	EE557	Project4		
Project name	SA4			
Document ref				
Version				
Release date				
Author	Fei			
Classification	[Document classific	ation]		
Distribution List	[Distribution list]			
Approved by	Name	Signature	Date	

Fei Wu 6897429283 wufei@usc.edu



Ming Hsieh Department of Electrical Engineering
University of Southern California, Los Angeles, CA 90089
Fall 2022

EE 557

1 Tables

					aí	L00		
2		Title Size	Matrix Size	A major	B major	Ex Time (ms)	FLOPs	GFLOPs
3	1	32*16*16	16*16*16	col	row	2.055641	8192.0000	0.003985
<u> </u>	2	32*16*16	16*16*32	col	row	1.359971	16384.0000	0.012047
5	3	32*16*16	16*16*16	col	col	2.052867	8192.0000	0.003991
5	4	32*16*16	16*16*32	col	col	1.346217	16384.0000	0.01217
,	5	32*16*16	16*16*16	row	row	2.055211	8192.0000	0.003986
3	6	32*16*16	16*16*32	row	row	2.020667	16384.0000	0.008108
)	7	32*16*16	16*16*16	row	col	2.053469	8192.0000	0.003989
0	8	32*16*16	16*16*32	row	col	1.350815	16384.0000	0.012129
1	9	32*16*16	32*32*32	col	row	1.345174	65536.0000	0.048719
2	10	32*16*16	32*32*32	col	col	1.930839	65536.0000	0.033942
3	11	32*16*16	32*32*32	row	row	1.367125	65536.0000	0.047937
4 DWMMA	12	32*16*16	32*32*32	row	col	1.936322	65536.0000	0.033846
5	13		128*128*128	col	row	1.69345	4194304.0000	2.476778
5	14	32*128*128	128*128*128	col	col	1.689855	4194304.0000	2.48205
7	15	32*128*128	128*128*128	row	row	1.708981	4194304.0000	2.454272
3	16	32*128*128	128*128*128	row	col	1.709111	4194304.0000	2.454085
9	17	32*128*128	256*256*128	col	row	1.372225	16777216.0000	12.226287
	18	32*128*128	256*256*128	col	col	1.35374	16777216.0000	12.393234
<u> </u>	19	32*128*128	256*256*128	row	row	1.351045	16777216.0000	12.417955
2	20	32*128*128	256*256*128	row	col	1.359801	16777216.0000	12.337994
3	21	64*128*128	256*256*256	col	row	1.363435	33554432.0000	24.610218
1	22	64*128*128	256*256*256	col	col	1.346884	33554432.0000	24.912637
5	23	64*128*128	256*256*256	row	row	1.352596	33554432.0000	24.809265
5	24	64*128*128	256*256*256	row	col	1.350902	33554432.0000	24.838539
7	1	8*128*128	1024*512*8	col	row	1.351123	8388608.0000	6.208619
SGEMM	2	8*128*128	128*81*1	col	row	1.340984	20736.0000	0.015463
	3	8*128*128	128*112*8	col	row	1.348889	229376.0000	0.170048
			64*32*8	col	row	2.007764	32768.0000	0.016321
	1	8*32*64				4.000055	1404004000	0.07074.0
DGEMM	2	8*32*64	256*128*64	col	row	1.362355	4194304.0000	3.078716
9 0 1 DGEMM		 			row	1.362355 1.939757	4194304.0000 65536.0000	3.078716 0.033786
DGEMM 2	2	8*32*64 8*64*64	256*128*64 64*64*8	col	row	1.939757 40	65536.0000	0.033786
DGEMM	2 3	8*32*64 8*64*64 Title Size	256*128*64 64*64*8 Matrix Size	col col	row a B major	1.939757 40 Ex Time (ms)	65536.0000 FLOPs	0.033786 GFLOPs
DGEMM DGEMM	2 3	8*32*64 8*64*64 Title Size 32*16*16	256*128*64 64*64*8 Matrix Size 16*16*16	col col	row a B major row	1.939757 40 Ex Time (ms) 1.044462	65536.0000 FLOPs 8192.0000	0.033786 GFLOPs 0.007843
DGEMM DGEMM	2 3 1 2	8*32*64 8*64*64 Title Size 32*16*16 32*16*16	256*128*64 64*64*8 Matrix Size 16*16*16 16*16*32	COI COI A major COI COI	B major row row	1.939757 40 Ex Time (ms) 1.044462 0.769658	FLOPs 8192.0000 16384.0000	0.033786 GFLOPs 0.007843 0.021287
DGEMM 22 DGEMM 22 2 3 3 4 5 5	2 3 1 2 3	8*32*64 8*64*64 Title Size 32*16*16 32*16*16 32*16*16	256*128*64 64*64*8 Matrix Size 16*16*16 16*16*32 16*16*16	A major col col col	B major row row	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622	FLOPs 8192.0000 16384.0000 8192.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336
DGEMM	2 3 1 2 3 4	8*32*64 8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16	256*128*64 64*64*8 Matrix Size 16*16*16 16*16*32 16*16*16	A major col col col col	B major row row col col	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101
DGEMM 2 2 2 3 3 4 5 5 6 7 7	1 2 3 4 5	8*32*64 8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16	256*128*64 64*64*8 Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16	A major col col col col row	B major row row col col row	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514
DGEMM 2	1 2 3 4 5 6	8*32*64 8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16	256*128*64 64*64*8 Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32	A major col col col col col row row	B major row row col col row row	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148 0.77629	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514 0.021106
DGEMM 2	2 3 1 2 3 4 5 6 7	8*32*64 8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16	256*128*64 64*64*8 Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16	A major col col col col row row	B major row row col col row row	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148 0.77629 0.714236	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514 0.021106 0.01147
DGEMM 2	2 3 1 2 3 4 5 6 7 8	8*32*64 8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16	256*128*64 64*64*8 Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16	A major col col col col row row row	B major row col col row row col col col col	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148 0.77629 0.714236 0.803424	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514 0.021106 0.01147 0.020393
DGEMM	2 3 1 2 3 4 5 6 7 8	8*32*64 8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16	256*128*64 64*64*8 Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32*32	A major col col col col row row row col	B major row col col row col col row row col col row	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148 0.77629 0.714236 0.803424 0.782515	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514 0.021106 0.01147 0.020393 0.08375
DGEMM	2 3 1 2 3 4 5 6 7 8 9	8*32*64 8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16	256*128*64 64*64*8 Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32 32*32*32	A major col col col col row row row col col col	B major row col col row col col row col col row col	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148 0.77629 0.714236 0.803424 0.782515 0.791541	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514 0.021106 0.01147 0.020393 0.08375 0.082795
DGEMM	2 3 1 2 3 4 5 6 7 8 9 10	8*32*64 8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32 32*32*32 32*32*32	A major col col col col row row row col col row row row row row row row	B major row row col col row	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148 0.77629 0.714236 0.803424 0.782515 0.791541 0.783696	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514 0.021106 0.01147 0.020393 0.08375 0.082795 0.083624
DGEMM	2 3 1 2 3 4 5 6 7 8 9 10 11 12	8*32*64 8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16	256*128*64 64*64*8 Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32 32*32*32 32*32*32 32*32*32	A major col col col col row row row col col row row row row row row row	B major row row col col row	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148 0.77629 0.714236 0.803424 0.782515 0.791541 0.783696 0.785251	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514 0.021106 0.01147 0.020393 0.08375 0.082795 0.083624 0.083395
DGEMM	2 3 1 2 3 4 5 6 7 8 9 10 11 12 13	8*32*64 8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32	A major col col col col row row row col col row row row col col row row	B major row row col col row col row col col row col row col row col row	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148 0.77629 0.714236 0.803424 0.782515 0.791541 0.783696 0.785251 0.72180	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514 0.021106 0.01147 0.020393 0.08375 0.082795 0.083624 0.083395 5.810863
