

ENSC 254 Lab #5: Hardware-Aware Software Optimization

Jovan Radovanovic - 301467968

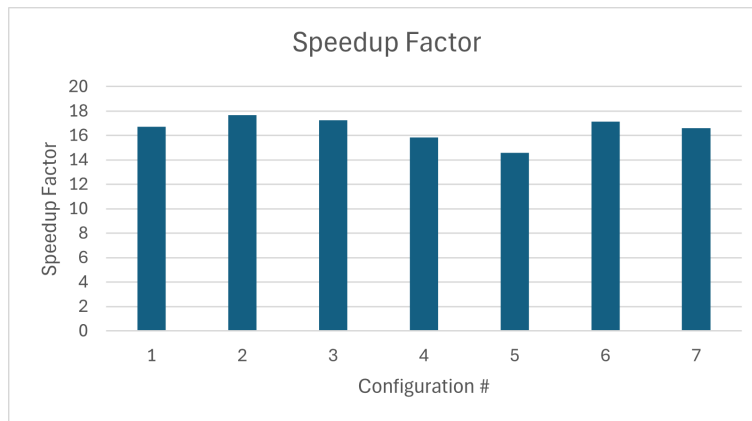
Kannan Jayakumar - 301462843

July 16th, 2024

Baseline time: average across a few runs: 566.8099443 seconds (9.446832405 minutes)

Data Tiling Optimization:

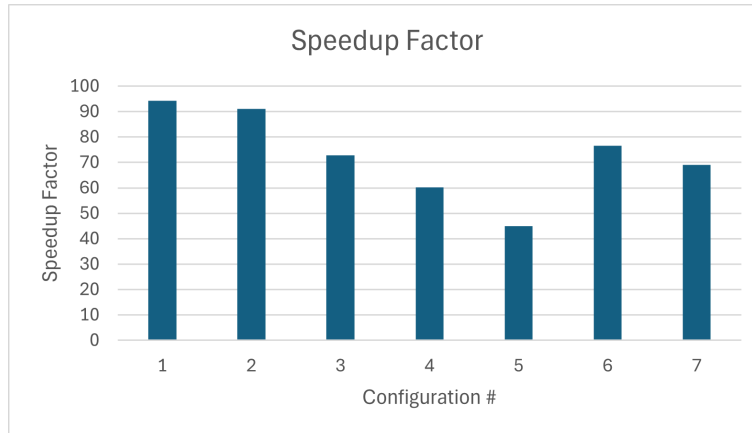
Configuration	TILE_SIZE_I	TILE_SIZE_J	TILE_SIZE_K	Time	Speedup
1	256	256	256	33.92325986	16.70859306
2	128	128	128	32.08853334	17.66394052
3	64	64	64	32.84377669	17.25775783
4	32	32	32	35.76844314	15.84664846
5	16	16	16	38.85567895	14.5875702
6	128	64	32	33.05015027	17.14999598
7	32	64	128	34.1585026	16.59352434



Looking at the table and the graph, the configuration that gives us the highest speedup factor is configuration 2, where the tile size is 128 for all three variables, i, j, and k. This gives us an execution time of ~32.0885 seconds and a speedup factor of ~17.6639.

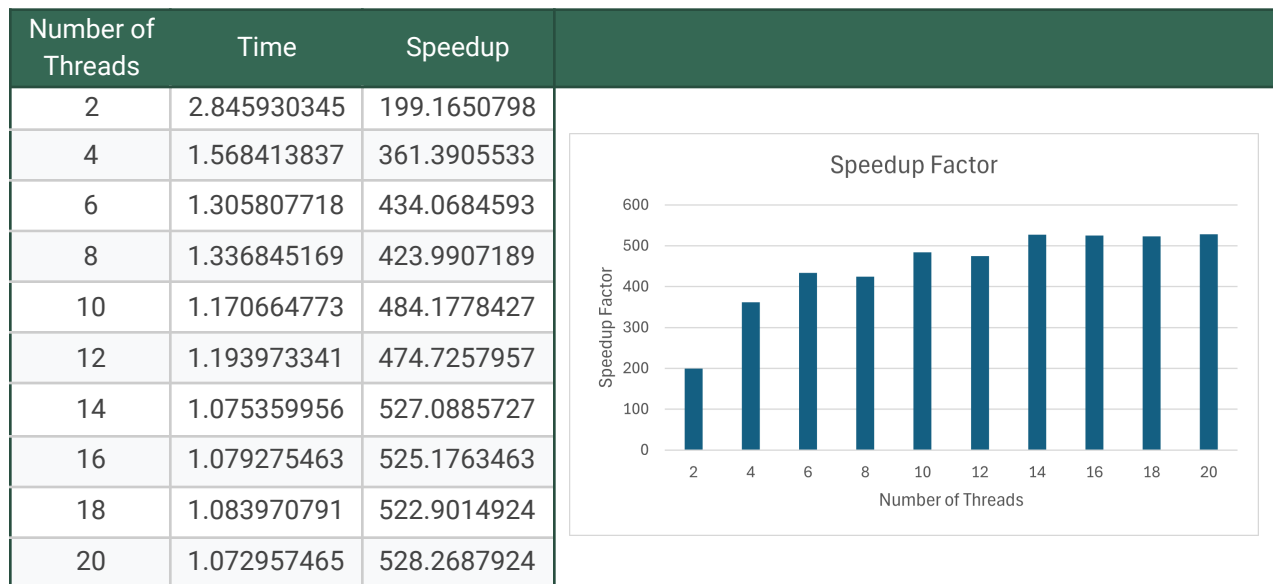
Vectorization (SIMD) Optimization:

