, A. Venkatesh, **A. A. Awan**, H. Subramoni, and D. K. Panda, *Exploiting GPUDirect RDMA in Designing High Performance OpenSHMEM for NVIDIA GPU Clusters*, IEEE International Conference on Cluster Computing (Cluster '15), Sep 2015.
16. **A. A. Awan**, K. Hamidouche, C. Chu, and D. K. Panda, *A Case for Non-Blocking Collectives in OpenSHMEM: Design, Implementation, and Performance Evaluation using MVAPICH2-X*, Workshop on OpenSHMEM and Related Technologies (OpenSHMEM '15), Aug 2015.
17. H. Subramoni, **A. A. Awan**, K. Hamidouche, D. Pekurovsky, A. Venkatesh, S. Chakraborty, K. Tomko, and D. K. Panda, *Designing Non-Blocking Personalized Collectives with Near Perfect Overlap for RDMA-Enabled Clusters*, ISC High Performance (ISC '15), Jul 2015.
18. S. Chakraborty, H. Subramoni, J. Perkins, **A. A. Awan**, and D. K. Panda, *On-demand Connection Management for OpenSHMEM and OpenSHMEM+MPI* (HIPS '15), IPDPS Workshop, May 2015.
19. **A. A. Awan**, M. S. Ayub, A. Shafi and S. Lee, *Towards Efficient Support for Parallel I/O in Java HPC*, 13th International Conference on Parallel and Distributed Computing, Applications and

Technologies (PDCAT '12), Dec 2012.

20. M. B. Amin, W. A. Khan, **A. A. Awan**, and S. Y. Lee, "Intercloud Message Exchange Middleware", 6th International Conference on Ubiquitous Information Management and Communication (ICUIMC '12), Sep 2012.

Posters

1. **A. A. Awan** and DK Panda, *Co-designing Communication Middleware and Deep Learning Frameworks for High-Performance DNN Training on HPC Systems*, Doctoral Showcase at SC '19, Nov 2019.
2. **A. A. Awan**, H. Subramoni, and DK Panda, *Exploiting CUDA Unified Memory for Efficient Out-of-Core DNN Training*, Poster at NVIDIA GTC '19, April 2019.
3. **A. A. Awan**, C-H Chu, X. Lu, H. Subramoni, and DK Panda, *Can Unified-Memory support on Pascal and Volta GPUs enable Out-of-Core DNN Training?*, ISC High-Performance (ISC '18), Jun 2018. **Best Student Poster Award**.
4. **A. A. Awan** and DK Panda, *Co-designing MPI Runtimes and Deep Learning Frameworks for Scalable Distributed Training on GPU Clusters*, ACM Student Research Competition (SRC) poster at SC '17, Nov 2017.
5. **A. A. Awan**, M. B. Amin, S. Hussain, A. Shafi, S. Y. Lee, *An MPI-IO Compliant Java based Parallel I/O Library*, Poster at 13th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid '13), May 2013.