May 31, 2022

Intel Editorial: Accelerated Innovations for Sustainable, Open HPC

Intel meets insatiable computing demand with sustainability as a priority in the next era of supercomputing for all.

HAMBURG, Germany--(BUSINESS WIRE)-- *The following is an opinion editorial from Jeff McVeigh of Intel Corporation:*

This press release features multimedia. View the full release here: https://www.businesswire.com/news/home/20220531005771/en/

During the International Supercomputing Conference on May 31, 2022, in Hamburg, Germany, Jeff McVeigh, vice president and general manager of the Super Compute Group at Intel Corporation, announced Rialto Bridge, Intel's data center graphics processing unit (GPU). Using the same architecture as the Intel data center GPU Ponte Vecchio and combining enhanced tiles with Intel's next process node, Rialto Bridge will offer up to 160 Xe cores, more FLOPs, more I/O bandwidth and higher TDP limits for significantly increased density, performance and efficiency. (Credit: Intel Corporation)

As we embark on the exascale era and sprint towards zettascale, the technology industry's contribution to global carbon emissions is also growing. It has been estimated that by 2030, between 3% and 7%[1] of global energy production will be consumed by data centers, with computing infrastructure being a top driver of new electricity use.

This year, Intel committed to achieve net-zero greenhouse gas emissions in our global operations by 2040 and to develop more sustainable technology solutions. Keeping up with the insatiable demands for computing while creating a sustainable future is one of the biggest challenges for high performance computing (HPC). While daunting, it is achievable if we address every part of the HPC compute stack – silicon, software, and systems.

This is at the heart of my keynote at ISC 2022 in Hamburg, Germany.

Start with Silicon and Heterogeneous Compute Architecture

We have an aggressive HPC roadmap planned through 2024 that will deliver a diverse portfolio of heterogeneous architectures. These architectures will allow us to improve performance by orders of magnitude while reducing power demands across both general-purpose and emerging workloads such as AI, encryption and analytics.

The Intel® Xeon® processor code-named Sapphire Rapids with High Bandwidth Memory (HBM) is a great example of how we are leveraging advanced packaging technologies and silicon innovations to bring substantial performance, bandwidth and power-saving improvements for HPC. With up to 64 gigabytes of high-bandwidth HBM2e memory in the package and accelerators integrated into the CPU, we're able to unleash memory

bandwidth-bound workloads while delivering significant performance improvements across key HPC use cases. When comparing 3rd Gen Intel® Xeon® Scalable processors to the upcoming Sapphire Rapids HBM processors, we are seeing two- to three-times performance increases across weather research, energy, manufacturing and physics workloads[2]. At the keynote, Ansys CTO Prith Banerjee also shows that Sapphire Rapids HBM delivers up to 2x performance increase on real-world workloads from Ansys Fluent and ParSeNet[3].

Compute density is another imperative as we push for orders of magnitude performance gains across HPC and AI supercomputing workloads. Our first flagship Intel data center graphics processing unit (GPU), code-named Ponte Vecchio, is already outperforming competition for complex financial services applications and AI inference and training workloads**.** We also show that Ponte Vecchio is accelerating high-fidelity simulation by 2x with OpenMC[4].

We are not stopping here. Today we are announcing our successor to this powerhouse data center GPU, code-named Rialto Bridge. By evolving the Ponte Vecchio architecture and combining enhanced tiles with next process node technology, Rialto Bridge will offer significantly increased density, performance and efficiency, while providing software consistency.

Looking ahead, Falcon Shores is the next major architecture innovation on our roadmap, bringing x86 CPU and X^e GPU architectures together into a single socket. This architecture is targeted for 2024 and projected to deliver benefits of more than 5x performance-per-watt, 5x compute density, 5x memory capacity and bandwidth improvements[5].

Tenets of a Successful Software Strategy: Open, Choice, Trust

Silicon is just sand without software to bring it to life. Our approach to software is to facilitate open development across the entire stack and to provide tools, platforms and software IP to help developers be more productive and to produce scalable, better-performing, more efficient code that can take advantage of the latest silicon innovations without the burden of refactoring code. The oneAPI industry initiative provides HPC developers with cross-architecture programming so code can be targeted to CPUs, GPUs and other specialized accelerators transparently and portably.

