The HPE Slingshot 400 Expedition

Duncan Roweth
Slingshot Chief Architect
HPC/AI BU, HPE, Inc
Bristol, UK
duncan.roweth@hpe.com

Greg Faanes
VP of Advanced Technology
HPC/AI BU, HPE, Inc
Chippewa Falls, WI, USA
gregory.faanes@hpe.com

Bob Alverson
Advanced Technology
HPC/AI BU, HPE, Inc.
Seattle, WA, USA
robert.alverson@hpe.com

Houfar Azgomi
Product Management
HPC/AI BU, HPE, Inc.
San Jose, CA, USA
houfar.azgomi@hpe.com

Forest Godfrey
Advanced Technology
HPC/AI BU, HPE, Inc
Bloomington, MN, USA
aaron.godfrey@hpe.com

Marten Terpstra
Product Management
HPC/AI BU, HPE, Inc.
Vero Beach, FL, USA
marten.terpstra@hpe.com

Abstract— **HPE Slingshot 400 is a high-performance interconnect for HPC and AI supercomputing clusters. As the successor of HPE Slingshot, it comprises a PCIe Gen5 NIC and a 64-port switch, connected using standard 400 Gbps Ethernet physical interfaces, and enabling dragonfly and fat-tree networks with over 250,000 endpoints. HPE Slingshot is currently [deployed in 7 of the 10 largest supercomputers worldwide and dominates the top 3 list as the interconnect solution for El Capitan, Frontier, and Aurora machines. The Slingshot Transport (ST) protocol has become the cornerstone for HPC-optimized Ethernet networking standardization efforts led by the Ultra Ethernet Consortium (UEC). HPE Slingshot 400 builds on the foundational adaptive routing and congestion management feature set that underwrites performance of the exascale systems, while doubling network bandwidth and adding significant enhancements: network overlays for security and cloud isolation; improved quality of service with additional traffic classes; and support for in-band network telemetry. Slingshot 400 is supported across HPE's portfolio of rack- and chassis-based supercomputing platforms including HPE Cray XD, HPE Cray EX, and the latest HPE Cray GX. This paper presents the key features and some early performance results for Slingshot 400 systems.**

Keywords— HPC network, MPI, Slingshot Cassini, Rosetta

I. Introduction

HPE Slingshot 400 is a modern, highly scalable, connectionless high-performance interconnect for high-performance computing (HPC) and artificial intelligence (AI) clusters that delivers industry leading performance, bandwidth, and low latency for HPC, AI/ML, and data analytics applications. Pioneered by the HPE Cray team over nine generations of supercomputing interconnect silicon, it brings together the best of HPC optimized fabrics with the ubiquity of standard Ethernet. Addressing the need to converge RDMA operations, MPI messages, and high-performance IP traffic over a single network, Slingshot 400 extends and improves first generation Slingshot features such as fine-grain adaptive routing, advanced congestion control [1], and quality-of-service (QoS) capabilities By leveraging standard IP/Ethernet technology and software, Slingshot 400 delivers cost-effective connectivity and interoperability with third-party NICs and switches, providing IP storage without gateway nodes, and a broad eco-system of software.

Slingshot fabrics provide flow-based congestion management, automatically detecting flows that cause congestion and regulating their injection. Network buffers are left free for non-congesting traffic – that would be the victim of congestion in other fabrics.

A new objective of the Slingshot 400 generation is to enable the secure integration of supercomputing resources into a larger cloud data center infrastructure using standard VLAN, VXLAN, and ACL implementations. This is achieved through use of a new packet engine that implements the P4[1] switch architecture [2].

Slingshot 400 based systems (Figure 1) are constructed from a switch ASIC, Rosetta-2, top-of-rack or blade switches using Rosetta-2; a network adapter ASIC Cassini-2, and NIC cards using Cassini-2; a mix of passive and active cables; and switch management software, together with NIC drivers and a Libfabric provider [3].

