

Esercitazione 4: Prodotto Scalare Vettori

GPU: Tesla T4

Compute capability: 7.5

Massimo numero di thread per blocco per SM: 1024

Numero massimo di blocchi residenti per SM: 16

Massimo numero di registri a 32 bit per blocco di thread: 64K

Configurazione 1 : 32

N	Tempo CPU	Tempo GPU	Sp
100000	0.437	0.021	20.349
200000	0.896	0.027	33.815
400000	1.813	0.049	36.925
800000	3.571	0.087	40.849
1600000	7.193	0.168	42.843
3200000	14.398	0.326	44.139

32 thread per blocco: $1024/32 = 32$ blocchi residenti.

Con un massimo di 16 blocchi per SM : $32 \times 16 = 512$ thread per SM su un totale di 1024 disponibili.

Uso dei registri

Eseguendo l'istruzione "nvcc -Xptxas -v" ottengo che il numero di registri utilizzato da ogni thread è 8.

Dunque, moltiplicando il numero di registri, per il numero di thread e per il numero di blocchi ottengo:

$$8 \times 32 \times 16 = 4096 < 64K$$

Configurazione 2 : 64

N	Tempo CPU	Tempo GPU	Sp
100000	0.441	0.019	23.816
200000	0.992	0.045	22.195
400000	2.056	0.088	23.366
800000	3.970	0.168	23.701
1600000	8.317	0.324	25.710
3200000	17.612	0.627	28.103

64 thread per blocco: $1024/64 = 16$ blocchi residenti.

Con 16 blocchi: $64 \times 16 = 1024$ thread per SM.

Piena occupazione dello SM!

Uso dei registri

Eseguendo l'istruzione "nvcc -Xptxas -v" ottengo che il numero di registri utilizzato da ogni thread è 8.

Dunque, moltiplicando il numero di registri, per il numero di thread e per il numero di blocchi ottengo:

$$8 \times 64 \times 16 = 8192 < 64K$$

Configurazione 3 : 128

N	Tempo CPU	Tempo GPU	Sp
100000	0.70	0.03	23.33
200000	1.00	0.05	22.06
400000	2.09	0.09	24.08
800000	4.10	0.17	24.72
1600000	7.94	0.32	24.87
3200000	17.77	0.63	28.34

128 thread per blocco: $1024/128 = 8$ blocchi residenti.

Con 8 blocchi: $128 \times 8 = 1024$ thread per SM. Piena occupazione dello SM ma minore parallelismo.

Uso dei registri

Eseguendo l'istruzione "nvcc -Xptxas -v" ottengo che il numero di registri utilizzato da ogni thread è 8.

Dunque, moltiplicando il numero di registri, per il numero di thread e per il numero di blocchi ottengo:

$$8 \times 128 \times 8 = 8192 < 131K$$



