

# Case Study

HPC+AI Case study

Intel® Xeon® CPU Max Series

Intel® Data Center GPU Max Series

intel  
xeon

## Key HPC Codes Run Up to 3.8x Faster on Intel® Xeon® CPU Max Series

**Texas Advanced Computing Center (TACC) benchmarked 14 scientific applications on the latest Intel® HPC CPU with significant gains over Frontera and uplift for memory-bandwidth-bound codes.**

### Highlights:

- TACC selects Dell PowerEdge C6620 servers powered by Intel® Xeon® CPU Max Series and Dell PowerEdge XE9640 servers featuring Intel® Data Center GPU Max Series for its new Stampede3 supercomputer which will provide almost 10 petaflops of peak capability.
- The selections followed an assessment of the performance of 14 leading HPC codes on the latest Intel Xeon CPU Max Series.
- 2.6x average speed up on Intel Xeon CPU Max Series<sup>1</sup> in high bandwidth memory mode.
- New subsystem powered by 40 Intel® Data Center GPU Max Series for AI, ML and GPU-friendly applications.

### Executive Summary

In July 2023, the [Texas Advanced Computing Center \(TACC\)](#) at the University of Texas at Austin announced that the U.S. National Science Foundation (NSF) had awarded the institution a \$10 million grant for new hardware for the Stampede3 supercomputer to support academic research across the U.S.

The Stampede systems have, for over a decade, been the flagships in the NSF academic supercomputing ecosystem. Stampede3 will consist of:

- New 4 petaflop capability system for high-end simulation powered by 560 nodes built on Intel® Xeon® CPU Max Series with high bandwidth memory. These nodes add nearly 63,000 cores for the largest, most performance-intensive compute jobs.
- New GPU/Artificial Intelligence (AI) subsystem including 10 Dell PowerEdge XE9640 servers powered by 40 Intel® Data Center GPU Max Series for AI/Machine Learning (ML) and other GPU-friendly applications.
- Reintegration of 224 3rd Gen Intel® Xeon® Scalable processor nodes for higher memory applications and more than 1,000 existing Intel® Xeon® Scalable processors from Stampede2. These processors will support high-throughput computing, interactive workloads, and other smaller workloads.
- Addition of Cornelis Networks' new Omni-Path Express 400 Gb/s fabric technology with 24 TB/s backplane bandwidth. The new fabric offers a high-performance interconnect to enable low-latency and excellent scalability for applications and high connectivity to the I/O subsystem.



DELL Technologies

**TACC**  
TEXAS ADVANCED COMPUTING CENTER



The Community Earth System Model on Intel Xeon CPU Max Series with DDR5 was 2.5x faster than on Frontera; the code achieved a further 30 percent improvement on Intel Max Series CPU in HBM-only mode.

- PowerEdge C6620 servers and the XE9640 servers that will be installed in the newly designed Dell Technologies DLC7000 rack, supporting direct liquid cooling to each CPU and GPU, providing near room-neutral temperatures.
- Dell Technologies networking that will be the management platform for Stampede3.

“We believe the high bandwidth memory of the Xeon Max CPU nodes will help deliver better performance than any other CPU that our users have seen before.”

—Dan Stanzione, TACC Director

Stampede3, in aggregate, will consist of 1,858 compute nodes with more than 140,000 Intel cores, more than 330 terabytes of RAM, 13 petabytes of new storage, and almost 10 petaflops of peak capability. All components will be integrated into the same fabric, file systems, and allocations.

“We believe the high bandwidth memory of the Intel Xeon CPU Max Series nodes will help deliver better performance than any other CPU that our users have seen before,” TACC director Dan Stanzione said. “They offer more than double the memory bandwidth performance per core over the current 2nd and 3rd Gen Intel Xeon processor nodes in Stampede2.”

## Challenge

TACC is a leading supercomputing facility for academic researchers in the U.S. The center is always looking to the next generation of computing capabilities to continue to support the grand challenges facing science. When looking to replace the Stampede2 system—an Intel/Dell Technologies system that is the workhorse of the U.S. academic HPC community—TACC evaluated performance of scientific codes on the Intel® Xeon® CPU Max Series, a processor family with High Bandwidth Memory (HBM).

HBM has been one of the key ingredients in the rise of GPUs. It was also instrumental in the [2020 and 2021 Top500 #1 world ranking of the Fugaku supercomputer](#), which includes HBM-powered processors. The Intel Xeon CPU Max Series is the first x86 CPU to integrate HBM.