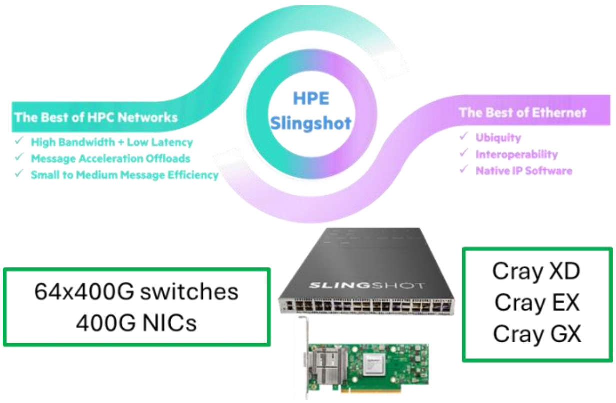

Figure 1: HPE Slingshot adopts 400 Gbps

II. Network Topologies and Cabling

HPE Cray EX systems use the dragonfly network topology, which provides cost effective global bandwidth at high scale. Slingshot 400 expands the range of supported topologies. Customers accustomed to clusters using fat-tree networks can select this topology – which can provide higher bisection bandwidth at higher cost. In high-end HPC systems, Slingshot 400 is being deployed with one NIC per CPU socket or one NIC per GPU. Where this level of injection bandwidth is not required, Rosetta-2 provides port-splitting. Each 4-lane switch port can provide one link at 400 Gbps, two links at 200 Gbps, or four links at 100 Gbps.

[1] P4 is shorthand for programming protocol-independent packet processors.

A. Dragonfly Topology

Slingshot 400 continues to support the dragonfly network topologies used in first generation systems. The dragonfly topology optimizes price performance of large systems while optical links are more expensive than electrical.

Low-cost electrical links are used for most of the connections – all the links within a Cray EX cabinet. Optical links, which are much more expensive, are reserved for the longer links between cabinets. Ports on every switch are configured either as downlinks (L0) that connect to the NICs, "intra-group" local links (L1) that connect to other switches in the dragonfly group, or "inter-group" global links (L2) that connect each group to all the other groups. Any endpoint can reach any other endpoint with a maximum of three switch-to-switch hops. The minimal path between any pair of nodes in a large system traverses up five links, four of which are electrical.

Figure 2 illustrates a small dragonfly network comprising six groups each with eight endpoints. Slingshot 400 supports up to 512 groups, each with 512 endpoints.

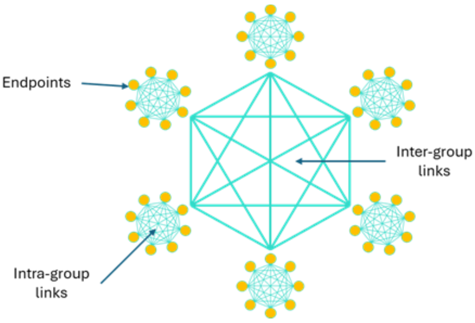

Figure 2: Dragonfly topology

Dragonfly group size is chosen based on factors such as the maximum system size, the desired scalable unit(s) for building and expanding the cluster, density and rack size impact to the economic tradeoffs between switches and optical cables, and preferences such as physical grouping, placement of resources, and cable routing policies.

A Slingshot 400 based system can scale up to over 256K physical endpoints with 16K switches, keeping a ratio of 16 endpoints per switch. For reference, the largest first-generation Slingshot system shipped, Aurora, has 10,624 compute blades and 2048 storage servers using 84,992 NICs connected by 5600 switches in 175 groups of 32 switches.

HPE Cray EX systems use up to eight NICs per node, with all the NICs connected to a single large rail. The single rail dragonfly design optimizes cost and performance.

B. Fat Tree Topology

In a Slingshot fat-tree or Clos topology every leaf switch can provide 32 endpoints. Each leaf switch is connected to 16 different spine switches using dual-density uplinks. A single Rosetta-2, radix-64, spine switch connects 32 leaf switches, 16 such spine switches provide a full bisection bandwidth network connecting 1,024 endpoints (see Figure 3).

