Omni-Path Express (OPX) Libfabric Provider: Overview & Case Study

Douglas Fuller
Director, Software Engineering

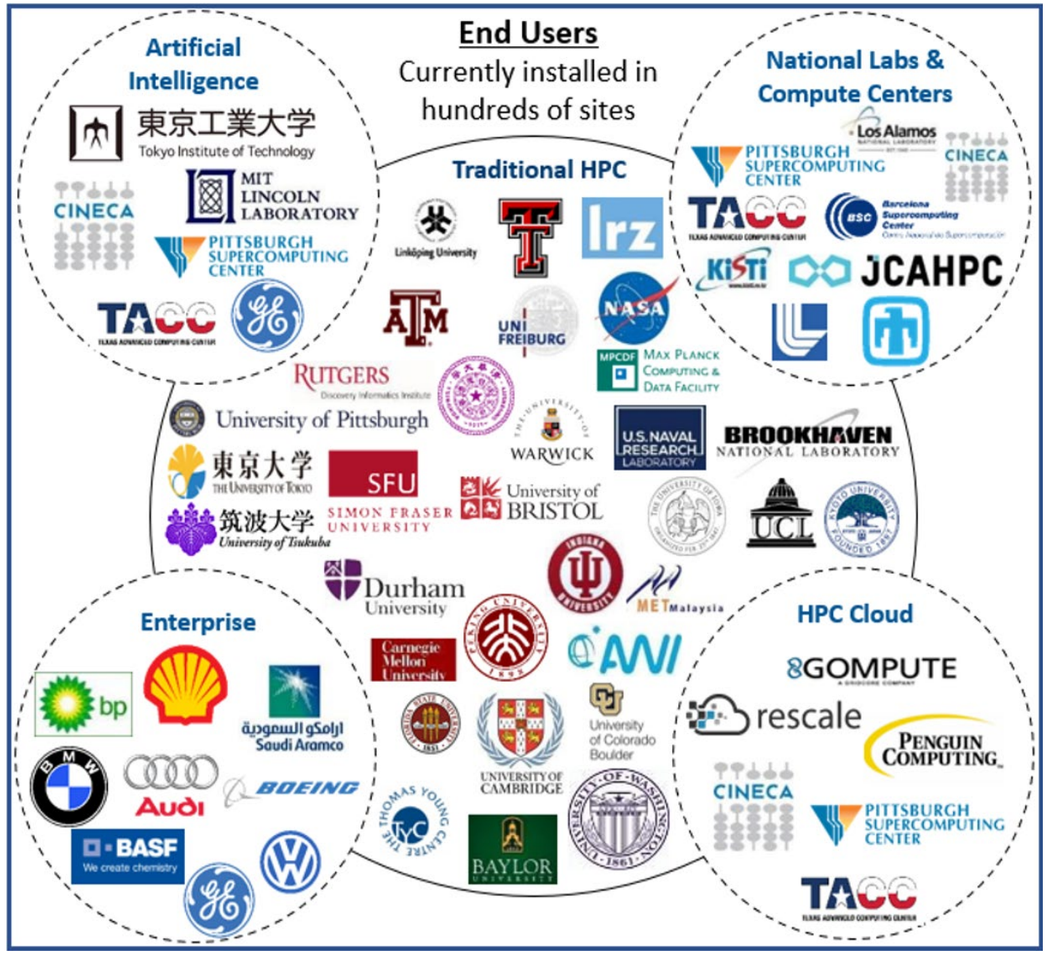
Thomas Steinke
NHR/ZIB, Head of Supercomputing

March 2022

CORNELIS™ NETWORKS

Cornelis Networks Overview

- ✓500+ deployments
- ✓In 42 of TOP500 Supercomputers
- ✓Across the largest government and academic research labs
- ✓Across all verticals
- ✓Across all HPC and AI applications



