

Supermicro GPU SuperServer SYS-821GE-TNHR

8x H200

Hopper

rack

2024

OVERVIEW

15.8 FP8 PFLOPS Dense	1.1 TB Aggregate Memory	38.4 TB/s Aggregate Bandwidth	— Aggregate Power
---	---	---	---

PERFORMANCE METRICS

Precision	Peak (PFLOPS)	Per GPU (TFLOPS)	TFLOPS/W
FP64	0.3	34.0	68.00
FP32	0.5	67.0	134.00
TF32	4.0	494.5	989.00
FP16	7.9	989.5	1979.00
BF16	7.9	989.5	1979.00
FP8	15.8	1979.0	3958.00
INT8	15.8	1979.0	3958.00

All figures are dense. Vendors commonly headline the sparse number, which is twice the dense one.

CONFIGURATION

Accelerator	H200 SXM 141GB
Count	8
Memory per GPU	141 GB
Bandwidth per GPU	4.8 TB/s

POWER & EFFICIENCY

Aggregate Power	—
TDP per GPU	700 W
FP8 Efficiency	3958.00 TFLOPS/W
Aggregate Bandwidth	38.4 TB/s

SYSTEM DETAILS

Vendor: Supermicro	Family: GPU SuperServer	Architecture: Hopper
Form Factor: rack	Release Year: 2024	Process Node: 4nm

INTERCONNECT

NVSwitch + NVLink

NOTES

8U GPU SuperServer built on the NVIDIA HGX H200 baseboard with 4x NVSwitch. Two Intel Xeon Scalable sockets (4th or 5th Gen, up to 64 cores each at 350W, air-cooled), 32 DIMM slots to an 8 TB maximum of DDR5-5600 ECC RDIMM, 19 front hot-swap 2.5-inch NVMe/SATA bays plus two M.2 NVMe boot slots, and 8x PCIe Gen5 x16 low-profile slots with 2x FHHL. Also sold in an HGX H100 configuration. System-level power draw is not published.

