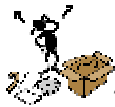




Ενότητα 3: Η Απόδοση της κρυφής μνήμης

Σκοπός Ο σκοπός της ενότητας αυτής είναι να εξετάσει τους παράγοντες που επηρεάζουν την απόδοση της κρυφής μνήμης και να παρουσιάσει τις τεχνικές που χρησιμοποιούνται για να μειώσουμε τις αποτυχίες, την ποινή αποτυχίας και το χρόνο επιτυχίας της κρυφής μνήμης

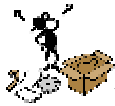
Προσδοκώμενα Αποτελέσματα Όταν θα έχετε μελετήσει την ενότητα, θα είστε σε θέση να:



υπολογίζετε την απόδοση της κρυφής μνήμης,



προτείνετε τρόπους για τη βελτίωση της απόδοσης της κρυφής μνήμης,



κατηγοριοποιείτε τις αποτυχίες της κρυφής μνήμης ανάλογα με την αιτία δημιουργίας τους,



περιγράφετε τις τεχνικές μείωσης των αποτυχιών, του ρυθμού αποτυχίας και της ποινής αποτυχίας της κρυφής μνήμης,



προσδιορίζετε τον τρόπο μείωσης του χρόνου επιτυχίας,



μέσος χρόνος πρόσβασης, χρόνος επιτυχίας, ρυθμός αποτυχίας, ποινή αποτυχίας, αναγκαστικές αποτυχίες, αποτυχία χωρητικότητας, αποτυχία σύγκρουσης, κρυφή μνήμη θύμα, υπο-μπλοκ, δεικτοδότηση κρυφής μνήμης



Η απόδοση της κρυφής μνήμης

Τόσο το πλήθος των εντολών όσο και ο ρυθμός αποτυχίας είναι ανεξάρτητα από το υλικό και την ταχύτητά του. Γι' αυτό το λόγο έχουν χρησιμοποιηθεί για να εκτιμηθεί η απόδοση της ιεραρχίας μνήμης. Οι έμμεσοι αυτοί τρόποι μέτρησης της απόδοσης έχουν προβληματίσει τους σχεδιαστές των ηλεκτρονικών υπολογιστών, με κατάληξη το καλύτερο μέτρο για την απόδοση της κρυφής μνήμης να θεωρείται ο μέσος χρόνος πρόσβασης στη μνήμη (τύπος 4.6):

$$\text{Μέσος χρόνος πρόσβασης στη μνήμη} = \text{χρόνος επιτυχίας} \times \text{ρυθμός αποτυχίας} + \text{ποινή αποτυχίας}$$

(τύπος 4.6)

Όπου :

- **χρόνος επιτυχίας** είναι ο χρόνος πρόσβασης στη κρυφή μνήμη που είναι επιτυχής, δηλαδή όταν τα δεδομένα που ζητάμε βρίσκονται στη κρυφή μνήμη.
- **ρυθμός αποτυχίας** είναι το κλάσμα των προσβάσεων στη μνήμη για δεδομένα που δε βρίσκονται στο ανώτερο επίπεδο της ιεραρχίας. Δηλαδή είναι, τα δεδομένα που ζητάμε και δε βρίσκονται στην κρυφή μνήμη, προς το σύνολο των προσβάσεων.
- **ποινή αποτυχίας** είναι ο επιπρόσθετος χρόνος εξυπηρέτησης μιας αποτυχίας.



Ο μέσος χρόνος πρόσβασης στη μνήμη αποτελεί και αυτός έμμεσο μέτρο της απόδοσης, αλλά πολύ καλύτερο απ' τον ρυθμό αποτυχίας. Ωστόσο σε καμία περίπτωση δεν αντικαθιστά το χρόνο εκτέλεσης.



ΔΡΑΣΤΗΡΙΟΤΗΤΑ 1

Θυμάστε τον τύπο ο οποίος μας δίνει την απόδοση της ΚΜΕ σε σχέση με το χρόνο μιας επιτυχίας ή μιας αποτυχίας της κρυφής μνήμης; Να καταγράψετε τον τύπο αυτό και να τον συγκρίνετε με τον τύπο 4.7 που ακολουθεί. Για περισσότερες λεπτομέρειες να ανατρέξετε στην 2^η ενότητα του κεφαλαίου 1 και συγκεκριμένα στην ενότητα: «Η τοπικότητα της αναφοράς και η απόδοση της ΚΜΕ».

$$\text{Χρόνος ΚΜΕ} = (\text{Κύκλοι Ρολογιού εκτέλεσης ΚΜΕ} + \text{Κύκλοι ρολογιού καθυστέρησης μνήμης}) \times \text{χρόνος κύκλου ρολογιού}$$

(τύπος 4.7)



Οι σχεδιαστές στην προσπάθειά τους να απλοποιήσουν τις εναλλακτικές επιλογές τους στον υπολογισμό της απόδοσης της κρυφής μνήμης, συχνά υποθέτουν ότι οι καθυστερήσεις της μνήμης οφείλονται κυρίως στις αποτυχίες της κρυφής μνήμης. Η υπόθεσή τους ευσταθεί, αφού στην ιεραρχία μνήμης κύριοι υπαίτιοι για τις καθυστερήσεις, είναι άλλοι, όπως ο ισχυρισμός για τις συσκευές Εισόδου/Εξόδου που χρησιμοποιούν τη μνήμη. Ωστόσο, είναι σημαντικό να λαμβάνουμε υπόψη μας όλες τις καθυστερήσεις της μνήμης στον υπολογισμό της τελικής απόδοσης του συστήματος.

Ο τύπος (4.7) για το χρόνο της ΚΜΕ, εισάγει ερωτήματα, όπως "πότε οι κύκλοι ρολογιού για επιτυχία της κρυφής μνήμης λαμβάνονται υπόψη στους κύκλους ρολογιού εκτέλεσης ΚΜΕ και πότε στους κύκλους ρολογιού καθυστέρησης μνήμης; ". Και οι δύο συμβάσεις έχουν τους υποστηρικτές τους, ωστόσο ισχυρότερη παρουσιάζεται η πρώτη θέση. Οι κύκλοι ρολογιού καθυστέρησης μνήμης προσδιορίζονται απ' τον αριθμό των προσβάσεων στη μνήμη ανά πρόγραμμα, την ποινή αποτυχίας (σε κύκλους ρολογιού) και του ρυθμού αποτυχίας για τις αναγνώσεις και τις εγγραφές (τύπος 4.8):

$$\text{Κύκλοι ρολογιού Καθυστέρησης μνήμης} = \text{Αναγνώσεις} \times \text{Ρυθμός Αποτυχίας Ανάγνωσης} \times \text{Ποινή Αποτυχίας Ανάγνωσης} + \text{Εγγραφές} \times \text{Ρυθμός Αποτυχίας Εγγραφής} \times \text{Ποινή Αποτυχίας Εγγραφής} \quad (\text{τύπος 4.8})$$

Αν συνδυάσουμε τις εγγραφές και τις αναγνώσεις και υπολογίζοντας το μέσο ρυθμό πρόσβασης, τη μέση ποινή αποτυχίας για αναγνώσεις και εγγραφές, ο τύπος 4.8 απλοποιείται στην μορφή του τύπου 4.9:

$$\text{Κύκλοι ρολογιού Καθυστέρησης μνήμης} = \text{Προσπελάσεις στη μνήμη} \times \text{Ρυθμός Αποτυχίας} \times \text{Ποινή Αποτυχίας} \quad (\text{τύπος 4.9})$$

Ο τύπος 4.9 προσεγγίζει τους κύκλους ρολογιού καθυστέρησης της μνήμης, αφού ο ρυθμός αποτυχίας και η ποινή αποτυχίας είναι διαφορετικός για τις εγγραφές και τις αναγνώσεις.

Αν στα δεδομένα γίνεται αναφορά στο πλήθος εντολών, στο χρόνο εκτέλεσης και στους κύκλους για την καθυστέρηση της μνήμης, υπολογίζουμε το χρόνο της ΚΜΕ με τον τύπο 4.10 που περιλαμβάνει τις προσπελάσεις της μνήμης ανά εντολή, το ρυθμό αποτυχίας και την ποινή αποτυχίας:

$$\text{Χρόνος ΚΜΕ} = IC \times (CP \cdot I_{\text{Εκτέλεσης}} + \text{Προσπελάσεις Μνήμης ανά εντολή} \times \text{ρυθμός αποτυχίας} \times \text{ποινή αποτυχίας}) \times \text{Χρόνος κύκλου ρολογιού} \quad (\text{τύπος 4.10})$$

Όπου :

CPI : κύκλοι ρολογιού ανά εντολή

IC : πλήθος εντολών



Κάποιοι σχεδιαστές στηριζόμενοι στην ισότητα :

$$\frac{\text{Αποτυχίες}}{\text{Εντολή}} = \frac{\text{Προσπελάσεις στη μνήμη} \times \text{Ρυθμός αποτυχίας}}{\text{Εντολή}}$$

έχουν μία ακόμα επιλογή ανεξάρτητη από την υλοποίηση του υλικού. Το μειονέκτημα εντοπίζεται στην εξάρτηση του λόγου *Αποτυχίες / Εντολή*, απ' την αρχιτεκτονική του υπολογιστικού συστήματος. Για παράδειγμα, ο μέσος όρος προσπελάσεων στη μνήμη διαφέρει για έναν 80x86 και για έναν DLX.