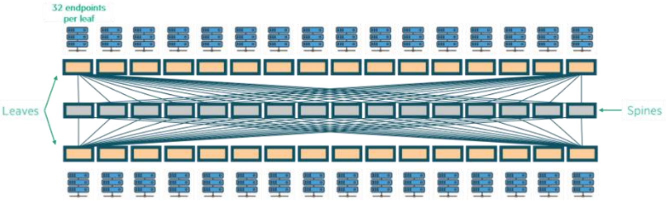

Figure 3: Two-tier fat-tree with 1K endpoints

The Slingshot team is defining additional options for larger leaf and spine switches, and multi-tier fat-trees to address requirements for 4K to 16K endpoints that are typical in AI/ML environments. Slingshot adaptive routing and congestion management operates over dragonfly or fat-tree networks.

C. Systems and cabling

The HPE Slingshot 400 portfolio comprises switch systems and NIC adapter cards for various HPE Cray platforms. Form factors used in Cray EX and XD systems are the same as those in the first-generation product. Slingshot 400 will also be available for the upcoming Cray GX systems, with switch blades, switch chassis, and NIC cards. These systems are cabled with a mix of passive and active cables, according to the link length. As data rates increase, the reach of passive electrical cables falls, and active cables become necessary for cables longer than 2 meters. Active optical cables are required for long paths between cabinets.

The Slingshot team designs to optimize for cost effective bandwidth. The switch chassis used in Cray GX systems brings the switch blades closer together, allowing all of the L1 links to be provided by a single wired electrical backplane, see Figure 4. Variants of this backplane connect 8, 16 or 32 switches in a GX cabinet.

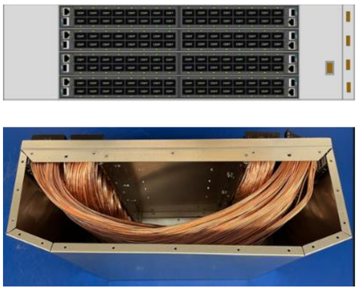

Figure 4: GX 8-slot switch chassis and wired midplane

OSFP connectors on the switch chassis provide connections to the NICs and global links between cabinets. The switch chassis is positioned mid-rack to maximize the number of NIC to switch links that can be cabled electrically.

Slingshot draws on the market for Ethernet standard optical links. While discrete transceivers have traditionally been more expensive than active optical cables, costs are converging. There are many benefits for transceivers including ease of maintenance, interoperability between networking gear with different connector types and simplified inventory management - as discrete transceivers can be deployed with passive fiber cables of different lengths. Serviceability is perhaps the most important factor – extracting faulty AOCs from a large system is difficult and sometimes impossible.

Slingshot 400 uses 800G SR8 transceivers and MPO16 fiber connectors with eight pairs of fibers. These fibers can be split out using a combination of passive fiber splitters and custom fiber patch panels to provide dual-density connector splitting, two- or four-way port splitting, or simply to manage pass-through links in an organized way. HPE has qualified transceivers from a variety of vendors and has deployed a proof-of-concept system using first generation parts. This option will be fully supported in Slingshot 400.

III. Rosetta-2 and Cassini-2 Specifications

HPE Slingshot 400 consists of switch systems built with the HPE-developed Rosetta-2 ASIC, and NIC adapter cards built with the HPE-developed Cassini-2 ASIC.

A. Rosetta-2

Rosetta-2 is a 64-port Ethernet switch, implemented as a single monolithic (800 mm2) ASIC fabricated in the TSMC 7nm FinFet process. All main switching logic utilizes a 1GHz clock, which results in a typical power dissipation of 190 watts and a maximum of around 300 watts.

All Rosetta-2 ports support IEEE 802.3 Ethernet standard signaling. Each port has 4 lanes operating at 100 or 50 Gbps using pulse amplitude modulated (PAM-4) signaling or 25 Gbps non-return to zero (NRZ) signaling to provide up to 400 Gbps Ethernet (IEEE 802.3ck). When used at the fabric edge, port splitting allows each of the 64 ports to provide 1,2, or 4 links. Port splitting is not used between switches. While providing an optimized, low latency, high throughput HPC infrastructure, Slingshot is fully Ethernet standards compliant and interoperable, operating as a converged HPC / Ethernet network. Rosetta-2 ports can be connected to a wide variety of third-party equipment. They have been qualified with devices from Aruba, Broadcom, and NVIDIA. Ethernet compliance testing is assured using Ixia testers.

