OpenFOAM
Performance Testing and Profiling

October 2017

Note

- **The following research was performed under the HPC Advisory Council activities**
  - Participating vendors: Huawei, Mellanox
  - Compute resource - HPC Advisory Council Cluster Center

- **The following was done to provide best practices**
  - OpenFOAM performance overview
  - Understanding OpenFOAM communication patterns
  - Ways to increase OpenFOAM productivity
  - MPI libraries comparisons

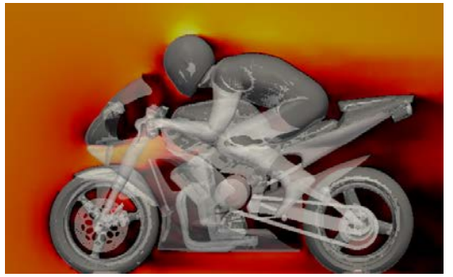
- **For more info please refer to**
  - http://www.huawei.com
  - http://www.mellanox.com
  - https://www.openfoam.com/

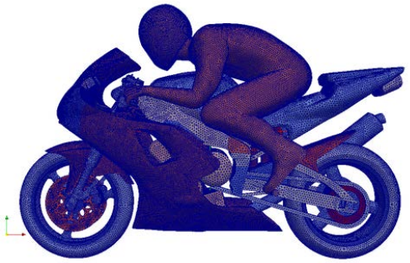
OpenFOAM®

OpenFOAM

- **OpenFOAM® (Open Field Operation and Manipulation) CFD**

- **Toolbox in an open source CFD applications that can simulate**
  - Complex fluid flows involving
  - Chemical reactions
  - Turbulence
  - Heat transfer
  - Solid dynamics
  - Electromagnetics
  - The pricing of financial options





- **OpenFOAM support can be obtained from OpenCFD Ltd**

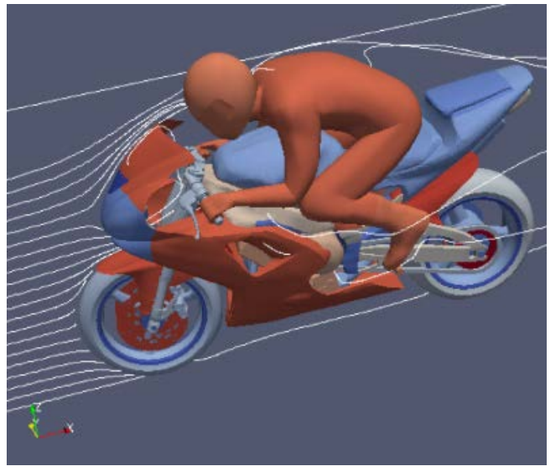
Objectives

- **The presented research was done to provide best practices**
  - OpenFOAM performance benchmarking
    - MPI Library performance comparison
    - Interconnect performance comparison
    - Compilers comparison
    - Optimization tuning

- **The presented results will demonstrate**
  - The scalability of the compute environment/application
  - Considerations for higher productivity and efficiency

Test Cluster Configuration

- **Huawei FusionServer E9000 with FusionServer CH121 V5 16-node (640-core) “Skylake” cluster**
  - Dual-Socket 20-Core Intel Xeon Gold 6138 @ 2.00 GHz CPUs (27.5MB L3 Cache, Turbo @3.70 GHz)
  - Dual-Socket 18-core Intel Xeon Gold 6140 @ 2.30 GHz CPUs (24.75MB L3 Cache, Turbo @3.70 GHz)
  - Dual-Socket 20-core Intel Xeon Gold 6148 @ 2.40 GHz CPUs (27.5MB L3 Cache, Turbo @3.70 GHz)
  - Memory: 192GB memory, DDR4 2666 MHz RDIMMs per node
  - OS: RHEL 7.3, MLNX_OFED_LINUX-4.1-1.0.2.0 InfiniBand SW stack
- **Mellanox ConnectX-4 and ConnectX-5 EDR 100Gb/s InfiniBand Adapters**
- **Mellanox Switch-IB SB7800 36-port EDR 100Gb/s InfiniBand Switch**
- **Huawei OceanStor 9000 Scale-out NAS storage system**
- **Compilers: Intel Parallel Studio XE 2018**
- **MPI: Intel MPI 2018, Mellanox HPC-X MPI Toolkit v1.9.7**
- **Application: OpenFOAM v1612+, single precision**
- **Benchmarks: MotorBike, 160K elements, 100 steps**