To evaluate performance of the new processor, TACC used a host of real-world HPC applications that are part of the NSF-funded Characteristic Science Applications (CSA) program. Through the CSA program, TACC collaborates with researchers to prepare scientific applications for the [Leadership-Class Computing Facility \(LCCF\)](#), which will host the agency’s flagship supercomputer, codenamed Horizon, expected to arrive in 2026. The applications were identified by the community of large-scale scientific computing users. They reflect the broad range of science domains and computational approaches—



<a href="#">Weather Research and Forecasting (WRF)</a>	Mesoscale numerical weather prediction system.
<a href="#">Parsec1/2</a>	A generic framework for architecture aware scheduling and management of micro-tasks on distributed many-core heterogeneous architectures.
<a href="#">Amber</a>	A suite of biomolecular simulation codes. It contains publicly available molecular mechanical force fields for the simulation of biomolecules and provides a package of molecular simulation programs.
<a href="#">Anelastic Wave Propagation, Olsen, Day, Cui (AWP-ODC)</a>	Simulates wave propagation in a 3D viscoelastic or elastic solid.
<a href="#">Seissol</a>	Code that simulates seismic wave phenomena and earthquake dynamics.
<a href="#">Community Earth System Model (CESM)</a>	A fully coupled global climate model that simulates Earth’s past, present, and future climate states.
EWP	3D deterministic wave propagation code
<a href="#">MIMD Lattice Computation (MILC)</a>	Runs simulations of four dimensional SU(3) lattice gauge theory.
<a href="#">Enzo-E</a>	Runs extreme scale numerical simulations to address current scientific questions in astrophysics and cosmology.
<a href="#">Ice Sheet and Sea Level System Model (ISSM)</a>	Used to model the evolution of the polar ice caps in Greenland and Antarctica.
<a href="#">Nanoscale Molecular Dynamics (NAMD)</a>	Simulates large biomolecular systems.
MuST	Code used to detect and report MPI errors
<a href="#">Parallel Spatial Direct Numerical Simulation (PSDNS)</a>	Simulation approach used to compute spatially evolving disturbances associated with the laminar-to-turbulent transition in boundary-layer flows.
<a href="#">athena++</a>	An astrophysical magnetohydrodynamics code.

Table 1. List of Characteristic Science Applications (CSA) and Weather Research and Forecasting (WRF) codes for benchmarking

from language to method to workflow—that researchers will run on future supercomputers.

### Solution

TACC researchers benchmarked 13 of the CSA codes and the Weather Research and Forecasting code (WRF) on the Intel Xeon CPU Max Series. Table 1 lists the codes used. The same codes were benchmarked on 2nd Gen Intel® Xeon® processors of Frontera—TACC’s most powerful capability computing systems and currently #21 on the June 2023 Top500 list.

The Intel Xeon CPU Max Series can run in a variety of modes—including an HBM-only mode and a flat mode where HBM can be turned off, relying only on DDR5. TACC tested the efficacy of the Intel Xeon CPU Max Series in both of these memory modes to understand the performance characteristics and benefits of HBM vs. DDR5. The Intel Xeon CPU Max Series delivered significant performance gains in both modes, especially for memory-bandwidth-bound applications.



The 3D earthquake code Anelastic Wave Propagation code ran 3.7x faster on Intel Max Series CPU than on Frontera and showed a 100 percent boost with HBM.<sup>1</sup>

### Results

Both modes delivered significant gains over the 2nd Gen Intel Xeon processors that power the TACC Frontera supercomputer. For example, with DDR5 memory only, the codes ran 2x faster on average than the previous version.<sup>1</sup> For massively parallel, data-hungry, and memory-bandwidth-limited problems, however, the Intel Xeon CPU Max Series with HBM excelled even more—with a 2.6x average speed-up.<sup>1</sup>

More than a third of the codes run on the Intel Xeon CPU Max Series with HBM saw 50 percent or more performance improvements over running only DDR5. Some codes saw up to 2x faster performance with the addition of HBM.

“The new Intel Xeon CPU Max Series has exactly twice as many cores as the 2nd Gen Intel Xeon processor, so I expect it will be at least two times better,” said John Cazes, TACC Director of HPC. “With HBM, however, it’s 2.6x, so it’s a great multiplier. It’s got enough memory bandwidth that the cores on the Intel Xeon CPU Max Series cannot saturate the memory bandwidth that HBM provides. This is a very rare problem to have on a CPU.”

### Faster... climate projections, materials discovered, universes modeled

Among the 14 applications that were assessed are software for large international experiments, like the [IceCube Neutrino Observatory](#), widely used codes from the earthquake and astrophysics communities, and custom codes that explore innovative approaches to machine learning and black hole modeling. Refer to Figure 1.

