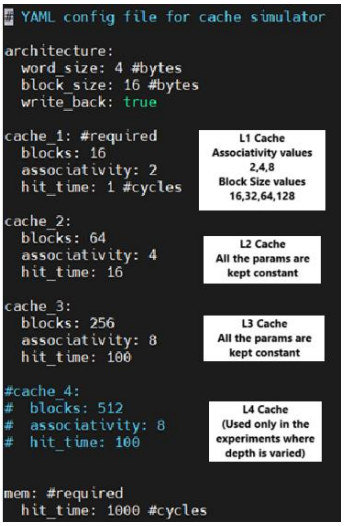
Assignment 1 - Cache System Performance Analysis

Assumptions Made and Architecture Block Diagram



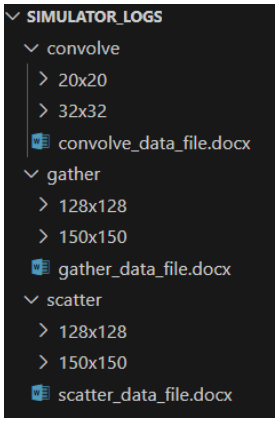
- For the Scatter and Gather operations, 128 x 128 and 150 x 150 matrices are used.
- For the Convolution operation, 32 x 32 and 20 x 20 matrices are used because of hardware issues.
- I have varied the depth of the cache from 1 to 4 and the results (# clock cycles, # misses, hit rate, miss rate and AMAT) are measured for L1 cache only.
- Parameters such as block size and associativity are varied for the L1 cache and the results are plotted accordingly in the graphs that follow. The results are reported for the L1 cache only.
- For the L1 cache, the block size is varied from 16 blocks to 128 blocks i.e. 16, 32,64 and 128 blocks respectively, keeping everything else constant.
- For the L1 cache, the associativity is varied from 2 to 8 i.e. 2,4 and 8 keeping everything else constant.

Bonus Question (Discussed in the 17th Sept Class)–

- For a memory endurance of 10^6 , the number of times we can repeat an operation is reported in the table below -

Operation	No. of Writes	No. of time operations are repeated (10^6/ # writes)
Scatter(128)	62268	16
Scatter(150)	63901	15.96 ~ 16
Gather(128)	120956	8.26 ~ 8
Gather(150)	130415	7.66 ~ 8
Convolution(32)	230077	4.34 ~ 4
Convolution(20)	94969	10.52 ~ 11

