KTH BREAKS NEW GROUND IN SCIENCE USING AMD EPYC™ CPUS AND INSTINCT™ GPUS

Six times the compute performance in the same power envelope using AMD EPYC processors and AMD Instinct accelerators

CUSTOMER

KTH VETENSKAP OCH KONST

INDUSTRY

Research and Education

CHALLENGES

Increase performance for computational science in Sweden

SOLUTION

Deploy HPE Cray EX HPC supercomputer cluster powered by AMD EPYC™ processors and Instinct™ accelerators

RESULTS

About six times as much overall performance for comparable overall power consumption, about 25 times as many GFLOPS per watt

AMD TECHNOLOGY AT A GLANCE

2nd Gen AMD EPYC™ CPUs
AMD Instinct™ GPUs

TECHNOLOGY PARTNER

Hewlett Packard Enterprise

CRAY

High-performance computing (HPC) is becoming increasingly important for science research. Now that the advancement of human knowledge is frequently determined by complex calculations, the more processing power you can deploy, the more discoveries you can make. So, universities around the world are increasingly investing in HPC, and one of the leading research institutions focusing on this area is the KTH Royal Institute of Technology (KTH) in Sweden. When KTH procured new supercomputing resources to enhance its HPC services to the Swedish and European research communities, an HPE Cray EX cluster powered by AMD EPYC™ processors and AMD Instinct™ GPUs offered exceptional performance per watt for the needs of the users.

"The AMD hardware was the enabling technology to get to the next level of resolution, but in less time."

Niclas Jansson, Researcher at KTH

Serving a diverse range of applications areas

"High-performance computing is a very important field for us," says Professor Dirk Pleiter, Director of the PDC Center for High Performance Computing at KTH. "We play two roles. On the one hand, PDC at KTH is a provider of HPC resources at the national level in Sweden. But we also provide these services to European users. HPC is a means to facilitate new science and engineering discoveries. But the challenge for PDC's supercomputing systems is that we have a huge diversity of workloads. We can't just optimize for one specific type of workload. We serve any research communities that can benefit from using HPC resources. For example, our system is used by Scania, which is a leading provider of transport solutions with headquarters in Sweden."

KTH, as a partner in the Swedish National Infrastructure for Computing (SNIC) consortium which provides a national HPC infrastructure for Swedish research, needed to leverage modern HPC architectures to deliver the best possible resources. "The only way was to make our resources more and more powerful," says Pleiter. "And to do that we needed to use GPU accelerators, not just CPUs. That means we had to consider the portability of the applications that researchers use, as the leading software has been developed over a very long time and therefore has primarily been geared to CPU-based systems. It's a challenge to keep the code bases up to date and adjustable for different architectures." Power consumption was another key consideration when KTH and PDC were deciding on the architecture for their latest HPC system.

The range of applications in use on KTH's HPC facilities is extensive. "Many researchers are running computational fluid dynamics (CFD) applications, as well as more general engineering applications," says Pleiter. "In the CFD area, there are extremely scalable applications, such as Nek5000, alongside powerful applications that are notorious for their scalability issues, like OpenFOAM. Another major group using our HPC systems are researchers from the fields of biophysics and genomics. Other large user groups are from materials science and computational chemistry."

"We have 25 applications pre-installed on our systems," continues Pleiter. "This includes, for instance, codes like: Neko, which is a reimplementation of Nek5000; GROMACS, a quite well-known molecular dynamics application; and VeloxChem, which is a newcomer in the area of computational chemistry. All these applications relate to high impact societal challenges.

For example, a lot of projects are focused on drug design in molecular dynamics using GROMACS. Neko has been used in engineering to understand the flows involved in airplane design or wind turbines to improve energy efficiency. Thus many of these applications play important roles in supporting the transition to carbon neutral economies. Applications in the materials science area are also frequently related to optimizing our use of renewable energy, such as creating more efficient materials for use in solar power systems."

