

Evaluation of MPI's Derived Datatypes on modern GPU Systems

Mike Söhner, Christoph Niethammer, and José Gracia

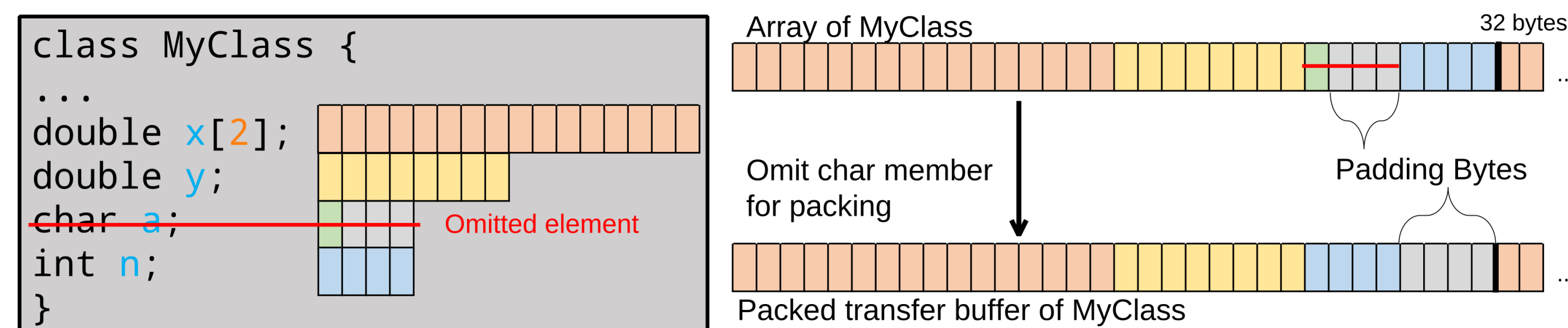
High-Performance Computing Center Stuttgart (HLRS), University of Stuttgart

Motivation

How do MPI's derived datatypes perform on systems ...

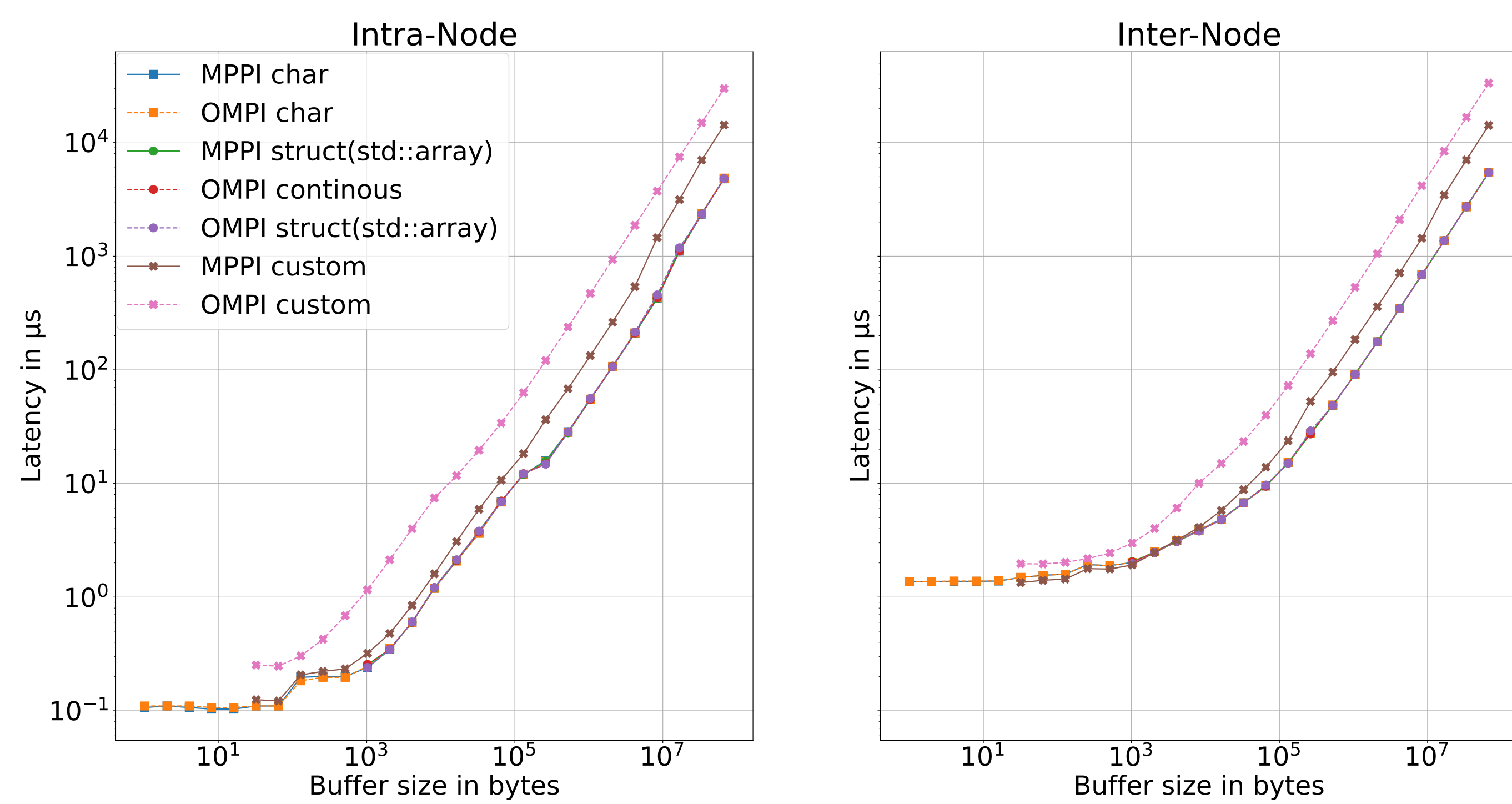
- ... with discrete GPUs, employing separate memory spaces for CPUs and GPUs?
- ... with Unified-Memory that provides a common address space for CPUs and GPUs?