DGEMM 2 2 3 4 5 6 6 DWMMA	2 3 1 2 3 4 5 6 7 8 9 10 11 12 13 14	8*32*64 8*64*64 8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32	A major col col col col row row row col col row row col col col row row col col	B major row row col col row col col row col col row col row col row col row col row col	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148 0.77629 0.714236 0.803424 0.782515 0.791541 0.783696 0.785251 0.72180 0.780555	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514 0.021106 0.01147 0.020393 0.08375 0.082795 0.08395 5.810863 5.373489
DGEMM 2 2 3 4 5 6 7 7 8 9 0 0 1 1 2 2 3 4 DWMMA 6 7	2 3 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15	8*32*64 8*64*64 8*64*64 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*18 32*12*128*128	256*128*64 64*64*8 Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 128*128*128*128*128*128*128*128*128*128*	A major col col col col row row col col row row col col row row col row col row	B major row row col col row col col row col col row col row col row col row col row col row	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148 0.77629 0.714236 0.803424 0.782515 0.791541 0.783696 0.785251 0.72180 0.780555 0.749086	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514 0.021106 0.01147 0.020393 0.08375 0.082795 0.083624 0.083395 5.810863 5.373489 5.599229
DGEMM	2 3 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	8*32*64 8*64*64 8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*18 32*128*128 32*128*128	256*128*64 64*64*8 Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32 32*32*32 32*32*32 32*32*32 128*128*128*128 128*128*128*128	COI COI COI COI COI COI COI FOW FOW COI COI COI COI FOW FOW COI COI FOW	row B major row row col col row col col row	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148 0.77629 0.714236 0.803424 0.782515 0.791541 0.783696 0.785251 0.72180 0.780555 0.749086 0.717948	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514 0.021106 0.01147 0.020393 0.08375 0.082795 0.083624 0.083395 5.810863 5.373489 5.599229 5.842072
DGEMM	2 3 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	8*32*64 8*64*64 8*64*64 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*18 32*128*128 32*128*128 32*128*128	256*128*64 64*64*8 Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32 32*32*32 32*32*32 32*32*32 128*128*128 128*128*128 128*128*128 128*128*128	COI	row B major row row col col row col col row	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148 0.77629 0.714236 0.803424 0.782515 0.791541 0.783696 0.785251 0.72180 0.780555 0.749086 0.717948 0.793625	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514 0.021106 0.01147 0.020393 0.08375 0.082795 0.083624 0.083395 5.810863 5.373489 5.599229 5.842072 21.139979
DGEMM	2 3 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18	8*32*64 8*64*64 8*64*64 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128	256*128*64 64*64*8 Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32 32*32*32 32*32*32 32*32*32 128*128*128 128*128*128 128*128*128 128*128*128 256*256*128	A major col col col col row row col col row row col col row row col col row col col col row row col col col row row	row B major row row col col row col col row	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148 0.77629 0.714236 0.803424 0.782515 0.791541 0.783696 0.785251 0.72180 0.780555 0.749086 0.717948 0.793625 0.780471	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 16777216.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514 0.021106 0.01147 0.020393 0.08375 0.082795 0.083624 0.083395 5.810863 5.373489 5.599229 5.842072 21.139979 21.496271
DGEMM	2 3 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19	8*32*64 8*64*64 8*64*64 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128	256*128*64 64*64*8 Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32 32*32*32 32*32*32 32*32*32 128*128*128 128*128*128 128*128*128 128*128*128	A major col col col col row row	B major row row col col row	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148 0.77629 0.714236 0.803424 0.782515 0.791541 0.783696 0.785251 0.72180 0.780555 0.749086 0.717948 0.793625 0.780471 0.804676	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 16777216.0000 16777216.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514 0.021106 0.01147 0.020393 0.08375 0.082795 0.083624 0.083395 5.810863 5.373489 5.599229 5.842072 21.139979 21.496271 20.849654
DGEMM	2 3 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18	8*32*64 8*64*64 8*64*64 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*18 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128	256*128*64 64*64*8 Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 128*128*128 128*128*128 128*256*256*128 256*256*128 256*256*128	A major col col col col row row col col col row row col col row row col col row row col col row row row	B major row col col row	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148 0.77629 0.714236 0.803424 0.782515 0.791541 0.783696 0.785251 0.72180 0.780555 0.749086 0.717948 0.793625 0.789471 0.804676 0.828801	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 16777216.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514 0.021106 0.01147 0.020393 0.08375 0.082795 0.083624 0.083395 5.810863 5.373489 5.599229 5.842072 21.139979 21.496271 20.849654 20.242755
DGEMM	2 3 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20	8*32*64 8*64*64 8*64*64 8*64*64 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*18 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 216*16*32 216*18*2 216*18*2 216*22 22*32*32*32 232*32*32 232*32*32 2128*128*128 2128*128*128 2128*128*128 228*256*128 2256*256*128 2256*256*128 2256*256*256	A major col col col col row row	B major row row col col row	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148 0.77629 0.714236 0.803424 0.782515 0.791541 0.783696 0.785251 0.72180 0.780555 0.749086 0.717948 0.793625 0.780471 0.804676	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 16777216.0000 16777216.0000 16777216.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514 0.021106 0.01147 0.020393 0.08375 0.082795 0.083624 0.083395 5.810863 5.373489 5.599229 5.842072 21.139979 21.496271 20.849654
DGEMM DG	2 3 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21	8*32*64 8*64*64 8*64*64 8*64*64 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*18 32*128*128	256*128*64 64*64*8 Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 128*128*128 128*128*128 128*256*256*128 256*256*128 256*256*128	A major col col col col row row col col col row row col col row row col col row row row col col row row col col row row col col row row col	B major row col col row	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148 0.77629 0.714236 0.803424 0.782515 0.791541 0.783696 0.785251 0.72180 0.780555 0.749086 0.717948 0.793625 0.789471 0.804676 0.828801 0.74797	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 16777216.0000 16777216.0000 16777216.0000 16777216.0000 33554432.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514 0.021106 0.01147 0.020393 0.08375 0.082795 0.083624 0.083395 5.810863 5.373489 5.599229 5.842072 21.139979 21.496271 20.849654 20.242755 44.860666
