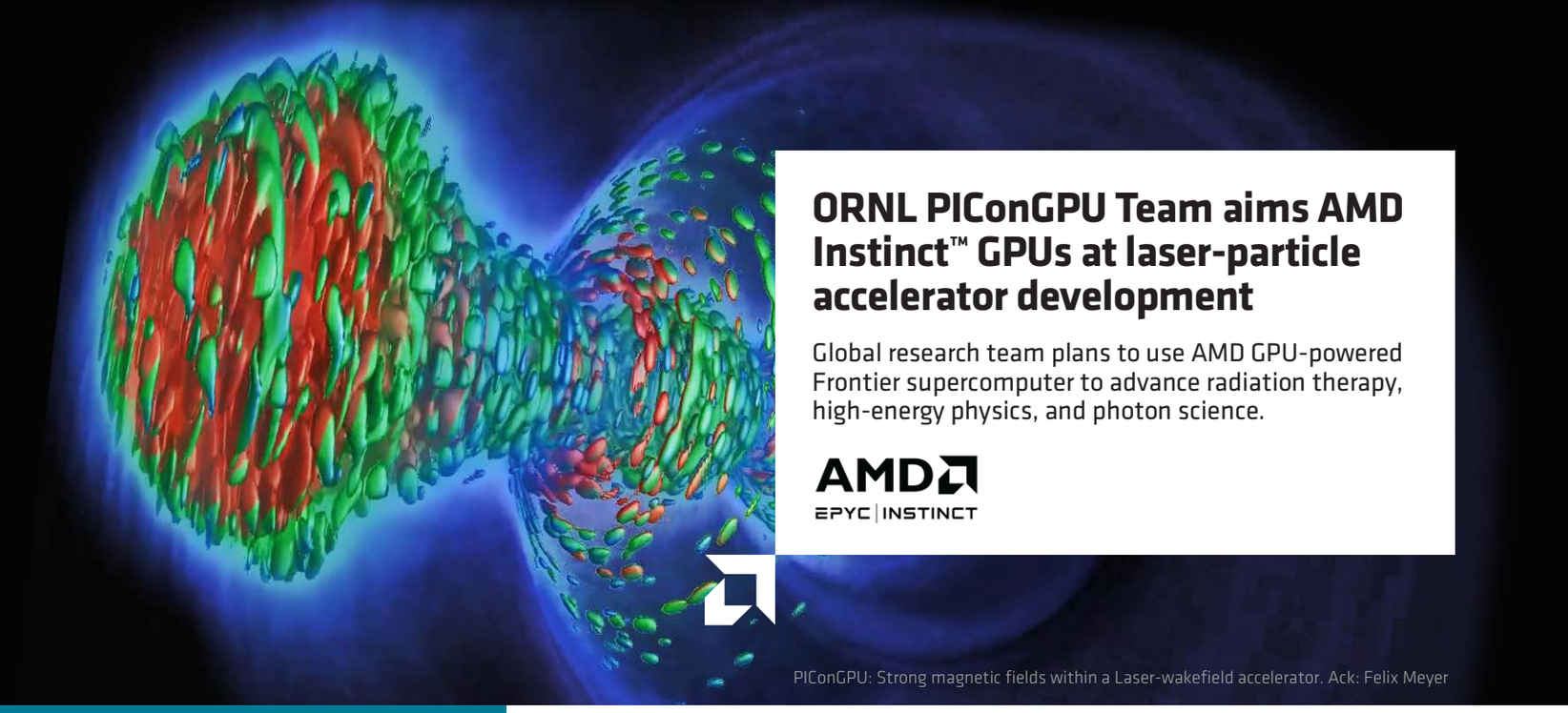


PIConGPU: Strong magnetic fields within a Laser-wakefield accelerator. Ack: Felix Meyer

ORNL PIConGPU Team aims AMD Instinct™ GPUs at laser-particle accelerator development

Global research team plans to use AMD GPU-powered Frontier supercomputer to advance radiation therapy, high-energy physics, and photon science.

AMD EPYC | INSTINCT

CUSTOMER

Oak Ridge National Laboratory

INDUSTRY

Scientific Research

CHALLENGES

Simulate large-scale simulations over millions of time-steps without monopolizing the resources of the world's leading high-power computers for weeks at a time

SOLUTION

Leverage the open source ROCm platform to tap the immense compute capacity of the forthcoming Frontier exascale supercomputer powered by AMD Instinct MI200 series accelerators

RESULTS

A simulation that took two months on the previous Summit system now takes less than two weeks on the AMD GPU-powered Frontier system while allowing the team to run several 10-million time-step simulations

AMD TECHNOLOGY AT A GLANCE

AMD Instinct MI200 accelerator

TECHNOLOGY PARTNER

Hewlett Packard Enterprise

In preparation for the AMD Instinct™ GPU-powered Frontier exascale supercomputer, Oak Ridge National Laboratory (ORNL) created the Center for Accelerated Application Readiness (CAAR). The CAAR program sought grand challenges in science that would test the computational prowess of the new Frontier supercomputer and provide guidance to application developers.