Introducing Huawei FusionServer E9000 – CH121 V5

High-Performance 2-Socket Blade Unlocks Supreme Computing Power

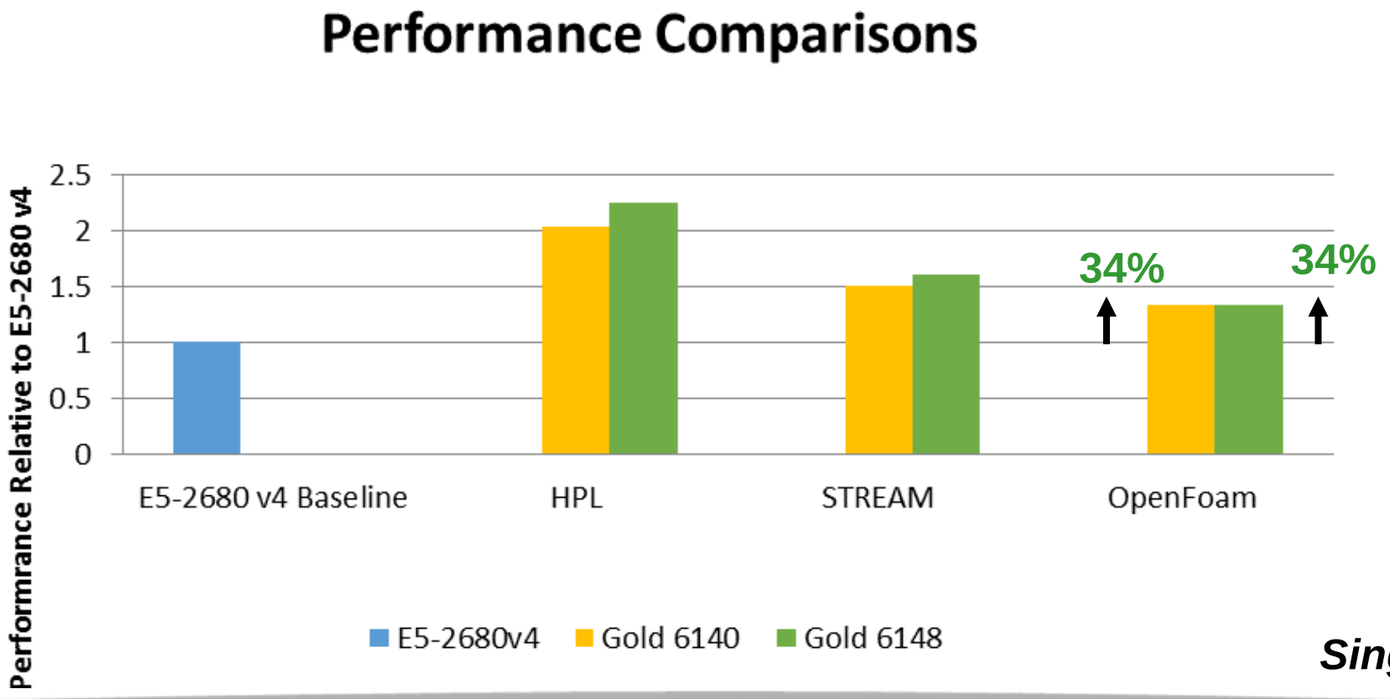
FusionServer

HUAWEI

Full-series Intel® Xeon® Scalable Processors, **24** DDR4 DIMMs, **AEP memory supported, 1** PCIe slot, 2 SFF/2 NVMe SSDs/4 **M.2 SSDs high-performance storage, multi-plane network**, LOM supported

OpenFOAM Performance – CPU SKUs and Generation

- **OpenFOAM performance gain by larger core counts and better memory throughput**
  - "Gold 6140" demonstrates a 34% of performance gain (29% more cores) vs E5-2680v4
  - "Gold 6148" demonstrates a 34% of performance gain (42% more cores) vs E5-2680v4
  - Base clock are the same on E5-2680 v4 and Gold 6148, while Gold 6140 runs slightly slower
  - Skylake supports 6 memory channels and faster DIMMs which impacts on memory performance