DGEMM DG	2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22	8*32*64 8*64*64 8*64*64 8*64*64 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*18 32*128*128	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 216*16*32 216*16*16 216*16*32 216*128*232 232*32*32 232*32*32 232*32*32 2128*128*128*128 128*128*128 256*256*128 256*256*128 256*256*128 256*256*256	A major col col col col row row col col col row row col col col row row col col col row col col col row row	B major row col col row	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148 0.77629 0.714236 0.803424 0.782515 0.791541 0.783696 0.785251 0.72180 0.780555 0.749086 0.717948 0.793625 0.780471 0.804676 0.828801 0.74797 0.793095	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 16777216.0000 16777216.0000 16777216.0000 13554432.0000 33554432.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514 0.021106 0.01147 0.020393 0.08375 0.082795 0.083624 0.083395 5.810863 5.373489 5.599229 5.842072 21.139979 21.496271 20.849654 20.242755 44.860666 42.308213
DGEMM 2 2 3 4 5 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6	2 3 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23	8*32*64 8*64*64 8*64*64 8*64*64 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*18 32*128*128	256*128*64 64*64*8 Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 232*32*32 32*32*32 32*32*32 32*32*32 128*128*128 128*128*128 128*128*128 256*256*128 256*256*128 256*256*128 256*256*256 256*256*256	COI	B major row col col row	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148 0.77629 0.714236 0.803424 0.782515 0.791541 0.783696 0.785251 0.72180 0.780555 0.749086 0.717948 0.793625 0.780471 0.804676 0.828801 0.74797 0.793095 0.773748	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 16777216.0000 16777216.0000 16777216.0000 16777216.0000 33554432.0000 33554432.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514 0.021106 0.01147 0.020393 0.08375 0.082795 0.08375 0.08395 5.810863 5.373489 5.599229 5.842072 21.139979 21.496271 20.849654 20.242755 44.860666 42.308213 43.366099
DGEMM 22 DGEMM 22 33 44 56 66 7 88 99 00 1 1 2 2 3 3 4 4 5 5 6 6 6 7 7	2 3 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24	8*32*64 8*64*64 8*64*64 8*64*64 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*18 32*128*128	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 20 20 20 20 20 20 20 20 20 20 20 20 20	COI	B major row col col row	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148 0.77629 0.714236 0.803424 0.782515 0.791541 0.783696 0.785251 0.72180 0.780555 0.749086 0.717948 0.793625 0.780471 0.804676 0.828801 0.74797 0.793095 0.773748 0.782936	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 33554432.0000 33554432.0000 33554432.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514 0.021106 0.01147 0.020393 0.08375 0.082795 0.08375 0.08395 5.810863 5.373489 5.599229 5.842072 21.139979 21.496271 20.849654 20.242755 44.860666 42.308213 43.366099 42.857184
DGEMM DG	2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 1	8*32*64 8*64*64 8*64*64 8*64*64 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*128*128	256*128*64 64*64*8 Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 128*128*128 128*128*128 128*128*128 256*256*128 256*256*128 256*256*128 256*256*128 256*256*256 256*256*256	A major col col col col col row row	row B major row col col row col col row	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148 0.77629 0.714236 0.803424 0.782515 0.791541 0.783696 0.785251 0.72180 0.780555 0.749086 0.717948 0.793625 0.780471 0.804676 0.828801 0.74797 0.793095 0.773748 0.782936 0.930532	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 65536.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 16777216.0000 16777216.0000 16777216.0000 16777216.0000 33554432.0000 33554432.0000 33554432.0000 33554432.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514 0.021106 0.01147 0.020393 0.08375 0.082795 0.083624 0.083395 5.810863 5.373489 5.599229 5.842072 21.139979 21.496271 20.849654 20.242755 44.860666 42.308213 43.366099 42.857184 9.014852
DGEMM 22 DGEMM 22 DGEMM 24 DWMMA 66 7 8 9 0 0 1 2 2 3 4 4 4 5 6 6 7 8 SGEMM 9	2 3 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 1 2	8*32*64 8*64*64 8*64*64 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*18 32*128*128	256*128*64 64*64*8 Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32 32*32*32 32*32*32 32*32*32 128*128*128 128*128*128 128*128*128 128*128*128 256*256*128 256*256*128 256*256*256 256*256*256 256*256*256 1024*512*8 128*81*1	A major col col col col col row row col col col row row col col col row row	row B major row col col row	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148 0.77629 0.714236 0.803424 0.782515 0.791541 0.783696 0.785251 0.72180 0.780555 0.749086 0.717948 0.793625 0.780471 0.804676 0.828801 0.74797 0.793095 0.773748 0.782936 0.930532 0.778397	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 16777216.0000 16777216.0000 16777216.0000 133554432.0000 33554432.0000 33554432.0000 8388608.0000 20736.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514 0.021106 0.01147 0.020393 0.08375 0.082795 0.083624 0.083395 5.810863 5.373489 5.599229 5.842072 21.139979 21.496271 20.849654 20.242755 44.860666 42.308213 43.366099 42.857184 9.014852 0.026639
DGEMM 22 DGEMM 22 33 44 56 66 7 88 99 00 11 22 33 44 55 66 7 7 88 99 00 17 22 33 44 65 66 7 7 88 99 00 17 22 33 34 44 65 66 7 7 8 8 99 00 17 22 33 34 44 65 66 7 7 8 8 99 00 17 22 33 34 44 65 66 7 7 8 8 99 00 17 22 33 34 44 65 66 7 7 8 8 99 00 17 22 33 34 44 65 66 7 7 8 8 99 00 17 22 33 34 44 65 66 67 7 7 8 99 00 17 22 33 34 44 65 66 67 7 7 8 99 00 17 22 33 34 44 65 66 67 7 7 8 99 00 17 22 33 34 44 65 66 67 7 7 8 99 00 17 22 33 34 44 65 66 67 7 7 8 99 00 17 22 33 34 44 65 66 67 7 7 8 99 00 17 22 33 34 44 65 66 67 7 7 8 90 00 17 22 33 34 65 66 67 7 7 8 90 00 17 22 33 34 65 66 67 7 7 8 90 00 17 22 33 34 65 66 67 7 7 8 90 00 17 22 33 34 65 66 67 7 7 8 90 00 17 22 33 34 65 66 67 7 7 8 90 00 17 22 33 34 65 66 67 7 7 8 90 00 17 22 33 34 65 66 67 7 7 8 90 00 17 22 33 34 65 66 67 7 7 8 90 00 17 22 33 34 65 66 67 7 7 8 90 00 17 22 33 34 65 66 67 7 7 8 90 00 17 22 33 34 65 66 67 7 7 8 90 00 17 22 33 34 65 66 67 7 7 8 90 00 17 22 33 34 65 66 67 7 7 8 90 00 17 22 30 7 7 7 8 90 00 17 22 30 7 7 7 7 8 90 00 17 22 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7	2 3 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 1 2 3	8*32*64 8*64*64 8*64*64 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*128*128	256*128*64 64*64*8 Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32 32*32*32 32*32*32 32*32*32 228*128*128*128 128*128*128 128*128*128 256*256*128 256*256*128 256*256*256 256*256*256 256*256*256	COI	B major row row col row	1.939757 40 Ex Time (ms) 1.044462 0.769658 0.680622 0.779821 0.71148 0.77629 0.714236 0.803424 0.782515 0.791541 0.783696 0.785251 0.72180 0.780555 0.749086 0.717948 0.793625 0.780471 0.804676 0.828801 0.74797 0.793095 0.773748 0.782936 0.930532 0.778397 0.758239	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 16777216.0000 16777216.0000 16777216.0000 16777216.0000 33554432.0000 33554432.0000 33554432.0000 33554432.0000 8388608.0000 20736.0000	0.033786 GFLOPs 0.007843 0.021287 0.0120336 0.02101 0.011514 0.021106 0.01147 0.020393 0.08375 0.082795 0.083624 0.083395 5.810863 5.373489 5.599229 5.842072 21.139979 21.496271 20.849654 20.242755 44.860666 42.308213 43.366099 42.857184 9.014852 0.026639 0.302511

