With the advent of general purpose GPU devices, it is now straightforward to perform massively parallel computations in a diversity of applications. This course will focus on the algorithmic approach to writing programs in the CUDA extensions to the C/C++ languages for GPUs. Access is available to a network of SGI hosts with 96 Nvidia Fermi GPUs at the CUNY High Performance Computing Center, the College of State Island.

Each Fermi has 448 thread processors allowing an unprecedented amount of parallel computation. Topics to be covered involve a review of SIMD programming, a software view of the multi-threaded programming model, the interplay of the GPU hardware and the CUDA language, how to increase efficiency in CUDA codes, using multiple GPUs on one host and distributing calculations to multiple GPUs over a network. Several applications will be discussed in detail. Students will be expected to write several smaller programs and develop one major project.