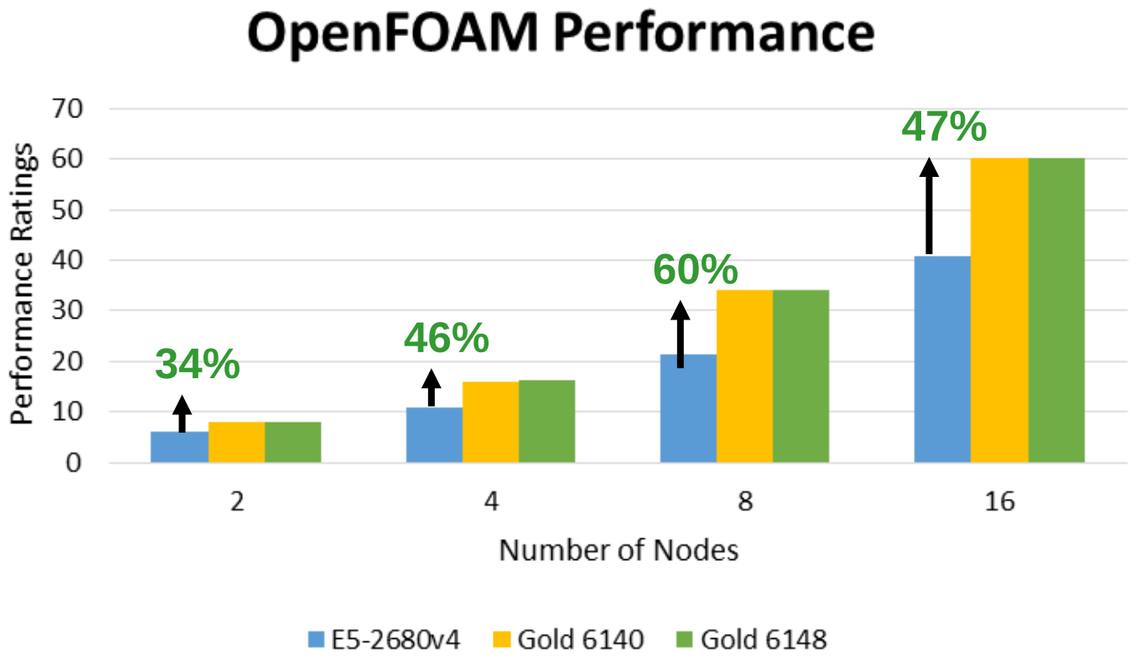


Higher is better

Single Node Performance

OpenFOAM Performance – CPU SKUs and Generation

- **Performance benefits of Skylake CPU grows as cluster scales (on per-node basis)**
  - Performance gain of ~60% by “Skylake” CPUs compared to “Broadwell” CPUs
    - Difference may be mainly due to additional cores, and newer CPU architecture available in Skylake
  - No difference in using “Gold 6140” versus “Gold 6148”, despite slightly higher clock and more cores
    - Gold 6148 has 2 more CPU cores, and slight increase in CPU base clock
    - Possible reason may be memory bandwidth saturation, and turbo clock are same for both CPUs