Introducing Omni-Path Express: Libfabric Provider

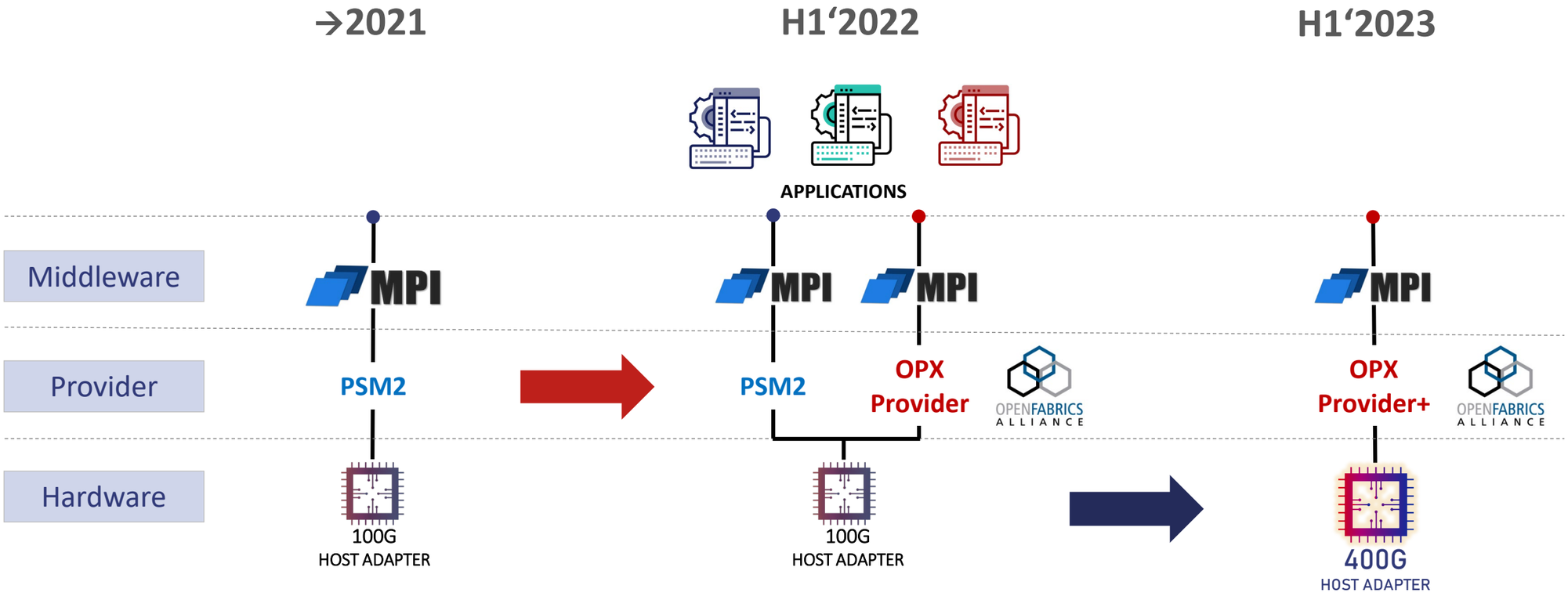
Enabling Dramatic Performance Improvements

- Optimized for high-performance converged infrastructures
  - Host architecture based on OpenFabrics Interfaces (OFI)
  - Access to industry standard frameworks and ongoing open-source development
  - Significant application performance gains resulting from accelerated fabric performance
    - Improved time-to-solution and return on investment
  - Foundational for next generation Omni-Path fabric architecture
    - Seamless transition to future Omni-Path platforms
  - Broad support coming for application-critical technologies
    - All popular MPIs, AI frameworks, Object Storage file systems like DAOS, and all popular GPUs

Broad Support Targeted by Omni-Path Libfabric Provider

Communication APIs	Storage	AI/ML	GPU
Open MPI	Intel	TensorFlow	NVIDIA
Intel MPI		PyTorch	Intel
MVAPICH			AMD
MPICH			
Sandia OpenSHMEM			
Charm++			
CHAPEL			
GASNet			

