

Dynamic Parallelism Performance Evaluation on Image Segmentation Algorithms

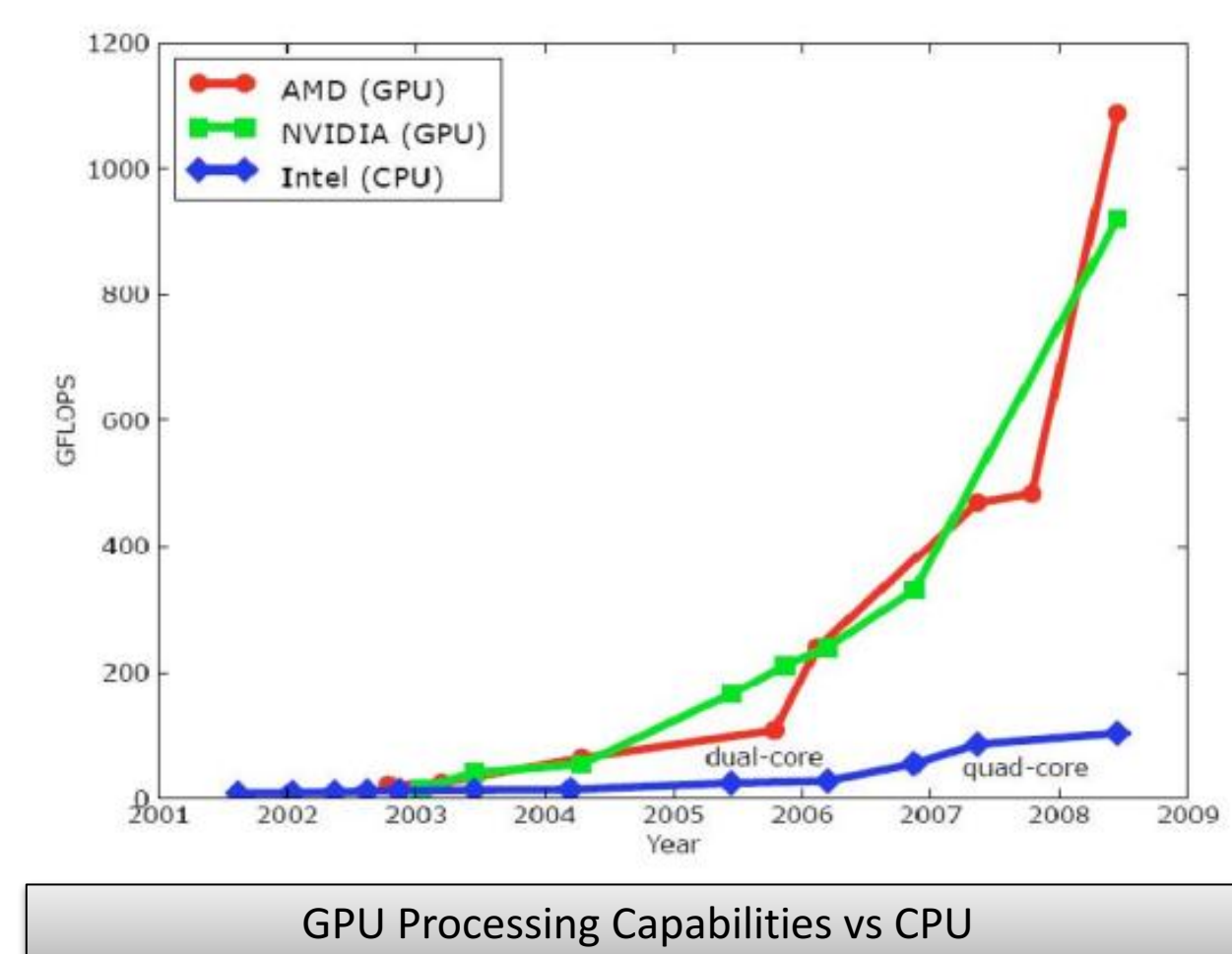
Daniel Goldstein, Julian Gutierrez, goldstein.d@husky.neu.edu

Northeastern University Computer Architecture Research Laboratory, *Northeastern University*, Boston, MA