2 Report

Set the GPU machine in the CARC system:

```
[wufei@discovery2 gemm-test]$ module purge
[wufei@discovery2 gemm-test]$ module load gcc/8.3.0 cuda/10.1.243
[wufei@discovery2 gemm-test]$ salloc --partition=gpu --time=4:00:00 --cpus-per-task=8 --gres=gpu:a100:1 --mem=3u
fei@discovery2 gemm-test]$ salloc --partition=gpu --time=4:00:00 --cpus-per-task=8 --gres=gpu:a100:1 --mem=32GB
salloc: error: Invalid --mem specification
[wufei@discovery2 gemm-test]$ salloc --partition=gpu --time=4:00:00 --cpus-per-task=8 --gres=gpu:a100:1 --mem=32
GB#fei@discovery2 gemm-test]$ salloc --partition=gpu --time=4:00:00 --cpus-per-task=8 --gres=gpu:a100:1 --mem=32 GB#fei@discovery2 gemm-test]$ salloc --partition=gpu --time=4:00:00 --cpus-per-task=8 --gres=gpu:a100:1 --mem=32 salloc: error: Invalid --mem specification
[wufei@discovery2 gemm-test]$ salloc --partition=gpu --time=4:00:00 --cpus-per-task=8 --gres=gpu:a100:1 --mem=32
GB
salloc: Granted job allocation 12418981
salloc: Waiting for resource configuration
salloc: Nodes b01-06 are ready for job
[wufei@b01-06 gemm-test]$ nvcc --version
nvcc: NVIDIA (R) Cuda compiler driver
Copyright (c) 2005-2019 NVIDIA Corporation
Built on Sun_Jul_28_19:07:16_PDT_2019
Cuda compilation tools, release 10.1, V10.1.243
[wufei@b01-06 gemm-test]$ nvidia-smi
Sun Nov 20 19:33:41 2022
     NVIDIA-SMI 510.73.08
                                                                 Driver Version: 510.73.08
                                                                                                                                              CUDA Version: 11.6
                                                                                                                                                   Volatile Uncorr. ECC
GPU-Util Compute M.
      GPU Name
                                                Persistence-M| Bus-Id
                                                                                                                           Disp.A
                 Temp Perf Pwr:Usage/Cap
                                                                                                            Memory-Usage
                                                                                                                                                                                       MIG M.
          0 NVIDIA A100-PCI... Off
/A 31C P0 38W / 250W
                                                                                       00000000:25:00.0 Off
0MiB / 40960MiB
                                                                                                                                                                                     Default
                                                                                                                                                                                  Disabled
     Processes:
                       GI CI
ID ID
                                                             PID Type Process name
       No running processes found
 [wufei@b01-06 gemm-test]$ ■
 SISTERED VERSION - Please support MobaXterm by subscribing to the professional edition here: https://mobaxterm.mobatek.net
```

