Insight into Intel innovations that will bring enhancement for HPC applications

Andrea Luiselli
Technical Solution Specialist – HPC
Intel Corporation
Acknowledgements:

Andrey Ovsyannikov, Nitya Hariharan, Chris Allison, Stephen Van Doren, Johann Lombardi
Notices & Disclaimers

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit http://www.intel.com/benchmarks.

Optimization Notice: Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice. Notice Revision #2011080

Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase. For more complete information about performance and benchmark results, visit www.intel.com/benchmarks.

Intel technologies’ features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration.

Intel® Advanced Vector Extensions (Intel® AVX)* provides higher throughput to certain processor operations. Due to varying processor power characteristics, utilizing AVX instructions may cause a) some parts to operate at less than the rated frequency and b) some parts with Intel® Turbo Boost Technology 2.0 to not achieve any or maximum turbo frequencies. Performance varies depending on hardware, software, and system configuration and you can learn more at http://www.intel.com/go/turbo.

Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

© 2019 Intel Corporation.
Intel, the Intel logo, and Intel Xeon are trademarks of Intel Corporation in the U.S. and/or other countries.

*Other names and brands may be claimed as property of others.
### NWP and ESM HPC centers in Top500

**From TOP500 List - June 2018**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Site</th>
<th>System</th>
<th>Cores</th>
<th>Rmax (TFlop/s)</th>
<th>Rpeak (TFlop/s)</th>
<th>Power (kW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>United Kingdom Meteorological Office</td>
<td>Cray XC40, Xeon ES-2695v4 18C 2.1GHz, Aries interconnect Cray Inc.</td>
<td>241,920</td>
<td>7,038.9</td>
<td>8,128.5</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Japan Meteorological Agency</td>
<td>Cray XC50, Xeon Platinum 8160 24C 2.1GHz, Aries interconnect Cray Inc./Hitachi</td>
<td>135,792</td>
<td>5,730.5</td>
<td>9,125.2</td>
<td>1,354</td>
</tr>
<tr>
<td>26</td>
<td>Japan Meteorological Agency</td>
<td>Cray XC50, Xeon Platinum 8160 24C 2.1GHz, Aries interconnect Cray Inc./Hitachi</td>
<td>135,792</td>
<td>5,730.5</td>
<td>9,125.2</td>
<td>1,354</td>
</tr>
<tr>
<td>31</td>
<td>National Center for Atmospheric Research (NCAR)</td>
<td>Cheyenne - SGI ICE XA, Xeon E5-2697v4 18C 2.3GHz, Infiniband EDR HPE</td>
<td>144,900</td>
<td>4,788.2</td>
<td>5,332.3</td>
<td>1,727</td>
</tr>
<tr>
<td>36</td>
<td>ECMWF</td>
<td>Cray XC40, Xeon ES-2695v4 18C 2.1GHz, Aries interconnect Cray Inc.</td>
<td>126,468</td>
<td>3,944.7</td>
<td>4,249.3</td>
<td>1,897</td>
</tr>
<tr>
<td>37</td>
<td>ECMWF</td>
<td>Cray XC40, Xeon ES-2695v4 18C 2.1GHz, Aries interconnect Cray Inc.</td>
<td>126,468</td>
<td>3,944.7</td>
<td>4,249.3</td>
<td>1,897</td>
</tr>
</tbody>
</table>

**www.top500.org**

- **47** supercomputing centers from TOP500 list with a dedicated mission for operational and research weather prediction, environmental and climate science. It includes NWSC Cheyenne system.
- + Multi-disciplinary supercomputing centers which allocate a lot of compute hours for ESM (e.g. ALCF, OLCF, NERSC, TACC, KISTI, KAUST, BSC…)
- + Medium size HPC centers (members of weather consortiums) and Cloud Service Providers
Evolution of ESM/NWP HPC center architecture over the last decade

Data source: www.top500.org

Data represents an architecture view of supercomputing centers from Top500 which are 100% dedicated to weather/climate.
NWP and ESM HPC centers in Top500

Distribution by OEM

- Cray: 47%
- HPE: 24%
- Sugon: 9%
- Bull: 6%
- Dell: 4%
- IBM: 4%
- Fujitsu: 2%
- Huawei: 2%
- Inspur: 2%

- Cray covers almost a half of NWP/ESM HPC in Top500

Distribution by arch

- Intel Broadwell: 39%
- Intel Haswell: 27%
- Intel Ivy Bridge: 2%
- Intel Sandy Bridge: 11%
- AMD Opteron: 4%
- Intel Skylake: 16%
- Fujitsu Sparc64: 1%

