

Introduction of Open MPI+UCX+UCC

<u>Shinji Sumimoto</u> The University of Tokyo

2024/6/28



Presentation Overview

Introduction of

- Open MPI
- UCX
- •UCC



What is Open MPI

- <u>Open MPI</u> is a community-based opensource implementations of the <u>MPI</u> <u>standard</u>
- Short list of features
 - Full MPI-4.1 standards conformance
 - Thread safety and concurrency
 - Dynamic process spawning
 - Network and process fault tolerance
 - Run-time instrumentation
 - Many job schedulers supported
 - Accelerators ready (CUDA/ROCM/ZE)
- Composed of dynamic components that can be configured in/out depending on the setup and user needs

← → C ⋒ 😁 open-mpi.org		@ ☆	C 🕄	💩 💋 🖆	L 📓 🕹	Ď	L	
□ bookmark □ Firefox ブックマーク	□ IE ブックマーク □ Tech □ Well □ メ	ディア 🗅 リンク G	e @DIME 7	ットダイム	»	🗅 কৃশ	10	
	Open MPI: Open Source Higl	ו Perfor	manc	e Cor	nputi	ng		
	Home Support FAQ Search >>							
About Presentations Open MPI Team FAQ Videos Performance Open MPI Software Download Documentation Source Code Access Bug Tracking Regression Testing Version Information Sub-Projects	A High Performance Message Passing Library The Open MPI Project is an open source <u>Message Passing Interface</u> implementation that is developed and maintained by a consortium of academic, research, and industry partners. Open MPI is therefore able to combine the expertise, technologies, and resources from all across the High Performance Computing community in order to build the best MPI library available. Open MPI offers advantages for system and software vendors, application developers and computer science researchers. Features implemented or in short-term development for Open MPI include:				V5.0.3 release Major no release > <u>Read</u> Open M v4.1.6 release	released Major new release > <u>Read more</u> Open MPI v4.1.6 released		
Hardware Locality Network Locality Network Locality Open Tool for Parameter Optimization Community Mailing Lists Getting Help/Support Contribute Controlbute Contact License	conformance 64 bit) • Thread safety and concurrency • Production quality software • Dynamic process spawning • High performance on all platforms • Network and process fault tolerance • Portable and maintainable users h • Support network heterogeneity • Component-based design.					Bug fix release > Read more hwloc 2.11.0 Major release > Read more		

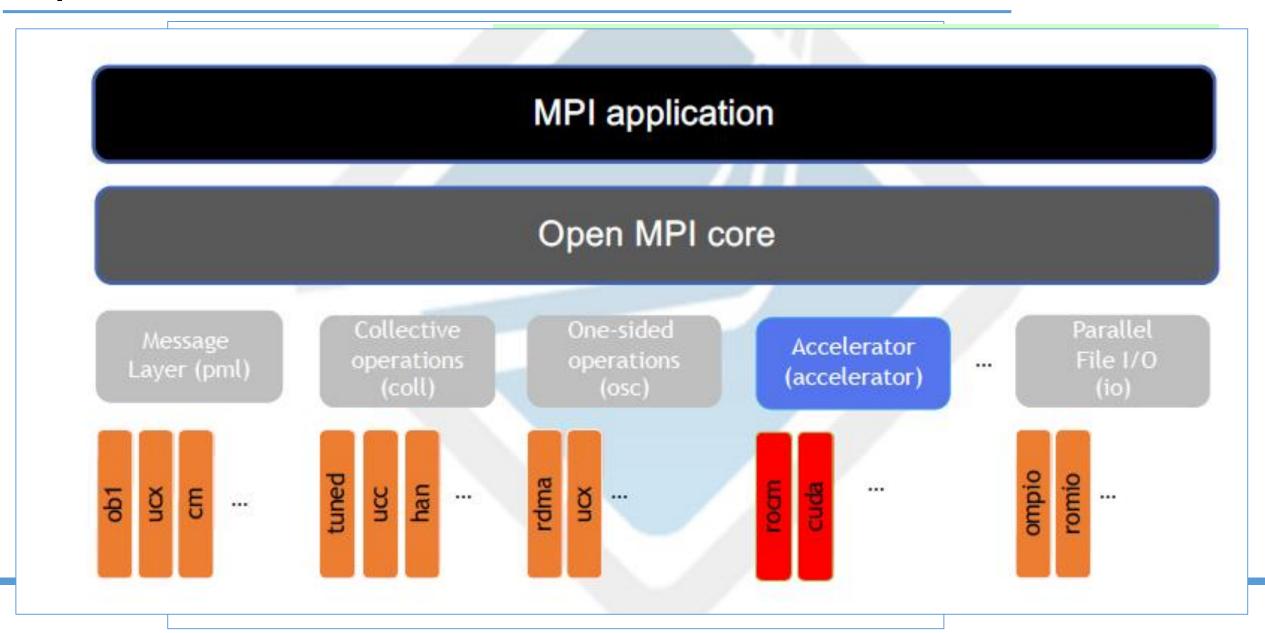
Open MPI is developed in a true open source fashion by a consortium of research, academic, and industry partners. The <u>Open MPI Team</u> page has a comprehensive listing of all contributors and active members.