When a Rosetta-2 switch is connected to another Rosetta-2 switch, they communicate using an enhanced fabric frame format, which provides additional control and status fields to support a multi-switch fabric. It is this fabric frame format that provides the unique features of Slingshot – adaptive routing, flow-based congestion management and credit flow control – within the fabric, while all externally facing ports operate using standard protocols.

Rosetta switches provide adaptive routing of all traffic. New paths are selected for ordered traffic on a flow-by-flow basis, each time an Ethernet flow, or a stream of small MPI messages enters the network. Long-lived flows are re-routed if they encounter mid-fabric congestion. Packet-by-packet adaptive routing is used for unordered traffic, e.g. the data packets of large RDMA transfers, the bulk of all traffic in an HPC/AI network. HPE made enhancements to the adaptive routing protocols used in Rosetta-2 based on experience from large first-generation systems.

Rosetta switches are unique in providing flow-based congestion management in hardware. Packets are classified into flows by their source and destination, together with a programmable selection of fields in both inner and outer headers. Packet headers are parsed on ingress and the selected flow steering and traffic class fields are used to generate a match value that becomes a signature for the flow. Different packets with the same match values belong to the same flow.

Separate queues are provided for each flow, 2048 of them per port in Rosetta-2, this allows each flow to make independent progress even in the presence of congestion. Each flow is separately managed and its contribution to fabric loading is measured.

In the default configuration flows are created at network ingress and cross the network to egress[1]. Ordered packets follow the path selected for the first packet in a flow. Acknowledgements take the return path from egress back to the source. These acknowledgements carry information on load and network congestion. Flows are created dynamically and torn down when their extent, the data flow pending between ingress and egress returns to zero. This data enables the ingress port to determine whether a flow is congested and whether that congestion is mid-fabric or at the endpoint. Where mid-fabric congestion is detected, a flow can be rerouted on a more lightly loaded path. Where endpoint congestion is detected, Slingshot networks back-pressure the individual flows causing congestion. This back-pressure is applied at ingress, regulating the flow to its fair share of egress bandwidth. Rosetta provides one more feature that is essential for high performance under load, injection limits. Each ingress port is allowed to inject enough data to sustain bandwidth but no more. In a 400 Gbps network that sustains a 3μs round-trip-time, each port is allowed to inject 150KB of data. Allowing more causes network buffers to fill, adding latency.

The RDMA workloads common in storage and AI applications play havoc with standard Ethernet networks, causing network wide congestion. Pause flow control (PFC) is not effective and operators opt to drop packets or their payload when congestion is encountered. Flow-based congestion management solves these problems, and RDMA workloads perform extremely well on Slingshot networks. PFC is used between the NIC and network ingress to regulate injection [2] . The novel adaptive routing and congestion management techniques mechanisms in Slingshot 400 fabrics operate with both Slingshot NICs and those from third parties.

The Slingshot supports QoS control / assurance for all network traffic. QoS can be used to provide performance isolation between different classes of traffic, independent of the congestion control mechanisms. Network traffic is tagged with a traffic class by the NIC or at fabric ingress port. Traffic shaping operates on the active classes to balance and optimize performance.

Slingshot 400 provides highly configurable end-to-end support for twelve network traffic classes, four more than the first-generation product. Each HPC/RDMA traffic class uses a pair of network classes, one for requests and one for responses. With twelve network classes Slingshot 400 can support four HPC/RDMA classes and four Ethernet classes. Administrators are provided with predefined traffic class policy sets, the *HPC* set, for example, supports the OFI Libfabric schema: *Low Latency*, *Dedicated Access*, *Best Effort* and *Bulk Data*. HPC applications and I/O services may be assigned to different traffic classes.

Each traffic class can be assigned assured bandwidth, guaranteed bandwidth, priority and forwarding preferences.

[1] Slingshot 400 hardware can forward traffic within a subnet (groups of switches) and route it between them. This capability is designed for use by a network operation system supporting multi-switch chassis.

[2] Rosetta switches use credit flow control on fabric links. HPE will be moving to use of credit flow control on NIC to switch links with Cassini-3.

