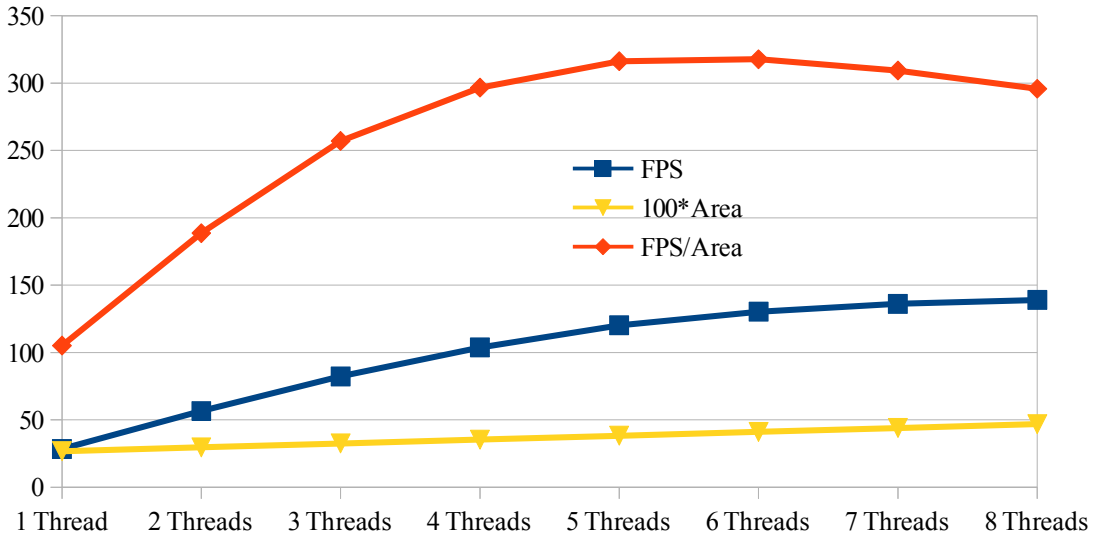
Ian Mallett
CS 6960.001

Assignment 2 Analysis

Analysis:

(a) Performance and Performance/Area:

•



- The optimal value of performance to area occurs at n=6 threads.
- A large portion of the chip area is functional units; these do not increase when the number of threads is increased. Thus, adding another thread to a single-threaded processor does not double area. However, performance per area can increase since those functional units may be underutilized; when the new thread is added, their usage increases. This corresponds to a more efficient use of area.

(b) Bottlenecks

- LOAD causes the most data dependence stalls, but mul_s is the instruction causing the most resource conflicts (38.504%), followed by sub_s (23.651%).
- Attempting to remove bottleneck by modifying functional units, even though the icache bottleneck is probably bigger. Doubled count of FPADD, FPMIN, INTADD, and FPMUL (the four most stalling instructions). Performance is 141.3744 FPS, area is 1.178969mm^2; ratio is ~119.9.
- Assuming removing bottlenecks by any means necessary. There is a huge icache conflict and some of the other instructions have peaked up, too. Final configuration: 32 threads, icaches:2, icache banks:8, FPADD 8, FPMIN 4, FPCMP 2, INTADD 4, FPMUL 8, INTMUL 1, FPINV 1, CONV 1, BLT 2, BITWISE 2. Performance 936.6944 FPS, area 1.469441mm^2, FPS/area ~637.45

(c) Traversal Algorithm

FPS with and without early exit optimization for various scenes:

Conference		Hairball		Dragon	
With Opt	Without Opt	With Opt	Without Opt	With Opt	Without Opt
61.07	31.26	20.58	7.87	22.5	10.16

- All of the scenes benefit significantly from the optimization, although the hairball scene benefits the most. This is because without early exit, the entire tree is traversed (so it is as bad—worse, actually—than brute-force checking every triangle). With early exit, we get logarithmic behavior.
- Energy consumption is by far the highest in DRAM. With usimm power about ~14W to ~18W. Comparing to previous section rendering Cornell box without usimm, ~0.6W, mostly to register files and instruction caches. Energy is a function of how long each render takes; since different scenes were done, they are not directly comparable.
- DRAM read latency:

Conference		Hairball		Dragon	
With Opt	Without Opt	With Opt	Without Opt	With Opt	Without Opt
42.45	43.62	55.33	60.37	54.66	58.79

It is difficult to say how the latency changes when using versus not using the optimization. It looks like with the optimization, latency is reduced, which means the scene is being traversed in a more row-buffer-friendly way. This is probably counteracted somewhat by skipping more data, and therefore less coherent access.