

## Introduction

- The complexity of HPC systems and applications, fueled by data-driven AI and ML workloads, poses challenges for I/O performance, impacting overall system efficiency.
- Thorough I/O analysis is essential to identify potential I/O bottlenecks, but it's challenging due to multiple metrics involved.
- Studies demonstrate that the causes of low I/O performance in applications can be diverse.
- This work presents a methodology that uses application I/O traces and simultaneously employs multiple metrics to identify I/O performance issues.
- Three scientific workloads with diverse I/O behaviors were analyzed using I/O time, I/O bandwidth, and IOPS metrics.
- Our key findings can be summarized as follows:
  - Different metrics uncover different I/O bottlenecks.
  - Specific I/O behaviors can only be captured by certain metrics.

## Methodology

- Chosen scientific workloads: HACC, Montage, and CM1.
- Performance metrics used: I/O time, I/O bandwidth, and IOPS.
- Criteria for detecting I/O bottlenecks:
  - I/O time: Records where time exceeded 90% of maximum I/O time per process.
  - I/O bandwidth: Records with throughput below 10MB/s.
  - IOPS: Records with operation rates less than 10% of maximum IOPS per process.

## Impacts of Multiple I/O Metrics

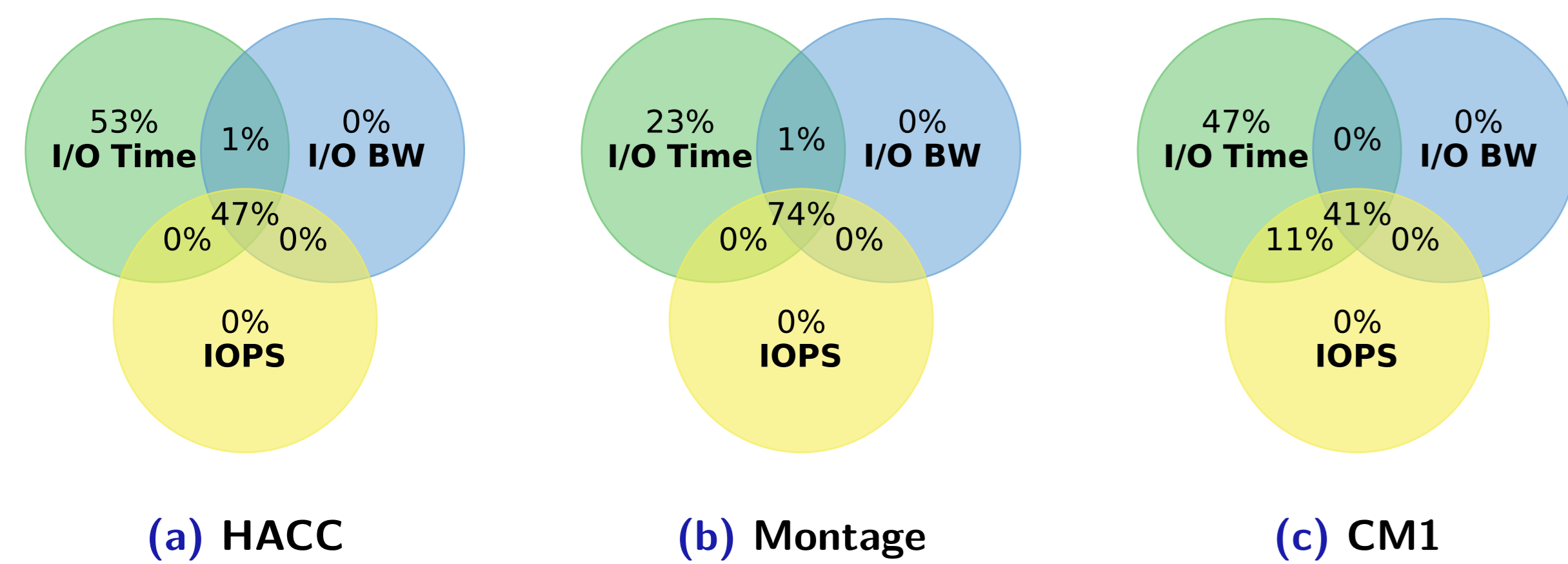
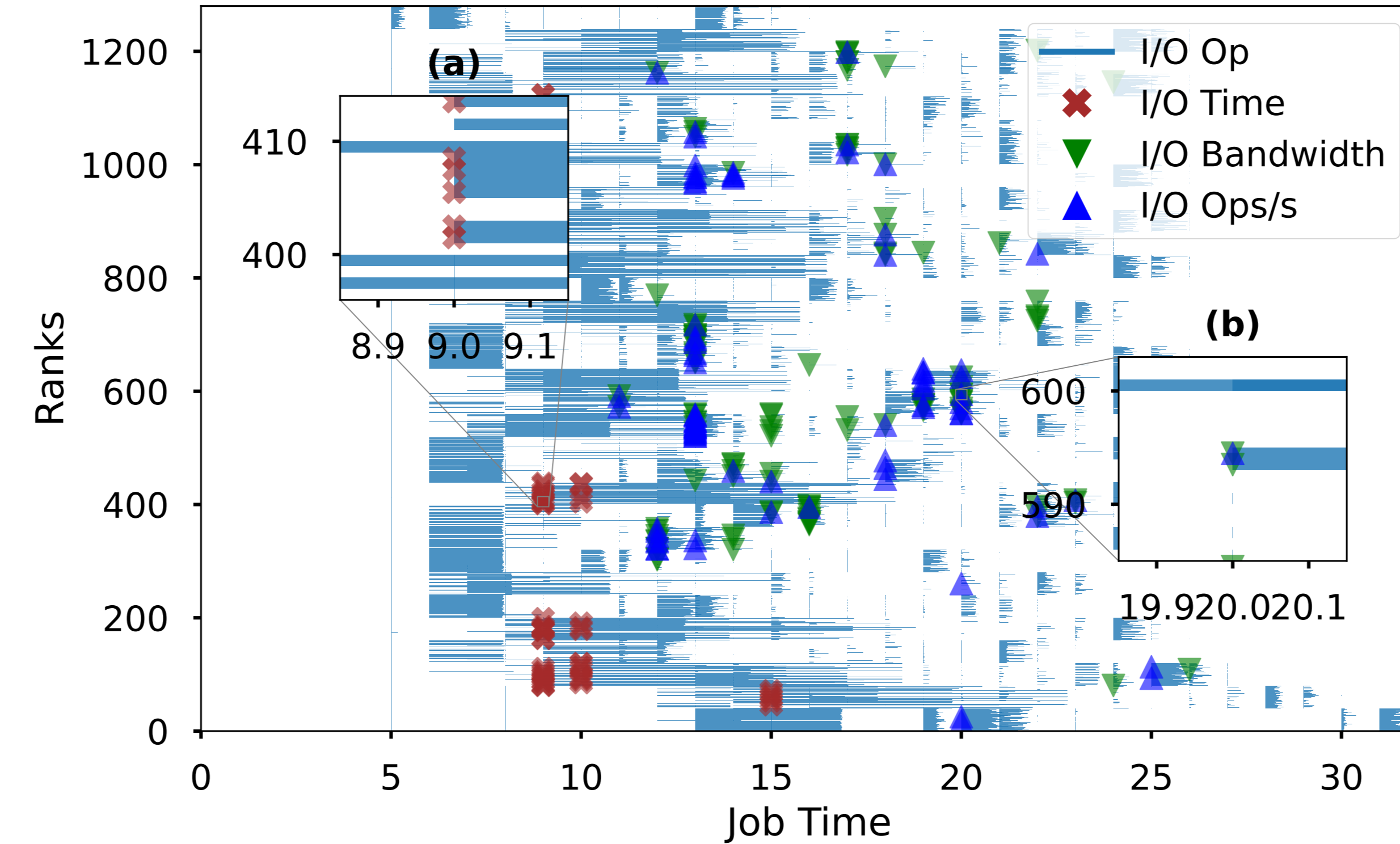