Omni-Path Evolution



New Software

Significant functionality and performance enhancements via

Libfabric over Omni-Path Express

New Hardware

Optimized performance via

Premiere OFI Adapter

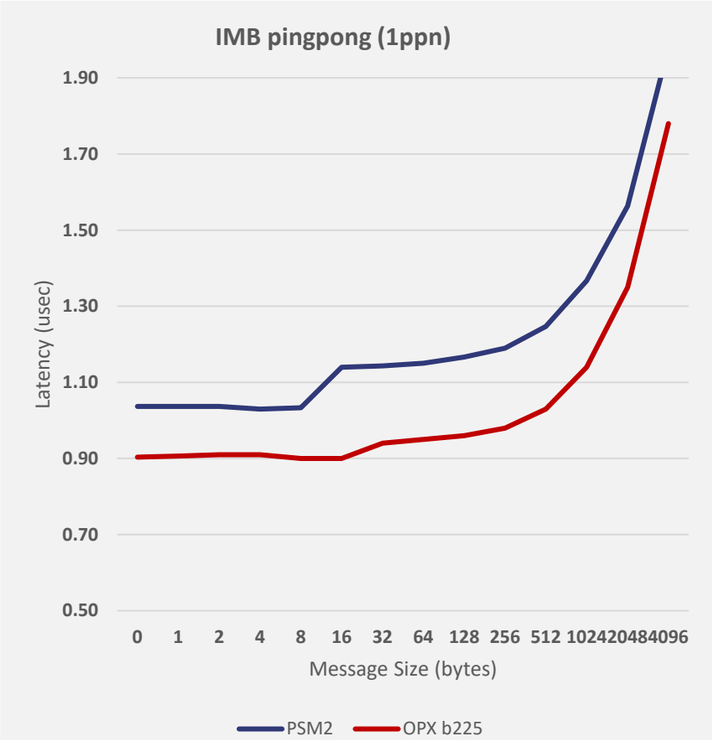
PSM: Performance Scaled Messaging
OPX: Libfabric over Omni-Path Express