Check the script and check the output files which are going to be generated:

```
make CFLAGS=-DWMMA 1
./cutlass-test>>log dwmma1
make clean
make CFLAGS=-DWMMA_2
./cutlass-test>>log_dwmma2
make clean
make CFLAGS=-DWMMA 3
./cutlass-test>>log_dwmma3
make clean
make CFLAGS=-DWMMA_4
./cutlass-test>>log_dwmma4
make clean
make CFLAGS=-DWMMA 5
./cutlass-test>>log dwmma5
make clean
make CFLAGS=-DWMMA 6
./cutlass-test>>log dwmma6
make clean
make CFLAGS=-DWMMA 7
./cutlass-test>>log dwmma7
```

```
[wufei@discovery1 cutlass-gpapu-sim]smodule purge
[wufei@discovery1 cutlass-gpapu-sim]smodule load gcc/8.3.0 cuda/10.1.2433
[wufei@discovery1 cutlass-gpapu-sim]smodule load gcc/8.3.0 cuda/10.1.2433
[wufei@discovery1 cutlass-gpapu-sim]smodule load gcc/8.3.0 cuda/10.1.2433
[salloc: pol ballocation 12433778
salloc: pol 2433778 pueued and watting for resources
salloc: job 12433778 has been allocated resources
salloc gob 12433778 has been allocated resource gob 1243378 has been allocated resource gob 1243378 has been allocated resource gob 1243378 has been allocated resource gob 1243278 has been allocated reso
```

Check all the generated log files:

Modified the gemm.h to get the execution time and GFLOPs:

```
time = (stop.tv_sec - start.tv_sec) + (stop.tv_nsec - start.tv_nsec) / 1e9;
printf("excution time is %f ms\n", (time * 1000));

double FLOPS = testbed.flops();
printf("FLOPS is %1f\n", FLOPS);
double GFLOPS = testbed.GFLOPs_per_sec(time*1000);
printf("GFLOPS is %1f\n", GFLOPS);
```

Check the complete coding inside the submission.

Check the three test code and record the parameters like tile size, matrix size, A major, B major.

```
-rw-rw----. 1 wutei wutei 866 Nov 20 19:53 dgemm_tests.h

-rw-rw----. 1 wufei wufei 5645 Nov 21 20:39 gemm.h

-rw-rw----. 1 wufei wufei 12546 Nov 20 19:53 gemm_testbed.h

-rw-rw----. 1 wufei wufei 747 Nov 20 19:53 sgemm_tests.h

-rw-rw----. 1 wufei wufei 13932 Nov 20 19:53 wmma_tests.h

[wufei@b05-10 gemm-test]$ ■
```

Run the script again and check the log:

Sample

```
[wufei@b05-13 cutlass-gpgpu-sim]$ cat log_dwmma1
Successfully Launched
excution time is 2.055641 ms
FLOPs is 8192.000000
GFLOPs is 0.003985
Result Verified
```

Reset the real GPU on the CARC to a40 with the similar command and rerun the script again:

on No	v 21 2						
NVIC				Version: 510.73.08			
				Bus-Id Disp Memory-Usa			
0 0% 	NVIDI 27C	A A40 P0	Off 69W / 300W	00000000:21:00.0 0 0MiB / 46068M	ff IiB	 2% 	0 Default N/A
	esses: GI ID			pe Process name			GPU Memory Usage
No	runnin	g proc	esses found				
[wufei	.@b05 - 10	0 cutl	ass-gpgpu-sim]	.			

get all the required parameters and make the table.

