Efficiency breakthroughs in HPC to drive "production science," global research

HPC needed economic disruption; computer scientists in the UK obliged

Case Study

Company snapshot

Name: **GW4, UK Met Office**

Description: **U.K. weather experts wanted more accurate models to support forecasts; computer scientists delivered Isambard, powered by Arm-based Marvell ThunderX2 SoCs, with unprecedented performance per dollar.**

Website:
https://gw4.ac.uk/isambard/

Goal

Design and deliver Arm-based supercomputers that are highly competitive on performance, efficiency and price.

Solution

The rate of innovation in the HPC CPU segment has slowed of late, and this is a concern for scientists around the world who require access to increasingly fast, more-efficient compute machines to help deliver new science to the world.

Looking back over the last half century, raw performance has always been a leading driver of computer architecture innovations. But today, because of increasing limitations in Dennard scaling at the silicon level and pure economics at the silicon and systems levels, performance increases are slowing. It's becoming clear that the economics of supercomputing need to be disrupted.

Fortunately, that disruption is underway. Power efficiency has risen to the fore as a crucial HPC design consideration, and this promises to change how and where we deliver massive compute capabilities to the world's scientists.

Supercomputer systems not only are expensive to design and build, they're also expensive to run. Most require megawatts of power, akin to the power usage in small towns. Power costs vary around the world and can be expensive. For example, a system that requires 2 megawatts of power costs around $2 million per year for electricity alone. This problem is only expected to get worse in the future. The world can't afford to continue to increase compute performance if power bills also keep increasing. Most of these systems are at universities or labs with limited budgets, so higher energy bills mean less money for scientists or hardware to aid scientific exploration.

That means all eyes are now focused on compute efficiency – delivering, for example, the same compute performance for half the power consumption or double the compute performance without increasing power consumption. This is today's power-efficiency design challenge.

These are the challenges that a group of computer scientists and domain scientists set out to overcome with the Isambard supercomputer.

"The rate of innovation in computer hardware in HPC has been slowing down, and so there is more opportunity for disruption than ever before," says Simon McIntosh-Smith, professor of High Performance Computing at the University of Bristol UK, and leader of the Isambard project.

The team's initial observations were around HPC's status quo: most HPC machines have been built around the same commodity processors, which means everyone, more or less, enjoys the same incremental performance improvements but is limited in terms of design flexibility and system optimization.

"We believe that Arm's ease of customization should result in a more diverse range of highly optimized processors, giving the HPC hardware space its next big boost," McIntosh-Smith said.

The team began exploring HPC alternatives around Arm partly because of the company's technological success in the mobile industry: Arm processor IP technology has had a long history of continuous innovation and as the company began to offer IP for the server and HPC space, that meant designers could count on similar innovation.

"We've seen how the Arm ecosystem has brought lots of competition and choice and very rapid innovation, such as in vector instruction sets," McIntosh-Smith said in a 2018 presentation.

The idea for Isambard came from McIntosh-Smith's work with the Mont Blanc project, a European effort begun in 2011 to explore the use of Arm technology in supercomputers. McIntosh-Smith approached colleagues from the Great Western Four (GW4) universities – made up from the universities of Bristol, Bath, Exeter and Cardiff – and the idea for Isambard was born. McIntosh-Smith and his GW4 partners teamed with the UK's weather forecasting agency, the Met Office, along with Cray, EPSRC and Arm to design the machine.

This consortium began the design in 2016, with a goal of delivering the world's first Arm-based production supercomputer. It would consist of more than 10,000 Armv8 cores in what ultimately would be a system based on Marvell ThunderX2 processors. Key to the decision to use Cray was that company's long history of high-performance, massively parallel compute but also its extensive software tools, from compilers to libraries and other tools.

When the Cray XC50 system was installed and complete, it contained 168 compute nodes or servers, packaged into four nodes in each of the 42 blades, which were connected by Cray's Aries interconnect technology. Each Marvell ThunderX2 CPU contains 32 cores each, running at 2.5 GHz for a total of 10,752 Arm-based cores. At 100 percent utilization, the machine uses 100kW.

It's one thing to achieve a computing milestone such as this, but the purpose, as McIntosh-Smith has said, is to deliver "production science" to researchers who need it. That's working already: As of July 2020, there were more than 400 registered global users on Isambard, and the system is continually running at nearly full capacity and at near 100% availability.

The Met Office has used Isambard for weather research to help make forecasts more accurate so authorities can better prepare and help communities avoid causalities and property damage during major storms.

In one fascinating health care application, researchers looked at how they might leverage a protein that regulates bone density in the human body. This particular protein can mutate, causing either osteoporosis or, to the positive, additional bone density. Scientists wanted to understand how they might develop therapies for osteoporosis. They simulated more than 8 million potential drugs – using 1 million core hours on Isambard – to understand how they might interact with the protein. From those simulations, researchers are working on a range of compounds that look promising as therapies, and those are being developed in the labs.

Additionally, scientists are now using Isambard to explore potential drugs to treat COVID-19. Molecular dynamics simulations are being used to investigate the 'spike' protein the decorates the surface of the SARS-CoV-2 virus, and its interactions with potential drugs.

The work on Isambard continues apace. In February 2020, the UK's Engineering and Physical Sciences Research Council (EPSRC) awarded the Isambard consortium £4.1 million to build Isambard 2. The design will be double the size of Isambard 1, with up to 21,504 high performance cores and 336 nodes to provide the work-horse grunt required by demanding scientific applications. To complement this production system, Isambard 2 will also be adding the very latest Arm-based CPUs optimized for HPC – Fujitsu's A64FX. These latest 7nm-based CPUs feature increased efficiency, delivering more compute performance and memory bandwidth than any other CPU to-date. Fujitsu's A64FX CPUs recently grabbed headlines around the world, powering Japan's Fugaku system to the coveted number one position on the Top500 league table of international supercomputers.

Isambard 2 is expected to be production-ready by October 2020, with the A64FX system being made available soon afterwards.

Learn more about Isambard and about Arm Neoverse computing solutions.