

X13 Universal GPU

Multi-Architecture Flexibility with Future-Proof Open-Standards-Based Design



Ultimate modularity and customization options for AI and HPC environments

- Dual 5th/4th Gen Intel® Xeon® Scalable processors
- Support for the latest industry standards including PCIe 5.0 and DDR5
- Innovative modular architecture designed for flexibility and futureproofing with 4U and 8U form factors
- Supports next-generation GPUs including NVIDIA H200 8-GPU, H100 8-GPU/4-GPU and Intel® Data Center GPU Max 1550
- Optimized thermal capacity and airflow to support CPUs up to 385W and GPUs up to 700W with air cooling
- PCIe 5.0 x16 networking slots and up to 16 U.2 NVMe drive bays

Open-Standards-Based Platform

Supermicro X13 Universal GPU systems feature an open, modular, standards-based architecture designed for maximum flexibility. Support for multiple industry-standard GPUs allows organizations to take advantage of different GPU configurations based on workload while only deploying a single server architecture, reducing infrastructure complexity and simplifying future upgrades. Designed for serviceability with hot-swappable, tool-less components in a modular construction, the chassis design is optimized for thermal capacity.

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
- Deep Learning for image and video detection/recognition, life sciences & drug discovery, autonomous driving and robotics

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), IT administrators can choose the GPU architecture best suited for their HPC or AI workloads. Additionally, support for GPU interconnects including the latest NVIDIA NVLink® v4.0 with a total bandwidth of 900 GB/s facilitates ultra-fast GPU-to-GPU communication, reducing bottlenecks caused by traditional GPU interlinks.

High Performance Networking and Storage

Extraordinary data throughput from next-generation GPUs and CPUs is nothing without networking and storage to match. All Universal GPU systems support high-speed networking up to 400Gb/s directly to each GPU via PCIe 5.0 x16 slots with additional OCP 3.0 compliant AIOM options for DPUs and other accelerator cards. For storage, systems support up to 16 front-accessible hot-swap U.2 NVMe drive bays. There is also an additional front I/O variation available for selected models for enhanced maintenance and management.

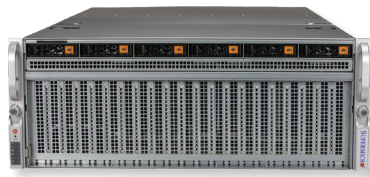
AI Accelerated with 5th Gen Intel Xeon Processors

The latest 5th Gen Intel Xeon processors are the most powerful Xeon processors ever, with up to 64 cores per CPU at the same power envelope as the previous generation, and a new high-power liquid cooled CPU up to 385W TDP. 5th Gen Intel Xeon includes built-in accelerator engines optimized for AI and HPC workloads including the purpose-built Intel Advanced Matrix Extensions (Intel AMX) accelerator which improves the performance of deep learning workloads to deliver robust AI capabilities for AI training and inference.



Universal GPU	SYS-821GE-TNHR/FTNHR	SYS-821GV-TNR
Processor Support	Dual 5th/4th Gen Intel® Xeon® Scalable processors Up to 350W TDP (air cooled) [†] Up to 385W TDP (liquid cooled) [†]	Dual 5th/4th Gen Intel® Xeon® Scalable processors Up to 350W TDP (air cooled) [†] Up to 385W TDP (liquid cooled) [†]
Memory Slots & Capacity	32 DIMM slots; up to 8TB DDR5-5600MT/s	32 DIMM slots; up to 8TB DDR5-5600MT/s
GPU Support	HGX H200 8-GPU Multi-GPU Board HGX H100 8-GPU Multi-GPU Board NVIDIA® NVLink™ with NVSwitch™	8 Intel Data Center GPU Max 1550
I/O Ports	2x 10GbE BaseT with Intel® X550-AT2 (optional) 2x 25GbE SFP28 with Broadcom® BCM57414 (optional) 2x 10GbE BaseT with Intel® X710-AT2 (optional) 1 VGA port (rear)	2x 10GbE RJ45 ports with Intel® X550-AT2 Ethernet controller 1 VGA port (rear)
Motherboard	X13DEG-OAD	X13DEG-PVC
Form Factor	8U Rackmount 843.28mm/33.2" depth	4U Rackmount 843.28mm/33.2" depth
Expansion Slots	8 PCIe 5.0 x16 LP slots 2 PCIe 5.0 x16 FHFL slots	8 PCIe 5.0 x16 LP slots 4 PCIe 5.0 x16 FHFL slots
Drive Bays	16 hot-swap 2.5" NVMe drive bays 3 hot-swap 2.5" NVMe/SATA drive bays	16 hot-swap 2.5" NVMe drive bays 3 hot-swap 2.5" NVMe/SATA drive bays
Cooling	10 heavy duty fans	10 removeable heavy duty 8cm fans
Power	6 3000W (4+2) Redundant power supplies, Titanium Level Optional: 8x 3000W (4+4) Redundant power supplies, Titanium Level	6 3000W Redundant power supplies, Titanium Level

[†] CPUs with high TDP supported under specific conditions. Contact Technical Support for details.



Universal GPU	SYS-421GU-TNXR
Processor Support	Dual 5th/4th Gen Intel® Xeon® Scalable processors Up to 350W TDP (air cooled) [†] Up to 350W TDP (liquid cooled) [†]
Memory Slots & Capacity	32 DIMM slots; up to 8TB DDR5-5600MT/s
GPU Support	HGX H100 4-GPU Multi-GPU Board NVIDIA® NVLink™
I/O Ports	2x 10GbE RJ45 ports with Intel® X710 Ethernet controller 1 VGA port (rear)
Motherboard	X13DGU-P
Form Factor	4U Rackmount 833mm/32.79" depth
Expansion Slots	8 PCIe 5.0 x16 LP slots
Drive Bays	6 hot-swap 2.5" NVMe/SATA drive bays
Cooling	5 heavy duty fans
Power	4x 3000W Redundant power supplies, Titanium Level

[†] CPUs with high TDP supported under specific conditions. Contact Technical Support for details.