Revisiting: MPI's derived datatypes (DDT)



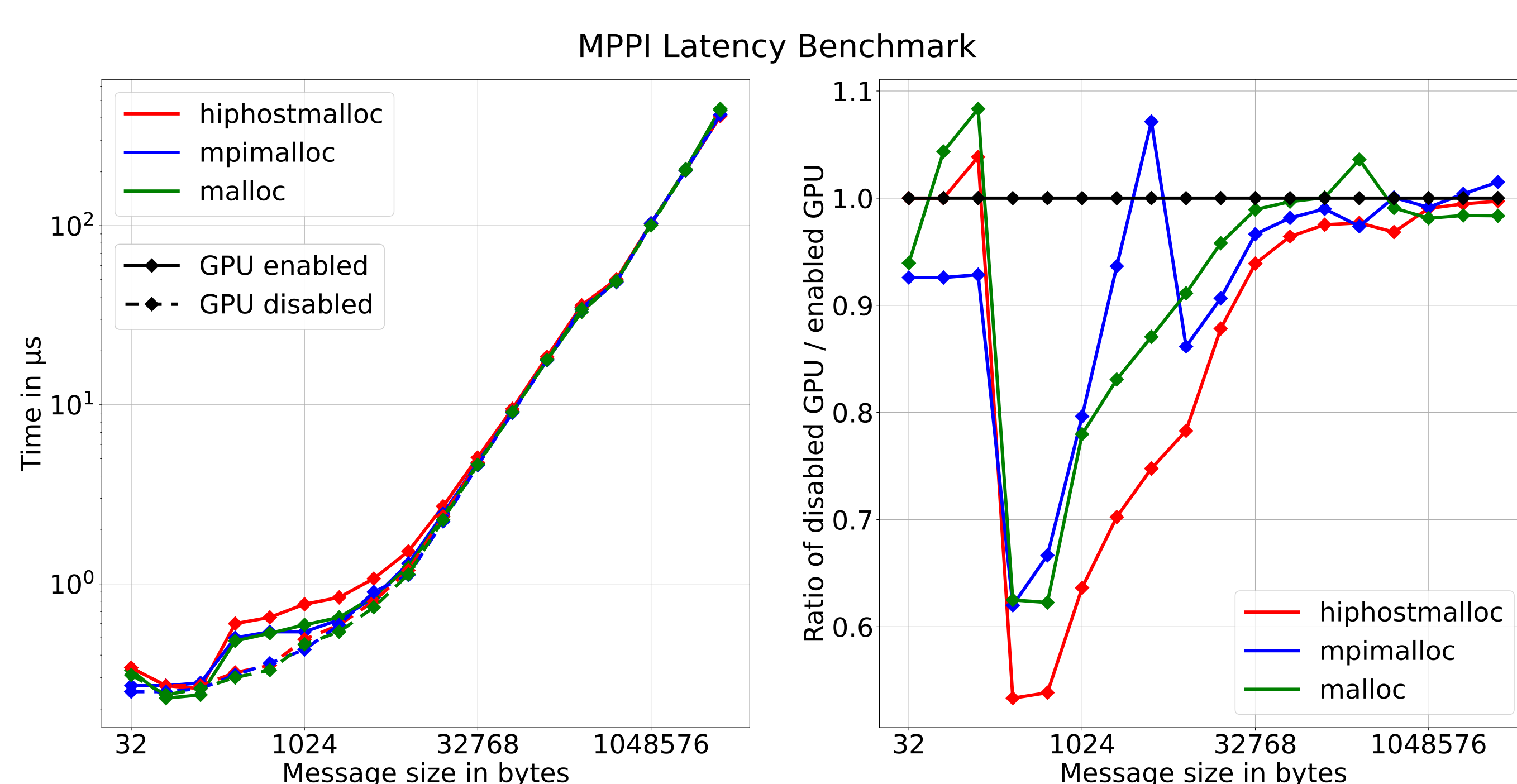
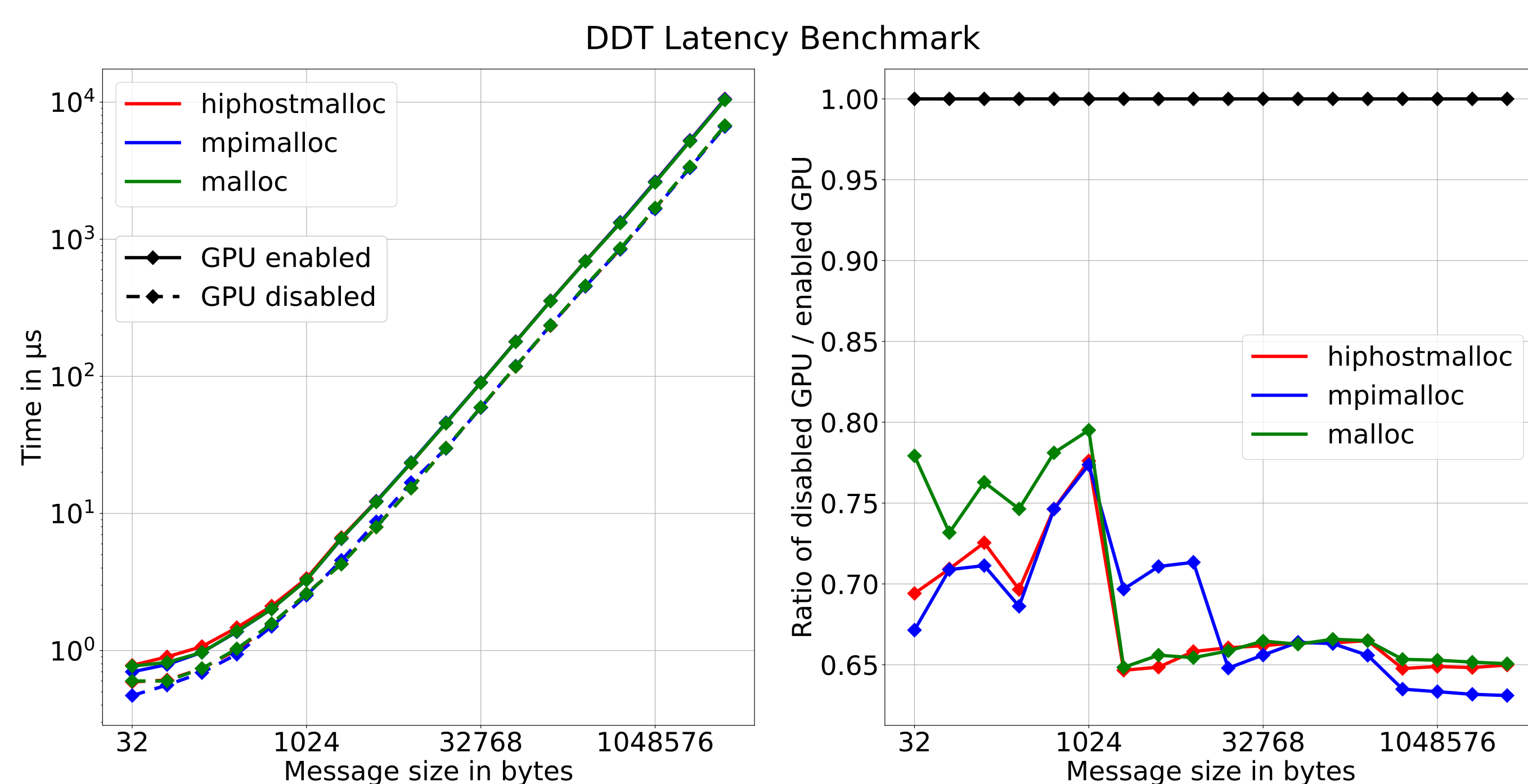
Revisiting: MPPI, C++ library for derived datatypes [SN25]

```
class MyClass {  
private:  
    std::array<double, 2> x;  
    double y;  
    char a;  
    int n;  
};  
  
constexpr mppl::Pattern<MyClass, "x", "y", "n"> pattern;
```

