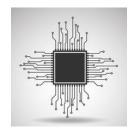
ΕΘΝΙΚΟ & ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ & ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ ΜΑΘΗΜΑ: **ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΥΠΟΛΟΓΙΣΤΩΝ ΙΙ**





ΑκαΔΗΜΑΪΚΟ ΕΤΟΣ 2024 - 2025

(εκφωνηση) Τρίτη 3 Δεκεμβρίου 2024

(ΠΑΡΑΔΟΣΗ ΣΤΟ ECLASS MEXPI) **ΔΕΥΤΕΡΑ 23 ΔΕΚΕΜΒΡΙΟΥ 2024**

Εργασία 2 (υποχρεωτική) – Κρυφές Μνήμες

Επώνυμο – Όνομα	Πατρώνυμο – Μητρώνυμο	Αριθμός Μητρώου	Email
Κουλλόλλι Μικαέλα	Γκεζίμ-Ετλέβα	1115202100071	sdi2100071@di.uoa.gr
Αντρέας Παύλου	Θεόδωρος-Γεωργία	1115202100223	sdi2100223@di.uoa.gr

Πληροφορίες για τις Υποχρεωτικές Εργασίες του μαθήματος

- Οι υποχρεωτικές εργασίες του μαθήματος είναι δύο. Σκοπός τους είναι η κατανόηση των εννοιών του μαθήματος με χρήση αρχιτεκτονικών προσομοιωτών. Η πρώτη υποχρεωτική εργασία αφορά τη διοχέτευση (pipelining) και η δεύτερη (αυτή) αφορά τις κρυφές μνήμες (cache memories).
- Οι δύο εργασίες είναι υποχρεωτικές και η βαθμολογία του μαθήματος θα προκύπτει από το γραπτό (60%), την εργασία της διοχέτευσης (20%), και την εργασία των κρυφών μνημών (20%).
- Κάθε ομάδα μπορεί να αποτελείται από 1 έως και 2 φοιτητές (η υποβολή να γίνει μόνο από έναν φοιτητή εκ μέρους όλης της ομάδας μην κάνετε υποβολές όλοι). Συμπληρώστε τα στοιχεία των μελών της ομάδας στον παραπάνω πίνακα. Τα μέλη της ομάδας πρέπει να έχουν ισότιμη συμμετοχή και να γνωρίζουν τις λεπτομέρειες της υλοποίησης.
- Για την εξεταστική Σεπτεμβρίου δεν θα δοθούν άλλες εργασίες. Το Σεπτέμβριο εξετάζεται μόνο το γραπτό.
- Σε περίπτωση αντιγραφής θα μηδενίζονται όλες οι ομάδες που μετέχουν σε αυτή.
- Η παράδοση της Εργασίας Κρυφών Μνημών πρέπει να γίνει ηλεκτρονικά μέχρι τα μεσάνυχτα της προθεσμίας και μόνο στο eclass (να ανεβάσετε ένα μόνο αρχείο zip ή rar με την τεκμηρίωσή σας σε PDF και τους κώδικές σας).
 Μην περιμένετε μέχρι την τελευταία στιγμή κάθε εργασία απαιτεί τον χρόνο της.

Ζητούμενο

Το ζητούμενο της εργασίας είναι η αξιολόγηση της ιεραρχίας μνήμης επεξεργαστών MIPS για συγκεκριμένες υπολογιστικές εργασίες με βάση το μοντέλο Roofline (γραμμή στέγης) όπως το είδαμε στο Κεφάλαιο 6 του COD6e. Η σχεδίαση του υλικού και του λογισμικού και η αξιολόγηση θα γίνει στον προσομοιωτή QtMips.

Οι υπολογιστικές εργασίες είναι: (α) πολλαπλασιασμός διανύσματος η στοιχείων επί βαθμωτό, (β) πολλαπλασιασμός τετραγωνικού πίνακα ηχη επί βαθμωτό, (γ) πολλαπλασιασμός δύο τετραγωνικών πινάκων ηχη. Για κάθε εργασία, οι αρχικοί πίνακες και τα αποτελέσματα θα βρίσκονται στο τμήμα δεδομένων του προγράμματος. Να εξετάζεται το ενδεχόμενο υπερχείλισης κατά τη διάρκεια εκτέλεσης των πράξεων και το πρόγραμμα να τερματίζει εάν συμβεί υπερχείλιση. Τα στοιχεία των πινάκων είναι ακέραιοι αριθμοί. Να παρουσιάσετε τα αποτελέσματα της μελέτης σας για τις τιμές η=8, 16, 32 και προαιρετικά για μεγαλύτερες δυνάμεις του 2.