A traffic class can exceed its assured bandwidth when other classes are dormant, and bandwidth is available. Guaranteed bandwidth is reserved for a class, it cannot be used by other classes. Priority operates amongst the classes that have assured bandwidth available. A traffic class for latency-sensitive traffic, for example, is given the highest priority, but with minimal assured bandwidth.

An application or service can use multiple traffic classes and where multiple classes are available it can select between them, switching to a low latency class for time critical collectives or queries for example. The administrator configures which classes are available to which user / job using the workload scheduler. For example, jobs in a given queue might be assigned to the best effort class, while those in another queue can use the Dedicated Access and Low Latency classes. I/O would normally be assigned to the Bulk Data class but can be given access to a Low Latency class for metadata operations.

B. Cassini-2 NIC

Cassini-2 is a custom ASIC designed for the Slingshot 400 network. Each chip provides two independent NICs. Use of two NICs per package is a cost and density optimization, Cray HPE Ex blades support up to 16 NICs so packaging density matters. Cassini NICs connect to the host using a 16x PCIe Gen5 host interface operating at signaling rates of 32 Gbps per lane. NICs connect to the network using a single link with 1, 2, or 4 lanes. This link supports IEEE 802.3 protocols at aggregate rates of up to 400Gbps in each direction. The NIC core is a faster implementation of the Cassini-1 design, ensuring a high degree of software compatibility, binary compatibility for user applications and workloads.

Cassini-2 provides the network endpoint for a converged fabric, with HPC/AI traffic sharing the same physical link as Ethernet and storage traffic. The NIC is designed to support a wide variety of use cases in HPC and AI:

- MPI point-to-point message passing
- One-sided remote memory access operations
- Ethernet packet processing
- RDMA access to storage

Applications perform inter-process communication using programing models that include MPI, NCCL, RCCL and SHMEM. The various programming models access Cassini-2 and hence the network using the Libfabric network API. Source for the Cassini provider is available from the Libfabric GitHub [2].

MPI message matching is performed in the NIC. Offload is provided for both the eager message case (small messages) and the rendezvous case (large messages). Matching in the NIC ensures progression, allowing effective overlap of communications and computation.

Remote memory access (RMA) operations include Put, Get, and atomic memory operations (AMO). A fast path is provided for simple RMA operations, those that create no target side state or in the case of AMOs operations, those that return an error if a response is lost in the network. Reliable AMOs are provided for applications that use AMOs in locking or work distribution.

Cassini-2 provides a reliable transport, protecting applications from any network errors. Its implementation is an evolution of the Cassini-1 design used and proven in the US exascale systems. A source-based retry mechanism is provided for idempotent operations. A stateful error detection and retry mechanism is used for ordered operations that manipulate target state; operations such as message matching and event generation that must be executed once and once only in the presence of errors. Hardware maintains the state necessary for retry, freeing software from the (heavy) burden of implementing reliability. Retry policy is implemented in software.

Ethernet packet processing is in three parts: efficient transmission on the sender side, receive side scaling (RSS) on the target side, and checksum computation/validation on both sides. Hardware computes IP checksums and the RoCE invariant CRC. HPE supercomputer systems use Lustre or DAOS based storage. Access to storage is provided using Libfabric and its kernel counterpart, kfabric. Use of RoCE is an option for Ethernet connected storage.

Cassini-2 NICs provide hardware support for container and VM-based virtualization. When combined with hardware implementation of VXLAN-based overlays in Rosetta-2 and focused software enablement, Slingshot 400 is well-positioned to deliver a fully integrated, virtualization-optimized solution for modern multi-tenant HPC and enterprise environments.

C. Slingshot 400 and 200 interoperability

Slingshot-400 supports the ability to carry ST traffic between fabrics, for example, between a compute fabric and a storage fabric. This new capability supports use of Libfabric/Kfabric based storage traffic between fabrics, allowing storage fabrics to be independently managed while ensuring performance end-to-end.

