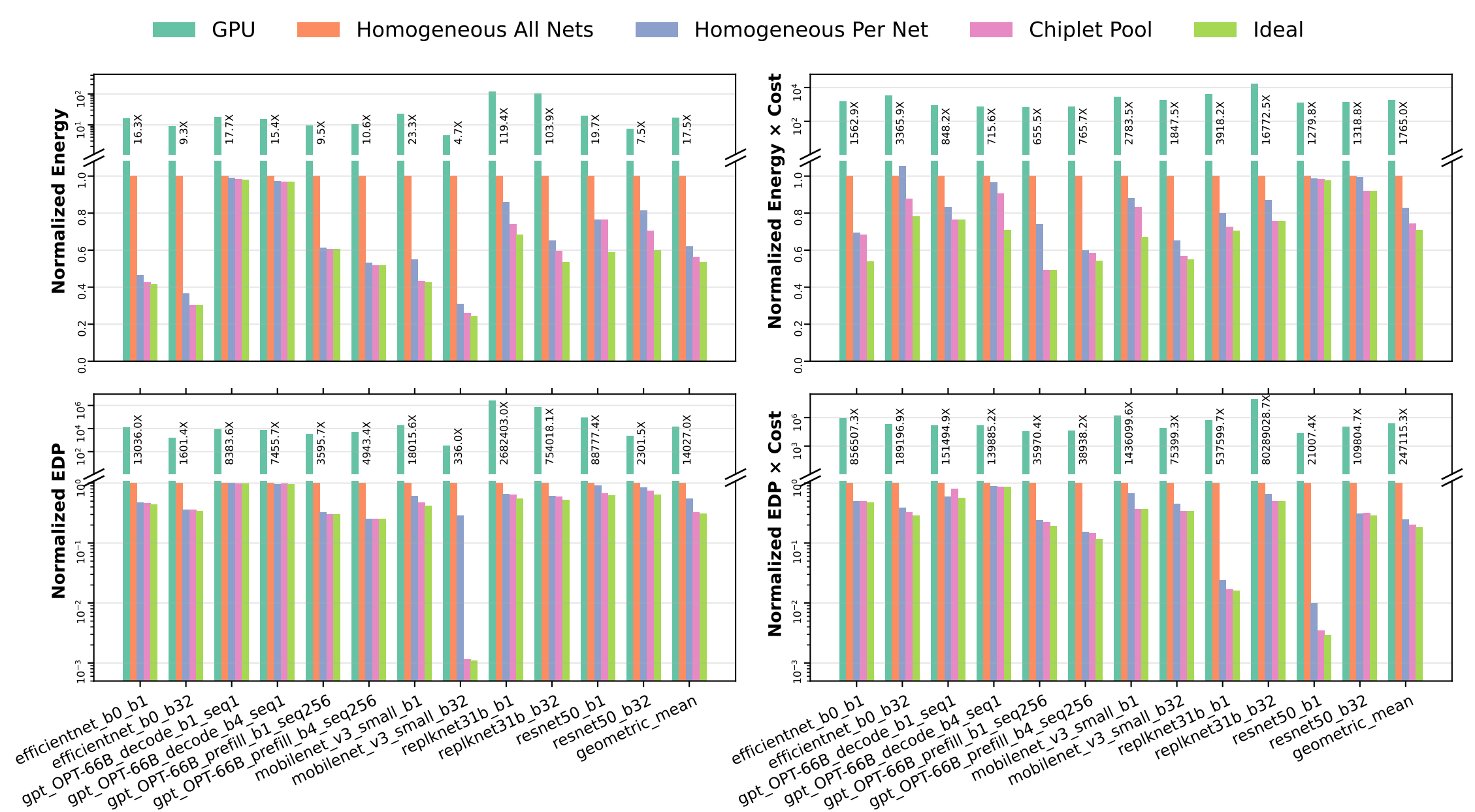
GPU
Homogeneous All Nets
Homogeneous Per Net
Chiplet Pool
Ideal
Normalized Energy
16.3X
9.3X
17.7X
15.4X
9.5X
10.6X
23.3X
4.7X
119.4X
103.9X
19.7X
7.5X
17.5X
Normalized Energy × Cost
1562.9X
3365.9X
848.2X
715.6X
655.5X
765.7X
2783.5X
1847.5X
3918.2X
16772.5X
1279.8X
1318.8X
1765.0X
Normalized EDP
13036.0X
1601.4X
8383.6X
7455.7X
3595.7X
4943.4X
18015.6X
336.0X
2682403.0X
754018.1X
88777.4X
2301.5X
14027.0X
Normalized EDP × Cost
856507.3X
189196.9X
151494.9X
139885.2X
35970.4X
38938.2X
1436099.6X
75399.3X
537599.7X
80289028.7X
21007.4X
109804.7X
247115.3X
efficientnet_b0_b1
efficientnet_b0_b32
gpt_OPT-66B_decode_b1_seq1
gpt_OPT-66B_decode_b4_seq1
gpt_OPT-66B_prefill_b1_seq256
gpt_OPT-66B_prefill_b4_seq256
mobilenet_v3_small_b1
mobilenet_v3_small_b32
replknet31b_b1
replknet31b_b32
resnet50_b1
resnet50_b32
geometric_mean