Για τη χρήση του μοντέλου Roofline ως «απόδοση» να χρησιμοποιήσετε το πλήθος των πολλαπλασιασμών ανά δευτερόλεπτο (MPS – multiplications per second) Ω ς αριθμητική ένταση (arithmetic intensity) να χρησιμοποιήσετε τον λόγο πολλαπλασιασμοί ανά byte (MPB – multiplications per byte).

Τα συστήματα ΜΙΡS που θα αξιολογήσετε είναι:

MIPS-A: διοχέτευση πέντε σταδίων με πλήρη μονάδα κινδύνων με ανίχνευση και προώθηση και branch predictor των 2-bit με BHT που προσπελάζεται με 5 bit και επίλυση των διακλαδώσεων στο στάδιο ΕΧ. Δεν διαθέτει καθόλου κρυφή μνήμη, έχει ρυθμό ρολογιού 100 MHz, και η προσπέλαση της κύριας μνήμης διαρκεί 60 κύκλους ρολογιού.

MIPS-B: διοχέτευση ίδια με την παραπάνω, αλλά το σύστημα μνήμης περιέχει και κρυφή μνήμη ενός επιπέδου με μέγεθος 8KB για εντολές και ξεχωριστά 8KB για δεδομένα. Μπορείτε να επιλέξετε το μέγεθος μπλοκ της καθεμίας μεταξύ 2, 4, και 8 λέξεων και τη συσχετιστικότητα τους μεταξύ 1, 2, και 4 (πολιτική αντικατάστασης LRU όπου απαιτείται). Η πολιτική εγγραφής να είναι ετερόχρονη (write back) και με κατανομή σε εγγραφή (write allocate). Ο ρυθμός ρολογιού του επεξεργαστή παραμένει σταθερός και το ίδιο ο χρόνος προσπέλασης της κύριας μνήμης.

MIPS-Γ: διοχέτευση ίδια με τους παραπάνω, και πρώτο επίπεδο κρυφής μνήμης όπως ο προηγούμενος. Διαθέτει όμως και δεύτερο επίπεδο κρυφής μνήμης με μέγεθος 64KB ενιαία για εντολές και δεδομένα. Μπορείτε να επιλέξετε το μέγεθος μπλοκ μεταξύ 2, 4, και 8 λέξεων και τη συσχετιστικότητα μεταξύ 1, 2, και 4 (πολιτική αντικατάστασης LRU όπου απαιτείται). Η πολιτική εγγραφής να είναι ετερόχρονη (write back) και με κατανομή σε εγγραφή (write allocate). Η προσπέλαση του δεύτερου επιπέδου κρυφής μνήμης διαρκεί 6 κύκλους ρολογιού. Ο ρυθμός ρολογιού του επεξεργαστή παραμένει σταθερός και το ίδιο ο χρόνος προσπέλασης της κύριας μνήμης.

Να συμπληρώσετε τον παρακάτω πίνακα με **τις σχεδιαστικές αποφάσεις σας** για τις κρυφές μνήμες κάθε επεξεργαστή που διαθέτει τέτοιες.

Χαρακτηριστικό	MIPS-A	MIPS-B	MIPS-T
L1 caches μέγεθος μπλοκ (2, 4, 8 λέξεις)	-	8	8
L1 caches συσχετιστικότητα (1, 2, 4)	-	4	4
L2 cache μέγεθος μπλοκ (2, 4, 8 λέξεις)	-	-	8
L2 cache συσχετιστικότητα (1, 2, 4)	-	-	4

Για κάθε CPU να σχεδιάσετε τα διαγράμματα Roofline και το σημείο που βρίσκονται σε αυτά οι υπολογιστικές εργασίες για τις διάφορες τιμές της παραμέτρου n. Να σχολιάσετε τις συμπεριφορές που παρατηρείτε και πιθανές βελτιώσεις του κάθε συστήματος για τις συγκεκριμένες υπολογιστικές εργασίες.

