

Η ιεραρχία της μνήμης

- Οι περιορισμοί στο σχεδιασμό της μνήμης συνοψίζονται σε τρεις ερωτήσεις :

1) Πόση

2) Πόσο γρήγορη

3) Πόσο ακριβή

Ερωτήματα-Απαντήσεις

- Ερώτημα πόση μνήμη. Είναι ανοικτό. Αν υπάρχει χωρητικότητα οι εφαρμογές θα αναπτυχθούν προκειμένου να χρησιμοποιούν αυτή τη μνήμη.
- Ερώτημα πόσο γρήγορη. Για να πετύχουμε μεγαλύτερη απόδοση, η μνήμη θα πρέπει να είναι σε θέση να ακολουθεί τον επεξεργαστή. Ο επεξεργαστής δεν θα πρέπει κατά την εκτέλεση εντολών να σταματά για να περιμένει να έρθουν εντολές ή δεδομένα από τη μνήμη.
- Ερώτημα πόσο ακριβή μνήμη. Το κόστος της μνήμης θα πρέπει να είναι λογικό σε σχέση με εκείνο των άλλων στοιχείων.
- Τα τρία βασικά χαρακτηριστικά της μνήμης ...
Χωρητικότητα, χρόνος προσπέλασης, κόστος
- Σχετίζονται μεταξύ τους.

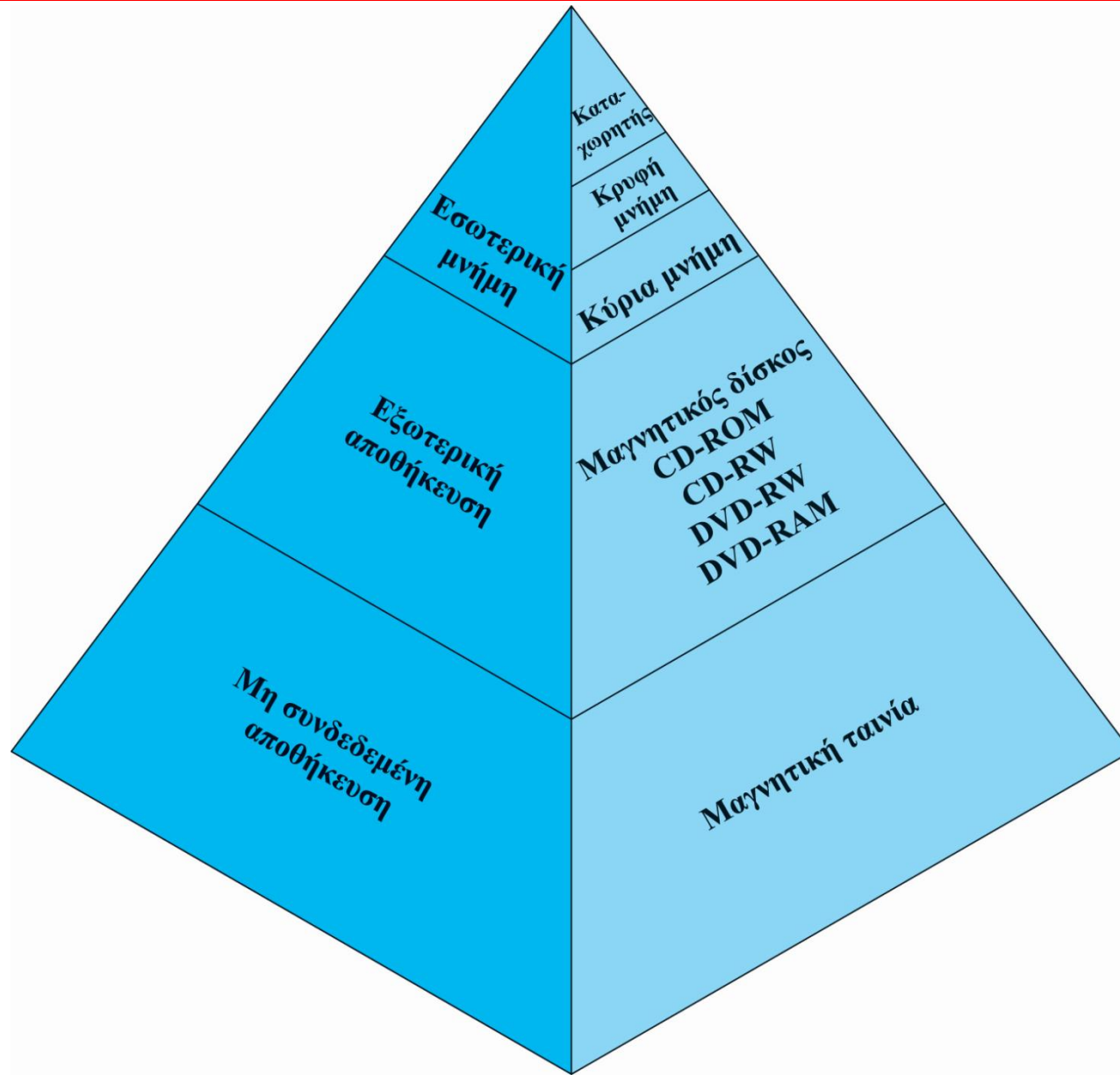
Βασικές αρχές στην κατασκευή μνήμης

- Για την υλοποίηση συστημάτων μνήμης, μέσα από διάφορες τεχνολογίες ισχύουν οι παρακάτω βασικές σχέσεις:
- Ταχύτερη προσπέλαση σημαίνει μεγαλύτερο κόστος ανά bit.
- Μεγαλύτερη χωρητικότητα σημαίνει μικρότερο κόστος ανά bit.
- Μεγαλύτερη χωρητικότητα σημαίνει πιο αργή προσπέλαση.

Επιλογές σχεδίασης

- Ο σχεδιαστής θα επιθυμούσε να χρησιμοποιήσει μνήμη μεγάλης χωρητικότητας γιατί η χωρητικότητα είναι αναγκαία και γιατί το κόστος/bit είναι χαμηλό.
- Για εξασφάλιση υψηλής απόδοσης πρέπει να χρησιμοποιηθεί μνήμη ...
- Σχετικά μικρής χωρητικότητας
- Με μικρούς χρόνους προσπέλασης
- Σχετικά ακριβή

Η ιεραρχία μνήμης - Διάγραμμα



Η ιεραρχία της μνήμης

- Καταχωρητές
 - Εντός του επεξεργαστή
- Εσωτερική ή Κύρια Μνήμη
 - Μπορεί να περιλαμβάνει ένα ή περισσότερα επίπεδα ενδιάμεσης μνήμης - cache
 - RAM
- Εξωτερική Αποθήκευση
 - Μέσα Αποθήκευσης
- Μη συνδεδεμένη αποθήκευση
 - Μαγνητική ταινία

Σειρά Ιεραρχίας

- Καταχωρητές
- L1 Cache
- L2 Cache
- Κύρια Μνήμη
- Cache Δίσκου
- Δίσκος
- Οπτικό μέσο
- Μαγνητικό μέσο

Χαρακτηριστικά Ιεραρχίας

- Διασχίζοντας την ιεραρχία ...
 - A. Μειώνεται το κόστος ανά bit
 - B. Αυξάνεται η χωρητικότητα
 - C. Αυξάνεται ο χρόνος προσπέλασης
 - D. Μειώνεται η συχνότητα της προσπέλασης της μνήμης από τον επεξεργαστή.

Θέλουμε Ταχύτητα?

- Είναι δυνατό να κατασκευάσουμε υπολογιστή που θα χρησιμοποιεί μόνο SRAM
- Θα είναι πραγματικά γρήγορος
- Θα ήταν πραγματικά πανάκριβος!!!!
- Οι μικρότερες μνήμες συμπληρώνονται από μεγαλύτερες πιο φθηνές και πιο αργές μνήμες.
- Το κλειδί της επιτυχίας αυτής της οργάνωσης είναι το D: μειωμένη συχνότητα προσπάθειας.
- Η χρήση δύο επιπέδων μνήμης προκειμένου να μειωθεί ο μέσος χρόνος προσπάθειας λειτουργεί γενικά αλλά μόνο με την προϋπόθεση ότι ισχύουν τα A έως D
- Χρησιμοποιώντας μια ποικιλία από τεχνολογίες υπάρχει ένα φάσμα από συστήματα μνήμης τα οποία ικανοποιούν τις συνθήκες από A έως C

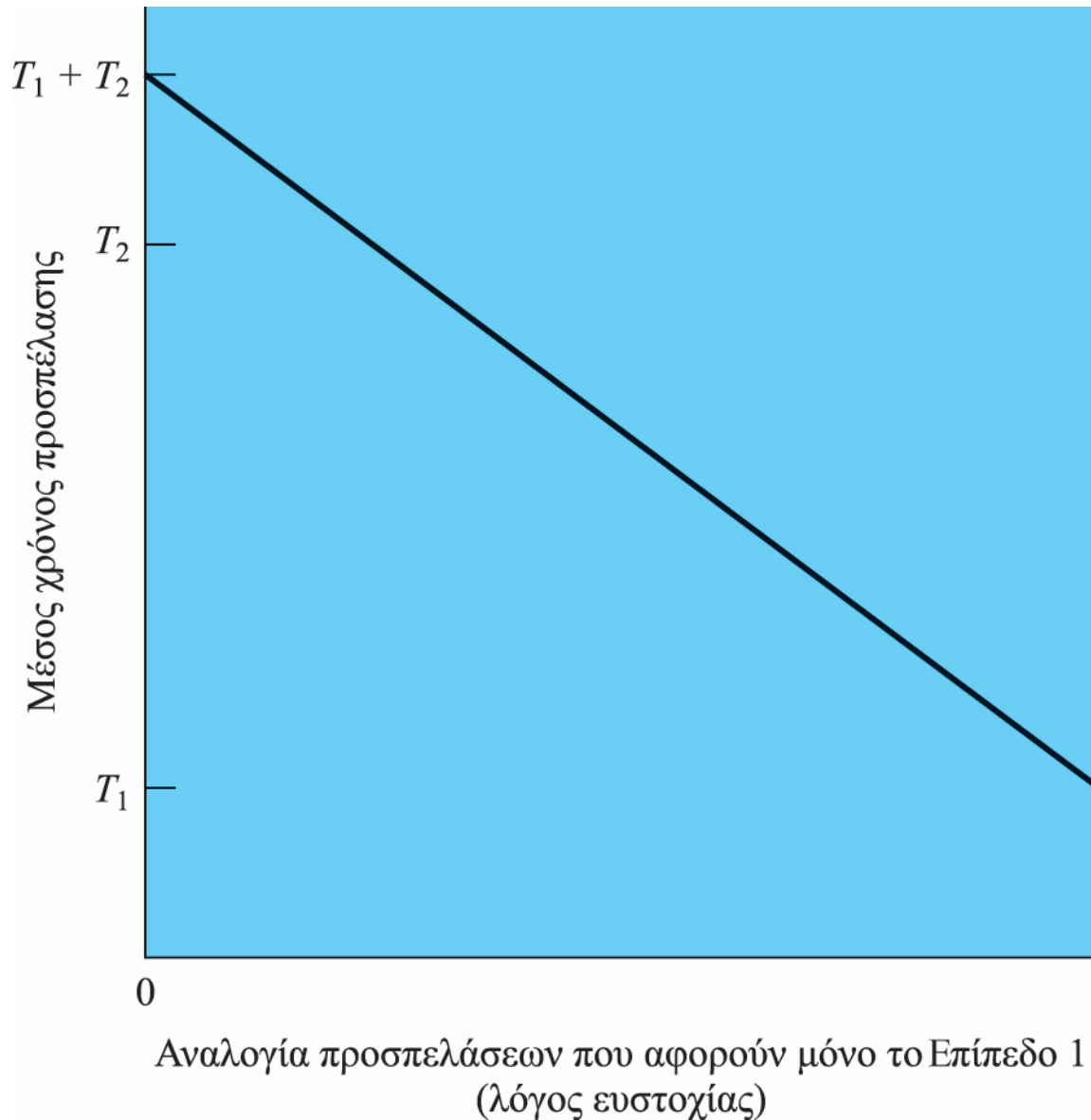
Η αρχή της Τοπικότητας της Αναφοράς

- Η βάση της εγκυρότητας της συνθήκης (D) είναι μια αρχή γνωστή ως τοπικότητα αναφοράς.
- Κατά τη πορεία εκτέλεσης ενός προγράμματος, οι αναφορές που κάνει ο επεξεργαστής στη μνήμη για δεδομένα ή εντολές, τείνουν να ομαδοποιούνται.
- π.χ. οι βρόγχοι και οι υπορουτίνες
- Γίνονται επαναλαμβανόμενες αναφορές σε ένα μικρό σύνολο από εντολές.
- Όμοια και οι πράξεις που εκτελούνται σε πίνακες
- Ο επεξεργαστής με σταθερές ομάδες αναφορών στη μνήμη.

