AbstractID: 14592 Title: Radiation Therapy Dose Calculation using Graphics Processing Units

Dose computation methods in radiotherapy have traditionally been developed for central processing units (CPUs). Methods used in radiation therapy treatment planning must balance clinical accuracy requirements with practical planning time limits. The first commercial 3D treatment planning systems in the early 1990s took between 5-30 minutes to compute a beam depending on the dose grid size and computation method. In the years that followed, CPU performance increased with Moore's law and more reasonable dose computation speeds of several seconds to minutes were achieved. However, in recent years CPUs have increased computing power with additional cores, not clock speed. At the same time, graphic processing units (GPUs) have evolved from fixed function devices into flexible, general purpose hardware. Today, general purpose GPU computing (GPGPU) has turned desktops into supercomputers and languages such as OpenCL, CUDA (Compute Unified Device Architecture) Fortran, CUDA C and DirectCompute have allowed for fast GPU program development.

Currently, convolution/superposition algorithms are the primary computation methods used in treatment planning systems due to their combination of reasonable performance and accuracy. Monte Carlo methods are becoming more prevalent, but they still have performance problems. Porting these algorithms to the GPU promises to provide a several orders of magnitude performance increase; resulting in near real-time convolution/superposition dose computation.

The implementation of these algorithms on the many-core architecture of the GPU poses several challenges and opportunities. The first concern is dividing the algorithm up into many separate parallel processes that can be run independently of one another. The second concern is memory access patterns; for example the manually caching via shared memory, the avoidance of write on write conflicts and the maximization of memory bandwidth efficiency. The third concern is how to best leverage the specific processing capabilities of the GPU, such as hardware exponentials.

Several approaches to implementing dose computation algorithms on the GPU will be discussed. The performance increase in these newer algorithms provides new opportunities in the way we approach treatment planning – whether it be real time treatment planning or enabling dosimetric guidance to treatment delivery.

Learning Objectives;

1) To understand the architecture differences between CPU and GPU
2) To understand approaches, complications and advantages to implementing pencil beam, convolution/superposition algorithm and Monte Carlo algorithms on the GPU
3) To understand the implications of the performance gain on treatment planning processes