

Special Analysis

Forecast Update: GPU and Accelerator Growth in HPC

Melissa Riddle, Tom Sorensen, and Earl Joseph
February 2023

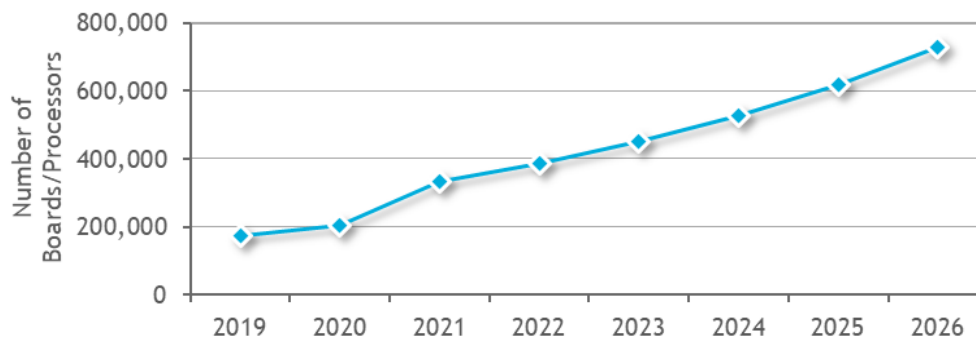
HYPERION RESEARCH OPINION

Historically, GPUs and accelerators have been growing roughly 26% per year (a five-year CAGR). Over the next five years, installations are expected to increase by at least 17% a year, greatly outpacing the revenue growth for HPC servers over the same forecast period. As AI workloads become an increasingly important part of overall HPC workload, both enabling new innovative techniques to solve new problems and to supplement traditional modelling and simulation workloads, GPU usage will likely continue to have strong growth. The figure below shows Hyperion Research's projection for accelerators in the HPC market.

Accelerators initially had a slow HPC adoption curve but have been steadily gaining momentum with GPUs emerging as the most popular option. Today, GPUs have become nearly synonymous with accelerators in the HPC space, given that GPUs comprise the overwhelming majority of accelerators in HPC clusters worldwide. Although originally designed as gaming processors for graphics rendering, GPUs were found to be well suited to a wide range of HPC applications and critical underlying AI functions, such as matrix multiplication, driving impressive performance gains. As a result, many HPC sites across a broad range of verticals are now using accelerators to speed up a larger portion of their increasingly AI-based workloads.

FIGURE 1

Technical Computer Market: GPUs/Accelerators/FPGA and Other Types of Accelerators



Note: High growth is strongly driven by the supercomputers segment -- not any one particular large system, but a variety of supercomputer systems.

Source: Hyperion Research, January 2023

GPU AND ACCELERATOR ADOPTION IN HPC

Note: The tables and figures included in this document are shown in number of boards/processors sold each year.

The use of accelerators, and specifically GPUs, has grown rapidly over the last decade (see table below). Many HPC sites and workloads are increasingly using accelerators in system procurements to support both traditional HPC and HPC-enabled AI workloads. NVIDIA is the historical GPU leader; however, AMD and Intel have either released GPU product lines in the past, or are about to bring a new GPU to market. Although GPUs constitute the majority of accelerator boards installed in HPC, others (such as FPGAs) are also gaining popularity. FPGAs (field-programmable gate arrays) can be more complex to integrate and program when compared to GPUs but are being adopted in several HPC areas, such as bio-sciences and finance. Xilinx is the major provider of FPGA technology for HPC although Intel is also increasing in adoption. One of the key differentiators of FPGAs is their ability to be reprogrammed for different applications, providing users with more flexibility in the utilization of such technology.

TABLE 1

Technical Computer Market: GPUs/Accelerators/FPGA and Other Types of Accelerators

	2014	2015	2016	2017	2018	2019	2020	2021	CAGR '16-'21
Total Attached Processors or Boards	65,092	88,123	106,988	109,501	181,694	175,394	202,293	334,037	25.6%
Yearly Growth		35.4%	21.4%	2.3%	65.9%	-3.5%	15.3%	65.1%	

Note: Several very large systems were installed in 2018.

Source: Hyperion Research, January 2023

The table below shows the four major competitive segments tracked in the HPC market. Each price band has experienced a five-year CAGR exceeding 15% over the last five years. The supercomputer segment, which represents the most expensive machines in the world priced at \$500,000 and above, consistently exhibits the highest use of accelerators due to the large-scale workloads typical for these systems as well as a greater emphasis on performance over price than in other HPC price bands. Accordingly, supercomputers had the highest growth between 2016 and 2021 (32.2%) and also represented nearly two thirds (3.9%) of all installed accelerator boards in 2021.

