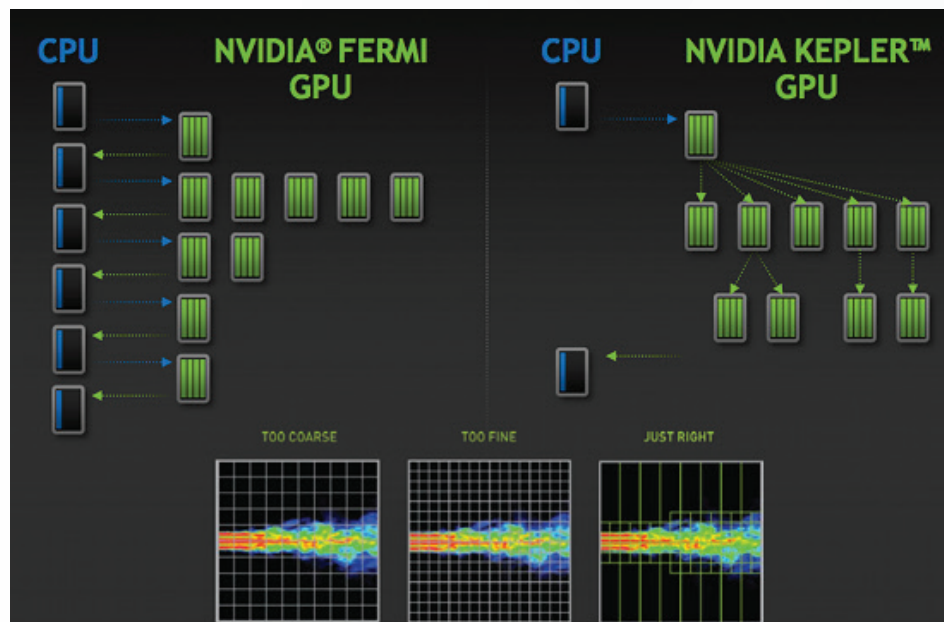




katedra
informatiky



Implementácia algoritmu násobenia matic na GPGPU s optimalizáciou prenosu údajov

Miroslav Kováč

ABSTRAKT

Táto práca sa zaoberá optimalizáciou násobenia dvoch matic s dvojnásobnou presnosťou na GPGPU architektúre s názvom Kepler. Zistili sme, že pri násobení matic použitím GPGPU zaberá prenos dát na GPU podstatnú časť behu programu. Skrátene tohto času by značne pomohlo k rýchlejšiemu získaniu výsledkov násobenia. My sme tento čas neskrátili, len sme redizajnovali algoritmus tak, aby kopírovanie bežalo čo najviac súbežne s počítaním. Predstavujeme Vám algoritmus s optimalizáciou prenosu údajov. Naša optimalizačná stratégia je ďalej porovnávaná na základe referenčných kritérií výkonnostného modelu. Vytvorili sme funkčnú metódu, ktorá môže byť použitá v každom CUDA programe. Naša optimalizácia zahŕňa používanie asynchrónnej metódy na každej matematickej operácii a Strassenov algoritmus pri operáciách na rozdelených maticiach. Prenášanie údajov počas toho, ako sa robia iné potrebné operácie je kľúčovou časťou pre zvýšenie výkonu. Najlepší CUDA algoritmus, ktorý sme vytvorili prekonáva algoritmus z knižnice CUBLAS. Dosiahli sme priemerne 26 % zrýchlenie.

Kľúčové slová: GPGPU, optimalizácia prenosu údajov, vysoko-výkonné počítanie, CUDA, DGEMM

Strassenov algoritmus

10 krát sčítanie alebo odčítanie

$$\begin{bmatrix} 10 & 20 & 30 & 40 \\ 50 & 60 & 70 & 80 \\ 90 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix} \times \begin{bmatrix} 10 & 20 & 30 & 40 \\ 50 & 60 & 70 & 80 \\ 90 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix}$$

7 násobení

$$\begin{aligned} S_1 &= B_{12} - B_{22}, \\ S_2 &= A_{11} + A_{12}, \\ S_3 &= A_{21} + A_{22}, \\ S_4 &= B_{21} - B_{11}, \\ S_5 &= A_{11} + A_{22}, \\ S_6 &= B_{11} + B_{22}, \\ S_7 &= A_{12} - A_{22}, \\ S_8 &= B_{21} + B_{22}, \\ S_9 &= A_{11} - A_{21}, \\ S_{10} &= B_{11} + B_{12}, \end{aligned}$$

$$\begin{aligned} P_1 &= A_{11} \cdot S_1, \\ P_2 &= S_2 \cdot B_{22}, \\ P_3 &= S_3 \cdot B_{11}, \\ P_4 &= A_{22} \cdot S_4, \\ P_5 &= S_5 \cdot S_6, \\ P_6 &= S_7 \cdot S_8, \\ P_7 &= S_9 \cdot S_{10}, \end{aligned}$$

Výsledný výpočet

$$\begin{aligned} C_{11} &= P_5 + P_4 - P_2 + P_6, \\ C_{12} &= P_1 + P_2, \\ C_{21} &= P_3 + P_4, \\ C_{22} &= P_5 + P_1 - P_3 - P_7, \end{aligned}$$
