

PHYS-8061 HW4 Discussions

B11902156 陳浩文

I start from block size = 8 to block size = 512, and will keep testing each grid size until no improvements are made in terms of speed-up.

Speed-up with rows as grid sizes, columns as block sizes (2 GPUs):
(Partially shown, as I did not test every possible configuration)

	8	16	32	64	128	256	512
16	0.273308	0.465009	0.882186	0.593727	0.645496	1.084016	
32	0.661126	0.551400	0.978119	1.062450	0.599821	0.665266	1.202287
64	0.745773	1.054121	0.663164	0.682051	1.001727	1.214445	0.660243
128	0.871878	1.097569		0.796749	0.667091	1.157730	0.679319
256	0.610229	0.530208		0.766335	0.659494	0.803486	0.777339
512							0.665068
1024							

From my observations, the larger the grid size, the better the speed-up is, until around 64~128, it starts to show a negative impact on speed-up.

Also, the larger the block size, the better the speed-up is.

But the result seems a bit inconsistent, and I mark them in blue. It seems like even with 2 openMP threads, some of the results are based on two virtual threads only running on one single physical thread. And the other results show to be almost two times better in terms of speed-up, this means these results are actually benefiting from the power of multiple threads in practice, and are run by two different physical threads.

The optimal block size and grid size is (256, 64), with a speed-up of 1.214445