Καταλήγουμε στο τύπο 4.11, για πιο βελτιωμένη εκτίμηση της απόδοσης της κρυφής μνήμης:

$$\text{Χρόνος KME} = \text{IC} \times (\text{CP I}_{\text{Εκτέλεσης}} + \text{Κύκλοι ρολογιού καθυστέρησης Μνήμης ανά εντολή}) \times \text{Χρόνος κύκλου ρολογιού}$$

(τύπος 4.11)



Παράδειγμα

Θεωρούμε ένα υπολογιστικό σύστημα στο οποίο ισχύουν τα ακόλουθα:

Ποινή αποτυχίας = 50 κύκλοι ρολογιού,

Όλες οι εντολές εκτελούνται σε 2.0 κύκλοι ρολογιού (αγνοώντας τις καθυστερήσεις),

Ρυθμός αποτυχίας = 2% και,

Αποτυχίες

Εντολή = 1,33 : Μέση τιμή αναφορών στη μνήμη ανά εντολή

Ποιο είναι το κόστος στην απόδοση λαμβάνοντας υπόψη μας τη συμπεριφορά της κρυφής μνήμης;

Απάντηση:

Από τον τύπο 4.11, βάση του οποίου υπολογίζεται ο Χρόνος εκτέλεσης της KME έχω:

$$\text{Χρόνος KME} = \text{IC} \times (\text{CP I}_{\text{Εκτέλεσης}} + \text{Κύκλοι ρολογιού καθυστέρησης Μνήμης ανά εντολή}) \times \text{Χρόνος κύκλου ρολογιού}$$

Η Απόδοση συμπεριλαμβάνοντας τις αποτυχίες κρυφής μνήμης, δίνεται από τον τύπο:

$$\begin{aligned} \text{Χρόνος Εκτέλεσης KME} + \text{Κρυφής μνήμης} &= \text{IC} \times [2.0 + (1,33 \times 2 \% \times 50)] \times \text{Χρόνος κύκλου Ρολογιού} \\ &= \text{IC} \times 3,33 \times \text{Χρόνος κύκλου ρολογιού} \end{aligned}$$

Η παρουσία της κρυφής μνήμης δεν επηρεάζει τον Χρόνο κύκλου ρολογιού ή το πλήθος των εντολών. Κατά συνέπεια ο χρόνος της KME αυξάνεται με το CPI από 2.0, για μία ιδανική κρυφή μνήμη, σε 3.33 για μία κρυφή μνήμη με αποτυχίες. Αν συμπεριλάβω στους υπολογισμούς του CPI και την ιεραρχία μνήμης, παρατηρώ ότι ο χρόνος της KME

αυξάνει κατά έναν παράγοντα 1.67. Αν απ' την άλλη αγνοήσω την ιεραρχία μνήμης, το CPI αυξάνεται κατά : $2.0 + 50 \times 1.33 = 68.5$, δηλαδή γίνεται 30 φορές μεγαλύτερη. 🚩

Η επίδραση της συμπεριφοράς της κρυφής μνήμης στην Απόδοση έγινε φανερή μέσα από το παράδειγμα. Συμπεραίνουμε λοιπόν ότι με χαμηλό CPI και γρήγορο ρολόι, οι αποτυχίες της κρυφής μνήμης έχουν διττή επιρροή στη ΚΜΕ:

- Όσο μικρότερο το $CPI_{\text{Εκτέλεσης}}$, τόσο περισσότερο επηρεάζεται ένας συγκεκριμένος αριθμός κύκλου ρολογιού για αποτυχία κρυφής μνήμης.
- Στον υπολογισμό του CPI η ποινή αποτυχίας της κρυφής μνήμης προσμετράται στους κύκλους ρολογιού της ΚΜΕ για μία αποτυχία.. Ωστόσο, ακόμα και αν έχω την ίδια ιεραρχία μνήμης για δύο συστήματα, η ΚΜΕ με τον μεγαλύτερο ρυθμό ρολογιού, θα έχει μεγαλύτερο αριθμό κύκλων ρολογιού ανά αποτυχία, έτσι και η αναλογία του CPI στη μνήμη είναι υψηλότερη.

Ένας εύλογος σκοπός αποτελεί η μείωση του μέσου χρόνου προσπέλασης στη μνήμη, ωστόσο ας έχουμε κατά νου ότι ο απώτερος και τελικός σκοπός μας είναι να μειώσουμε το χρόνο εκτέλεσης της ΚΜΕ.



Παράδειγμα

Ποια είναι η επιρροή σε δύο διαφορετικές οργανώσεις κρυφής μνήμης στην επίδοση μιας ΚΜΕ; Υποθέστε ότι η CPI μιας τέλει κρυφής μνήμης είναι 2.0 και ο κύκλος ρολογιού 2 ns, η αναφορά μνήμης είναι 1.3 ανά εντολή, το μέγεθος και των δύο κρυφών μνημών είναι 64Kb ενώ και οι δύο έχουν ένα μπλοκ μεγέθους 32 bytes. Η μία κρυφή μνήμη είναι άμεσης αντιστοίχισης, ενώ η άλλη 2-δρόμων συνολο-συσχετιστική, η ποινή αποτυχίας κρυφής μνήμης είναι 70 ns και για τις δύο περιπτώσεις. Πρώτα υπολογίστε το Μέσο χρόνο προσπέλασης στη μνήμη και έπειτα την Απόδοση της ΚΜΕ. Υποθέστε ότι το χρόνος επιτυχίας είναι ένα κύκλος ρολογιού και ότι το ρυθμός αποτυχίας του πρώτου 1.4%, ενώ του δευτέρου 1%.

Απάντηση:

Γνωρίζουμε ότι:

$$\text{Μέσος χρόνος προσπέλασης στη μνήμη} = \text{Χρόνος επιτυχίας} + \text{Ρυθμός αποτυχίας} \times \text{Ποινή αποτυχίας}$$

Οπότε για τις δύο κρυφές μνήμες έχουμε:

$$\text{Μέσος χρόνος προσπέλασης στη μνήμη}_{1\text{-δρόμου}} = 2.0 + (0.014 \times 70) = 2.98\text{ns}$$

$$\text{Μέσος χρόνος προσπέλασης στη μνήμη}_{2\text{-δρόμου}} = 2.0 \times 1.10 + (0.010 \times 70) = 2.9\text{ns}$$

Συνεπώς ο Μέσος χρόνος προσπέλασης στη μνήμη είναι καλύτερος στη 2-δρόμων συνολο-συσχετιστική

Τώρα έχουμε:

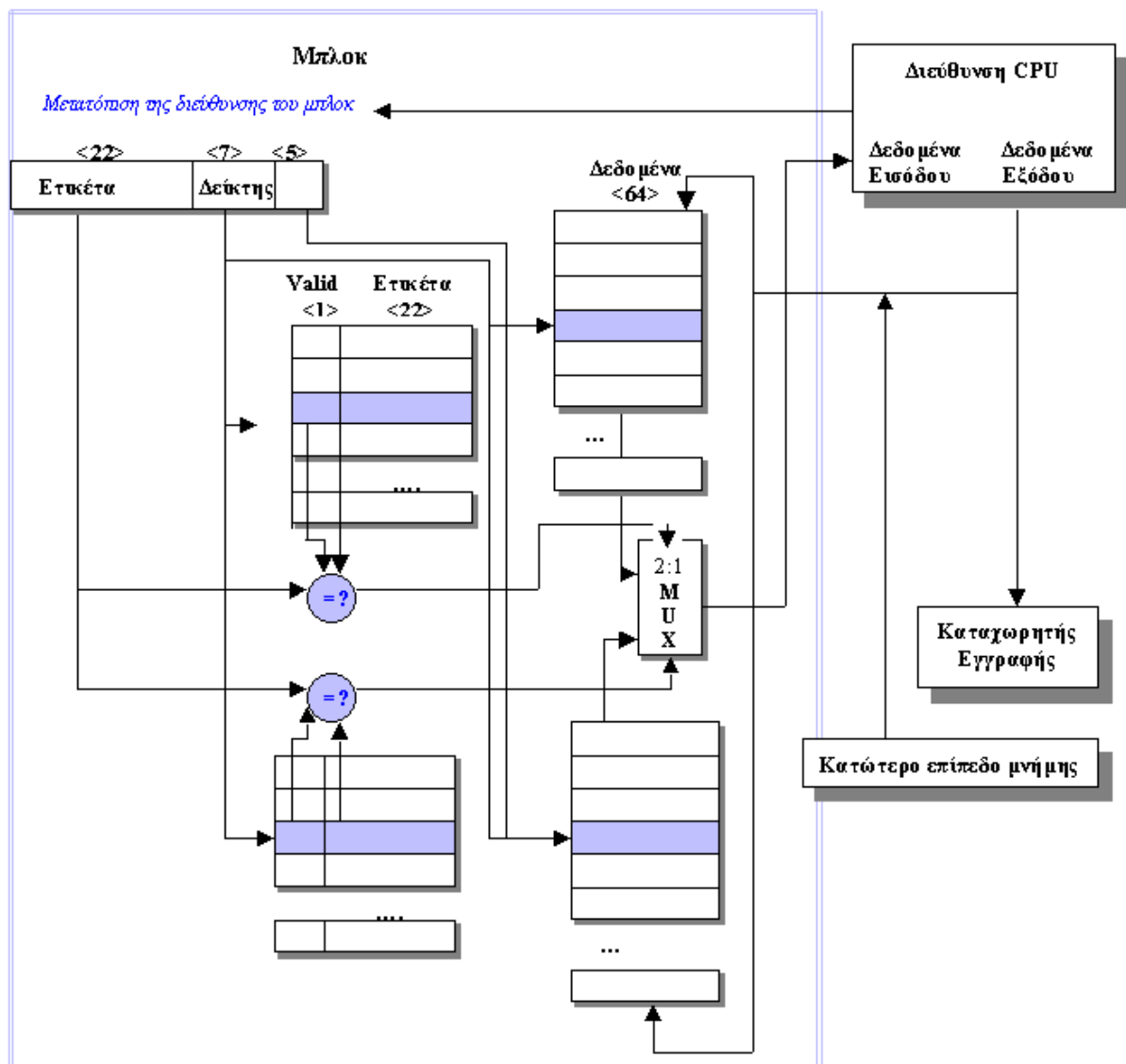
$$\begin{aligned} \text{ΚΜΕ}_{\text{χρόνος}} &= \text{ICx}(\text{CPI}_{\text{εκτέλεσης}} + (\text{Αποτυχίες/εντολές}) * \text{ποινή αποτυχίας}) \times \text{Χρόνος} \\ &\quad \text{κύκλου ρολογιού} \\ &= \text{ICx}(\text{CPI}_{\text{εκτέλεσης}} \times \text{Χρόνος κύκλου ρολογιού}) + [(\text{Προσπελάσεις στη μνήμη} \\ &\quad \text{/εντολές}) \times \text{Ρυθμός αποτυχίας} \times \text{Χρόνος κύκλου ρολογιού}] \end{aligned}$$