IV. PERFORMANCE MEASUREMENTS

Benchmark testing of Slingshot 400 used a 64 dual-socket AMD EPYC 7654 (Genoa) based nodes with a NIC per socket. A larger but older system (Huygens), with 512 dual-socket Intel Sapphire nodes is being used for scale testing. It is also equipped with a NIC per socket. Both systems run Cray OS, the Cray Programming Environment (CPE), and the open-source Libfabric provider.

A. Latency

MPI point to point latency was measured at 1.45µs for a NIC-switch-NIC network path. Additional network hops add 300ns per switch and 100ns or more per digital optical cable. The longest minimal path in a large system adds approximately 1µs.

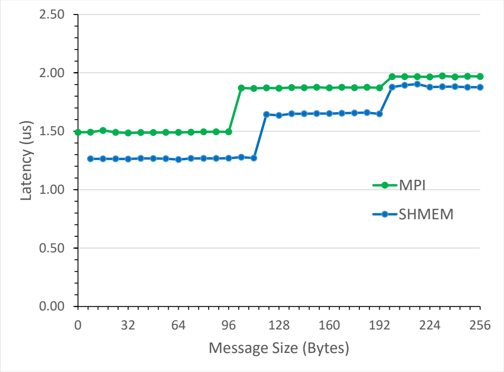

Figure 5: Slingshot 400 latency

Latency does depend on the CPU type and the internal design of the node – with latencies being higher in a socket that is remote from the NIC. It also depends on the programming model and transfer size as illustrated in Figure 5.

The Libfabric provider uses a low-latency path in which command and data are written directly to the NIC for small payloads, switching to DMA as payload size increases.

B. Bandwidth

Figure 6 illustrates MPI bandwidth achieved between a pair of nodes using the 200 Gbps and 400 Gbps generations of Slingshot. Tests use 32 processes per node. The figure includes the message size at which half bandwidth is achieved, a little over 256 bytes for Slingshot 400.

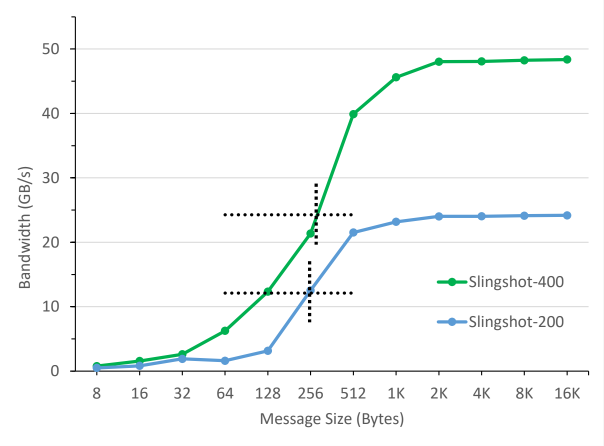

Figure 6: MPI bandwidth for increasing message size

As network bandwidths increase, performance of the hosts' on-chip network can become a critical factor; not all CPU types perform as well as the AMD Genoa CPUs used in these tests. A poorly performing host generates back-pressure on the network and can cause congestion – users can see this via counters in Cassini-2 that report cycles blocked waiting for writes to the host to complete.

The performance of quiet network latency and bandwidth tests is interesting, but not that relevant to application performance on a loaded network. Figure 7 shows bandwidth achieved on MPI bisection and all-to-all tests operating over the whole of our 64-node AMD test system. This system has a Cassini-2 NIC per socket. Measured performance is between 85% and 95% of line rate.

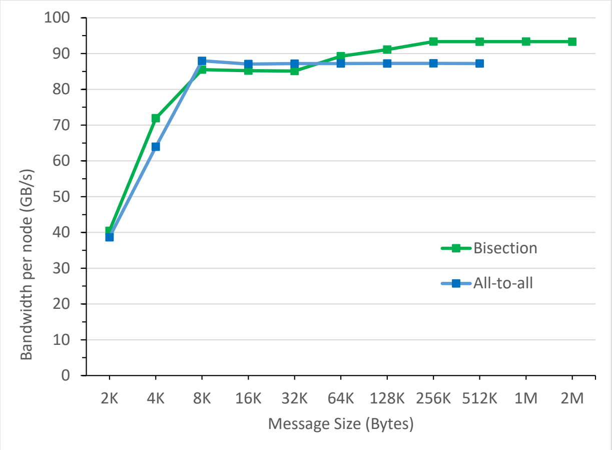

Figure 7: Per-node bandwidth on MPI all-to-all and bisection tests

Recent tests on an early customer system show these high levels of performance being achieved at much larger system size. Testing of nodes with 1600 Gbps of injection bandwidth has just started.

