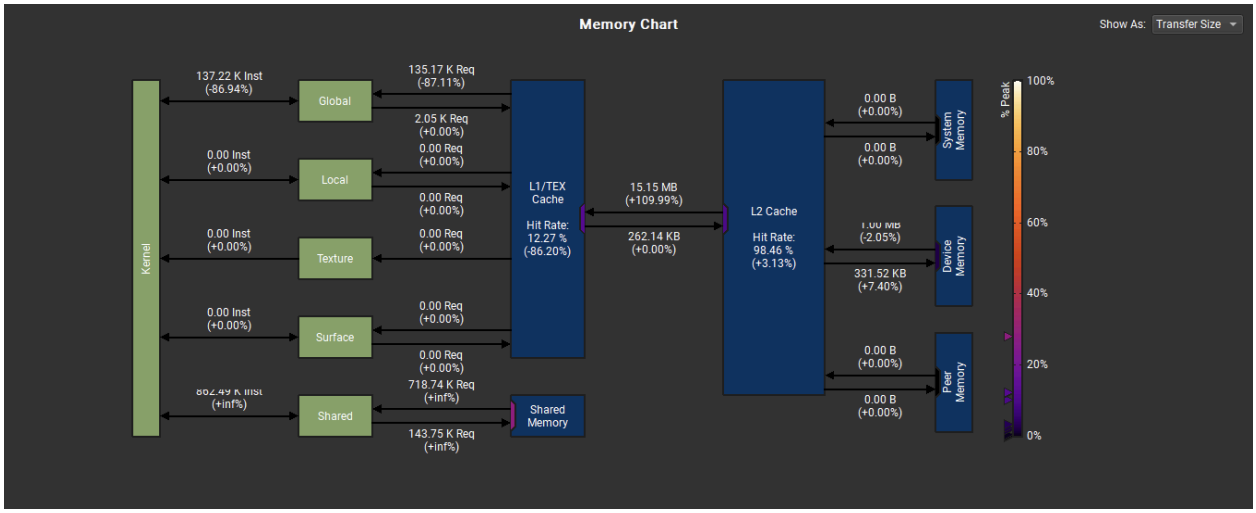
8x8 analysis
Between Naive and tiled, there are quite a few differences. Particularly, when you see the Shared mem, L2, L1, and device mem statistics.

We see that in the naive, there is no shared memory usage, as expected.

In tiled, there is a lot more transfer of memory from L2 to L1 but the same vice versa. From the kernel, there is a lot less transfer of data through global, and a lot fewer bytes of data requested from L1 to Global. Device memory to L2 cache transfer is also different. Lesser data to L2 and more data to Device Mem.

There is a significant drop in hitrate for L1 Cache. And a slight improvement in hitrate for L2 cache. There are fewer actions and operations for sectors and wavefronts in the L1 Level. L2 Level also has more sector operations.

Tiled



Shared Memory

	Instructions	Requests	Wavefronts	% Peak	Bank Conflicts
Shared Load	718,740 (+inf%)	718,740 (+inf%)	862,488 (+inf%)	22.30 (+inf%)	0 (+0.00%)
Shared Load Matrix	0 (+0.00%)	0 (+0.00%)			
Shared Store	143,748 (+inf%)	143,748 (+inf%)	143,748 (+inf%)	3.72 (+inf%)	0 (+0.00%)
Shared Atomic	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
Other	-	-	84,810 (+1,734.13%)	2.19 (+1,739.00%)	0 (+0.00%)
Total	862,488 (+inf%)	862,488 (+inf%)	1,091,046 (+23,495.29%)	28.21 (+23,558.05%)	0 (+0.00%)