Figure 1: Distribution and overlap of I/O bottlenecks in scientific workloads

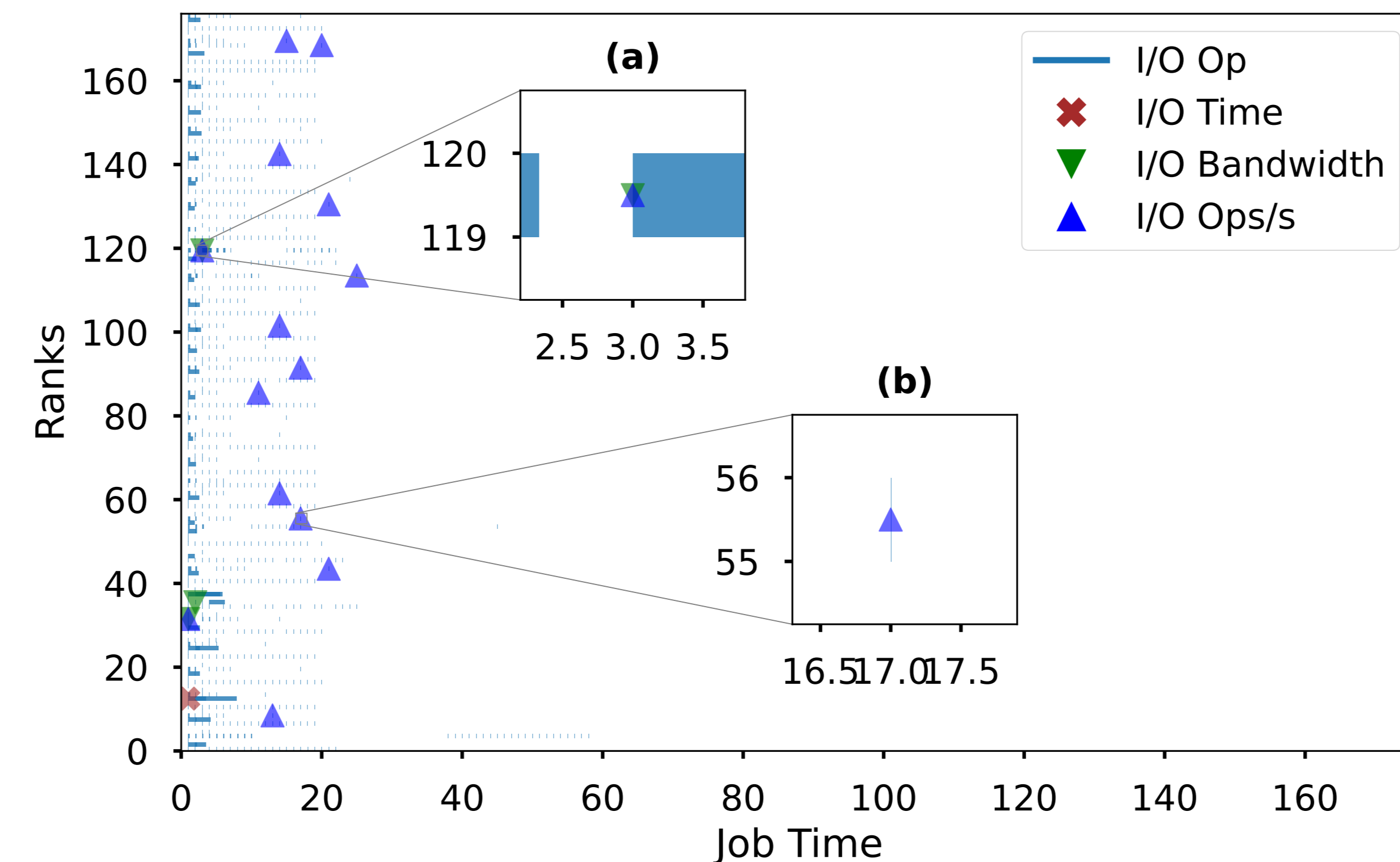
- The workloads exhibit diverse I/O behaviors, driven by their unique characteristics and functionalities.
- Figure 1 shows the distribution of I/O bottlenecks observed in the workloads per each criteria and an overlap analysis between them.
- The use cases demonstrate comprehensive I/O analysis on a timeline, showcasing different I/O bottlenecks are detected by different metrics.