**Simplification for concept illustration. Future features/options are subject to change without notice.*

Significant Performance Improvements

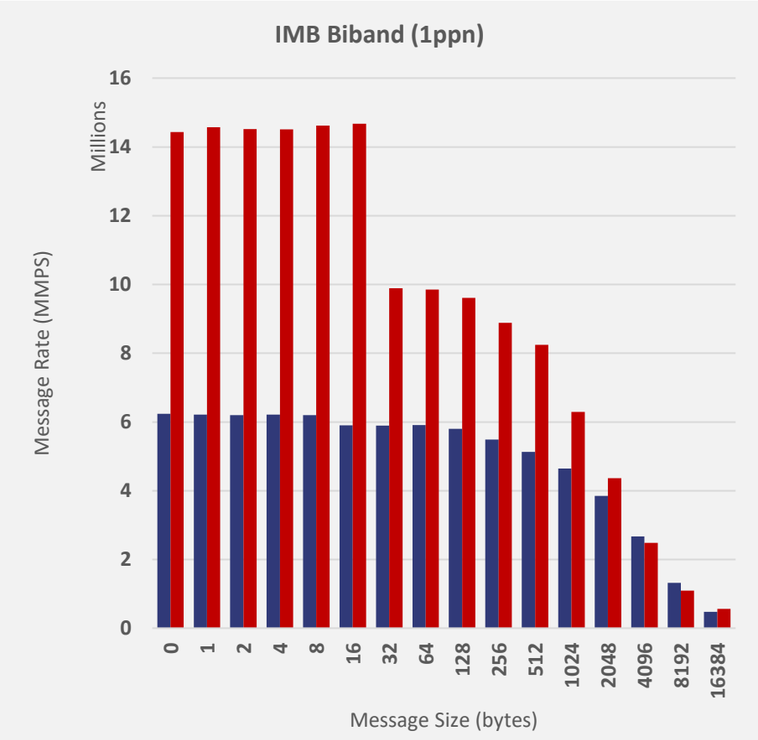
Latency

Up to 20% latency improvement



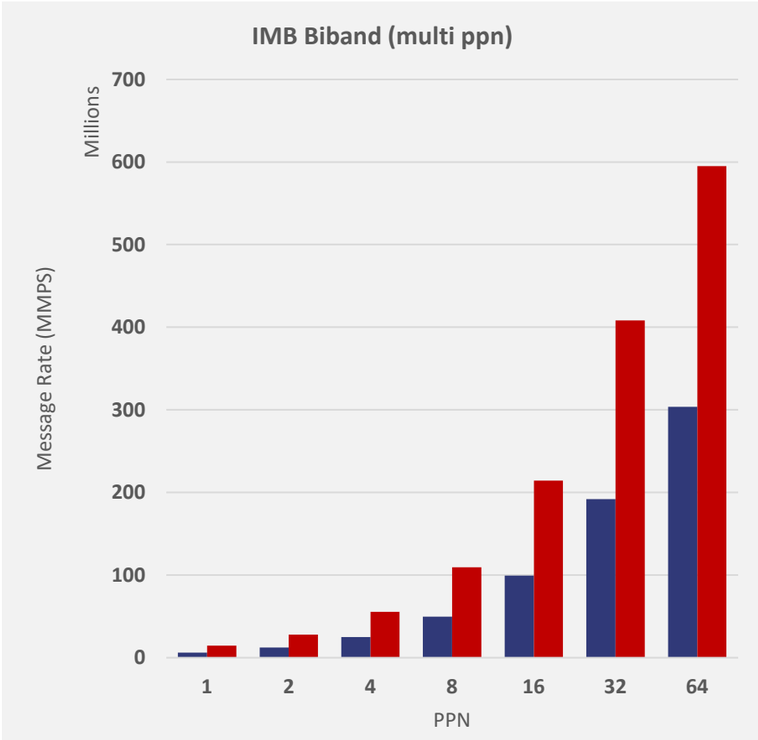
Message Rate

Up to 2.4X messages/sec per core



Scalability

Linear Scalability at double the throughput



PSM2 Provider OPX Provider

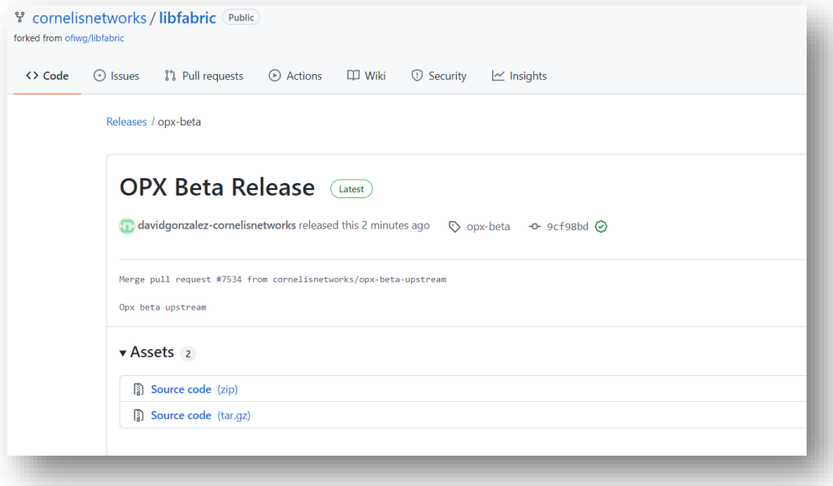
Test Configuration:
2-socket Intel® 3rd Generation Xeon® Scalable (Icelake) Platinum 8358, Dual Rail OPA100, BIOS: Snoop Hold-off Response Timer=11, Energy Efficient Turbo=DISABLED, C-States=DISABLED
Rocky Linux 8.4 (Green Obsidian), Kernel 4.18.0-305.19.1.el8_4.x86_64, IntelMPI 2019.6, IMB 2019.6, IFS 10.11.1.1.1, OPX Build 225

Beta Release & Download

- Validation criteria

Functional MPIs Tested	Supported CPUs	Supported OS
Intel MPI 2021.3 or later	Intel Skylake, Ice Lake, Cascade Lake (CLX-AP)	RHEL 8.5

- MPI microbenchmark testing
  - IMPI, IMB, and OSU: bandwidth test, message rate, and latency test
- Application testing
  - GROMACS, OpenFOAM, NAMD, FESOM2, VASP, Siesta, and PALM
- Upstreamed to Github
  - Link: https://github.com/cornelisnetworks/libfabric/releases/tag/opx-beta
  - Review Read Me and Application Notes to get started
  - Beta is now accepted by the OFI Libfabric community!

