

ΣΥΣΤΗΜΑ ΜΝΗΜΗΣ Η/Υ

Μονάδες μνήμης Η/Υ



Hard Disk



RAM



ROM



CD/DVD



Floppy



Memory Card



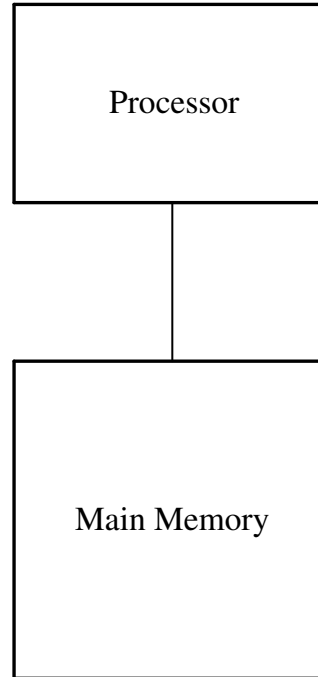
Pen Drive



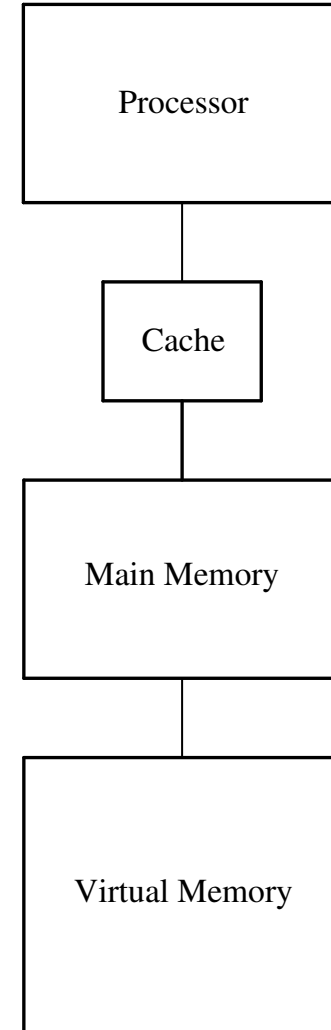
Tape

Ιεραρχία του συστήματος μνήμης

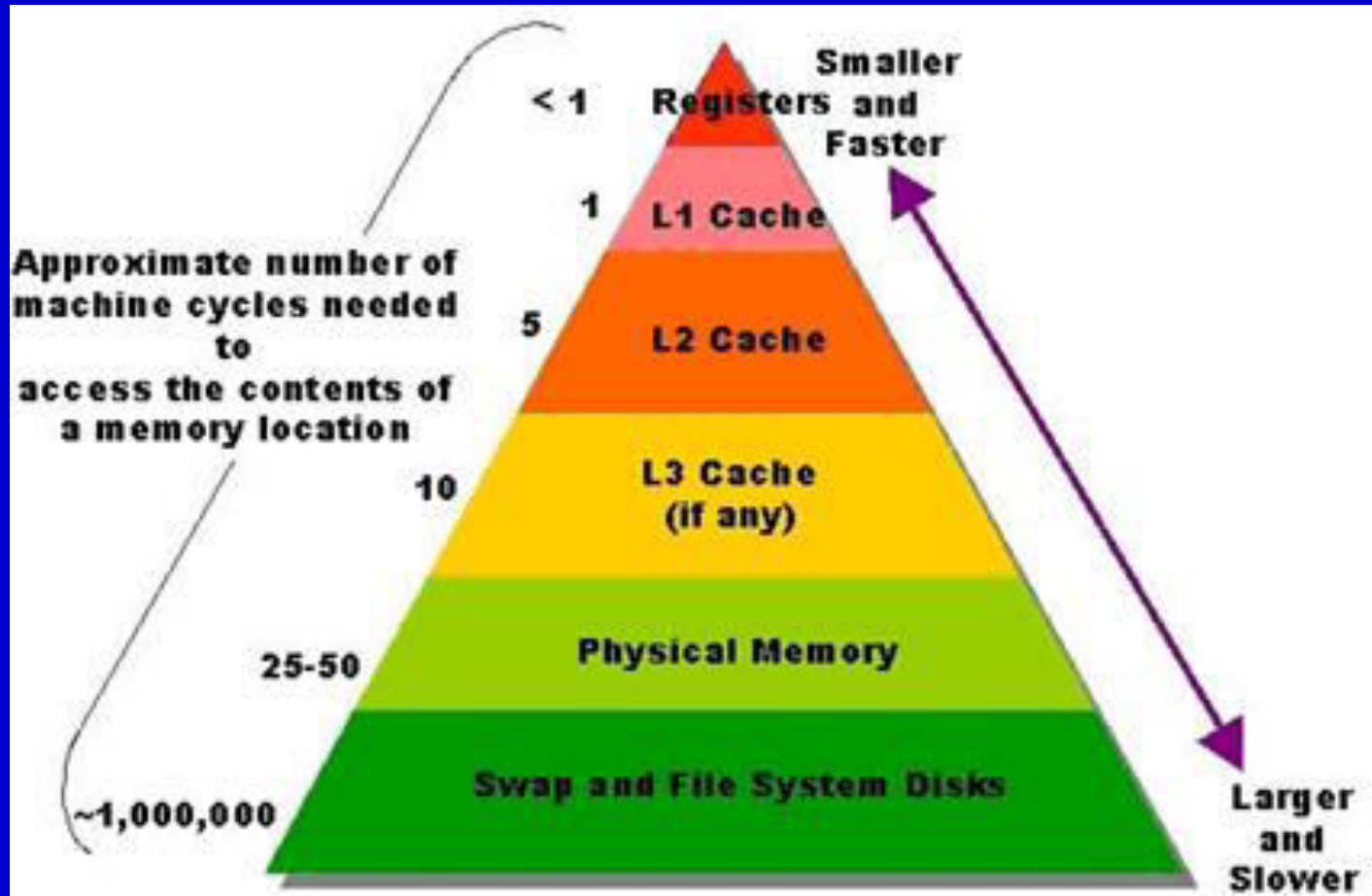
Single-Level
Memory System



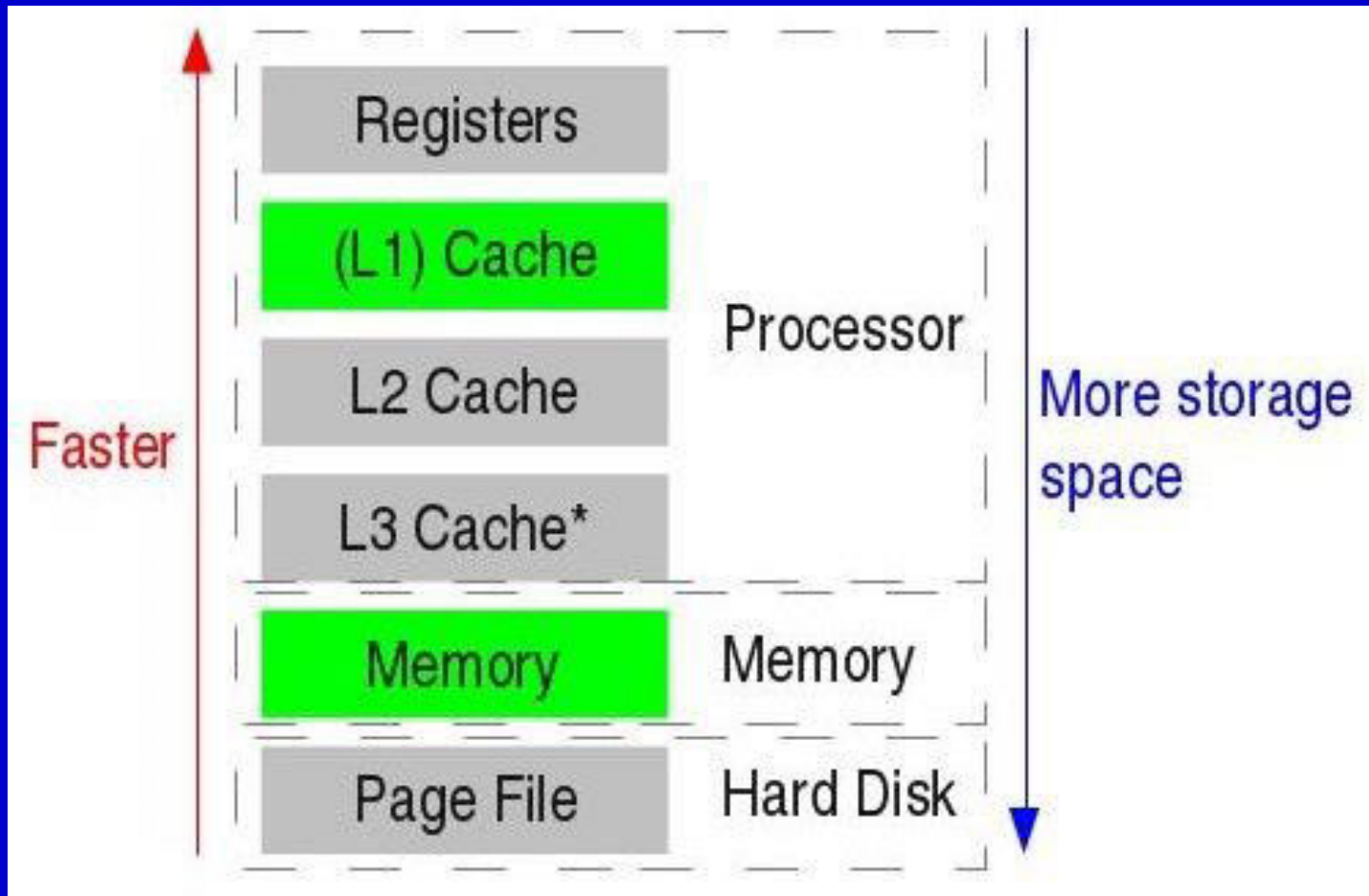
Multilevel
Memory Hierarchy



Memory Hierarchy



Memory Hierarchy

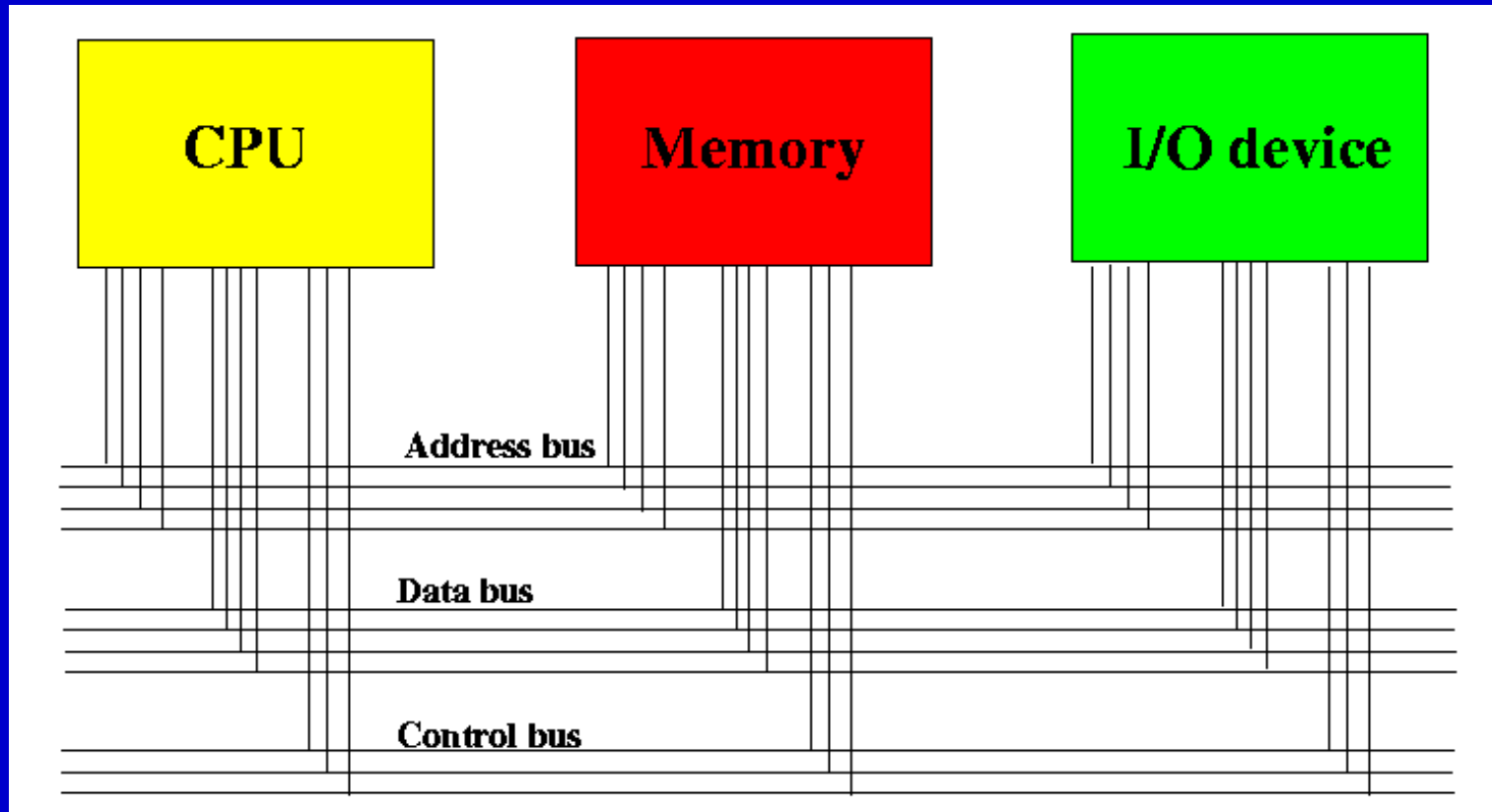


Κύρια Μνήμη

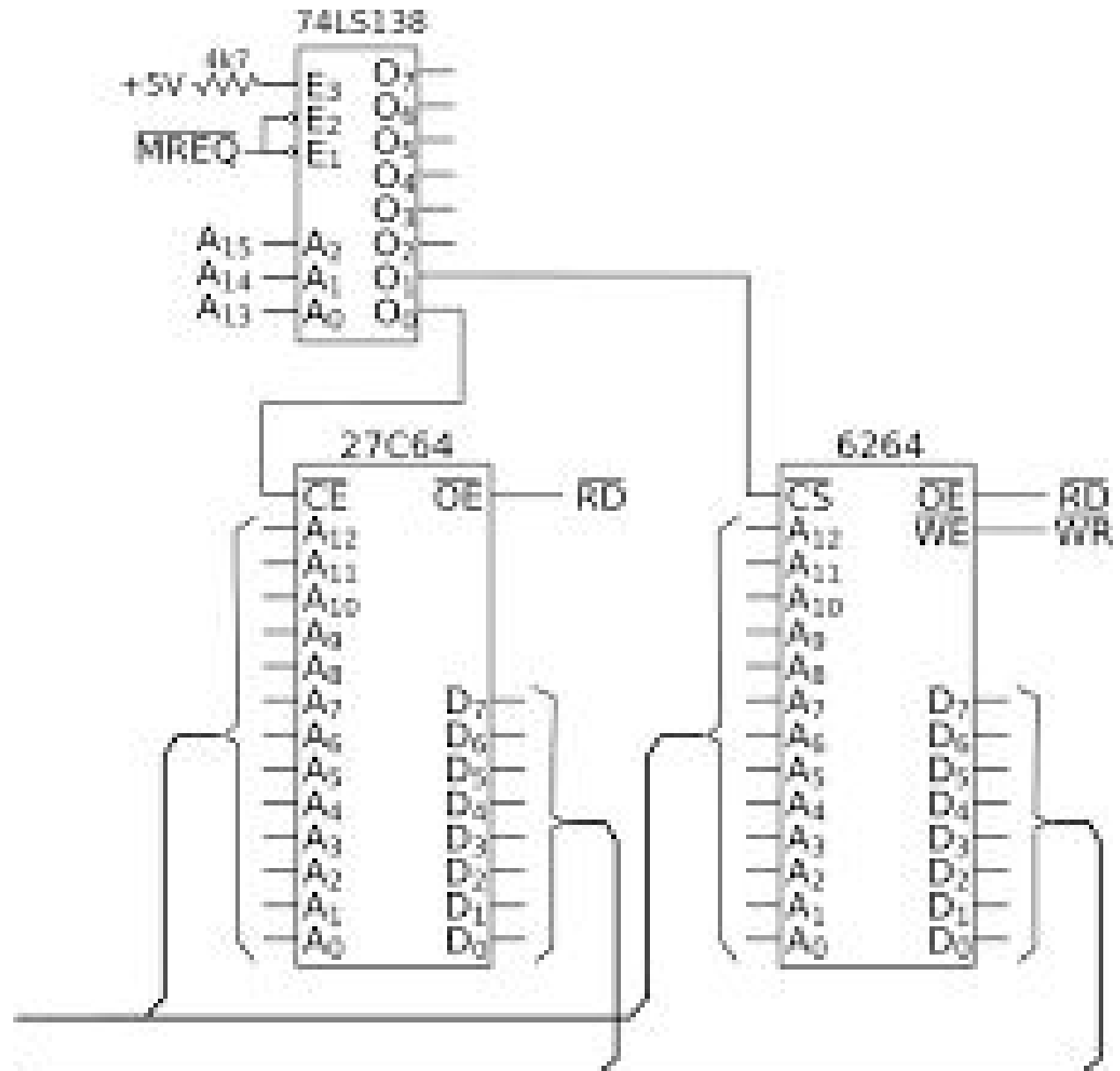
Η κύρια μνήμη (*main memory*) των Η/Υ μπορεί να θεωρηθεί ότι αποτελείται από διαδοχικές θέσεις στις οποίες αποθηκεύονται πληροφορίες (εντολές, δεδομένα ή τμήματα αυτών) οι οποίες είναι συνδυασμοί δυαδικών ψηφίων. Η κάθε θέση της κυρίας μνήμης προσδιορίζεται από έναν αριθμό που καλείται *διεύθυνση* (*address*) αυτής. Εάν η μνήμη έχει N θέσεις τότε οι διευθύνσεις της αριθμούν από 0 έως $N-1$.

Οι πληροφορίες που βρίσκονται σε κάποια θέση της κύριας μνήμης μπορούν να προσπελασθούν άμεσα από την CPU.