Format of Log Files



Directory structure of the Simulator logs folder -

For example –

Inside every operation's folder –

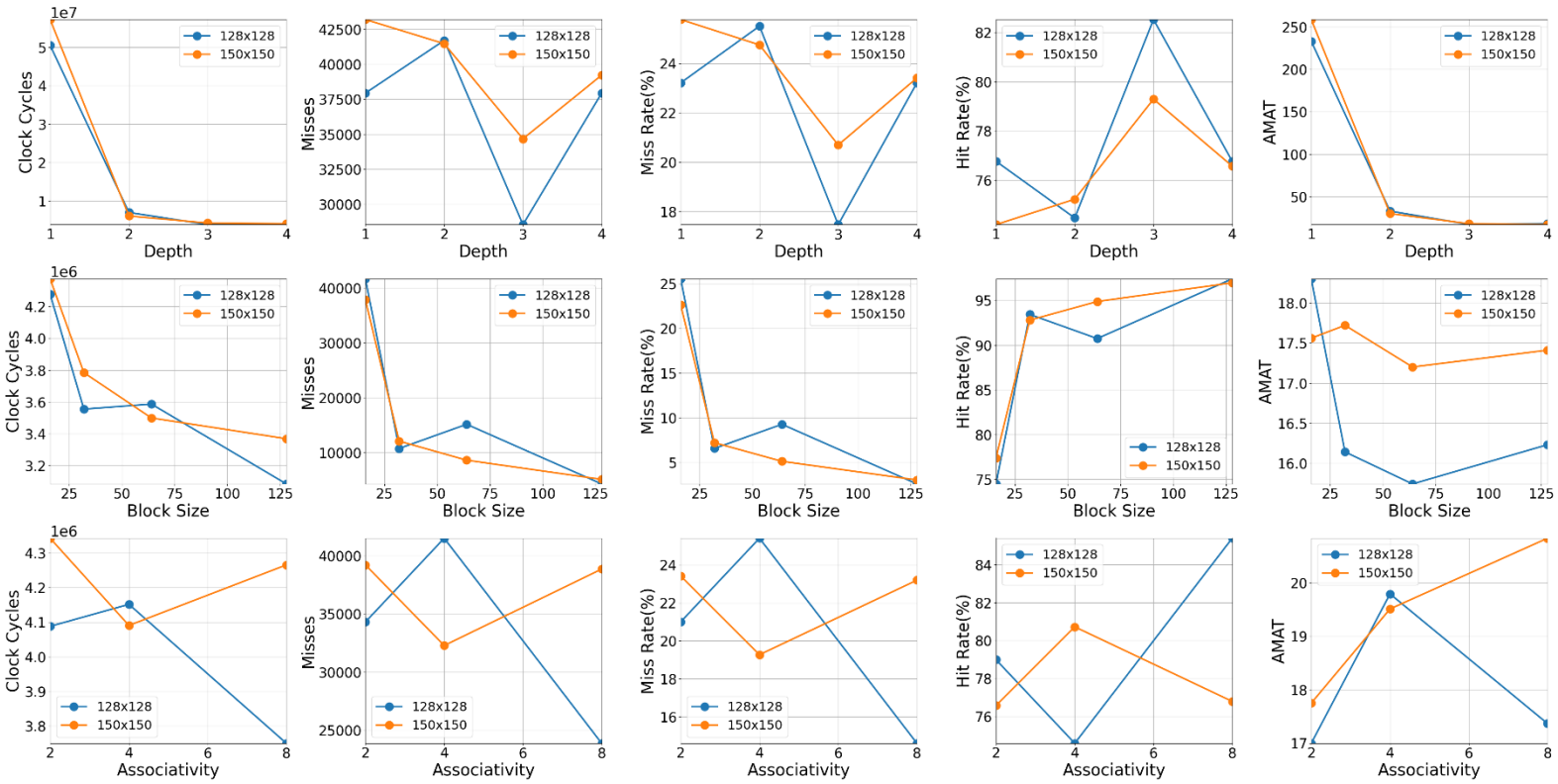
Each log file is labelled as -

operation_**matrix_size**_**parameter_varied**.txt

- **operation** – scat(Scatter),gath(Gather),conv(Convolution)
- **matrix_size** – 128, 150, 32, 20
- **parameter_varied** – assoc_#, block_#, depth_#_level

Submitted By – Kapil Ravi Rathod
PSU ID - 973212163
Email – kvr5715@psu.edu

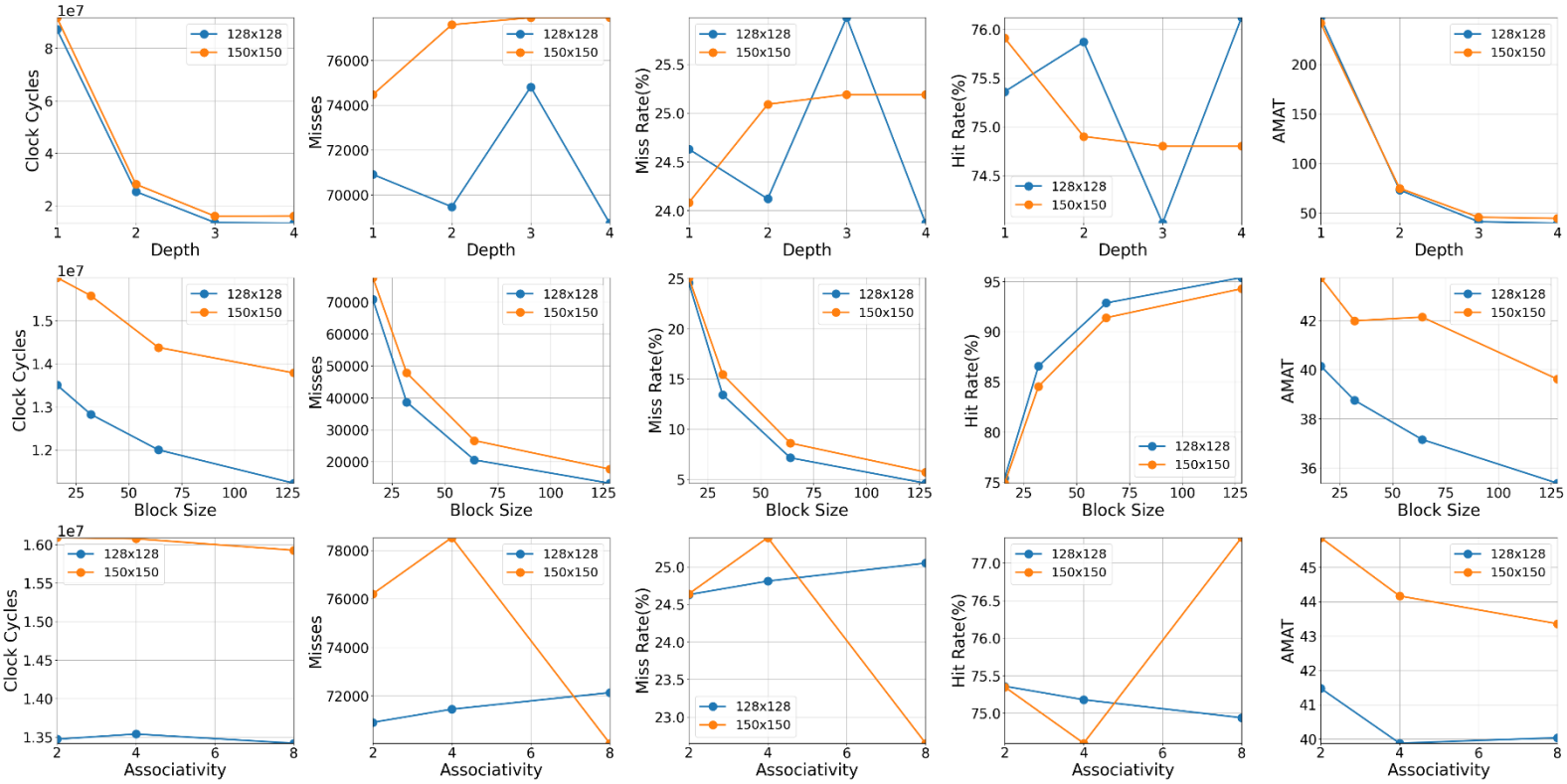
Scatter Operation (for 128 x 128 and 150 x 150 matrices