https://www.open-mpi.org/

From George Bosilca



Open MPI Software Stacks for GPU



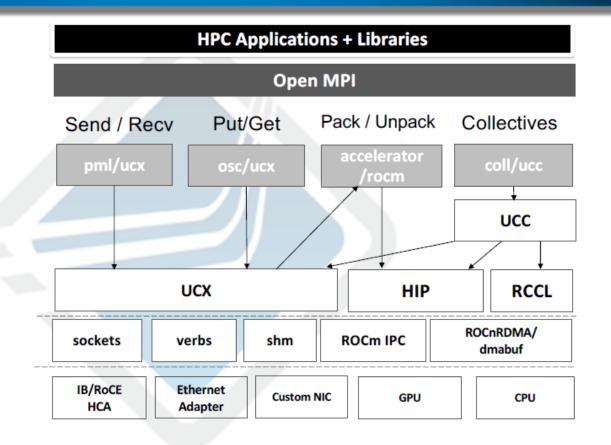


Open MPI Software Stacks

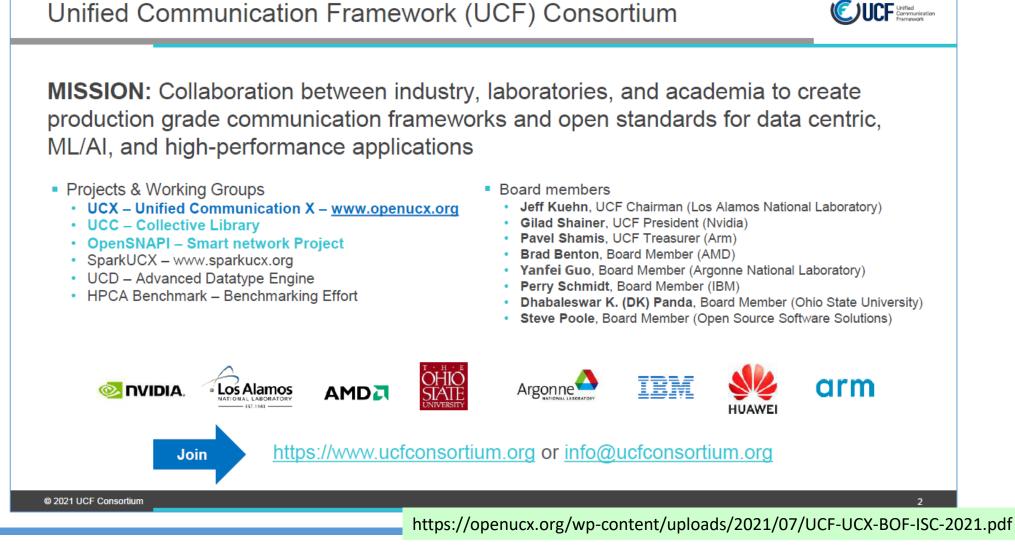
https://www-lb.open-mpi.org/papers/sc-2023/Open-MPI-SC23-BOF.pdf

