Special Analysis

Exploring Select GPU Software Options

Mike Heroux and Tom Sorensen
May 2025

OVERVIEW

Industry publications often focus on GPU hardware's raw performance potential. While competitive performance is essential in assessing GPUs' capabilities for high-performance and AI computing, software capabilities are equally important. This article explores software capabilities for GPUs from three major GPU software providers: NVIDIA, AMD, and Intel.

The NVIDIA Software Ecosystem: CUDA

NVIDIA provides a software ecosystem that supports AI, simulation for gaming, scientific computing, cloud and enterprise, edge computing, and autonomous systems, with the Compute Unified Device Architecture (CUDA) underpinning everything.

CUDA is a parallel computing software ecosystem that enables developers to leverage GPU parallelism for general-purpose computing. By providing a standardized framework for NVIDIA GPU acceleration, CUDA has become NVIDIA's foundation for high-performance computing (HPC), artificial intelligence (AI), and deep learning applications.

Building on CUDA, NVIDIA has invested in AI and deep learning by developing software frameworks optimized for its GPUs:

- cuDNN (CUDA Deep Neural Network library) accelerates deep learning model training and inference
- cuDNN underpins both PyTorch and TensorFlow, the two most popular AI frameworks
- Libraries such as cuBLAS (optimized linear algebra), cuSPARSE (sparse matrix operations) and NCCL (multi-GPU communication) provide researchers with tools for accelerating complex simulations

The AMD Software Ecosystem: HIP/ROCm

AMD's Heterogeneous-Computing Interface for Portability (HIP) is designed as a drop-in replacement for CUDA, allowing developers to port CUDA applications to AMD GPUs with minimal code changes using hipify tools. HIP provides a programming model nearly identical to CUDA, supporting parallel computing constructs such as kernels, memory management, and thread execution while abstracting the underlying hardware differences. This is targeted to support organizations that have invested in CUDA-based applications to migrate to AMD hardware without a complete rewrite of their codebase.

FIGURE 2

Software Ecosystems for GPUs

	On NVIDIA GPUs	On AMD GPUs	On Intel GPUs
CUDA code	CUDA is officially supported by NVIDIA, providing a proprietary software ecosystem for their GPUs	CUDA is not available on AMD GPUs, but CUDA source code can be translated to HIP using hipify tools	CUDA is not available on Intel GPUs
HIP/ROCm code	HIP/ROCm is open source and used on NVIDIA GPUs, but NVIDIA does not provide support. Performance and compatibility issues are a barrier for production use	HIP/ROCm is officially supported by AMD, providing an optimized version of the open-source software for their GPUs	HIP/ROCm is open source, but neither AMD nor Intel support it on Intel GPUs
oneAPI code	oneAPI is an open standard, but neither NVIDIA nor Intel support oneAPI on NVIDIA GPUs	oneAPI is an open standard, but neither Intel nor AMD support oneAPI on AMD GPUs	oneAPI is officially supported by Intel, providing an optimized implementation of the open standard for Intel GPUs

Source: Hyperion Research, 2025

Note: NVIDIA, AMD, and Intel each develop their own ecosystem of compilers, libraries, and tools. NVIDIA's ecosystem is exclusive to its hardware. AMD and Intel both promote cross-platform approaches but for production use, their ecosystems are not officially supported on other vendors' hardware.

In addition to HIP, AMD's ROCm (Radeon Open Compute) product suite offers a set of libraries that closely mirror NVIDIA's CUDA libraries. Libraries such as hipBLAS, hipSPARSE, and RCCL provide functionality equivalent to their CUDA counterparts (cuBLAS, cuSPARSE, and NCCL), ensuring that applications requiring high-performance linear algebra, sparse matrix operations, or communication can run efficiently on AMD GPUs with minimal modification. Similarly, MIOpen, AMD's deep learning acceleration library, is an alternative to cuDNN, enabling optimized training and inference of deep neural networks.

The HIP/ROCm software ecosystem is open source and could, in principle, work on NVIDIA and Intel GPUs. By offering these nearly functionally identical libraries, AMD provides a future alternative to NVIDIA's software ecosystem, seeking to reduce vendor lock-in and encouraging broader adoption of AMD GPUs in AI, HPC, and scientific computing.