Parameter	Metric	128x128 Trend	150x150 Trend	Trend Summary	Comments
Depth	Clock Cycles	Decreases as depth increases	Decreases as depth increases	Both matrix sizes show improved performance as depth increases.	Higher depth allows for more data to be stored, reducing misses and accesses to slower memory.
Depth	Misses	Irregular trend	Decreases, then increases	Both sizes show lowest misses at depth 3.	3 is the optimal depth for the entire cache for minimal misses.
Depth	Hit Rate	Irregular trend	Increases, then decreases	Maximum hit rate occurs at depth 3 for both sizes.	Hit rates improve until depth 3, where the cache is optimally sized for fast lookups.
Depth	AMAT	Decreases as depth increases	Decreases as depth increases	Both sizes show lowest AMAT at depth 4.	Memory access time is minimized at higher depths since more data is stored in the cache.
Block Size	Clock Cycles	Overall decreases with larger blocks	Overall decreases with larger blocks	Performance improves as block size increases for both sizes.	Larger blocks reduce the number of accesses by fetching more data at once.
Block Size	Misses	Overall decrease	Overall decrease	Larger block sizes lead to fewer misses for both sizes.	Larger blocks capture more data in fewer accesses, reducing the chance of a miss.
Block Size	Hit Rate	Sharp increase, then stabilizes	Sharp increase, then stabilizes	Hit rate improves with larger block sizes.	Hit rates increase as more data fits in the cache per access, up to an optimal size.
Block Size	AMAT	Decreases sharply, then stabilizes	Irregular trend	AMAT drops with larger block sizes.	Fewer accesses due to larger blocks mean lower average memory access time.
Associativity	Clock Cycles	Fluctuates	Fluctuates	Opposite trends for both sizes.	Increased associativity reduces conflict misses but increases lookup overhead, explaining divergent trends.
Associativity	Misses	Fluctuates	Fluctuates	Opposite trends for both sizes.	For 150x150, associativity effectively reduces conflicts, while for 128x128, cache size might dominate performance.
Associativity	Hit Rate	Fluctuates	Fluctuates	Opposite trends for both sizes.	128x128 benefits the most from higher associativity since it has the highest hit rate at 8-way associativity.
Associativity	AMAT	Fluctuates	Increases	AMAT fluctuates for 128x128 and increases steadily for 150x150.	Associativity reduces misses but increases lookup time, leading to mixed effects depending on cache size and matrix size.