Τεκμηρίωση

[Σύντομη τεκμηρίωση της λύσης σας ενδεικτικά (όχι αυστηρά) μέχρι 10 σελίδες ξεκινώντας από την επόμενη σελίδα – μην αλλάζετε τη μορφοποίηση του κειμένου (και παραδώστε την τεκμηρίωση σε αρχείο PDF). Η τεκμηρίωσή σας πρέπει να περιλαμβάνει παραδείγματα εκτέλεσης και σχολιασμό για την επίλυση του προβλήματος και την επίτευξη του ζητούμενου. Μπορείτε να χρησιμοποιήσετε εικόνες, διαγράμματα και ό,τι άλλο μπορεί να βοηθήσει στην εξήγηση της δουλειάς σας.]

[ΚΩΔΙΚΑΣ]

Όσον αφορά τον αλγόριθμο πολλαπλασιασμού πινάκων, υλοποιήθηκε αλγόριθμος με χρήση της τεχνικής tiling, δηλαδή να γίνεται διαχωρισμός των πινάκων σε πολλούς υποπίνακες, όπου για πίνακες διαστάσεων X * Y, είναι απαραίτητο ο Y να είναι ο ανάστροφος (transpose) πίνακας. Στην περίπτωση μας, επιλέχθηκε το kernel 4x4, ώστε να βελτιστοποιηθεί η απόδοση ως προς τη χρήση LRU caching. Παρακάτω παρατίθενται στιγμιότυπα οθόνης (screenshots), τα οποία αποδεικνύουν την ορθότητα των προγραμμάτων μας, τα οποία έχουν ελεγχθεί μέσω ενός Python script, το οποίο είναι επίσης διαθέσιμο στα προγράμματα.

**Για τα άλλα δύο προγράμματα - εργασίες η εκτέλεση τους για τα διαφορετικά η γίνεται αλλάζοντας την μεταβλητή η Στο data segment.

N = 8	N = 16

```
Console
  Console
                                                                                                                                                                        8 16 24 32 40 48 56 64
16 32 48 64 80 96 112 128
24 48 72 96 120 144 168 192
32 64 96 128 160 192 224 256
40 80 120 160 200 240 280 320
48 96 144 192 240 288 336 384
56 112 168 224 280 336 392 448
64 128 192 256 320 384 448 512
                                                                                                                                                                                                                       128
256
384
512
640
768
896
1024
128
256
384
512
640
768
896
                                                                                                                                                                                                                             16 32
32 64
48 96
64 128
80 160
96 192
112 224
128 256
16 32
32 64
48 96
64 128
80 160
96 192
112 224
                                                                                                                                                                                                                                                64 88
128 169
192 248
192 249
320 480
384 488
564
64 89
128 169
192 249
326 480
384 480
448 569
448 569
                                                                                                                                                                                                                                                                   224
336
448
560
672
784
896
112
224
336
448
560
672
784
                                                                                                                                                                                                    89
169
249
329
409
489
569
649
329
499
489
560
649
                                                                                                                                                                                                           96 112
192 224
288 336
384 448
480 560
576 672
784
768 896:
96 112
192 224
288 336
384 448
480 560
576 672 784
Matrix Y (Original):

[[1 1 1 1 1 1 1 1 1]

[2 2 2 2 2 2 2 2 2]

[3 3 3 3 3 3 3 3]

[4 4 4 4 4 4 4 4]

[5 5 5 5 5 5 5 5]
[6 6 6 6 6 6 6 6]
[7 7 7 7 7 7 7 7 7]
[8 8 8 8 8 8 8 8 8]]
Matrix Y (Transposed):
     2 3 4 5 6 7 8]
 [1 2 3 4 5 6 7 8]
[1 2 3 4 5 6 7 8]
[1 2 3 4 5 6 7 8]
[1 2 3 4 5 6 7 8]
[1 2 3 4 5 6 7 8]
[1 2 3 4 5 6 7 8]
                        32 40 48 56 64]
64 80 96 112 128]
96 120 144 168 192]
           16
32
48
     8
                  24
    24
          64 96 128 160 192 224 256
80 120 160 200 240 280 320
    32
40
          96 144 192 240 288 336 384
112 168 224 280 336 392 448
    48
         128 192 256 320 384 448 512
```