- 95% of NWP/ESM HPC runs on Intel architecture

Data source: www.top500.org
INTEL DATA CENTER GROUP

MOVE | STORE | PROCESS

LEADERSHIP WORKLOAD PERFORMANCE

GROUNDBREAKING MEMORY INNOVATION

EMBEDDED ARTIFICIAL INTELLIGENCE ACCELERATION

HARDWARE ENHANCED SECURITY

ENHANCED AGILITY & UTILIZATION

BUILT-IN VALUE

UNINTERRUPTED LEADERSHIP WORKLOAD PERFORMANCE

GROUNDBREAKING MEMORY INNOVATION

EMBEDDED ARTIFICIAL INTELLIGENCE ACCELERATION

HARDWARE ENHANCED SECURITY

ENHANCED AGILITY & UTILIZATION

INTEL XEON® SCALEABLE PROCESSORS

INTEL® XEON® PLATINUM 9200 PROCESSORS

A NEW CLASS OF ADVANCED PERFORMANCE

INTEL® XEON® PLATINUM 8200 PROCESSORS

INTEL® XEON® GOLD 6200 PROCESSORS

INTEL® XEON® GOLD 5200 PROCESSORS

INTEL® XEON® SILVER 4200 PROCESSORS

INTEL® XEON® BRONZE 3200 PROCESSORS

INTEL.COM/XEONSCALEABLE
INCREASING THE PACE OF INNOVATION

2014: INTEL® XEON® PROCESSOR E5 V3 HASWELL
2015: INTEL® XEON® PROCESSOR E5 V4 BROADWELL
2016: INTEL® XEON® SCALABLE PROCESSOR SKYLAKE
2017: 2ND GEN INTEL® XEON® SCALABLE PROCESSOR CASCADE LAKE
2018: COOPER LAKE & ICE LAKE
2019: SAPPHIRE RAPIDS
2020: NEXT GEN

DRIVING LEADERSHIP WORKLOAD PERFORMANCE

5 TO 7 QUARTER CADENCE
MOVING TO 4 TO 5 QUARTER CADENCE
Single node gen-to-gen study on Intel® Xeon® CPU

Selected applications: WRF, MPAS-A, HOMME, NEMO
Selected suite of NWP/ESM workloads

<table>
<thead>
<tr>
<th>Application</th>
<th>Version</th>
<th>Dataset</th>
<th>Compiler/MPI</th>
<th>Run</th>
</tr>
</thead>
<tbody>
<tr>
<td>WRF</td>
<td>3.9.1.1</td>
<td>CONUS-12km, CONUS-2.5km</td>
<td>Intel 2018 update 3</td>
<td>Out-of-the-box</td>
</tr>
<tr>
<td>MPAS-A</td>
<td>6.1</td>
<td>120km_L56, dycore+physics</td>
<td>Intel 2018 update 3</td>
<td>Out-of-the-box</td>
</tr>
<tr>
<td>NEMO</td>
<td>4.0</td>
<td>ORCA2_ICE_PISCES</td>
<td>Intel 2018 update 3</td>
<td>Out-of-the-box</td>
</tr>
<tr>
<td>HOMME</td>
<td>dungeon28</td>
<td>WACCM, NE=8</td>
<td>Intel 2018 update 3</td>
<td>Out-of-the-box</td>
</tr>
</tbody>
</table>
Best performance on IA with Intel® tools

Example from NCAR: Intel® compiler outperforms GNU compiler by 30% on WRF CONUS-12km workload on NWSC Cheyenne supercomputer powered by Intel® Xeon® E5-2697v4 CPU
Performance results are based on testing as of January 30, 2019 to April 30, 2019 and may not reflect all publicly available security updates. See configuration disclosure for details. No product can be absolutely secure.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations, and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases. Intel does not control or audit the design or implementation of third party benchmark data or Web sites referenced in this document. Intel encourages all of its customers to visit the referenced Web sites or others where similar performance benchmark data are reported and confirm whether the referenced benchmark data are accurate and reflect performance of systems available for purchase. Refer to Configuration: HPC Workloads – WRF CONUS-12km, WRF CONUS-2.5km, MPAS-A, NEMO, HOMME, STREAM Triad, HPL - Other names and brands may be claimed as the property of others.
INTERCONNECT DAY

CXL Compute Express Link
CXL: A NEW CLASS OF INTERCONNECT FOR DEVICE CONNECTIVITY
Why a New Class of Interconnect?

Move past the PCIe limiters on heterogeneous computing and server disaggregation usages

- PCIe creates isolated memory pools with an inefficient mish-mash of access mechanisms.
- Moving operands and results back and forth between accelerators and devices is painful and inefficient.
- Resource sharing is all but disallowed.
- Latencies are an order of magnitude off of what is needed to enable disaggregated memory.
WHAT IS CXL?