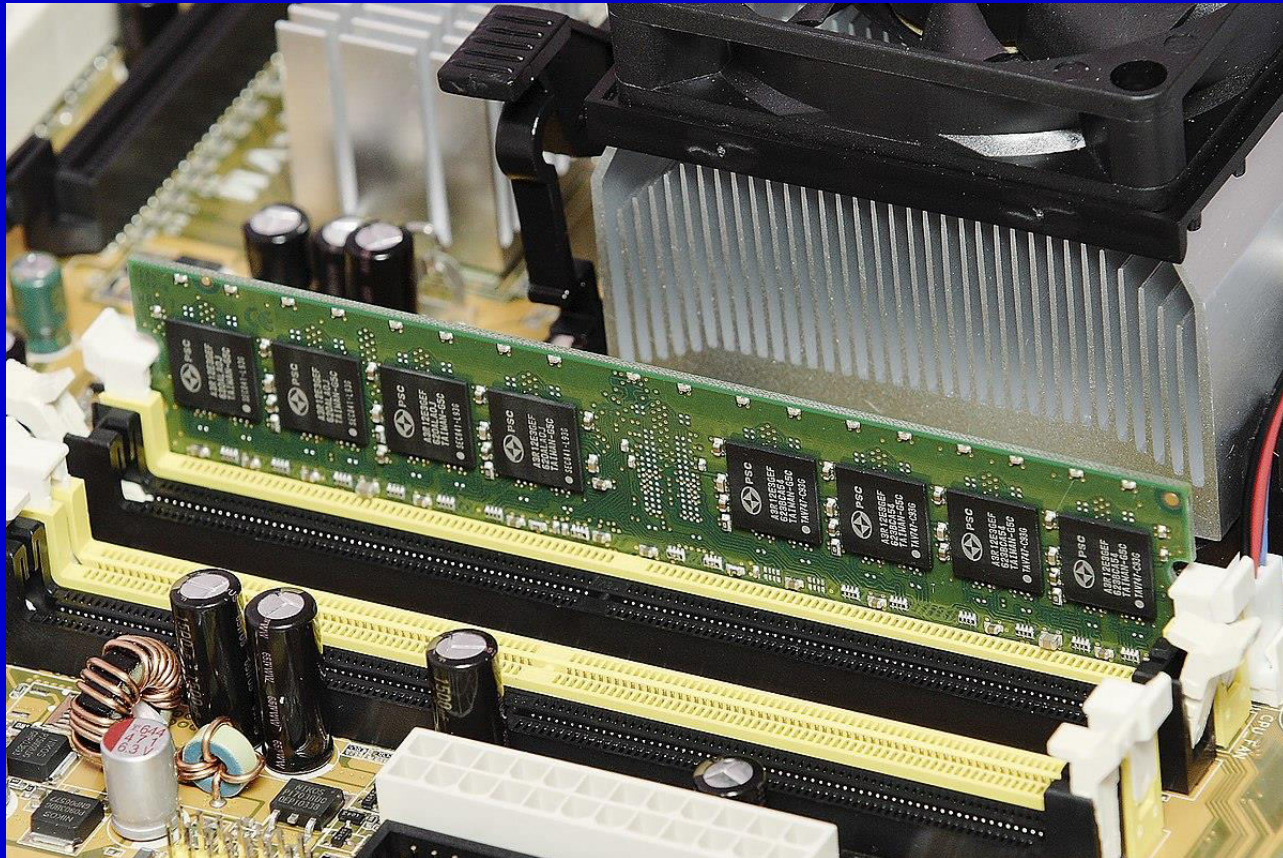
Σύνδεση CPU και μνήμης μέσω διαύλων



Σύστημα κύριας
μνήμης με RAM
και ROM



Main memory slots of a modern PC



Χρόνος προσπέλασης, Χρόνος κύκλου κύριας μνήμης

Ο *χρόνος προσπέλασης* (*access time or latency*) κύριας μνήμης ισούται με τον χρόνο που παρέρχεται από την έναρξη της διαδικασίας ανάγνωσης μιας θέσης μνήμης μέχρι την στιγμή που η απαιτούμενη πληροφορία είναι διαθέσιμη στις γραμμές δεδομένων της μνήμης.

Ορισμένες κατηγορίες μνημών απαιτούν να διέλθει ένα χρονικό διάστημα μεταξύ διαδοχικών προσπελάσεων. Αυτός ο χρόνος χρειάζεται για να προφορτιστούν τα κυκλώματά τους για την επόμενη προσπέλαση.

Ο *χρόνος κύκλου* (*cycle time*) ισούται με τον χρόνο που πρέπει να περάσει από την στιγμή που θα ξεκινήσει μια προσπέλαση μέχρι την στιγμή που επιτρέπεται να αρχίσει μια άλλη προσπέλαση. Όταν για την προσπέλαση μιας θέσης μνήμης απαιτείται *προφόρτιση* (*precharging*) ο χρόνος κύκλου είναι ίσος με το άθροισμα του χρόνου προσπέλασης και του χρόνου προφόρτισης.

Ρυθμός μεταφοράς, **bandwidth**

Ο ρυθμός μεταφοράς (*transfer rate*) είναι ο ρυθμός με τον οποίο τα δεδομένα μεταφέρονται εντός ή εκτός της κύριας μνήμης και είναι το αντίστροφο του χρόνου κύκλου.

Bandwidth είναι το γινόμενο του ρυθμού μεταφοράς επί την ποσότητα των δεδομένων που μεταφέρονται σε κάθε προσπέλαση μνήμης.

Memory interleaving

Η τεχνική *memory interleaving* χρησιμοποιείται σε ένα σύστημα κύριας μνήμης για την αύξηση της ταχύτητας επικοινωνίας με την CPU για διαδοχικές προσπελάσεις. Οι διαδοχικές θέσεις μνήμης δεν βρίσκονται στο ίδιο module μνήμης, αλλά σε διαφορετικά. Με την τεχνική αυτή όταν γίνεται προσπέλαση στο ένα ολοκληρωμένο γίνεται προετοιμασία για την προσπέλαση στο άλλο ώστε αυτή να μπορεί να ολοκληρωθεί γρηγορότερα.

SDRAM

Κατά την εκτέλεση των προγραμμάτων από τους Η/Υ οι προσπελάσεις στην κύρια μνήμη γίνονται στην πλειοψηφία τους σε διαδοχικές θέσεις (burst). Με βάση αυτή την παρατήρηση σχεδιάστηκαν οι **Synchronous DRAM (SDRAM)** που χρησιμοποιούνται για την υλοποίηση της κύριας μνήμης των Η/Υ.

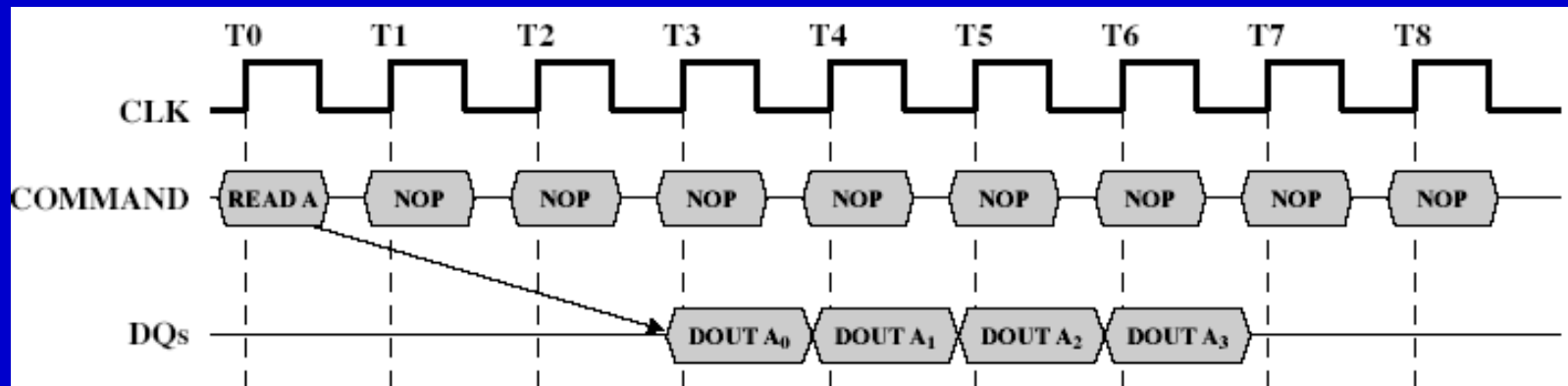
Η SDRAM είναι σύστημα κύριας μνήμης που βασίζεται σε DRAM και σχεδιάσθηκε ώστε να στέλνει δέσμες δεδομένων σε υψηλές ταχύτητες χρησιμοποιώντας αυτόματα διευθυνσιοδότηση, εσωτερικό interleaving, pipelining και ένα σύγχρονο (ή χρονιζόμενο) σύστημα διασύνδεσης.

SDRAM

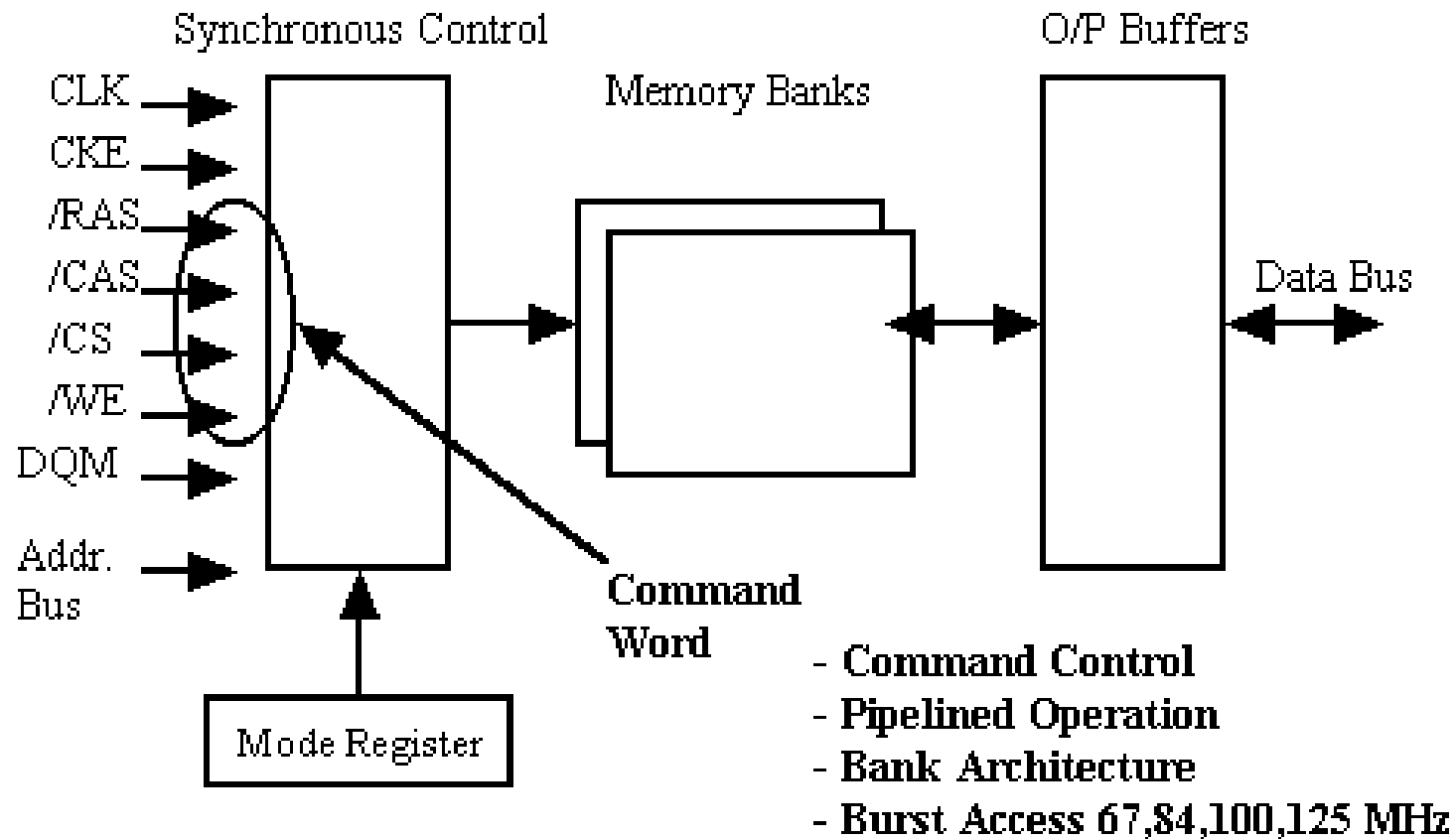
Στις SDRAM ένας ενσωματωμένος burst counter επιτρέπει οι διευθύνσεις διαδοχικών γραμμών δεδομένων να παράγονται ταχύτατα.