N = 32

```
Console
   32 64 96 128 160 192 224 256 32 64 96 128 160 192 224 256 32 64 96 128 160 192 224 256 32 64 96 128 160 192 224 256 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 448 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 192 256 320 384 488 512 64 128 1
192 384 576 768 960 1152 1344 1536 192 384 576 768 960 1152 1344 1536 192 384 576 768 960 1152 1344 1536 192 384 576 768 960 1152 1344 1536 224 448 672 896 1120 1344 1568 1792 224 448 672 896 1120 1344 1568 1792 224 448 672 896 1120 1344 1568 1792 224 448 672 896 1120 1344 1568 1792 2256 512 768 1024 1280 1536 1792 2048 256 512 768 1024 1280 1536 1792 2048 256 512 768 1024 1280 1536 1792 2048 256 512 768 1024 1280 1536 1792 2048 256 512 768 1024 1280 1536 1792 2048 256 512 768 1024 1280 1536 1792 2048
                                                                                                                              Result of Matrix Multiplication (X * Y Transposed):
                                                                                                                                                                                                               96 ...
                                                                                                                                                     32
                                                                                                                                                                                  64
                                                                                                                                                                                                                                                             192
                                                                                                                                                                                                                                                                                           224
                                                                                                                                                                                                                                                                                                                        256]
                                                                                                                                                                                                          192 ...
                                                                                                                                                                                                                                                                                                                        512]
                                                                                                                                                                                                                                                              384
                                                                                                                                                                                                                                                                                           448
                                                                                                                                                      64
                                                                                                                                                                            128
                                                                                                                                                      96
                                                                                                                                                                             192
                                                                                                                                                                                                          288 ...
                                                                                                                                                                                                                                                             576
                                                                                                                                                                                                                                                                                           672
                                                                                                                                                                                                                                                                                                                        768]
                                                                                                                                                                                                         576 ... 1152 1344 1536]
                                                                                                                                               192
                                                                                                                                                                             384
                                                                                                                                                224
                                                                                                                                                                                                         672 ... 1344 1568 1792]
                                                                                                                                                                            448
                                                                                                                                               256
                                                                                                                                                                            512
                                                                                                                                                                                                        768 ... 1536 1792 2048]]
```

[MIPS-A]

[Cycle Statistics]

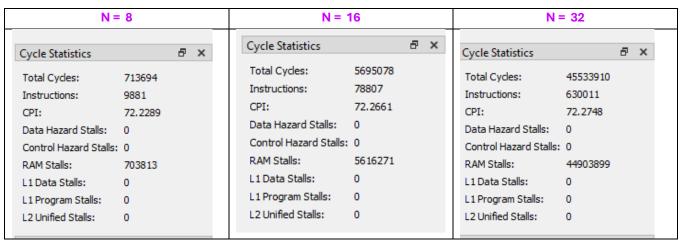
[Πολλαπλασιασμός διανύσματος επί βαθμωτού]

N =	8	N = 1	16	N = 3	32
Cycle Statistics		Cycle Statistics		Cycle Statistics	
Total Cycles: Instructions: CPI:	160 112 1.42857	Total Cycles: Instructions: CPI:	288 208 1.38462	Total Cycles: Instructions: CPI:	544 400 1.36
Data Hazard Stalls: Control Hazard Stalls:	42	Control Hazard Stalls:	_		138
RAM Stalls: L1 Data Stalls: L1 Program Stalls:	0 0 0	RAM Stalls: L1 Data Stalls: L1 Program Stalls:	0 0	RAM Stalls: L1 Data Stalls: L1 Program Stalls:	0 0 0
L2 Unified Stalls:	0	L2 Unified Stalls:	0	L2 Unified Stalls:	0

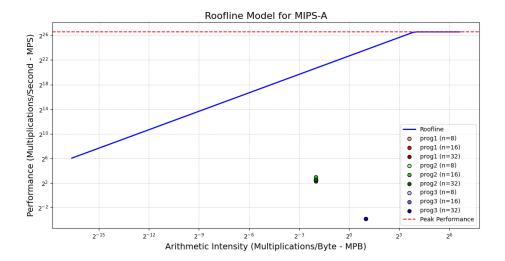
[Πολλαπλασιασμός πίνακα επί βαθμωτού]