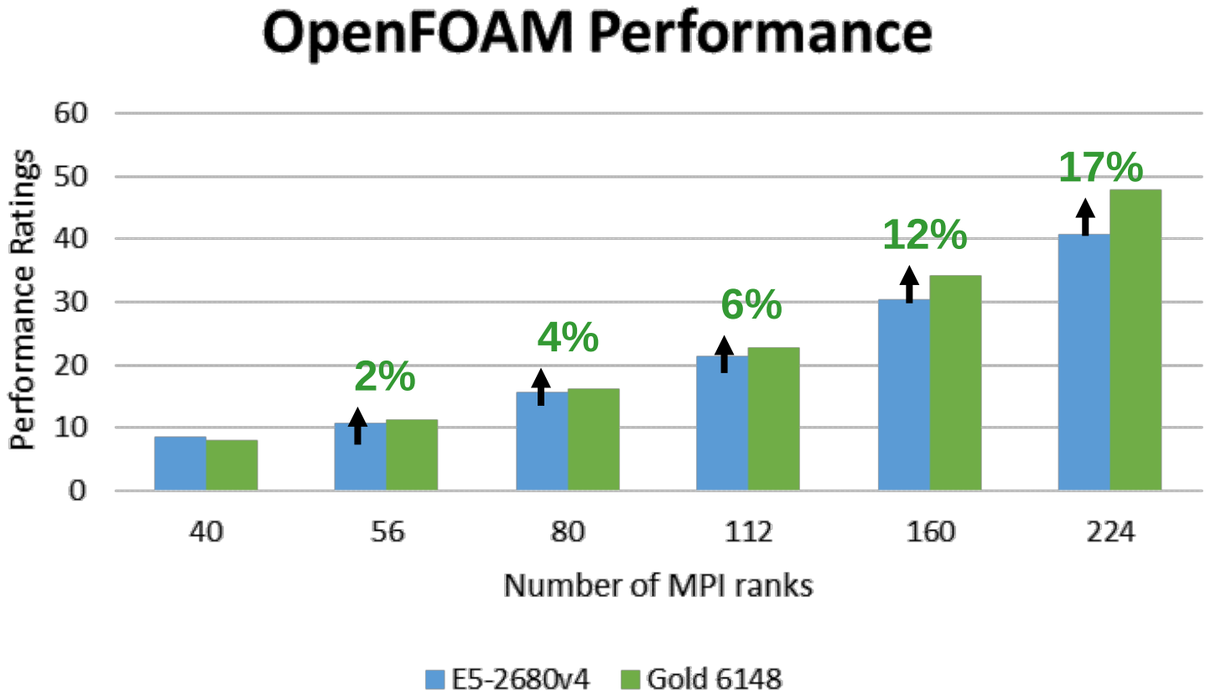


Higher is better

40 MPI Processes / Node

OpenFOAM Performance – CPU SKUs and Generation

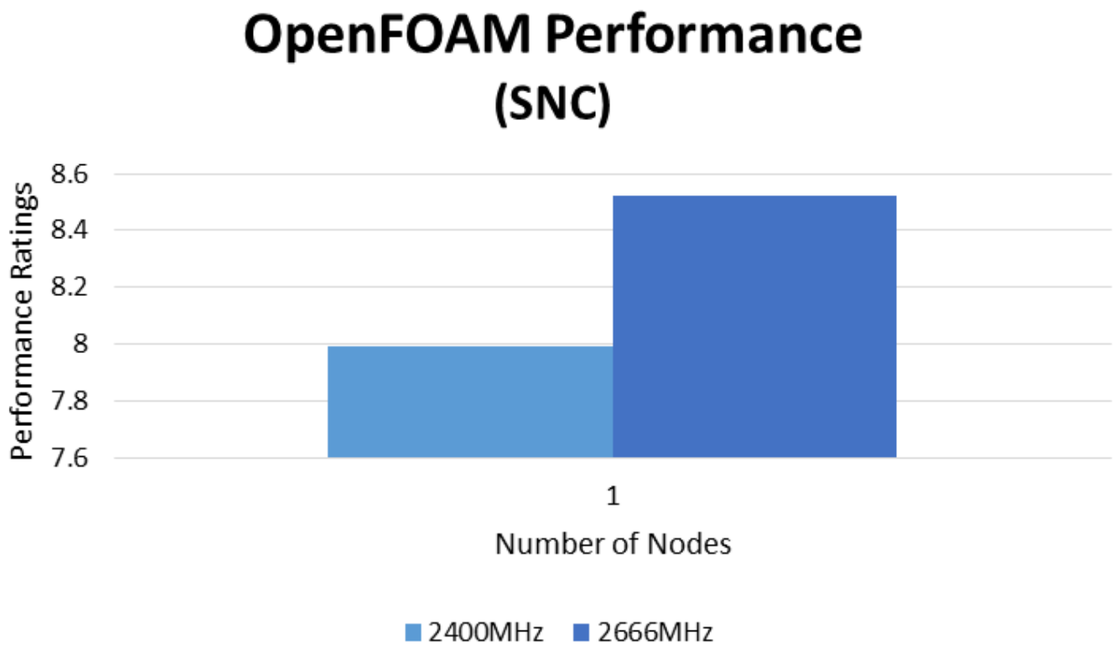
- **Performance benefits of Skylake CPU grows as cluster scales (on by-core basis)**
  - Performance is even between “Skylake” CPUs and “Broadwell” CPUs at 40 cores
  - Performance gain become much more apparent at scale
  - CPU Information:
    - Gold 6148: Dual Socket 20-core Intel Xeon 6148 @ 2.4GHz (Turbo @ 3.7GHz)
    - E5-2680v4: Dual Socket 14-core Intel Xeon E5-2680v4 @2.4GHz (Turbo @ 3.3GHz)



