



AWS Trainium3

NeuronCore-v4 2025 3nm

OVERVIEW

183.0 FP32 TFLOPS	144 GB VRAM HBM3e	4.9 TB/s Memory Bandwidth	— TDP
---	---	---	-----------------------------

PERFORMANCE METRICS

Precision	Peak TFLOPS	Bits	Type
FP8	2517.0	8	Standard
FP4	2517.0	4	Standard
FP16	671.0	16	Standard
BF16	671.0	16	Standard
TF32	671.0	32	Standard
FP32	183.0	32	Standard

MEMORY SPECIFICATIONS

Capacity	144 GB
Type	HBM3e
Bandwidth	4.9 TB/s
Interface Width	—

POWER & EFFICIENCY

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

HARDWARE DETAILS

Vendor: AWS	Architecture: NeuronCore-v4	Process Node: 3nm
Form Factor: —	Launch Year: 2025	

COMPUTE ENGINE

NeuronCore-v4: 8

INTERCONNECT

GPU-to-GPU: NeuronLink-v4	Bandwidth: 2.6 TB/s
---------------------------	---------------------

ENTERPRISE FEATURES

Sparsity: Yes