### Performance Highlights

One code seeing significant performance improvements with HBM is a special configuration of the Community Earth System Model (CESM) being developed by the NSF-sponsored EarthWorks project, led by Colorado State University, to

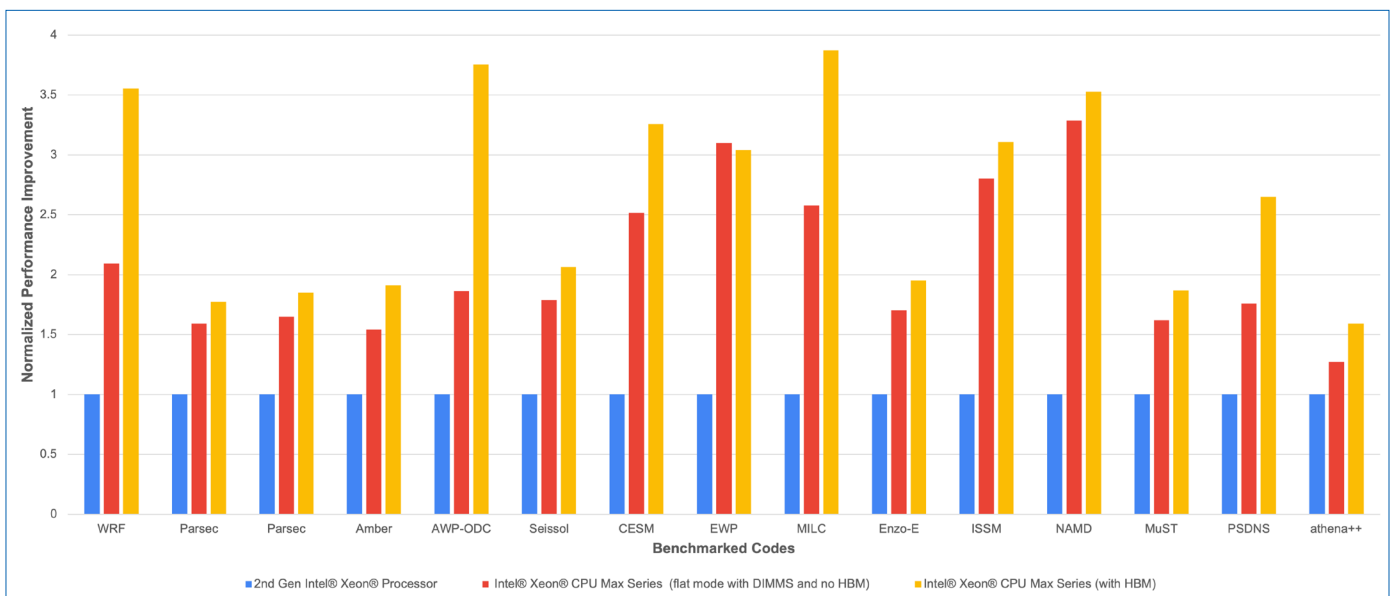


Figure 1. Normalized performance comparison of Characteristic Science Applications (CSA) and Weather Research and Forecasting (WRF) codes<sup>1</sup>





The Weather Research and Forecasting Model (WRF) saw 2.09x speedup on the Intel Xeon CPU Max Series processor with DDR5 compared to Frontera’s CPUs. On Intel Xeon CPU Max Series with HBM, WRF ran 3.5x faster than 2nd Gen Intel Xeon processors—a 70 percent speed-up over DDR5.<sup>1</sup>

study seasonal weather and climate phenomena at ultra-high resolutions. CESM is one of the principal climate codes used by the earth science community. CESM is developed and maintained by the National Center for Atmospheric Research (NCAR) in collaboration with the research community. The EarthWorks configuration of CESM was 2.5x faster on the Intel Xeon CPU Max Series with DDR5 than on Frontera;<sup>1</sup> the code achieved a further 30 percent improvement (to 3.2x) in HBM-only mode.<sup>1</sup>

“Applying the power of new technologies will enable us to develop global storm-resolving models that will help us better understand the risks that come with climate change,” said Colorado State University professor David Randall, one of the developers of the EarthWorks configuration. “A 2.5x to 3x speedup means we can find answers faster or increase the resolution and accuracy of our models even further.”

The Weather Research and Forecasting Model (WRF) is another state-of-the-art numerical weather prediction system

designed for both atmospheric research and operational forecasting applications. WRF saw 2.09x speedup on the Intel Xeon CPU Max Series processor with DDR5 compared to Frontera’s CPUs.<sup>1</sup> On Intel Xeon CPU Max Series with HBM, WRF ran 3.5x faster than 2nd Gen Intel Xeon processors—a 70 percent speed-up over DDR5.<sup>1</sup>

Another code that is showing exceptional performance on both Intel Xeon CPU Max Series memory modes is the 3D earthquake code, Anelastic Wave Propagation (AWP). The code was developed by Yifeng Cui of the San Diego Supercomputer Center. The code ran 3.7x faster on Intel Xeon CPU Max Series than on Frontera and showed a 100 percent boost with HBM.<sup>1</sup>

For applications that are not yet optimized to take advantage of HBM, Cazes believes the availability of Intel Xeon CPU Max Series will lead to code and algorithmic changes.

