ESERCITAZIONE 5

Marrazzo Vincenzo
Spagna Zito Marika

Caratteristiche GPU:

GPU Colab: Tesla T4

Compute Capability: 7.5

Technical specifications	Compute Capability
Maximum x-dimension of a grid of thread blocks	2^{31}-1
Maximum y-, or z-dimension of a grid of thread blocks	65535
Maximum number of threads per block	1024
Maximum number of resident blocks per multiprocessor	16
Maximum number of resident threads per multiprocessor	1024
Number of 32-bit registers per multiprocessor	64K

CONFIGURAZIONE 1: Blocco 4x4

Ogni SM può gestire fino a 1024 thread ed un numero massimo di 16 blocchi.

Considerando il **blocco 4x4** abbiamo un totale di 16 thread per blocco.

$(4x4) = 16 < 1024$ **thread** per blocco, vincolo della dimensione del blocco soddisfatto.

Thread per SM ← $\frac{1024}{16} = 64\ blocchi$ (non si può fare)

Thread per blocco ←

Non si può fare poiché possiamo considerare solo un massimo di 16 blocchi per SM. Quindi otteniamo:

$16\ (\text{blocchi})\ \text{x}\ 16(\text{thread per blocco}) = 256$ thread in totale per SM su un totale di 1024 thread utilizzabili per SM.

In questa configurazione utilizziamo il numero massimo di blocchi ma il numero di thread risulta troppo basso rispetto a quelli disponibili.

CONFIGURAZIONE 2: Blocco 8x8

Ogni SM può gestire fino a 1024 thread ed un numero massimo di 16 blocchi.

Considerando adesso un **blocco 8x8** abbiamo un totale di 64 thread per blocco.

$(8x8) = 64 < 1024$ **thread** per blocco**,** vincolo della dimensione del blocco soddisfatto.

$$\frac{1024}{64} = 16\ blocchi$$

Siccome soddisfa il massimo numero di blocchi per SM consideriamo tutti e 16 i blocchi.

16(blocchi) x 64(thread per blocco)= 1024 thread per SM (Otteniamo la piena occupazione dello SM sia per quanto riguarda i thread che per i blocchi)

VALORE OTTIMALE

Controlliamo se questo valore rispetta il vincolo della memoria attraverso l'istruzione `!nvcc -Xptxas -v Ese5.cu`. che ci fornisce il numero di registri utilizzato da ogni thread. Otteniamo il valore 8 e notiamo:

n° registri per ogni SM ← 1024 x 8 = 8.192 < 64.000

n° di thread per ogni SM

n° registri per ogni thread

CONFIGURAZIONE 3: Blocco 16x16

Ogni SM può gestire fino a 1024 thread ed un numero massimo di 16 blocchi.

Considerando il **blocco 16x16** abbiamo un totale di 256 thread per blocco.

$(16x16) = 256 < 1024$ thread per blocco, vincolo della dimensione del blocco soddisfatto.

$$\frac{1024}{256} = 4\ blocchi$$

Siccome consideriamo 4 blocchi per SM avremo $256 \times 4 = 1024$ thread per ogni SM.

In questa configurazione, nonostante riusciamo ad ottenere il numero massimo di thread, il numero di blocchi utilizzati sono soltanto 4 su un totale massimo di 16 blocchi. Con un numero così basso di blocchi otteniamo un parallelismo minore rispetto a quello ottenuto nella configurazione 8 x 8 (dove consideriamo 16 blocchi).

SPEED-UP CONFIGURAZIONE 4x4

N	tempo CPU (s)	tempo GPU (s)	Sp
1.024	0,003394	0,000238	14,26
2.048	0,014658	0,000721	20,33
4.096	0,072563	0,002929	24,77
8,19E+03	2,14E-01	0,012296	17,42

SPEED-UP CONFIGURAZIONE 8x8

N	tempo CPU (s)	tempo GPU(s)	Sp
1.024	0,003204	0,000097	33,03
2.048	0,012828	0,000216	59,39
4.096	0,051591	0,000839	61,49
8,19E+03	2,13E-01	0,003251	65,52

SPEED-UP CONFIGURAZIONE 16x16

N	tempo CPU (s)	tempo GPU (s)	Sp
1.024	0,003170	0,000084	37,74
2.048	0,012704	0,000214	59,36
4.096	0,054817	0,000828	66,20
8,19E+03	2,15E-01	0,003255	66,19