N = 8	}	N = 10	6	N = 32	2
Cycle Statistics		Cycle Statistics		Cycle Statistics	
Total Cycles: Instructions: CPI: Data Hazard Stalls: Control Hazard Stalls: RAM Stalls: L1 Data Stalls: L1 Program Stalls: L2 Unified Stalls:	1147 837 1.37037 282 28 0 0	Total Cycles: Instructions: CPI: Data Hazard Stalls: Control Hazard Stalls: RAM Stalls: L1 Data Stalls: L1 Program Stalls: L2 Unified Stalls:	4307 3197 1.3472 1066 44 0 0 0	Total Cycles: Instructions: CPI: Data Hazard Stalls: Control Hazard Stalls: RAM Stalls: L1 Data Stalls: L1 Program Stalls: L2 Unified Stalls:	16771 12525 1.339 4170 76 0 0

[Πολλαπλασιασμός τετραγωνικών πινάκων]

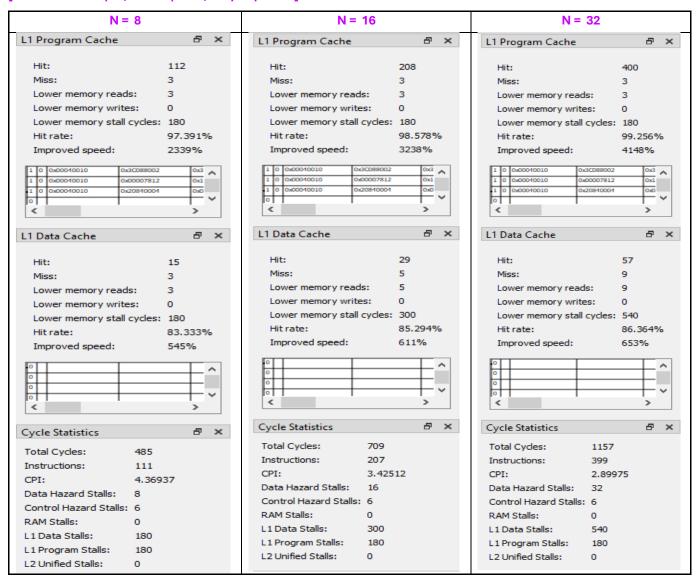


[Roofline Analysis]



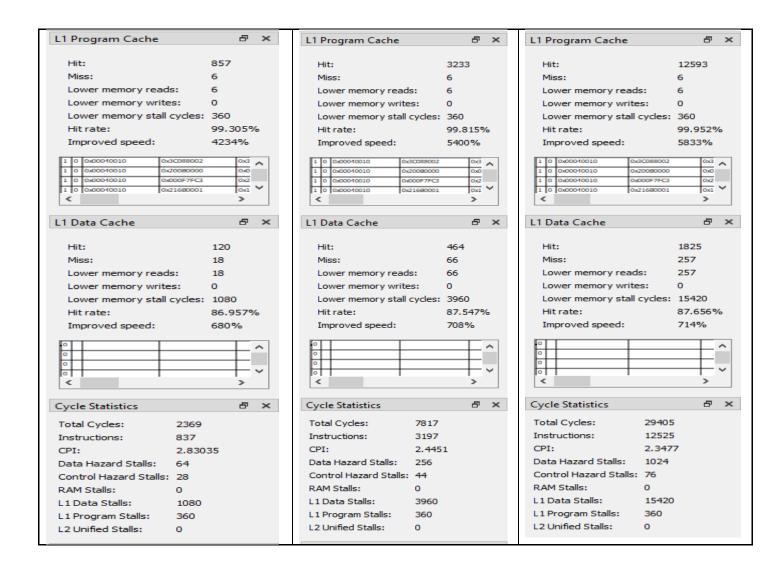
[MIPS-B]

[Πολλαπλασιασμός διανύσματος επί βαθμωτού]



