

DRAM  
reads

100K  
100K  
100K  
100K  
100K  
100K  
100K  
100K  
100K  
100K

MatMul  
f32  
 $32 \times 32 \times 32$   
1 block

MatMul  
f32  
 $64 \times 64 \times 64$   
4 blocks

MatMul  
f32  
 $128 \times 128 \times 128$   
16 blocks

MatMul  
f32  
 $256 \times 256 \times 256$   
64 blocks

MatMul  
f32  
 $512 \times 512 \times 512$   
256 blocks

