

NVIDIA going full stack for ARM boosts supercomputing presence

June 18 2019, by Nancy Cohen



Credit: CC0 Public Domain

NVIDIA and ARM make one power couple for supercomputing. NVIDIA has announced its chips will work with ARM processors. Outside observers got busy earlier this week assessing why this was a big deal to empower both companies and the effort to explain was not at all difficult.

Paul Lilly, *HotHardware*, found this a win-win for NVIDIA and ARM. "As it pertains to the former, once stack [optimization](#) is complete, the company can boast support for every major CPU platform, including ARM, IBM Power, and of course x86. And for ARM, access to NVIDIA's CUDA stack is a major boost in GPU horsepower, and a selling point to clients."

ARM has been gaining traction in a number of HPC centers, so NVIDIA's move to embrace this processor architecture was not a surprise, said Timothy Prickett [Morgan](#) in *The Next Platform*.

The accent can be on the word *full*. NVIDIA's full stack of AI and HPC software is being made available to the ARM ecosystem. The way that Timothy Prickett Morgan saw it, a GPU compute software stack as complex as an [operating system](#) was coming to ARM.

What does that include? (1) all of its CUDA X AI and HPC libraries, (2) GPU-accelerated AI frameworks, and (3) software development tools (e.g., PGI compilers with OpenACC support).

NVIDIA said the stack of AI and HPC software, accelerating more than 600 HPC applications and all AI frameworks, is being made available by year's [end](#).

CUDA stands for Compute Unified Device Architecture. Chris Mellor in *The Register* wrote, "For the uninitiated, CUDA is a programming interface for turning Nvidia graphics processors into math accelerators for neural network applications, simulations, and other [vector-intensive](#) software."

Gareth Halfacree talked about it in *bit-tech*. It allows for general-purpose and typically non-graphics [workloads](#) to be executed on an NVIDIA graphics processor, and for workloads that parallelize well, he said, "this

can result in a dramatic boost to performance."

CUDA was partially supported on ARM before, he added, but the difference is that now "Nvidia has confirmed it will port its entire stack of AI- and [high performance](#) compute-focused software to Arm by the end of the year."

NVIDIA's news release also affirmed the two companies were no strangers in working on support for HPC systems, with over 10 years of collaboration. So, *The Register* might be forgiven for yawning when they wrote their headlines. "Nvidia 'brings CUDA to Arm' – Translation: We're still doing a thing we've already done but now doing more of it."

NVIDIA has a long history of collaborating with ARM, said *HotHardware*, so this announcement was not surprising, even if "a bit overdue."

Mellor said, "Anyone who has used one of, say, Nvidia's automotive and internet-of-things development boards, which sport system-on-chips that feature a mix of Arm Cortex CPU cores and Nv CUDA-based GPU engines, will know Nvidia already develops and distributes Arm-compatible CUDA libraries and toolkits."

Stephen Nellis, Reuters, meanwhile, noted how the newest move could deepen NVIDIA's push into systems that are used for modeling issues such as climate change predictions. Already, he said, "researchers now also use its chips inside data centers to speed up artificial intelligence computing work such as [training](#) computers to recognize images."

Halfacree said Nvidia confirmed it was making the full CUDA stack available on ARM "to follow industry interest in using the low-power chips in future exascale systems."

Nellis noted that this move to work with ARM on supercomputers follows a deal to buy Israeli firm Mellanox Technologies. "Mellanox makes [high-speed](#) networking chips that help stitch together many smaller computers into a larger one and is found in some of the world's most powerful supercomputers."

© 2019 Science X Network

Citation: NVIDIA going full stack for ARM boosts supercomputing presence (2019, June 18)
retrieved 25 April 2024 from

<https://techxplore.com/news/2019-06-nvidia-full-stack-arm-boosts.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.