Παράδειγμα μνήμης δυο επιπέδων

- Ο επεξεργαστής έχει πρόσβαση σε δυο επίπεδα μνήμης.
- Το επίπεδο 1 περιλαμβάνει 1000 λέξεις με χρόνο προσπέλασης 0.01μs
- Το επίπεδο 2 περιλαμβάνει 100.000 λέξεις και έχει χρόνο προσπέλασης 0.1μs
- Λέξη που βρίσκεται στο επίπεδο 1, ο επεξεργαστής την προσπελαίνει αμέσως.
- Λέξη που βρίσκεται στο επίπεδο 2, η λέξη μεταφέρεται στο επίπεδο 1 και μετά προσπελαίνεται από τον επεξεργαστή.
- Ο χρόνος αναγνώρισης από τον επεξεργαστή του επιπέδου δεν λαμβάνεται υπ' όψη.

Παράδειγμα μνήμης δυο επιπέδων



Παράδειγμα μνήμης δυο επιπέδων

- Το σχήμα δείχνει το μέσο χρόνο προσπέλασης μιας μνήμης 2 επιπέδων ως συνάρτηση του λόγου ευστοχίας (Hit ratio) H όπου το H ορίζεται ως το ποσοστό όλων των προσπελάσεων της μνήμης, οι οποίες βρέθηκαν στην γρήγορη μνήμη (κρυφή μνήμη)
- T_1 χρόνος προσπέλασης του επιπέδου 1
- T_2 χρόνος προσπέλασης του επιπέδου 2.
- Έστω $H = 95\%$ στο παράδειγμά μας
- Ο μέσος χρόνος προσπέλασης μιας λέξης είναι :
- $0.95 \cdot 0.01 + 0.05 \cdot (0.01 + 0.1) = 0.0095 + 0.0055 = 0.015 \mu s$
- Ο μέσος χρόνος προσπέλασης είναι πιο κοντά στα $0.01 \mu s$ όπως και είναι επιθυμητό.

Ιεραρχική οργάνωση δεδομένων

- Τα δεδομένα είναι δυνατόν να οργανωθούν έτσι ώστε το ποσοστό των προσπελάσεων σε κάθε διαδοχικά χαμηλότερο επίπεδο να είναι πιο μικρό από το ποσοστό του αμέσως υψηλότερου επιπέδου.
- Το παράδειγμα με την κρυφή μνήμη.... Η μνήμη επιπέδου 2 έχει όλα τα δεδομένα και εντολές.
- Οι τρέχουσες συστάδες τοποθετούνται προσωρινά στο επίπεδο 1.
- Μετά μια συστάδα από το επίπεδο 1 θα μεταφερθεί στο επίπεδο 2 και έτσι μια νέα θα μεταφερθεί στο 1.
- Η μέση κατάσταση είναι ότι οι περισσότερες αναφορές θα γίνονται στα δεδομένα του επιπέδου 1

Κύρια μνήμη – κρυφή μνήμη

- Ένας επεξεργαστής έχει συνήθως μερικές δεκάδες καταχωρητές.
- Η κύρια μνήμη αποτελεί το βασικό εσωτερικό σύστημα μνήμης του υπολογιστή.
- Κάθε θέση της έχει μοναδική διεύθυνση.
- Επεκτείνεται με την προσθήκη μιας μικρότερης κρυφής μνήμης υψηλής ταχύτητας.
- Η κρυφή μνήμη δεν είναι ορατή στον προγραμματιστή ή στην πραγματικότητα, στον επεξεργαστή.
- Πρόκειται για μια συσκευή που ρυθμίζει την κίνηση των δεδομένων μεταξύ κύριας μνήμης και καταχωρητών για βελτίωση της απόδοσης.

Δευτερεύουσα μνήμη

- Οι τρεις μορφές μνήμης που έχουν περιγραφεί, είναι πτητικές και κατασκευασμένες με τεχνολογία ημιαγωγών.
- Τα δεδομένα αποθηκεύονται σε εξωτερικές συσκευές αποθήκευσης μεγάλου όγκου, τους δίσκους.
- Η μνήμη αυτή ονομάζεται και δευτερεύουσα μνήμη ή βοηθητική.
- Χρησιμοποιούνται για αποθήκευση αρχείων και δεδομένων των προγραμμάτων.
- Είναι ορατές στον προγραμματιστή με τη μορφή αρχείων και εγγραφών και όχι bytes και words.
- Είναι δυνατόν να προστεθούν και άλλα επίπεδα με λογισμικό. Ένα τμήμα της κύριας μνήμης να γραφτεί σε κάποιο δίσκο (Disk Cache) αναφέρεται λανθάνων δίσκος.

Αρχιτεκτονική και Οργάνωση Υπολογιστών

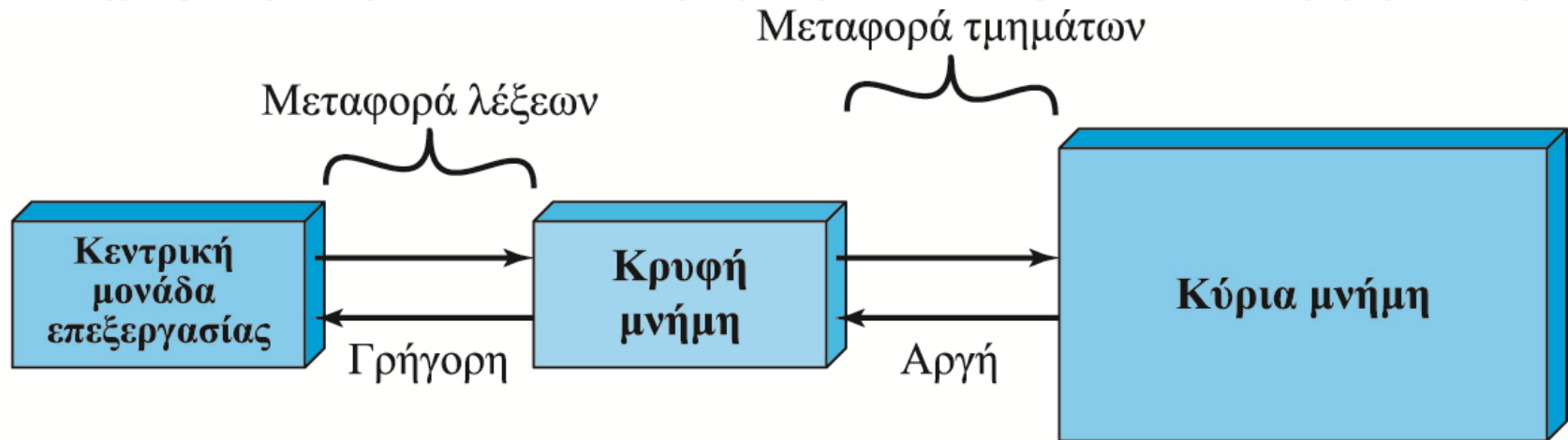
Κεφάλαιο 4.2

Αρχές κρυφής μνήμης

- Έχει στόχο να δώσει στην κύρια μνήμη ταχύτητα που προσεγγίζει εκείνη των πιο γρήγορων μνημών. Σε κάποιες περιπτώσεις να παρέχει μεγάλο μέγεθος με τιμή ανάλογη των πιο φθηνών μνημών.
 - Υπάρχει μια σχετικά μεγάλη και αργή κύρια μνήμη.
 - Υπάρχει μια πιο μικρή υψηλότερης ταχύτητας.
 - Στην περίπτωση ανάγνωσης από τον επεξεργαστή μιας λέξης μνήμης, η λέξη αναζητείται στην κρυφή μνήμη.
-
- ✓ Εάν υπάρχει, η λέξη μεταφέρεται στον επεξεργαστή.
 - ✓ Αλλιώς η κρυφή μνήμη διαβάζει ένα τμήμα της κύριας μνήμης και η λέξη παραδίδεται στον επεξεργαστή.

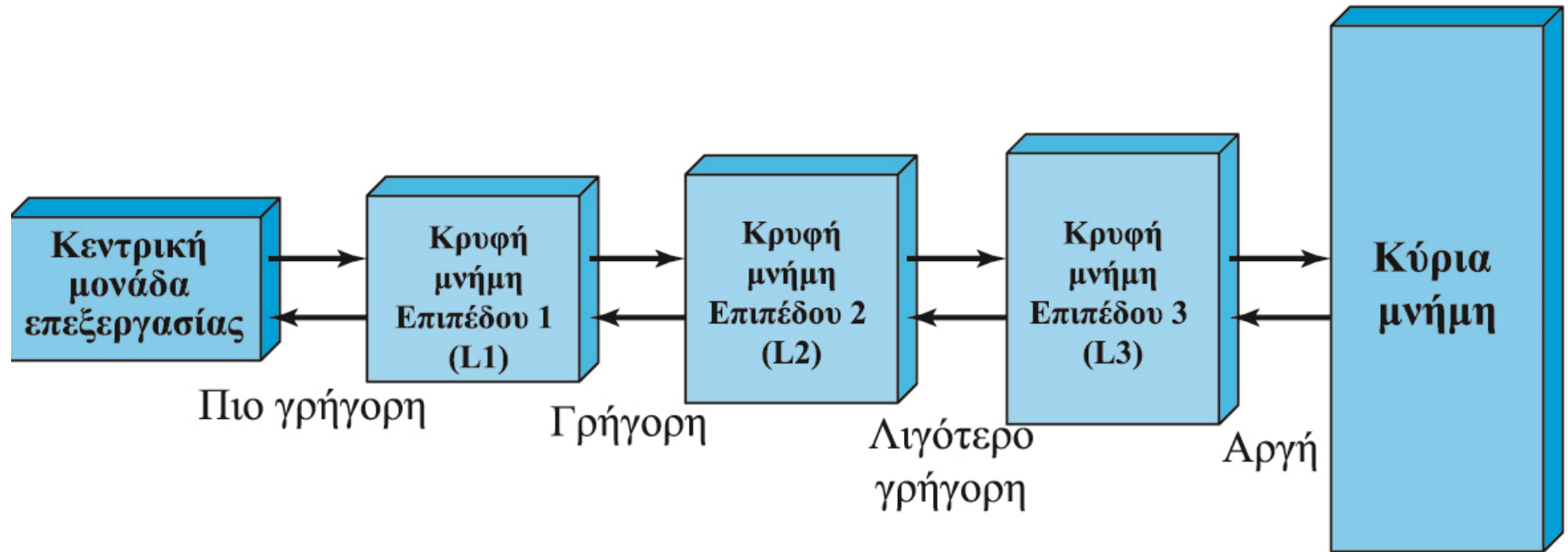
Αρχές κρυφής μνήμης

- Λόγω του φαινομένου της τοπικότητας αναφοράς είναι πιθανό να υπάρχουν μελλοντικές αναφορές σ' αυτό το μπλοκ της μνήμης.
- Τοποθετείται μεταξύ της Κύριας Μνήμης και του Επεξεργαστή
- Σήμερα μπορεί να είναι μέρος του chip του επεξεργαστή