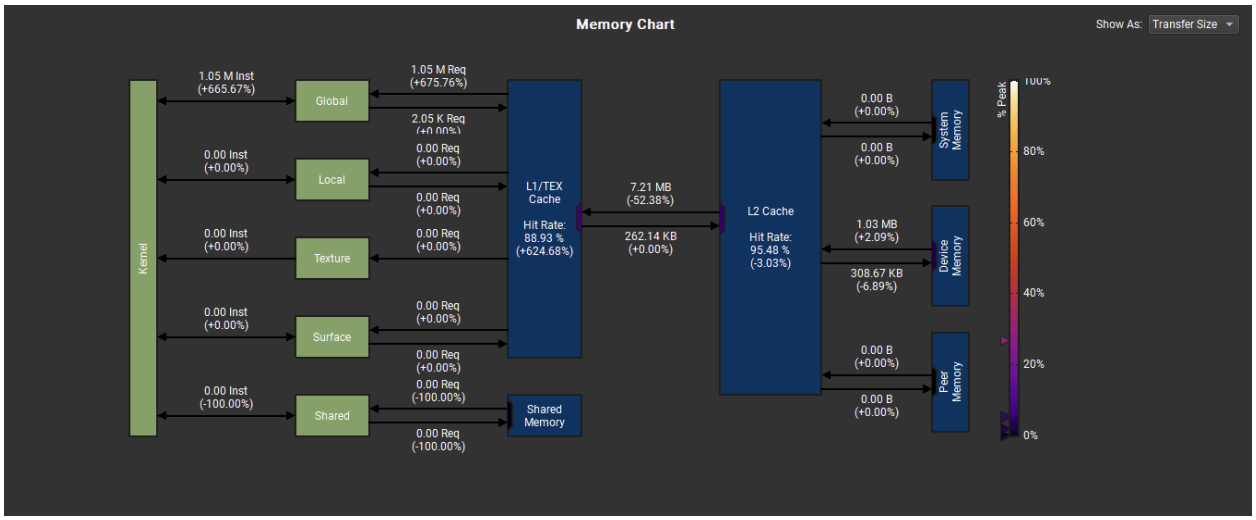
L1/TEX Cache

	Instructions	Requests	Wavefronts	% Peak	Sectors	Sectors/Req	Hit Rate	Bytes	Sector Misses to L2	% Peak
Local Load									470,972 (+109.44%)	
Global Load	135,168 (-87.11%)	135,168 (-87.11%)	135,168 (-87.11%)	3.49 (-87.08%)	536,853 (-74.40%)	3.97 (+98.59%)	12.83 (-85.62%)	17,179,296 (-74.40%)		(+1
Surface Load	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Texture Load	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Store	2,048 (+0.00%)	2,048 (+0.00%)	2,048 (+0.00%)	0.05 (+0.27%)	8,192 (+0.00%)	4 (+0.00%)	13.13 (+15.95%)	262,144 (+0.00%)	8,192 (+0.00%)	0.21 (
Local Store	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)		
Surface Store	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Reduction	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
DSMEM Reduction						-	-		0 (+0.00%)	0 (
Surface Reduction	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Atomic ALU	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Atomic CAS	0 (+0.00%)									
Surface Atomic ALU	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Surface Atomic CAS	0 (+0.00%)									
Loads	135,168 (-87.11%)	135,168 (-87.11%)	135,168 (-87.11%)	3.49 (-87.08%)	536,853 (-74.40%)	3.97 (+98.59%)	12.83 (-85.62%)	17,179,296 (-74.40%)	470,972 (+109.44%)	(+1
Stores	2,048 (+0.00%)	2,048 (+0.00%)	2,048 (+0.00%)	0.05 (+0.27%)	8,192 (+0.00%)	4 (+0.00%)	13.13 (+15.95%)	262,144 (+0.00%)	8,192 (+0.00%)	0.21 (
Atomics & Reductions	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Total	137,216 (-86.94%)	137,216 (-86.94%)	137,216 (-86.94%)	3.55 (-86.90%)	545,045 (-74.11%)	3.97 (+98.22%)	12.83 (-85.57%)	17,441,440 (-74.11%)	479,164 (+105.59%)	(+1

L2 Cache

	Requests	Sectors	Sectors/Req	% Peak	Hit Rate	Bytes	Throughput	Sector Misses to Device	Sector Misses to System
L1/TEX Load				(+107.30%)	(+3.99%)				
L1/TEX Store	8,128 (+98.44%)	8,192 (+0.00%)	1.01 (-49.61%)	0.18 (-0.18%)	100 (+0.00%)	262,144 (+0.00%)	1,542,459,047.26 (-0.15%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Atomic ALU	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Atomic CAS	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Reduction	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Total	478,172 (+185.46%)	480,433 (+107.38%)	1.00 (-27.35%)	10.33 (+107.01%)	96.64 (+3.72%)	15,373,856 (+107.38%)	90,459,988,702.69 (+107.06%)	16,384 (+0.00%)	0 (+0.00%)
ECC Total	-	8 (+0.00%)	-	0.00 (-0.18%)	-	256 (+0.00%)	1,506,307.66 (-0.15%)	8 (+0.00%)	-
GPU Total	477,820 (+185.51%)	480,642 (+103.60%)	1.01 (-28.69%)	10.34 (+103.24%)	96.63 (+3.73%)	15,380,544 (+103.60%)	90,499,340,990.40 (+103.29%)	16,491 (-0.15%)	30 (+0.00%)

Naive



Shared Memory

	Instructions	Requests	Wavefronts	% Peak	Bank Conflicts
Shared Load	0 (-100.00%)	0 (-100.00%)	0 (-100.00%)	0 (-100.00%)	0 (+0.00%)
Shared Load Matrix	0 (+0.00%)	0 (+0.00%)			
Shared Store	0 (-100.00%)	0 (-100.00%)	0 (-100.00%)	0 (-100.00%)	0 (+0.00%)
Shared Atomic	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
Other	-	-	4,624 (-94.55%)	0.12 (-94.56%)	0 (+0.00%)
Total	0 (-100.00%)	0 (-100.00%)	4,624 (-99.58%)	0.12 (-99.58%)	0 (+0.00%)

