# NVIDIA GB200 NVL72 GPU 186GB vs NVIDIA GeForce RTX 5080 16GB

Side by side datacenter GPU specifications from the Flopper.io database.

## OVERVIEW

| | NVIDIA GB200 NVL72 GPU 186GB | NVIDIA GeForce RTX 5080 16GB |
|---|---|---|
| FP32 TFLOPS | **80.0** | **56.3** |
| Memory | **186 GB** | **16 GB** |
| Memory Bandwidth | **8.0 TB/s** | **960 GB/s** |
| TDP | – | **360 W** |

## PERFORMANCE BY PRECISION

| Precision | NVIDIA GB200 NVL72 GPU 186GB peak TFLOPS (dense) | NVIDIA GeForce RTX 5080 16GB peak TFLOPS (dense) |
|---|---|---|
| FP64 | 40.0 | 0.9 |
| FP32 | 80.0 | 56.3 |
| TF32 | 1250.0 | 56.3 |
| BF16 | 2500.0 | 112.6 |
| FP16 | 2500.0 | 225.1 |
| FP8 | 5000.0 | 450.2 |
| FP6 | 5000.0 | – |
| FP4 | 10000.0 | 900.5 |
| INT8 | 5000.0 | 450.2 |

## SPECIFICATIONS

| | NVIDIA GB200 NVL72 GPU 186GB | NVIDIA GeForce RTX 5080 16GB |
|---|---|---|
| Architecture | Blackwell | Blackwell |
| Process Node | 4nm | 4nm |
| Launch Year | 2024 | 2025 |
| Form Factor | SXM | PCIe |
| Memory Type | HBM3e | GDDR7 |

| | NVIDIA GB200 NVL72 GPU 186GB | NVIDIA GeForce RTX 5080 16GB |
|---|---|---|
| Interface Width | – | 256-bit |
| Interconnect | NVLink | – |
| TDP | – | 360 W |
| FP32 Efficiency | – | 0.16 TFLOPS/W |