(α) Απλή κρυφή μνήμη

Οργάνωση κρυφής μνήμης τριών επιπέδων



(β) Οργάνωση κρυφής μνήμης τριών επιπέδων

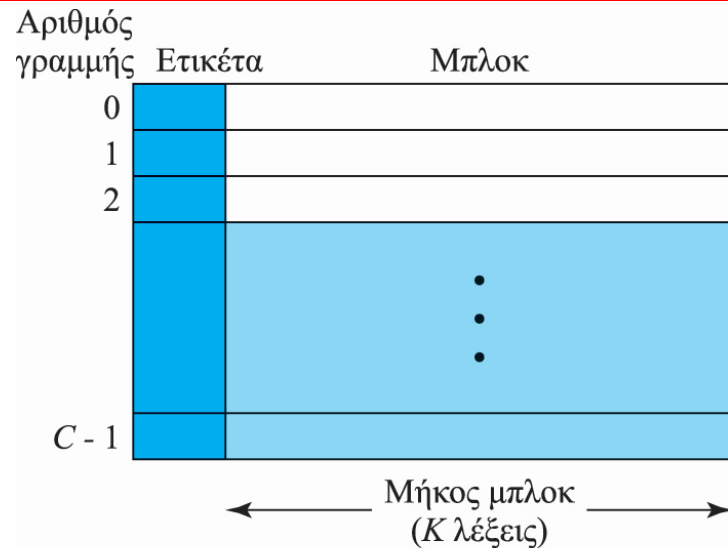
Κρυφή μνήμη με πολλά επίπεδα

- Η κρυφή μνήμη L2 είναι πιο αργή και τυπικά μεγαλύτερη από την L1.
- Η κρυφή μνήμη L3 είναι πιο αργή και μεγαλύτερη από την L2.
- Αν υπάρχει και επόμενο/α επίπεδο/α θα είναι πιο αργά σε ταχύτητα και μεγαλύτερο/α σε μέγεθος από τα προηγούμενα.

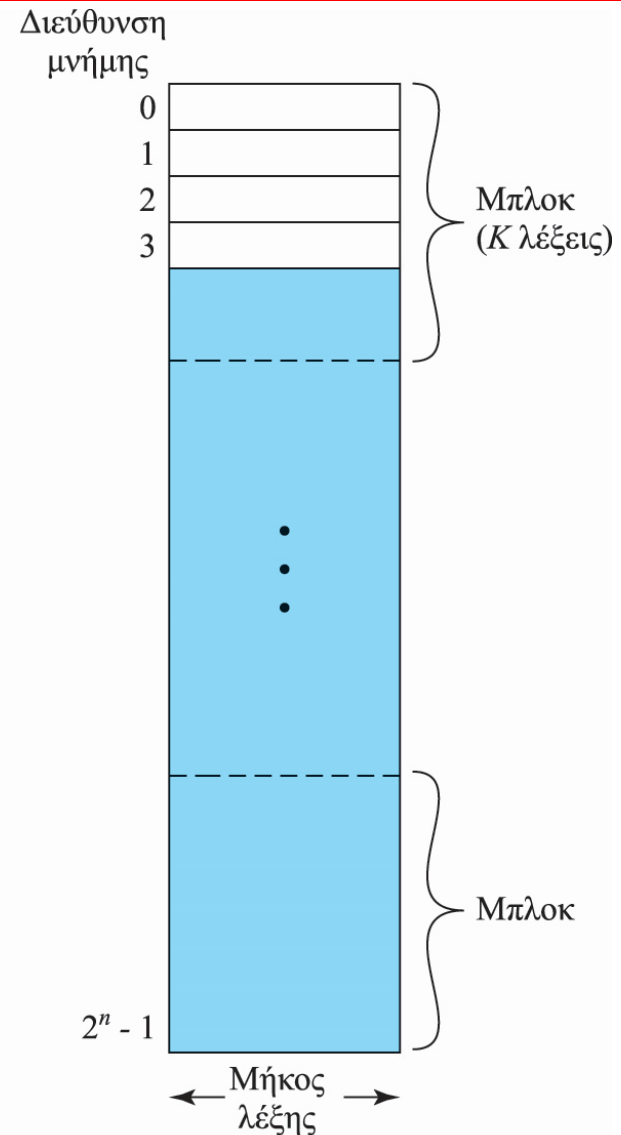
Η λειτουργία της Cache – Γενικά

- Η CPU στέλνει αίτημα για μεταφορά δεδομένων από μια θέση μνήμης
- Γίνεται έλεγχος στην cache για τα δεδομένα
- Αν υπάρχουν, διατίθενται γρήγορα και άμεσα από την cache
- Αν δεν υπάρχουν, μεταφέρεται το απαιτούμενο τμήμα της Κ.Μ. στην cache και ταυτόχρονα στη CPU
- Σε άλλες περιπτώσεις έχουμε μεταφορά δεδομένων πρώτα στην κρυφή και μετά από την cache στην CPU
- Η Cache περιλαμβάνει ετικέτες για να αναγνωρίζει ποιο τμήμα της μνήμης βρίσκεται σε ποια γραμμή της.

Δομή συστήματος κύριας/κρυφής μνήμης



(α) Λανθάνουσα μνήμη



(β) Κύρια μνήμη

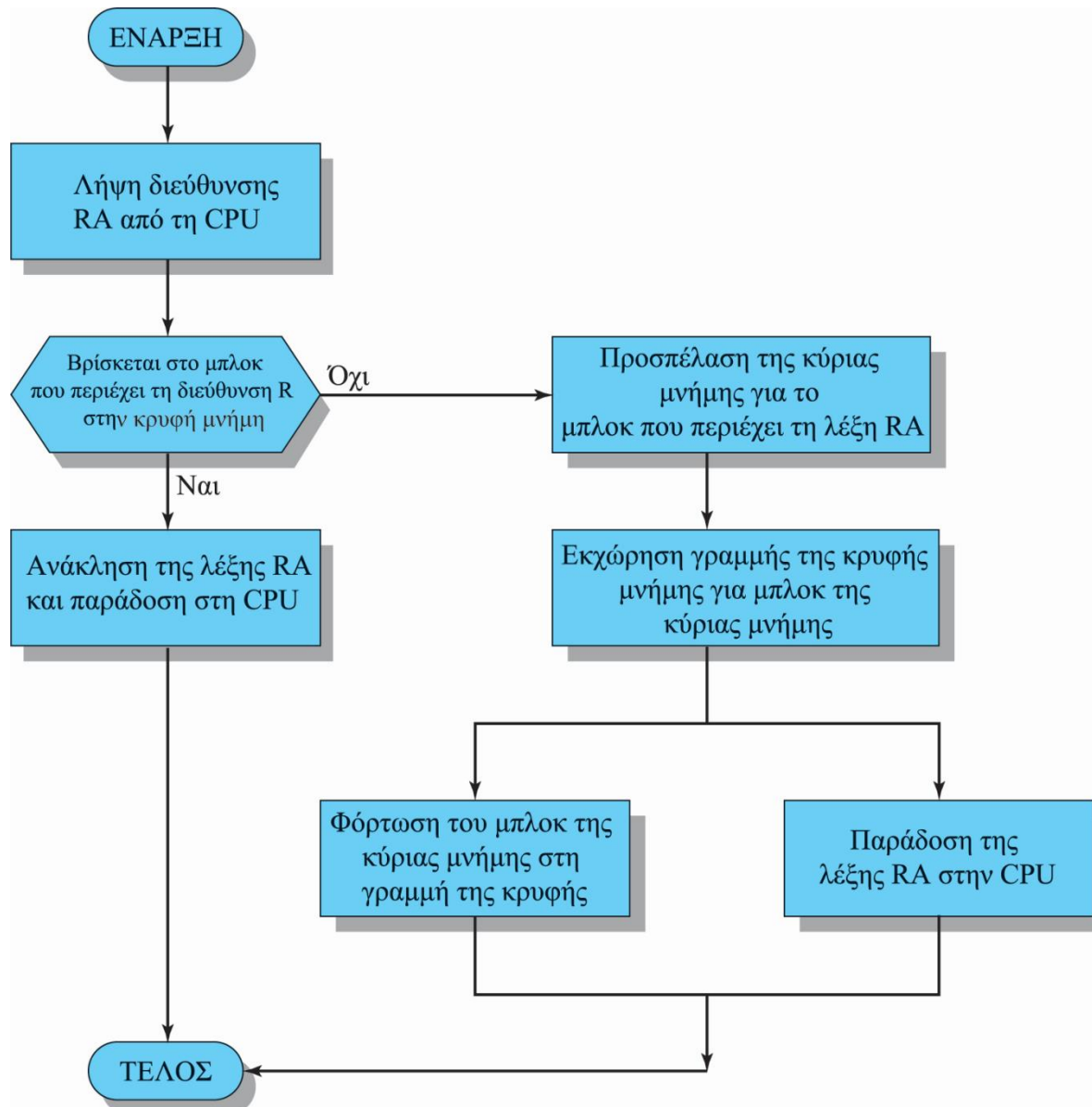
Δομή συστήματος κύριας/κρυφής μνήμης

- Η κύρια μνήμη αποτελείται από 2^n διευθυνσιοδοτούμενες λέξεις.
- Κάθε λέξη έχει μια μοναδική διεύθυνση n -bits.
- Για εξυπηρέτηση της αντιστοίχισης, η μνήμη θεωρείται ότι αποτελείται από ένα πλήθος μπλοκ σταθερού μήκους με **K** λέξεις το καθένα, δηλ. $2^n = N * K$, όπου **N** : αριθμός των μπλόκ
- Η κρυφή μνήμη αποτελείται από **m** μπλοκ που ονομάζονται **γραμμές**.
- Κάθε γραμμή περιέχει K λέξεις και μια ετικέτα που περιλαμβάνει ένα μικρό πλήθος bits για τη διεύθυνση του μπλοκ και για έλεγχο.
- Έλεγχος μεταβολής περιεχομένου της λέξης.

Δομή συστήματος κύριας/κρυφής μνήμης

- Το μήκος μιας γραμμής (χωρίς τα bits ετικέτας και ελέγχου) είναι το **μέγεθος γραμμής**.
- Μπορεί να είναι μόλις 32 bits με κάθε «λέξη» ίση με 1 byte. Επομένως το μέγεθος γραμμής = 4 bytes
- Το πλήθος των γραμμών είναι σημαντικά μικρότερο από τον αριθμό των μπλοκ της κύριας μνήμης.
- Κάθε μπλοκ περιλαμβάνει μια ετικέτα για να γνωρίζει ο επεξεργαστής ποιο συγκεκριμένο μπλοκ βρίσκεται στην κρυφή μνήμη την τρέχουσα χρονική στιγμή.
- Η ετικέτα είναι μέρος της διεύθυνσης της κύριας μνήμης.
- Τα μπλοκ, ανάλογα με τη ζήτησή τους από τον επεξεργαστή, εισάγονται στην κρυφή μνήμη από την κύρια μνήμη.