Επειδή $70 \text{ ns} = \text{Ρυθμός αποτυχίας} \times \text{Χρόνος κύκλου ρολογιού}$ έχουμε:


$$\mathbf{KME}_{\chi\rho\acute{o}\nu\omicron\varsigma} (1\text{-}\delta\rho\acute{o}\mu\omicron\upsilon) = \mathbf{IC}*(2.0*2 + (1,3*0,014*70)) = 5,27*\mathbf{IC}$$

$$\text{KME}_{\text{χρόνος}} (2\text{-δρόμων}) = \text{IC} * (2.0 * 2 * 1,10 + (1,3 * 0,010 * 70)) = 5,31 * \text{IC}$$

$$\frac{\text{ΚΜΕ}_{\text{χρόνος}} (2\text{- δρόμων})}{\text{ΚΜΕ}_{\text{χρόνος}} (1\text{- δρόμου})} = 1,01$$



Σχήμα 4.3.1 – Μία 2-δρόμων συνολο-συσχετιστική κρυφή μνήμη 8KB με την προσθήκη ενός πολυπλέκτη. Τα δύο πιο αριστερά ψηφία της μετατόπισης του μπλοκ συνδυάζονται μέσω του πολυπλέκτη με τον δείκτη, για να διευθυνσιοδοτήσουν την λέξη των 64-bits στη μνήμη, η οποία στη συνέχεια στέλνεται στην ΚΜΕ

Σε αντίθεση με τα αποτελέσματα από την σύγκριση του μέσου χρόνου προσπέλασης στη μνήμη, η κρυφή μνήμη άμεσης αντιστοίχισης οδηγεί σε ελαφρώς καλύτερη μέση απόδοση. Αυτό οφείλεται στο ότι ο κύκλος ρολογιού συρρικνώνεται για να αντεπεξέλθει σε όλες τις εντολές στην περίπτωση των 2-δρόμων, ακόμα και για λιγότερες αποτυχίες. Απ' τη στιγμή που ο χρόνος της ΚΜΕ είναι η τελική μας εκτίμηση και η άμεση αντιστοίχιση είναι ευκολότερη στην υλοποίηση, η προαναφερόμενη κρυφή μνήμη θεωρείται άμεσης αντιστοίχισης. 



ΔΡΑΣΤΗΡΙΟΤΗΤΑ 2

Σύμφωνα με όσα γνωρίζετε (θεωρία, τύπους) για την απόδοση της κρυφής μνήμης, να προτείνετε τρόπους για τη βελτίωσή της. Να συγκρίνετε την απάντησή σας με την παράγραφο που ακολουθεί: «Βελτίωση της απόδοσης της κρυφής μνήμης».

Βελτίωση της απόδοσης της κρυφής μνήμης

Ο τύπος για τον υπολογισμό του μέσου χρόνου πρόσβασης στη μνήμη:

Μέσος χρόνος πρόσβασης στη μνήμη = χρόνος επιτυχίας + ρυθμός αποτυχίας x ποινή αποτυχίας

μας έδωσε ένα πλαίσιο εργασίας για να παρουσιάσουμε τις βελτιστοποιήσεις στη κρυφή μνήμη αλλά και τις διάφορες τεχνικές βελτίωσης της λειτουργίας της. Μπορούμε να τις εντάξουμε σε τρεις κατηγορίες :

- Μείωση του ρυθμού αποτυχίας
- Μείωση της ποινής αποτυχίας
- Μείωση του χρόνου επιτυχίας



Χρόνος επιτυχίας είναι ο χρόνος που μεσολαβεί για να συμβεί μία επιτυχία στη κρυφή μνήμη.



Η μείωση των αποτυχιών της κρυφής μνήμης

Για να μειωθεί ο ρυθμός αποτυχίας των προσβάσεων στην κρυφή μνήμη, θα πρέπει να αναζητηθούν οι αιτίες που τις προκαλούν. Οι αποτυχίες στην κρυφή μνήμη χωρίζονται σε τρεις κατηγορίες, ανάλογα με την αιτία δημιουργίας τους :

- **Αναγκαστικές ή αποτυχίες κρύας αρχής ή αποτυχίες πρώτης αναφοράς:** Η κρυφή μνήμη αρχικά είναι κενή, οπότε η πρώτη πρόσβαση σε αυτήν θα είναι αναγκαστικά ανεπιτυχής, αφού το μπλοκ δε βρίσκεται στη κρυφή μνήμη, και πρέπει να μεταφερθεί σε αυτή.
- **Χωρητικότητα:** Αν η κρυφή μνήμη δεν μπορεί να συμπεριλάβει όλα τα μπλοκ που χρειάζονται κατά τη διάρκεια της εκτέλεσης ενός προγράμματος, τότε έχουμε αποτυχίες χωρητικότητας. Αυτό συμβαίνει όταν ζητάμε κάποια μπλοκ το οποία δεν μεταφέρθηκαν στην κρυφή μνήμη ή αντικατεστάθηκαν και δε βρίσκονται πια στην κρυφή μνήμη.
- **Σύγκρουσης ή αποτυχίες θραύσης ή αποτυχίες παρεμβολής:** Αν η στρατηγική τοποθέτησης μπλοκ είναι συνολο-συσχετιστική ή άμεσης αντιστοίχισης, αποτυχίες σύγκρουσης θα συμβούν, επειδή ένα μπλοκ μπορεί να αποβληθεί και μετά να ανακτηθεί αν πολλά μπλοκ απεικονίζονται στο σύνολό του.

Οι αποτυχίες σύγκρουσης επιπλέον διακρίνονται σε 4 είδη αποτυχιών ανάλογα με τη μείωση της συσχετιστικότητας. Πιο συγκεκριμένα οι διακρίσεις αυτές είναι :

- 8-δρόμων : οι αποτυχίες σύγκρουσης οφείλονται στη μετάβαση από την πλήρως συσχετιστική (δεν έχω καμία σύγκρουση) στην 8-δρόμων συσχετιστική.
- 4-δρόμων : οι αποτυχίες σύγκρουσης οφείλονται στη μετάβαση από την 8-δρόμων συσχετιστική στην 4-δρόμων συσχετιστική.
- 2-δρόμων : οι αποτυχίες σύγκρουσης οφείλονται στη μετάβαση από την 4-δρόμων συσχετιστική στην 2-δρόμων συσχετιστική
- 1-δρόμων : οι αποτυχίες σύγκρουσης οφείλονται στη μετάβαση από την 2-δρόμων συσχετιστική στην 1-δρόμου συσχετιστική (άμεσης αντιστοίχισης).

Η πλήρως συσχετιστική τοποθέτηση αποφεύγει όλες τις αποτυχίες σύγκρουσης, είναι ακριβή στο υλικό, μπορεί να επιβραδύνει τον ρυθμό του ρολογιού του επεξεργαστή και οδηγεί σε χαμηλότερη ολική απόδοση.



Υπάρχουν λίγα πράγματα που μπορούμε να κάνουμε για τη χωρητικότητα, εκτός απ' το να μεγαλώσουμε την κρυφή μνήμη. Κάνοντας τα μπλοκ μεγαλύτερα, μειώνεται το πλήθος των αναγκαστικών αποτυχιών, επειδή μειώνεται και ο συνολικός αριθμός των μπλοκ που χωράνε στη μνήμη, αλλά μεγαλύτερα μπλοκ μπορούν να αυξήσουν άλλων ειδών αποτυχίες.