There are now more than 20 oneAPI Centers of Excellence at leading research and academic institutions around the world, and they are making significant progress. For example, Simon MacIntosh-Smith and his team at Bristol University's Science Department are developing best practices for achieving performance portability at exascale using oneAPI and the Khronos Group's SYCL abstraction layer for cross-architecture programming. Their work will ensure that scientific code can achieve high performance on massive heterogeneous supercomputing systems around the world.

Tying It Together: Systems for Sustainable Heterogeneous Computing

As the data center and HPC workloads increasingly move toward disaggregated architectures and heterogeneous computing, we will need tools that can help us effectively manage these complex and diverse computing environments.

Today, we are introducing Intel® XPU Manager, an open-source solution for monitoring and managing Intel data center GPUs locally and remotely. It was designed to simplify administration, to maximize reliability and uptime by running comprehensive diagnostics, to improve utilization and to perform firmware updates.

A Distributed Asynchronous Object Storage (DAOS) file system provides system-level optimizations for the power-hunger tasks of moving and storing data. DAOS has an enormous impact on file system performance, both improving overall access time and reducing the capacity required for storage to reduce data center footprints and increase energy efficiency. In I/O 500 results relative to Lustre, DAOS achieved a 70x increase[6] in hard write file system performance.

Addressing the HPC Sustainability Challenge

We are proud to be partnering with like-minded customers and leading research institutions around the world to achieve a more sustainable and open HPC. Recent examples include our partnership with the Barcelona Supercomputing Center to set up a pioneering RISC-V zettascale lab, and our continued collaboration with the University of Cambridge and Dell to evolve the current Exascale Lab into the new Cambridge Zettascale Lab. These efforts build on our plans to create a robust EU innovation ecosystemfor the future of compute.

The bottom line is no single company can do it alone. The entire ecosystem needs to equally lean in, across manufacturing, silicon, interconnect, software and systems. By doing this together, we can turn one of the biggest HPC challenges of the century into the opportunity of the century – and change the world for future generations.

Jeff McVeigh is vice president and general manager of the Super Compute Group at Intel Corporation.

About Intel

Intel (Nasdaq: INTC) is an industry leader, creating world-changing technology that enables global progress and enriches lives. Inspired by Moore's Law, we continuously work to advance the design and manufacturing of semiconductors to help address our customers' greatest challenges. By embedding intelligence in the cloud, network, edge and every kind of computing device, we unleash the potential of data to transform business and society for the better. To learn more about Intel's innovations, go to newsroom.intel.com and intel.com.

Notices and Disclaimers:

[1] Andrae Hypotheses for primary energy use, electricity use and CO2 emissions of global computing and its share of the total between 2020 and 2030, WSEAS Trans Power Syst, 15 (2020)

[2] As measured by the following:

CloverLeaf

- Test by Intel as of 04/26/2022. 1-node, 2x Intel® Xeon® Platinum 8360Y CPU, 72 cores, HT On, Turbo On, Total Memory 256GB (16x16GB DDR4 3200 MT/s),

SE5C6200.86B.0021.D40.2101090208, Ubuntu 20.04, Kernel 5.10, 0xd0002a0, ifort 2021.5, Intel MPI 2021.5.1, build knobs: -xCORE-AVX512 –qopt-zmm-usage=high
- Test by Intel as of 04/19/22. 1-node, 2x Pre-production Intel® Xeon® Scalable Processor codenamed Sapphire Rapids Plus HBM, >40 cores, HT ON, Turbo ON, Total Memory 128 GB (HBM2e at 3200 MHz), BIOS Version EGSDCRB1.86B.0077.D11.2203281354, ucode revision=0x83000200, CentOS Stream 8, Linux version 5.16, ifort 2021.5, Intel MPI 2021.5.1, build knobs: -xCORE-AVX512 –qopt-zmm-usage=high

