# Analyzing Hardware Parameters in GPU based HPC Platform

Saptarshi Bhowmik[1], Nikhil Jain[2],, Abhinav Bhatele[3] and Xin Yuan[1]

1. Florida State University, 2. NVIDIA, Inc, 3. University of Maryland

## Goals

- Analyze if a HPC system with fewer nodes each with more compute capability perform better than a system with more nodes each with lesser compute capability.
- Use discrete-event simulations to study several parameters including network bandwidth, number of GPUs per node in the context of two most popular network topology Fat-Tree and 1D-Dragonfly.

## Introduction

Today's high-end HPC clusters employ many GPUs per node. The performance of applications on such a platform depends heavily on the interconnection network performance [1]. As such, it is important to understand the impact of hardware parameters on the overall application and system performance. In this research, we perform extensive simulation study to understand hardware parameters and their impact on the performance of HPC workload.

## Method

**Tracer-CODES** We use the discrete event driven simulator, TraceR-CODES[**?**] to replay the application traces. The simulator network model is used to implement the interconnect topology, on top of which the traces are replayed.

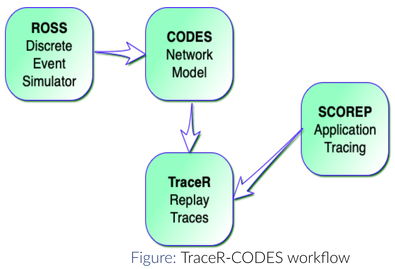


Figure: TraceR-CODES workflow

**Applications**

Six popular HPC Applications are use in the simulation.

| Traces | Computation Intensive | Communication Intensive |
|---|---|---|
| Stencil4d | ✗ | ✓ |
| Kripke | ✓ | ✗ |
| Laghos | ✓ | ✗ |
| Subcomm-a2a | ✗ | ✓ |
| Sw4lite | ✓ | ✓ |
| Amg | ✓ | ✓ |

Table: Profiles of Application Traces

## Simulation Environment Parameters

**Network Topologies**

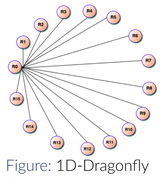

Figure: 1D-Dragonfly Group

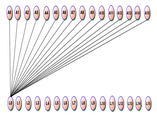

Figure: Fat-Tree Pod

| Topology | 1-GPU/Node | 2-GPU/Node | 4-GPU/Node | 8-GPU/Node |
|---|---|---|---|---|
| 1D-Dragonfly | 16 Group | 8 Group | 4 Group | 2 Group |
| Fat-Tree | 8 Pods | 8 Pods | 4 Pods | 2 Pods |

Table: Profiles of Application Traces

**Bandwidth**

**Default Setting :**

Link Bandwidth = 11.9 Gb/s

Internal Bandwidth = 11.9 Gb/s

We use 8 more bandwidth, x/16, x/8, x/4, x/2, 2x, 4x, 8x, 16x, for our simulations.

**GPUs per node**


Figure: 1-GPU/Node


Figure: 2-GPU/Node


Figure: 4-GPU/Node

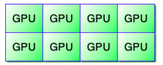


Figure: 8-GPU/Node

From 1-GPU/Node, 2-GPU/Node, 4-GPU/Node and 8-GPU/Node configuration.

## Result

**Impact of GPUs per Node**

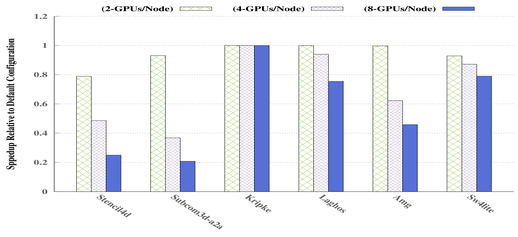


Figure: Fat-Tree

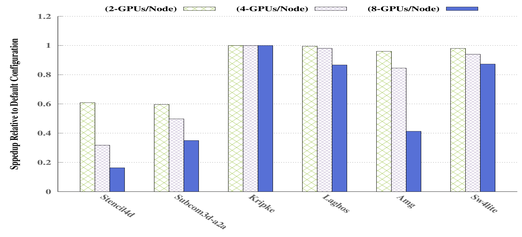


Figure: 1D-Dragonfly

Communication intensive applications experience more slowdown than applications with computation.

