

Τεχνολογίες Κύριας Μνήμης (και η ανάγκη για χρήση ιεραρχιών μνήμης)

<http://mixstef.github.io/courses/comparch/>



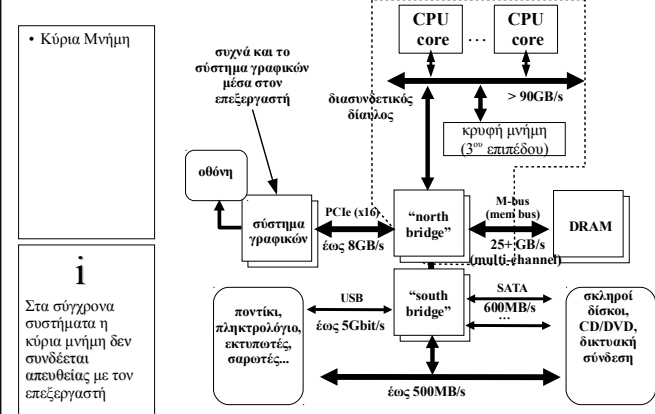
Μ.Στεφανιδάκης

Τεχνολογίες Κύριας Μνήμης

• Κύρια Μνήμη

- Στους πρώτους υπολογιστές
 - Ιστορικά, η κατασκευή κύριας μνήμης ήταν πολύ πιο δύσκολη από την κατασκευή των πρώτων υπολογιστών!
- Αρχικές τεχνολογίες
 - Flip-Flop με λυχνίες κενού
 - Γραμμές καθυστέρησης υδραργύρου
 - Μαγνητικές μνήμες (core memories - 1950)
 - Η πρώτη αξιόπιστη και σχετικά φθηνή τεχνολογία
 - Κυριάρχησε για 20 περίπου χρόνια
- Ημιαγωγικές μνήμες (Intel – 1970)
 - 1Kbit DRAM - “core killer”

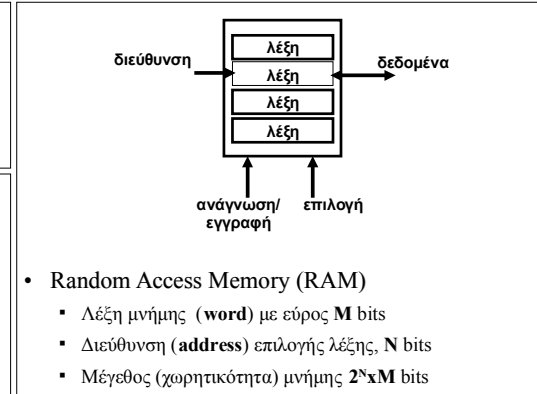
Κύρια Μνήμη



Το μοντέλο της Μνήμης Τυχαίας Προσπέλασης

• Κύρια Μνήμη • RAM

• Η λέξη είναι η μικρότερη προσπελάσιμη ομάδα bits.
• Το εύρος των μεταφερόμενων δεδομένων σε κάθε ανάγνωση ή εγγραφή ισούται με το εύρος της λέξης μνήμης



• Random Access Memory (RAM)

- Λέξη μνήμης (word) με εύρος M bits
- Διεύθυνση (address) επιλογής λέξης, N bits
- Μέγεθος (χωρητικότητα) μνήμης $2^N \times M$ bits

Διευθυνσιοδότηση μνήμης RAM

- Κύρια Μνήμη
- RAM

0x80154FF0	byte	byte	byte	byte	Λέξη μνήμης
0x80154FF4	byte	byte	byte	byte	
0x80154FF8	byte	byte	byte	byte	

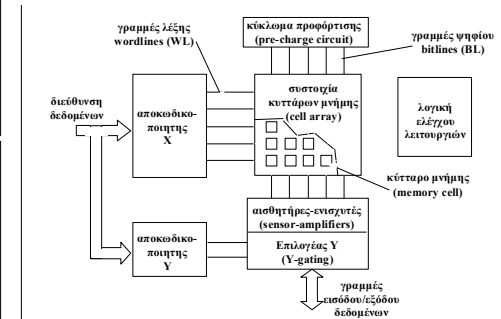
- Byte addressing
 - Οι διευθύνσεις μνήμης που παράγει ο επεξεργαστής αυξάνονται ανά byte
 - Ακόμα κι όταν η λέξη μνήμης έχει πολλαπλάσιο εύρος!
- Εναλλακτικά: word addressing
 - Υπερυπολογιστές – εδώ η προσπέλαση ανά byte είναι σπάνια