Ο ελεγκτής μνήμης του υπολογιστή παράγει την θέση και το μέγεθος του block δεδομένων που πρέπει να μεταφερθεί και τα δεδομένα μεταφέρονται μέσω του διαύλου διαδοχικά και σε συγχρονισμό με το clock του διαύλου (synchronous).

Χρονισμός ανάγνωσης από SDRAM



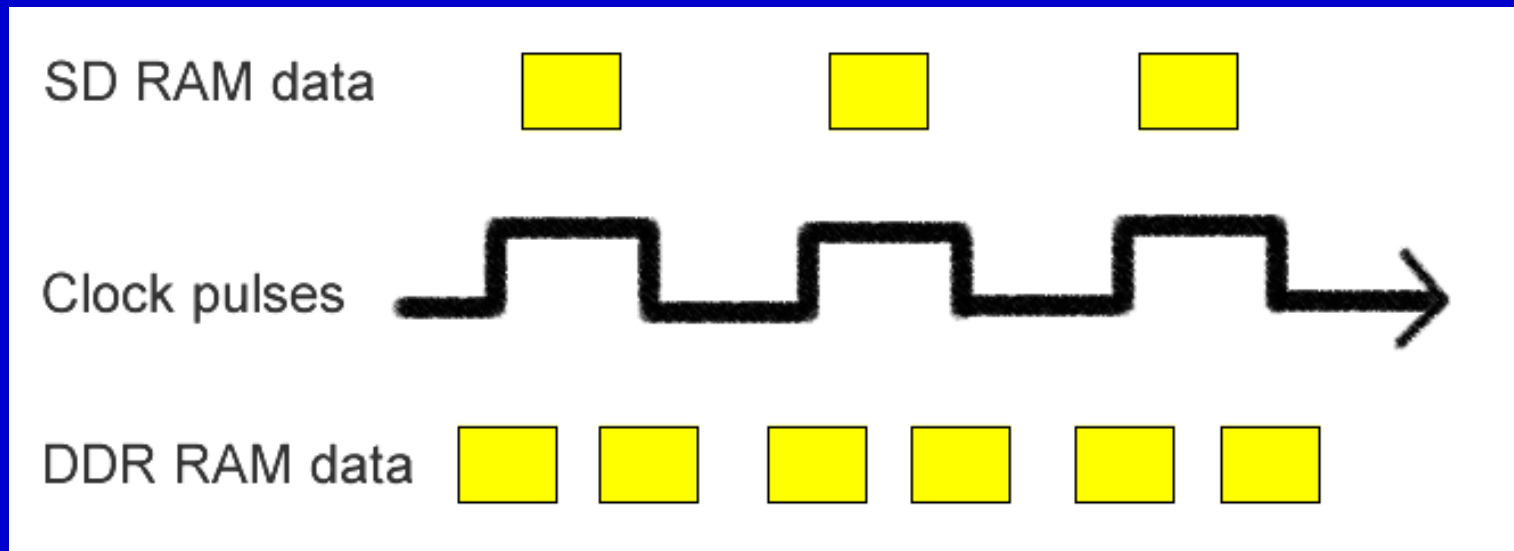
SDRAM Block Diagram



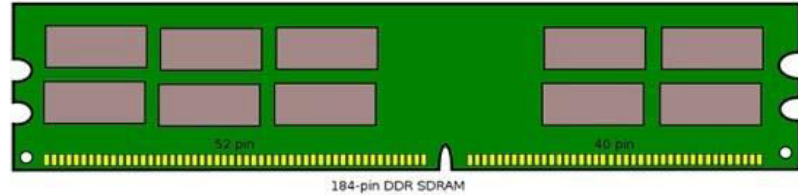
DDR SDRAM

Οι DDR SDRAM (Double Data Rate SDRAM) είναι πιο εξελιγμένοι τύποι SDRAM. Στην συνέχεια δίδεται ο χρονισμός προσπέλασης ανάγνωσης ή εγγραφής των SDRAM και των DDR SDRAM. Στις SDRAM γίνεται μία προσπέλαση στην μνήμη σε κάθε παλμό του ωρολογίου, ενώ στις DDR SDRAM διπλή προσπέλαση σε κάθε παλμό ωρολογίου, μία στην άνοδο και μία στην κάθοδο του σήματος.

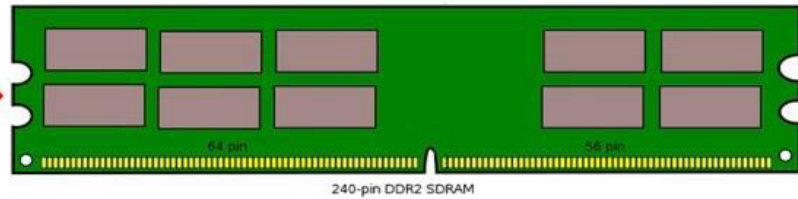
Χρονισμός των SDRAM και των DDR SDRAM



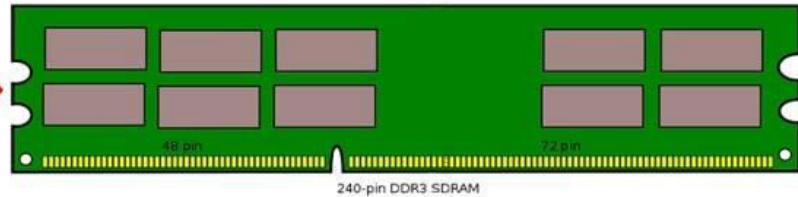
DDR



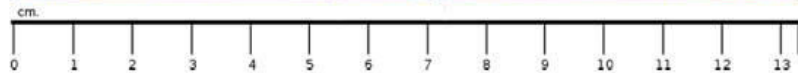
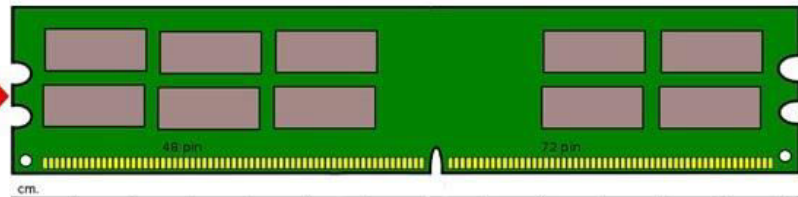
DDR2



DDR3



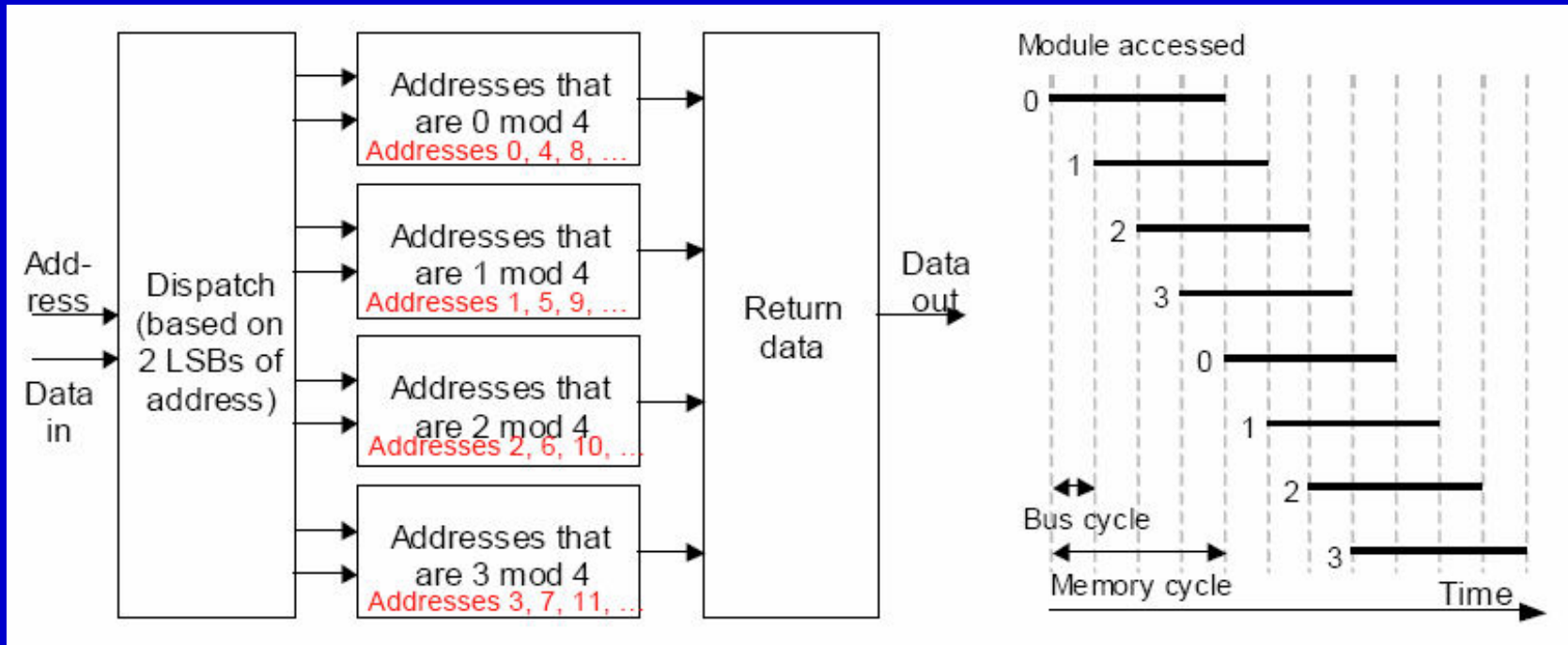
DDR4



Τεχνικά χαρακτηριστικά μνημών SDRAM

DDR SDRAM Standard	Internal rate (MHz)	Bus clock (MHz)	<u>Prefetch</u>	Data rate (MT/s)	Transfer rate (GB/s)	Voltage (V)
SDRAM	100-166	100-166	1n	100-166	0.8-1.3	3.3
DDR	133-200	133-200	2n	266-400	2.1-3.2	2.5/2.6
DDR2	133-200	266-400	4n	533-800	4.2-6.4	1.8
DDR3	133-200	533-800	8n	1066-1600	8.5-14.9	1.35/1.5
DDR4	133-200	1066-1600	8n	2133-3200	17-21.3	1.2

Memory interleaving



Wide-access memory

Η τεχνική *wide-access memory* χρησιμοποιείται για την αύξηση της ταχύτητας επικοινωνίας μεταξύ της CPU και της κύριας μνήμης. Με την τεχνική αυτή χρησιμοποιείται κύρια μνήμη με λέξεις πολύ μεγάλου μήκους έτσι ώστε σε κάθε προσπέλαση να ανακαλείται από την κύρια μνήμη μια μεγάλη ποσότητα δεδομένων. Για τη επιλογή και την μεταφορά στον επεξεργαστή της κατάλληλης λέξης χρησιμοποιείται πολυπλέκτης.