The Intel Software Ecosystem: oneAPI

Intel provides software for its GPUs via oneAPI, which provides a unified, open, and standards-based programming model designed to support multiple hardware architectures, including CPUs, GPUs, and FPGAs. Intel's Data Parallel C++ (DPC++) is built on SYCL, an open standard for parallel programming and is part of oneAPI, enabling developers to write code that runs on Intel's GPUs while targeting portability across different hardware platforms. Intel Data Center GPUs, such as Ponte Vecchio and Gaudi, are designed for AI training, deep learning, and HPC applications using oneAPI.

Intel provides oneDNN (like cuDNN and MIOpen), an optimized deep learning library for AI model training and inference, and oneMKL (analogous to cuBLAS and related libraries) for high-performance mathematical operations. These libraries offer capabilities similar to NVIDIA's software stack, enabling AI and scientific computing workloads to leverage Intel's hardware efficiently. Like HIP/ROCm, oneAPI is designed to provide an alternative to CUDA, making it easier for developers to write applications for Intel GPUs.

ANALYST COMMENT

The NVIDIA CUDA software ecosystem is the largest and most widely used GPU software stack. Porting CUDA source code to HIP can be straightforward, using hipify tools that automatically translate CUDA to HIP. However, this is effectively a one-way translation since NVIDIA does not support HIP use on its hardware. AMD's HIP/ROCm software is open-source, and Intel's DPC++/oneAPI follows open standards. Both are designed to be capable of running on any GPU.

However, none of the GPU providers support any software environment except their own. Therefore, it is not feasible to maintain a single source code base using vendor-supported software ecosystems that works well across more than one GPU platform.

NVIDIA, AMD, and Intel all provide strong core software capabilities for AI, scientific computing, and HPC. At the same time, there is a clear difference in scope and presence:

- NVIDIA CUDA has been available for nearly 20 years. CUDA code is embedded in many software products, and CUDA libraries and tools provide significant vertical integration, performance, correctness debugging tools, and support for industry standard functionality.
- AMD HIP/ROCm has been around for nearly 10 years. It has a growing collection of libraries and tools, even if the general community consensus is that more needs to be done to reach parity with NVIDIA.
- Intel, however, is a relative newcomer to developing high-end GPU capabilities. Leveraging its CPU software libraries and tools resources provides them with a good starting point, but the general consensus is that there are significant gaps in GPU capabilities beyond core features. In addition, concerns about the future availability of Intel high-end GPUs leads to some reluctance to invest in using Intel's oneAPI ecosystem for now.

All three major GPU vendors, NVIDIA, AMD, and Intel, provide software ecosystems that facilitate the optimized use of their devices using standard programming languages and an extensive collection of optimized libraries. CUDA is supported only on NVIDIA GPUs. HIP/ROCm is open source but is not well supported on any hardware except AMD's. Similarly, oneAPI is affiliated with the SYCL open standard but is officially supported only on Intel hardware.

Practically speaking, there is no portability possible using vendor-endorsed programming environments. While not highlighted in this article, performance portability is possible using vendor-neutral approaches such as parallel programming features in modern C++, embedded markup via OpenMP, and C++ libraries and tools of the Kokkos project. These approaches are promising but are still emerging and represent a small portion of the overall GPU software base.

Currently, most production GPU applications rely heavily on the vendor software stacks (CUDA, HIP/ROCm, oneAPI). Until robust portable GPU programming approaches emerge, NVIDIA's software advantage will remain an important aspect of its market competitiveness in addition to its hardware.

About Hyperion Research, LLC

Hyperion Research provides data-driven research, analysis and recommendations for technologies, applications, and markets in high performance computing and emerging technology areas to help organizations worldwide make effective decisions and seize growth opportunities. Research includes market sizing and forecasting, share tracking, segmentation, technology, and related trend analysis, and both user & vendor analysis for multi-user technical server technology used for HPC and HPDA (high performance data analysis). Hyperion Research provides thought leadership and practical guidance for users, vendors, and other members of the HPC community by focusing on key market and technology trends across government, industry, commerce, and academia.

Headquarters

365 Summit Avenue
St. Paul, MN 55102
USA
612.812.5798
www.HyperionResearch.com and www.hpcuserforum.com