Πώς προσπελάζω bytes όταν η λέξη μνήμης έχει πολλαπλάσιο εύρος;

Οργάνωση Μνήμης Τυχαίας Προσπέλασης

- Κύρια Μνήμη
- RAM

i
Οι μεγαλύτερες μνήμες RAM διαθέτουν πολλαπλές συστοιχίες κυττάρων μνήμης



Ταχύτητα Προσπέλασης RAM

- Κύρια Μνήμη
- RAM

- Access Time (χρόνος προσπέλασης)
 - Ο απαιτούμενος χρόνος για την ολοκλήρωση μιας αίτησης προς τη μνήμη RAM
 - Ανάγνωση - Εγγραφή
- Cycle Time (χρόνος κύκλου προσπέλασης)
 - Ο ελάχιστος απαιτούμενος χρόνος μεταξύ διαδοχικών αιτήσεων προς τη μνήμη RAM
 - Πρόβλεψη ενδιάμεσων λειτουργιών
- Ταχύτητα διαφορετική για ανάγνωση - εγγραφή!

Τύποι Μνήμης Τυχαίας Προσπέλασης

- Κύρια Μνήμη
- RAM
- SRAM

Ο χρόνος προσπέλασης μιας μνήμης SRAM βρίσκεται μεταξύ 0,5 και 5 ns

- Στατική Μνήμη RAM (SRAM)
 - Κάθε bit αποθηκεύεται σε κύτταρο (“cell”) 6 τρανζίστορ
 - Ανάλογο ενός flip-flop
 - Διατήρηση όσο υπάρχει τροφοδοσία της μνήμης
- Η προσπέλαση είναι γρήγορη αλλά:
 - Μεγαλύτερο κόστος
 - Πολυπλοκότερο κύκλωμα
 - Δεν επιτρέπει μεγάλη ολοκλήρωση
 - Μεγαλύτερη κατανάλωση ενέργειας
- Χρησιμοποιείται στις κρυφές μνήμες (caches)

Τύποι Μνήμης Τυχαίας Προσπέλασης

- Κύρια Μνήμη
- RAM
- SRAM
- DRAM

Ο χρόνος προσπέλασης μιας μνήμης DRAM βρίσκεται μεταξύ 50 και 70 ns

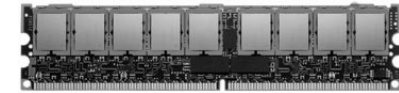
- **Δυναμική Μνήμη RAM (DRAM)**
 - Κάθε bit αποθηκεύεται ως φορτίο
 - Διατήρηση μόνο με συχνή ανανέωση του φορτίου
 - Κάθε 16 έως 128 ms (5% συνολικού χρόνου)
- **Απλούστερο κύκλωμα – μεγάλη ολοκλήρωση**
 - Πολύ μεγάλες χωρητικότητες (1 Gbit/chip και πλέον)
 - Η προσπέλαση είναι αργή
 - Αρχιτεκτονικές βελτιώσεις για αύξηση ρυθμού μεταφοράς δεδομένων
- Χρησιμοποιείται για τη συγκρότηση της κύριας μνήμης όλων των σύγχρονων υπολογιστικών συστημάτων
 - Μνήμη = ασύγχρονη λειτουργία αλλά: προσθήκη ρολογιού για διασύνδεση με το υπόλοιπο σύστημα

Αρχιτεκτονική Υπολογιστών – “Τεχνολογίες Κύριας Μνήμης”

9

Τμήματα (modules) μνήμης DRAM

- Κύρια Μνήμη
- RAM
- SRAM
- DRAM



- 64 - 72 bits δεδομένων (χωρητικότητα έως 4GB)
- Μεταφορά δεδομένων στις 2 ακμές ρολογιού
 - Double Data Rate (DDR) RAM
- Σήματα ανίχνευσης και αναγνώρισης
- Διατάξεις βελτίωσης ηλεκτρικών χαρακτηριστικών σημάτων
- Ρυθμός μεταφοράς > 8.5GB/s

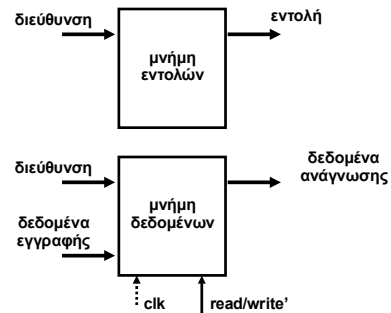
Αρχιτεκτονική Υπολογιστών – “Τεχνολογίες Κύριας Μνήμης”

10

Η “ιδανική μνήμη”

- Κύρια Μνήμη
- RAM
- SRAM
- DRAM
- Ιεραρχίες Μνήμης

Πόσο απέχει η ιδανική εικόνα από την πραγματικότητα;



- Ολοκλήρωση ανάγνωσης-εγγραφής σε έναν κύκλο ρολογιού...

