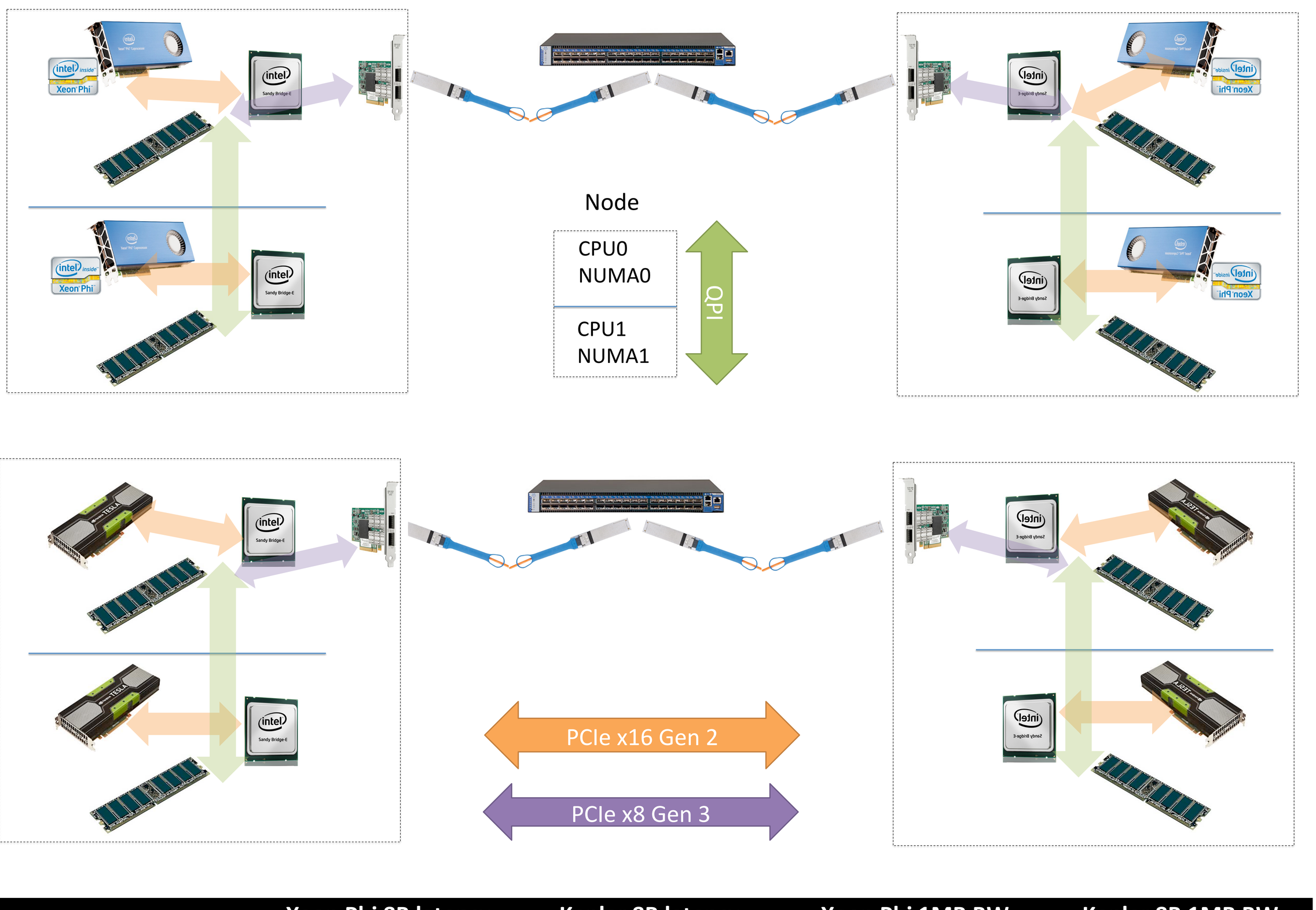
Auto-tuning MPI Execution Models on Clusters with Accelerator Devices and NUMA Memories

Sadaf R Alam, Swiss National Supercomputing Centre

Abstract—Symmetric Multiprocessing (SMP) aware MPI implementations on multi-socket, multi-core processors provide higher communication and scaling efficiencies for a wide range of payloads. Increasingly, these multi-socket and multi-core clusters incorporate accelerator devices such as Intel Xeon Phi and NVIDIA Kepler devices for higher performance and energy efficiencies. On these systems, MPI communication can be performed on data structures and pointers that reside on device memories, which are typically based on a different memory technology than the CPU or host memory. While an accelerator aware MPI implementation could offer opportunities to achieve higher code productivity by eliminating the need for explicit data transfers between the host and device memories, achieving high performance productivity on these systems is still work in progress. Since a number system configuration parameters can significantly influence MPI performance, both at inter and intra node levels, we attempt to device an auto-tuning methodology for application developers. This poster will present early micro-benchmark results demonstrating how our methodology can be applied to systems with NUMA memories and contemporary accelerator devices.



Execution model	Xeon Phi 8B latency (usec)	Kepler 8B latency (usec)	Xeon Phi 1MB BW (GB/s)	Kepler 8B 1MB BW (GB/s)
Inter-node Host-Host	1.09	1.09	6.4	6.4
Intra-node Host-Device	4.1	9.11	5.1	4.95
Inter-node Host-Device	7.9	16.5	0.4	5.7
Intra-node Device-Device	3.4	15.8	2.2	5.8
Inter-node Device-Device	9.2	18.08	0.3	5.7

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre