NVIDIA Brings CUDA to Arm, Enabling New Path to Exascale Supercomputing

Global HPC Leaders Join to Support New Platform

International Supercomputing Conference -- NVIDIA today announced its support for Arm CPUs, providing the high performance computing industry a new path to build extremely energy-efficient, AI-enabled exascale supercomputers.

NVIDIA is making available to the Arm® ecosystem its full stack of AI and HPC software -- which accelerates more than 600 HPC applications and all AI frameworks -- by year's end. The stack includes all NVIDIA CUDA-X AI™ and HPC libraries, GPU-accelerated AI frameworks and software development tools such as PGI compilers with OpenACC support and profilers.

Once stack optimization is complete, NVIDIA will accelerate all major CPU architectures, including x86, POWER and Arm.

"Supercomputers are the essential instruments of scientific discovery, and achieving exascale supercomputing will dramatically expand the frontier of human knowledge,'' said Jensen Huang, founder and CEO of NVIDIA. "As traditional compute scaling ends, power will limit all supercomputers. The combination of NVIDIA's CUDA-accelerated computing and Arm's energy-efficient CPU architecture will give the HPC community a boost to exascale.''

"Arm is working with our ecosystem to deliver unprecedented compute performance gains and exascale-class capabilities to Arm-based SoCs,'' said Simon Segars, CEO of Arm. "Collaborating with NVIDIA to bring CUDA acceleration to the Arm architecture is a key milestone for the HPC community, which is already deploying Arm technology to address some of the world's most complex research challenges.''

According to the Green500 list released today, NVIDIA powers 22 of the world's 25 most energy-efficient supercomputers.

Key factors making this possible are: the ability of NVIDIA GPU-powered supercomputers to offload heavy processing jobs to more energy-efficient parallel processing CUDA® GPUs; NVIDIA's collaboration with Mellanox to optimize processing across entire supercomputing clusters; and NVIDIA's invention of SXM 3D-packaging and NVIDIA NVLink® interconnect technology, which allows for extremely dense scale-up nodes.

NVIDIA's support for Arm-based HPC systems builds on more than 10 years of collaboration. NVIDIA uses Arm for several of its system on a chip products available for portable gaming, autonomous vehicles, robotics and embedded AI computing.

Strong Ecosystem Support
World leaders of the HPC industry and Arm ecosystem, including supercomputing centers and systems providers and system-on-a-chip manufacturers, have voiced their support.

"Our customers are looking for high-performance, Arm-based processors to run their most demanding workloads. We are thrilled that NVIDIA is moving CUDA and the rich ecosystem built around NVIDIA to Arm. This will accelerate our work in building out the software ecosystem for Arm-based servers and enable breakthrough Ampere platforms with NVIDIA GPUs for efficiency and performance.''

-- Renee James, chairman and CEO, Ampere Computing

"Atos is a pioneer in the ARM ecosystem through the Montblanc project and an Arm computing blade design available for the exascale supercomputers, BullSequana X. We are really pleased to support NVIDIA's major announcement to turbo-boost Arm, which will accelerate the convergence of the AI and simulation worlds while optimizing energy efficiency.''

-- Pierre Barnabé, senior executive vice president, head of Big Data & CyberSecurity Division, Atos

"At Cray our vision for the exascale era is systems that integrate AI and analytics with modeling and simulation, systems that enable and often require a diversity of processor architectures and systems that are built for data-intensive workloads required in science, engineering and digital transformation. We are excited to partner with NVIDIA to help realize this vision in our supercomputers by leveraging their CUDA and CUDA-X HPC and AI software stack to the Arm platform and integrating it closely with our Cray system management and programming environment (compilers, libraries and tools) already enabled to support Arm processors across our XC and future Shasta supercomputers.''

-- Peter Ungaro, president and CEO, Cray

"EuroHPC enables European collaboration on high performance computing to advance research, innovation and industrial growth. We are very interested in testing NVIDIA's GPU-accelerated computing platform for HPC and AI on Arm as a potential building block for future pre-exascale solutions.''

-- Kimmo Koski, managing director, CSC

"The European Processor Initiative aims to endow the European Union with its own high-end, low-power, general purpose and accelerator solutions. EPI and SiPearl, its industrial hand, consider very positively the new possibilities offered by NVIDIA. The combination between the EPI Arm-based microprocessor and NVIDIA accelerator could make a perfect match for equipping building blocks in the future European exascale modular supercomputers.''

-- Philippe Notton, general manager, EPI

"Both NVIDIA and Arm leverage technologies that offer high performance computing customers greater levels of energy efficiency. NVIDIA's support for Arm complements our latest developments on the HPE Apollo 70, an Arm-based, purpose-built HPC system, and now, NVIDIA GPU-enabled. With the HPE Apollo 70 supporting a 2U GPU tray and multiple energy-efficient cooling options, we can further help the HPC industry address increasingly unsustainable levels of power consumption.''

-- Bill Mannel, vice president and general manager of HPC and AI, HPE

"The Jülich Supercomputing Centre is driving developments at the forefront of supercomputing and establishing modular technologies to make the best resources available most effectively to researchers in Europe to help them solve the world's greatest challenges. Particularly in view of the exascale systems planned for the coming years and the rise of large-scale AI calculations, NVIDIA's support of the Arm processor is a very exciting development, which is essential for the establishment of true modularity for supercomputers and composable data centers of the future. It will help to advance supercomputing in Europe.''