**Sockets motherboard με wide memory access
των 2 DIMM (2x2) (DDR3)**



Μνήμη cache

Η **μνήμη cache** παρεμβάλλεται μεταξύ της CPU και της κύριας μνήμης.

Είναι πολύ μικρότερου μεγέθους, αλλά μεγαλύτερης ταχύτητας συγκρινόμενη με την κύρια μνήμη.

Τα κύτταρα αποθήκευσης αυτής είναι, για λόγους ταχύτητας, στατικής τεχνολογίας.

Η μνήμη cache γεφυρώνει την διαφορά ταχύτητας μεταξύ της CPU και της κύριας μνήμης.

Είναι διάφανη στον προγραμματιστή, δηλαδή οι θέσεις της δεν μπορούν να προσπελασθούν άμεσα με εντολές γλώσσας μηχανής.

Μνήμη cache

Σε κάθε προσπάθεια μνήμης εξετάζεται πρώτα εάν το δεδομένο υπάρχει στην μνήμη cache και εάν δεν υπάρχει μεταφέρεται από την κύρια μνήμη. Η μνήμη cache χωρίζεται σε *πλαίσια* (γραμμές) στα οποία τοποθετούνται μπλοκ δεδομένων της κύριας μνήμης. Η μεταφορά δεδομένων μεταξύ της κύριας μνήμης και της μνήμης cache γίνεται ανά μπλοκ. Η μνήμη cache περιέχει επίσης και τμήματα των διευθύνσεων των μπλοκ της κύριας μνήμης που περιέχονται σε αυτή και τα οποία ονομάζονται *ετικέτες* (*tag*).

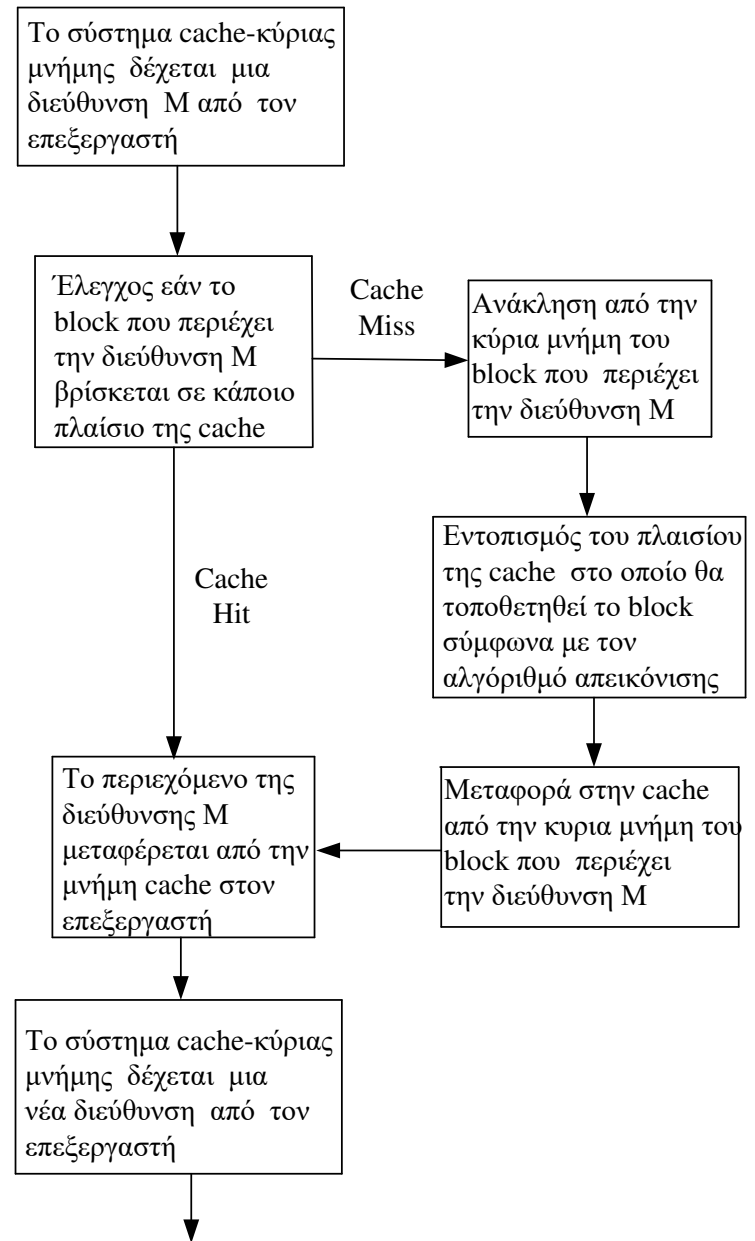
Αρχή της τοπικότητας

Η αύξηση της απόδοσης των συστημάτων Η/Υ με χρήση μνήμης cache βασίζεται σε μία ιδιότητα των προγραμμάτων που ονομάζεται *αρχή της τοπικότητας (principle of locality)*. Σύμφωνα με αυτή στα προγράμματα των υπολογιστών οι αναφορές στην μνήμη δεν γίνονται τυχαία, αλλά τείνουν να εντοπίζονται. Υπάρχουν δύο τύποι τοπικότητας αναφοράς στην μνήμη.

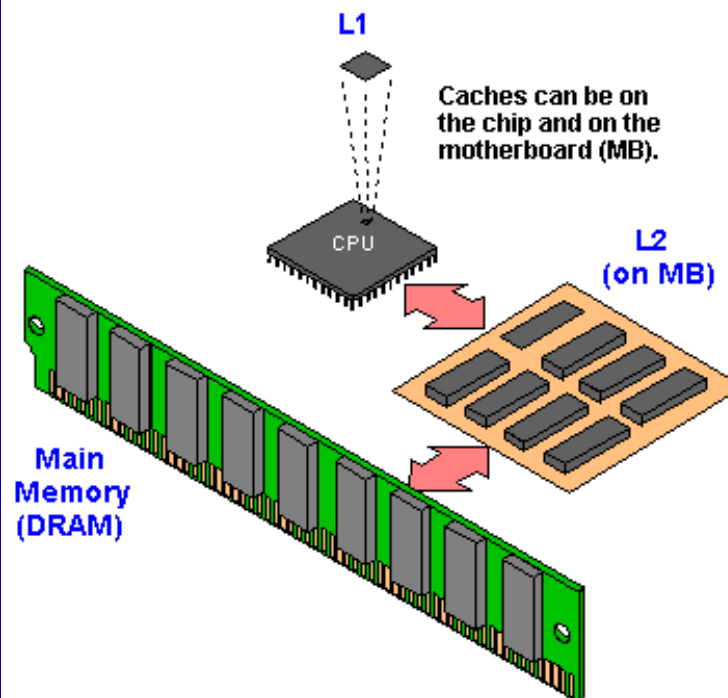
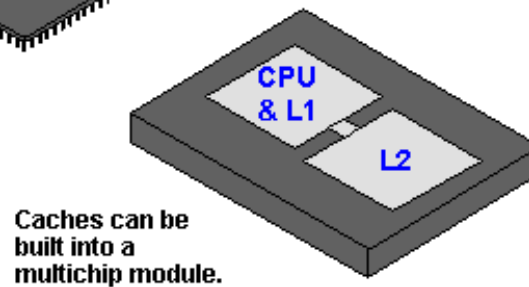
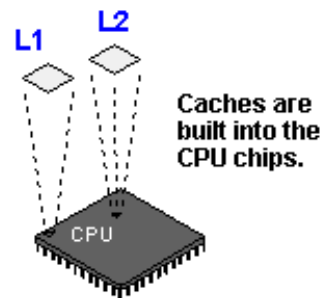
Temporal locality. Όταν γίνει μία προσπέλαση στην μνήμη υπάρχει πολύ μεγάλη πιθανότητα η θέση αυτή να προσπελασθεί στο άμεσο μέλλον.

Spatial locality. Όταν γίνει προσπέλαση σε μία θέση μνήμης είναι πολύ πιθανόν γειτονικές θέσεις μνήμης να προσπελασθούν άμεσα.

Διαδικασία ανάγνωσης δεδομένων σε ένα υπολογιστικό σύστημα με μνήμη cache ενός επιπέδου.



Μνήμη cache



Χαρακτηριστικά των μνημών cache

Οι μνήμες cache έχουν ορισμένα χαρακτηριστικά που επηρεάζουν το κόστος υλοποίησης και την απόδοσή τους. Τα κύρια χαρακτηριστικά μιας μνήμης cache είναι

Το μέγεθός της.

Το μέγεθος των πλαισίων δεδομένων που τοποθετούνται σε αυτή.

Η τακτική τοποθέτησης *block* δεδομένων από την κύρια μνήμη σε αυτή.

Η τακτική αντικατάστασης των *block* δεδομένων που είναι τοποθετημένα σε αυτή.

Η τακτική ενημέρωσης της κύριας μνήμης με τα δεδομένα που εγγράφονται στην μνήμη cache.

Blocks και Lines

Η μεταφορά δεδομένων μεταξύ της κύριας μνήμης και της μνήμης cache γίνεται ανά ομάδες δεδομένων που ονομάζονται *μπλοκ*.

Ουσιαστικά η κύρια μνήμη ενός υπολογιστικού συστήματος που διαθέτει μνήμη cache χωρίζεται σε *μπλοκ* που το κάθε ένα αποτελείται από συγκεκριμένο αριθμό byte.