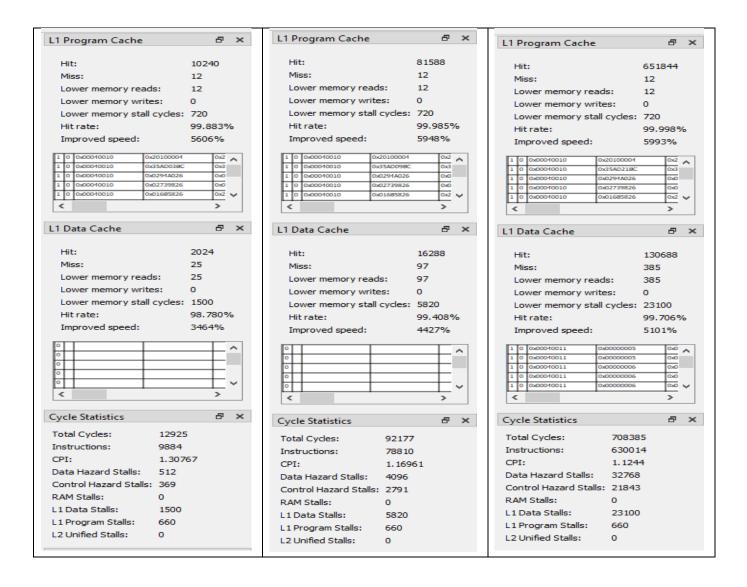
[Πολλαπλασιασμός πίνακα επί βαθμωτού]

N = 8	N = 16	N = 32
-------	--------	--------

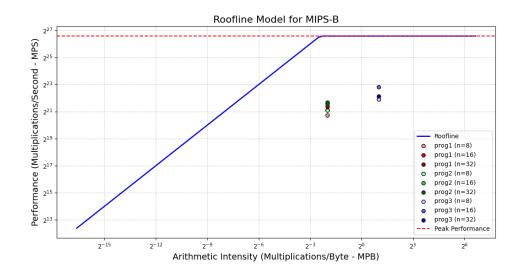


[Πολλαπλασιασμός τετραγωνικών πινάκων]

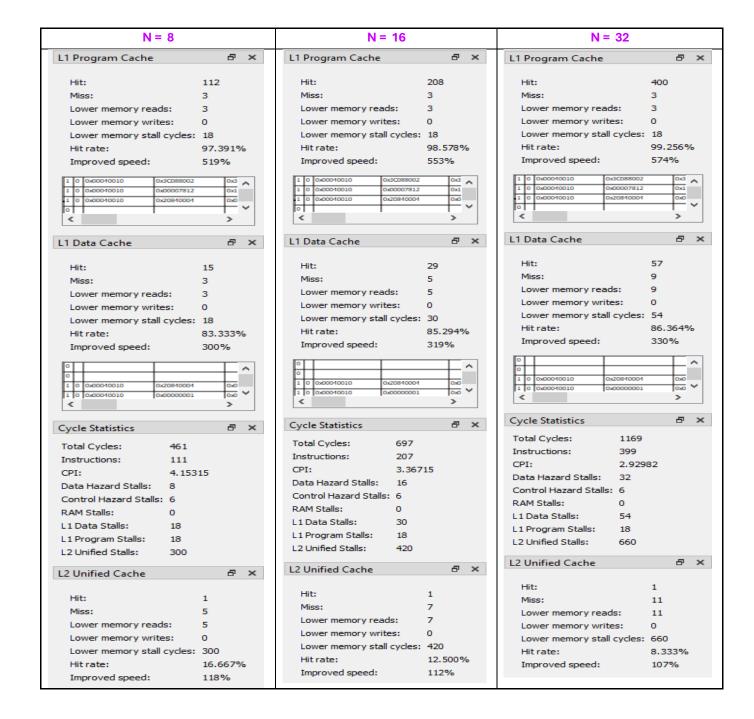
N = 8	N = 16	N = 32
-------	--------	--------



[Roofline Analysis]

