Αναφορά για τα ερωτήματα

1)

Για το ερώτημα αυτό χρησιμοποίησα τον κώδικα από το προηγούμενο lab,με μόνη διαφορά ότι τοποθέτησα τους timers όπως αναφέρεται στην εκφώνηση.Οι μετρήσεις για κάθε μέγεθος εικόνας είναι ο μέσος όρος 5 τρεξιμάτων για μεγαλύτερη ακρίβεια.Όπως αναμέναμε οι χρόνοι εκτέλεσης είναι μικρότεροι όταν είναι ενεργοποιημένα όλα τα επίπεδα των caches,ενώ αυξάνονται όταν τα αφαιρούμε.Οι μεγαλύτεροι χρόνοι παρατηρούνται με την ταυτόγχρονη απενεργοποίηση των L1 και L2 caches.Οι χρόνοι αναπαρίστανται στο παρακάτω διάγραμμα(σε sec):

Χρησιμοποίησα τα συγκεκριμένα μεγέθη εικόνας ώστε οι χρόνοι να είναι μεγαλύτεροι και να αναπαρίστανται οι διαφορές καλύτερα.Στο αρχείο 2.1 metriseis,υπάρχουν αναλυτικά όλοι οι χρόνοι για όλα τα μεγέθη εικόνας και για όλες τις caches.

2)

Επειδή χρησιμοποιώ 2D tile,παρουσιάζεται bank conflict στο warp αφού κάθε thread κάνει access στη memory την ίδια στιγμή.Αυτό το βελτιώνω στο βήμα 4 όπου και υλοποιώ 1D tile.

3)

Αυτές είναι οι ποσότητες των πόρων που καταγράφηκαν με την παράμετρο:

soutis@inf-zeus2:~$ nvcc -o lab2\_2 lab2\_2.cu -O4 --ptxas-options=-v

ptxas info : 0 bytes gmem, 132 bytes cmem[0]

ptxas info : Compiling entry function '\_Z16tiled\_row\_KernelPfS\_iii' for 'sm\_10'

ptxas info : Used 11 registers, 4140 bytes smem, 4 bytes cmem[1]

ptxas info : Compiling entry function '\_Z10row\_KernelPfS\_S\_iii' for 'sm\_10'

ptxas info : Used 9 registers, 52 bytes smem, 4 bytes cmem[1]

ptxas info : Compiling entry function '\_Z19tiled\_column\_KernelPfS\_iii' for 'sm\_10'

ptxas info : Used 14 registers, 4140 bytes smem, 4 bytes cmem[1]

ptxas info : Compiling entry function '\_Z13column\_KernelPfS\_S\_iii' for 'sm\_10'

ptxas info : Used 9 registers, 52 bytes smem, 8 bytes cmem[1]

Το device μας έχει 49152 bytes shared memory(από device Query).

Οπότε το μέγιστο μέγεθος tile που μπορούμε να υποστηρίξουμε είναι 96\*96.

Επειδή όμως το block size = tile size,και επειδή έχουμε 1024 threads/block

το γραμμικό tile που μπορούμε να υποστηρίξουμε είναι 32χ32(=1024).

4)

Στο optimization που υλοποίησα,χρησιμοποίησα 1D tile ώστε να αποφύγω το bank conflict στο warp όπως είπα και προηγουμένως.Οι χρόνοι κατέβηκαν πολύ.Είναι μικρότεροι από την μη tiled έκδοση,πράγμα που αναμέναμε με την χρήση της shared memory.Ακόμα και όταν χρησιμοποιούμε 48kb L1 cache για να πετύχουμε τον μικρότερο δυνατό χρόνο για την μη tiled έκδοση,οι χρόνοι με τις tiled συναρτήσεις είναι οριακά μικρότεροι.Αναλυτικά οι χρόνοι βρίσκονται στο lab2\_4\_metriseis.Βέβαια παρατηρούμε ότι διαφορές αντιληπτές εμφανίζονται για πιo μεγάλα μεγέθη εικόνας.Στα μικρά οι χρόνοι είναι σχεδόν ίδιοι.

Επίσης για ακρίβεια στους χρόνους η main καλεί 10 φορές τους kernels και βγάζει μέσο όρο.

Στο διάγραμμα χρησιμοποιώ 16kb L1 cache και μεγέθη εικόνας 1024,2048,4096 και 8192(σε sec):

6)

Αυτό που παρατηρούμε είναι ότι οι απενεργοποιήσεις των caches έχουν πολυ μεγάλο αντίκτυπο στους χρόνους των μη tiled συναρτήσεων με αποτέσμα να αυξάνεται ο χρόνος εκτέλεσης.Αντίθετα οι tiled συναρτήσεις επειδή εκμεταλλεύονται την shared memory και όχι τόσο την επίδραση των caches,δεν επηρεάζονται τόσο.Οι χρόνοι ναι μεν αυξάνονται αλλά πολύ πιο λίγο σε σχέση με τις μη tiled συναρτήσεις.

Διάγραμμα με απενεργοποιημένη την L1 cache(σε sec):

Διάγραμμα με απενεργοποιημένες τις L1 & L2 caches(σε sec):

Αναλυτικά οι μετρήσεις για όλα τα μεγέθη εικόνας υπάρχουν στο lab2\_6\_metriseis.

Παρατηρούμε ότι μεγαλύτερη επίπτωση έχει η απενεργοποίηση της L1 cache στις μη tiled συναρτήσεις ενώ στις tiled μεγαλύτερη επίδραση έχει η ταυτόγχρονη απενεργοποίηση των L1/L2 caches.Μάλιστα οι μη tiled συναρτήσεις με απενεργοποιημένα και τα 2 επίπεδα των caches είναι πιο γρήγορες σε σχέση με απενεργοποιημένο μόνο το επίπεδο της L1 cache.

7)

Οι doubles καταλαμβάνουν 8 bytes ενώ οι float 4 bytes.Οπότε με τους περιορισμούς από το device μας(49152 bytes per block),μπορώ να χρησιμοποιήσω μέχρι 64\*64 γραμμικό block / tile.Οι χρόνοι αυξάνονται σε σχέση με τους floats.

Στο διάγραμμα αντιπαραβολής με τους floats χρησιμοποιώ 48kb L1 cache για τις μη tiled συναρτήσεις.

Διάγραμμα:

Παρατηρούμε ότι οι double συναρτήσεις είναι πιο αργές αλλά η double non tile έκδοση είναι πιο γρήγορη απο την αντίστοιχη tile.Επίσης δεν υποστηρίζεται μέγεθος εικόνας 8192\*8192 στις doubles συναρτήσεις.Οι αναλυτικές μετρήσεις υπάρχουν στο lab2\_7\_metriseis.