The Frontier supercomputer design is the result of a collaboration between the U.S. Department of Energy, ORNL, HPE, and AMD. Expected to deliver more than 1.5 exaflops of peak processing power. Frontier leverages the AMD Infinity Architecture for high-speed data transfer between GPUs, CPUs, and memory to help ensure low latency at the scale of this unique system.

Eight research projects gained early access to Frontier. Among them is an international team, led by Sunita Chandrasekaran , Associate Professor at the Computational Research Programming Lab at the University of Delaware, which includes members from several U.S. institutions as well as Germany's Helmholtz Zentrum Dresden Rossendorf (HZDR). The team uses PIConGPU, a critical, open-source simulation tool used to study plasma and laser-plasma physics. Their research objective is to develop advanced particle accelerators for radiation therapy, high-energy physics, and photon science.

Achieving continuous performance increases throughout development

PIConGPU code is of particular interest to CAAR because it was purpose-built to take advantage of the many GPU acceleration cores available in the newest generations of supercomputers. This heritage has enabled PIConGPU to keep pace with the generational iterations of AMD GPUs used in Frontier test nodes as the system scales on its journey to the AMD MI200 Instinct series accelerators being deployed in the completed production supercomputer.

With this in mind, CAAR further challenged the PIConGPU team to optimize their code and techniques to achieve a 4X performance improvement over what was possible with Summit, ORNL's previous generation supercomputer. The PIConGPU teams use a tool from the AMD ROCm™ open software ecosystem called AMD rocProf to build a roofline model, a performance analysis of how many integer/floating computations are performed relative to the amount of data moved from one place to another.

A simulation that took two months on the previous Summit system might take less than two weeks on the [AMD AMD Instinct MI200 GPU-powered] Frontier system…"

Sunita Chandrasekaran, Associate Professor, University of Delaware

"Over the past two years, we have seen consistent improvements in PIConGPU case studies that will run on Frontier as they moved from MI25 to MI60 to MI100," said Chandrasekaran. For instance, a case study optimized scientifically and computationally showed a 1.11x speedup from ROCm 4.1 to 4.2 on MI100."

Speaking in broad terms, Chandrasekaran continued, "At every step of the way, we were able to say that there is a significant increase in performance. A simulation that took two months on the previous Summit system might take less than two weeks on the [AMD AMD Instinct MI200 GPU-powered] Frontier system and could let us run several 10-million time-step simulations."

Fast simulation and precision are the keys to productive outcomes

Exascale systems such as Frontier are crucial to exploring advanced laser-plasma accelerators that reach electron energies far beyond 10GeV. These plasma accelerators require more time steps in simulations, higher resolutions, and more compute-intensive numerical algorithms to maintain the desired accuracy of results. According to Chandrasekaran, "Running large-scale simulations over millions of time steps is challenging. Monopolizing all the resources of world-leading high-power computers for weeks at a time is tough to justify."

"At every step of the way, we were able to say that there is an increase in performance."

Sunita Chandrasekaran, Associate Professor, University of Delaware

Physicist Dr. Alexander Debus from the Institute of Radiation Physics at HZDR added, "Frontier's vast compute capabilities enable us to allocate sufficient compute resources to speed up simulations for faster completion of advanced laser-plasma accelerator models at high-resolution and with many time steps. These models are designed to reach towards the highest particle energies. The results will provide us with critical insights on the laser-plasma and electron beam dynamics that will continue to guide our theoretical and experimental work on these compact particle accelerators."

Powered by the 2nd Generation AMD CDNA™ architecture, AMD Instinct MI200 series accelerators will allow researchers to process parallel codes up to 47.9TFLOPS of peak FP64 performance. AMD's Matrix Core technology also supports optimized BF16, INT4, INT8, FP16, FP32, and FP32 matrix capabilities to enable a full range of mixed-precision operations when working with large models.

Chandrasekaran said, "We have kernels within PIConGPU that are memory bound for FP64 and FP32. One is *move and mark*, which has FP64. The other, *compute current*, is FP32. The ability to measure single and double precision performances of such kernels could be very useful to us."