C. Message rate

The rate at which the NIC can perform small one-sided operations (Put, Get and AMO) is important for SHMEM applications, as is the ability to sustain high performance when these operations perform global uniform-random memory access patterns. Figure 8 illustrates the rate at which a Cassini-2 NIC can perform such operations. Our benchmark codes generate one-to-one (symmetric) traffic with asymptotic rates of 200-250 million operations/sec. These tests are limited by the performance of the CPU, its host interface and memory system. Our tests also generate many-to-one and one-to-many traffic, allowing us to investigate where the bottlenecks lie. These tests show that operation rates on symmetric traffic are limited by the CPU's PCIe write performance. Read limited tests achieve higher performance, e.g. asymptotic rates of 465M Puts/sec on 1-to-4 traffic and 450M Gets/sec on 4-to-1 traffic. The performance of the AMD CPUs used in our test system is markedly better than that of other CPUs on this metric.

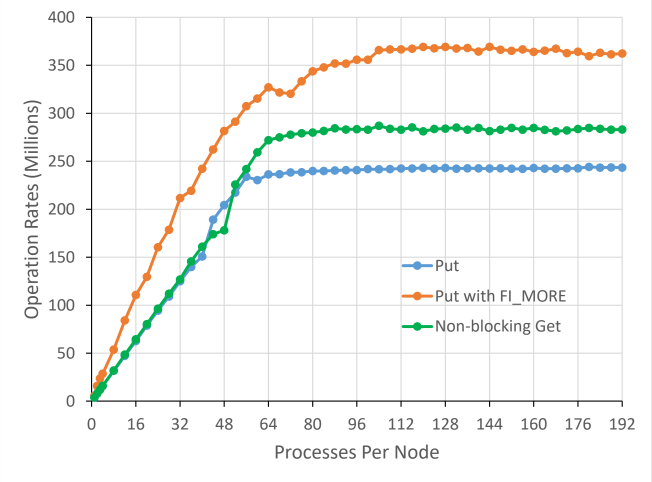

Figure 8: Slingshot 400 SHMEM operation rates

Cassini-2 provides greater flexibility in its use of relaxed ordering on the host interface. This boosts the performance of non-blocking Gets to that of non-blocking Puts.

Figure 8 also shows the benefit of using the Libfabric FI_MORE option. This option allows multiple commands to be issued to the NIC at the same time. Its use amortizes the (high) cost of the CPU's SFENCE. The FI_MORE option is most valuable when upper layer software has multiple operations to perform between synchronization points.

D. Ethernet performance

Figure 9 illustrates Ethernet performance of the Slingshot 400 test system. Throughput measurements were made with the iperf benchmark. Tests use increasing numbers of sending threads. Throughput of 40 Gbps per thread/core is higher than expected. This reflects effective distribution of interrupts and packet processing workload over the cores. Cassini-2 uses the same internal structures to process Ethernet packets as are used for MPI messages, enabling the Ethernet driver to achieve good NUMA/core locality.

