This discussion will introduce “Compute Unified Device Architecture” also known as CUDA. CUDA is a compute engine that was developed by NVidia to perform massively parallel computations on Graphics Processing Units (GPS). CUDA is a framework to access the parallel computing power of Nvidia GPUs using a preprocessor for a standard C compiler. CUDA typically allows for a performance gain of about 10X in typical applications. We will cover some of computational architectures and how CUDA relates to them. This will be followed by an introduction to CUDA, its API, and some sample code. We will also go over the implementation of Fast Fourier Transform (FFT) using CUDA. We will finish with talk about benefits and drawbacks of using CUDA.

BIO: Peter Zalutski is pursuing his MS in Computer Science at the University of Missouri – St. Louis. His general field of interest is high performance computer systems. Professionally, he is involved in development of low latency, high throughput infrastructure for financial industry.