					a	40		
		Title Size	Matrix Size	A major	B major	Ex Time (ms)	FLOPs	GFLOPs
	1	32*16*16	16*16*16	col	row	1.044462	8192.0000	0.007843
	2	32*16*16	16*16*32	col	row	0.769658	16384.0000	0.021287
	3	32*16*16	16*16*16	col	col	0.680622	8192.0000	0.0120336
	4	32*16*16	16*16*32	col	col	0.779821	16384.0000	0.02101
	5	32*16*16	16*16*16	row	row	0.71148	8192.0000	0.011514
	6	32*16*16	16*16*32	row	row	0.77629	16384.0000	0.021106
	7	32*16*16	16*16*16	row	col	0.714236	8192.0000	0.01147
	8	32*16*16	16*16*32	row	col	0.803424	16384.0000	0.020393
	9	32*16*16	32*32*32	col	row	0.782515	65536.0000	0.08375
	10	32*16*16	32*32*32	col	col	0.791541	65536.0000	0.082795
	11	32*16*16	32*32*32	row	row	0.783696	65536.0000	0.083624
	12	32*16*16	32*32*32	row	col	0.785251	65536.0000	0.083395
DWMMA -	13	32*128*128	128*128*128			0.783231	4194304.0000	5.810863
<u> </u>				col	row			
\vdash	14		128*128*128	col	col	0.780555	4194304.0000	5.373489
<u> </u>	15	32*128*128		row	row	0.749086	4194304.0000	5.599229
⊢	16	32*128*128		row	col	0.717948	4194304.0000	5.842072
	17	32*128*128	256*256*128	col	row	0.793625	16777216.0000	21.139979
	18	32*128*128	256*256*128	col	col	0.780471	16777216.0000	21.496271
	19	32*128*128		row	row	0.804676	16777216.0000	20.849654
	20	32*128*128	256*256*128	row	col	0.828801	16777216.0000	20.242755
	21	64*128*128		col	row	0.74797	33554432.0000	44.860666
	22	64*128*128	256*256*256	col	col	0.793095	33554432.0000	42.308213
L	23	64*128*128	256*256*256	row	row	0.773748	33554432.0000	43.366099
	24	64*128*128	256*256*256	row	col	0.782936	33554432.0000	42.857184
	1	8*128*128	1024*512*8	col	row	0.930532	8388608.0000	9.014852
SGEMM	2	8*128*128	128*81*1	col	row	0.778397	20736.0000	0.026639
	3	8*128*128	128*112*8	col	row	0.758239	229376.0000	0.302511
	1	8*32*64	64*32*8	col	row	0.930532	32768.0000	9.014852
DGEMM	2	8*32*64	256*128*64	col	row	0.778397	4194304.0000	0.026639
	3	8*64*64	64*64*8	col	row	0.758239 100	65536.0000	0.302511
	3	8*64*64 Title Size	64*64*8 Matrix Size	col	row a B major	0.758239 100 r Ex Time (ms)	65536.0000 FLOPs	0.302511 GFLOPs
SCLAIM	3	8*64*64 Title Size 32*16*16	64*64*8 Matrix Size 16*16*16	A majo	row a B major row	0.758239 100 r Ex Time (ms) 2.055641	65536.0000	0.302511 GFLOPs 0.003985
	1 2	8*64*64 Title Size 32*16*16 32*16*16	64*64*8 Matrix Size 16*16*16 16*16*32	A majo	row R B major row row	0.758239 100 r Ex Time (ms) 2.055641 1.359971	65536.0000 FLOPs 8192.0000 16384.0000	0.302511 GFLOPs 0.003985 0.012047
	1 2 3	8*64*64 Title Size 32*16*16 32*16*16 32*16*16	Matrix Size 16*16*16 16*16*32 16*16*16	A majo col col col	row R B major row row col	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867	FLOPs 8192.0000 16384.0000 8192.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991
	1 2 3 4	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32	A major col col col col	row B major row row col col	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867 1.346217	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217
	3 1 2 3 4 5	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16	Matrix Size 16*16*16 16*16*32 16*16*32 16*16*32 16*16*16	A majo col col col col row	row R B major row row col col row	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867 1.346217 2.055211	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986
	1 2 3 4 5 6	7itle Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16	Matrix Size 16*16*16 16*16*32 16*16*32 16*16*32 16*16*16 16*16*32	A majo col col col col row row	row R B major row row col col row row row	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867 1.346217 2.055211 2.020667	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108
	1 2 3 4 5 6 7	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16	Matrix Size 16*16*16 16*16*32 16*16*32 16*16*16 16*16*32 16*16*32 16*16*32	A major col col col col row row	row R B major row row col col row row col col col	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867 1.346217 2.055211 2.020667 2.053469	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108 0.003989
	1 2 3 4 5 6 7	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32	A majo col col col col row row	row row row col col row row col col col col col col	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867 1.346217 2.055211 2.020667 2.053469 1.350815	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108 0.003989 0.012129
	1 2 3 4 5 6 7 8	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32	A majo col col col col row row row col	r B major row row col col row col col row row	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867 1.346217 2.055211 2.020667 2.053469 1.350815 1.345174	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108 0.003989 0.012129 0.048719
	1 2 3 4 5 6 7 8 9	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 16*16*32 32*32*32 32*32*32	A majo col col col col row row row col col col	row R major row row col col row row col col row row col	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867 1.346217 2.0525211 2.020667 2.053469 1.350815 1.345174 1.930839	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108 0.003989 0.012129 0.048719 0.033942
	1 2 3 4 5 6 7 8 9	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32 32*32*32 32*32*32	A major col col col col row	row A r B major row row col col row row col col row col col row col row col row	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867 1.346217 2.055211 2.020667 2.053469 1.350815 1.345174 1.930839 1.367125	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108 0.003989 0.012129 0.048719 0.033942 0.047937
DWMMA	3 1 2 3 4 5 6 7 8 9 10 11 12	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32 32*32*32 32*32*32 32*32*32	A major col col col row row col col row	row R major row row col col row row col col row row col col col col row col col col col col col col col col	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867 1.346217 2.055211 2.020667 2.053469 1.350815 1.345174 1.930839 1.367125 1.936322	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 65536.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108 0.003989 0.012129 0.048719 0.033942 0.047937 0.033846
DWMMA	3 1 2 3 4 5 6 7 7 8 9 10 11 12 13	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32	A majo col col col col row row row row row col col col col col	row R major row row col col row row col row row col col row col row col row row col row row	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867 1.346217 2.055211 2.020667 2.053469 1.350815 1.345174 1.930839 1.367125 1.936322 1.69345	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 65536.0000 4194304.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108 0.003989 0.012129 0.048719 0.033942 0.047937 0.033846 2.476778
DWMMA	3 1 2 3 4 5 6 7 8 9 10 11 12 13	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*128*128*128*128*128*128*128*128*128*12	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 16*16*12 16*16*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32	A majo col col col col row row col col col col col col col c	row R major row row col col row row col col row row col col row col row col row col row col row col row	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867 1.346217 2.055211 2.020667 2.053469 1.350815 1.345174 1.930839 1.367125 1.936322 1.69345 1.689855	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108 0.003989 0.012129 0.048719 0.033942 0.047937 0.033846 2.476778 2.48205
DWMMA -	3 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*128*128*128*128*128*128*128*128*128*12	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32	A majo col col col col row row col col col row row row col col row row row row row row row r	row R major row row col col row row col col row col col row col col row col row col row col row col row	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867 1.346217 2.055211 2.020667 2.053469 1.350815 1.345174 1.930839 1.367125 1.936322 1.69345 1.689855 1.708981	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108 0.003989 0.012129 0.048719 0.033942 0.047937 0.033846 2.476778 2.48205 2.454272
DWMMA	3 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*128*128*128*128*128*128*128*128*128*12	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32	A majo col col col row row row col col col col row row row row row row row r	row R major row row col col row col col row col col row col	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867 1.346217 2.055211 2.020667 2.053469 1.350815 1.345174 1.930839 1.367125 1.936322 1.69345 1.689855 1.708981 1.709111	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108 0.003989 0.012129 0.048719 0.033942 0.047937 0.033846 2.476778 2.48205 2.454272 2.454085
DWMMA	3 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*128*128 32*128*128 32*128*128	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32	COI A majo COI COI COI COI COI TOW TOW COI COI TOW	row R major row row col col row col col row col col row col col row col row col row col row col row col row row col row row	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867 1.346217 2.055211 2.020667 2.053469 1.350815 1.345174 1.930839 1.367125 1.936322 1.69345 1.689855 1.708981 1.709111 1.372225	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108 0.003989 0.012129 0.048719 0.033942 0.047937 0.033846 2.476778 2.48205 2.454272 2.454085 12.226287
DWMMA	3 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*12 32*128*128 32*128*128 32*128*128 32*128*128	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32	COI A majo COI COI COI COI FOW FOW FOW COI COI FOW	row Residue to the column of	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867 1.346217 2.055211 2.020667 2.053469 1.350815 1.345174 1.930839 1.367125 1.936322 1.69345 1.689855 1.708981 1.709111 1.372225 1.35374	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 16777216.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108 0.003989 0.012129 0.048719 0.033942 0.047937 0.033846 2.476778 2.48205 2.454272 2.454085 12.226287 12.393234