Αρχιτεκτονική Υπολογιστών – “Τεχνολογίες Κύριας Μνήμης”

11

Η πραγματική εικόνα

- Κύρια Μνήμη
- RAM
- SRAM
- DRAM
- Ιεραρχίες Μνήμης

Η ιδανική μνήμη είναι πρακτικά αδύνατο να υλοποιηθεί. Ποια η πιθανή λύση;

- Ένας σύγχρονος επεξεργαστικός πυρήνας
 - με ρολόι 3 GHz
 - και έναρξη εκτέλεσης έως και 8 εντολών ανά κύκλο
 - απαιτεί από τη μνήμη 24G εντολές/sec!
- Η “ιδανική μνήμη” θα έπρεπε να είναι
 - Πολύ γρήγορη
 - Πολύ φθηνή
 - Με πολύ μεγάλη χωρητικότητα

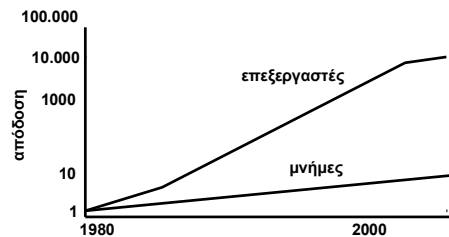
Αρχιτεκτονική Υπολογιστών – “Τεχνολογίες Κύριας Μνήμης”

12

Το χάσμα απόδοσης μεταξύ επεξεργαστή-μνήμης

- Κύρια Μνήμη
- RAM
- SRAM
- DRAM
- Ιεραρχίες Μνήμης

Οι μνήμες ακολουθούν τον νόμο του Moore στην αύξηση της χωρητικότητάς τους, όχι όμως και στην απόδοση



- Επεξεργαστές: αύξηση απόδοσης 35%-55% /έτος
 - Μνήμες: αύξηση απόδοσης 7% /έτος
- [Patterson-Hennessy]

Η αρχή της τοπικότητας

- Κύρια Μνήμη
- RAM
- SRAM
- DRAM
- Ιεραρχίες Μνήμης

“Ένα πρόγραμμα εκτελεί το 90% των εντολών του μέσα στο 10% του κώδικά του”

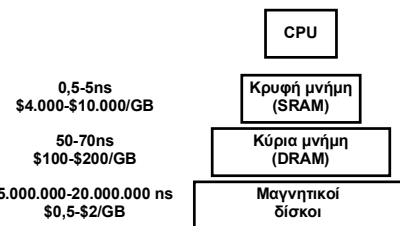
- Χρονική Τοπικότητα
 - Εάν προσπελαστεί μια θέση μνήμης, είναι πολύ πιθανό να προσπελαστεί ξανά στο άμεσο μέλλον
 - Π.χ. για εντολές ενός βρόχου (loop)
- Χωρική Τοπικότητα
 - Εάν προσπελαστεί μια θέση μνήμης, είναι πολύ πιθανό να προσπελαστούν και οι γειτονικές θέσεις στο άμεσο μέλλον
 - Εντολές προγραμμάτων
 - Δεδομένα σε πίνακες κλπ

Ιεραρχίες Μνήμης

- Κύρια Μνήμη
- RAM
- SRAM
- DRAM
- Ιεραρχίες Μνήμης

Και οι δικτυακές τοποθεσίες μπορούν να θεωρηθούν μέρος της ιεραρχίας μνήμης (το χαμηλότερο)

- Πολλαπλά επίπεδα μνήμης
 - Διαφορετικής τεχνολογίας
 - Με διαφορετική ταχύτητα και μέγεθος
 - Γρηγορότερη μνήμη κοντά στον επεξεργαστή

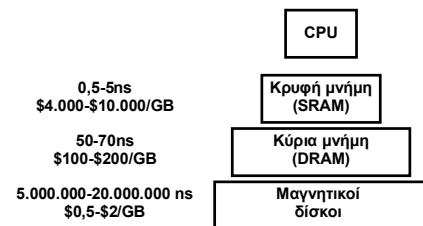


Σκοπός της Ιεραρχίας Μνήμης

- Κύρια Μνήμη
- RAM
- SRAM
- DRAM
- Ιεραρχίες Μνήμης

Για να επιτύχει τον σκοπό της η ιεραρχία μνήμης εκμεταλλεύεται την αρχή της τοπικότητας