More performance per watt from AMD

KTH had to accommodate this diverse application ecosystem when procuring a new supercomputer for its PDC facility. AMD CPUs and GPUs were considered for the deployment from the very beginning. "AMD has always been a major technology supplier in the area of HPC," says Pleiter. "What makes their CPUs very attractive is the very high core count. It was also helpful that the memory bandwidth was higher than the competitors'."

"Our new system powered by AMD will be at least six times faster than our previous one, in a comparable power envelope."

Professor Dirk Pleiter, Director of the PDC Center for High Performance Computing at KTH

"On the GPU side, there are two aspects," adds Pleiter. "On the one hand, memory bandwidth is very important. AMD has a lead here due to their high-memory bandwidth (HBM) technology. But AMD GPUs also provide compute performance within a limited power envelope, which allows us to squeeze out much more performance per watt. The available FLOPS, per watt, is an important aspect that made us choose AMD Instinct GPUs."

KTH's needs have for some years been well served by solutions from the hardware integrator HPE Cray. "We have a history of working with Cray," says Pleiter. "It was very important to work with an OEM that could supply a complete multi-server HPC system, including integrated management. HPE has strong research and computational science know-how available internally, too, so we can discuss porting strategies with them. Experts from AMD also helped with this, by organizing hackathons. The experts from HPE and AMD helped us port our applications from an early stage."

"We also have a limited power envelope within which to operate," continues Pleiter. "It would be difficult for our infrastructure to go far beyond a megawatt of power consumption. The AMD EPYC CPU-only nodes produce about 6 GFLOPS per watt, the GPU-accelerated nodes boost this to about 60 GFLOPS per watt. These results meant the outcome of the KTH public procurement process was that AMD EPYC CPUs plus AMD Instinct GPUs were the best solution for us. That was not only demonstrated by simple numerical benchmarks, but it was also the case when considering application benchmarks, for example running GROMACS."

Power-efficient HPC

KTH is deploying a total of 1270 nodes in the CPU partition and 56 nodes in the GPU partition of its new cluster, an HPE Cray EX system called Dardel. Each CPU node is powered by dual 2.25GHz 64-core AMD EPYC™ 7742 processors. The majority (524) are equipped with 256GB of memory, but there are also some nodes with 512GB or 1TB and two have 2TB RAM. The GPU nodes will also use 64-core AMD EPYC processors, but in single-socket configurations with 512GB of memory. Each will be equipped with four AMD Instinct™ MI250X GPUs.

"Migrating to the new cluster from our previous systems was straightforward," says Pleiter. "Our new system powered by AMD will be at least six times faster than our previous one, in a comparable power envelope. The peak performance of our previous system was 2.28 PFLOPS, and the new system's will be about 13.5 PFLOPS. In the green performance metric, our previous system had an efficiency of 1.8 GFLOPS per watt. In comparison, the most extreme figures for our new cluster will be from the GPU partition, where we expect to end up with similar figures to the Frontier and LUMI supercomputers. Today, with the figures which have been published in the June 2022 edition of the Top 500 green performance report, that's about 50 GFLOPS per watt. So that's an improvement by more than a factor of 25."

Detecting Alzheimer's earlier

This huge leap in performance from Dardel promises significant enhancements in research capabilities. Patrick Norman, Professor of Theoretical Chemistry at KTH, is expecting major benefits for his areas of study, which include research into Alzheimer's disease via the VeloxChem software his team has developed. "The niche I've been working on for my entire career has been the interaction between molecular materials and radiation in different wavelengths and radiances," says Norman. "The spectrum spans all the way from visible to UV and up to X-rays. We refer to this as theoretical spectroscopy. We investigate light harvesting, where the light from the sun excites the molecule from a ground state to an excited state. Then you can harvest that energy into chemical energy and use it in the future."