ΔΡΑΣΤΗΡΙΟΤΗΤΑ 3

Να κατηγοριοποιήσετε τις αποτυχίες της κρυφής μνήμης ανάλογα με την αιτία δημιουργίας τους. Να συγκρίνετε την απάντησή σας με την παράγραφο που μόλις διαβάσατε: «Η μείωση των αποτυχιών της κρυφής μνήμης».



Τεχνικές Μείωσης του Ρυθμού Αποτυχίας

Χρησιμοποιούμε συνολικά επτά τεχνικές προκειμένου να εξισορροπήσουμε τη μείωση του ρυθμού αποτυχίας και την επιτάχυνση όλου του υπολογιστικού συστήματος:

- **1^η -Μπλοκ μεγαλύτερου μεγέθους** : μειώνει τις αναγκαστικές αποτυχίες. Η μείωση αυτή οφείλεται στην τοπικότητα του χρόνου και κυρίως στη τοπικότητα του χώρου.
- **2^η - Μεγαλύτερες συσχετιστικότητες** : το τίμημα που πληρώνουμε για τη τεχνική αυτή είναι η αύξηση του χρόνου επιτυχίας.
- **3^η - Κρυφή μνήμη Θύμα** : από τις πιο πρόσφατες επινοήσεις χωρίς να επηρεάζεται ο χρόνος του κύκλου ρολογιού ή την ποινή αποτυχίας.
- **4^η - Ψευδοσυσχετιστική Κρυφή μνήμη** : συνδυάζουν το ρυθμό αποτυχίας της συνολο-συσχετιστικής οργάνωσης και τον χρόνο επιτυχίας της άμεσης αντιστοίχισης.
- **5^η - Προανάκληση εντολών και δεδομένων** : γίνεται προανάκληση εντολών και δεδομένων πριν ζητηθούν από τον επεξεργαστή είτε απευθείας στην κρυφή μνήμη, είτε σε έναν γρήγορο εξωτερικό καταχωρητή.
- **6^η - Προανάκληση εντολών ελεγχόμενη από Μεταγλωττιστή** : η προανάκληση των εντολών και των δεδομένων ελέγχεται από τον μεταγλωττιστή.
- **7^η - Βελτιστοποίηση λογισμικού** : χωρίς επεμβάσεις στο υλικό αλλά στο λογισμικό του συστήματος, επιτυγχάνεται η βελτίωση της απόδοσης του συστήματος.



Παράδειγμα

Όσο πηγαίνουμε σε υψηλότερες συσχετιστικότητες ο κύκλος ρολογιού αυξάνεται σύμφωνα με τους παρακάτω τύπους:

$$\text{κύκλος ρολογιού}_{2-\delta\rho\acute{o}\mu\omega\nu} = 1.10 * \text{κύκλος ρολογιού}_{1-\delta\rho\acute{o}\mu\omega\nu}$$

$$\text{κύκλος ρολογιού}_{4-\delta\rho\acute{o}\mu\omega\nu} = 1.12 * \text{κύκλος ρολογιού}_{1-\delta\rho\acute{o}\mu\omega\nu}$$

$$\text{κύκλος ρολογιού}_{8-\delta\rho\acute{o}\mu\omega\nu} = 1.14 * \text{κύκλος ρολογιού}_{1-\delta\rho\acute{o}\mu\omega\nu}$$

Θεωρήστε ότι ο χρόνος επιτυχίας είναι 1 κύκλος ρολογιού, ότι η ποινή αποτυχίας για κρυφή μνήμη άμεσης αντιστοίχισης είναι 50 κύκλοι ρολογιού και χρησιμοποιώντας τον πίνακα 1 να απαντήσετε στις παρακάτω ερωτήσεις.

Ρυθμός αποτυχίας (πίνακας 1)
(άθροισμα 100% του ολικού ρυθμού αποτυχίας)

Μέγεθος Κρυφής μν.	Βαθμός Συσχετιστικότητας	Ολικός Ρυθμός αποτυχίας	Αναγκαστικές	Χωριτηκότητα	Συγκρούσεις
1-KB	1-δρόμου	0.133	0.002 1%	0.080 60%	0.052 39%
1-KB	2-δρόμων	0.105	0.002 2%	0.080 76%	0.023 22%
1-KB	4- δρόμων	0.095	0.002 2%	0.080 84%	0.013 14%
1-KB	8- δρόμων	0.087	0.002 2%	0.080 92%	0.005 6%
128-KB	1- δρόμων	0.010	0.002 20%	0.004 40%	0.004 40%
128-KB	2- δρόμων	0.007	0.002 29%	0.004 58%	0.001 14%
128-KB	4- δρόμων	0.006	0.002 31%	0.004 61%	0.001 8%
128-KB	8- δρόμων	0.006	0.002 31%	0.004 62%	0.000 7%

Για το ρυθμό αποτυχίας, να βρεθεί ποιο μέγεθος κρυφής μνήμης αληθεύει κάθε μια από τις ακόλουθες προτάσεις 1, 2, 3:

- 1) Μέσος χρόνος προσπέλασης μνήμης συνολο-συσχετιστική οργάνωση 8-δρόμων < Μέσος χρόνος προσπέλασης μνήμης συνολο-συσχετιστική οργάνωση 4-δρόμων
- 2) Μέσος χρόνος προσπέλασης μνήμης συνολο-συσχετιστική οργάνωση 4-δρόμων < Μέσος χρόνος προσπέλασης μνήμης συνολο-συσχετιστική οργάνωση 4-δρόμων
- 3) Μέσος χρόνος προσπέλασης μνήμης συνολο-συσχετιστική οργάνωση 2-δρόμων < Μέσος χρόνος προσπέλασης μνήμης συνολο-συσχετιστική οργάνωση 1-δρόμου

Απάντηση:

Μέσος χρόνος προσπέλασης μνήμης για κάθε συσχετιστικότητα είναι :

$$\begin{aligned} \text{Μέσος χρόνος προσπέλασης μνήμης} & \text{συνολο-συσχετιστική οργάνωση 8-δρόμων} = \\ & = \text{Χρόνος επιτυχίας 8-δρόμων} + \text{Ρυθμός αποτυχίας 8-δρόμων} \times \text{Ποινή αποτυχίας 1-} \\ & \text{δρόμου} \\ & = 1.14 + \text{Ρυθμός αποτυχίας 8-δρόμων} \times 50 \end{aligned}$$

$$\text{Μέσος χρόνος προσπέλασης μνήμης 4-δρόμων} = 1.12 + \text{Ρυθμός αποτυχίας} \times 50$$

$$\text{Μέσος χρόνος προσπέλασης μνήμης 2-δρόμων} = 1.10 + \text{Ρυθμός αποτυχίας} \times 50$$

$$\text{Μέσος χρόνος προσπέλασης μνήμης 1-δρόμου} = 1.00 + \text{Ρυθμός αποτυχίας} \times 50$$

Η ποινή αποτυχίας είναι η ίδια σε κάθε περίπτωση, για αυτό την αφήνουμε ίση με 50 κύκλους ρολογιού. Για παράδειγμα, ο **Μέσος χρόνος προσπέλασης μνήμης** για κρυφή μνήμη άμεσης αντιστοίχισης μεγέθους 1-KB είναι :

$$\text{Μέσος χρόνος προσπέλασης μνήμης 1-δρόμου} = 1.00 + (0.133 \times 50) = 7.65$$

Και ο χρόνος για μία συνολο-συσχετιστική 8-δρόμων κρυφή μνήμη μεγέθους 128-KB είναι :

$$\text{Μέσος χρόνος προσπέλασης μνήμης 8-δρόμων} = 1.14 + (0.006 \times 50) = 1.44$$

Χρησιμοποιούμε αυτούς του τύπους και τις προβλέψεις του ρυθμού αποτυχίας από τον πίνακα που ακολουθεί:

**Ρυθμός αποτυχίας (σχετικά ποσοστά)
(άθροισμα 100% του ολικού ρυθμού αποτυχίας)**

Μέγεθος	Βαθμός	Ολικός Ρυθμός							
Κρυφής μν.	Συσχετιστικότητα	αποτυχίας	Αναγκαστικές			Χωριτηκότητα		Συγκρούσεις	
1-KB	1-δρόμου	0.133	0.002	1%		0.080	60%	0.052	39%
1-KB	2-δρόμων	0.105	0.002	2%		0.080	76%	0.023	22%
1-KB	4- δρόμων	0.095	0.002	2%		0.080	84%	0.013	14%
1-KB	8- δρόμων	0.087	0.002	2%		0.080	92%	0.005	6%
128-KB	1- δρόμων	0.010	0.002	20%		0.004	40%	0.004	40%
128-KB	2- δρόμων	0.007	0.002	29%		0.004	58%	0.001	14%
128-KB	4- δρόμων	0.006	0.002	31%		0.004	61%	0.001	8%
128-KB	8- δρόμων	0.006	0.002	31%		0.004	62%	0.000	7%

Μπορούμε να δημιουργήσουμε ένα πίνακα που θα δείχνει το Μέσο χρόνο προσπέλασης μνήμης για κάθε κρυφή μνήμη και συσχετιστικότητα. Ο πίνακας αυτός δείχνει ότι οι τρεις προτάσεις τις εκφώνησης ισχύουν για κρυφές μνήμες μικρότερες ή ίσες των 16KB. Αρχίζοντας με 32KB παρατηρούμε ότι ο Μέσος χρόνος προσπέλασης μνήμης 4-δρόμων είναι μικρότερος από ότι για 2-δρόμων, για 2-δρόμων είναι μικρότερος από ότι για 1-δρόμου, αλλά για μια 8-δρόμων κρυφή μνήμη δεν είναι μικρότερος από μια 4-δρόμων.