Higher is better

OpenFOAM Performance – Memory Speed

- **Memory speed provides some benefits to OpenFOAM performance**
  - Skylake platform supports DIMM speed up to 2666MHz DIMMs
  - 2666MHz DIMMs is theoretically ~11% faster than the 2400MHz DIMMs
  - OpenFOAM reports only about ~5% of the improvement on a single node
  - Only part of the benefits in speed is translated into performance gain for OpenFOAM

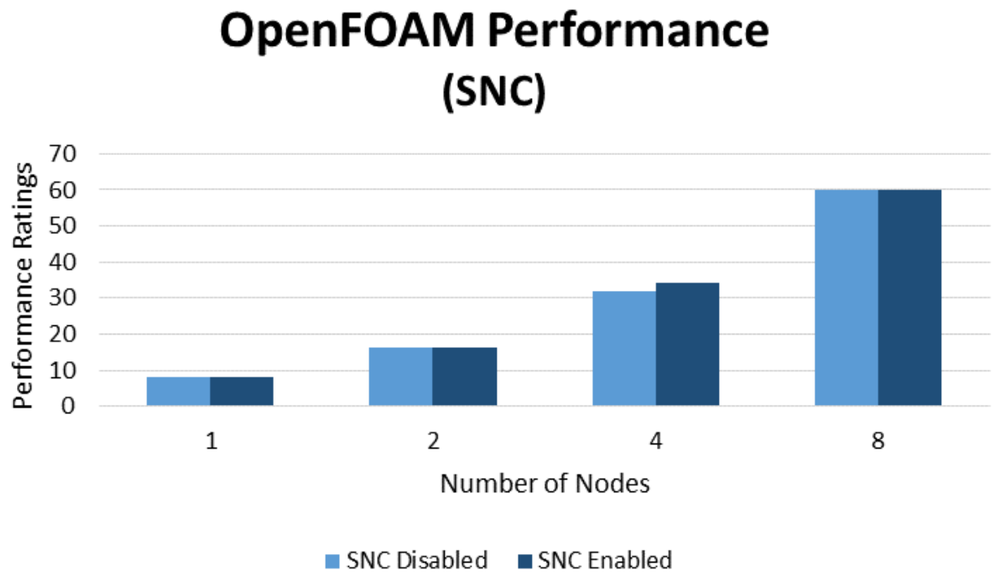


Higher is better

40 MPI Processes / Node

OpenFOAM Performance – Sub-NUMA Clustering

- **Enabling SNC provides marginal benefits for OpenFOAM**
  - Sub-NUMA Clustering (SNC) is similar to a cluster-on-die (COD) in Haswell/Broadwell generation
  - CPU cores and memory would be split into 2 separate NUMA domains when SNC is enabled
  - SNC generally should demonstrate some benefits for applications that requires good NUMA locality
  - SNC only demonstrates small marginal gain when SNC is enabled