Η μνήμη cache διαιρείται αντίστοιχα σε *γραμμές (lines)*. Σε κάθε γραμμή της μνήμης cache τοποθετείται ένα block δεδομένων της κύριας μνήμης. Σε κάθε γραμμή τοποθετείται επίσης και ένα τμήμα της διεύθυνσης του μπλοκ που αποθηκεύεται σε αυτό και ονομάζεται *ετικέτα (tag)*.

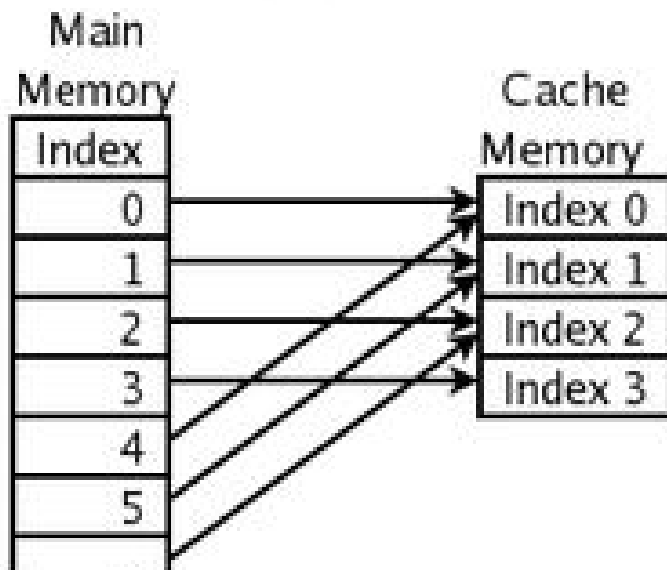
Τεχνικές τοποθέτησης μπλοκ δεδομένων στην μνήμη cache

Direct mapped

Full Associative

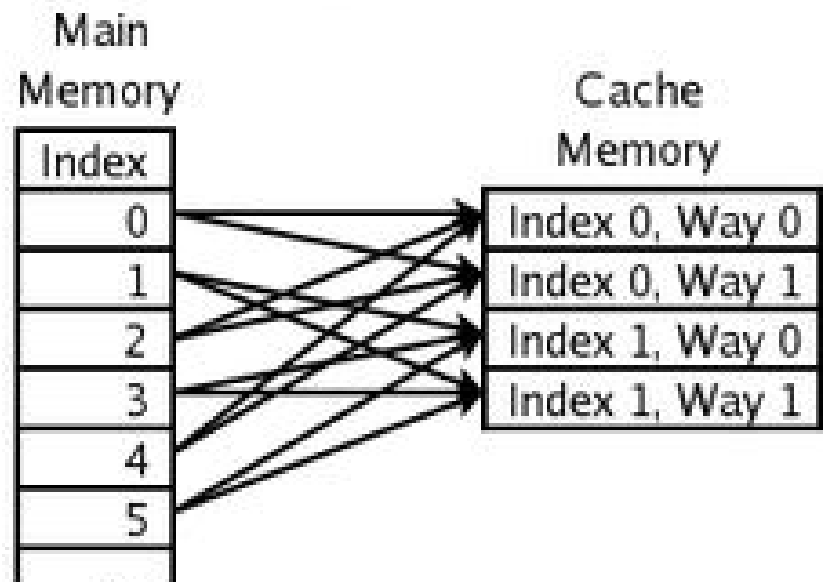
t-way set associative

Direct Mapped Cache Fill



...
Each location in main memory can be cached by just one cache location.

2-Way Associative Cache Fill



...
Each location in main memory can be cached by one of two cache locations.

Direct mapped cache

Με την τεχνική άμεσης απεικόνισης (*direct-mapped*) τα δεδομένα ενός μπλοκ της κύριας μνήμης τοποθετούνται σε συγκεκριμένη γραμμή (πλαίσιο) της κρυφής μνήμης. Ο αριθμός της γραμμής της κρυφής μνήμης, στην οποία θα τοποθετηθούν τα δεδομένα ενός μπλοκ της κύριας μνήμης με διεύθυνση A , προκύπτει από το υπόλοιπο της διαίρεσης του αριθμού A με το πλήθος των πλαισίων της κρυφής μνήμης.

Virtual memory

Η *virtual memory* (ιδεατή μνήμη) είναι τεχνολογία οργάνωσης μνήμης που επιτρέπει ένα τμήμα του σκληρού δίσκου να φαίνεται σαν ένα επίπεδο στο σύστημα μνήμης. Προσφέρει επίσης προστασία μεταξύ των προγραμμάτων που τρέχουν στο ίδιο σύστημα ώστε το ένα πρόγραμμα να μην τροποποιεί τα δεδομένα του άλλου.

Η μονάδα διαχείρισης μνήμης (MMU) και το λειτουργικό σύστημα (OS) είναι υπεύθυνα για την υλοποίηση της ιδεατής μνήμης.

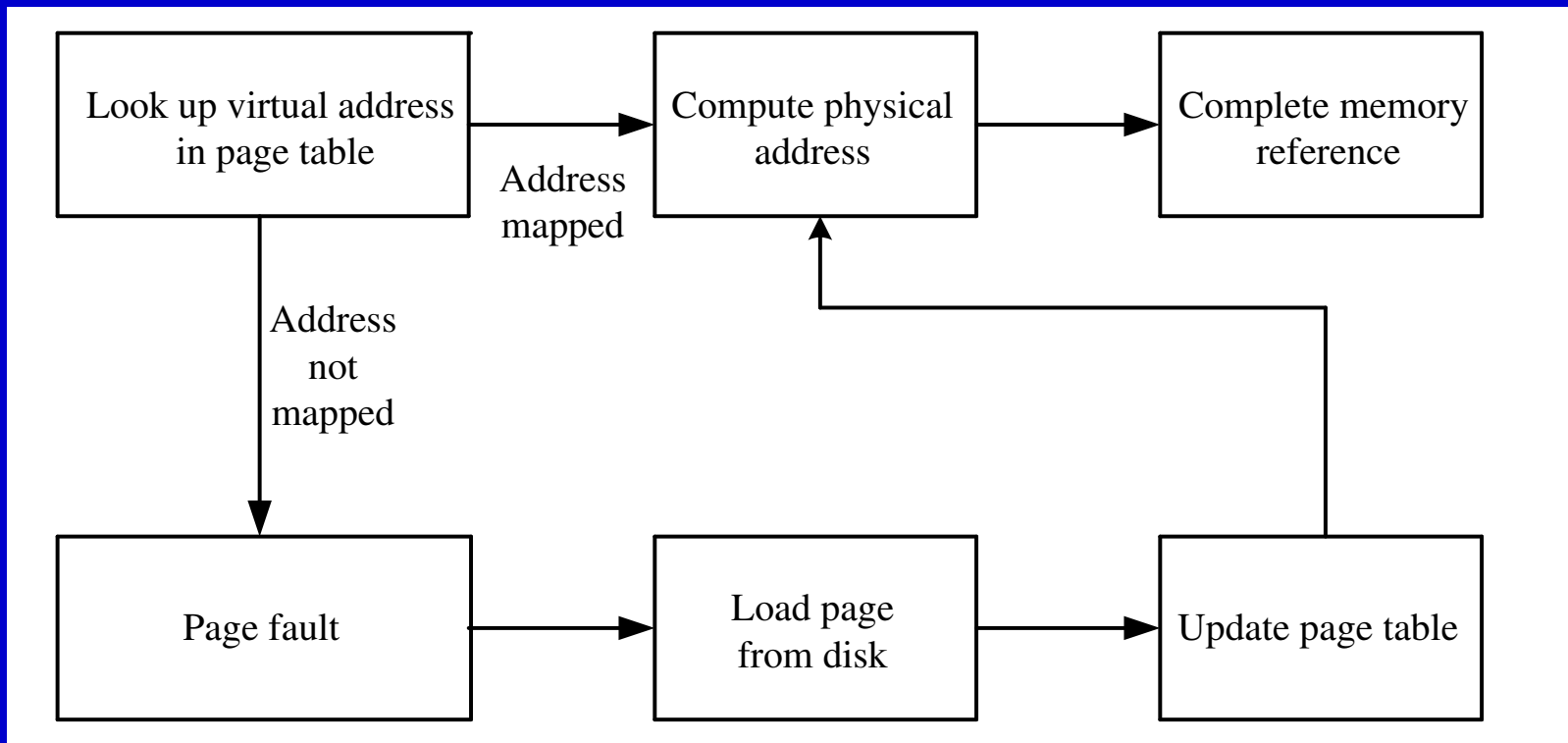
Στα συστήματα με virtual memory τα προγράμματα δεν μπορούν να προσπελάσουν απευθείας δεδομένα στον σκληρό δίσκο. Αντίθετα το πρόγραμμα χωρίζεται σε σελίδες (pages) και αποθηκεύονται στον σκληρό δίσκο.

Σε σύστημα με virtual μνήμη οι διευθύνσεις που παράγει η CPU ονομάζονται λογικές διευθύνσεις και το σύνολο των λογικών διευθύνσεων χώρος λογικών διευθύνσεων. Οι διευθύνσεις της πραγματικής κύριας μνήμης ονομάζονται φυσικές διευθύνσεις και το σύνολό τους χώρος φυσικών διευθύνσεων.