## HACC (Use Case: Simulation with Checkpoint & Restart)



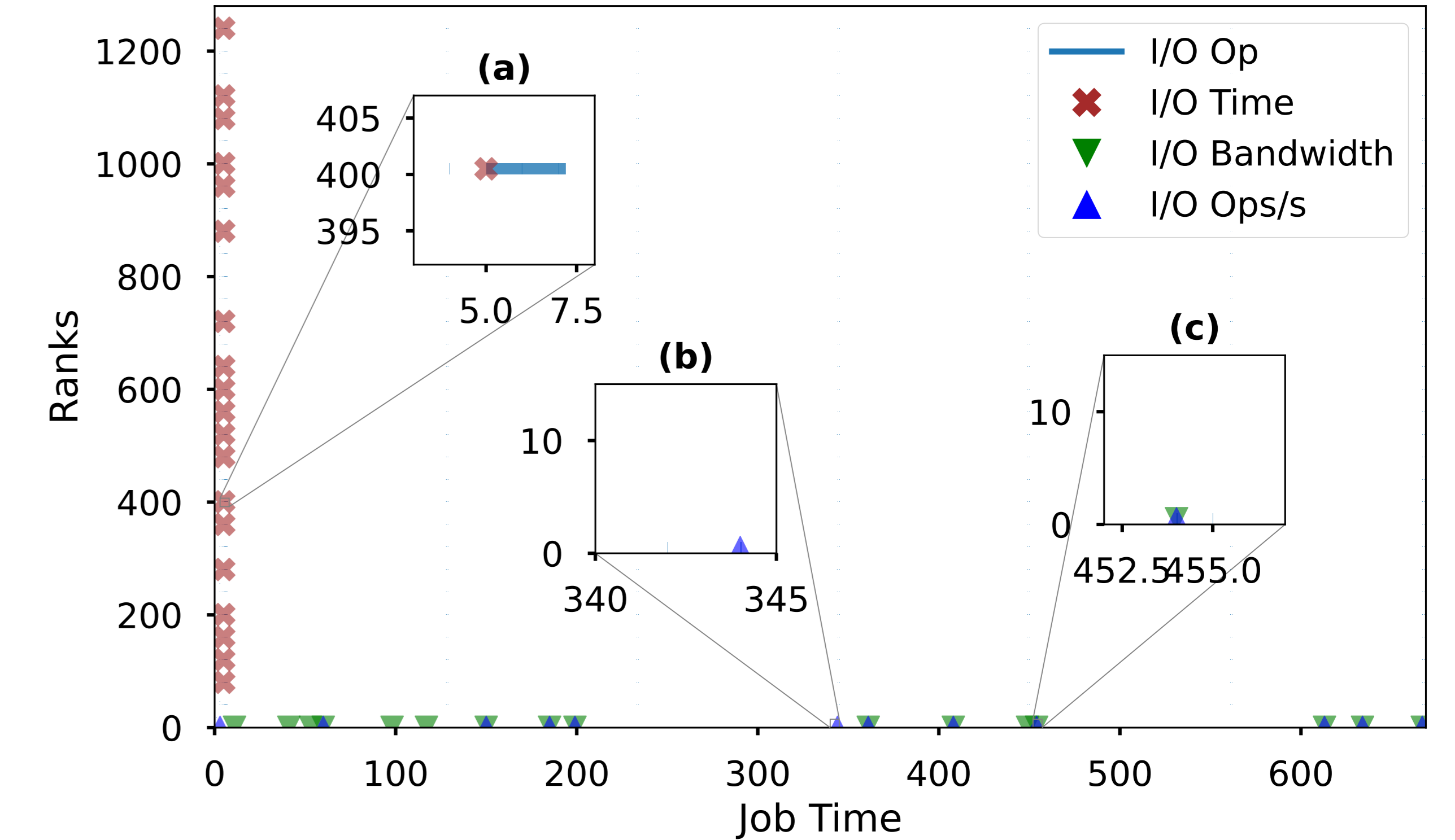
- Multiple ranks "open"ing simulation files on GPFS concurrently lead to over 90% of I/O time consumed by metadata operations and resulting in I/O bottlenecks *per I/O time*.
- High parallelism during checkpointing causes GPFS contention, resulting in I/O bottlenecks *per both I/O BW and IOPS* due to very low I/O bandwidth and IOPS.

## Montage (Use Case: Workflow with Complex Dependencies)



- Small "read"s (<3KB) on FITS files leads to I/O bottlenecks *per both I/O bandwidth and IOPS* due to very low rates.
- Slow "open"s during PNG image generation causes I/O bottlenecks *per IOPS* due to very low IOPS.

## CM1 (Use Case: Simulation with Separate I/O Phases)



- Simultaneously "open"ing the same configuration file leads to metadata contention, causing over 90% of I/O time to be spent on metadata operations and resulting in I/O bottlenecks *per I/O time*.
- Simulation data writes are dominated by metadata operations, resulting in very low IOPS and, hence, I/O bottlenecks *per IOPS*.
- Simulation data writes dominated by small "write"s exhibit very low I/O BW and IOPS, hence are detected as I/O bottlenecks *per both I/O BW and IOPS*.

## Conclusion

- In this work, we presented a comprehensive I/O analysis using multiple metrics, namely I/O time, I/O bandwidth, and IOPS.
- Through the evaluation of three diverse scientific workloads, we demonstrated that different metrics uncover different I/O bottlenecks.
- Our findings demonstrate that specific I/O behaviors, such as contention on GPFS, can only be identified through certain metrics, further highlighting the need for considering multiple metrics.

## Acknowledgments

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research under the DOE Early Career Research Program (LLNL-POST-854293). Also, the material is based upon work supported by the National Science Foundation under Grant no. NSF OAC-2104013, OCI-1835764, CSR-1814872, and CSSI-2104013.

## Poster QR

