

# ECE 60827 CUDA Assignment 1 Report

Setsuna

January 30, 2025

Runtime breakdown of significant components for different vector sizes for SAXPY are reported in figure 1.

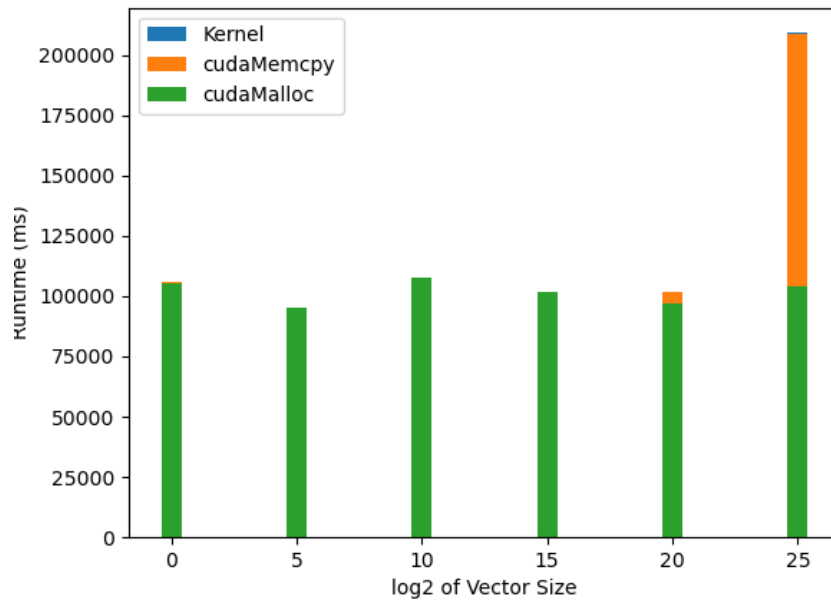


Figure 1: Runtime of SAXPY with different vector size

Runtime breakdown when varying the number of points in  $\pi$  estimator is reported in figure 2. Varying the number of threads yields runtime result figure 3.

Only significant API calls are reported in the breakdown.

As seen in all figures, the runtime of `cudaMalloc()` stays relatively constant regardless of the amount of data, and is significant especially when the amount of work is small. As expected, when the amount of data increases beyond

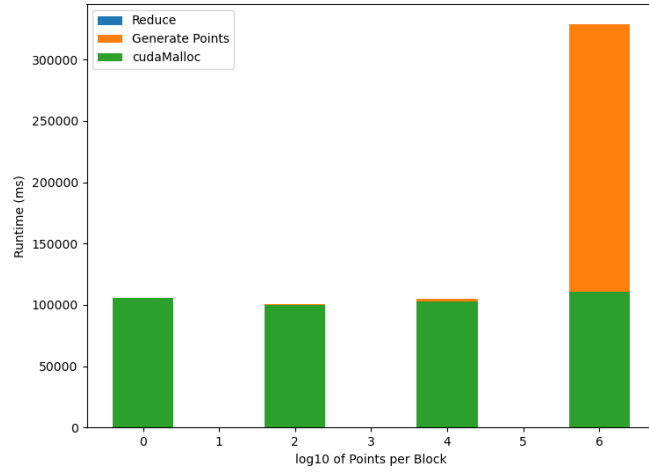


Figure 2: Runtime of Pi Estimate with different number of points per thread

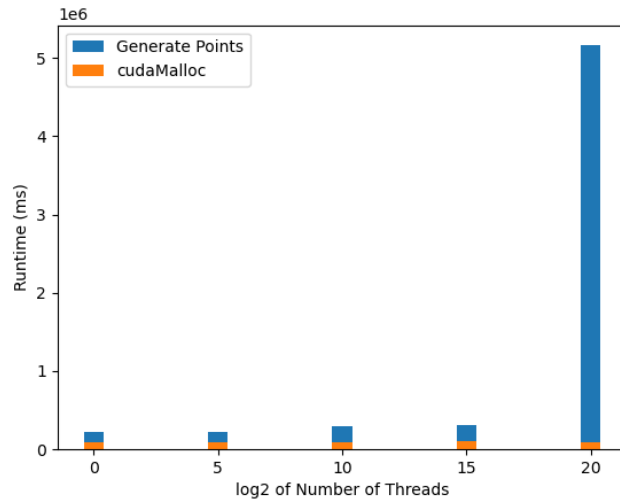


Figure 3: Runtime of Pi Estimate with different number of threads

a certain saturating point, the `cudaMemcpy()` overhead becomes much more significant. Similar saturating effects are also observed when increasing the number of threads or the amount of work for each thread.