L1/TEX Cache

	Instructions	Requests	Wavefronts	% Peak	Sectors	Sectors/Req	Hit Rate	Bytes	Sector Misses to L2	% Peak
Local Load	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	224,872 (-52.25%)	5.80 (-
Global Load	1,048,576 (+675.76%)	1,048,576 (+675.76%)	1,048,576 (+675.76%)	27.04 (+673.70%)	2,097,147 (+290.64%)	2.00 (-49.64%)	89.24 (+595.56%)	67,108,704 (+290.64%)		
Surface Load	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Texture Load	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Store	2,048 (+0.00%)	2,048 (+0.00%)	2,048 (+0.00%)	0.05 (-0.27%)	8,192 (+0.00%)	4 (+0.00%)	11.33 (-13.75%)	262,144 (+0.00%)	8,192 (+0.00%)	0.21
Local Store	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)		
Surface Store	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Reduction	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
DSMEM Reduction						-	-		0 (+0.00%)	0 (
Surface Reduction	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Atomic ALU	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Atomic CAS	0 (+0.00%)									
Surface Atomic ALU	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Surface Atomic CAS	0 (+0.00%)									
Loads	1,048,576 (+675.76%)	1,048,576 (+675.76%)	1,048,576 (+675.76%)	27.04 (+673.70%)	2,097,147 (+290.64%)	2.00 (-49.64%)	89.24 (+595.56%)	67,108,704 (+290.64%)	224,872 (-52.25%)	5.80 (-
Stores	2,048 (+0.00%)	2,048 (+0.00%)	2,048 (+0.00%)	0.05 (-0.27%)	8,192 (+0.00%)	4 (+0.00%)	11.33 (-13.75%)	262,144 (+0.00%)	8,192 (+0.00%)	0.21
Atomics & Reductions	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Total	1,050,624 (+665.67%)	1,050,624 (+665.67%)	1,050,624 (+665.67%)	27.09	2,105,339 (+286.27%)	2.00 (-49.55%)	88.93	67,370,848 (+286.27%)	233,064 (-51.36%)	6.01 (-

L2 Cache

	Requests	Sectors	Sectors/Req	% Peak	Hit Rate	Bytes	Throughput	Sector Misses to Device	Sector Misses to System
L1/TEX Load	[illegible]	[illegible]	(+35.52%)	[illegible]	(-3.84%)	[illegible]	[illegible]	[illegible]	[illegible]
L1/TEX Store	4,096 (-49.61%)	8,192 (+0.00%)	2 (+98.44%)	0.18 (+0.18%)	100 (+0.00%)	262,144 (+0.00%)	1,544,785,970.21 (+0.15%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Atomic ALU	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Atomic CAS	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Reduction	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Total	167,509 (-64.97%)	231,671 (-51.78%)	1.38 (+37.65%)	4.99 (-51.69%)	93.18 (-3.58%)	7,413,472 (-51.78%)	43,686,781,067.32 (-51.71%)	16,384 (+0.00%)	0 (+0.00%)
ECC Total	-	8 (+0.00%)	-	0.00 (+0.18%)	-	256 (+0.00%)	1,508,580.05 (+0.15%)	8 (+0.00%)	-
GPU Total	167,355 (-64.98%)	236,077 (-50.88%)	1.41 (+40.24%)	5.09 (-50.80%)	93.15 (-3.60%)	7,554,464 (-50.88%)	44,517,631,529.32 (-50.81%)	16,515 (+0.15%)	30 (+0.00%)

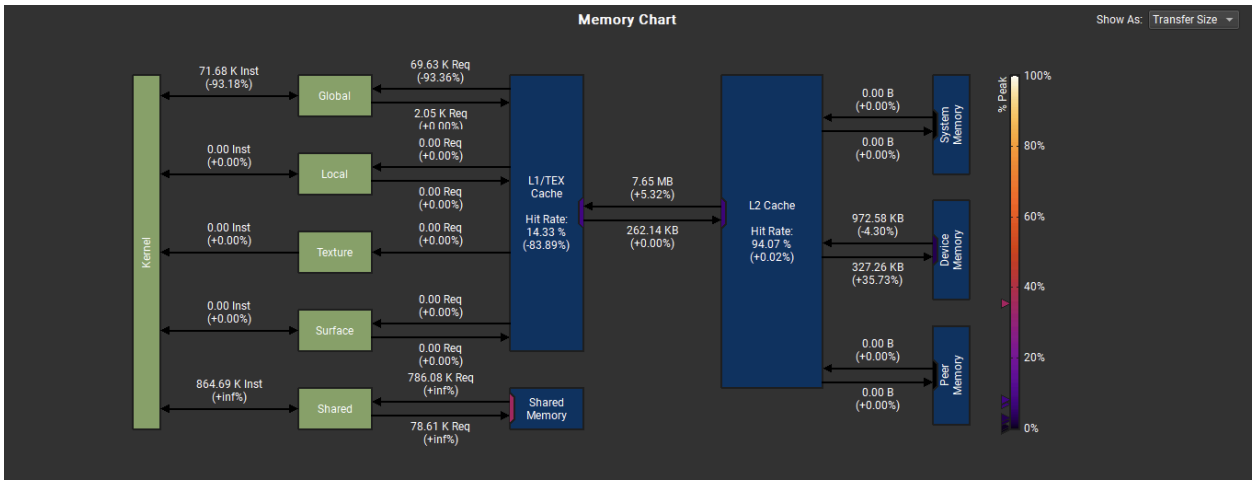
16x16 analysis
Between Naive and tiled, there are quite a few differences. Particularly, when you see the Shared mem, L2, L1, and device mem statistics.

We see that in the naive, there is no shared memory usage, as expected.