- CXL is an alternate protocol that runs across the standard PCIe physical layer
- CXL uses a flexible processor port that can auto-negotiate to either the standard PCIe transaction protocol or the alternate CXL transaction protocols
- First generation CXL aligns to 64 GB/s PCIe Gen5
- CXL usages expected to be key driver for an aggressive timeline to PCIe Gen6
The CXL transaction layer is comprised of 3 dynamically multiplexed sub-protocols on a single link:

- CXL.io – Discovery, configuration, register access, interrupts, etc.
- CXL.cache – Device access to processor memory
- CXL.memory – Processor access to device attached memory
HETEROGENEOUS COMPUTING REVISIT, WITH CXL

CXL ENABLES A MORE FLUID AND FLEXIBLE MEMORY MODEL
SINGLE, COMMON, MEMORY ADDRESS SPACE ACROSS PROCESSORS AND DEVICES

- More efficient population and update of operands
- More efficient extraction of results
- Memory resource “borrowing”
- User/Kernel level data access and data movement.
- Low latency to memory, host to device and device to host.

- PCIe DMA
- CPU-ATTACHED MEMORY (OS MANAGED)
- Memory Load/Store
- "Writeback" Memory
- CPU
- CPU

- ACCELERATOR-ATTACHED MEMORY (RUNTIME MANAGED CACHE)
- PCIe DMA
- NIC
- NIC

- "Writeback" Memory
- Memory Load/Store
- GPU
- FPGA
- AI
DIVERSE WORKLOADS REQUIRE DIVERSE ARCHITECTURES

The future is a diverse mix of scalar, vector, matrix, and spatial architectures deployed in CPU, GPU, AI, FPGA and other accelerators.
PROGRAMMING CHALLENGE

Diverse set of data-centric hardware

No common programming language or APIs

Inconsistent tool support across platforms

Each platform requires unique software investment
INTEL’S ONEAPI
CORE CONCEPT

Project oneAPI delivers a unified programming model to simplify development across diverse architectures

Common developer experience across Scalar, Vector, Matrix and Spatial architectures (CPU, GPU, AI and FPGA)

Uncompromised native high-level language performance

Based on industry standards and open specifications
ONEAPI FOR CROSS-ARCHITECTURE PERFORMANCE

Optimized Applications

Optimized Middleware & Frameworks

oneAPI Product

Direct Programming

Data Parallel C++

API-Based Programming

Libraries

Analysis & Debug Tools

CPU

SCALAR

GPU

VECTOR

AI

MATRIX

FPGA

SPATIAL

Some capabilities may differ per architecture.

Optimization Notice

Copyright © 2019. Intel Corporation. All rights reserved.

*Other names and brands may be claimed as the property of others.
DATA PARALLEL C++
STANDARDS-BASED, CROSS-ARCHITECTURE LANGUAGE

Language to deliver uncompromised parallel programming productivity and performance across CPUs and accelerators

Based on C++ with language enhancements being driven through community project

Open, cross-industry alternative to single architecture proprietary language

There will still be a need to tune for each architecture.
ADVANCED ANALYSIS & DEBUG TOOLS

Productive performance analysis across SVMS architectures

Performance Profiler
Parallelization Assistant
Debugger
SUMMARY

Diverse workloads for data-centric computing are driving the need for diverse compute architectures including CPUs, GPUs, FPGAs, and AI accelerators.

OneAPI unifies and simplifies programming of Intel CPUs and accelerators, delivering developer productivity and full native language performance.

OneAPI is based on industry standards and open specifications to encourage ecosystem collaboration and innovation.
DAOS: SCALE-OUT SOFTWARE-DEFINED STORAGE FOR HPC/BIG DATA/AI CONVERGENCE
Intel® Optane DC Persistent Memory is revolutionizing high performance storage by providing low-latency and fine-grained persistent storage.

Intel® Optane DC SSDs bring very high IOPS, and handle reads & writes concurrently without degradation.

... but existing distributed software is a bottleneck

- Optimized for millisecond rotating media
- POSIX constraints limit performance

Scale out storage needs to be built from the ground up for new NVM technology.
DAOS: Distributed Asynchronous Object Storage

DAOS is the scale-out software-defined storage platform for HPC, Big Data, and AI convergence.

- High throughput/IOPS @ arbitrary alignment/size
- Low-latency I/O
- Data access time orders of magnitude faster (µs vs ms)
- Primary storage on Aurora exascale supercomputer at Argonne National Labs, with a capacity of 230PB and bandwidth >25TB/s.