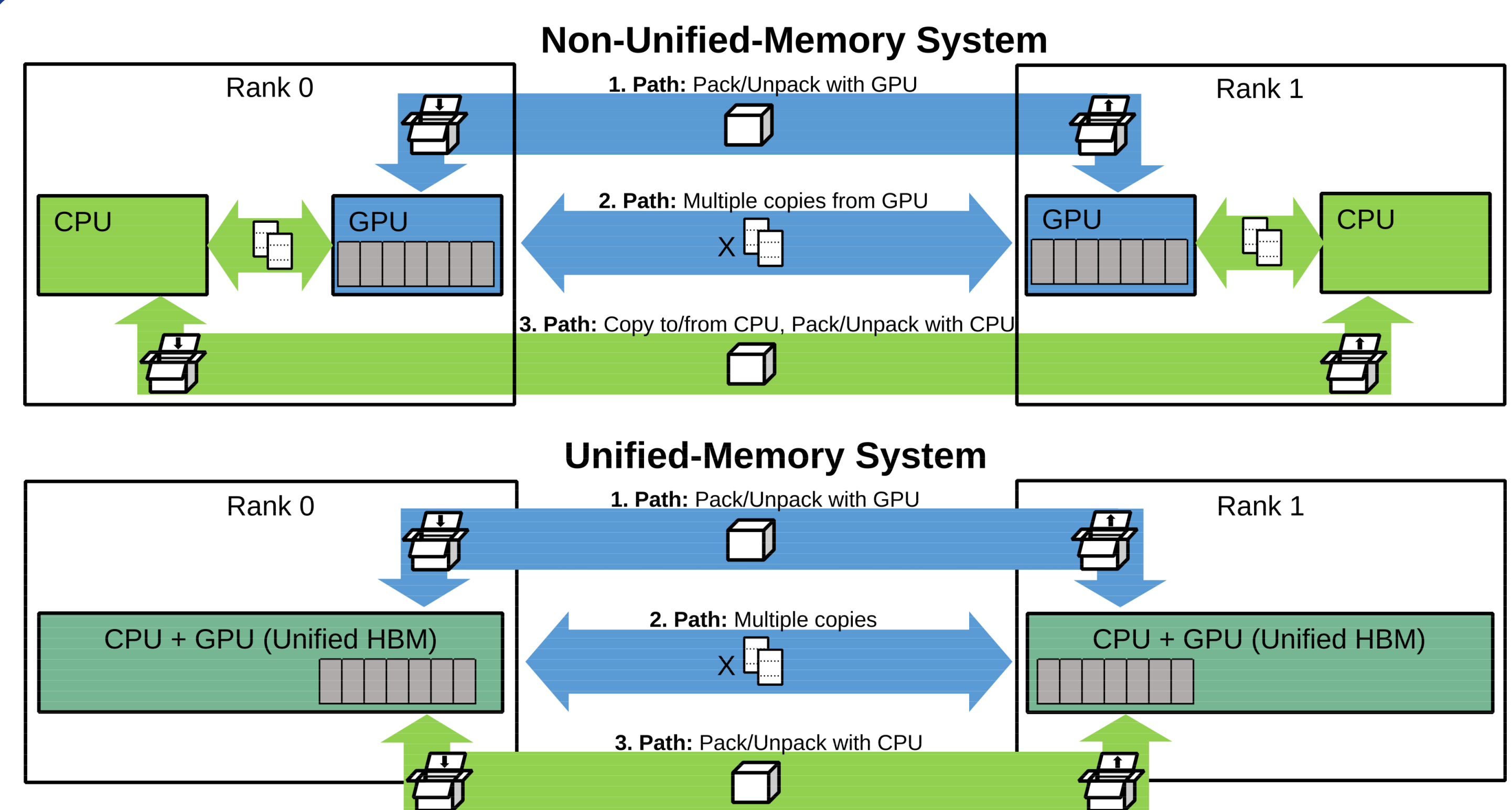


Derived datatypes performances on Unified Memory Systems