DWMMA	3 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32	COI A majo COI COI COI COI FOW FOW COI COI FOW FOW FOW COI COI FOW	row Residue to the column of	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867 1.346217 2.052211 2.020667 2.053469 1.350815 1.345174 1.930839 1.367125 1.936322 1.69345 1.689855 1.708981 1.709111 1.372225 1.35374 1.351045	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 16777216.0000 16777216.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108 0.003989 0.012129 0.048719 0.033942 0.047937 0.033846 2.476778 2.48205 2.454272 2.454085 12.226287 12.393234 12.417955
DWMMA	3 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32	COI A major COI COI COI COI FOW FOW COI COI FOW FOW COI COI FOW FOW COI FOW	row Residue to the column of	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052667 1.346217 2.055211 2.020667 2.053469 1.350815 1.345174 1.930839 1.367125 1.936322 1.69345 1.689855 1.708981 1.709111 1.372225 1.35374 1.351045 1.359801	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 16777216.0000 16777216.0000 167777216.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108 0.003989 0.012129 0.048719 0.033942 0.047937 0.033846 2.476778 2.48205 2.454272 2.454085 12.226287 12.393234 12.417955 12.337994
DWMMA	3 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*12 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32	COI A major COI COI COI COI FOW FOW COI COI COI FOW FOW FOW COI COI FOW FOW FOW FOW FOW COI FOW	row Residue to the column of	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052667 1.346217 2.055211 2.020667 2.053469 1.350815 1.345174 1.930839 1.367125 1.936322 1.69345 1.689855 1.708981 1.709111 1.372225 1.35374 1.351045 1.359801 1.363435	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 16777216.0000 16777216.0000 16777216.0000 16777216.0000 16777216.0000 16777216.0000 16777216.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108 0.008108 0.003989 0.012129 0.048719 0.033942 0.047937 0.033846 2.476778 2.48205 2.454272 2.454085 12.226287 12.393234 12.417955 12.337994 24.610218
DWMMA	3 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32*32 32*32*32*32 32*32*32 32*32*32*32 32*32*32*32 32*32*32*32 32*32*32*32 32*32*32*32 32*32*32*32 32*32*32*32	COI A major COI COI COI COI FOW FOW COI COI FOW FOW COI COI FOW	row Residue to the column of	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867 1.346217 2.055211 2.020667 2.053469 1.350815 1.345174 1.930839 1.367125 1.936322 1.69345 1.689855 1.708981 1.709111 1.372225 1.35374 1.351045 1.359801 1.363435 1.346884	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 16777216.0000 16777216.0000 167777216.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108 0.003989 0.012129 0.048719 0.033942 0.047937 0.033846 2.476778 2.48205 2.454272 2.454085 12.226287 12.393234 12.417955 12.337994
DWMMA	3 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*12 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32*32 32*32*32*32 32*32*32 32*32*32*32 32*32*32*32 32*32*32*32 32*32*32*32 32*32*32*32 32*32*32*32 32*32*32*32 32*32*32*32 32*32*32*32	COI A major COI COI COI COI FOW FOW COI COI FOW FOW COI COI FOW	row Residue to the column of	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052667 1.346217 2.055211 2.020667 2.053469 1.350815 1.345174 1.930839 1.367125 1.936322 1.69345 1.689855 1.708981 1.709111 1.372225 1.35374 1.351045 1.359801 1.363435	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 16777216.0000 16777216.0000 16777216.0000 16777216.0000 16777216.0000 16777216.0000 16777216.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108 0.008108 0.003989 0.012129 0.048719 0.033942 0.047937 0.033846 2.476778 2.48205 2.454272 2.454085 12.226287 12.393234 12.417955 12.337994 24.610218
DWMMA	3 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128 32*128*128	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32 32*	A major col col col col row row col col row	row Residue to the column of	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867 1.346217 2.055211 2.020667 2.053469 1.350815 1.345174 1.930839 1.367125 1.936322 1.69345 1.689855 1.708981 1.709111 1.372225 1.35374 1.351045 1.359801 1.363435 1.346884	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 66777216.0000 16777216.0000 16777216.0000 16777216.0000 33554432.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108 0.008108 0.003989 0.012129 0.048719 0.033942 0.047937 0.033846 2.476778 2.48205 2.454272 2.454085 12.226287 12.393234 12.417955 12.337994 24.610218 24.912637
DWMMA	3 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*128*128	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32*32 32*32*32 32*32*32 32*32*32 32*32*32 32*32*32*32 32*32*32*32 32*32*32*32 32*32*32*32 32*32*32*32 32*32*32*32 32*32*32*32 32*32*32*32*32 32*32*32*32*32*32*32*32*32*32*32*32*32*3	A majo col col col col row row row col col col col col row row col col row row col col row row col col row row row row row row row r	row R major row row col col row	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867 1.346217 2.055211 2.020667 2.053469 1.350815 1.345174 1.930839 1.367125 1.936322 1.69345 1.689855 1.708981 1.709111 1.372225 1.35374 1.351045 1.359801 1.363435 1.346884 1.352596	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 16777216.0000 16777216.0000 16777216.0000 16777216.0000 33554432.0000 33554432.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108 0.003989 0.012129 0.048719 0.033942 0.047937 0.033846 2.476778 2.48205 2.454272 2.454085 12.226287 12.393234 12.417955 12.337994 24.610218 24.912637 24.809265
DWMMA	3 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*128*128	Matrix Size 16*16*16 16*16*32 16*16*16 16*16*32 16*16*16 16*16*32 32*32*32 32*	A majo col col col col row row row row col col row row row row row row row r	row Residue to the column of	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867 1.346217 2.055211 2.020667 2.053469 1.350815 1.345174 1.930839 1.367125 1.936322 1.69345 1.689855 1.708981 1.709111 1.372225 1.35374 1.351045 1.35374 1.363435 1.346884 1.352596 1.350902	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 16777216.0000 16777216.0000 16777216.0000 16777216.0000 33554432.0000 33554432.0000 33554432.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108 0.003989 0.012129 0.048719 0.033942 0.047937 0.033846 2.476778 2.48205 2.454272 2.454085 12.226287 12.393234 12.417955 12.337994 24.610218 24.912637 24.809265 24.838539
DWMMA	3 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 1	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*12 32*128*128	Matrix Size	A majo col col col col row row row row row row row r	row Residue to the column of	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867 1.346217 2.055211 2.020667 2.053469 1.350815 1.345174 1.930839 1.367125 1.936322 1.69345 1.69345 1.709111 1.372225 1.35374 1.351045 1.359801 1.363435 1.346884 1.352596 1.350902 1.351123 1.340984	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 16777216.0000 16777216.0000 16777216.0000 16777216.0000 33554432.0000 33554432.0000 33554432.0000 8388608.0000 20736.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108 0.003989 0.012129 0.048719 0.033942 0.047937 0.033846 2.476778 2.48205 2.454272 2.454085 12.226287 12.393234 12.417955 12.337994 24.610218 24.912637 24.809265 24.838539 6.208619 0.015463
DWMMA	3 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 1 2 3	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*12 32*128*128	Matrix Size	COI A majo COI COI COI COI COI FOW	row Residue to the column of	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867 1.346217 2.055211 2.020667 2.053469 1.350815 1.345174 1.930839 1.367125 1.936322 1.69345 1.689855 1.708981 1.709111 1.372225 1.35374 1.351045 1.359801 1.363435 1.346884 1.352596 1.350902 1.351123 1.340984 1.340984 1.340984	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 4194304.0000 16777216.0000 16777216.0000 16777216.0000 16777216.0000 33554432.0000 33554432.0000 33554432.0000 8388608.0000 20736.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108 0.003989 0.012129 0.048719 0.033942 0.047937 0.033846 2.476778 2.48205 2.454272 2.454085 12.226287 12.393234 12.417955 12.337994 24.610218 24.912637 24.809265 24.838539 6.208619 0.015463 0.170048
DWMMA	3 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 1	8*64*64 Title Size 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*16 32*16*12 32*128*128	Matrix Size	A majo col col col col row row row row row col col col col col col col c	row R major row row col col row col col row row col row row col row row row col row row	0.758239 100 r Ex Time (ms) 2.055641 1.359971 2.052867 1.346217 2.055211 2.020667 2.053469 1.350815 1.345174 1.930839 1.367125 1.936322 1.69345 1.69345 1.709111 1.372225 1.35374 1.351045 1.359801 1.363435 1.346884 1.352596 1.350902 1.351123 1.340984	FLOPs 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 8192.0000 16384.0000 65536.0000 65536.0000 65536.0000 4194304.0000 4194304.0000 4194304.0000 16777216.0000 16777216.0000 16777216.0000 16777216.0000 33554432.0000 33554432.0000 33554432.0000 8388608.0000 20736.0000	0.302511 GFLOPs 0.003985 0.012047 0.003991 0.01217 0.003986 0.008108 0.003108 0.003989 0.012129 0.048719 0.033942 0.047937 0.033846 2.476778 2.48205 2.454272 2.454085 12.226287 12.393234 12.417955 12.337994 24.610218 24.912637 24.809265 24.838539 6.208619 0.015463