Συσχετιστικότητα

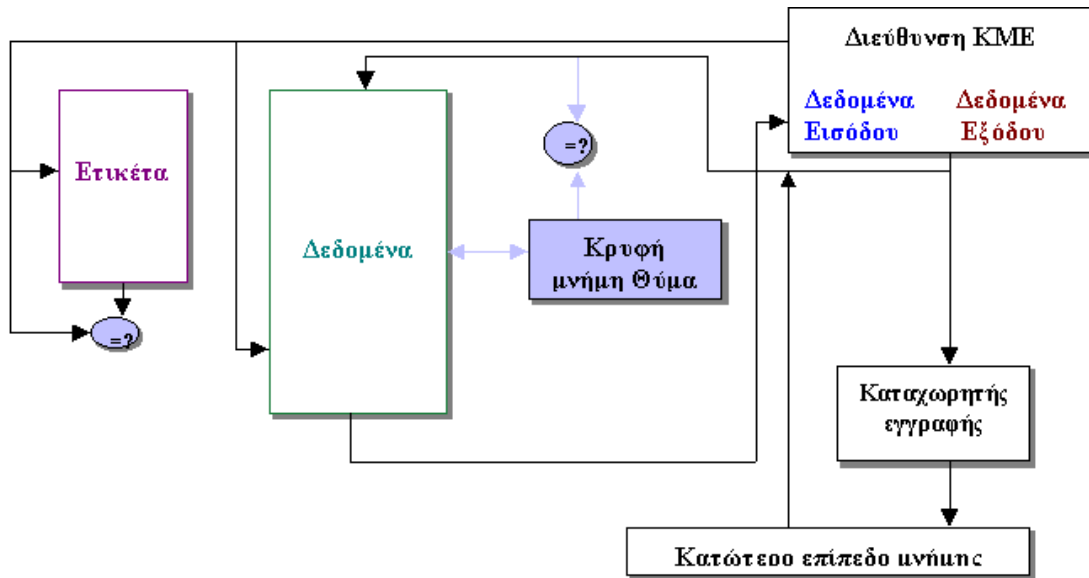
Μέγεθος				
Κρυφής μν.	1-δρόμου	2 –δρόμων	4 - δρόμων	8 – δρόμων
1	7.65	6.60	6.22	5.44
2	5.90	4.90	4.62	4.09
4	4.60	3.95	3.57	3.19
8	3.30	3.00	2.87	2.59
16	2.45	2.20	2.12	2.04
32	2.00	1.80	1.77	1.79
64	1.70	1.60	1.57	1.59
128	1.50	1.45	1.42	1.44



Κρυφή μνήμη Θύμα

Η κρυφή μνήμη Θύμα αποτελεί μία καλή εναλλακτική επιλογή με σκοπό τη μείωση του ρυθμού αποτυχίας. Υλοποιείται με την προσθήκη μιας μικρής πλήρως συσχετιστικής μνήμης ανάμεσα στη κρυφή μνήμη και το μονοπάτι ανατροφοδότησης της με δεδομένα. Το σχήμα 4.3.2 παρουσιάζει την θέση της κρυφής μνήμης θύμα στην ιεραρχία μνήμης.

Η κρυφή μνήμη θύμα περιέχει μόνο εκείνα τα μπλοκ που έχουν απορριφθεί σε μία αποτυχία απ' την κρυφή μνήμη, τα οποία ονομάζονται μπλοκ "θύματα". Σε μια αποτυχία κρυφής μνήμης, ο έλεγχος για τον εντοπισμό των επιθυμητών δεδομένων μετατοπίζεται στα μπλοκ "θύματα". Ακολουθεί -αν χρειάζεται- η μετάβαση στη κύρια μνήμη. Εάν εντοπιστεί το ζητούμενο μπλοκ, γίνεται εναλλαγή του μπλοκ θύμα με το μπλοκ της κρυφής μνήμης.



Σχήμα 4.3.2 - Η θέση της κρυφής μνήμης θύμα στην ιεραρχία μνήμης.



ΔΡΑΣΤΗΡΙΟΤΗΤΑ 4

- Να απαριθμήσετε τουλάχιστον 5 τεχνικές μείωσης του ρυθμού αποτυχίας.
- Να εξηγήσετε περιληπτικά τι είναι η κρυφή μνήμη θύμα και σε ποια θέση τοποθετείται στην ιεραρχία της μνήμης. Να συγκρίνετε την απάντησή σας με την παράγραφο: «Κρυφή μνήμη θύμα».



Μείωση της ποινής αποτυχίας της κρυφής μνήμης

Ο τύπος απόδοσης της κρυφής μνήμης μας επιβεβαιώνει ότι οι βελτιώσεις στις ποινές αποτυχίας μπορούν να είναι το ίδιο ωφέλιμες όπως οι βελτιώσεις στο ρυθμό αποτυχίας. Συνολικά διαθέτουμε 5 τεχνικές βάσει των οποίων προσπαθούμε να μειώσουμε την ποινή αποτυχίας. Επιγραμματικά είναι οι ακόλουθες:

1. Προτεραιότητα στις αποτυχίες ανάγνωσης έναντι των αποτυχιών εγγραφής
2. Τοποθέτηση υπο-μπλοκ για μειωμένη ποινή αποτυχίας
3. Αποστολή πρώτα της αρχικής κρίσιμης λέξης στη ΚΜΕ
4. Υλοποίηση κρυφής μνήμης χωρίς μπλοκ για μείωση των καθυστερήσεων κατά τις αποτυχίες κρυφής μνήμης.
5. Κρυφές μνήμες 2⁰⁰ επιπέδου



Από τις παραπάνω τεχνικές μόνο η 5^η επικεντρώνεται στη διεπαφή μεταξύ της κρυφής μνήμης και της κύριας μνήμης, ενώ οι πρώτες τέσσερις μετατοπίζουν τη βαρύτητα στην ΚΜΕ, προκειμένου να μειώσουν την ποινή αποτυχίας. Για το λόγο αυτό θα αναφερθούμε στην 5^η τεχνική, η οποία είναι και η πιο σημαντική, ενώ από τις υπόλοιπες θα μελετήσουμε την 1^η και 2^η ως τις πιο ενδιαφέρουσες.

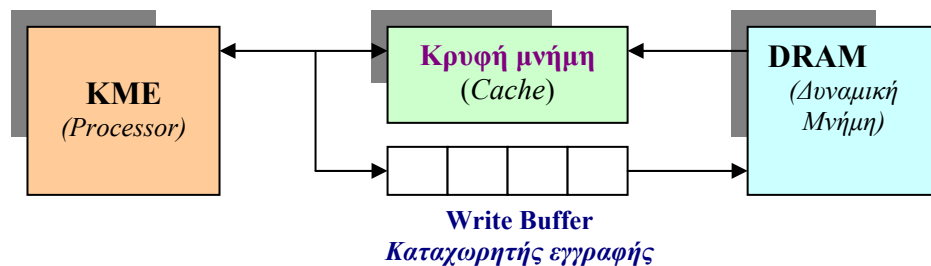
► Προτεραιότητα στις αποτυχίες ανάγνωσης έναντι των αποτυχιών εγγραφής

Μια τεχνική μείωσης της ποινής αποτυχίας είναι να δοθεί προτεραιότητα στις αποτυχίες ανάγνωσης έναντι των αποτυχιών εγγραφής.



ΔΡΑΣΤΗΡΙΟΤΗΤΑ 5

Θυμάστε τον τρόπο με τον οποίο συνδέεται ο καταχωρητής εγγραφής με την ΚΜΕ και την κρυφή μνήμη; Να σχεδιάσετε την υλοποίηση της διεγγραφής με χρήση ενός καταχωρητή εγγραφής. Να συγκρίνετε το σχήμα σας με το σχήμα 4.3.3 που ακολουθεί. Για περισσότερες λεπτομέρειες να ανατρέξετε στην 3^η ενότητα αυτού του κεφαλαίου και συγκεκριμένα στην υποενότητα: «Η λειτουργία της Ανάγνωσης και της Εγγραφής».



Σχήμα 4.3.3- Η θέση ενός καταχωρητή εγγραφής

Σε μία κρυφή μνήμη που υλοποιείται η τεχνική της διεγγραφής για τις εγγραφές, η πιο σημαντική βελτίωση είναι η χρήση ενός καταχωρητή εγγραφής. Οι καταχωρητές εγγραφής ωστόσο περιπλέκουν τις προσπελάσεις στη μνήμη, επειδή μπορεί να καθυστερήσουν την ενημέρωση κάποιου μπλοκ της κύριας μνήμης με διεύθυνση αναγκαία σε μία αποτυχία ανάγνωσης της κρυφής μνήμης.