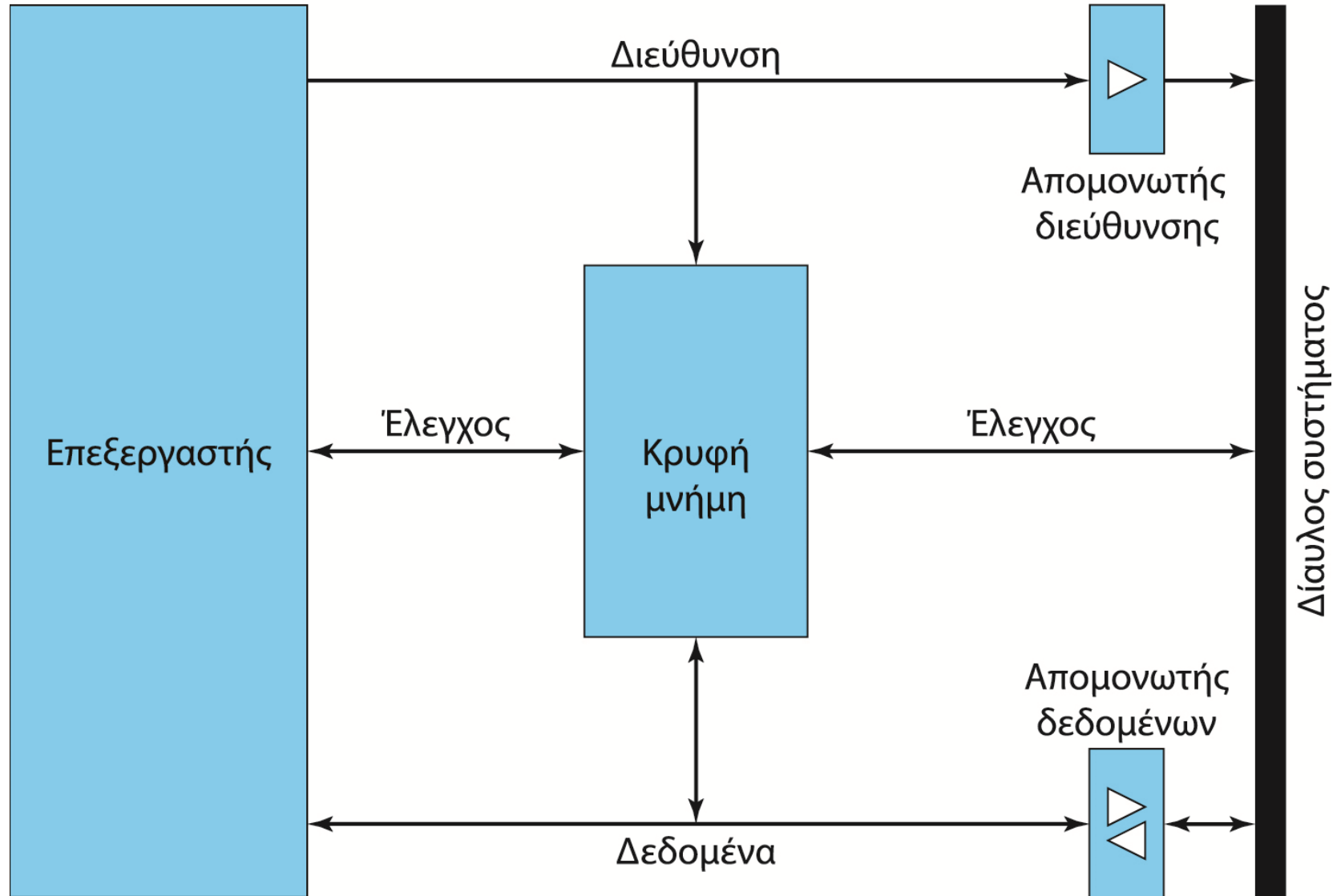
Λειτουργία ανάγνωσης κρυφής μνήμης



Λειτουργία ανάγνωσης κρυφής μνήμης

- Ο επεξεργαστής παράγει την διεύθυνση ανάγνωσης (Read Address – RA).
- Αν η λέξη περιέχεται στην κρυφή μνήμη παραδίδεται στον επεξεργαστή.
- Σε διαφορετική περίπτωση το μπλοκ ο οποίο περιέχει αυτή τη λέξη φορτώνεται στην κρυφή μνήμη και η λέξη παραδίδεται στον επεξεργαστή.
- Η κρυφή μνήμη συνδέεται με τον επεξεργαστή μέσα από γραμμές δεδομένων, ελέγχου και διευθύνσεων.
- Οι γραμμές δεδομένων και διευθύνσεων συνδέονται με απομονωτές δεδομένων και διευθύνσεων οι οποίοι συνδέονται μέσω διαύλου με την κύρια μνήμη.

Τυπική Οργάνωση Cache



Λειτουργία ανάγνωσης κρυφής μνήμης

- Στην περίπτωση της ευστοχίας, οι απομονωτές απενεργοποιούνται και η επικοινωνία γίνεται αποκλειστικά ανάμεσα στην κρυφή μνήμη και τον επεξεργαστή χωρίς να υπάρχει κυκλοφορία στο δίαυλο.
- Στην περίπτωση αστοχίας, η επιθυμητή διεύθυνση φορτώνεται στο δίαυλο και τα δεδομένα επιστρέφονται μέσω του απομονωτή δεδομένων τόσο στη κρυφή μνήμη όσο και στον επεξεργαστή.
- Σε άλλες μορφές οργάνωσης, η κρυφή μνήμη παρεμβάλλεται με φυσικό τρόπο ανάμεσα στον επεξεργαστή και την κύρια μνήμη για όλες τις γραμμές.
- Σε περίπτωση αστοχίας σε αυτή την οργάνωση, πρώτα λαμβάνει χώρα η αποθήκευση στην κρυφή μνήμη και μετά από αυτή στον επεξεργαστή.

Αρχιτεκτονική και Οργάνωση Υπολογιστών

Κεφάλαιο 4.3

Στοιχεία σχεδίασης της κρυφής μνήμης

Διευθύνσεις κρυφής μνήμης	Πολιτική εγγραφής
Λογική	Διά μέσου
Φυσική	Πίσω εγγραφής
Μέγεθος κρυφής μνήμης Λειτουργία αντιστοίχισης	Μέγεθος γραμμής Πλήθος κρυφών μνημών
Άμεση	Ενός ή δύο επιπέδων
Συσχετιστική	Ενοποιημένη ή διαχωρισμένη
Συσχέτισης συνόλων	
Αλγόριθμος αντικατάστασης	
Λιγότερο πρόσφατα χρησιμοποιημένα (LRU)	
Πρώτη μέσα πρώτη έξω (FIFO)	
Λιγότερο συχνά χρησιμοποιούμενη (LFU)	
Τυχαία	

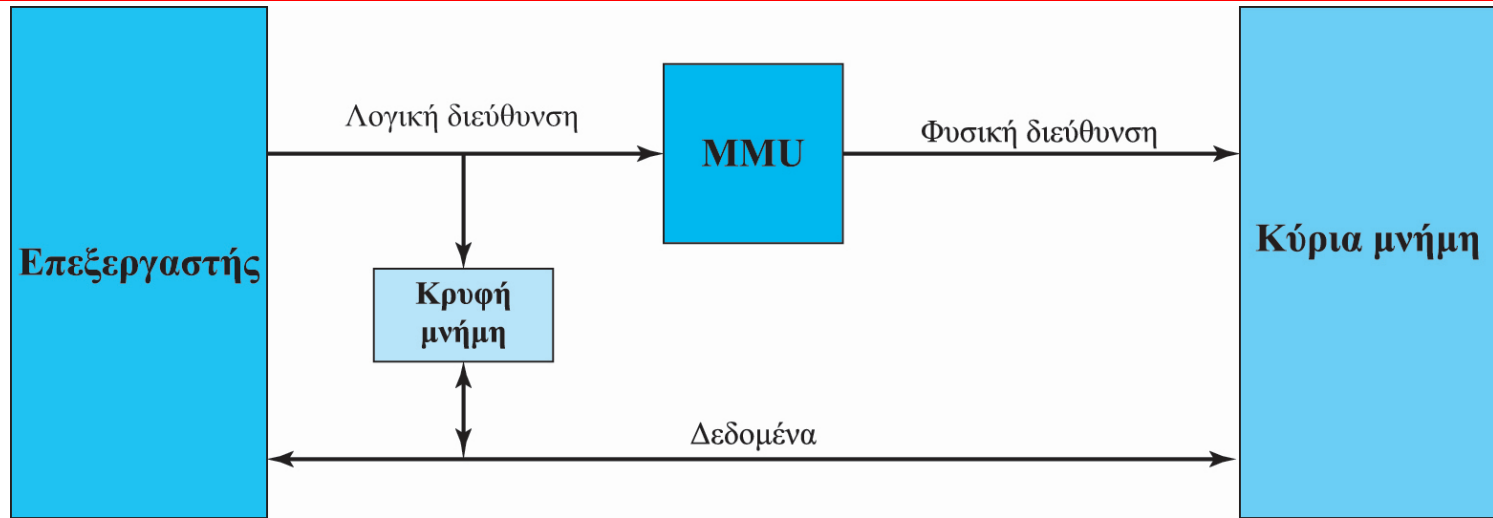
Στοιχεία σχεδίασης της κρυφής μνήμης

- Διευθύνσεις
- Μέγεθος
- Λειτουργία Απεικόνισης
- Αλγόριθμος Αντικατάστασης
- Πολιτική Εγγραφής
- Μέγεθος γραμμής
- Πλήθος Κρυφών Μνημών

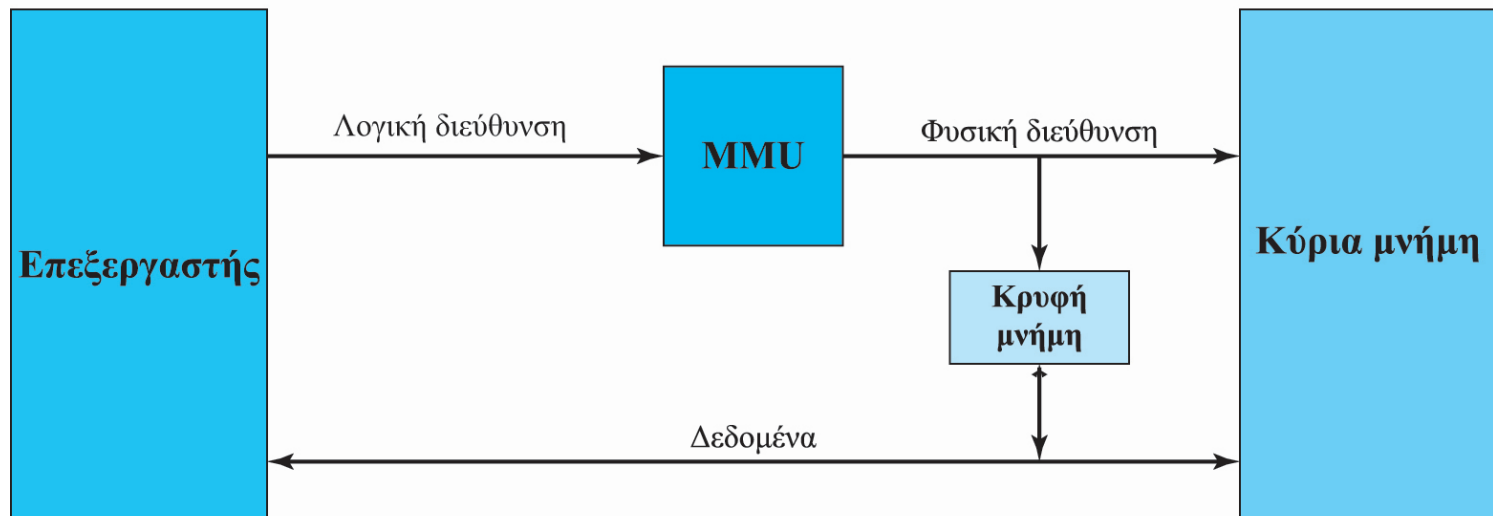
Διευθύνσεις κρυφής μνήμης

- Σχεδόν όλοι οι επεξεργαστές υποστηρίζουν την ιδεατή μνήμη.
- Η ιδεατή μνήμη εξυπηρετεί τα προγράμματα, επιτρέποντάς τα να αναφερθούν στην κύρια μνήμη από λογική πλευρά, χωρίς να ενδιαφέρονται για την ποσότητα μνήμης η οποία είναι φυσικά διαθέσιμη.
- Για να γίνει ανάγνωση/εγγραφή προς την κύρια μνήμη μια Μονάδα Διαχείρισης της Μνήμης (Memory Management Unit – MMU) μεταφράζει κάθε ιδεατή διεύθυνση σε φυσική διεύθυνση της κύριας μνήμης.
- Όταν χρησιμοποιούνται ιδεατές διευθύνσεις η κρυφή μνήμη τοποθετείται ανάμεσα στον επεξεργαστή και την MMU ή ανάμεσα στην MMU και την κύρια μνήμη.