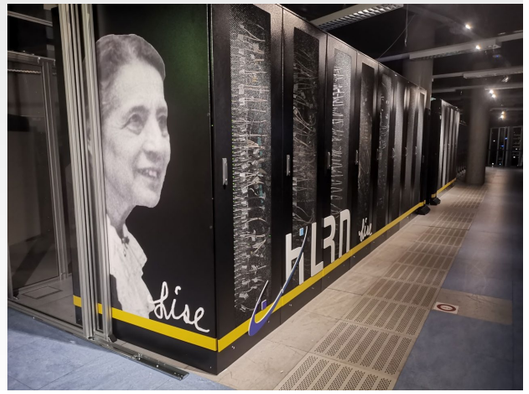


Evaluation of Omni-Path Express Software Stack

- Applied mathematical research
- Scientific services: HPC (national scale)

“Lise” System (Atos/Bull/Intel cluster)

- 1270 nodes with 2S Intel CLX-AP
- Omni-Path interconnect
- 10 PB Lustre storage
- 0.5 PB DAOS



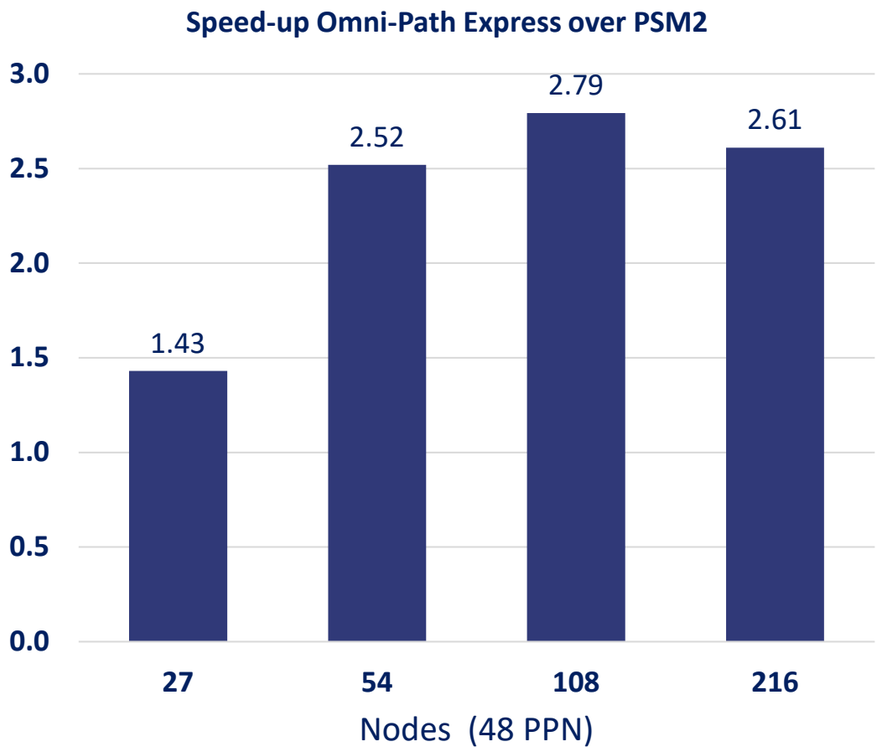
- ca. 200 projects
- ca. 4720 users

Omni-Path Express Evaluation

- Stability and scalability on 100+ nodes
- 6 Real-world applications +
  2 synthetic benchmarks
- Performance improvements of 10-20% for latency sensitive applications

Evaluation of Omni-Path Express Software Stack

OpenFOAM, potentialFOAM solver:
Speed-up Omni-Path Express over PSM2
(weak scaling experiment)



Summary

- Omni-Path Express delivers significant performance improvements leveraging native OFI framework
- Foundational hardware and software co-design innovation
- Fully engaged in open-source collaboration and interoperability
  - Linux, MPI Forum, and OpenFabrics Alliances
- Fully open-source messaging software stack
- Optimized for direct semantic match between MPI applications and OpenFabrics Interfaces
- Ease of access and broad support for communication libraries
- Customers gain significant improvements in…
  - Time-to-solution
  - Return on investment

Thank You

www.cornelisnetworks.com