**Impact of Bandwidth**

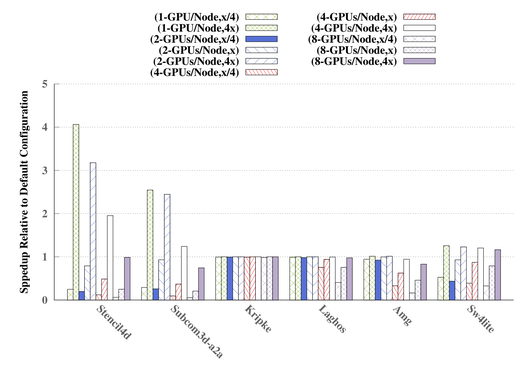


Figure: Fat-Tree

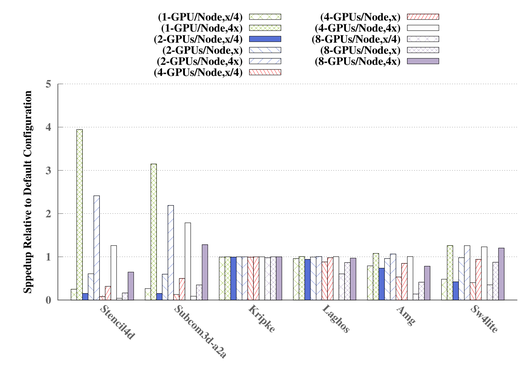


Figure: 1D-Dragonfly

Applications performance speedup when the bandwidth is increased.
The increase is more pronounced is full computational intensive applications.

## Conclusion

- As the number of GPU per node increases, the node becomes more intensive, and thus, there is a slowdown in application performance as the communication/computation capacity of the network reduces.
- As the number of GPU per node increases, more bandwidth is needed to substantiate the slowdown in application performance.
- Every application has a sweet spot where it is performing the best.

## Future Works

- Study how other simulation environment and hardware design, such as NIC scheduling policies effect the performance of applications
- Profile more HPC applications and find the performance of those applications across the currently deployed GPU based interconnect topology,

## References

[1] Jain, N., Bhatele, A., Howell, L. H., Böhme, D., Karlin, I., León, E. A., ... Leininger, M. L. (2017, November). Predicting the performance impact of different fat-tree configurations. In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (pp. 1-13).

[2] Kim, J., Dally, W. J., Scott, S., Abts, D. (2008, June). Technology-driven, highly-scalable dragonfly topology. In 2008 International Symposium on Computer Architecture (pp. 77-88). IEEE.

[3] Knüpfer, A., Rössel, C., an Mey, D., Biersdorff, S., Diethelm, K., Eschweiler, D., ... Nagel, W. E. (2012). Score-p: A joint performance measurement run-time infrastructure for periscope, scalasca, tau, and vampir. In Tools for High Performance Computing 2011 (pp. 79-91). Springer, Berlin, Heidelberg.

[4] Acun, B., Jain, N., Bhatele, A., Mubarak, M., Carothers, C. D., Kale, L. V. (2015, August). Preliminary evaluation of a parallel trace replay tool for hpc network simulations. In European Conference on Parallel Processing (pp. 417-429). Springer, Cham.

[5] Alzaid, Z. S. A., Bhowmik, S., Yuan, X., Lang, M. (2020, June). Global link arrangement for practical Dragonfly. In Proceedings of the 34th ACM International Conference on Supercomputing (pp. 1-11). Magnetics Japan, p. 301, 1982].