

Παράλληλα και Διανεμημένα Συστήματα

Εργασία 3 Ising Model

Φοιτητής : Ζηκόπης Ευάγγελος / Παπαγεωργίου Δημήτριος
ΑΕΜ : 8808 / 8884

Email : ezikopis@ece.auth.gr / dkpapageo@ece.auth.gr

15/01/2020

Κώδικας

Ο κώδικας βρίσκεται στο παρακάτω github repository. Για την ορθή εκτέλεσή του παρακαλώ ακολουθήστε τα βήματα που περιγράφονται στο README του repository: <https://github.com/vagzikopis/IsingCuda>

Sequential

Η σειριακή υλοποίηση αποτελεί μια απλή προσέγγιση στην λύση του προβλήματος, χωρίς την επιστράτευση τεχνικών και μεθόδων παραλληλοποίησης. Τρέχει στην CPU και ικανοποιεί πλήρως τα ζητούμενα της εκφώνησης.

V1

Η V1 έκδοση χρησιμοποιεί την GPU προκειμένου να πετύχει ταχύτερη εκτέλεση του προγράμματος. Σε αυτήν την έκδοση, η αναλογία threads-moments είναι 1-1, δηλαδή κάθε thread δουλεύει για τον υπολογισμό του atomic spin ενός moment στο πλέγμα(square lattice). Χρησιμοποιήθηκαν 2D Grid και 2D Blocks. Η παράμετρος BLOCKSIZE, που καθορίζει τη διάσταση των Block γίνεται define στην αρχή του προγράμματος. Με βάση αυτήν την τιμή υπολογίζονται οι διαστάσεις του Grid προκειμένου να προκύπτει πάντα 1-1 αναλογία thread-moment, ανεξάρτητα από την τιμή της παραμέτρου n. Ο τύπος υπολογισμού είναι:

```
if(n % BLOCK_SIZE == 0)
    GRID_SIZE = n/BLOCK_SIZE;
else
    GRID_SIZE = n/BLOCK_SIZE + 1;
```

n = 2000 k = 100				n = 1000 k = 100		
TIME(ms)	GRID	BLOCK		TIME(ms)	GRID	BLOCK
521.85	500	4		166.42	250	4
332.20	250	8		85.89	125	8
334.17	125	16		87.20	63	16
338.46	84	24		88.07	42	24
341.42	63	32		90.33	32	32

Από τα τεστ που έγιναν, ο μικρότερος χρόνος καταγράφηκε για BLOCK_SIZE = 8, οπότε αφέθηκε αυτή η τιμή για τα επόμενα τεστ. Οι παραπάνω χρόνοι αποτελούν τους μέσους όρους από τα πειράματα που έγιναν.

V2

Η V2 έκδοση αποτελεί την βάση της V3. Σε αυτήν την υλοποίηση κάθε thread υπολογίζει το atomic spin για περισσότερα από ένα σημεία. Προκειμένου να πετύχουμε όσο το δυνατόν μεγαλύτερο efficiency στις προσπελάσεις μνήμης που κάνουν τα thread(memory coalescing), κάθε block of threads υπολογίζει το atomic spin σημείων που είναι αποθηκευμένα σε διπλανάς θέσεις μνήμης. Όταν υπολογιστούν αυτά τα moments τότε το block of threads προχωρά στον υπολογισμό των atomic spins από άλλη ομάδα απο moment. Για παράδειγμα, αν το square lattice είναι 4x4 και έχουμε GRID 4x4 και BLOCK 4x4, τότε οι προσπελάσεις των moment από τα thread θα ακολουθούν την παρακάτω εικόνα(με διαφορετικό χρώμα απεικονίζονται τα διαφορετικά block).

Παρατηρούμε ότι οι ελάχιστοι χρόνοι καταγράφηκαν για GRID_SIZE 32x32 και BLOCK_SIZE 32x32. Το γεγονός αυτό οφείλεται και στο ότι η κάρτα που χρησιμοποιήθηκε(Nvidia Tesla P100, συστοιχία hpc) έχει wrap size 32 thread. Επομένως αυτές οι παράμετροι χρησιμοποιήθηκαν και στα πειράματα για το V3.