- Προσέγγιση της ιδανικής μνήμης
 - Ο επεξεργαστής να βλέπει “μνήμη”
 - Με την ταχύτητα του υψηλότερου επιπέδου
 - Και το μέγεθος του χαμηλότερου



Αποθήκευση δεδομένων στην Ιεραρχία Μνήμης

- Κύρια Μνήμη
- RAM
- SRAM
- DRAM
- Ιεραρχίες Μνήμης

- **Αποθήκευση δεδομένων**
 - Τα υψηλότερα επίπεδα είναι υποσύνολα των χαμηλότερων
 - Όλα τα δεδομένα αποθηκεύονται τελικά στο χαμηλότερο επίπεδο
- **Μεταφορά δεδομένων**
 - Αντιγραφή από επίπεδο σε επίπεδο
 - Το ελάχιστο σύνολο δεδομένων που μεταφέρεται μεταξύ δύο επιπέδων ονομάζεται μπλοκ
 - Πολλαπλά bytes

Αναζήτηση δεδομένων στην Ιεραρχία Μνήμης

- Κύρια Μνήμη
- RAM
- SRAM
- DRAM
- Ιεραρχίες Μνήμης

- **Αναζήτηση δεδομένων**
 - Ο επεξεργαστής ζητά πάντοτε τα δεδομένα από το κοντινότερο σε αυτόν επίπεδο
 - Τα δεδομένα υπάρχουν στο επίπεδο αυτό: **hit**
 - Τα δεδομένα δεν βρίσκονται στο επίπεδο αυτό: **miss**
 - Η αίτηση προωθείται στο επόμενο (χαμηλότερο) επίπεδο
 - Και το μπλοκ που περιέχει τα δεδομένα αντιγράφεται στο ανώτερο επίπεδο

Μετρήσεις απόδοσης στην Ιεραρχία Μνήμης

- Κύρια Μνήμη
- RAM
- SRAM
- DRAM
- Ιεραρχίες Μνήμης

- **Hit Rate**
 - Ποσοστό προσπελάσεων μνήμης, όπου τα δεδομένα βρίσκονται στο ανώτερο επίπεδο
- **Miss Rate**
 - Ποσοστό προσπελάσεων μνήμης, όπου τα δεδομένα δεν βρίσκονται στο ανώτερο επίπεδο
 - (1-hit rate)
- **Hit Time**
 - Ο χρόνος για την προσπέλαση δεδομένων σε hit
- **Miss Penalty**
 - Ο χρόνος για την προσπέλαση, μεταφορά και τοποθέτηση των δεδομένων miss από το χαμηλότερο στο ανώτερο επίπεδο

Εισαγωγή στις κρυφές μνήμες (caches)

- Κύρια Μνήμη
- RAM
- SRAM
- DRAM
- Ιεραρχίες Μνήμης
- Κρυφές Μνήμες

- **Κρυφή μνήμη**
 - Μεταξύ του επεξεργαστή και της κύριας μνήμης
 - Εμφάνιση στη δεκαετία του 60
 - Σήμερα δεν υπάρχει υπολογιστικό σύστημα χωρίς κρυφή μνήμη
- **Αποθήκευση δεδομένων στην κρυφή μνήμη**
 - Όχι ανά λέξη μνήμης ή ανά byte...
 - ...αλλά ανά μπλοκ (64-512bits)
 - Μεταφορά δεδομένων από την κύρια προς την κρυφή μνήμη σε ριπές (bursts)
 - Το σύστημα κύριας μνήμης έχει βελτιστοποιηθεί αρχιτεκτονικά για αυτού του τύπου τις μεταφορές

Θέματα κρυφών μνημών

- Κύρια Μνήμη
- RAM
- SRAM
- DRAM
- Ιεραρχίες Μνήμης
- Κρυφές Μνήμες

- Πού αποθηκεύεται ένα μπλοκ στην κρυφή μνήμη;
- Πώς εντοπίζεται ένα μπλοκ στην κρυφή μνήμη;
- Ποιο μπλοκ θα αντικατασταθεί όταν χρειαστεί;
- Τι συμβαίνει στην εγγραφή νέων δεδομένων;
- Πώς υπολογίζεται η απόδοση της ιεραρχίας μνήμης;

(στο επόμενο μάθημα..)