Διευθύνσεις κρυφής μνήμης



(α) Λογική κρυφή μνήμη



(β) Φυσική κρυφή μνήμη

Διευθύνσεις κρυφής μνήμης

- Μια λογική κρυφή μνήμη (logical cache) γνωστή και ως ιδεατή κρυφή μνήμη (virtual cache) αποθηκεύει τα δεδομένα χρησιμοποιώντας ιδεατές διευθύνσεις. Ο επεξεργαστής προσπελαύνει την κρυφή μνήμη άμεσα χωρίς χρήση της MMU.
- Μια φυσική κρυφή μνήμη αποθηκεύει τα δεδομένα χρησιμοποιώντας τις φυσικές διευθύνσεις της κύριας μνήμης.
- Η λογική κρυφή μνήμη έχει μεγαλύτερη ταχύτητα προσπέλασης. Μειονέκτημα ότι δίνουν σε κάθε εφαρμογή τον ίδιο χώρο διευθύνσεων ιδεατής μνήμης. Κάθε εφαρμογή βλέπει μια ιδεατή μνήμη που αρχίζει από τη διεύθυνση 0. Σε 2 εφαρμογές η ίδια ιδεατή διεύθυνση αναφέρεται σε 2 διαφορετικές φυσικές διευθύνσεις. Πρέπει να προστεθούν επιπλέον bits για διαχωρισμό.

Μέγεθος Κρυφής Μνήμης

- Κόστος
 - Επιθυμούμε να είναι αρκετά μικρή ώστε το συνολικό μέσο κόστος/bit να είναι σχεδόν ίδιο με της κύριας
- Ταχύτητα
 - Αρκετά μεγάλη ώστε ο συνολικός μέσος χρόνος προσπέλασης να είναι σχεδόν ίδιος με της κρυφής μόνο
 - Ο έλεγχος της cache για την ύπαρξη των δεδομένων απαιτεί κάποιο χρόνο..
- Οι μεγάλες κρυφές μνήμες έχουν μεγάλο πλήθος πυλών για τη διευθυνσιοδότησή τους και έτσι τείνουν να είναι πιο αργές.
- Είναι αδύνατον να καταλήξουμε σε ένα «βέλτιστο» μέγεθος της κρυφής μνήμης.

Μεγέθη κρυφής μνήμης ορισμένων επεξεργαστών

Processor	Type	Year of Introduction	L1 Cachea	L2 cache	L3 Cache
IBM 360/85	Mainframe	1968	16 to 32 kB	—	—
PDP-11/70	Minicomputer	1975	1 kB	—	—
VAX 11/780	Minicomputer	1978	16 kB	—	—
IBM 3033	Mainframe	1978	64 kB	—	—
IBM 3090	Mainframe	1985	128 to 256 kB	—	—
Intel 80486	PC	1989	8 kB	—	—
Pentium	PC	1993	8 kB/8 kB	256 to 512 KB	—
PowerPC 601	PC	1993	32 kB	—	—
PowerPC 620	PC	1996	32 kB/32 kB	—	—
PowerPC G4	PC/server	1999	32 kB/32 kB	256 KB to 1 MB	2 MB
IBM S/390 G6	Mainframe	1999	256 kB	8 MB	—
Pentium 4	PC/server	2000	8 kB/8 kB	256 KB	—
IBM SP	High-end server/ supercomputer	2000	64 kB/32 kB	8 MB	—
CRAY MTA ^b	Supercomputer	2000	8 kB	2 MB	—
Itanium	PC/server	2001	16 kB/16 kB	96 KB	4 MB
Itanium 2	PC/server	2002	32 kB	256 KB	6 MB
IBM POWER5	High-end server	2003	64 kB	1.9 MB	36 MB
CRAY XD-1	Supercomputer	2004	64 kB/64 kB	1MB	—
IBM POWER6	PC/server	2007	64 kB/64 kB	4 MB	32 MB
IBM z10	Mainframe	2008	64 kB/128 kB	3 MB	24-48 MB
Intel Core i7 EE 990	Workstaton/ server	2011	6 × 32 kB/32 kB	1.5 MB	12 MB
IBM zEnterprise 196	Mainframe/ Server	2011	24 × 64 kB/ 128 kB	24 × 1.5 MB	24 MB L3 192 MB L4

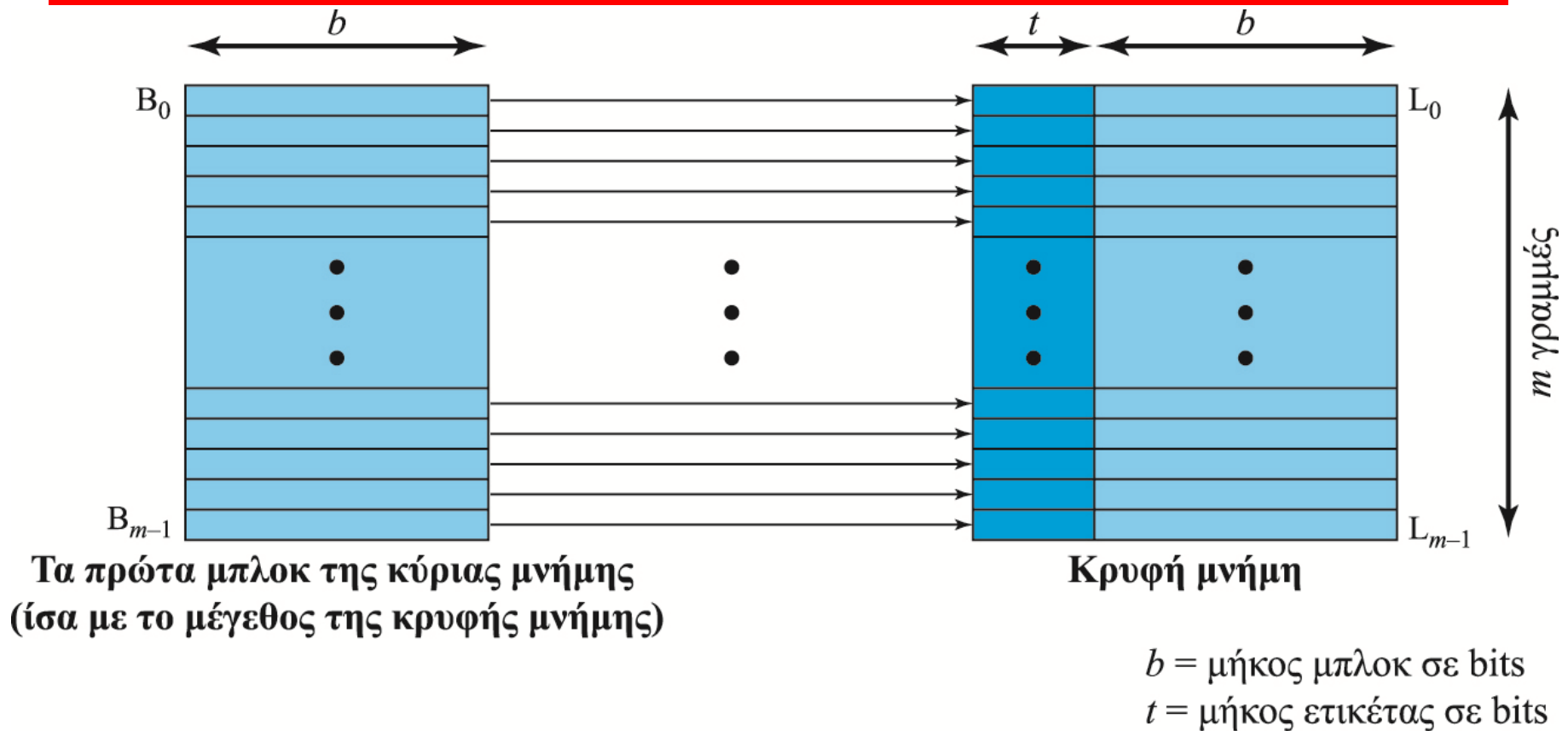
Η Λειτουργία της Απεικόνισης

- Cache των 64kBytes
- Cache block των 4 bytes
 - άρα η cache αποτελείται 16k (2^{14}) γραμμές των 4 bytes
- 16MBytes Κύρια Μνήμη ($2^{24} = 16\text{M}$)
- 24 bit λέξη διεύθυνσης
- Η κύρια μνήμη αποτελείται από
 - 4M μπλοκ μεγέθους 4 bytes

Η Λειτουργία της Απεικόνισης

- Υπάρχουν λιγότερες γραμμές κρυφής σε σχέση με μπλοκ κύριας. Είναι αναγκαίο ένας αλγόριθμος απεικόνισης μπλοκ της κύριας σε γραμμές της κρυφής.
- Είναι αναγκαίος ένας τρόπος προσδιορισμού του μπλοκ της κύριας το οποίο καταλαμβάνει μια γραμμή της κρυφής την τρέχουσα στιγμή.
- Η επιλογή του τρόπου απεικόνισης καθορίζει και την οργάνωση της κρυφής μνήμης.
- Υπάρχουν τρεις τεχνικές απεικόνισης:
 - 1) Άμεση
 - 2) Συσχετιστική
 - 3) Συσχέτισης συνόλων

Άμεση Απεικόνιση



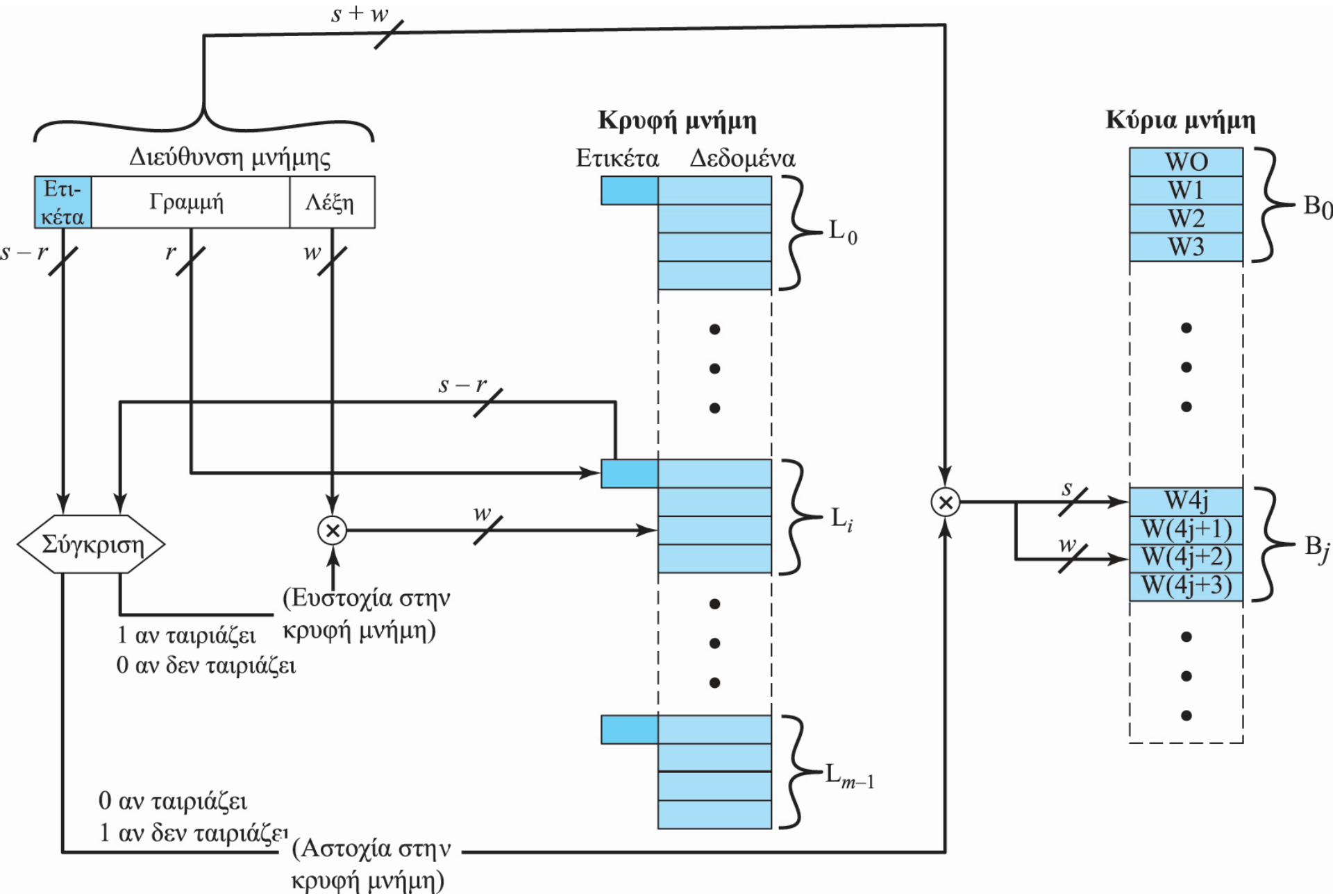
(α) Άμεση απεικόνιση

Άμεση Απεικόνιση

- Η πιο απλή τεχνική όπου αντιστοιχεί κάθε μπλοκ της κύριας μνήμης με μία μόνο γραμμή της κρυφής. Εκφράζεται ως:

$$i = j \text{ modulo } m$$

- i = αριθμός γραμμής της κρυφής μνήμης
 - j = αριθμός μπλοκ της κύριας μνήμης
 - m = πλήθος των γραμμών της κρυφής μνήμης
-
- Το μπλοκ B_m της κύριας μνήμης απεικονίζεται στη γραμμή L_0 , το B_{m+} στη γραμμή L_1 κ.ο.κ.