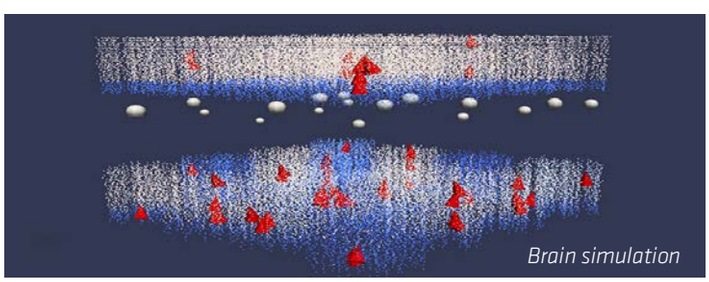
"My flagship application is a collaboration that I've had for 12 years within an experimental group at Linköping University, 200 kilometers southwest of Stockholm," says Norman. "They are designing ligands that you inject into the blood vessels, which then spread around in the body. Once they get close to amyloid protein aggregates, which are the cause of Alzheimer's, they attach themselves to those misfolded proteins and change their optical characteristics. You can then use fluorescence imaging to see colorful pictures of different stages of this disease. The key thing here is that you can do that some 15 years before clinical symptoms become apparent, in other words, well before a patient starts to lose memory and have cognitive problems."

"That is one of the flagship applications that I will be using Dardel for," says Norman. "The ambition here is to do early-stage detection, such that you could slow down the process. We are rewriting our code to make full use of the MI250X GPUs. One of the fantastic things about the AMD CPUs is the number of cores that are available. We're now working with nodes that have 128 cores. This will have a huge impact on how fast we can develop our research."

Optimizing sustainable sea transport

Dardel, particularly via its GPU nodes, also looks promising for computational fluid dynamics, a key scientific domain for KTH. Niclas Jansson, a researcher at KTH and lead developer of the highly scalable and highly portable Neko code, will be able to pursue his research in this area with much more depth. "This workload is very computationally demanding," he says. "We are trying to solve fundamental flow problems, which are very expensive to compute. We have been developing our own code here at KTH to run on future machines such as Dardel."

"A very timely topic that we're studying is a Flettner rotor," says Jansson. "This is a rotating cylinder that creates a force that can then act as a wind sail. It's called the Magnus Effect. The goal is to reduce carbon emissions in sustainable transport."



Brain simulation

Ship with Flettner rotors

"We are simulating the interaction of the incoming turbulence with the rotation of a cylinder representing this type of wind sail to see what kind of forces you would get to propel a ship, so we can make a prediction about how much that thrust would help to reduce the fuel cost, for example, in global shipping," continues Jansson.

Initial experiments with test nodes in advance of the full Dardel deployment have impressed Jansson greatly. "I was completely amazed when we ran large cases because I've been using large CPU clusters for my entire professional career," he says. "You must always request a large chunk of the machine and wait in the queue. But in one situation we were quite late on the deadline, and we managed to run a case in less than three days using a couple of nodes with AMD hardware, while we would have needed more than two weeks on the entire CPU partition of the previous system at PDC. The AMD hardware was the enabling technology to get to the next level of resolution but in less time."

"Most of the systems that have been deployed recently are based on AMD EPYC processors because, for typical scientific applications, they provide outstanding performance characteristics."

Professor Dirk Pleiter, Director of the PDC Center for High Performance Computing at KTH

"Instead of just doing detailed studies on fundamental geometries, you can start to move it into more complex shapes. For example, in the past, you had to run an air flow representation of a building on the CPUs. Now, with GPUs added, you can do an analysis of the air flow on an entire city block, which has not been feasible before. It's really a game changer with a very large impact. More cities, for example London, now have building guidelines where you need to do a detailed wind comfort analysis in the project phase of a construction project. If an engineering firm sticks with CPUs, they need to build a small cluster. But with a small investment in buying a couple of GPU accelerators, they can get back results that would otherwise have required 1,000 or 2,000 cores."

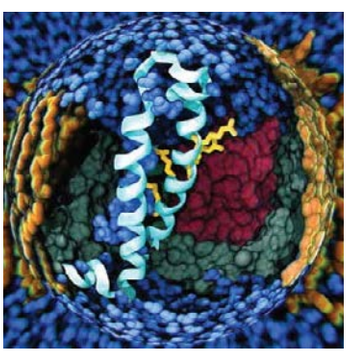
In an environment where power consumption is at a premium, AMD GPUs have proven to be an excellent choice. "We found that the performance of an AMD Instinct MI250X is comparable to that of two NVIDIA A100 GPUs, but it consumes about half the power. The performance is good and less power is used, which is a win-win situation for us. Another huge benefit is the interconnect on the MI250X, where we have the network going directly into the GPUs."