Opportunity

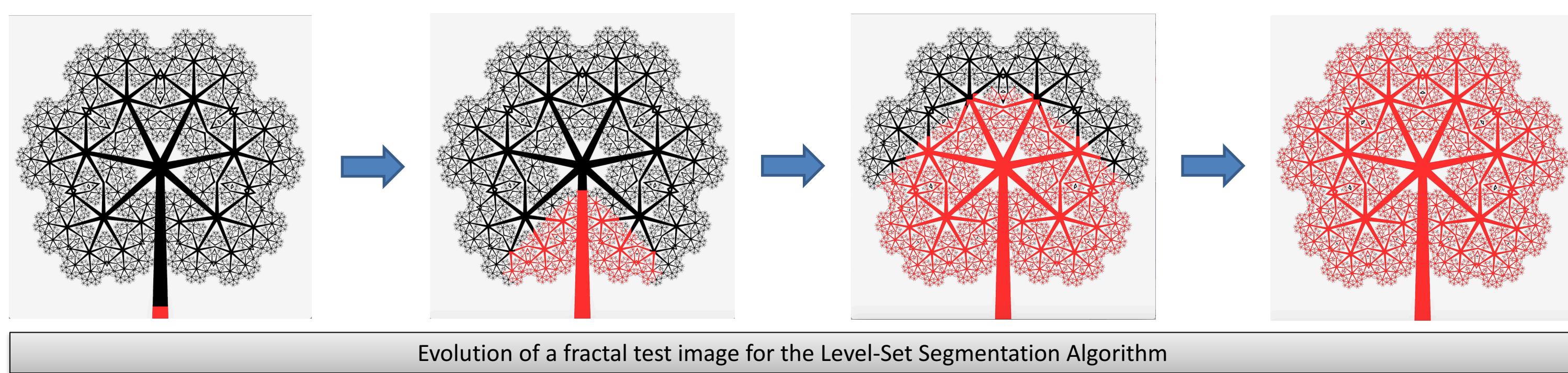
Abstract

- CPU speeds are reaching a plateau
- Parallel computing through GPUs is a method for achieving faster computing speeds
- We compare CPU and GPU parallelization to analyze impact of dynamic parallelism in image-processing algorithms



Goal

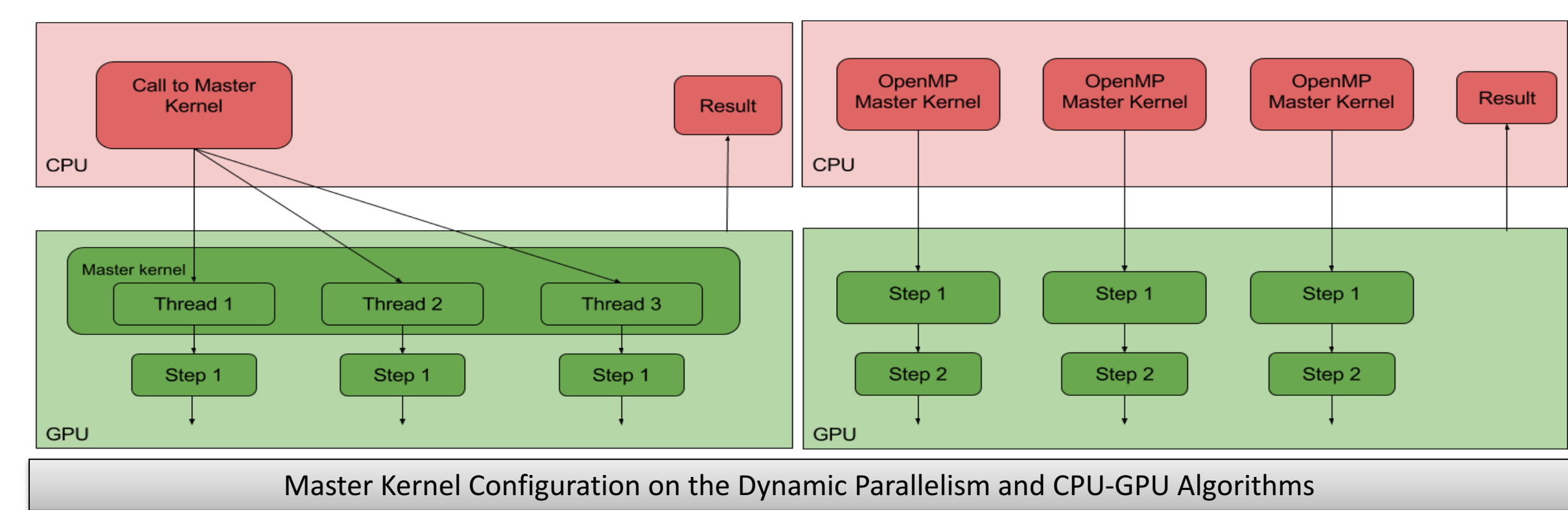
- Compare dynamic parallelism in CUDA to a split CPU-GPU parallelization to see which method is fastest for a Level-Set Segmentation algorithm



