



NVIDIA GeForce RTX 5070 12GB

Blackwell PCIe 2025 4nm

OVERVIEW

30.9 FP32 TFLOPS	12 GB VRAM GDDR7	672 GB/s Memory Bandwidth	250 W TDP
----------------------------	----------------------------	-------------------------------------	---------------------

PERFORMANCE METRICS

Precision	Peak (dense)	2:4 Sparse	Bits	Type
FP64	0.5	N/A	64	Standard
FP32	30.9	N/A	32	Standard
TF32	30.9	61.7	32	Standard
BF16	61.7	123.5	16	Standard
FP16	123.5	246.9	16	Standard
FP8	246.9	493.9	8	Standard
FP6	—	not published	—	Standard
FP4	493.9	987.8	4	Standard
INT8	246.9	493.9	8	Standard

MEMORY SPECIFICATIONS

Capacity	12 GB
Type	GDDR7
Bandwidth	672 GB/s
Interface Width	192-bit

POWER & EFFICIENCY

TDP	250 W
Max Power	250 W
FP32 Efficiency	0.12 TFLOPS/W
FP16 Efficiency	0.49 TFLOPS/W

HARDWARE DETAILS

Vendor: NVIDIA	Architecture: Blackwell	Process Node: 4nm
Form Factor: PCIe	Launch Year: 2025	Variant: 12GB

COMPUTE ENGINE

CUDA Cores: 6,144	Tensor Cores: 192 (5th Gen)	SMs: 48
Base Clock: 2.33 GHz	Boost Clock: 2.51 GHz	

INTERCONNECT

PCIe: 5.0

CHIP DESIGN

Transistors: 31.1 billion Die Size: 263 mm²

ENTERPRISE FEATURES

Sparsity: Yes