

Performance  
(TFLOPs)

Memory Bandwidth  
1008 GB/s

Peak FP16 FLOPS  
147 TFLOPS

$10^2$

$10^1$

Memory  
Bound

Compute  
Bound

Arithmetic Intensity  
(FLOPS/Byte)

$10^2$

145.79

$10^3$

● Encoder ■ VLM Prefill ▲ Action Denoise