Approach

Algorithm Modifications

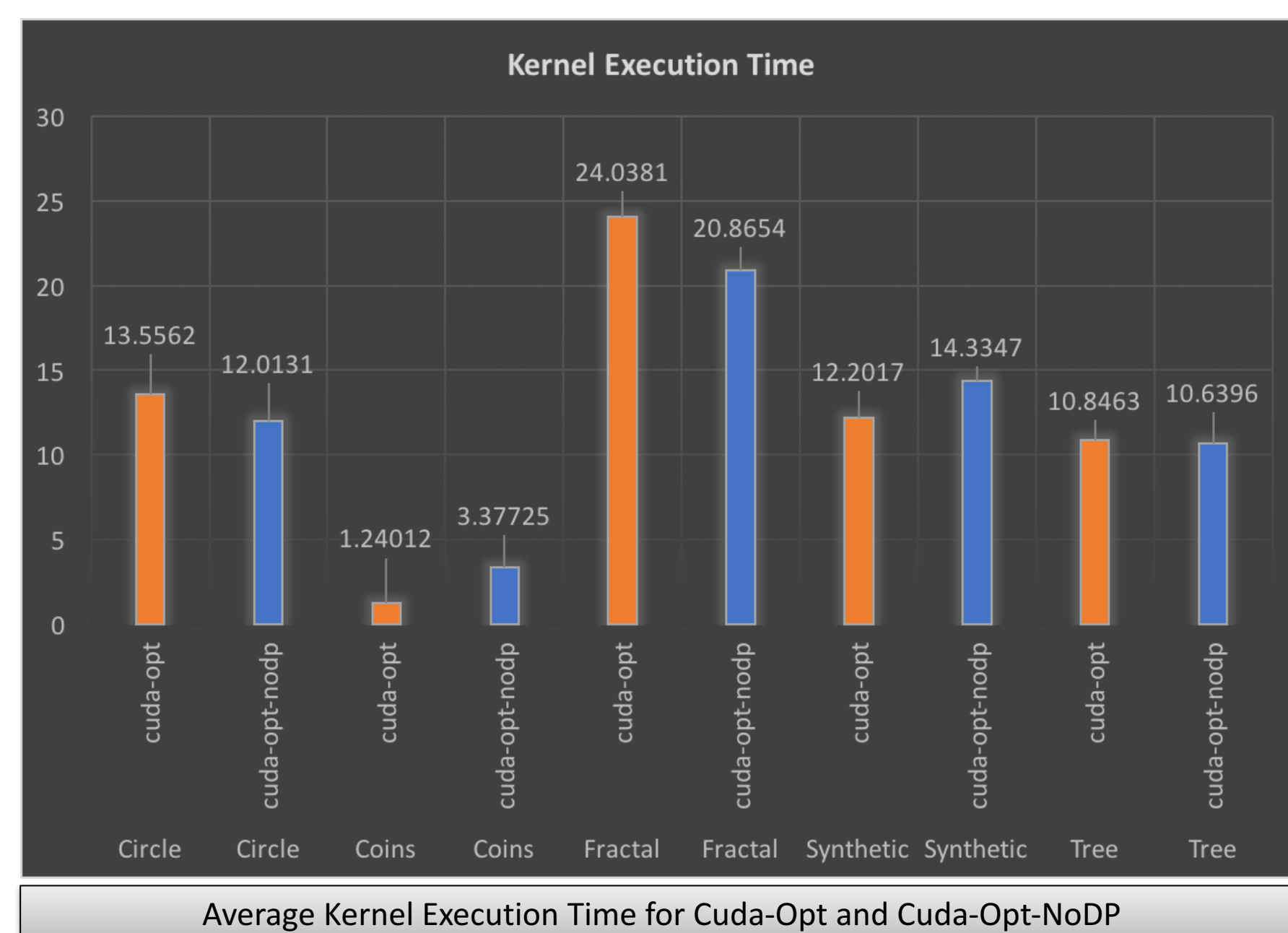
- Largest factors in performance are: size of the image, number of objects, and the complexity of the objects
- Original algorithm employ's *dynamic parallelism*
 - A master kernel on the GPU calls subsequent kernels that perform the curve evolution (shown on left)
- Reconfigure the master kernel so that it runs on the CPU, entirely removing dynamic parallelism
 - Kernel is constructed into a for-loop on the CPU side which is parallelized using the OpenMP runtime (shown on right)
 - No longer any CUDA kernels spawning other kernels
- Adapt master kernel grid size to compare using one block with one thread per object vs one block per object with one thread each