-- Thomas Lippert, director, Jülich Supercomputing Centre

"We are excited to work with NVIDIA and server OEMs to couple the CUDA-X platform and NVIDIA GPUs with the Marvell ThunderX2 family of server processors. The combination of ThunderX2's best-in-class 64-bit Armv8 performance and NVIDIA GPUs offers breakthrough levels of energy efficiency and application performance, enabling world-class HPC and AI solutions for exascale computing.''

-- Matt Murphy, president and CEO, Marvell

"As the leader in HPC networks, our InfiniBand and Ethernet technologies connect many of the largest supercomputers in the world, including the first generation of Arm-based systems. We look forward to continuing to work with NVIDIA to deploy our advanced 200Gb/s HDR and computational-networking technologies to optimize HPC and artificial intelligence workloads and to super-connect the next generation of Arm-based supercomputers.''

-- Eyal Waldman, founder and CEO, Mellanox Technologies

"We have been a pioneer in using NVIDIA GPUs on large-scale supercomputers for the last decade, including Japan's most powerful ABCI supercomputer. At Riken R-CCS, we are currently developing the next-generation, Arm-based, exascale Fugaku supercomputer and are thrilled to hear that NVIDIA's GPU acceleration platform will soon be available for Arm-based systems.''

-- Satoshi Matsuoka, director, Riken Center for Computational Sciences, and professor, Tokyo Institute of Technology

About NVIDIA
NVIDIA's (NASDAQ: NVDA) invention of the GPU in 1999 sparked the growth of the PC gaming market, redefined modern computer graphics and revolutionized parallel computing. More recently, GPU deep learning ignited modern AI — the next era of computing — with the GPU acting as the brain of computers, robots and self-driving cars that can perceive and understand the world. More information at http://nvidianews.nvidia.com/.

Certain statements in this press release including, but not limited to, statements as to: support for Arm CPU architecture enabling a new path to exascale supercomputing; NVIDIA's software stack being optimized for Arm and enabling a new wave of exascale supercomputers; the availability of NVIDIA's full stack of software to Arm and its ecosystem; NVIDIA accelerating all major CPU architectures; the world's supercomputers becoming power constrained; NVIDIA's support for Arm being a step forward to provide high performance computing a more power-efficient future; the benefits, performance, impact and abilities of NVIDIA's technologies, including NVIDIA GPU-powered supercomputers; how NVIDIA powers the energy-efficient supercomputers; the Arm community requesting NVIDIA CUDA acceleration and the announcement thrilling the world's HPC ecosystem; customers looking forward to Arm-based processors, NVIDIA moving CUDA to Arm and it enabling breakthrough Ampere platforms; computing at exascale enabling a new world of AI and science; ATOS solutions with NVIDIA GPUs and Arm architecture providing energy efficiencies and delivering science at exascale and beyond; EuroHPC enabling collaboration and their interest in testing NVIDIA's platform on Arm as a building block for future pre-exascale solutions; Cray's vision for the exascale era and their excitement to and how they will work with NVIDIA to realize this vision; the combination of EPI microprocessors and NVIDIA equipping the building blocks of future European exascale modular supercomputers; the combination of Arm-based CPU processors with NVIDIA GPUs being a game-changer for the industry; NVIDIA and Arm helping to contend with increasing levels of power consumption; the Jülich Supercomputing Center driving developments at the forefront of supercomputing and establishing modular technologies to help solve the world's greatest challenges; the excitement for NVIDIA's support of the Arm processor and it helping to advance supercomputing in Europe; Arm and NVIDIA computing enabling Lenovo to target solutions as the world races to exascale; Arm and NVIDIA enabling world-class HPC and AI solutions; Mellanox looking forward to working with NVIDIA to deploy technologies to optimize HPC and AI workloads and to connect the next generation of Arm-based supercomputers; and Riken's development of the next generation Arm-based exascale supercomputer and excitement about NVIDIA's GPU platform being available for Arm-based systems are forward-looking statements that are subject to risks and uncertainties that could cause results to be materially different than expectations. Important factors that could cause actual results to differ materially include: global economic conditions; our reliance on third parties to manufacture, assemble, package and test our products; the impact of technological development and competition; development of new products and technologies or enhancements to our existing product and technologies; market acceptance of our products or our partners' products; design, manufacturing or software defects; changes in consumer preferences or demands; changes in industry standards and interfaces; unexpected loss of performance of our products or technologies when integrated into systems; as well as other factors detailed from time to time in the most recent reports NVIDIA files with the Securities and Exchange Commission, or SEC, including, but not limited to, its annual report on Form 10-K and quarterly reports on Form 10-Q. Copies of reports filed with the SEC are posted on the company's website and are available from NVIDIA without charge. These forward-looking statements are not guarantees of future performance and speak only as of the date hereof, and, except as required by law, NVIDIA disclaims any obligation to update these forward-looking statements to reflect future events or circumstances.

Media Contacts

Kristin Bryson
+1-203-241-9190
kbryson@nvidia.com