

# Universal GPU Systems

## Multi-Architecture Flexibility, Future Proof Open-Standards Based Design



### Highly Flexible Platform

Supermicro's Universal GPU System is the industry's most advanced and flexible GPU server platform. Designed to deliver maximum compute power for large-scale AI deep learning and HPC workloads, this modular, open-standards based platform supports the industry's most popular GPU technologies in a variety of form factors and combinations both today and into the future.

### Designed for Demanding HPC and AI Workloads

The Supermicro Universal GPU platform has been designed from the ground up to support a combination of CPU and GPU configurations, allowing customization for specific HPC and AI workloads within the data center using a single platform.

- **High-performance computing** including energy, molecular dynamics, physics, computational chemistry and climate sciences
- **AI** for image and video detection/recognition, life sciences & drug discovery, autonomous driving and robotics

### All-New Multi-Standard Architecture

**Ultimate Modularity and Customization Options for AI and HPC Environments**

- All-new modular architecture designed for flexibility and future proofing
- Optimal storage, networking, power, and cooling mix for top performance configurations
- Supports all major GPUs and CPUs, maximum CPU and GPU performance configurations, and all major GPU form factors and interfaces
- 10 Gen4 PCIe slots for fast networking devices
- 4U chassis supports up to 10x 2.5" NVMe/SATA drives
- Optional 1U expansion module for improved thermal capacity (up to 700W GPUs) and 2x AIOMs for networking

### Supports Industry-Standard GPU Form Factors

The Supermicro Universal GPU platform is designed to work with a wide range of GPUs based on an open standards design. By adhering to an agreed-upon set of hardware design standards, such as Universal Baseboard (UBB) and OCP Accelerator Modules (OAM), as well as PCI-E and platform-specific interfaces, IT administrators can choose the GPU architecture best suited for their HPC or AI workloads.

NVIDIA HGX™



AMD Instinct™ and OAM



PCIe



Additionally, support for GPU interconnects including NVIDIA NVLink® and AMD Infinity Fabric™ facilitates ultra-fast GPU-to-GPU communication, reducing bottlenecks caused by traditional GPU interlinks. The ability to handle both high-speed fabrics and standard PCI-E-based GPUs is a first for the industry.

## Optional 1U Expansion Module

For the most demanding deployments, thermal performance and I/O of the standard 4U chassis can be further enhanced with an optional 1U add-on module. The resulting 5U configuration supports GPUs with TDP up to 700W and provides two additional OCP 3.0 AIOM slots for enhanced networking.



## Future Proof Design

The Universal GPU platform’s modular design and compatibility with a wide range of open-standard GPU interfaces allows for forward compatibility with next-generation GPU architectures without the need to replace or upgrade existing server infrastructure, thus reducing the cost of GPU upgrades. The platform has also been designed to easily accommodate next-generation CPU, memory and PCI-E technologies, reducing the cost and complexity of both development and deployment when such new technologies become available.



