Project 6: CUDA Monte Carlo

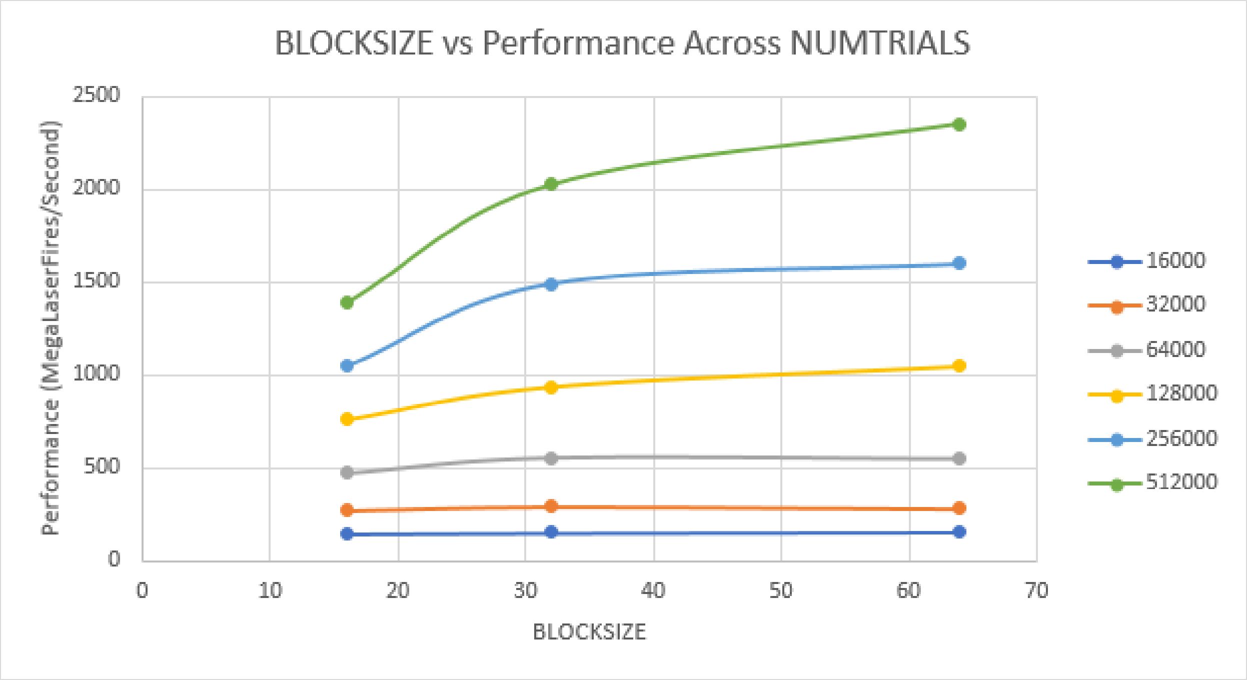
Joel Huffman huffmajo@oregonstate.edu

Tables and Graphs

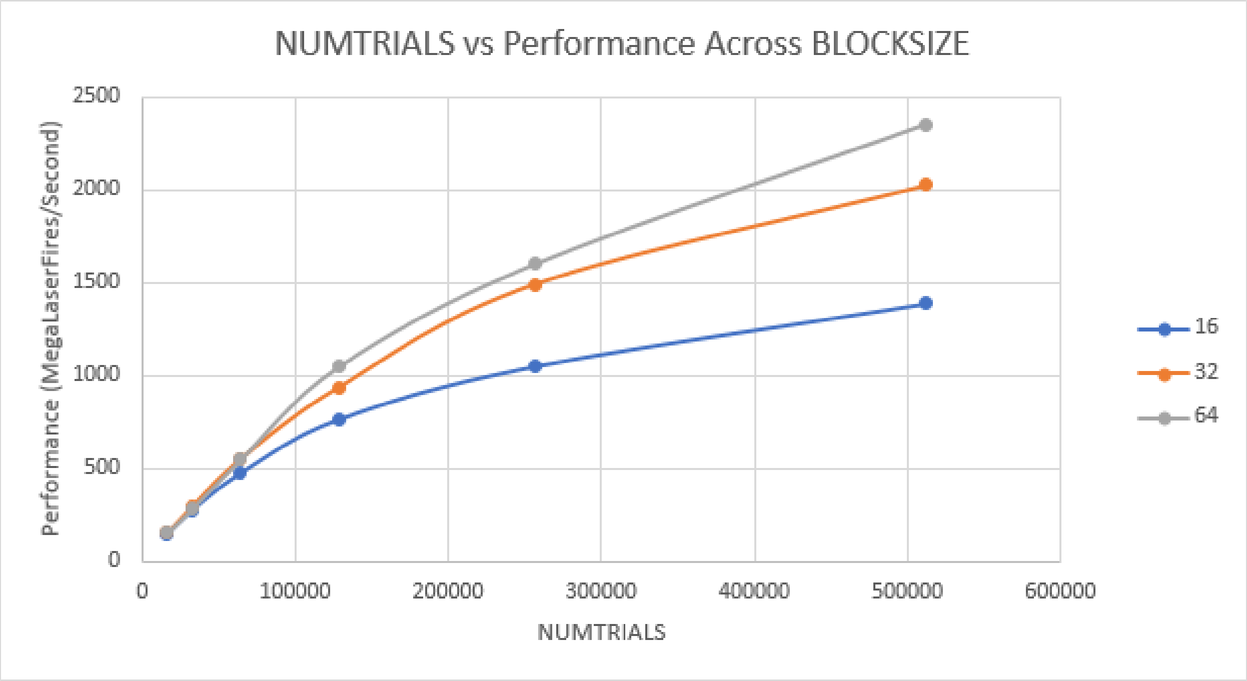
Table 1: Performance Across NUMTRIALS and BLOCKSIZE

		NUMTRIALS					
		16000	**32000**	**64000**	**128000**	**256000**	**512000**
BLOCK SIZE	**16**	150.4664	273.9726	474.9466	764.6721	1049.869	1389.613
	32	154.0832	295.1594	554.6312	940.0705	1494.21	2026.599
	64	157.6293	284.0102	550.6608	1049.593	1601.281	2354.326

Graph 1:



Graph 2:



Explanation and Analysis

1. What machine you ran this on

All benchmarking was performed on OSU rabbit server.

2. Show the tables and graphs

See Table 1, Graph 1 and Graph 2 above for all results.

3. What patterns are you seeing in the performance curves?

Performance increases with BLOCKSIZE. Performance also increases with NUMTRIALS.

4. Why do you think the patterns look this way?

The increase in performance as NUMTRIALS increases is due to these larger datasets giving us more calculations for the GPU to crunch through. Rabbit's Titan Black really gets to flex it's processing power muscles when it has a large number of calculations to run. The increase in performance as BLOCKSIZE increases is due to the larger number of available threads ready to run calculations as they are needed. When one warp finishes it's instruction, the more warps I have on hand, the more likely I am to have another warp that is ready to go as soon as my current instruction set is complete. This reduces downtime and keeps efficiency high.

5. **Why is a BLOCKSIZE of 16 so much worse than the other two?**

The BLOCKSIZE of 16 has worse performance than 32 or 64 because their warps (or half-warps in this case) are finishing their instructions before we have another warp ready to execute it's instructions. This causes downtime where we are waiting for operands to be prepared.

6. **What does that mean for the proper use of GPU parallel computing?**

For better performance we should use a larger BLOCKSIZE (that is a multiple of 32, e.g., 192, 256, etc.) to ensure we always have warps available and ready to process data when they're needed.