



COMPARAÇÃO ENTRE A PROGRAMAÇÃO SERIAL (C/C++) E A PARALELA UTILIZANDO CUDA.

Mauro Sergio Dos Santos Moura e Josecley Fialho Góes

Com o avanço da tecnologia surgiram os processadores de vários núcleos, chamados multicore, facilitando a introdução ao conceito de programação paralela, ou seja dividindo tarefas entre núcleos. A programação paralela utilizando GPU (Graphics Processing Unit), diminuiu o tempo e o número de máquinas necessários para a execução de códigos de grande escala, desde que os mesmos sejam paralelizáveis, já que as GPUs possuem mais núcleos que as CPUs (Central Processing Unit). Visando o aperfeiçoamento das GPUs como núcleos de processamento para cálculos, a NVIDIA investiu em uma API para extrair o máximo de potencial da GPU, criando assim a API CUDA (Compute Unified Device Architecture). Mesmo com vantagens em relação à programação serial, a programação com CUDA ainda é pouco explorada no meio da pesquisa, tendo em mente que o mesmo foi lançado em 2006. Logo, com o intuito de testar a diferença na velocidade de processamento entre a programação serial (C/C++) e a programação paralela, utilizando a CUDA, comparando os resultados obtidos, foram feitos testes benchmark de multiplicação de matrizes, nas versões serial e paralela. Pode-se observar que as versões em paralelo foram mais rápidas, principalmente nos casos com muitos valores ou muitas variáveis, mas nos casos com poucas variáveis, os códigos desenvolvidos com CUDA já não foram tão vantajosos. Uma das causas, é a necessidade de copiar os vetores e matrizes da CPU para GPU e vice-verso. Nesse contexto, o investimento em programação paralela é fundamental e a utilização da mesma é importante pelo fato de reduzir o custo computacional na maioria dos casos, principalmente para a área de modelagem computacional, onde existem vários vetores e variáveis físicas que precisam ser calculados ao mesmo tempo para a simulação de um problema real.