



NVIDIA GB200 NVL72 GPU 186GB

Blackwell SXM 2024 4nm

OVERVIEW

80.0 FP32 TFLOPS	186 GB VRAM HBM3e	8.0 TB/s Memory Bandwidth	— TDP
---------------------	----------------------	------------------------------	----------

PERFORMANCE METRICS

Precision	Peak (dense)	2:4 Sparse	Bits	Type
FP64	40.0	N/A	64	Standard
FP32	80.0	N/A	32	Standard
TF32	1250.0	2500.0	32	Standard
BF16	2500.0	5000.0	16	Standard
FP16	2500.0	5000.0	16	Standard
FP8	5000.0	10000.0	8	Standard
FP6	5000.0	10000.0	—	Standard
FP4	10000.0	20000.0	4	Standard
INT8	5000.0	10000.0	8	Standard

MEMORY SPECIFICATIONS

Capacity	186 GB
Type	HBM3e
Bandwidth	8.0 TB/s
Interface Width	—

POWER & EFFICIENCY

TDP	—
Est. Max Power	—
FP32 Efficiency	—
FP16 Efficiency	—

HARDWARE DETAILS

Vendor: NVIDIA	Architecture: Blackwell	Process Node: 4nm
Form Factor: SXM	Launch Year: 2024	Variant: NVL72 GPU 186GB

INTERCONNECT

GPU-to-GPU: NVLink	Bandwidth: 1.8 TB/s
--------------------	---------------------

CHIP DESIGN

Transistors: 208.0 billion	Chiplets: 2 (Two reticle-limited dies joined by a 10 TB/s die-to-die link,
----------------------------	--

presented to software as one GPU)

ENTERPRISE FEATURES

ECC: **Yes**

Sparsity: **Yes**

Flopper.io | Compare the GPUs Powering AI | Source: NVIDIA