Universal GPU	AS-4124GQ-TNMI	SYS-420GU-TNXR
Processor Support	<ul style="list-style-type: none"> <li>Dual AMD EPYC™ 7003 Series processors</li> <li>Up to 64 cores, up to 280W TDP per socket<sup>1</sup></li> </ul>	<ul style="list-style-type: none"> <li>Dual 3rd Gen Intel® Xeon® Scalable processors</li> <li>Up to 48 cores, up to 270W TDP per socket<sup>1</sup></li> </ul>
Memory Slots & Capacity	<ul style="list-style-type: none"> <li>32 DIMM slots for DDR4-3200 MHz RDIMM/LRDIMM; up to 8TB registered ECC</li> </ul>	<ul style="list-style-type: none"> <li>32 DIMM slots for DDR4-3200 MHz RDIMM/LRDIMM; up to 8TB registered ECC or up to 12TB with Intel® Optane® Persistent Memory</li> </ul>
GPU Compatibility	<ul style="list-style-type: none"> <li>4x AMD MI250</li> <li>4x NVIDIA HGX A100-4</li> <li>Up to 500W TDP per GPU or 700W with optional 1U expansion module</li> </ul>	<ul style="list-style-type: none"> <li>4x NVIDIA HGX A100-4</li> <li>Up to 500W TDP per GPU or 700W with optional 1U expansion module</li> </ul>
I/O Ports	<ul style="list-style-type: none"> <li>2x 10GbE LAN</li> <li>1 RJ45 Dedicated IPMI LAN port</li> <li>2 USB 3.0 ports (rear)</li> <li>1 VGA and 1 COM port header</li> </ul>	<ul style="list-style-type: none"> <li>2x 10GbE LAN</li> <li>1 RJ45 Dedicated IPMI LAN port</li> <li>2 USB 3.0 ports (rear)</li> <li>1 VGA and 1 COM port header</li> </ul>
BIOS	<ul style="list-style-type: none"> <li>AMI 256 Mb (32 MB) SPI Flash ROM</li> </ul>	<ul style="list-style-type: none"> <li>AMI 256 Mb (32 MB) SPI Flash ROM</li> </ul>
System Management	<ul style="list-style-type: none"> <li>Integrated IPMI 2.0 plus KVM with dedicated LAN</li> <li>Supermicro Server Manager (SSM) and Supermicro Update Manager (SUM)</li> <li>Supermicro SuperDoctor® 5 and Watch Dog</li> </ul>	<ul style="list-style-type: none"> <li>Integrated IPMI 2.0 plus KVM with dedicated LAN</li> <li>Supermicro Server Manager (SSM) and Supermicro Update Manager (SUM)</li> <li>Supermicro SuperDoctor® 5 and Watch Dog</li> </ul>
Chassis	CSE-458GTS-R000NDP	CSE-458GTS-R000NDP
Form Factor	<ul style="list-style-type: none"> <li>4U rackmount with optional 1U expansion module</li> </ul>	<ul style="list-style-type: none"> <li>4U rackmount with optional 1U expansion module</li> </ul>
Front Panel	<ul style="list-style-type: none"> <li>On/off and Universal Information (UID) buttons</li> <li>Power status, HDD activity, network activity, and UID LEDs</li> </ul>	<ul style="list-style-type: none"> <li>On/off and Universal Information (UID) buttons</li> <li>Power status, HDD activity, network activity, and UID LEDs</li> </ul>
Expansion Slots	<ul style="list-style-type: none"> <li>8x PCI-E 4.0 (x16) low-profile slots (via PLX switch)</li> <li>2x PCI-E 4.0 x16 LP or AIOM (via CPU with 1U expansion module)</li> </ul>	<ul style="list-style-type: none"> <li>8x PCI-E 4.0 (x16) low-profile slots (via PLX switch)</li> <li>2x PCI-E 4.0 x16 LP or AIOM (via CPU with 1U expansion module)</li> </ul>
Drive Bays	<ul style="list-style-type: none"> <li>10 hot-swap 2.5" NVMe U.2 (via PCI-E switch) or 10x hot-swap 2.5" SATA</li> <li>2 M.2 NVMe/SATA3 slots (max 80mm)</li> </ul>	<ul style="list-style-type: none"> <li>10 hot-swap 2.5" NVMe U.2 (via PCI-E switch) or 10x hot-swap 2.5" SATA</li> <li>2 M.2 NVMe/SATA3 slots (max 80mm)</li> </ul>
Shared Power & Cooling	<ul style="list-style-type: none"> <li>5 x hot-swap 11.5K RPM heavy duty fans</li> <li>4 Redundant 3000W (2+2) Titanium Level power supplies</li> </ul>	<ul style="list-style-type: none"> <li>5 x hot-swap 11.5K RPM heavy duty fans</li> <li>4 Redundant 3000W (2+2) Titanium Level power supplies</li> </ul>

<sup>1</sup>Certain CPUs with high TDP may be supported only under specific conditions. Please contact Supermicro Technical Support for additional information about specialized system optimization