Submitted By – Kapil Ravi Rathod
PSU ID - 973212163
Email – kvr5715@psu.edu

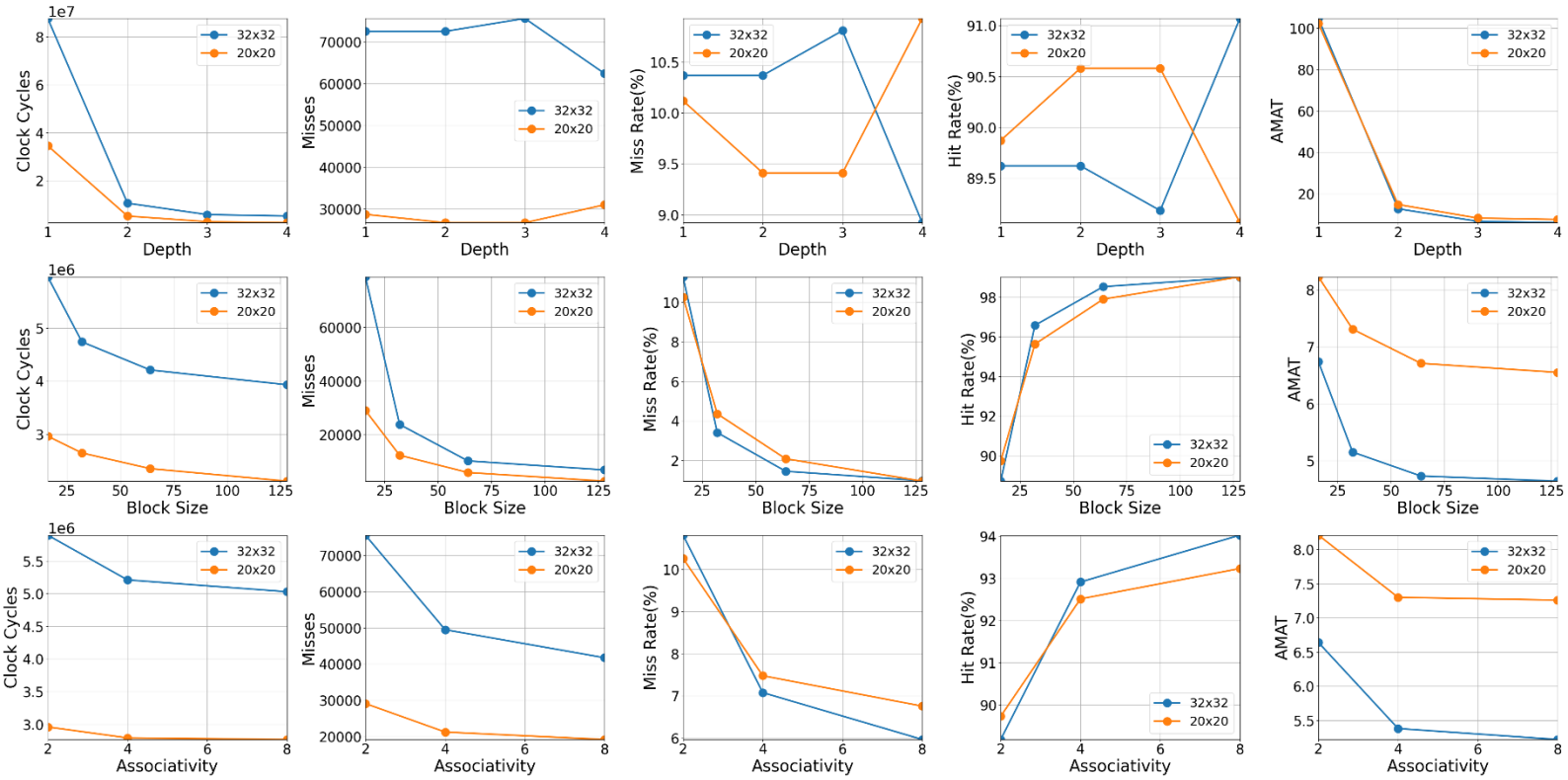
Gather Operation (for 128 x 128 and 150 x 150 matrices)