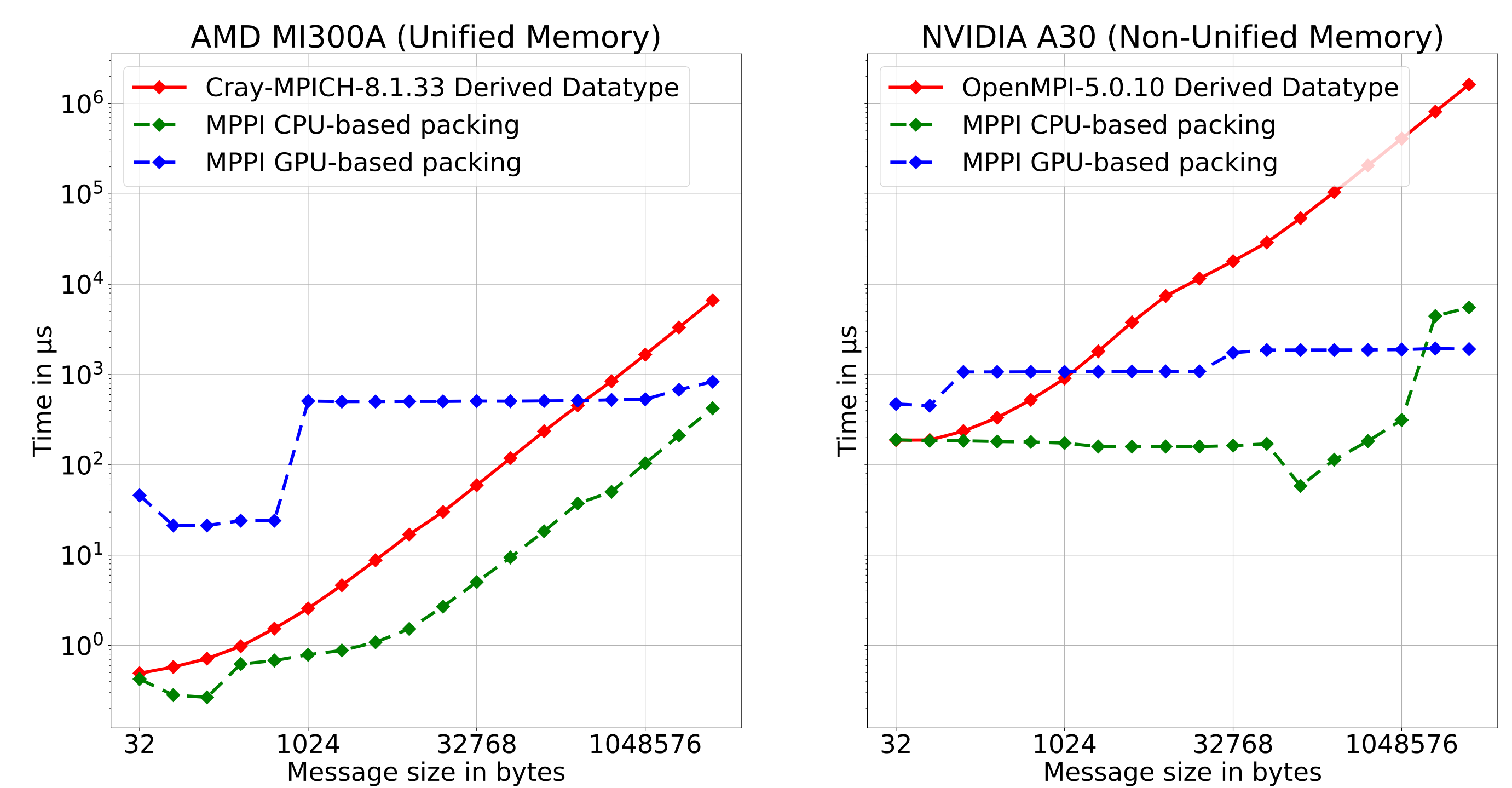
Experiments were done on an AMD MI300A with Cray-MPICH 8.1.33 and ROCm 6.4.1.