In tiled, there is a lot more transfer of memory from L2 to L1 but the same vice versa. From the kernel, there is a lot less transfer of data through global, and a lot fewer bytes of data requested from L1 to Global. Device memory to L2 cache transfer is also different. Lesser data to L2 and more data to Device Mem.

There is a significant drop in hitrate for L1 Cache. And a slight improvement in hitrate for L2 cache. There are fewer actions and operations for sectors and wavefronts in the L1 Level. L2 Level also has fewer sector operations.

Tiled



Shared Memory

	Instructions	Requests	Wavefronts	% Peak	Bank Conflicts
Shared Load	786,080 (+inf%)	786,080 (+inf%)	943,296 (+inf%)	32.19 (+inf%)	0 (+0.00%)
Shared Load Matrix	0 (+0.00%)	0 (+0.00%)			
Shared Store	78,608 (+inf%)	78,608 (+inf%)	78,608 (+inf%)	2.68 (+inf%)	0 (+0.00%)
Shared Atomic	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
Other	-	-	23,426 (+406.62%)	0.80 (+576.69%)	0 (+0.00%)
Total	864,688 (+inf%)	864,688 (+inf%)	1,045,330 (+22,506.62%)	35.67 (+30,095.65%)	0 (+0.00%)

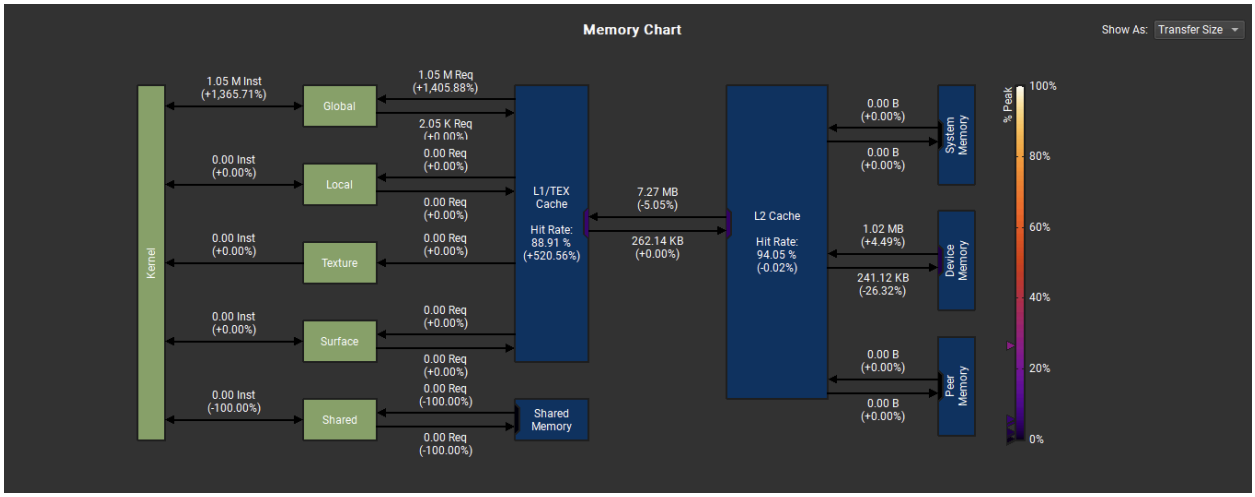
L1/TEX Cache

	Instructions	Requests	Wavefronts	% Peak	Sectors	Sectors/Req	Hit Rate	Bytes	Sector Misses to L2	% Peak
Global Load	69,632 (-93.36%)	69,632 (-93.36%)	69,632 (-93.36%)	2.38 (-91.13%)	276,126 (-86.83%)	3.97 (+98.28%)	15.17 (-83.00%)	8,836,032 (-86.83%)	238,674 (+5.23%)	8.15 (+
Surface Load	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Texture Load	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Store	2,048 (+0.00%)	2,048 (+0.00%)	2,048 (+0.00%)	0.07 (+33.57%)	8,192 (+0.00%)	4 (+0.00%)	3.96 (-61.79%)	262,144 (+0.00%)	8,192 (+0.00%)	0.28 (+
Local Store	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)		
Surface Store	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Reduction	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
DSMEM Reduction						-	-		0 (+0.00%)	0 (
Surface Reduction	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Atomic ALU	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Atomic CAS	0 (+0.00%)									
Surface Atomic ALU	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Surface Atomic CAS	0 (+0.00%)									
Loads	69,632 (-93.36%)	69,632 (-93.36%)	69,632 (-93.36%)	2.38 (-91.13%)	276,126 (-86.83%)	3.97 (+98.28%)	15.17 (-83.00%)	8,836,032 (-86.83%)	238,674 (+5.23%)	8.15 (+
Stores	2,048 (+0.00%)	2,048 (+0.00%)	2,048 (+0.00%)	0.07 (+33.57%)	8,192 (+0.00%)	4 (+0.00%)	3.96 (-61.79%)	262,144 (+0.00%)	8,192 (+0.00%)	0.28 (+
Atomics & Reductions	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Total	71,680 (-93.18%)	71,680 (-93.18%)	71,680 (-93.18%)	2.45 (-90.89%)	284,318 (-86.50%)	3.97 (+97.94%)	14.85 (-83.31%)	9,098,176 (-86.50%)	246,866 (+5.05%)	8.42 (+