Parameter	Metric	128x128 Trend	150x150 Trend	Trend Summary	Comments
Depth	Clock Cycles	Decreases as depth increases	Decreases as depth increases	Both matrix sizes show improved performance as depth increases.	Increasing depth allows for more data to be fetched efficiently, reducing clock cycles.
Depth	Misses	Fluctuates	Increases, then remains stable	128x128 shows rising misses after depth 2, while 150x150 remains stable.	The trend observed here is not explainable because of some underlying phenomena not explored in the data.
Depth	Hit Rate	Increases until depth 2, then decreases	Stable with a slight drop	128x128 hit rate peaks at depth 2, while 150x150 remains stable.	Depth 2 provides an optimal hit rate for 128x128, but further complexity reduces efficiency.
Depth	AMAT	Decreases	Decreases	Both sizes show lowest AMAT at depth 4.	More data is stored and accessed efficiently, lowering average memory access time (AMAT).
Block Size	Clock Cycles	Decreases consistently	Decreases consistently	Both sizes show improved performance with larger block sizes.	Larger blocks reduce the number of accesses required, reducing clock cycles.
Block Size	Misses	Decreases significantly	Decreases significantly	Misses reduce dramatically for both sizes with larger block sizes.	Larger block sizes fetch more data in a single access, reducing cache misses.
Block Size	Hit Rate	Increases sharply	Increases sharply	Hit rate increases significantly with block size for both sizes.	Larger blocks allow more data to be stored, increasing hit rates.
Block Size	AMAT	Decreases significantly	Decreases but stabilizes after block size 75	AMAT decreases for both sizes but stabilizes for 150x150 after block size 64.	Fewer accesses due to larger blocks mean lower average memory access time.
Associativity	Clock Cycles	Remains stable	Remains stable	Both sizes show minimal change in clock cycles with increasing associativity.	Associativity does not significantly impact the clock cycles.
Associativity	Misses	Remains relatively stable	Fluctuates	Misses reduce for 150x150 with higher associativity; 128x128 remains relatively stable.	Higher associativity reduces conflict misses for 150x150, with minimal effect on 128x128.
Associativity	Hit Rate	Remains relatively stable	Fluctuates	Irregular trends for both matrices, one remains stable other fluctuates.	150x150 has the maximum hit rate at 8-way associativity.
Associativity	AMAT	Decreases slightly and then remains stable	Decreases	AMAT decreases for 150x150; 128x128 decreases slightly then remains stable.	Higher associativity reduces cache misses, improving AMAT.

Submitted By – Kapil Ravi Rathod
PSU ID - 973212163
Email – kvr5715@psu.edu

Convolution Operation (for 32x 32 and 20 x 20 matrices)



Parameter	Metric	32x32 Trend	20x20 Trend	Trend Summary	Comments
Depth	Clock Cycles	Decreases rapidly from depth 1 to 2, then stabilizes	Decreases slightly with depth increases	Lower depth gives much higher clock cycles initially	Larger depths result in less frequent memory accesses, reducing clock cycles.
Depth	Misses	Fluctuates	Fluctuates	Misses stay stable till depth 3 then follow opposite trends after depth 3	Cannot generalize since the fluctuations are quite random.
Depth	Hit Rate	Fluctuates	Fluctuates	Trend is opposite in nature	Trend observed is irregular.
Depth	AMAT	Decreases significantly	Decreases as depth increases	Both show clear reductions in AMAT as depth increases	Deeper caches reduce miss penalties and overall memory access time, improving AMAT.
Block Size	Clock Cycles	Decreases with larger block size	Decreases steadily with block size	Larger block size results in fewer cycles	Larger blocks lead to fewer fetches from memory, saving cycles during cache accesses.
Block Size	Misses	Sharp drop with larger block size	Sharp drop as block size increases	Fewer misses with larger block size	Bigger blocks reduce the need for repeated memory accesses by accommodating more data.
Block Size	Hit Rate	Increases consistently with block size	Hit rate rises as block size increases	Hit rate improves with larger block size	Larger block sizes fetch more data per memory request, increasing chances of hits.
Block Size	AMAT	Decreases sharply with larger block size	Similar sharp decrease with larger block size	Larger blocks improve AMAT	Fewer accesses due to larger blocks mean lower average memory access time.
Associativity	Clock Cycles	Decreases with higher associativity	Decreases with higher associativity	Higher associativity results in fewer clock cycles	Higher associativity reduces conflicts in cache lines, improving cycle efficiency.
Associativity	Misses	Decreases with increasing associativity	Similar trend of decreasing misses	Misses decrease as associativity increases	Higher associativity reduces the likelihood of eviction and cache misses.
Associativity	Hit Rate	Rises consistently as associativity increases	Steady increase with associativity	Higher associativity improves hit rate	Higher associativity means less conflict for cache lines, increasing hit chances.
Associativity	AMAT	Decreases with associativity	Decreases with associativity	AMAT improves as associativity increases	Associativity minimizes cache conflicts, reducing average memory access time.

Submitted By – Kapil Ravi Rathod
PSU ID - 973212163
Email – kvr5715@psu.edu