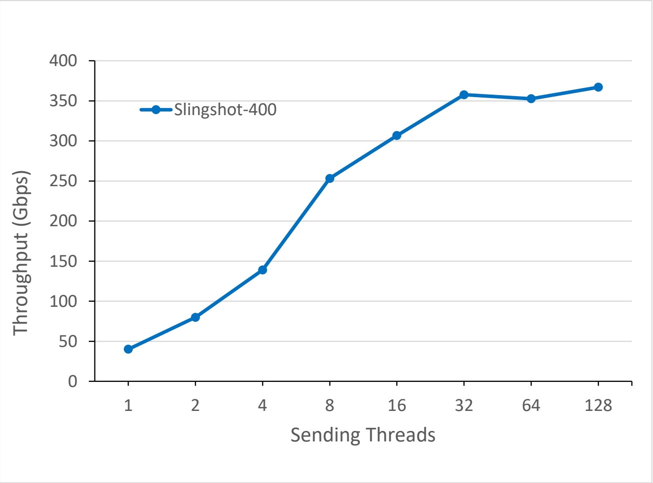

Figure 9: Slingshot 400 Ethernet performance

E. Comparison with NDR IB

The Slingshot team is occasionally asked to compare performance of our network with that of InfiniBand. Our preference is to do this using real-world application/system tests that show the benefits to the end user. Figure 10 provides results of one such test, a snapshot of FFTW performance, comparing performance of an NDR InfiniBand in a full-bandwidth Fat Tree configuration with that of an HPE Slingshot 400 full-bandwidth dragonfly configuration. Both systems use the same AMD EPYC 7654 processor and memory. Benchmarking was conducted with Cray MPI and Open MPI on both systems, with MPI rendezvous protocol on or off on both systems, and with local shared memory optimizations on or off on both systems. Data points in the figure record the highest performance measured for any combination of the options at each of the job sizes.

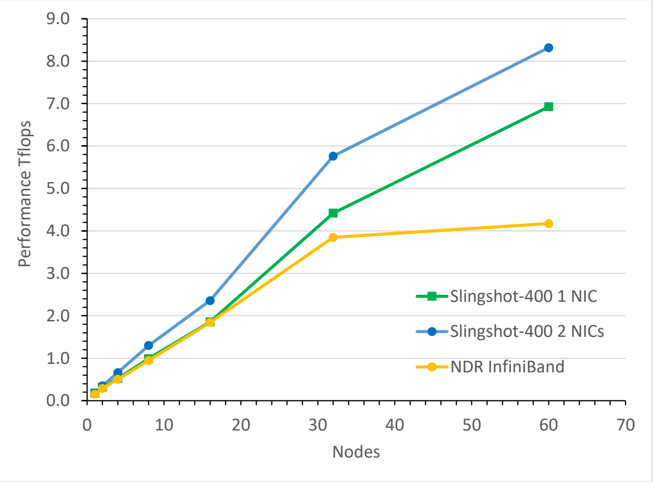

Figure 10: FFTW performance

Slingshot 400 outperforms NDR InfiniBand at 32 nodes and above. There is a notable gain in application performance (30%) with two Slingshot 400 NICs, one in each socket. Cost savings made through use of dragonfly contribute to better performance achieved through using two NICs per node.

V. Conclusions

The first-generation Slingshot network is in widespread use, with HPE Slingshot based systems occupying seven of the top ten positions in the Top500. Slingshot 400 builds on this base, increasing performance and adding new function. Use of Ethernet physicals, link layer, and protocols ensure that Slingshot can use and interoperate with a wide variety of standards compliant network components. Early performance tests demonstrate that Slingshot 400 achieves double the bandwidth of the first-generation product on a range of metrics. Latency improvements are in line with the modest increase in clock speed.

Slingshot runs a proprietary transport over this network, ST. This was essential for exascale as other Ethernet transports don't scale. Some of the ideas that Slingshot pioneered are being standardized by the Ultra Ethernet Consortium [4], use of link level retry and credit flow control for example. UEC is also developing a new transport (UET) that builds on the ideas of ST to provide connection-free RDMA between NICs from multiple vendors. HPE will start introducing UEC compliance in the Slingshot 800 generation. The congestion management and adaptive routing features that underwrite performance of Slingshot 400 are not part of UEC, they will remain negotiated features that HPE devices enable when connected to each other.

HPE is investing heavily in its own network IP. The Slingshot 800 ASICs will tape out as Slingshot 400 based systems move to general availability. Specification and architecture development is in progress for the fourth-generation product, with links operating at 1600 Gbps.

References

[1] The original HPE Slingshot white paper provides details on the benefits of this revolutionary Ethernet-based interconnect solution: CUG 2022: HPE Slingshot Launched into Networking Space.

[2] For details of the programming protocol-independent packet processors (P4) see the P4 website

[3] The Cassini Libfabric provider is available from GitHub

[4] For information on Ultra Ethernet is available from the UEC website