“We believe the high bandwidth memory of the Intel Xeon CPU Max Series nodes will help deliver better performance than any other CPU that our users have seen before,” Stanzone said. “They offer more than double the memory bandwidth performance per core over the current 2nd and 3rd Gen Intel Xeon processor nodes in Stampede2. We look forward to deploying Stampede3 as the next high capability and capacity HPC system in the national cyberinfrastructure available to all open science research projects in the U.S.”

### No Code Changes Required

Porting codes is always a consideration when looking at new CPU architectures. The time and effort it takes to develop and optimize a code reduces cycles available for the scientific effort. For many small teams, it is prohibitively hard to port complicated, multi-dependency legacy codes to GPUs.

**“Because we have the same system libraries, I could just lift the binaries that we ran on Frontera and run them on the Intel Xeon CPU Max Series and they just worked.”**

*—John Cazes, TACC Director of HPC*

It was easy for the TACC team to evaluate and compare the performance of the science codes. Little to no code changes were required to port the codes from Frontera CPUs to the latest generation of Intel data center processors. This is beneficial for the thousands of codes and billions of lines of

## Performance Benefits of Intel Xeon CPU Max Series

Here’s are a few examples of the performance TACC is seeing for codes running on the new Intel Xeon CPU Max Series:

- The EarthWorks configuration of CESM was 2.5x faster on the Intel Xeon CPU Max Series with DDR5 than on Frontera;<sup>1</sup> the code achieved a further 30 percent improvement (to 3.2x) in HBM-only mode.<sup>1</sup>
- WRF saw 2.09x speedup on the Intel Xeon Max Series processor with DDR5 compared to Frontera’s CPUs.<sup>1</sup> On Intel Xeon CPU Max Series with HBM, WRF ran 3.5x faster than 2nd Gen Intel Xeon processors—a 70 percent speed-up over DDR5.<sup>1</sup>
- The 3D earthquake code, Anelastic Wave Propagation (AWP) ran 3.7x faster on Intel Xeon CPU Max Series than on Frontera and showed a 100 percent boost with HBM.<sup>1</sup>

scientific software that scientists have optimized for x86 processors.

“Because we have the same system libraries, I could just lift the binaries that we ran on Frontera and run them on the Intel Xeon CPU Max Series and they just worked,” said John Cazes, head of HPC at TACC. This echoed the sentiment of other early customers, including researchers from [Los Alamos National Laboratory](#) and [Numenta](#).

Performance of these codes on the latest Intel Xeon processors is compelling. Adding to performance, the ease with which the codes can be taken from Frontera directly to the newest CPUs gives researchers both faster results without extra work.

## Summary

Assessing 13 of the CSA codes and WRF, TACC’s evaluation shows considerable performance boosts using both DDR5 and HBM-only modes of the Intel Xeon CPU Max Series compared to Frontera. Most interesting are the benefits of HBM to many of the codes when run on the Intel Xeon CPU Max Series. Speedup also comes in the form of scientists not needing to spend time on porting codes across different systems and their CPUs.

“The use of accelerators and GPUs are definitely on the rise in HPC and AI, but it’s not clear that much of the advantage isn’t provided by high bandwidth memory,” said Stanzione. “We need high performance CPUs too, and based on our benchmarks, the Intel Xeon CPU Max Series will provide clear advantages to our users.”



<sup>1</sup> Applications compared by TACC were WRF CONUS 2.5km, Parsec liquid\_water\_64H2O\_0.3A, Parsec Si1947H604\_0.9A, Amber STMV\_production\_BP4\_4fs, AWP-ODC Fortran, Seis-sol tpv5\_inode, CESM PFS\_Ld5\_P56.mpsal20\_mpsa, EWP MgB2\_16, MILC grid 32x32x32x32, Enzo-E 128^3 root mech 512 root blocks, ISSM test435, MuST muffin\_56x32, PSDNS 768x768x768, Plascom, and athena++. These were run on systems installed at TACC that use Intel® Xeon® 8280, Intel® Xeon® Max 9480 (Flat Mode, DDR only), and Intel® Xeon® Max 9480 (HBM only). TACC observed a 2.6x average performance increase from Frontera (Intel® Xeon® 8280) to Intel® Xeon® Max 9480 (HBM only) on these workloads.

Performance varies by use, configuration and other factors. Learn more at [www.Intel.com/PerformanceIndex](http://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.

For workloads and configurations visit [www.Intel.com/PerformanceIndex](http://www.Intel.com/PerformanceIndex). Results may vary.

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

Your costs and results may vary.

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

© 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.