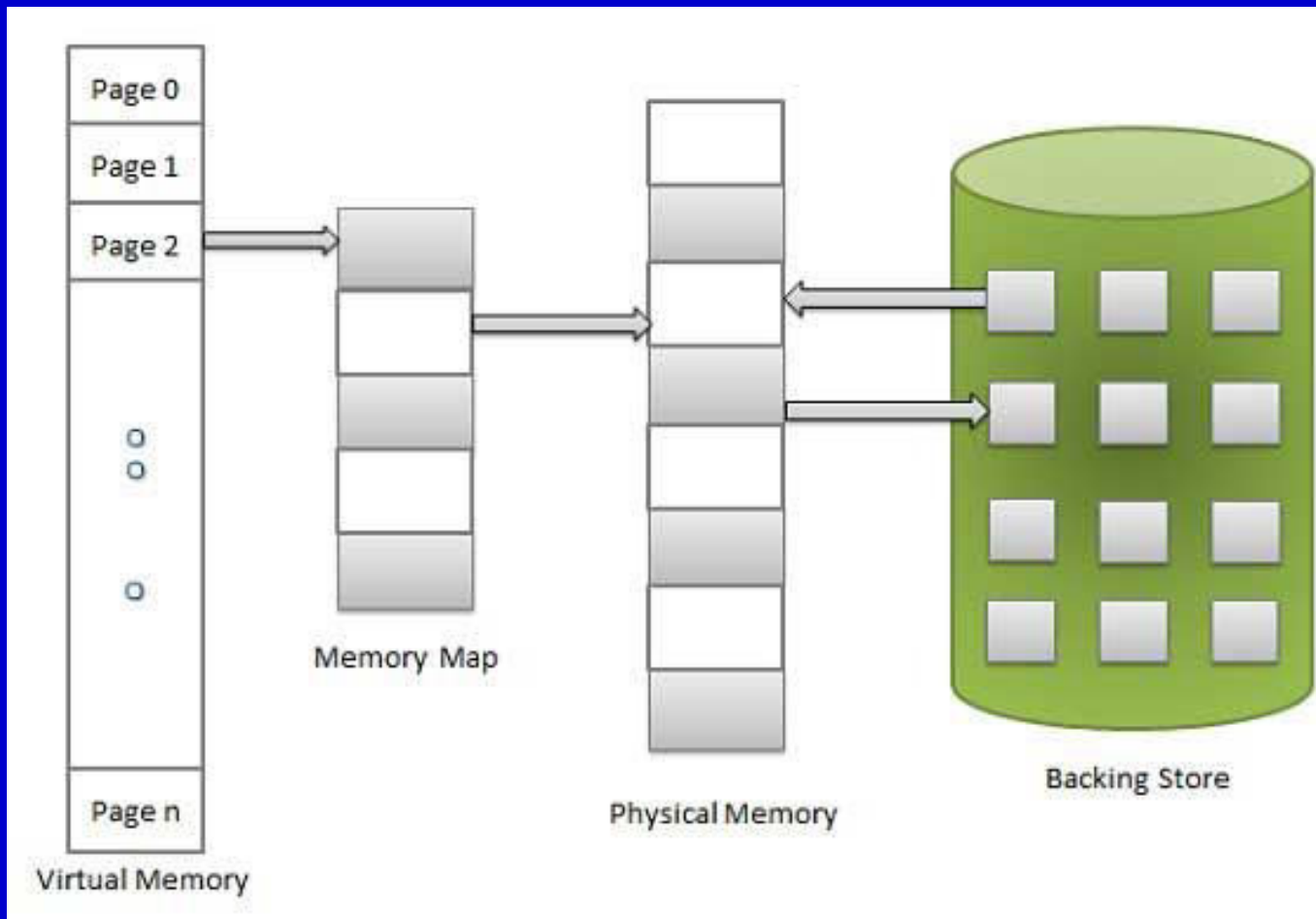
Page table

Το λειτουργικό σύστημα χρησιμοποιεί μία δομή δεδομένων που ονομάζεται *page table* για να παρακολουθεί πως οι *virtual addresses* απεικονίζονται σε φυσικές διευθύνσεις. Επειδή κάθε πρόγραμμα έχει το δικό του *virtual-physical address mapping* χρειάζεται διαφορετικό *page table* για κάθε πρόγραμμα.

Virtual memory access



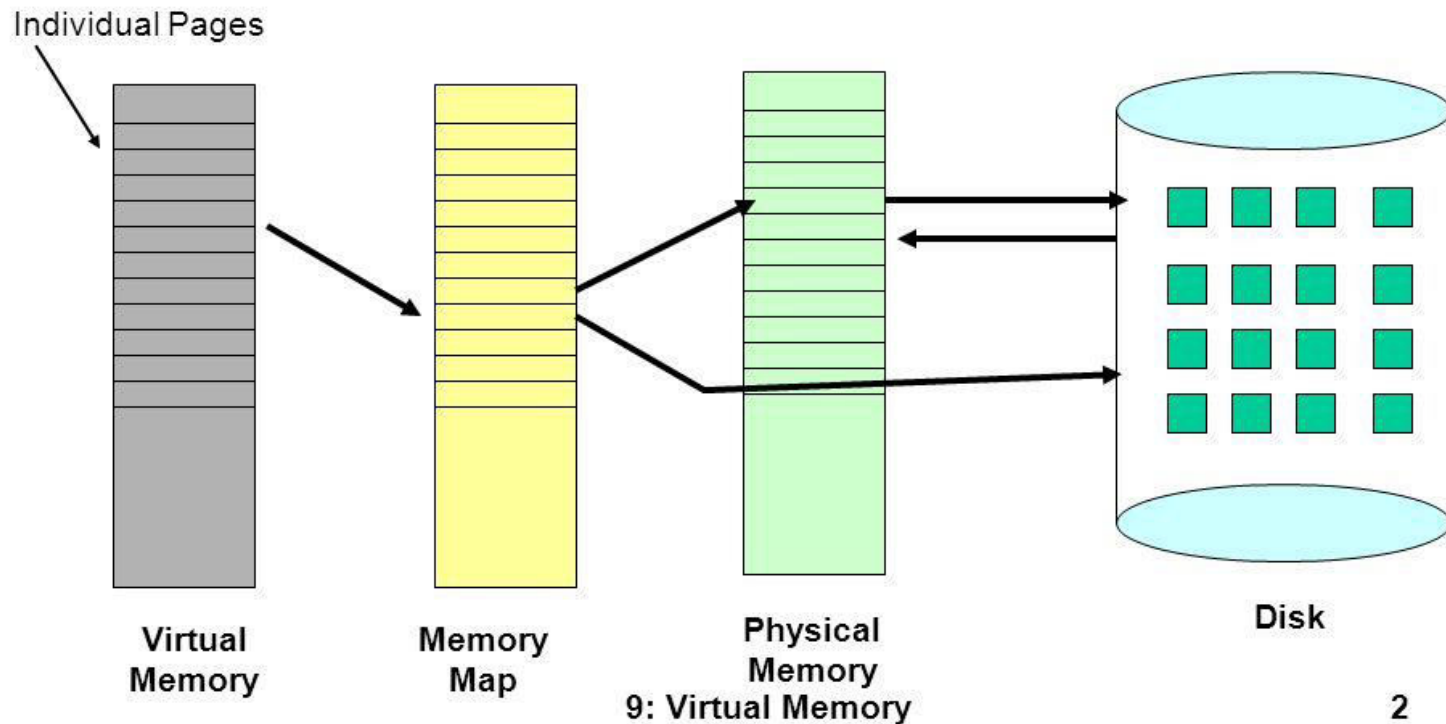
Σύστημα virtual memory



VIRTUAL MEMORY

Definitions

Virtual memory The conceptual separation of user logical memory from physical memory. Thus we can have large virtual memory on a small physical memory.



Μέσος χρόνος προσπέλασης ιεραρχικών συστημάτων μνήμης

Σε ένα ιεραρχικό σύστημα μνήμης, εάν γνωρίζουμε το hit rate και τον χρόνο προσπέλασης για κάθε επίπεδο, μπορούμε να υπολογίσουμε τον μέσο χρόνο προσπέλασής του. Για κάθε επίπεδο ο μέσος χρόνος προσπέλασης είναι

$$t_{av} = t_{hit} \times h_{hit} + t_{miss} \times (1 - h_{hit}),$$

όπου t_{hit} είναι ο χρόνος προσπέλασης που εκπληρώθηκε με επιτυχία στο συγκεκριμένο επίπεδο, h_{hit} είναι ο ρυθμός επιτυχίας για το συγκεκριμένο επίπεδο, t_{miss} είναι ο μέσος χρόνος προσπέλασης στην αμέσως κατώτερη επίπεδο στην ιεραρχία μνήμης. Στο κατώτατο επίπεδο ο ρυθμός επιτυχίας είναι 100%.

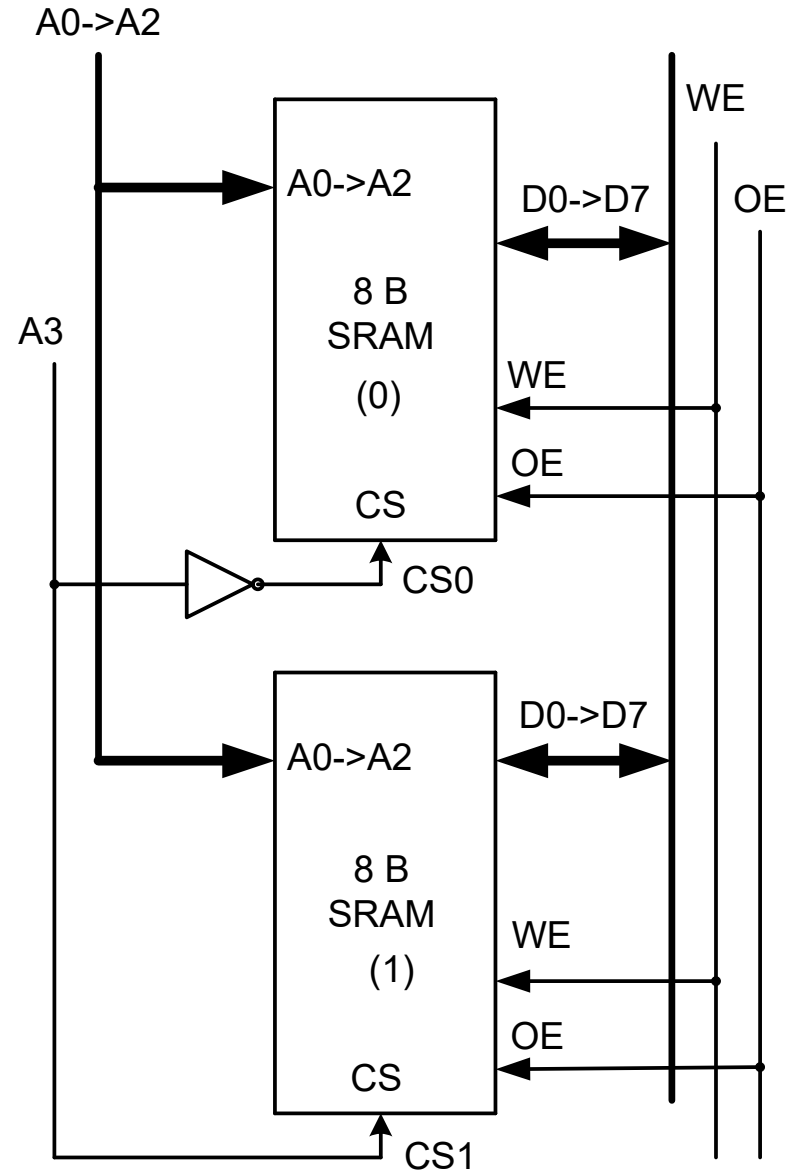
7.1. Ένα σύστημα κύριας μνήμης Η/Υ έχει χρόνο κύκλου 10 ns και εύρος διαύλου δεδομένων επικοινωνίας με την CPU 32 bit. Ποιος είναι ο ρυθμός μεταφοράς δεδομένων και ποιο το bandwidth αυτού του συστήματος.