OpenFOAM

- Test by Intel as of 01/26/2022. 1-node, 2x Intel® Xeon® Platinum 8380 CPU), 80 cores, HT On, Turbo On, Total Memory 256 GB (16x16GB 3200MT/s, Dual-Rank), BIOS Version SE5C6200.86B.0020.P23.2103261309, 0xd000270, Rocky Linux 8.5 , Linux version 4.18., OpenFOAM® v1912, Motorbike 28M @ 250 iterations; Build notes: Tools: Intel Parallel Studio 2020u4, Build knobs: -O3 -ip -xCORE-AVX512
- Test by Intel as of 01/26/2022 1-node, 2x Pre-production Intel® Xeon® Scalable Processor codenamed Sapphire Rapids Plus HBM, >40 cores, HT Off, Turbo Off, Total Memory 128 GB (HBM2e at 3200 MHz), preproduction platform and BIOS, CentOS 8, Linux version 5.12, OpenFOAM® v1912, Motorbike 28M @ 250 iterations; Build notes: Tools: Intel Parallel Studio 2020u4, Build knobs: -O3 -ip -xCORE-AVX512

WRF

- Test by Intel as of 05/03/2022. 1-node, 2x Intel® Xeon® 8380 CPU, 80 cores, HT On, Turbo On, Total Memory 256 GB (16x16GB 3200MT/s, Dual-Rank), BIOS Version SE5C6200.86B.0020.P23.2103261309, ucode revision=0xd000270, Rocky Linux 8.5, Linux version 4.18, WRF v4.2.2
- Test by Intel as of 05/03/2022. 1-node, 2x Pre-production Intel® Xeon® Scalable Processor codenamed Sapphire Rapids Plus HBM, >40 cores, HT ON, Turbo ON, Total Memory 128 GB (HBM2e at 3200 MHz), BIOS Version EGSDCRB1.86B.0077.D11.2203281354, ucode revision=0x83000200, CentOS Stream 8, Linux version 5.16, WRF v4.2.2

YASK

- Test by Intel as of 05/9/2022. 1-node, 2x Intel® Xeon® Platinum 8360Y CPU, 72 cores, HT On, Turbo On, Total Memory 256GB (16x16GB DDR4 3200 MT/s), SE5C6200.86B.0021.D40.2101090208, Rocky linux 8.5, kernel 4.18.0, 0xd000270, Build knobs: make -j YK_CXX='mpiicpc -cxx=icpx' arch=avx2 stencil=iso3dfd radius=8,
- Test by Intel as of 05/03/22. 1-node, 2x Pre-production Intel® Xeon® Scalable Processor codenamed Sapphire Rapids Plus HBM, >40 cores, HT ON, Turbo ON, Total Memory 128 GB (HBM2e at 3200 MHz), BIOS Version EGSDCRB1.86B.0077.D11.2203281354, ucode revision=0x83000200, CentOS Stream 8, Linux version 5.16, Build knobs: make -j YK_CXX='mpiicpc -cxx=icpx' arch=avx2 stencil=iso3dfd radius=8,

[3] Ansys Fluent

- Test by Intel as of 2/2022 1-node, 2x Intel ® Xeon ® Platinum 8380 CPU, 80 cores, HT

On, Turbo On, Total Memory 256 GB (16x16GB 3200MT/s, Dual-Rank), BIOS Version SE5C6200.86B.0020.P23.2103261309, ucode revision=0xd000270, Rocky Linux 8.5 , Linux version 4.18, Ansys Fluent 2021 R2 Aircraft_wing_14m; Build notes: Commercial release using Intel 19.3 compiler and Intel MPI 2019u

- Test by Intel as of 2/2022 1-node, 2x Pre-production Intel® Xeon® Scalable Processor code names Sapphire Rapids with HBM, >40 cores, HT Off, Turbo Off, Total Memory 128 GB (HBM2e at 3200 MHz), preproduction platform and BIOS, CentOS 8, Linux version 5.12, Ansys Fluent 2021 R2 Aircraft_wing_14m; Build notes: Commercial release using Intel 19.3 compiler and Intel MPI 2019u8