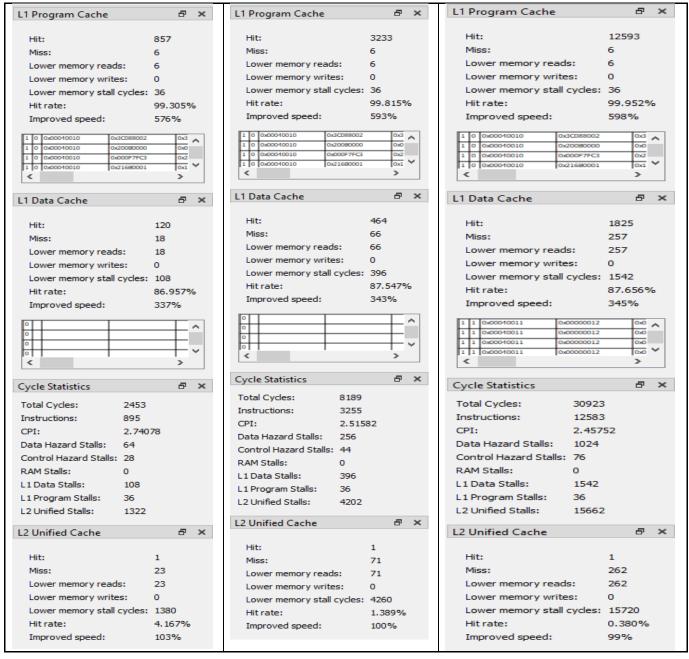


[MIPS-[]



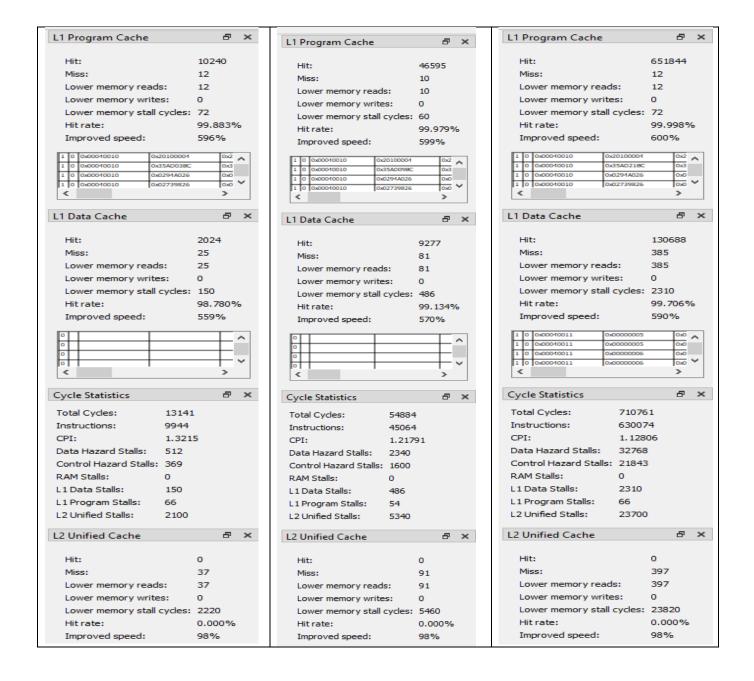
[Πολλαπλασιασμός πίνακα επί βαθμωτού]

N = 8	N = 16	N = 32
., ,	14 10	14 02

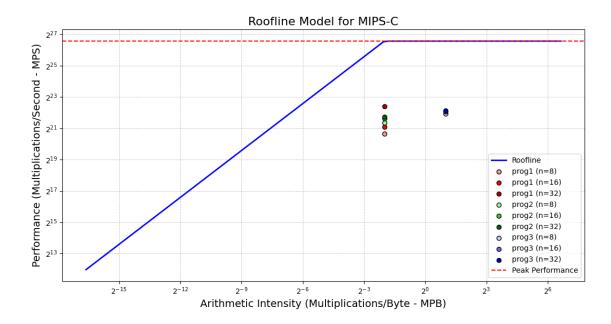


[Πολλαπλασιασμός τετραγωνικών πινάκων]

N = 8	N = 16	N = 32
-------	--------	--------



[Roofline Analysis]



[ΣΥΜΠΕΡΑΣΜΑ]

MIPS-A: Αυτός ο επεξεργαστής δεν έχει κρυφές μνήμες, οπότε κάθε φορά που χρειάζεται δεδομένα προσπελάζει την κύρια μνήμη, κάτι που είναι αργό. Αυτό προκαλεί μεγάλη καθυστέρηση και μειώνει την απόδοση, καθώς τα φορτία υπολογισμού καταλήγουν να είναι περιορισμένα από την ταχύτητα της μνήμης.

