

Esercitazione 5

Sviluppare e implementare in linguaggio CUDA-C tre algoritmi che calcolino il prodotto scalare di due vettori di dimensione N. Applicare le seguenti strategie:

1. strategia I (banale) con solo uso della global memory;
2. strategia II (versione 1-divergente), con uso della shared memory;
3. strategia III, con uso della shared memory, per evitare divergenza e conflitti di accesso ai banchi della shared memory.

Usare la stessa configurazione del kernel per i tre programmi (tenere conto delle limitazioni sia sulla griglia, che sulla memoria richiesta: registri e memoria condivisa). Per la configurazione del kernel usata, calcolare il numero di blocchi residenti in uno streaming multiprocessor, in base alla GPU utilizzata.

Calcolare i tempi di esecuzione e lo Speedup al variare della dimensione del problema N ($N > 100.000$, N multiplo di 32). Calcolare il tempo di esecuzione a partire da dopo la copia dei dati da host a device, fino al termine del calcolo del prodotto scalare.