Για να μην υπάρχει αυτό το πρόβλημα, η αποτυχία ανάγνωσης πρέπει να περιμένει μέχρι να αδειάσει καταχωρητής εγγραφής. Η εναλλακτική λύση είναι να ελέγχουμε τα περιεχόμενα καταχωρητή εγγραφής σε μια αποτυχία ανάγνωσης και εάν δεν υπάρχουν συγκρούσεις και το σύστημα μνήμης είναι διαθέσιμο, να επιτρέπεται να συνεχίσει η αποτυχία ανάγνωσης.



Παράδειγμα

Δίνεται ο ακόλουθος κώδικας :


SW 512 (R0), R3 ; M[512] ← R3 (δείκτης κρυφής μνήμης 0)

LW R1, 1024 (R0) ; R1 ← M[1024] (δείκτης κρυφής μνήμης 0)

LW R2, 512 (R0) ; R2 ← M[512] (δείκτης κρυφής μνήμης 0)

Εάν διαθέτουμε μία κρυφή μνήμη άμεσης αντιστοίχισης με διεγγραφή που αντιστοιχεί τις διευθύνσεις 512 και 1024 στο ίδιο μπλοκ και έναν καταχωρητή εγγραφής 4 λέξεων, τότε η τιμή του καταχωρητή R2 θα ισούται πάντοτε με τη τιμή του R3;

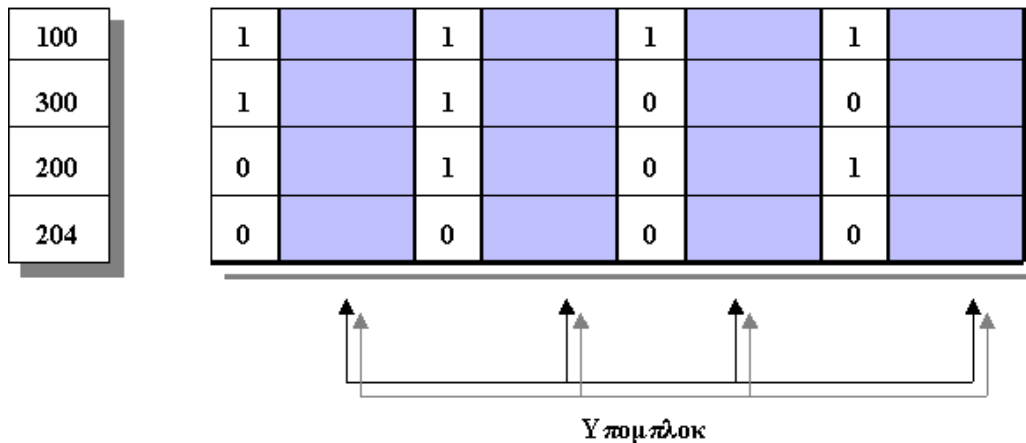
Απάντηση :

Αν ακολουθήσουμε την προσπέλαση στην κρυφή μνήμη, είναι φανερό πως μπορεί να προκύψει πρόβλημα. Μετά την αποθήκευση, τα δεδομένα του καταχωρητή **R3** τοποθετούνται στον καταχωρητή εγγραφής. Η επόμενη φόρτωση χρησιμοποιεί τον ίδιο δείκτη και είναι αποτυχία. Η φόρτωση της 2^{ης} εντολής προσπαθεί να θέσει την τιμή της θέσης 512 στον καταχωρητή **R2**, γεγονός που καταλήγει σε αποτυχία, στην περίπτωση που ο καταχωρητής εγγραφής δεν έχει ολοκληρώσει την εγγραφή της διεύθυνσης 512 στη μνήμη. Οπότε, η ανάγνωση της διεύθυνσης 512 θα θέσει τη παλιά, λανθασμένη τιμή στο μπλοκ της κρυφής μνήμης και στη συνέχεια στον καταχωρητή **R2**. Χωρίς τις απαραίτητες προφυλάξεις, το περιεχόμενο του **R3** δεν θα είναι ίσο με εκείνο του **R2**. 

➡ Τοποθέτηση υπο-μπλοκ για μειωμένη ποινή αποτυχίας

Τα μεγάλου μεγέθους μπλοκ περιορίζουν το πεδίο ετκέτας χωρίς να μειώνουν την πληροφορία που μπορεί να αποθηκευτεί στη κρυφή μνήμη. Επιπρόσθετα βελτιώνεται ο ρυθμός αποτυχίας με τίμημα την αύξηση της ποινή αποτυχίας. Η λύση δίνεται με την τοποθέτηση υπο-μπλοκ. Ένα bit εγκυρότητας, προστίθεται σε μονάδες πληροφορίας μικρότερες του μπλοκ που ονομάζονται υπο-μπλοκ. Μόνο ένα υπο-μπλοκ διαβάζεται σε μία αποτυχία μνήμης. Το bit εγκυρότητας χαρακτηρίζει κάποια τμήματα του μπλοκ ως έγκυρα και τα υπόλοιπα ως μη έγκυρα. Το γεγονός ότι η ετικέτα ταιριάζει δεν αποτελεί τη μοναδική εγγύηση ότι η λέξη ανήκει στην κρυφή μνήμη. Επιπλέον θα πρέπει να έχει τεθεί το bit εγκυρότητας. Η διαφορά στην ποινή αποτυχίας για τα υπο-μπλοκ και τα

μπλοκ είναι φανερή. Το σχήμα 4.3.4 παρουσιάζει την αποθήκευση της ετικέτας. Εάν τα bit εγκυρότητας αντικατασταθούν από τις πλήρεις ετικέτες, απαιτείται περισσότερη μνήμη.



Σχήμα 4.3.4 - Η κρυφή μνήμη άμεσης αντιστοίχισης με 4 υπο-μπλοκ ανά μπλοκ. Τα υπο-μπλοκ μπορούν να θεωρηθούν ως ένα επιπλέον επίπεδο διευθυνσιοδότησης εκτός από την ετικέτα της διεύθυνσης. Στο 1^ο μπλοκ (κορυφή) έχουν τεθεί όλα τα μπλοκ εγκυρότητας, ισοδύναμα έχουν τεθεί τα bit εγκυρότητας για ένα μπλοκ της κρυφής μνήμης. Στο τελευταίο μπλοκ (τελευταία γραμμή) δεν έχει τεθεί κανένα από τα bit. Στο 2^ο μπλοκ στις διευθύνσεις 300, 301 έχουμε επιτυχία ενώ στις 302, 303 αποτυχία. Στο 3^ο μπλοκ στις διευθύνσεις 201, 203 έχουμε επιτυχία. Εάν διαθέταμε 16 μπλοκ με μέγεθος υπο-μπλοκ, θα χρειαζόμασταν 16 αντί 4 ετικέτες. Σε κρυφές μνήμες που υλοποιείται τοποθέτηση υπο-μπλοκ, το μπλοκ δεν λογαριάζεται πλέον ως η ελάχιστη μονάδα πληροφορίας διακινήσιμη μεταξύ κρυφής και κύριας μνήμης, αλλά ως η μονάδα πληροφορίας που σχετίζεται με την ετικέτα διεύθυνσης.

➡ Κρυφές Μνήμες 2^{ου} επιπέδου

Με την προσθήκη ενός ακόμη επιπέδου της κρυφής μνήμης μεταξύ της αρχικής κρυφής μνήμης και της κύριας μνήμης, η κρυφή μνήμη 1^{ου} επιπέδου μπορεί να είναι αρκετά μικρή, τόσο ώστε να ταιριάζει με το κύκλο ρολογιού της γρήγορης ΚΜΕ, ενώ η κρυφή μνήμη 2^{ου} επιπέδου μπορεί να είναι αρκετά μεγάλη, ώστε να δεσμεύσει πολλές προσπελάσεις που προορίζονται για την κύρια μνήμη. Με αυτόν τον τρόπο μειώνεται η ισχύς της ποινής αποτυχίας.

Αν και η γενική ιδέα της προσθήκης ενός επιπλέον επιπέδου στην ιεραρχία μνήμης είναι ξεκάθαρη, περιπλέκει την ανάλυση της απόδοσης. Οι ορισμοί για μια κρυφή μνήμη δευτέρου επιπέδου δεν είναι πάντα ξεκάθαροι. Ας αρχίσουμε με τον ορισμό του μέσου χρόνου πρόσβασης στην κύρια μνήμη για μια κρυφή μνήμη δύο επιπέδων. Χρησιμοποιώντας τους συμβολισμούς L1 και L2 για να αναφερθούμε στο 1^ο και 2^ο επίπεδο αντίστοιχα της κρυφής μνήμης ισχύουν τα εξής:

(α). Μέσος χρόνος πρόσβασης στη μνήμη = Χρόνος επιτυχίας_{L1} + Ρυθμός αποτυχίας_{L1} x Ποινή αποτυχίας_{L1}

(β). Ποινή αποτυχίας_{L1} = Χρόνος επιτυχίας_{L2} + Ρυθμός αποτυχίας_{L2} x Ποινή αποτυχίας_{L2}

Οπότε:

(γ). Μέσος χρόνος πρόσβασης στη μνήμη = Χρόνος επιτυχίας_{L1} + Ρυθμός αποτυχίας_{L1} x (Χρόνος επιτυχίας_{L2} + Ρυθμός αποτυχίας_{L2} x Ποινή αποτυχίας_{L2})

Σε αυτόν τον τύπο ο ρυθμός αποτυχίας 2^{ου} επιπέδου υπολογίζεται βάσει του ρυθμού της κρυφής μνήμης 1^{ου} επιπέδου.