ROCm Aware Open MPI Software Stack with UCX and UCC

- Recommended software stack for InfiniBand and RoCE networks
- Most stable and best tested configuration

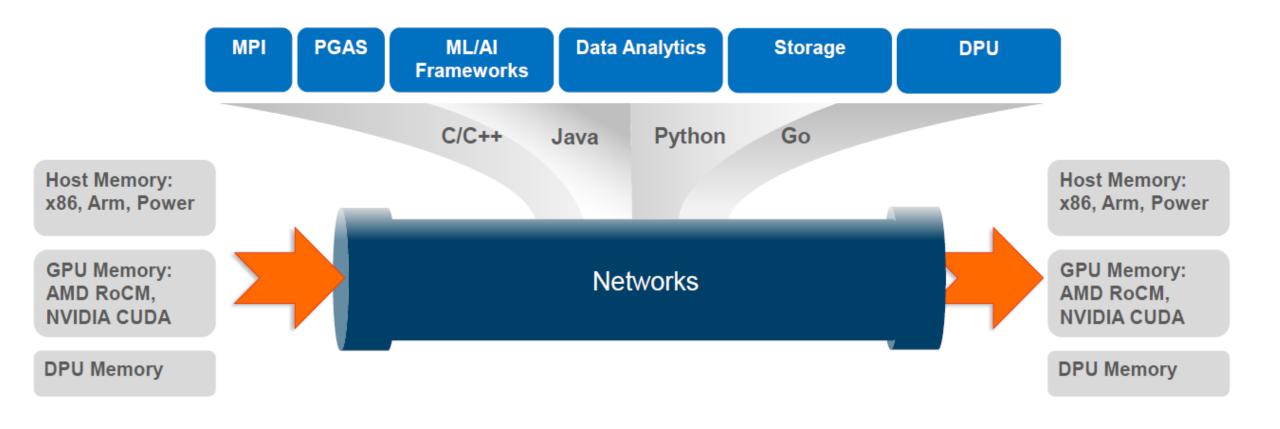


UCX and UCC by UCF Consortium



Why UCX ?