n=2000 & k=100			n=1000 & k=100																			
time(ms)	GRID	BLOCK	time(ms)	GRID	BLOCK																	
2037.25	4	8	349.432	4	16	#	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
768.032	4	16	171.93	8	16	0																
597.716	4	24	100.352	16	16	1																
547.153	4	32	97.5035	24	16	2																
714.409	8	8	92.304	32	16	3																
486.263	8	16	98.6815	32	8	4																
459.518	8	24	92.8885	32	24	5																
457.406	8	32	90.9796	32	32	6																
541.687	12	8	95.2458	24	32	7																
432.188	12	16	91.1731	24	24	8																
407.809	12	24	114.41	24	8	9																
443.766	12	32	152.62	16	8	10																
478.103	16	8	106.581	16	24	11																
380.069	16	16	97.3956	16	32	12																
428.647	16	24	139.629	8	32	13																
379.489	16	32	179.775	8	24	14																
373.035	32	8	351.26	8	8	15																
363.551	32	16	670.671	4	8																	
356.298	32	24	308.271	4	24																	
349.262	32	32	292.576	4	32																	

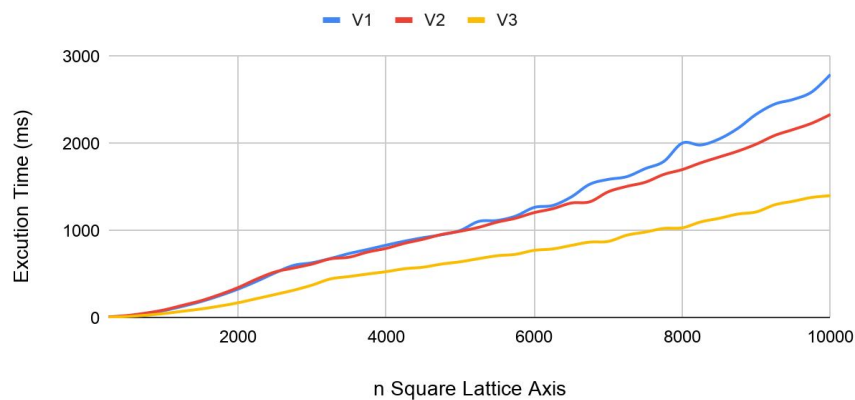
V3

Η V3 έκδοση ακολουθεί τα όσα περιγράφηκαν στην V2, με τη μόνη διαφορά ότι τα σημεία με τα οποία δουλεύει το κάθε block και οι απαραίτητοι γείτονες αποθηκεύονται πρώτα στην shared memory. Για να επιτευχθεί αυτό, κάθε thread αποθηκεύει το σημείο με το οποίο δουλεύει, και κάποια συγκεκριμένα threads αποθηκεύουν επίσης τους εξωτερικούς γείτονες απο το συνολικό block σημείων, οι οποίοι θα είναι απαραίτητοι για τα moments που βρίσκονται στις άκρες. Επίσης στην shared memory αποθηκεύεται και ο πίνακας με τα βάρη, κάτι που ωστόσο δεν επηρεάζει ιδιαίτερα τον συνολικό χρόνο εκτέλεσης.

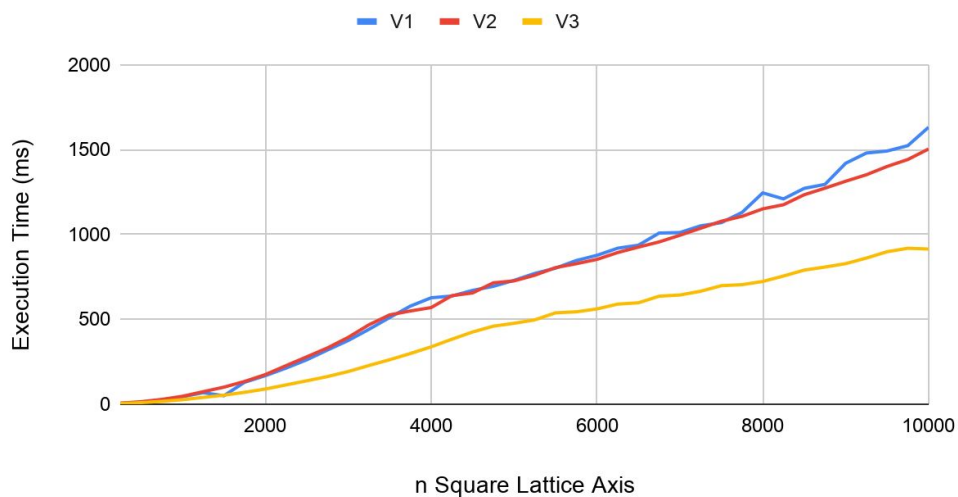
Πειράματα

Τα πειράματα τρέξανε στη συστοιχία hpc και στο partition pdlabs(Nvidia Tesla P100). Λήφθηκαν οι μέσοι όροι απο τους χρόνους που καταγράφηκαν συνολικά. Η σειριακή υλοποίηση τοποθετήθηκε σε άλλο διάγραμμα καθώς οι χρόνοι απέχουν τρείς τάξεις μεγέθους(millisecond - second)

K=100



K=50



Sequential