L2 Cache

	Requests	Sectors	Sectors/Req	% Peak	Hit Rate	Bytes	Throughput	Sector Misses to Device	Sector Misses to System
L1/TEX Load	119,103 (-25.04%)	238,894 (+6.23%)	2.01 (+41.71%)	6.89 (+41.84%)	93.13 (+0.20%)	7,644,608 (+6.23%)	60,326,767,676.77 (+41.82%)	16,384 (+0.00%)	0 (+0.00%)
L1/TEX Store	4,096 (+0.05%)	8,192 (+0.00%)	2 (-0.05%)	0.24 (+33.53%)	100 (+0.00%)	262,144 (+0.00%)	2,068,686,868.69 (+33.51%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Atomic ALU	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Atomic CAS	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Reduction	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Total	123,610 (-25.36%)	245,622 (+6.21%)	1.99 (+42.30%)	7.09 (+41.82%)	93.37 (+0.30%)	7,859,904 (+6.21%)	62,025,757,575.76 (+41.79%)	16,384 (+0.00%)	0 (+0.00%)
ECC Total	-	8 (+0.00%)	-	0.00 (+33.53%)	-	256 (+0.00%)	2,020,202.02 (+33.51%)	8 (+0.00%)	-
GPU Total	123,583 (-26.36%)	246,987 (+4.41%)	2.00	7.13 (+39.42%)	93.35	7,903,584 (+4.41%)	62,370,454,545.45 (+39.40%)	16,499 (-0.10%)	23 (-23.33%)

Naive



Shared Memory

	Instructions	Requests	Wavefronts	% Peak	Bank Conflicts
Shared Load	0 (-100.00%)	0 (-100.00%)	0 (-100.00%)	0 (-100.00%)	0 (+0.00%)
Shared Load Matrix	0 (+0.00%)	0 (+0.00%)			
Shared Store	0 (-100.00%)	0 (-100.00%)	0 (-100.00%)	0 (-100.00%)	0 (+0.00%)
Shared Atomic	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
Other	-	-	4,624 (-80.26%)	0.12 (-85.22%)	0 (+0.00%)
Total	0 (-100.00%)	0 (-100.00%)	4,624 (-99.56%)	0.12 (-99.67%)	0 (+0.00%)

L1/TEX Cache

	Instructions	Requests	Wavefronts	% Peak	Sectors	Sectors/Req	Hit Rate	Bytes	Sector Misses to L2	% Peak
Local Load	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	226,808 (-4.97%)	5.80 (-
Global Load	1,048,576 (+1,405.88%)	1,048,576 (+1,405.88%)	1,048,576 (+1,405.88%)	26.79 (+1,027.41%)	2,097,149 (+659.49%)	2.00 (-49.57%)	89.23 (+488.29%)	67,108,768 (+659.49%)		
Surface Load	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Texture Load	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Store	2,048 (+0.00%)	2,048 (+0.00%)	2,048 (+0.00%)	0.05 (-25.13%)	8,192 (+0.00%)	4 (+0.00%)	10.35 (+161.73%)	262,144 (+0.00%)	8,192 (+0.00%)	0.21 (-
Local Store	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)		
Surface Store	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Reduction	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
DSMEM Reduction						-	-		0 (+0.00%)	0 (
Surface Reduction	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Atomic ALU	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Atomic CAS	0 (+0.00%)									
Surface Atomic ALU	0 (+0.00%)									
Surface Atomic CAS	0 (+0.00%)									
Loads	1,048,576 (+1,405.88%)	1,048,576 (+1,405.88%)	1,048,576 (+1,405.88%)	26.79 (+1,027.41%)	2,097,149 (+659.49%)	2.00 (-49.57%)	89.23 (+488.29%)	67,108,768 (+659.49%)	226,808 (-4.97%)	5.80 (-
Stores	2,048 (+0.00%)	2,048 (+0.00%)	2,048 (+0.00%)	0.05 (-25.13%)	8,192 (+0.00%)	4 (+0.00%)	10.35 (+161.73%)	262,144 (+0.00%)	8,192 (+0.00%)	0.21 (-
Atomics & Reductions	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Total	1,050,624 (+1,365.71%)	1,050,624 (+1,365.71%)	1,050,624 (+1,365.71%)	26.84	2,105,341 (+640.49%)	2.00 (-49.48%)	88.93	67,370,912 (+640.49%)	235,000 (-4.81%)	6.00 (-