Ansys ParSeNet

- Test by Intel as of 05/24/2022. 1-node, 2x Intel® Xeon® Platinum 8380 CPU, 80 cores, HT On, Turbo On, Total Memory 256GB (16x16GB DDR4 3200 MT/s [3200 MT/s]), SE5C6200.86B.0021.D40.2101090208, Ubuntu 20.04.1 LTS, 5.10, ParSeNet (SplineNet), PyTorch 1.11.0, Torch-CCL 1.2.0, IPEX 1.10.0, MKL (2021.4-Product Build 20210904), oneDNN (v2.5.0)
- Test by Intel as of 04/18/2022. 1-node, 2x Pre-production Intel® Xeon® Scalable Processor codenamed Sapphire Rapids Plus HBM, 112 cores, HT On, Turbo On, Total Memory 128GB (HBM2e 3200 MT/s), EGSDCRB1.86B.0077.D11.2203281354, CentOS Stream 8, 5.16, ParSeNet (SplineNet), PyTorch 1.11.0, Torch-CCL 1.2.0, IPEX 1.10.0, MKL (2021.4-Product Build 20210904), oneDNN (v2.5.0)

[4]Test by Argonne National Laboratory as of 5/23/2022, 1-node, 2x AMD EPYC 7532, 256 GB DDR4 3200, HT On, Turbo On, ucode 0x8301038. 1x A100 40GB PCIe. OpenSUSE Leap 15.3, Linux Version 5.3.18, Libararies: CUDA 11.6 with OpenMP clang compiler. Build Knobs: cmake --preset=llvm_a100 -DCMAKE_UNITY_BUILD=ON -DCMAKE_UNITY_BUILD_MODE=BATCH -DCMAKE_UNITY_BUILD_BATCH_SIZE=1000 -DCMAKE_INSTALL_PREFIX=./install -Ddebug=off -Doptimize=on -Dopenmp=on -Dnew_w=on -Ddevice_history=off -Ddisable_xs_cache=on -Ddevice_printf=off. Benchmark: Depleted Fuel Inactive Batch Performance on HM-Large Reactor with 40M particles
Test By Intel as of 5/25/2022, 1-node, 2x Intel(r) Xeon(r) Scalable Processor 8360Y, 256GB DDR4 3200, HT On, Turbo, On, ucode 0xd0002c1. 1x Pre-production Ponte Vecchio. Ubunt 20.04, Linux Version 5.10.54, agama 434, Build Knobs: cmake -DCMAKE_CXX_COMPILER="mpiicpc" -DCMAKE_C_COMPILER="mpiicc" -DCMAKE_CXX_FLAGS="-cxx=icpx -mllvm -indvars-widen-indvars=false -Xclang -fopenmp-declare-target-global-default-no-map -std=c++17 -Dgsl_CONFIG_CONTRACT_CHECKING_OFF -fsycl -DSYCL_SORT -D_GLIBCXX_USE_TBB_PAR_BACKEND=0" --preset=spirv -DCMAKE_UNITY_BUILD=ON -DCMAKE_UNITY_BUILD_MODE=BATCH -DCMAKE_UNITY_BUILD_BATCH_SIZE=1000 -DCMAKE_INSTALL_PREFIX=../install -Ddebug=off -Doptimize=on -Dopenmp=on -Dnew_w=on -Ddevice_history=off -Ddisable_xs_cache=on -Ddevice_printf=off Benchmark: Depleted Fuel Inactive Batch Performance on HM-Large Reactor with 40M particles

[5] Falcon Shores performance targets based on estimates relative to current platforms in February 2022. Results may vary.

[6] Results may vary. Learn more atio500 and “DAOS Performance comparison with Lustre installation” on YouTube.

All product plans and roadmaps are subject to change without notice.

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

Intel technologies may require enabled hardware, software or service activation.

Performance varies by use, configuration and other factors. Learn more at www.Intel.com/PerformanceIndex.

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See backup for configuration details. No product or component can be absolutely secure.

Your costs and results may vary.

Statements that refer to future plans or expectations are forward-looking statements. These statements are based on current expectations and involve many risks and uncertainties that could cause actual results to differ materially from those expressed or implied in such statements. For more information on the factors that could cause actual results to differ materially, see our most recent earnings release and SEC filings at www.intc.com.

© Intel Corporation. Intel, the Intel logo and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.

View source version on businesswire.com:
https://www.businesswire.com/news/home/20220531005771/en/

Bats Jafferji
1-603-809-5145
bats.jafferji@intel.com

Source: Intel Corporation