MIPS-B: Αυτός ο επεξεργαστής με μόνο μνήμη cache L1 κατάφερε να επιλύσει τις προσβάσεις στη μνήμη με ελάχιστη καθυστέρηση. Χωρίς την ανάγκη πρόσβασης σε μια μνήμη cache L2, ο επεξεργαστής ωφελήθηκε από την μειωμένη υπερφόρτωση του ελέγχου μιας δευτερεύουσας cache. Η αυξημένη επιτυχία της L1 cache με μεγαλύτερους πίνακες υποδεικνύει ότι η L1 cache αποθήκευε δεδομένα με αποδοτικό τρόπο, επιτρέποντας στον επεξεργαστή να εκτελεί περισσότερες εντολές ανά κύκλο.

MIPS-C: Αυτός ο επεξεργαστής με μνήμες cache L1 και L2 σε αυτή την περίπτωση αντιμετώπισε αναποτελεσματικότητες λόγω του εξαιρετικά υψηλού ρυθμού επιτυχίας της μνήμης L1 του προγράμματος. Δεδομένου ότι η συντριπτική πλειονότητα των προσπελάσεων μνήμης εξυπηρετήθηκε από την μνήμη cache L1, η μνήμη L2, η οποία συνήθως σχεδιάζεται για να χειρίζεται τις αποτυχίες της μνήμης L1, προσπελάστηκε σπάνια. Αυτή χωρίς να εκμεταλλεύεται, την μνήμη L2, ως αποτέλεσμα οδήγησε σε χάσιμο κύκλων που ξοδεύτηκαν για τον έλεγχο της μνήμης L2.

Όπως φαίνεται στον MIPS B, με μνήμη cache L1 μόνο και στο MIPS C με μνήμες cache L1 και L2 — παρατηρήθηκε ότι ο MIPS B ήταν ταχύτερος. Αυτό συμβαίνει επειδή το πρόγραμμα πέτυχε έναν εξαιρετικά υψηλό ρυθμό επιτυχίας στη μνήμη cache L1 (99.9%), διασφαλίζοντας ότι σχεδόν όλες οι προσπελάσεις μνήμης εξυπηρετήθηκαν απευθείας από τη μνήμη L1. Ω ς αποτέλεσμα, ο επεξεργαστής με μόνο L1 απέφυγε την επιπλέον καθυστέρηση που σχετίζεται με την πρόσβαση στη μνήμη L2.

Αντίθετα, ο MIPS C, υπέστη πρόσθετη καθυστέρηση (αν και ελάχιστη) λόγω περιττών αναζητήσεων στη μνήμη L2 μετά από αποτυχίες στη μνήμη L1, παρόλο που η μνήμη L2 δεν προσέφερε ουσιαστικό όφελος σε αυτό το σενάριο (καθώς συχνά αποτύγχανε). Αυτή η καθυστέρηση προκάλεσε την επιβράδυνση του επεξεργαστή με μνήμες L1 και L2.

Συμπέρασμα:

Ο MIPS Β παρουσίασε ταχύτερη απόδοση επειδή ο εξαιρετικά υψηλός ρυθμός επιτυχίας της μνήμης L1 εξάλειψε την ανάγκη για ελέγχους σε βαθύτερα επίπεδα ιεραρχίας μνήμης, ενώ η πρόσθετη καθυστέρηση από τις αναζητήσεις στη μνήμη L2 στον άλλο επεξεργαστή μείωσε τη συνολική αποτελεσματικότητα. Τέλος αυτή η διάταξη αναδεικνύει το πλεονέκτημα της ύπαρξης μιας πολύ αποδοτικής L1 cache, ιδιαίτερα για φορτία εργασίας όπου το σύνολο των δεδομένων είναι μικρό ή καλά τοποθετημένο. Αποφεύγοντας τις πολυπλοκότητες και την καθυστέρηση της πρόσβασης σε επιπλέον επίπεδα cache, ο επεξεργαστής κατάφερε να διατηρήσει υψηλό επίπεδο απόδοσης.