Οι ακόλουθοι όροι υιοθετούνται για ένα σύστημα κρυφής μνήμης δύο επιπέδων

- Τοπικός ρυθμός αποτυχίας: Είναι ο αριθμός των αποτυχιών στην κρυφή μνήμη δια του συνολικού αριθμού των προσβάσεων στην κρυφή μνήμη. Αυτό είναι ο **ρυθμός αποτυχίας_{L2}** του παραπάνω τύπου για την κρυφή μνήμη 2^{ου} επιπέδου.

- Καθολικός ρυθμός αποτυχίας: Είναι ο αριθμός των αποτυχιών στην κρυφή μνήμη δια του συνολικού αριθμού των προσβάσεων στην μνήμη που παράγονται από τη ΚΜΕ. Χρησιμοποιώντας τους όρους των παραπάνω τύπων ο καθολικός ρυθμός αποτυχίας του 2^{ου} επιπέδου κρυφής μνήμης είναι:

(δ). Ρυθμός αποτυχίας_{L1} x Ρυθμός αποτυχίας_{L2}



ΔΡΑΣΤΗΡΙΟΤΗΤΑ 6

Υποθέτουμε ότι σε 1000 αναφορές στην μνήμη υπάρχουν 40 αποτυχίες σε μια κρυφή μνήμη 1^{ου} επιπέδου και 20 αποτυχίες σε μία κρυφή μνήμη 2^{ου} επιπέδου. Ποιοι είναι οι διάφοροι ρυθμοί αποτυχίας;



ΑΠΑΝΤΗΣΗ ΔΡΑΣΤΗΡΙΟΤΗΤΑΣ 6

Ο ρυθμός αποτυχίας (είτε τοπικός είτε καθολικός) για τη κρυφή μνήμη 1^{ου} επιπέδου είναι: $40/1000 = 4\%$

Για τη κρυφή μνήμη 2^{ου} επιπέδου έχουμε ότι:

- Τοπικός ρυθμός αποτυχίας = $20/40 = 50\%$
- Καθολικός ρυθμός αποτυχίας = $20/1000 = 2\%$



Παράδειγμα

Με βάση τα ακόλουθα δεδομένα, να βρείτε ποια είναι η επίδραση της κρυφής μνήμης 2^{ου} και 1^{ου} επιπέδου σχετικά με την ποινή αποτυχίας.

- Η 2-δρόμων τοποθέτηση αυξάνει τον χρόνο επιτυχίας έως 10% του κύκλου ρολογιού της ΚΜΕ.
- Χρόνος επιτυχίας_{L2} για άμεση αντιστοίχιση=10 κύκλους ρολογιού
- Τοπικός ρυθμός αποτυχίας_{L2} για άμεση αντιστοίχιση=25%
- Τοπικός ρυθμός αποτυχίας_{L2} για 2-δρόμων σύνολο- συσχετιστική τοποθέτηση=20%
- Ποινή αποτυχίας_{L2}=50 κύκλους ρολογιού

Απάντηση:

Για μία κρυφή μνήμη 2^{ου} επιπέδου άμεσης αντιστοίχισης η ποινή αποτυχίας 1^{ου} επιπέδου είναι:

$$\text{Ποινή αποτυχίας}_{1\text{-δρόμων}L1} = 10 + 25\% * 50 = 22,5 \text{ κύκλοι ρολογιού}$$

Προσθέτοντας το κόστος συσχετιστικότητας αυξάνεται το κόστος επιτυχίας μόνο 0.1 κύκλους ρολογιού με νέα ποινή αποτυχίας 1^{ου} επιπέδου την εξής :

$$\text{Ποινή αποτυχίας}_{2\text{-δρόμων}L1} = 10.1 + 20\% * 50 = 20,1 \text{ κύκλοι ρολογιού}$$

Στην πραγματικότητα η κρυφή μνήμη 2^{ου} επιπέδου είναι σχεδόν πάντα συγχρονισμένη με την κρυφή μνήμη 1^{ου} επιπέδου και την ΚΜΕ. Αναλογικά ο χρόνος επιτυχίας 2^{ου} επιπέδου θα πρέπει να είναι ένας ακέραιος αριθμός κύκλων ρολογιού. Αν είμαστε τυχεροί μπορούμε να περιορίσουμε το χρόνο επιτυχίας του 2^{ου} επιπέδου σε 10 κύκλους εάν όχι, μπορούμε γύρω στους 11 κύκλους. Μία άλλη επιλογή είναι μία βελτίωση πάνω στην κρυφή μνήμη άμεσης αντιστοίχισης 2^{ου} επιπέδου.

$$\text{Ποινή αποτυχίας}_{2\text{-δρόμων}L1} = 10 + 20\% * 50 = 20,0 \text{ κύκλοι ρολογιού}$$

$$\text{Ποινή αποτυχίας}_{2\text{-δρόμων}L1} = 11 + 20\% * 50 = 21,0 \text{ κύκλοι ρολογιού}$$



Προκειμένου να μειώσουμε την ποινή αποτυχίας, μειώνουμε το ρυθμό αποτυχίας της κρυφής μνήμης 2^{ου} επιπέδου χρησιμοποιώντας τις τεχνικές που αναφέρθηκαν στην 2^η υποενότητα της ενότητας αυτής (Η μείωση των αποτυχιών της κρυφής μνήμης). Αναφέρουμε ότι η υψηλή συσχετιστικότητα ή ψευδοσυσχετιστικότητα αποτελούν λύσεις γιατί έχουν μικρή επιρροή στον χρόνο επιτυχίας 2^{ου} επιπέδου και γιατί μέρος του χρόνου πρόσβασης οφείλεται στις αποτυχίες στην κρυφή μνήμη 2^{ου} επιπέδου. Το μεγαλύτερο μέγεθος του 2^{ου} επιπέδου κρυφής μνήμης εξαλείφει τις συγκρούσεις αποτυχιών διανέμοντας δεδομένα σε περισσότερα μπλοκ, αλλά και τις περισσότερες από τις ικανές αποτυχίες. Παρόλα αυτά το ποσοστό των αποτυχημένων συγκρούσεων είναι ακόμη σημαντικό στην άμεση αντιστοίχιση 2^{ου} επιπέδου κρυφής μνήμης.



Μια άλλη προσέγγιση στη μείωση αποτυχιών είναι η *αύξηση του μεγέθους μπλοκ στις κρυφές μνήμες 2^{ου} επιπέδου*. Αυξάνοντας το μέγεθος των μπλοκ μπορούμε να αυξήσουμε τις αποτυχίες σύγκρουσης που πραγματοποιούνται σε μικρές κρυφές μνήμες αφού μπορεί να μην υπάρχει αρκετός χώρος για τοποθέτηση δεδομένων και γι' αυτό το λόγο αυξάνεται ο ρυθμός αποτυχίας. Επειδή αυτό δεν αποτελεί πρόβλημα για τις κρυφές μνήμες 2^{ου} επιπέδου και επειδή η πρόσβαση στην μνήμη είναι σχετικά μεγάλη, μεγέθη μπλοκ των 64bytes, 128 bytes και ακόμη και των 256 bytes είναι δημοφιλή.



Μια άλλη σκέψη αναφέρεται στο εάν όλα τα δεδομένα στην κρυφή μνήμη 1^{ου} επιπέδου είναι πάντα στο 2^ο επίπεδο κρυφής μνήμης. Εάν ναι, το 2^ο επίπεδο κρυφής μνήμης λέγεται ότι έχει την πολυεπίπεδη έγκλειστη ιδιοκτησία. Ο εγκλεισμός είναι σκόπιμος γιατί είναι συνέπεια μεταξύ των μονάδων Εισόδου/Εξόδου και κρυφών μνημών (ή μεταξύ κρυφών μνημών και ενός πολυεπεξεργαστή) και μπορεί να επιτευχθεί με τον έλεγχο του 2^{ου} επιπέδου κρυφής μνήμης. Το μειονέκτημα σ' αυτόν τον φυσικό εγκλεισμό είναι ότι ο χαμηλός μέσος χρόνος πρόσβασης στη μνήμη προτείνει μικρότερα μπλοκ για το μικρότερο 1^ο επίπεδο κρυφής μνήμης και μεγαλύτερα μπλοκ για το μεγαλύτερο 2^ο επίπεδο κρυφής μνήμης. Το 2^ο επίπεδο κρυφής μνήμης ακόμη θα πρέπει να ακυρώσει όλα τα 1^{ου} επιπέδου μπλοκ τα οποία ταιριάζουν στο 2^{ου} επίπεδο μπλοκ αντικαθιστώντας τα, και προκαλώντας έτσι έναν λίγο υψηλότερο ρυθμό αποτυχίας 1^{ου} επιπέδου. Μπορεί επίσης να προκαλέσει περιττές ακυρώσεις στην κρυφή μνήμη. Ο εγκλεισμός αυξάνει την πολυπλοκότητα όταν συνδέεται με την εκτέλεση βελτιστοποίησης, όπως μια δευτερεύουσα κρυφή μνήμη χωρίς μπλοκ.



