

Uvod u numeričko računanje na grafičkim karticama uz pomoć CUDA-e

Domagoj Matijević

Odjel za matematiku, Sveučilište J.J. Strossmayera u Osijeku

Sažetak: CUDA ili *Compute Unified Device Architecture* je arhitektura za paralelno izvođenje poslova na grafičkom procesoru (GPU) razvijenom od tvrtke NVIDIA. CUDA omogućuje programeru jednostavan pristup virtualnom skupu instrukcija i memoriji NVIDIA-inih grafičkih kartica. Uz pomoć CUDA-e, najnoviji NVIDIA GPU su postali dostupni za računanje na sličan način na koji su to u tradicionalnom smislu bili CPU-ovi. Međutim, za razliku od CPU, GPU ima paralelnu arhitekturu s naglaskom na izvođenju mnogo niti (*threads*) „polako“ umjesto „brzog“ računanja u jednoj niti. Moderne NVIDIA kartice, poput TESLA C2070 koju imamo instaliranu na Odjelu, mogu vrtjeti 448 niti paralelno. U ovom seminaru mi je želja kroz jednostavne primjere, poput zbrajanja dva vektora ili skalarnog produkta dva vektora, objasniti osnove programiranja u CUDA C-u te logički koncept paralelizacije nekog numeričkog problema na grafičkoj kartici kroz grid, blokove i niti, te dijeljenu (*shared*) memoriju.