L2 Cache

	Requests	Sectors	Sectors/Req	% Peak	Hit Rate	Bytes	Throughput	Sector Misses to Device	Sector Misses to System
L1/TEX Load	158,894 (+33.41%)	224,894 (-5.86%)	1.42 (-29.44%)	4.86 (-29.50%)	92.95 (-0.20%)	7,196,608 (-5.86%)	42,537,166,635.14 (-29.49%)	16,384 (+0.00%)	0 (+0.00%)
L1/TEX Store	4,094 (-0.05%)	8,192 (+0.00%)	2.00 (+0.05%)	0.18 (-25.11%)	100 (+0.00%)	262,144 (+0.00%)	1,549,460,941.93 (-25.10%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Atomic ALU	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Atomic CAS	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Reduction	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Total	165,617 (+33.98%)	231,271 (-5.84%)	1.40 (-29.72%)	5.00 (-29.49%)	93.10 (-0.30%)	7,400,672 (-5.84%)	43,743,332,702.86 (-29.48%)	16,384 (+0.00%)	0 (+0.00%)
ECC Total	-	8 (+0.00%)	-	0.00 (-25.11%)	-	256 (+0.00%)	1,513,145.45 (-25.10%)	8 (+0.00%)	-
GPU Total	167,820 (+35.80%)	236,559 (-4.22%)	1.41 (-29.47%)	5.11 (-28.27%)	93.06	7,569,888 (-4.22%)	44,743,521,846.04 (-28.26%)	16,515 (+0.10%)	30 (+30.43%)

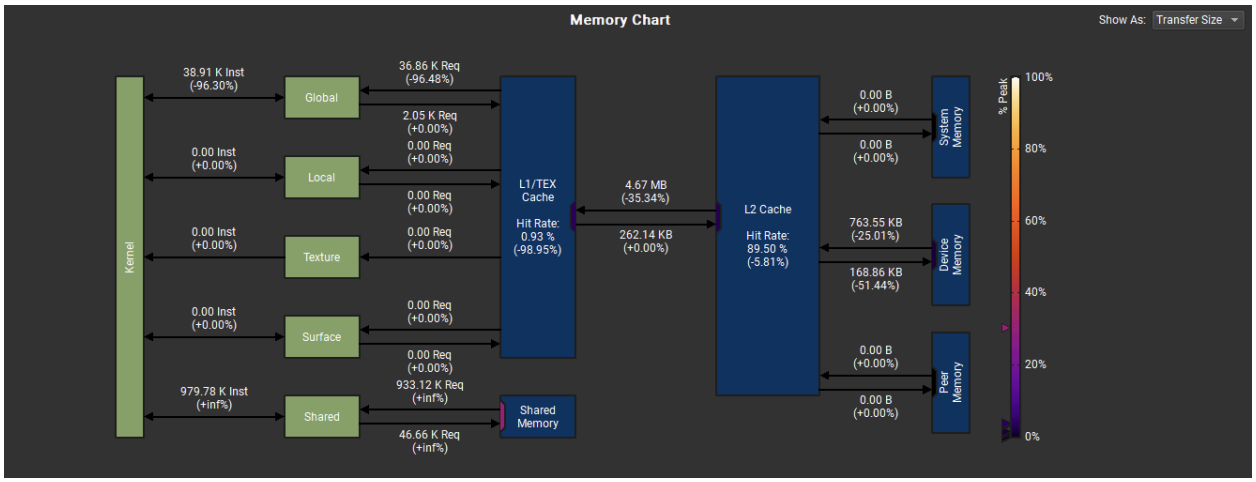
32x32 analysis
Between Naive and tiled, there are quite a few differences. Particularly, when you see the Shared mem, L2, L1, and device mem statistics.

We see that in the naive, there is no shared memory usage, as expected.

In tiled, there is a lot less transfer of memory from L2 to L1 but the same vice versa. From the kernel, there is a lot less transfer of data through global, and a lot fewer bytes of data requested from L1 to Global. Device memory to L2 cache transfer is also different. Lesser data to L2 and also lesser data to Device Mem.

There is a significant drop in hitrate for L1 Cache. And a slight reduction of hitrate for L2 cache. There are fewer actions and operations for sectors and wavefronts in the L1 Level. L2 Level also has fewer sector operations.

Tiled



Shared Memory

	Instructions	Requests	Wavefronts	% Peak	Bank Conflicts
Shared Load	933,120 (+inf%)	933,120 (+inf%)	1,119,744 (+inf%)	28.86 (+inf%)	0 (+0.00%)
Shared Load Matrix	0 (+0.00%)	0 (+0.00%)			
Shared Store	46,656 (+inf%)	46,656 (+inf%)	46,656 (+inf%)	1.20 (+inf%)	0 (+0.00%)
Shared Atomic	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
Other	-	-	16,434 (+255.41%)	0.42 (+258.52%)	0 (+0.00%)
Total	979,776 (+inf%)	979,776 (+inf%)	1,182,834 (+25,480.32%)	30.49 (+25,704.28%)	0 (+0.00%)