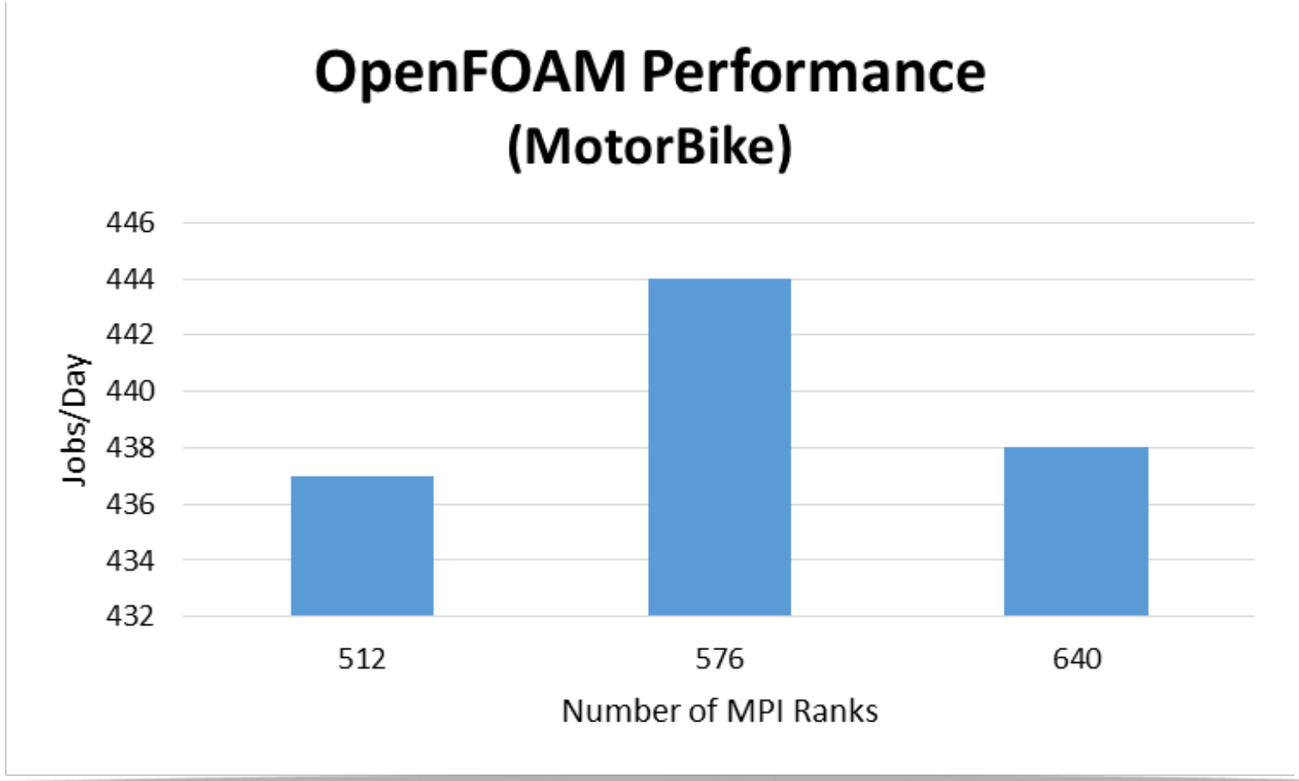


Higher is better

40 MPI Processes / Node

OpenFOAM Performance – Processes Per Node

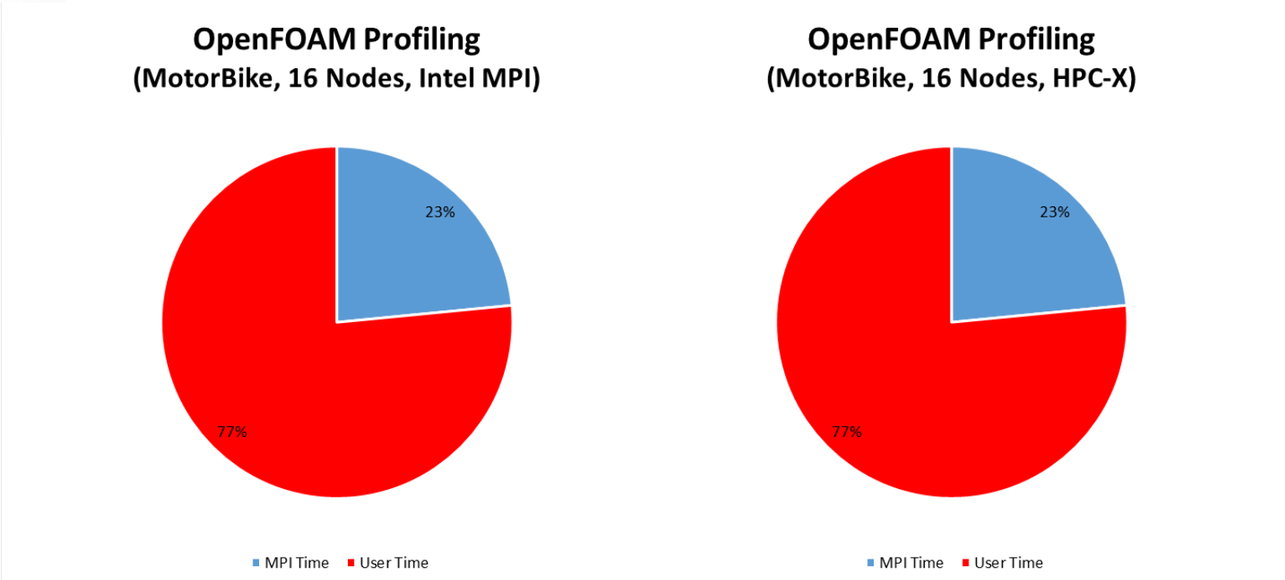
- **Observed best performance with using less CPU cores per node**
  - Some benefits by using 36 cores per node on a Gold 6138; compared to 40 or 32 PPN
  - Dual-socket “Gold 6138” provide up to 40 per node
  - Potentially due to memory bandwidth saturation for the number of Skylake cores