MPI on Unified-Memory Systems [Wu+16]



Performance on Unified- and Separate-Memory Systems



Conclusion

- Implemented GPU-based packing for MPPI.
- Shown impact of different memory-allocation strategies.
- Even seemingly unrelated options in software system settings can have unexpected implications.
- Exact reason for performance drop caused by GPU settings still unknown.
 - Possible cause: Cray MPICH's GPU-transfer layer (GTL) library checks all pointer given to MPI.
 - Though this does not explain why performance drop scales with problem size.

Outlook

Further investigations:

- Identify reasons for performance drop with Cray MPICH in unified-memory systems.
- Analyze how performance of derived datatypes can be more portable across all system setups.
- Improve of GPU-based packing of MPPI.
- Perform more in-depth comparisons with Jenkins et al. [Jen+14] and Wu et al. [Wu+16]
- Include more derived datatype creation strategies into MPPI. I.e. Subarrays for Ghost-Cell Exchanges.

References

- [Jen+14] John Jenkins et al. "Processing MPI Derived Datatypes on Noncontiguous GPU-Resident Data". In: *IEEE Transactions on Parallel and Distributed Systems* 25.10 (2014), pp. 2627–2637. DOI: 10.1109/TPDS.2013.234.
- [Wu+16] Wei Wu et al. "GPU-Aware Non-contiguous Data Movement In Open MPI". In: *Proceedings of the 25th ACM International Symposium on High-Performance Parallel and Distributed Computing*. HPDC '16. Kyoto, Japan: Association for Computing Machinery, 2016, pp. 231–242. ISBN: 9781450343145. DOI: 10.1145/2907294.2907317.
- [SN25] Mike Söhner and Christoph Niethammer. "MPPI - Type safe C++ Datatypes for MPI". In: *Proceedings of the SC '25 Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis*. SC Workshops '25. Association for Computing Machinery, 2025, pp. 439–448. ISBN: 9798400718717. DOI: 10.1145/3731599.3767390.

Acknowledgments

This work has been supported by the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 101143931 (POP3).