Υπόδειξη

$$\begin{aligned}\text{Ρυθμός_μεταφοράς} &= 1/\text{χρόνος_κύκλου} = 1/10\text{ns} = \\ &= 100 \times 10^6 \text{ μεταφορές/sec}\end{aligned}$$

$$\begin{aligned}\text{Bandwidth} &= \text{Ρυθμός_μεταφοράς} \times 32 \text{ bit} \\ &= \text{Ρυθμός_μεταφοράς} \times 4 \text{ byte} \\ &= 400 \times 10^6 \text{ byte/sec} \\ &= 400 \text{ MB/sec.}\end{aligned}$$

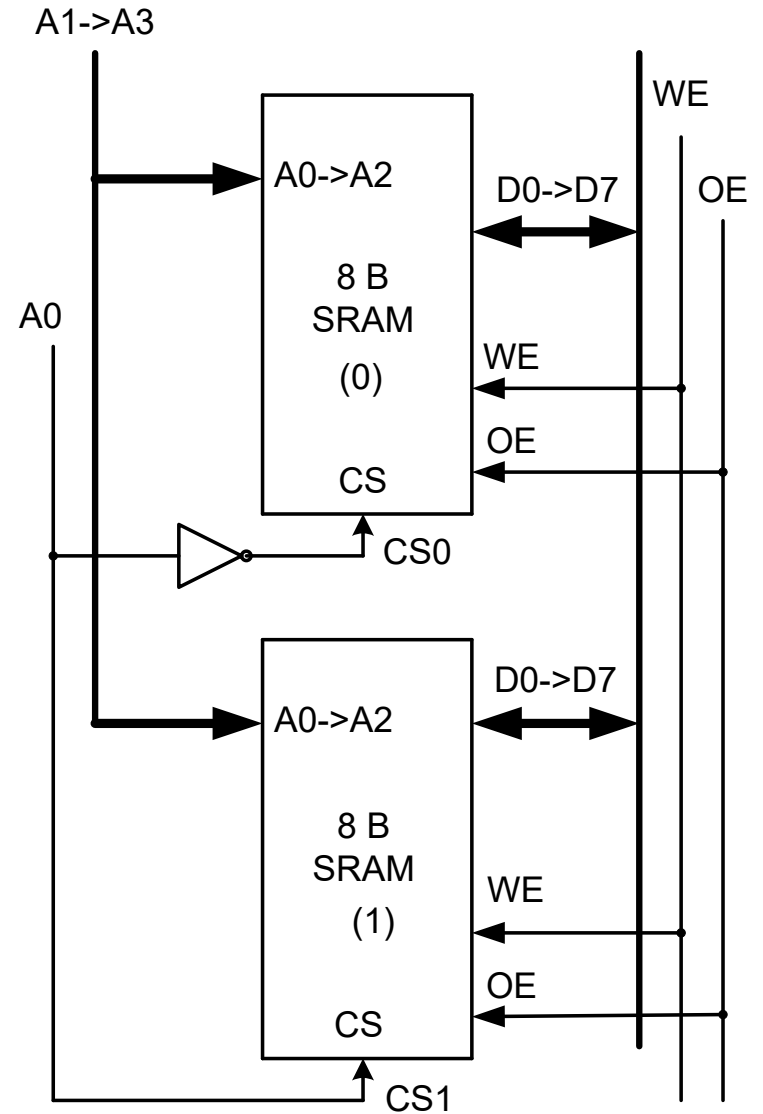
7.2. Να δοθούν οι διευθύνσεις μνήμης που αντιστοιχούν σε κάθε μονάδα μνήμης του δοσμένου συστήματος μνήμης.



Η επιλογή της μονάδας SRAM που θα χρησιμοποιεί τον δίαυλο δεδομένων γίνεται με την ενεργοποίηση του αντίστοιχου σήματος CS. Για $A_3=0$ ενεργοποιείται η μονάδα 0, ενώ για $A_3=1$ ενεργοποιείται η μονάδα 1.

Επομένως οι διευθύνσεις 0000→0111 αντιστοιχούν στην μονάδα 0, ενώ οι διευθύνσεις 1000→1111 στις στην μονάδα 1

7.3. Να δοθούν οι διευθύνσεις μνήμης που αντιστοιχούν σε κάθε module του συστήματος μνήμης που βασίζεται στην τεχνική 2-way memory interleaving που δίδεται στην συνέχεια.



Η επιλογή της μονάδας SRAM που θα χρησιμοποιεί τον δίαυλο δεδομένων γίνεται με την ενεργοποίηση του αντίστοιχου σήματος CS. Για $A0=0$ ενεργοποιείται η μονάδα 0, ενώ για $A0=1$ ενεργοποιείται η μονάδα 1.

Επομένως οι διευθύνσεις 0000, 0010, 0100, 0110, ..., 110 αντιστοιχούν στην μονάδα 0, ενώ οι διευθύνσεις 0001, 0011, 0101, ..., 1111, στις στην μονάδα 1.

7.4. Στο τμήμα motherboard που δίδεται στην συνέχεια σημειώστε τις θέσεις στις οποίες πρέπει να τοποθετηθούν 2 DIMM DDR3 SDRAM για επιτευχθεί η ταχύτερη λειτουργία του συστήματος.



7.5. Ένα σύστημα μνήμης περιλαμβάνει μνήμη cache και κύρια μνήμη. Η μνήμη cache έχει χρόνο προσπέλασης 5 ns και hit rate 80%. Η κύρια μνήμη έχει χρόνο προσπέλασης 100 ns. Ποιος είναι ο μέσος χρόνος προσπέλασης του συστήματος μνήμης.

Λύση

Έστω t_{av} ο μέσος προσπέλασης του συστήματος μνήμης. Επίσης έστω $t_c=5$ ns ο χρόνος προσπέλασης και $h_c=0,8$ το hit rate της μνήμης cache. Ο χρόνος προσπέλασης της κύριας μνήμης είναι $t_m=100$ ns. Ισχύει

$$t_{av}=t_c \times h_c + (1-h_c) \times t_m,$$

Επομένως,

$$\begin{aligned} t_{av} &= (5 \times 0,8 + 100 \times 0,2) \text{ ns} \\ &= (4 + 20) \text{ ns} = 24 \text{ ns} \end{aligned}$$

7.6. Ένα σύστημα μνήμης περιλαμβάνει μνήμη cache, κύρια μνήμη και virtual μνήμη. Η μνήμης cache έχει χρόνο προσπέλασης 5 ns και hit rate 80%. Η κύρια μνήμη έχει χρόνο προσπέλασης 100 ns και hit rate 99.5%. Ο χρόνος προσπέλασης της virtual memory είναι 10 ms. Ποιος είναι ο μέσος χρόνο προσπέλασης του συστήματος μνήμης..

Λύση

Έστω t_{av} ο μέσος προσπέλασης του συστήματος μνήμης, $t_c=5$ ns ο χρόνος προσπέλασης και $h_c=0,8$ το hit rate της μνήμης cache. Ανάλογα, έστω $t_m=100$ ns ο χρόνος προσπέλασης, $h_m=0,995$ το hit rate της κύρια μνήμης. και ο χρόνος προσπέλασης της virtual μνήμης είναι $t_v=10$ ms. Ισχύει

$$t_{av}=t_c \times h_c + (1-h_c) \times t_{avm},$$

όπου t_{avm} ο μέσος χρόνος προσπέλασης της κύριας μνήμης. Ο μέσος χρόνος προσπέλασης της κύριας μνήμης είναι

$$t_{avm}=t_m \times h_m + t_v \times (1-h_m)$$

Επομένως

$$\begin{aligned} t_{avm} &= (100 \times 0,995 + 10000000 \times 0,005) \text{ ns} = \\ &= (99,5 + 50000) \text{ ns} = 50099,5 \text{ ns} \end{aligned}$$

Κατά συνέπεια

$$\begin{aligned} t_{av} &= (5 \times 0,8 + 50099,5 \times 0,2) \text{ ns} \\ &= (4 + 1020) \text{ ns} = 1024 \text{ ns} \approx 10 \mu\text{s} \end{aligned}$$

7.7. Έστω υποθετικό υπολογιστικό σύστημα που έχει κύρια μνήμη μεγέθους 32 byte και μήκος λέξης 1 byte. Η μνήμη cache του υπολογιστικού συστήματος έχει 4 πλαίσια και είναι οργανωμένη ώστε το μέγεθος του τμήματος δεδομένων των πλαισίων της να είναι 4 byte.

α) Ποιο είναι το μήκος του διαύλου διευθύνσεων το επεξεργαστή.

β) Σε πόσα μπλοκ χωρίζεται η κύρια μνήμη.

γ) Ποια bit της διεύθυνσης ενός byte της κύριας μνήμης προσδιορίζουν τον αριθμό (διεύθυνση) του μπλοκ στο οποίο ανήκει το byte. Ποια bit των διευθύνσεων χρησιμοποιούνται για τον προσδιορισμό της θέσης του byte μέσα στο μπλοκ ή στο πλαίσιο.

7.8. Ένα εκπαιδευτικό σύστημα Η/Υ έχει μνήμη cache με αρχιτεκτονική άμεσης απεικόνισης. Η κύρια μνήμη διαθέτει 16 μπλοκ και η μνήμη cache 4 πλαίσια. Το τμήμα δεδομένων κάθε πλαισίου έχει 4 byte. Να βρεθεί σε ποιο πλαίσιο της μνήμης cache μπορεί να τοποθετηθεί κάθε μπλοκ της κύριας μνήμης. Η αρίθμηση των μπλοκ και των πλαισίων αρχίζει από 0.