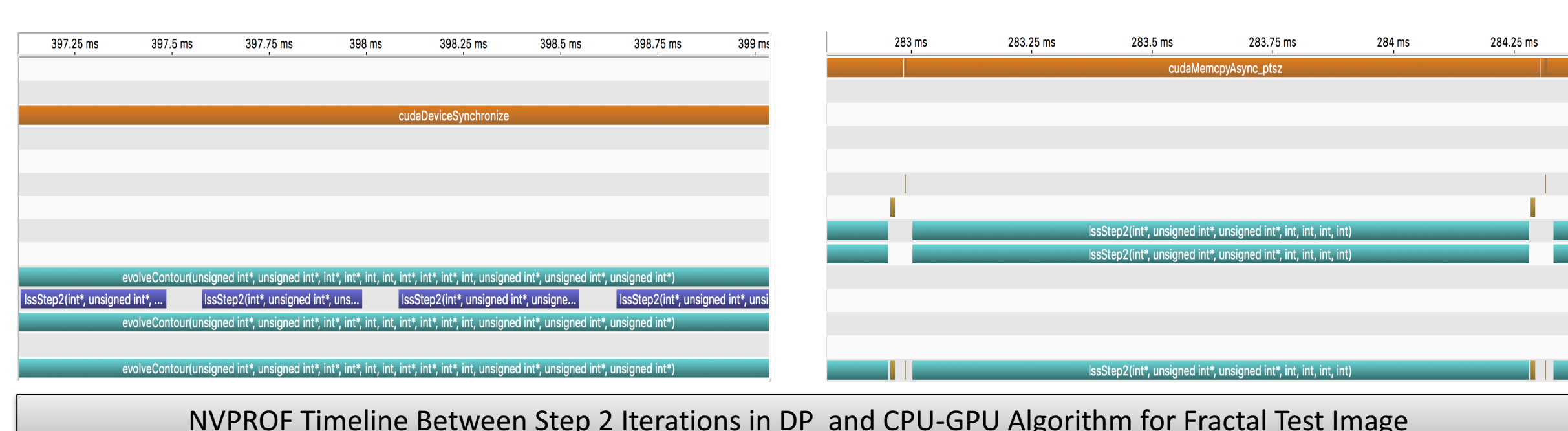
Data/Results

Kernel Execution Analysis

- Twenty trials were run for each sample image for the original and modified algorithm



- Cuda-opt performed better on images with multiple objects, e.g. coins and synthetic, while CPU-GPU implementation was faster for single-object images
- For multiple-object images, CPU-GPU implementation experienced between 17%-172% slower times
- Large improvement in fractal image is largely due to the overhead required by the large number of *atomic operations* necessary in the original algorithm that were removed in the modified algorithm
- Similar behavior observed when run on an NVIDIA Pascal GTX 1080



Impact

Conclusion

- Dynamic parallelism proved to be the most effective for images of multiple or an unknown number of objects
- CPU-GPU implementation would be best for more intricate images of one object
- Results can be extrapolated to other applications run on parallel hardware
- Further research would need to be done to conclude if these results remain consistent across different GPU architectures

Value Proposition

- This project is unique because it
 - combines CUDA parallelism and OpenMP, two different environments that serve similar purposes
 - found optimized paths to parallelization, an underexplored field of comparison
- This experiment solves the problem of finding the best implementations to parallelize algorithms

Image-Segmentation Applications

- Cancer radiation therapy
- Medical image processing
- Border security
- Self-driving cars