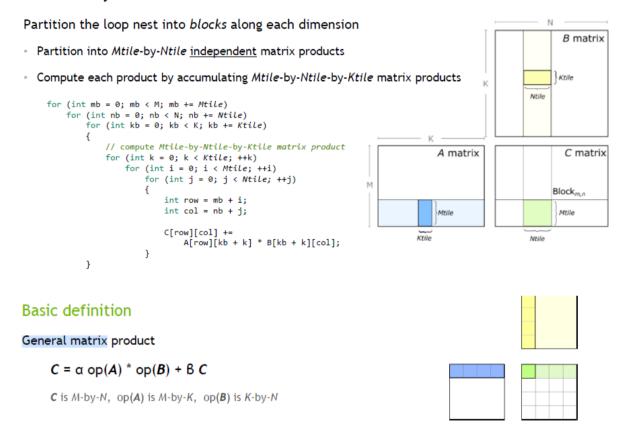
Check all the result inside the excel sheets submitted.

Analysis:

What is the tile size, matrix size, and column/row-major of the input matrices for each of the test?

ANS: The column/column major or row/row major will increase execution time. The pipeline stalls due to memory is not affected by the column/row-major order a lot.

Tile is the parameter shows that how the matric be partitioned into independent matric products in each dimension and the matrix is the products shows in three dimensions. And the matric a and b is the layout of the matrix defined in the cutlass.



Your analysis to the results, including the comparison of the two GPU performance and the influence of GEMM configurations to the computation performance. You may use tables and charts to assist your analysis.

ANS: In conclusion, the performance of a40 is much better than a100 with the comparation based on the GFLOPs.

As for the GEMM configuration, the GPU with a larger matrix size or saying with a lager c matrix size or K will have a higher GFLOPs; while with the same size configuration, the major

A/B do not affect much on the performance; the GPU with the larger matrix A/B will perform better; the performance is growing in a linear way with the parameter matrix C/K; and the tile size and matrix size should be matched or saying that the matrix should not be partitioned too much so that the performance can be promised, which means the tile size should be smaller than matrix to gain a better performance.