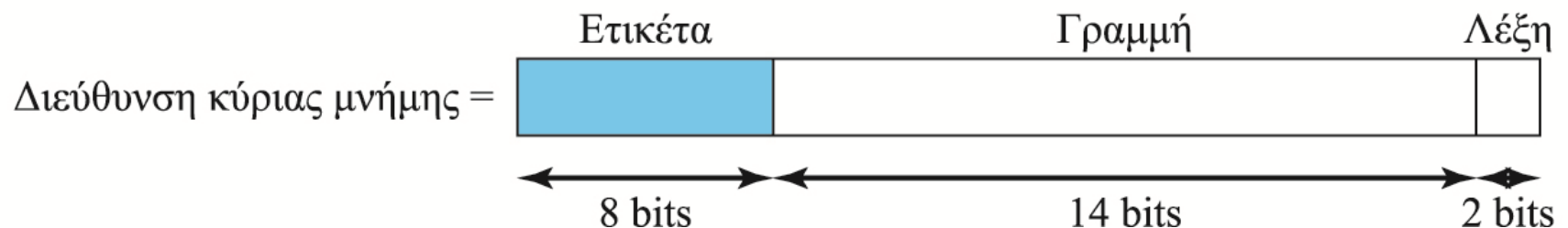


Άμεση Απεικόνιση

- Μήκος διεύθυνσης = $(s+w)$ bits
- Πλήθος διευθυνσιοδοτούμενων μονάδων = 2^{s+w} λέξεις ή bytes
- Μέγεθος μπλοκ = μέγεθος γραμμής = 2^w λέξεις ή bytes
- Πλήθος μπλοκ της κύριας μνήμης = $2^{s+w} / 2^w = 2^s$
- Πλήθος γραμμών της κρυφής μνήμης = $m = 2^r$
- Μέγεθος της κρυφής μνήμης = 2^{s+r} λέξεις ή bytes
- Μέγεθος ετικέτας = $(s-r)$ bits
- Κύρια μνήμη = $2^{s-r} * \text{Κρυφή μνήμη}$

Άμεση Απεικόνιση

- 24 bit Διεύθυνση
- 2 bit δείκτης λέξης (4 byte block)
- 22 bit δείκτης block
 - 8 bit ετικέτα (=22-14)
 - 14 bit γραμμή
- Δεν υπάρχουν δύο Block στην ίδια γραμμή με την ίδια ετικέτα
- Έλεγχος περιεχομένων Cache με εύρεση γραμμής και έλεγχο ετικέτας



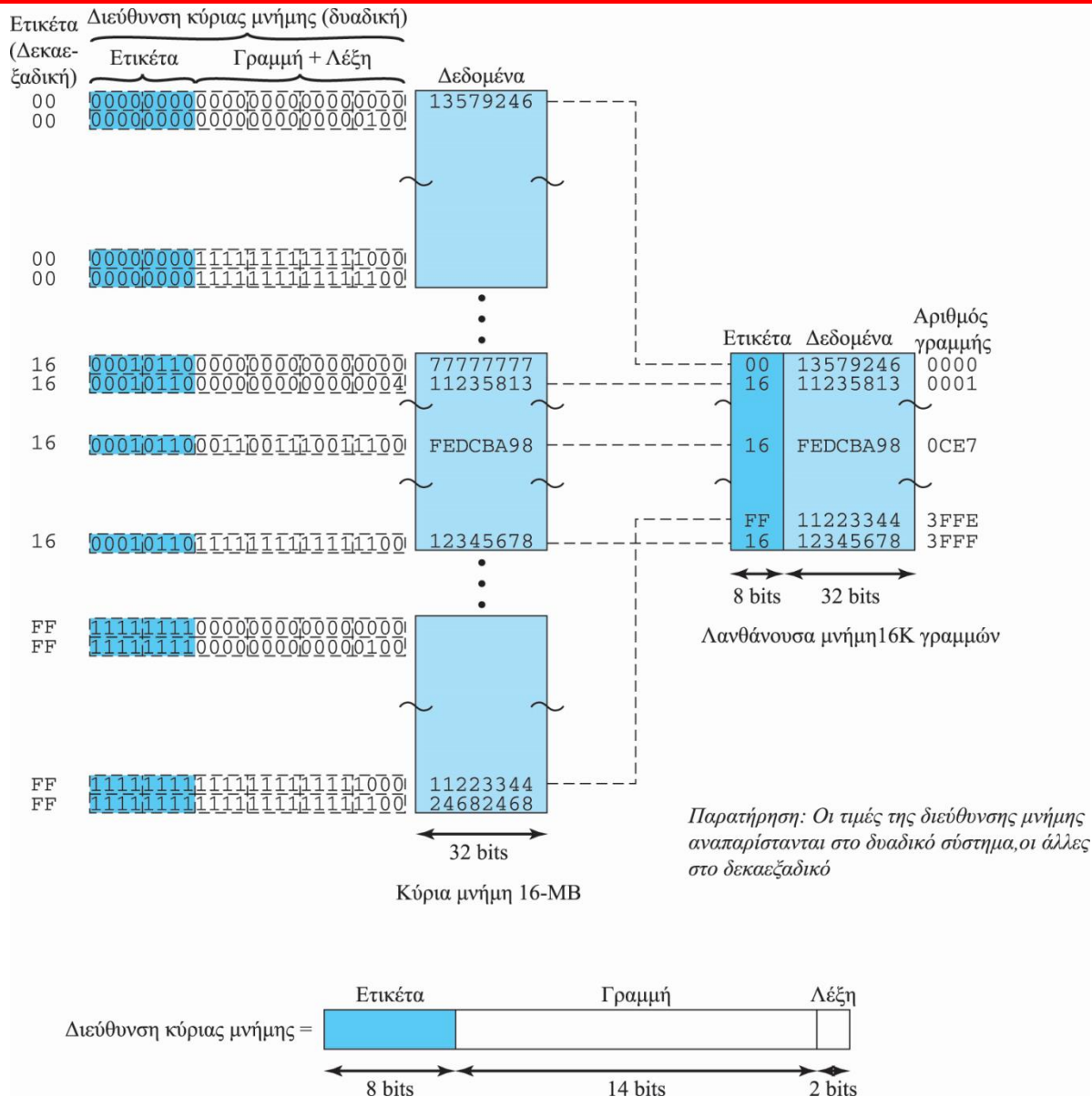
Άμεση Απεικόνιση

- Κάθε τμήμα (block) της κύριας μνήμης αντιστοιχεί σε μια συγκεκριμένη γραμμή της cache
 - Άρα, αν ένα τμήμα είναι μέσα στην cache, πρέπει να βρίσκεται σε μια συγκεκριμένη θέση
- Η διεύθυνση διακρίνεται σε δύο μέρη
- Τα λιγότερο σημαντικά (LSB) w bits δείχνουν σε μια μοναδική λέξη
- Τα περισσότερα σημαντικά (MSB) s bits δείχνουν ένα μοναδικό τμήμα μνήμης
- Τα s MSBs με την σειρά τους διακρίνονται σε ένα πεδίο γραμμών cache r και σε μια ετικέτα που περιλαμβάνει τα $s-r$ MSBs

Άμεση Απεικόνιση

- | | |
|--------------|-------------------------------|
| • Cache line | Main Memory blocks held |
| • 0 | 0, m, 2m, 3m... $2^s - m$ |
| • 1 | 1, m+1, 2m+1... $2^s - m + 1$ |
| • m-1 | m-1, 2m-1, 3m-1... $2^s - 1$ |

Άμεση Απεικόνιση- Παράδειγμα



Παράδειγμα άμεσης απεικόνισης

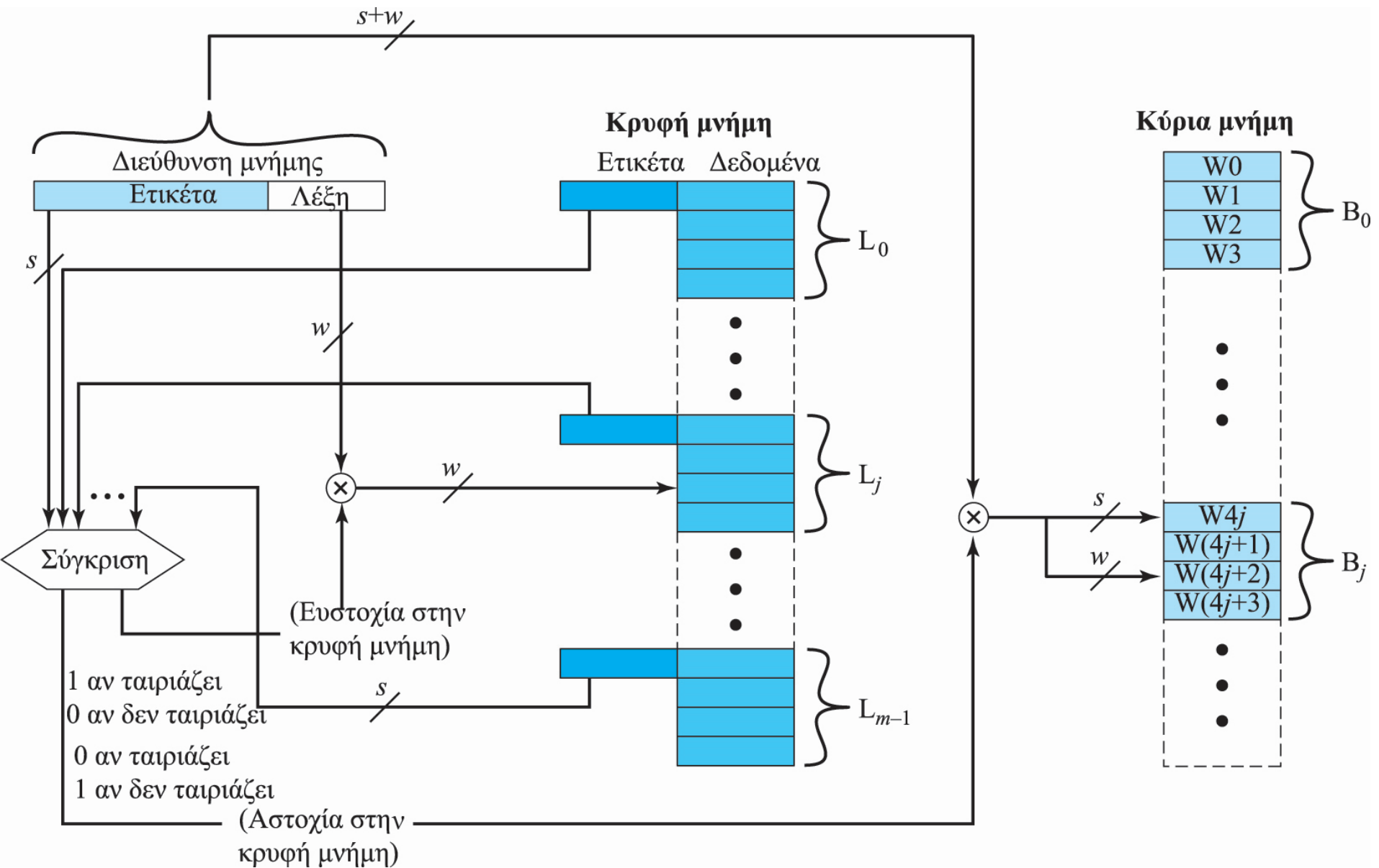
- $m=2^{14}$, $i=j \text{ modulo } 2^{14}$ για κρυφή μνήμη.
- η αντιστοίχιση είναι :
- Γραμ. Κρ. Μν. Διευθ. Μν. Έν. Μπλόκ
- 0 000000, 010000,...,FF0000
- 1 000004, 010004,...,FF0004
-
-
- $2^{14} - 1$ 00FFFC, 01FFFC,...,FFFFFC