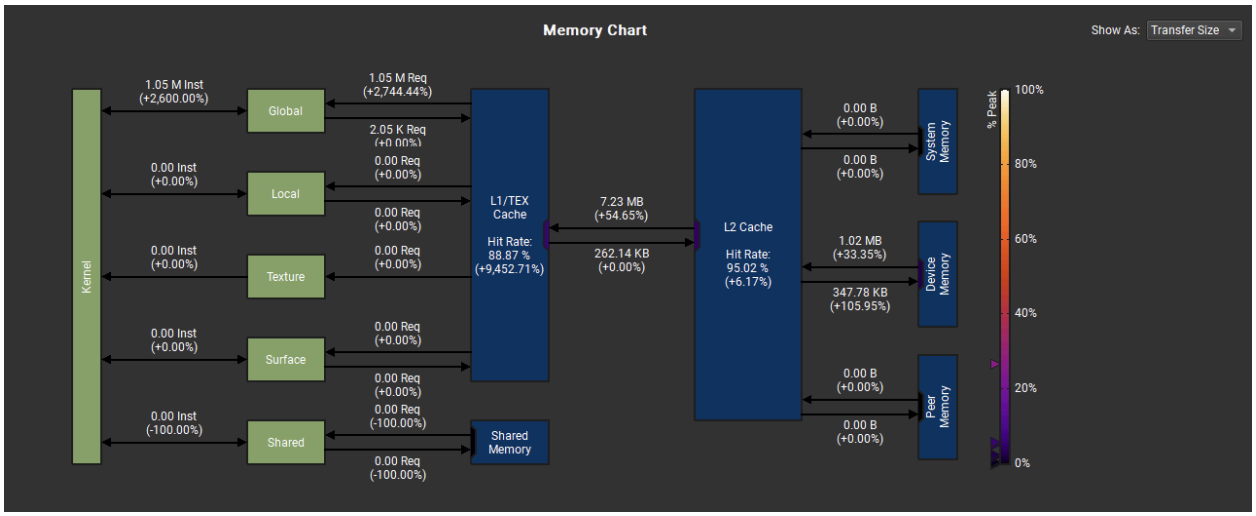
L1/TEX Cache

	Instructions	Requests	Wavefronts	% Peak	Sectors	Sectors/Req	Hit Rate	Bytes	Sector Misses to L2	% Peak
Local Load	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	145,120 (-35.91%)	3.74 (-
Global Load	36,864 (-96.48%)	36,864 (-96.48%)	36,864 (-96.48%)	0.95 (-96.45%)	147,456 (-92.97%)	4 (+100.00%)	1.48 (-98.35%)	4,718,592 (-92.97%)		
Surface Load	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Texture Load	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Store	2,048 (+0.00%)	2,048 (+0.00%)	2,048 (+0.00%)	0.05 (+0.88%)	8,192 (+0.00%)	4 (+0.00%)	0 (-100.00%)	262,144 (+0.00%)	8,192 (+0.00%)	0.21 (
Local Store	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)		
Surface Store	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Reduction	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
DSMEM Reduction						-	-		0 (+0.00%)	0 (
Surface Reduction	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Atomic ALU	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Atomic CAS	0 (+0.00%)									
Surface Atomic ALU	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Surface Atomic CAS	0 (+0.00%)									
Loads	36,864 (-96.48%)	36,864 (-96.48%)	36,864 (-96.48%)	0.95 (-96.45%)	147,456 (-92.97%)	4 (+100.00%)	1.48 (-98.35%)	4,718,592 (-92.97%)	145,120 (-35.91%)	3.74 (-
Stores	2,048 (+0.00%)	2,048 (+0.00%)	2,048 (+0.00%)	0.05 (+0.88%)	8,192 (+0.00%)	4 (+0.00%)	0 (-100.00%)	262,144 (+0.00%)	8,192 (+0.00%)	0.21 (
Atomics & Reductions	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (

L2 Cache

	Requests	Sectors	Sectors/Req	% Peak	Hit Rate	Bytes	Throughput	Sector Misses to Device	Sector Misses to System
L1/TEX Load	36,417 (-77.39%)	145,544 (-35.88%)	4.00 (+183.54%)	3.14 (-36.16%)	88.84 (-4.34%)	4,657,408 (-35.88%)	27,461,132,075.47 (-36.17%)	16,384 (+0.00%)	0 (+0.00%)
L1/TEX Store	2,048 (-49.98%)	8,192 (+0.00%)	4 (+99.90%)	0.18 (-0.43%)	100 (+0.00%)	262,144 (+0.00%)	1,545,660,377.36 (-0.45%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Atomic ALU	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Atomic CAS	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Reduction	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Total	38,782 (-76.86%)	154,864 (-33.91%)	3.99 (+185.60%)	3.34 (-34.20%)	89.44 (-4.01%)	4,955,648 (-33.91%)	29,219,622,641.51 (-34.21%)	16,384 (+0.00%)	0 (+0.00%)
ECC Total	-	8 (+0.00%)	-	0.00 (-0.43%)	-	256 (+0.00%)	1,509,433.96 (-0.45%)	8 (+0.00%)	-
GPU Total	45,168 (-73.07%)	172,040 (-27.38%)	3.81	3.71 (-27.70%)	90.28	5,505,280 (-27.38%)	32,460,377,358.49 (-27.71%)	16,523 (+0.05%)	30 (-6.25%)

Naive



Shared Memory

	Instructions	Requests	Wavefronts	% Peak	Bank Conflicts
Shared Load	0 (-100.00%)	0 (-100.00%)	0 (-100.00%)	0 (-100.00%)	0 (+0.00%)
Shared Load Matrix	0 (+0.00%)	0 (+0.00%)			
Shared Store	0 (-100.00%)	0 (-100.00%)	0 (-100.00%)	0 (-100.00%)	0 (+0.00%)
Shared Atomic	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
Other	-	-	4,624 (-71.86%)	0.12 (-72.11%)	0 (+0.00%)
Total	0 (-100.00%)	0 (-100.00%)	4,624 (-99.61%)	0.12 (-99.61%)	0 (+0.00%)