TABLE 2**GPU/Accelerator Historical Market by Competitive Segments**

	2014	2015	2016	2017	2018	2019	2020	2021	CAGR '16-'21
Supercomputer	34,517	42,368	52,802	49,239	112,592	95,070	110,474	213,409	32.2%
Divisional	11,928	18,695	25,166	32,896	41,506	47,660	49,992	51,533	15.4%
Departmental	10,319	17,479	20,613	18,847	19,952	23,333	29,117	43,136	15.9%
Workgroup	8,328	9,581	8,406	8,519	7,644	9,332	12,710	25,959	25.3%
Total	65,092	88,123	106,988	109,501	181,694	175,394	202,293	334,037	25.6%

Note: Several very large systems were installed in 2018.

Source: Hyperion Research, January 2023

Inhibitors to GPU/Accelerator Adoption

Despite their rapid growth, accelerators still face some barriers to adoption. Many workloads were not designed to exploit the capabilities of accelerated computing. Refactoring and modernizing to run on CPU-accelerator platforms can be complex and requires HPC talent that is in short supply. In addition, some machines have been designed, tuned, and optimized for a specific architecture that may not be well suited to take advantage of the particular performance capabilities of available accelerators. Furthermore, accelerators can be a costly option for an overall system design, requiring additional and expensive power and cooling capabilities.

FUTURE ACCELERATOR ADOPTION

Accelerators are anticipated to grow by at least 16.8% CAGR over the next five years, as can be seen in the table and figure below. Compared to the previous iteration of this forecast released in 2022, the expected yearly board counts for 2023-2025 have increased.

Over the forecast period, GPUs are anticipated to continue as the large majority of HPC accelerators. This is highlighted by pending high-profile exascale procurements, such as AMD's Radeon Instinct GPU featured in the Frontier system at Oak Ridge National Laboratory, Intel's Datacenter GPU Max (also referred to as Ponte Vecchio or Xe) to be deployed in the Aurora machine at Argonne National Laboratory, and many that use NVIDIA GPUs.

TABLE 3

Technical Computer Market: GPUs/Accelerators/FPGA and Other Types of Accelerators

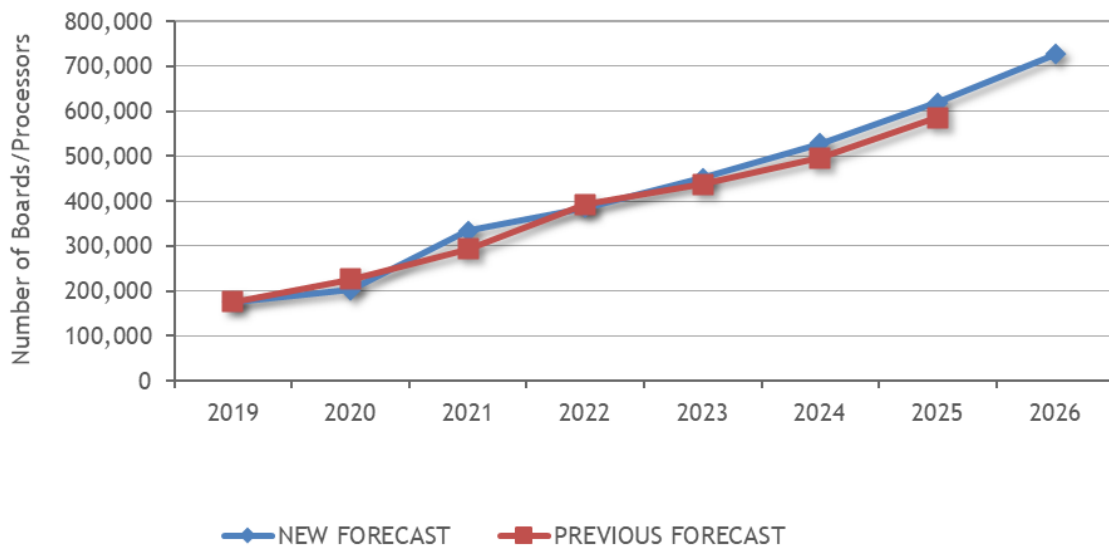
	2020	2021	2022	2023	2024	2025	2026	CAGR '21-'26
New Forecast	202,293	334,037	386,035	451,371	528,438	619,405	726,853	16.8%
Previous Forecast	226,588	293,185	392,876	437,284	496,030	585,315	-	-

Note: High growth is strongly driven by the supercomputers segment -- not any one particular large system, but a variety of supercomputer systems.