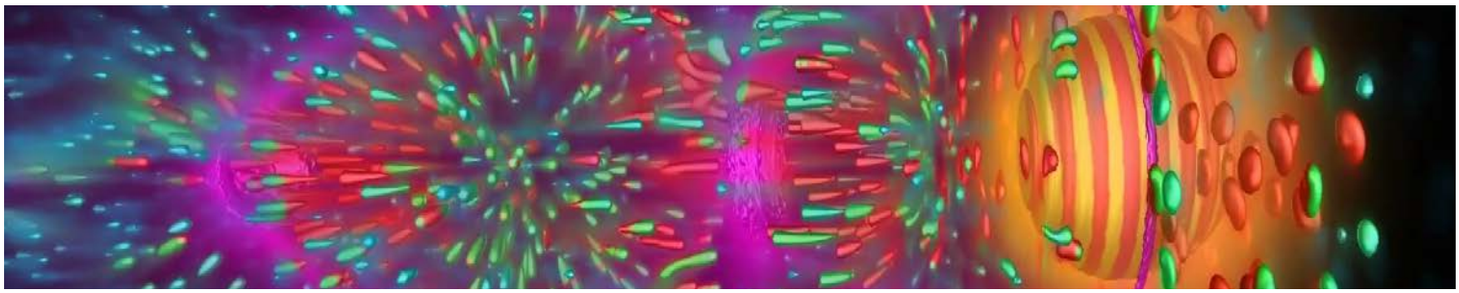
Open source is the foundation of scientific collaboration

One of the significant changes between the previous Summit system and Frontier is the shift to an entirely open-sourced environment. Unlike the previous, proprietary CUDA-based architecture, AMD ROCm™ open software platform enables open portability across accelerator vendors and architectures. Chandrasekaran and her team took advantage of the ROCm native open-source Heterogeneous Computing Interface for Portability, or HIP, to port and optimize PIConGPU code for the AMD GPUs. ROCm provides coding flexibility that enables developers to select and employ the best tools and run-time language for specific applications while making the most of the hardware performance to meet even the most demanding workloads.

"Because AMD's platform is open-source, it is much easier to do pull requests to get help during development," said Chandrasekaran. "As a software stack community, we can pinpoint issues, and then the AMD software team can work their magic. For example, we needed to create visualization videos and images of PIConGPU on AMD GPUs. We had submitted a feature request with AMD, and they put a release out for us to pull. We then built the ISAAC plugin needed to build the visualization software on AMD GPUs. This workflow is only possible because the environment is open source."

The PIConGPU team itself has been open source from the get-go. Chandrasekaran said, "PIConGPU uses a modular open-source software stack, so working with another open-source software like AMD HIP is tremendous. One key part, Alpaka, has the ability to target multiple platforms from conventional CPUs, GPUs, and even FPGAs by offering backends for HIP, CUDA, OpenMP host, OpenMP offloading, and OpenACC. To that end, AMD's commitment to open source is a huge plus."

Even before the PIConGPU team has access to the fully realized Frontier system, its work impacts the broader world of computer science. Chandrasekaran points out, "Novelty matters. Research matters. And publication matters." To that end, the team has already published papers and technical reports and has plans to deliver critical insights to upcoming conferences.



About Oak Ridge National Laboratory (ORNL)

Oak Ridge National Laboratory is the largest US Department of Energy science and energy laboratory, conducting basic and applied research to deliver transformative solutions to compelling problems in energy and security. ORNL's diverse capabilities span a broad range of scientific and engineering disciplines, enabling the Laboratory to explore fundamental science challenges and to caary out the research needed to accelerate the delivery of solutions to the marketplace. Within ORNL, the Oak Ridge Leadership Computing Facility was established to accelerate scientific discovery and engineering progress by providing world-leading computational performance and advanced data infrastructure, including Summit, the second most powerful supercomputer in the world. For more information, visit ornl.gov.

About AMD

For over 50 years AMD has driven innovation in high-performance computing, graphics, and visualization technologies—the building blocks for gaming, immersive platforms, and the data center. Hundreds of millions of consumers, leading Fortune 500 businesses, and cutting-edge scientific research facilities around the world rely on AMD technology daily to improve how they live, work, and play. AMD employees around the world are focused on building great products that push the boundaries of what is possible. For more information about how AMD is enabling today and inspiring tomorrow, visit amd.com/Instinct.

All performance and cost savings claims are provided by Oak Ridge National Laboratory (ORNL) and have not been independently verified by AMD. Performance and cost benefits are impacted by a variety of variables. Results herein are specific to ORNL and may not be typical. GD-181