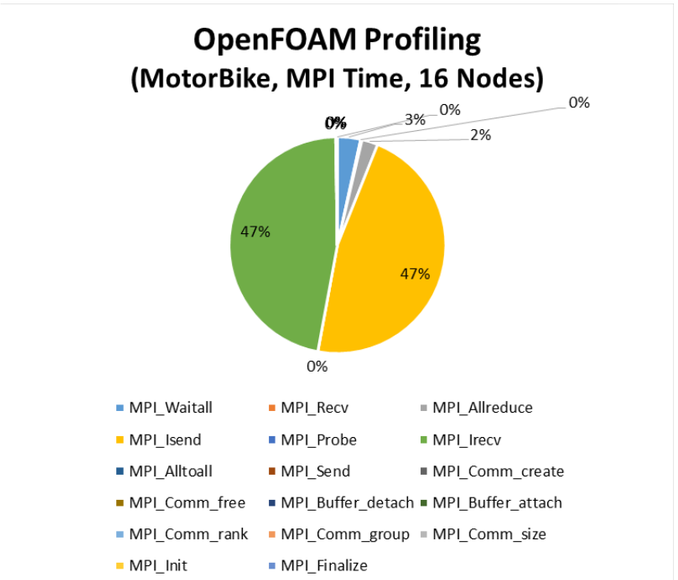


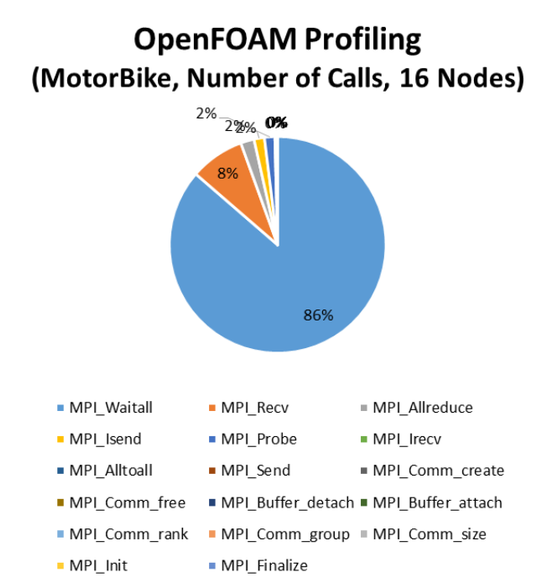
Gold 6138

OpenFOAM Profiling – MPI/User Time Ratio

- **OpenFOAM simpleFOAM solver uses mainly non-blocking communications**
  - 23% of overall runtime spent on MPI communication at 16 nodes / 640 MPI cores
  - Both Intel MPI and HPC-X spent the same time in overall runtime on MPI communications
  - Overall of MPI time spent in MPI non-blocking communications (MPI_Waitall 47%, MPI_Isend, 47%)
  - Most of the MPI calls made by OpenFOAM are MPI_Waitall







OpenFOAM Summary

- **OpenFOAM performance gain by larger core counts and better memory throughput**
  - "Gold 6140" demonstrates a 34% of performance gain (29% more cores) vs E5-2680v4 (on 1-node)
  - "Gold 6148" demonstrates a 34% of performance gain (42% more cores) vs E5-2680v4 (on 1-node)
  - Performance gain of ~60% by "Skylake" CPUs compared to "Broadwell" CPUs (on multi-node)
  - No difference in using "Gold 6140" versus "Gold 6148", despite slightly higher clock and more cores
- **Effect on Skylake generation on OpenFOAM performance**
  - Provides substantial performance gain due to the larger core count, support for memory channels
  - Faster 2666MHz DIMM (compares to 2400MHz) translates to increase of 5% in performance
- **Effect on SNC (Sub-NUMA Clustering) on performance**
  - Enabling Sub-NUMA Clustering provides little/marginal benefits
- **Observed best performance is using less cores than available per node**
  - Slight benefits by using 36 CPU cores per node; compared to 40 or 32 PPN

Thank You

HPC Advisory Council