# Evaluating Performance of Burst Buffer Models for Real-World Application Workloads in HPC Systems

Harsh Khetawat
Department of Computer Science
North Carolina State University
Raleigh, North Carolina
Email: hkhetaw@ncsu.edu

Frank Mueller
Department of Computer Science
North Carolina State University
Raleigh, North Carolina
Email: mueller@cs.ncsu.edu

Christopher Zimmer
Oak Ridge National Laboratory
Oak Ridge, Tennessee
Email: zimmercj@ornl.gov

***Abstract*—Burst buffers have become increasingly popular in HPC systems, allowing bursty I/O traffic to be serviced faster without slowing down application execution. The ubiquity of burst buffers creates opportunities for studying the ideal placement of these fast storage devices in the HPC topology. Furthermore, the topology of the network interconnect can also affect the performance of the storage hierarchy for different burst buffer placement schemes. As part of this work, we would like to simulate I/O from real-world application workloads across burst buffer architectures and network topologies. We will study the performance of these models for large mixed systems with varying application workloads and create a reproducible tool to allow individual centers to acquire models that fit their workload characteristics.**

## I. Introduction

With the increasing scale of computation and data in High Performance Computing (HPC), the existing storage systems are becoming a bottleneck in the progress of scientific and data-intensive applications. To alleviate this bottleneck, the shared parallel file system model is being augmented with a tier of intermediate, high-bandwidth storage. This intermediate burst buffer services application I/O for checkpoint/restart operations, staging data, or serves as a write-through cache for the parallel file system.

Various architectures have been explored for the placement of burst buffers in HPC systems:

- Burst buffers co-located with compute nodes: used in Summit, the next-generation OLCF system
- Burst buffers co-located with I/O nodes: used in Cori, an HPC system at NERSC
- Burst buffers on separate set of nodes

The advantages and disadvantages of these architectures have been studied in previous work[1]. Moreover, modern HPC systems employ a range of network topologies, e.g., fattree, dragonfly, slimfly, etc. As part of this work, we would like to simulate these network and burst buffer architectures with I/O traces from real-world applications. Our goal is to simulate a full system with multi-tenant workloads to assess the expected performance of a deployed HPC system. We also aim to enable tools with models that individual supercomputing centers can employ in order to fit their workloads.

The simulations would enable us to explore not only the different burst buffer architectures, but also performance under different striping and protection schemes.

### A. Simulation

We use the CODES simulation suite from Argonne National Laboratory (ANL)[2] to design and create our simulations. CODES is a Parallel Discrete Event Simulation (PDES) framework built on ROSS, which has support for the most popular network topologies. By combining these network topologies, and storage architectures we have created models for collecting various network and storage performance metrics. Fig. 1 shows the CODES simulation suite with the pluggable network and storage models, and workload generators.

CODES supports the replay of Darshan I/O traces as part of the simulation. Using this capability we can replay actual application workloads in order to measure performance of an HPC system under expected workloads. Furthermore, we can explore various strategies for striping data across burst buffers and resilience methods such as neighbor copy and other encoding techniques. We have run preliminary simulations of burst buffer architectures and various network topologies with small to medium sized mixed application traces.

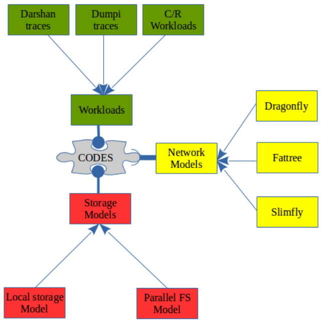

Fig. 1. The CODES simulation suite

## References

[1] Harms, Kevin and Oral, H Sarp and Atchley, Scott and Vazhkudai, Sudharshan S, *Impact of Burst Buffer Architectures on Application Portability*, Oak Ridge National Laboratory, Oak Ridge, TN, 2016

[2] Cope, Jason, et al., *Codes: Enabling co-design of multilayer exascale storage architectures.*, Proceedings of the Workshop on Emerging Supercomputing Technologies. Vol. 2011. 2011.