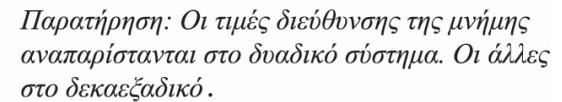
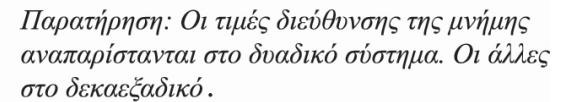
Άμεση Απεικόνιση— Πλεονεκτήματα & Μειονεκτήματα

- Απλή
- Φθηνή υλοποίηση
- Μείον ότι υπάρχει για κάθε μπλοκ μια σταθερή θέση της κρυφής μνήμης
 - Αν ένα πρόγραμμα αναφέρεται επαναληπτικά σε λέξεις δύο διαφορετικών μπλοκ, οι οποίες αντιστοιχίζονται στην ίδια γραμμή τότε τα μπλοκ θα εναλλάσσονται συνεχώς μέσα στην κρυφή μνήμη τότε ο λόγος επιτυχών ευρέσεων θα είναι χαμηλός → μειωμένη απόδοση (το φαινόμενο ονομάζεται αλώνισμα-thrashing)

Συσχετιστική Απεικόνιση

- Κάθε μπλοκ της Κ.Μ. μπορεί να «φορτωθεί» σε οποιαδήποτε γραμμή της cache
- Η διεύθυνση μνήμης ερμηνεύεται ως πεδίο ετικέτας και πεδίο λέξης.
- Η ετικέτα προσδιορίζει μοναδικά ένα τμήμα της κύριας μνήμης
- Η ετικέτα κάθε γραμμής στην cache ελέγχεται για πιθανό ταίριασμα (για να βρεθεί αν αυτό το μπλοκ βρίσκεται στην κρυφή)
- Ο έλεγχος της Cache αποδεικνύεται ακριβός





Συσχετιστική Απεικόνιση

- 22 bit ετικέτα αποθηκεύεται μαζί με 32 bit τμήμα δεδομένων
- Το πεδίο της ετικέτας συγκρίνεται με κάθε ετικέτα στην cache
- Τα LSB 2 bits διεύθυνσης αναγνωρίζουν ποια 8 bit λέξη απαιτείται από το 32 bit block δεδομένων
- e.g.

—Address	Tag	Data	Cache line
—FFFFFFC	3FFFFFF	24682468	3FFF

Συσχετιστική Απεικόνιση

- Μήκος διεύθυνσης = $(s + w)$ bits
- Αριθμός διευθυνσιοδοτούμενων μονάδων = 2^{s+w} λέξεις ή bytes
- Μέγεθος Block = Μέγεθος γραμμών = 2^w λέξεις ή bytes
- Αριθμός blocks στην Κ.Μ. = $2^{s+w}/2^w = 2^s$
- Αριθμός γραμμών στην cache = απροσδιόριστο
- Μέγεθος ετικέτας = s bits

Συσχετιστική Απεικόνιση— Πλεονεκτήματα & Μειονεκτήματα

- Υπάρχει ευελιξία σχετικά με το μπλοκ που θα αντικατασταθεί όταν ένα νέο τοποθετηθεί στην κρυφή.
- Έχουν σχεδιαστεί αλγόριθμοι αντικατάστασης με σκοπό τη μεγιστοποίηση του λόγου ευστοχίας.
- Μειονέκτημα είναι ότι απαιτούνται πολύπλοκα κυκλώματα για να εξεταστούν παράλληλα οι ετικέτες όλων των γραμμών της κρυφής μνήμης.

Απεικόνιση Συσχέτισης Συνόλων

- Είναι μια μορφή που εμφανίζει τα πλεονεκτήματα της άμεσης και της συσχετιστικής απεικόνισης ενώ παράλληλα μειώνει και τα μειονεκτήματά τους.
- Η κρυφή αποτελείται από ένα πλήθος συνόλων καθένα από τα οποία περιέχει έναν πλήθος γραμμών.
- Ένα δοσμένο block μπορεί να αντιστοιχηθεί σε κάθε γραμμή ενός δοσμένου συνόλου
 - π.χ. Το Block B μπορεί να τοποθετηθεί σε κάθε γραμμή του συνόλου i
- π.χ. 2 γραμμές
 - Συσχετιστική Απεικόνιση Συνόλων Δύο Γραμμών
 - Ένα δοσμένο block μπορεί να είναι σε μια από τις δύο γραμμές ενός μόνο συνόλου

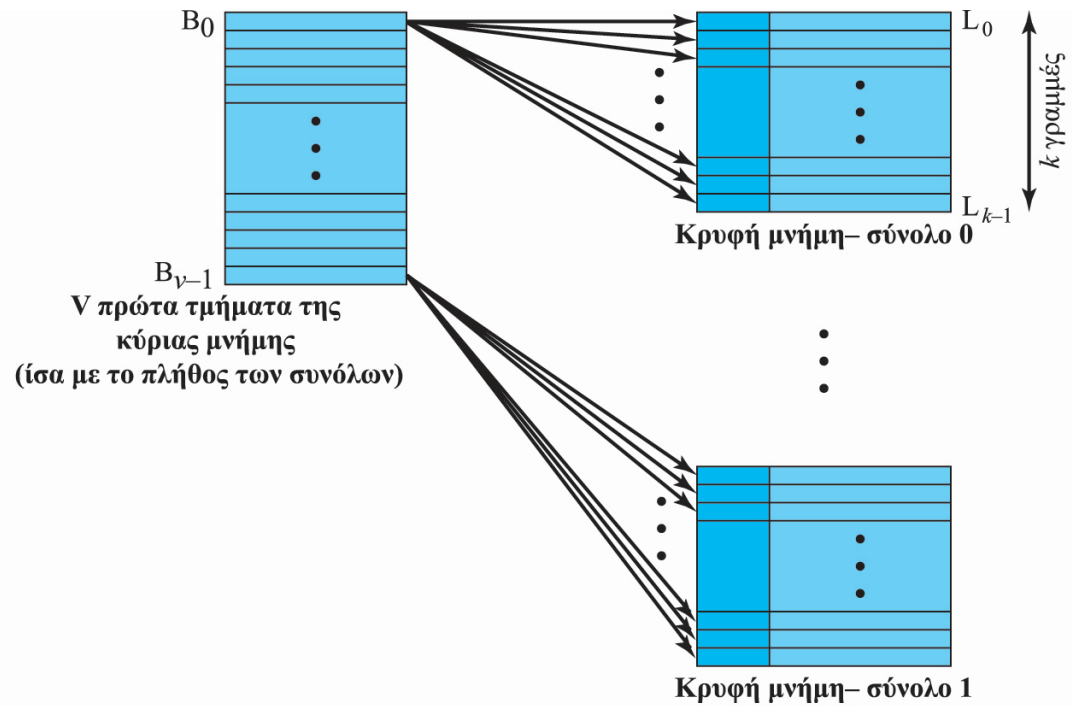
Απεικόνιση Συσχέτισης Συνόλων

- Εκφράζεται ως:

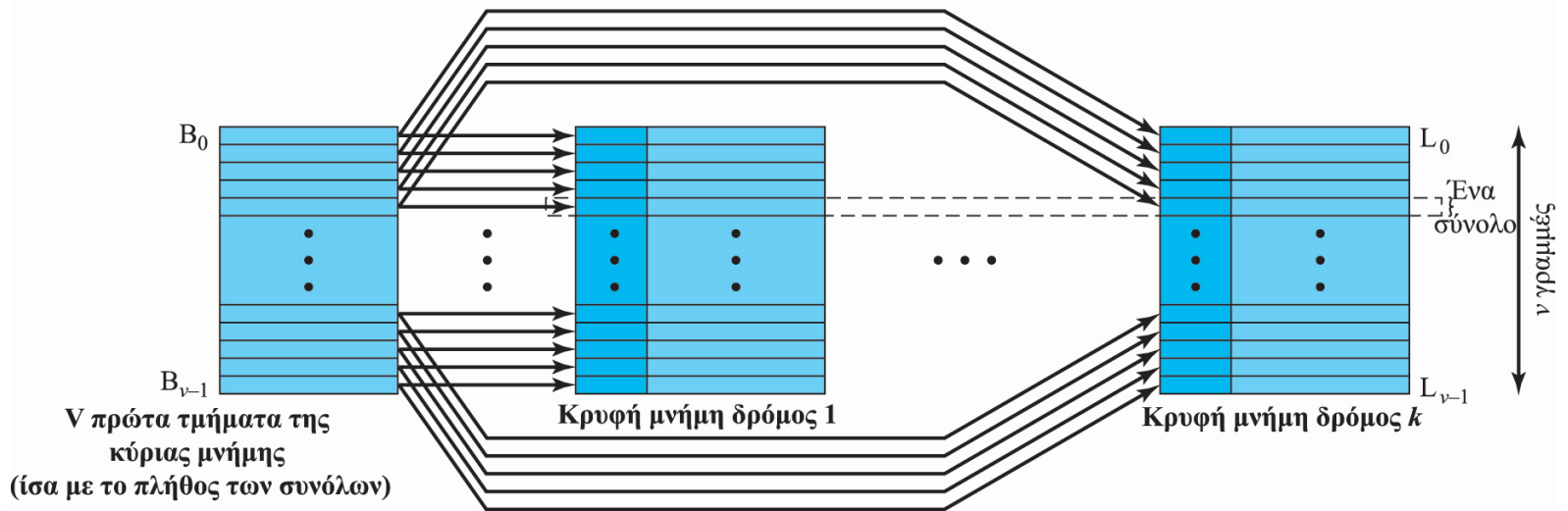
$$m = v * k$$

$$n = j \text{ modulo } v$$

- n = αριθμός συνόλου της κρυφής μνήμης
 - j = αριθμός μπλοκ της κύριας μνήμης
 - m = πλήθος των γραμμών της κρυφής μνήμης
 - v = πλήθος των συνόλων
 - k = πλήθος των γραμμών κάθε συνόλου
-
- Αναφέρεται και ως απεικόνιση συσχέτισης συνόλων k -δρόμων.
 - Το μπλοκ B_j μπορεί να απεικονιστεί σε οποιαδήποτε από τις γραμμές του συνόλου j .



