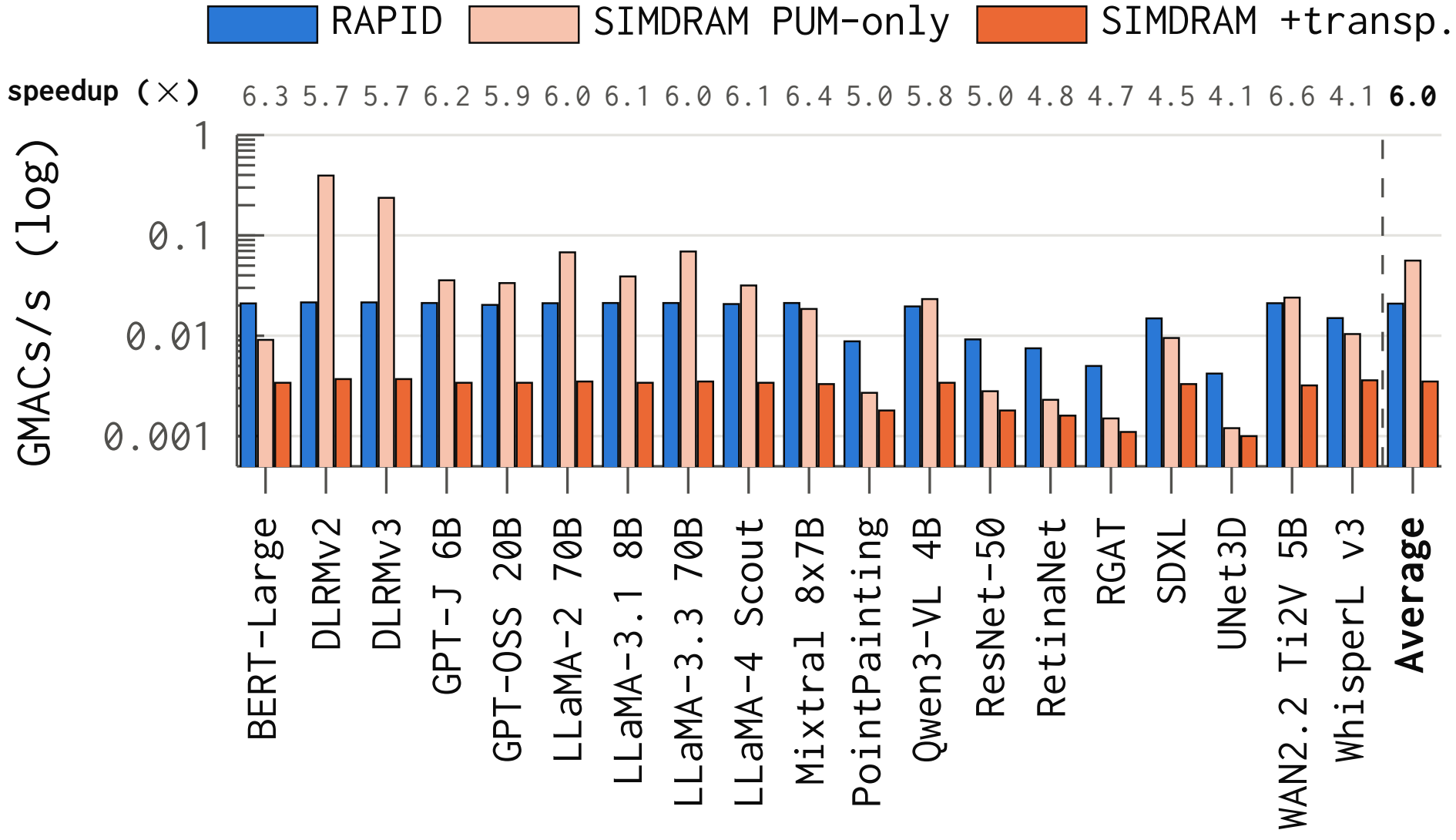

RAPID
SIMDRAM PUM-only
SIMDRAM +transp.
speedup (×)
6.3 5.7 5.7 6.2 5.9 6.0 6.1 6.0 6.1 6.4 5.0 5.8 5.0 4.8 4.7 4.5 4.1 6.6 4.1 6.0
GMACs/s (log)
1
0.1
0.01
0.001
BERT-Large
DLRMv2
DLRMv3
GPT-J 6B
GPT-OSS 20B
LLaMA-2 70B
LLaMA-3.1 8B
LLaMA-3.3 70B
LLaMA-4 Scout
Mixtral 8x7B
PointPainting
Qwen3-VL 4B
ResNet-50
RetinaNet
RGAT
SDXL
UNet3D
WAN2.2 Ti2V 5B
WhisperL v3
Average