Outstanding performance for science

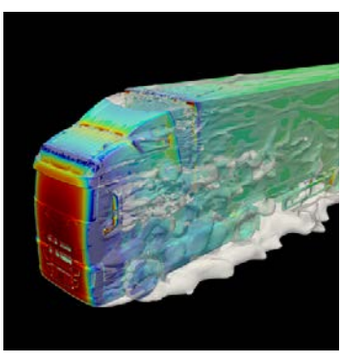
"The large number of cores on AMD EPYC CPUs has been the most significant feature for Dardel, along with the new packaging technology," says Pleiter. "This convinces us that the next generations of the EPYC CPUs will also be very interesting because the packaging allows the integration of even more cores by having more chiplets. It also enables an increase in the integrated cache, as we have seen recently in later AMD EPYC CPU generations."

"In the final configuration of Dardel, which should be ready by the end of the year, we will have more than 1,250 CPU-only nodes and 56 GPU nodes. We then hope to further increase the number of GPUs available in our system, as funding allows. One of the things that we would like to accomplish now with this new system is to attract more researchers focused on machine learning. That's a strand that we can leverage with the GPU nodes. Most of the workloads on our system have been simulation-based, but we are seeing a steady increase in the machine learning type of workloads and we are actively promoting that. We see good opportunities for further uptake of scientific use in this area, which would justify increasing the number of GPUs available."

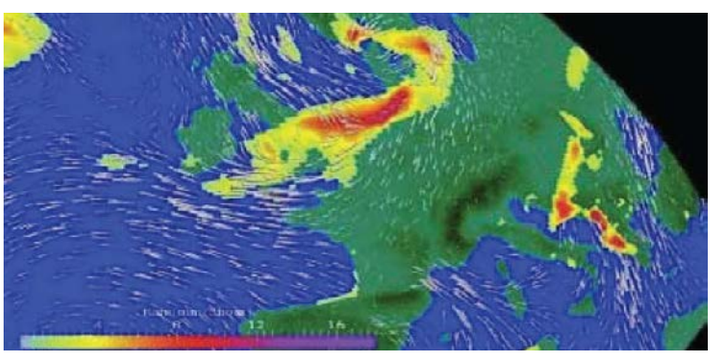
Pleiter expects AMD to continue to be the key hardware vendor to fulfil this expansion: "Most of the systems that have been deployed recently are based on AMD EPYC processors because, for typical scientific applications, they provide outstanding performance characteristics."



Biomolecular modeling



CFD simulation



Climate prediction

About KTH

Since its founding in 1827, the KTH Royal Institute of Technology in Stockholm has grown to become one of Europe's leading technical and engineering universities. KTH conducts research and education in engineering and technology and is Sweden's largest technical university. Currently, KTH consists of five schools with four campuses in and around Stockholm. The PDC Center for High Performance Computing at KTH is the leading provider of high-performance computing (HPC) services for academic research in Sweden as part of the Swedish HPC research infrastructure SNIC. PDC's HPC services are made available to Swedish and European academic researchers and to the commercial sector. The main HPC system at PDC is Dardel, an HPE Cray EX with an expected peak performance of about 13.5 petaflops when both phases of the system are operational. For more information visit kth.se.

About AMD

For more than 50 years AMD has driven innovation in high-performance computing, graphics and visualization technologies. Billions of people, leading Fortune 500 businesses and cutting-edge scientific research institutions around the world rely on AMD technology daily to improve how they live, work and play. AMD employees are focused on building leadership high-performance and adaptive products that push the boundaries of what is possible. For more information about how AMD is enabling today and inspiring tomorrow, visit the AMD (NASDAQ: AMD) website, amd.com/Instinct, blog, LinkedIn and Twitter pages.