(α) ν κρυφές μνήμες συσχετιστικής απεικόνισης



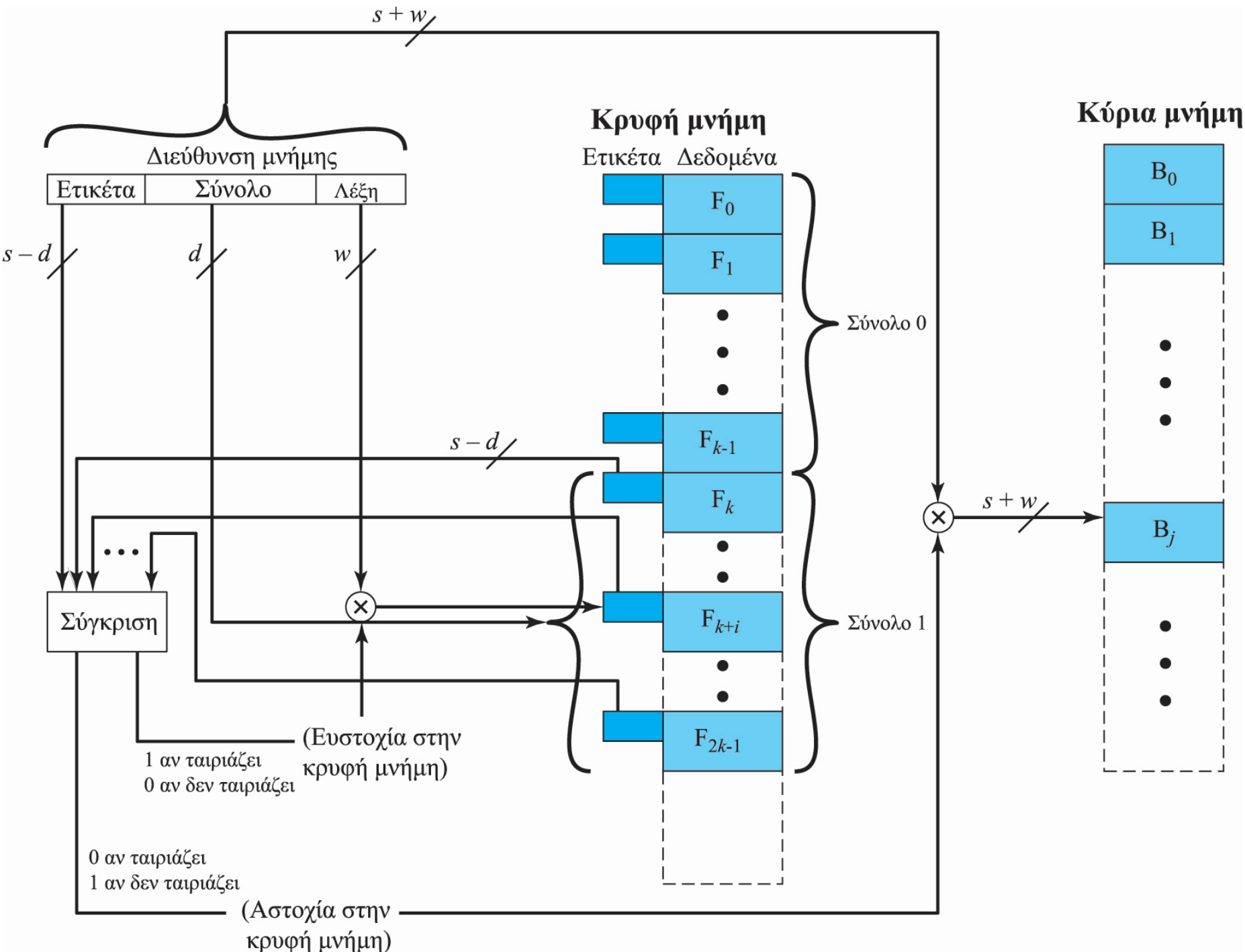
(β) k κρυφές μνήμες άμεσης απεικόνισης

Απεικόνιση Συσχέτισης Συνόλων

- Όπως και στη συσχετιστική απεικόνιση κάθε λέξη απεικονίζεται σε πολλαπλές γραμμές τις κρυφής μνήμης.
- Στην απεικόνιση συσχέτισης συνόλων κάθε λέξη απεικονίζεται σε όλες τις γραμμές της κρυφής μνήμης σε ένα συγκεκριμένο σύνολο, έτσι ώστε το τμήμα B_0 της κύριας να απεικονίζεται στο σύνολο 0 κ.ο.κ.
- Είναι δυνατόν να υλοποιηθεί με φυσικό τρόπο σαν ένα σύνολο ν συσχετιστικών κρυφών μνημών.
- Είναι δυνατόν να την υλοποιήσουμε σαν ένα σύνολο από k κρυφές μνήμες άμεσης απεικόνισης. Κάθε κρυφή άμεσης απεικόνισης αναφέρεται και ως *δρόμος* αποτελούμενος από ν γραμμές.
- Η υλοποίηση άμεσης απεικόνισης χρησιμοποιείται για μικρούς βαθμούς συσχέτισης (μικρές τιμές k).

Απεικόνιση Συσχέτισης Συνόλων

- Η υλοποίηση συσχετιστικής απεικόνισης χρησιμοποιείται για μεγαλύτερους βαθμούς συσχέτισης.
- Η λογική ελέγχου της κρυφής μνήμης ερμηνεύει μια διεύθυνση ως ένα σύνολο τριών πεδίων: Ετικέτα, Σύνολο και Λέξη.
- Τα d σε πλήθος bits συνόλου προσδιορίζουν ένα από τα $v = 2^d$ σύνολα.
- Τα s bits των πεδίων Ετικέτας και Συνόλου προσδιορίζουν ένα από τα 2^s μπλοκ της κύριας μνήμης.
- Η ετικέτα μιας διεύθυνσης μνήμης είναι πολύ πιο μικρή (σε σχέση με την συσχετιστική απεικόνιση) και συγκρίνεται μόνο με τις k ετικέτες εντός ενός μόνο συνόλου.



Απεικόνιση Συσχέτισης Συνόλων

- Μήκος Διεύθυνσης = $(s + w)$ bits
- Αριθμός διευθυνσιοδοτούμενων μονάδων = 2^{s+w} λέξεις ή bytes
- Μέγεθος Block = Μέγεθος Γραμμής = 2^w λέξεις ή bytes
- Πλήθος μπλοκ της κύριας μνήμης = $2^{s+w} / 2^w = 2^s$
- Πλήθος γραμμών στο σύνολο = k (από το σχεδιαστή)
- Πλήθος Συνόλων = $v = 2^d$ (από το σχεδιαστή)
- Πλήθος των γραμμών της cache = $m = v = k * 2^d$
- Μέγεθος της κρυφής μνήμης = $k * 2^d + w$ λέξεις ή bytes
- Μέγεθος ετικέτας = $(s - d)$ bits

Απεικόνιση Συσχέτισης Συνόλων

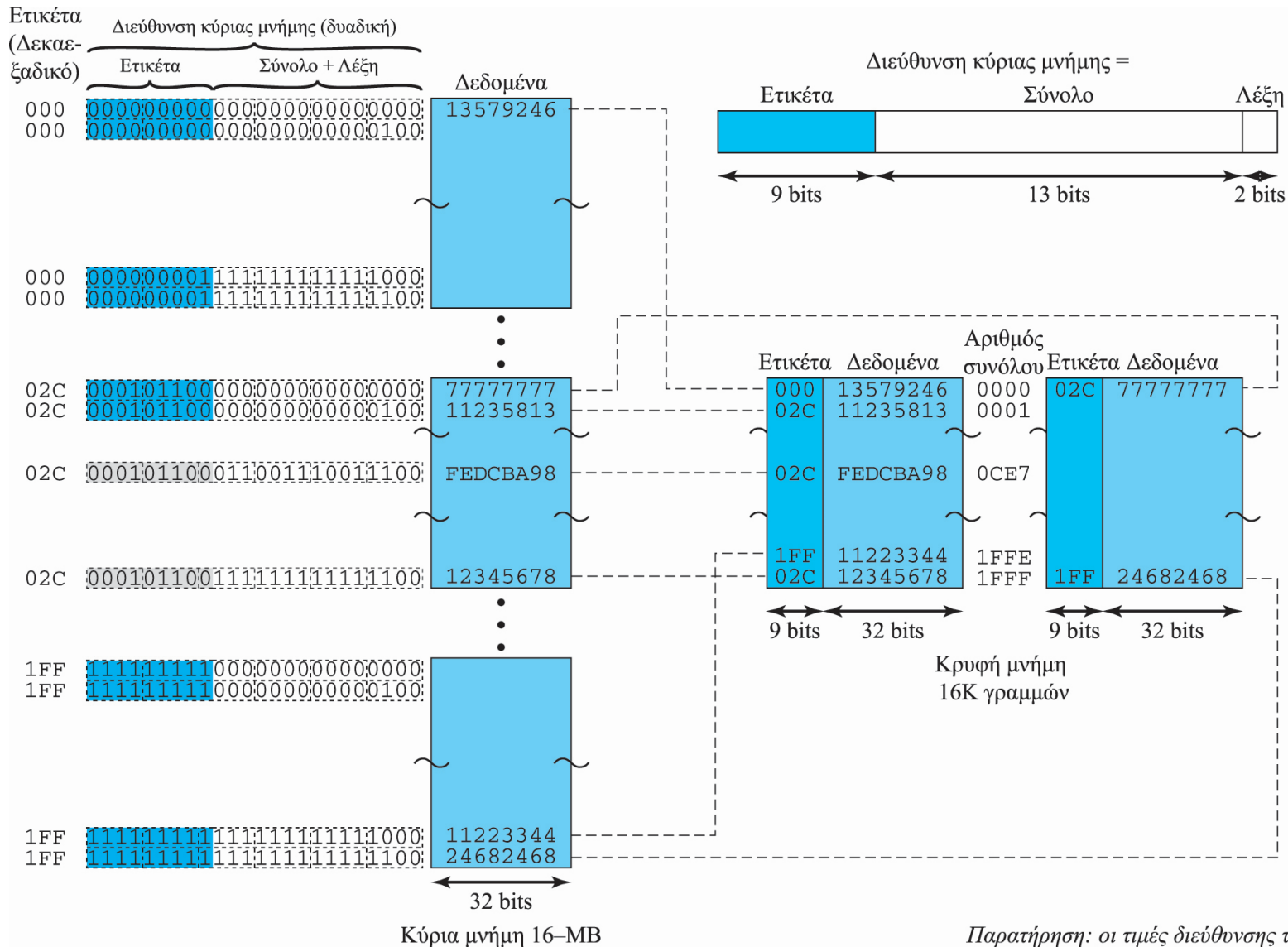
- Στην ακραία περίπτωση όπου $n=m$, $k=1$ η τεχνική ανάγεται σε εκείνη της άμεσης απεικόνισης.
- Για $n=1$, $k=m$ ανάγεται στη συσχετιστική.
- Η χρήση 2 γραμμών ανά σύνολο ($n=m/2$, $k=2$) αποτελεί την πιο συνηθισμένη οργάνωση συσχέτισης συνόλων.
- Βελτιώνει σημαντικά τον λόγο επιτυχίας σε σύγκριση με την άμεση απεικόνιση.
- Η χρήση 4 δρόμων επιτυγχάνει μια μέτρια περεταίρω βελτίωση με μικρό επιπλέον κόστος. Περαιτέρω αυξήσεις του πλήθους των γραμμών ανά σύνολο έχουν ελάχιστα αποτελέσματα.

Απεικόνιση Συσχέτισης Συνόλων

- Χρησιμοποιούμε το πεδίο συνόλου για να αποφασίσουμε πιο cache set θα εξετάσουμε
- Συγκρίνουμε την ετικέτα για πετυχημένη εύρεση
- π.χ

—Address	Tag	Data	Set number
—1FF 7FFC	1FF	12345678	1FFF
—02C 7FF8	02C	11223344	1FFF

Απεικόνιση Συσχέτισης Συνόλων- Παράδειγμα 2 δρόμων



Παρατήρηση: οι τιμές διεύθυνσης της μνήμης αντιπαρίστανται στο δυαδικό σύστημα. Οι άλλες τιμές στο δεκαεξαδικό.