Source: Hyperion Research, January 2023

FIGURE 2

Technical Computer Market: GPUs/Accelerators/FPGA and Other Types of Accelerators



Source: Hyperion Research, January 2023

The table below shows the price segments forecast for accelerators. The large US DOE systems and others like it around the world will be responsible for major accelerator deployments over the period. In addition, the broader market beyond the top one percent of the highest-end machines is also increasing their use of GPUs. While the middle two segments of the market, departmental and

divisional, are increasing their adoption, the supercomputer segment shows the highest growth of accelerator adoption. The larger budgets for supercomputers allow users to incorporate more heterogeneity into their system (e.g., more accelerators and different types of accelerators), addressing a broad set of workload requirements. Workgroup is expected to have minimal growth, a reflection of the overall stagnation of this segment in the HPC market writ large.

TABLE 4

GPU/Accelerator Forecast by Competitive Segments

	2020	2021	2022	2023	2024	2025	2026	CAGR '21-'26
Supercomputer	110,474	213,409	255,024	304,753	364,180	435,195	520,058	19.5%
Divisional	49,992	51,533	58,362	67,262	77,519	89,341	102,965	14.8%
Departmental	29,117	43,136	47,989	53,388	59,394	66,076	73,509	11.3%
Workgroup	12,710	25,959	24,661	25,968	27,344	28,794	30,320	3.2%
Total	202,293	334,037	386,035	451,371	528,438	619,405	726,853	16.8%

Source: Hyperion Research, January 2023

GPU and Accelerator Use by Verticals

Government Labs unsurprisingly has the largest accelerator counts of any vertical, fueled in part by exascale and near-exascale systems. By 2026, Government Labs is projected to make up 40.6% of all installed accelerators. Bio-Sciences is expected to show the highest growth over the forecast period with a 21.3% CAGR. These verticals are both adopting and integrating AI for many of their standard workloads, but also working to modernize code to take advantage of the performance gains possible with accelerated computing. All verticals are expected to grow over the forecast period, and all except one (Mechanical Design) are expected to grow at rates exceeding 12% CAGR.

TABLE 5

GPU/Accelerators Forecast by Industry/Application Segments (in Number of Boards or Processors)

	2020	2021	2022	2023	2024	2025	2026	CAGR '21-'26
Bio-Sciences	6,399	11,247	13,837	17,222	21,463	25,158	29,522	21.3%
CAE	26,908	43,678	49,611	57,451	66,087	76,844	90,174	15.6%
Chemical Engineering	2,493	4,219	4,361	4,604	5,497	6,443	7,561	12.4%

TABLE 5**GPU/Accelerators Forecast by Industry/Application Segments (in Number of Boards or Processors)**

	2020	2021	2022	2023	2024	2025	2026	CAGR '21-'26
DCC	4,609	7,702	8,358	8,789	10,361	12,144	14,251	13.1%
Economics/Financial	5,942	9,911	10,902	11,719	13,789	16,163	18,966	13.9%
EDA	12,422	20,907	21,803	24,274	28,821	33,782	39,642	13.7%
Geosciences	4,950	7,684	8,347	9,174	10,096	11,834	13,887	12.6%
Mechanical Design	73	110	103	97	92	108	126	2.9%
Defense	10,788	18,011	20,350	23,019	26,814	31,430	36,882	15.4%
Government Lab	93,811	153,913	182,095	217,049	253,118	296,691	348,157	17.7%
University/Academic	19,829	33,465	39,514	46,784	55,937	66,185	77,666	18.3%
Weather	3,333	5,604	6,593	7,804	9,203	10,788	12,659	17.7%
Other	10,737	17,587	20,163	23,386	27,160	31,836	37,358	16.3%
Total	202,293	334,037	386,035	451,371	528,438	619,405	726,853	16.8%

Source: Hyperion Research, January 2023

FUTURE OUTLOOK

The accelerator market will be a high growth market for the foreseeable future, as utilization and integration of new HPC and AI applications and the desire to modernize a wide base of existing code continues to expand.

A majority of accelerators deployed over the next five years will likely be GPUs, building on their existing market leadership and established mind share. FPGAs and ASICs are each expected to increase their market presence over the next five years, as both small and large companies alike seek to provide users with alternatives to main-stream GPU options. Many users will also take advantage of the cloud to experiment with and use accelerators ahead of future procurements or for subsets of their workloads that may not yet warrant an on-premises deployment of accelerators.

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). We provide 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

Copyright Notice

Copyright 2023 Hyperion Research LLC. Reproduction is forbidden unless authorized. All rights reserved. Visit www.HyperionResearch.com to learn more. Please contact 612.812.5798 and/or email info@hyperionres.com for information on reprints, additional copies, web rights, or quoting permission.