L1/TEX Cache

	Instructions	Requests	Wavefronts	% Peak	Sectors	Sectors/Req	Hit Rate	Bytes	Sector Misses to L2	% Peak
Local Load	Number of instructions issued. Equal to the number of requests made to the L1/TEX cache			%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)		
Global Load	1,048,576 (+2,744.44%)	1,048,576 (+2,744.44%)	1,048,576 (+2,744.44%)	26.79 (+2,719.76%)	2,097,148 (+1,322.22%)	2.00 (-50.00%)	89.30 (+5,951.17%)	67,108,736 (+1,322.22%)	226,431 (+56.03%)	5.79 (+
Surface Load	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Texture Load	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Store	2,048 (+0.00%)	2,048 (+0.00%)	2,048 (+0.00%)	0.05 (-0.87%)	8,192 (+0.00%)	4 (+0.00%)	10.30 (+inf%)	262,144 (+0.00%)	8,192 (+0.00%)	0.21
Local Store	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)		
Surface Store	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Reduction	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
DSMEM Reduction						-	-		0 (+0.00%)	0 (
Surface Reduction	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Atomic ALU	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Global Atomic CAS	0 (+0.00%)									
Surface Atomic ALU	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Surface Atomic CAS	0 (+0.00%)									
Loads	1,048,576 (+2,744.44%)	1,048,576 (+2,744.44%)	1,048,576 (+2,744.44%)	26.79 (+2,719.76%)	2,097,148 (+1,322.22%)	2.00 (-50.00%)	89.30 (+5,951.17%)	67,108,736 (+1,322.22%)	226,431 (+56.03%)	5.79 (+
Stores	2,048 (+0.00%)	2,048 (+0.00%)	2,048 (+0.00%)	0.05 (-0.87%)	8,192 (+0.00%)	4 (+0.00%)	10.30 (+inf%)	262,144 (+0.00%)	8,192 (+0.00%)	0.21
Atomics & Reductions	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (
Total	1,050,624 (+2,600.00%)	1,050,624 (+2,600.00%)	1,050,624 (+2,600.00%)	26.84	2,105,340 (+1,252.63%)	2.00 (-49.90%)	88.99	67,370,880 (+1,252.63%)	234,623 (+53.04%)	5.99 (+

L2 Cache

	Requests	Sectors	Sectors/Req	% Peak	Hit Rate	Bytes	Throughput	Sector Misses to Device	Sector Misses to System
L1/TEX Load	161,046 (+342.23%)	226,999 (+55.97%)	1.41 (-64.73%)	4.92 (+56.64%)	92.87 (+4.54%)	7,263,968 (+55.97%)	43,024,829,416.22 (+56.68%)	16,384 (+0.00%)	0 (+0.00%)
L1/TEX Store	4,094 (+99.90%)	8,192 (+0.00%)	2.00 (-49.98%)	0.18 (+0.43%)	100 (+0.00%)	262,144 (+0.00%)	1,552,691,432.90 (+0.45%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Atomic ALU	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Atomic CAS	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Reduction	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)	0 (+0.00%)
L1/TEX Total	167,598 (+332.15%)	234,331 (+51.31%)	1.40 (-64.99%)	5.07 (+51.97%)	93.17 (+4.18%)	7,498,592 (+51.31%)	44,414,518,574.68 (+52.00%)	16,384 (+0.00%)	0 (+0.00%)
ECC Total	-	8 (+0.00%)	-	0.00 (+0.43%)	-	256 (+0.00%)	1,516,300.23 (+0.45%)	8 (+0.00%)	-
GPU Total	167,738 (+271.36%)	236,916 (+37.71%)	1.41 (-62.93%)	5.13 (+38.30%)	93.14	7,581,312 (+37.71%)	44,904,473,085.67 (+38.34%)	16,515 (-0.05%)	32 (+6.67%)

Comparison Overall

There are differences between the 8x8, 16x16 and 32x32 tiled matmuls. But let us discuss similarities first.

There is a significant drop in L1 Hitrate. That probably has to do with the usage of shared memory. Not everything is done via L1 anymore, so more is used in shared, reducing need for L1. As the size of the tile was increasing, so too was the L2 cache hitrate for the tiled variants. It is slightly confusing as to why. The only reason I can think of is that the need for L2, or the singular shared memory across the SM, is reduced when everything is put in shared mem.

There is also the fact that all wavefronts in the L1 level are reduced in terms of num of operations, from 8x8 matrices to 32x32 matrices. This is probably due to the lower reliance on L1 cache.

Across all three, we see fewer number of bytes transferred, especially compared to the naive kernel, from kernel to global, and beyond. In 8x8, it is most and in 32x32 it is least. This trend is explained by the increased reliance on shared memory as well.

In terms of differences, we can see the 8x8 matrix has more sectors in L2 cache than the others do. This is confusing, since logically this should also be lesser, due to the reduced reliance on L2 cache. Maybe a certain threshold of size is not yet met?