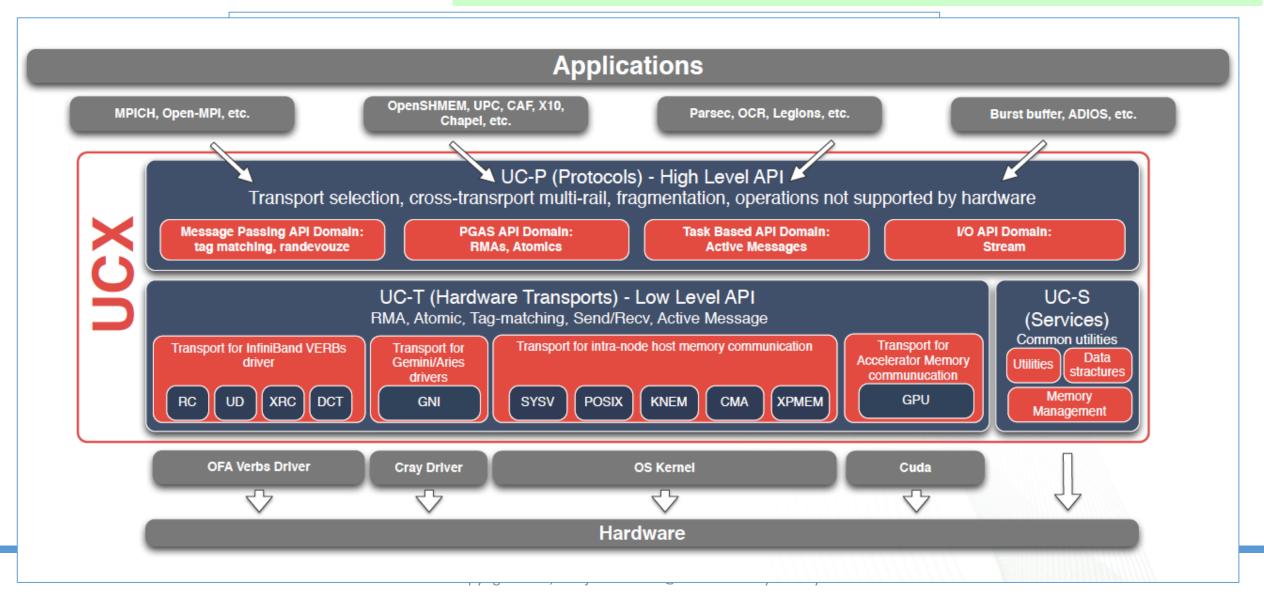
High-Performance Universal Data Mover

https://openucx.org/wp-content/uploads/2021/07/UCF-UCX-BOF-ISC-2021.pdf



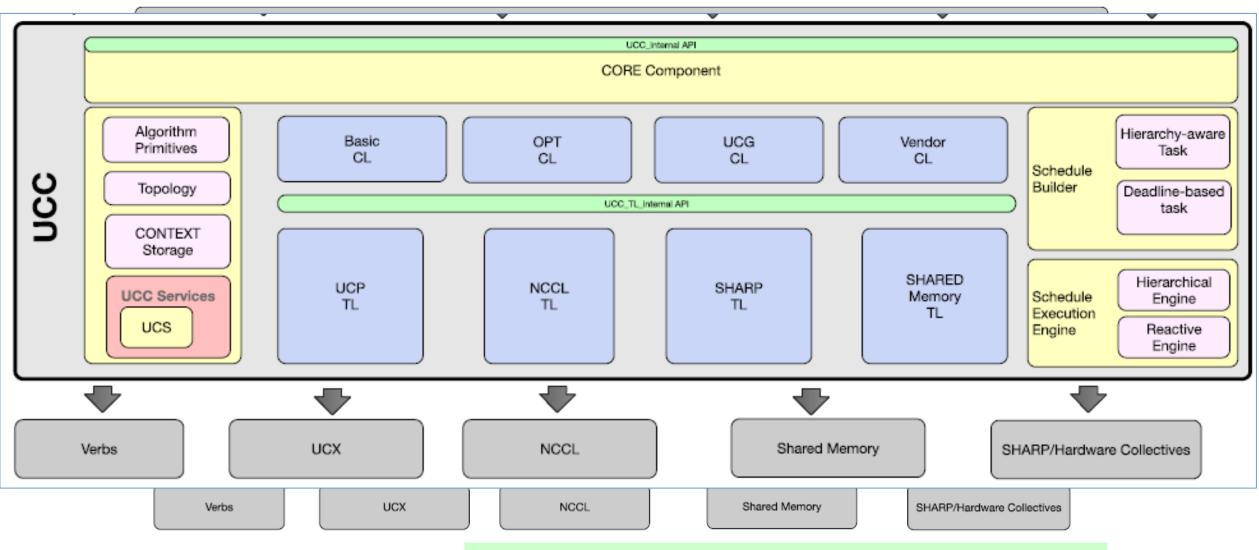
UCX: High-level Overview

https://openucx.org/wp-content/uploads/2015/11/UCX-booth-presentation.pdf



UCC Reference Implementation: Component Diagram





https://openucx.org/wp-content/uploads/2021/07/UCF-UCX-BOF-ISC-2021.pdf



- Unified collective stack for HPC and DL/ML workloads
 - Need to support a wide variety of semantics
 - · Need to optimize for different performance sensitives latency, bandwidth, throughput
 - · Need for flexible resource scheduling and ordering model
- Unified collective stack for software and hardware transports
 - Need for complex resource management scheduling, sharing, and exhaustion
 - Need to support multiple semantic differences reliability, completion
- Unify parallelism and concurrency
 - Concurrency progress of a collective and the computation
 - Parallelism progress of many independent collectives
- Unify execution models for CPU, GPU, and DPU collectives
 - Two-way execution model control operations are tightly integrated
 - Do active progress, returns values, errors, and callbacks with less overhead
 - One-way execution model control operations are loosely integrated
 - passive progress, and handle return values (GPU/DPUs)

https://openucx.org/wp-content/uploads/2021/07/UCF-UCX-BOF-ISC-2021.pdf



Questions?