

# PROFILE-GUIDED OPTIMIZATION OF CRITICAL MEDICAL IMAGING ALGORITHMS

David R. Kaeli  
Byunghyun Jang  
Perhaad Mistry  
Dana Schaa

Department of Electrical and Computer Engineering  
Northeastern University  
Boston, MA

## ABSTRACT

Given the rapid growth in computational requirements for medical image analysis, Graphics Processing Units (GPUs) have begun to be utilized to address these demands. But even though GPUs are well-suited to the underlying processing associated with medical image reconstruction, extracting the full benefits of moving to GPU platforms requires significant programming effort, and presents a fundamental barrier for more general adoption of GPU acceleration in a wider range of medical imaging applications.

In this paper we describe our experience in accelerating a number of challenging medical imaging applications, and discuss how we utilize profile-guided analysis to reap the full benefits available on GPU platforms. Our work considers different GPU architectures, as well as how to fully exploit the benefits of using multiple GPUs.

**Index Terms**— Profile-guided optimization, algebraic reconstruction, tomographic reconstruction.

## 1. INTRODUCTION

Multi-dimensional medical imaging techniques involve the processing of high resolution images commonly used in a number of clinical applications. A barrier commonly encountered in this class of application is the time required to reconstruct and render a single image captured from the imaging system. Current imaging systems utilize multiple dimensions and challenging geometries, resulting in large volumes of data to image vital organs and suspicious growths.

Partitioning these large datasets across multiple processors on a computing cluster can help to improve performance and lends the data to more efficient parallel execution. But partitioning leads to communication between the processing elements, and given the latency associated with typical cluster-based systems, can be a limiting factor in terms of image reconstruction time.

Graphics Processing Units are an interesting image reconstruction platform. These devices, which have been

primarily developed to serve the computer gaming community, are presently being used as co-processors in a variety of general-purpose applications. Given the massive amounts of parallelism provided on-chip, these hardware platforms can provide 1-2 orders of magnitude speedup over a serial execution of many image reconstruction tasks. Given this degree of acceleration, we can begin to consider how to leverage near real-time reconstruction performance.

## 2. PROFILE-GUIDED OPTIMIZATION

Profile-guided optimization has been used extensively to characterize the dynamics of program execution. Equipped with a profile or trace of the execution, a program can be optimized utilizing a range of source-level transformations and compiler optimizations. By effectively utilizing program execution profiles, we can begin to remove performance bottlenecks in reconstruction codes, which will allow us to better utilize the underlying compute resources present on a single GPU [1].

We can effectively utilize program profiles to also explore how best to leverage multiple GPUs in order to further accelerate the reconstruction process [2]. Our approach can consider both shared-memory GPUs as well as distributed-system GPUs. We discuss implementation tradeoffs and describe how profiles taken on a single GPU system can be used to arrive at the best multi-GPU configuration for the profiled reconstruction applications.

## REFERENCES

- [1] D. Rivera, D. Schaa, M. Moffie and D. Kaeli, "Exploring Novel Parallelization Technologies for 3-D Imaging Applications," *19<sup>th</sup> International Symposium on Computer Architecture and High Performance Computing*, pp. 26-33, Oct. 2007.
- [2] D Schaa and D. Kaeli, "Exploring the Multi-GPU Design Space," *Proc. of IPDPS*, Plenary Session, Rome, Italy, May 2009.