Configuration	TILE_SIZE_I	TILE_SIZE_J	TILE_SIZE_K	Time	Speedup
1	256	256	256	6.01334626	94.25865729
2	128	128	128	6.224976903	91.05414416
3	64	64	64	7.788610145	72.77420923
4	32	32	32	9.412000654	60.22204685
5	16	16	16	12.62808569	44.8848668
6	128	64	32	7.400302466	76.59280778
7	32	64	128	8.215342952	68.99406971



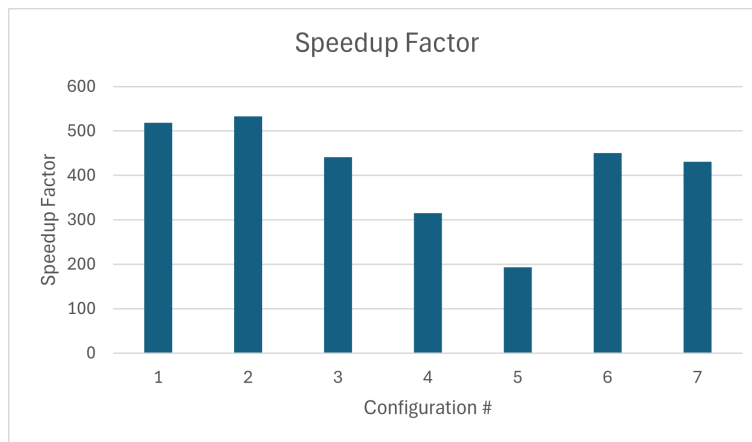
Looking at the table and the graph, the configuration that gives us the highest speedup factor is configuration 1/2 (Was hard to differentiate which one was actually faster, they were very similar), where the tile size is 128 or 256 for all three variables, i, j, and k. This gives us an execution time of ~6.0133 seconds and a speedup factor of ~94.2587.

Parallelization Optimization:



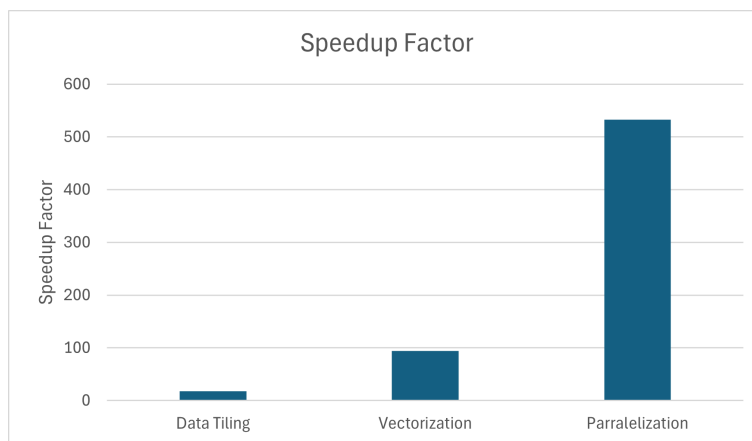
Looking at the table and the graph, the highest speedup factor is achieved with 20 threads. This gives us an execution time of ~1.0729 seconds and a speedup factor of ~528.2688.

Configuration	TILE_SIZE_I	TILE_SIZE_J	TILE_SIZE_K	Time	Speedup
1	256	256	256	1.093221467	518.4767784
2	128	128	128	1.063773135	532.8297225
3	64	64	64	1.283998779	441.4411864
4	32	32	32	1.796815868	315.4524369
5	16	16	16	2.938229961	192.9086395
6	128	64	32	1.257808658	450.632885
7	32	64	128	1.316583289	430.5158276



After choosing 20 threads, by looking at the table and graph, the tile size configuration that gives us the highest speedup factor is configuration 2, where the tile size is 128 for all three variables, i, j, and k. This gives us an execution time of ~1.0638 seconds and a speedup factor of ~532.8297.

Step-by-Step Speedup Over Baseline:



Performance Competition:

Using the code from `gemm_simd_par()`, we added loop unrolling and tuned the values of the tile size for i, j, and k to 64, 512, and 64 respectively. This got us execution times in the low 0.8xx seconds, with it sometimes dropping below 0.8 seconds. With an execution time of 0.8 seconds, that means our speedup factor is 708.512430375 compared to baseline.