Συμπερασματικά, η ουσία του σχεδιασμού κρυφής μνήμης 2^{ου} επιπέδου εξισορροπεί γρήγορες επιτυχίες και λίγες αποτυχίες. Σε αντίθεση με την τεχνική αυτή, οι περισσότερες βελτιστοποιήσεις που χρησιμοποιούνται και έχουν θετική επίδραση στις επιτυχίες, έχουν αρνητικές επιπτώσεις στις αποτυχίες και αντίστροφα. Για τις κρυφές μνήμες 2^{ου} επιπέδου υπάρχουν πολύ λιγότερες επιτυχίες από τις κρυφές μνήμες 1^{ου} επιπέδου. Αυτή η θεώρηση οδηγεί σε **μεγαλύτερες κρυφές μνήμες με υψηλή συσχετιστικότητα και μεγαλύτερα μπλοκ**.



ΔΡΑΣΤΗΡΙΟΤΗΤΑ 7

Να απαριθμήσετε τις τεχνικές που χρησιμοποιούνται για τη μείωση της ποινής αποτυχίας της κρυφής μνήμης. Ποια από τις τεχνικές αυτές θεωρείται η πιο σημαντική και με ποιον τρόπο αυτή εφαρμόζεται; Να συγκρίνετε την απάντησή σας με την υποενότητα που διαβάσατε: «Μείωση της ποινής αποτυχίας της κρυφής μνήμης».



Μείωση του χρόνου επιτυχίας

Ένας τρόπος για να βελτιώσουμε την απόδοση της κρυφής μνήμης είναι να μειώσουμε το μέσο χρόνο προσπέλασης της μνήμης.

Ο χρόνος επιτυχίας είναι ένας σημαντικός παράγοντας, γιατί επηρεάζει το ρυθμό ρολογιού του επεξεργαστή ως εξής: στα περισσότερα υπολογιστικά συστήματα ο χρόνος πρόσβασης στη κρυφή μνήμη περιορίζει το ρυθμό του κύκλου του ρολογιού, ακόμα και σε εκείνα τα συστήματα που χρησιμοποιούν πολλαπλούς κύκλους ρολογιού, προκειμένου να προσπελάσουν την κρυφή μνήμη. Στη συνέχεια εξετάζουμε δύο βασικές τεχνικές για τη μείωση του χρόνου επιτυχίας και μια βελτιστοποίηση για τη μείωση του χρόνου στις επιτυχίες εγγραφής.

► 1^η τεχνική μείωσης του χρόνου επιτυχίας: *Μικρή και απλή κρυφή μνήμη*

Ένα τμήμα του χρόνου που καταναλώνεται σε μια επιτυχία της κρυφής μνήμης χρησιμοποιεί το τμήμα δείκτη της διεύθυνσης για να διαβάσει την μνήμη ετικετών και μετά να το συγκρίνει με τη διεύθυνση. Είναι σημαντικό η κρυφή μνήμη να είναι αρκετά μικρή για να χωράει στο ίδιο ολοκληρωμένο κύκλωμα με τον επεξεργαστή. Έτσι αποφεύγεται η ποινή χρόνου που προστίθεται για την έξοδο από το ολοκληρωμένο κύκλωμα προκειμένου να εκλεχτούν οι ετικέτες διεύθυνσης. Επιπρόσθετα, η κρυφή μνήμη πρέπει να κρατηθεί απλή, όπως όταν χρησιμοποιείται άμεση αντιστοίχιση. Ένα κύριο πλεονέκτημα των κρυφών μνημών άμεσης αντιστοίχισης είναι ότι ανάλογα με τη σχεδίαση μπορεί να επικαλυφθεί ο έλεγχος ετικετών με την μετάδοση των δεδομένων. Αυτό μειώνει αποτελεσματικά τον χρόνο επιτυχίας.

► 2^η τεχνική μείωσης του χρόνου επιτυχίας: *Αποφυγή μεταφράσεως διεύθυνσης κατά τη διάρκεια δεικτοδότησης της κρυφής μνήμης*

Η κρυφή μνήμη πρέπει να κάνει μετάφραση μιας ιδεατής διεύθυνσης από την ΚΜΕ σε μια φυσική διεύθυνση για να προσπελάσει την μνήμη.



Η ιδεατή διεύθυνση αναφέρεται σε διευθύνσεις θέσεων της ιδεατής μνήμης. Για αυτήν θα μιλήσουμε στην 4^η ενότητα του κεφαλαίου.

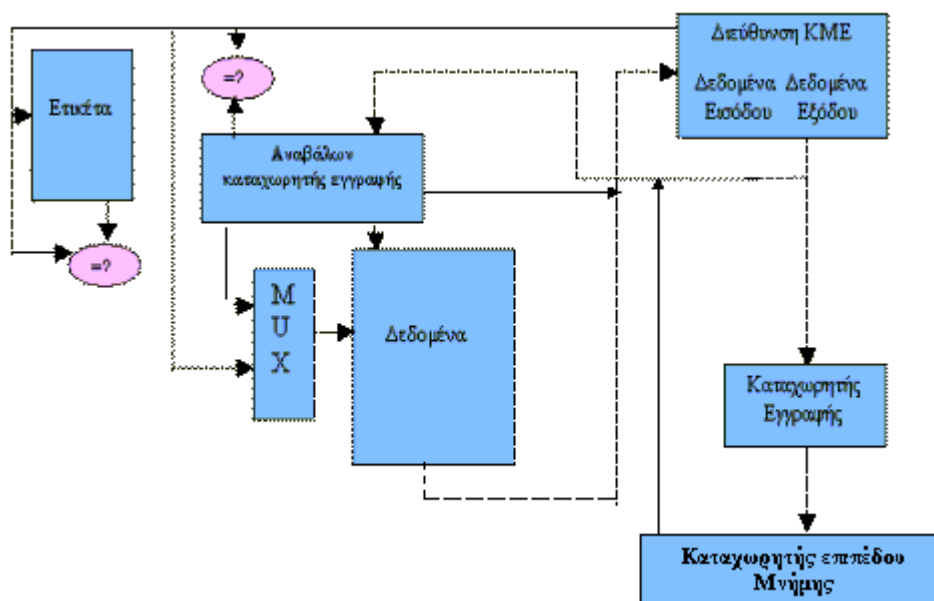
Μία τεχνική είναι να χρησιμοποιούνται εικονικές διευθύνσεις για την κρυφή μνήμη. Τέτοιες κρυφές μνήμες ονομάζονται εικονικές κρυφές μνήμες. Η ιδεατή διευθυνσιοδότηση εξαλείφει τον χρόνο μεταφράσεως διευθύνσεων από μια επιτυχία κρυφής μνήμης.

Κρατώντας τις κρυφές μνήμες μικρές και απλές και χρησιμοποιώντας τεχνικές για την αποφυγή των καθυστερήσεων που απαιτεί η μετάφραση της διεύθυνσης, οι επιτυχίες ανάγνωσης και οι επιτυχίες εγγραφής γίνονται γρηγορότερες.

► Βελτιστοποίηση για επιτυχίες εγγραφής: Σωλήνωση εγγραφών για πιο γρήγορες επιτυχίες εγγραφών

Οι επιτυχίες εγγραφών συνήθως παίρνουν μεγαλύτερο χρονικό διάστημα από αυτές των αναγνώσεων γιατί η ετικέτα πρέπει να ελεγχθεί πριν εγγραφούν τα δεδομένα, διαφορετικά θα μπορούσε να γίνει εγγραφή σε λανθασμένη διεύθυνση. Μια τεχνική χρησιμοποιούμενη στον υπολογιστή Alpha AXP 21064 και σε άλλους υπολογιστές, χρησιμοποιεί τη σωλήνωση των εγγραφών. Το σχήμα 4.3.4, απεικονίζει την οργάνωση του υλικού των σωληνωμένων εγγραφών.

Αρχικά οι ετικέτες και τα δεδομένα διαχωρίζονται έτσι ώστε να διευθυνσιοδοτηθούν ανεξάρτητα. Σε μία εγγραφή η κρυφή μνήμη συγκρίνει την ετικέτα με την τρέχουσα διεύθυνση εγγραφής όπως συνηθίζεται. Η διαφορά είναι ότι εκτελείται εγγραφή στο τμήμα των δεδομένων της κρυφής μνήμης το οποίο βρίσκεται κατά τη διάρκεια της σύγκρισης της ετικέτας. Θα πρέπει να χρησιμοποιεί κάποια άλλη διεύθυνση μέχρι η τρέχουσα διεύθυνση εγγραφής να έχει και πάλι ελεγχθεί. Το κόλπο είναι ότι η κρυφή μνήμη χρησιμοποιεί διεύθυνση και δεδομένα από την προηγούμενη εγγραφή η οποία έχει ήδη καθοριστεί ως μία επιτυχία. Έτσι η λογική σωλήνωση είναι μεταξύ των εγγραφών – η δεύτερη φάση της εγγραφής συμβαίνει κατά τη διάρκεια της πρώτης φάσης της επόμενης εγγραφής (ή κατά τη διάρκεια της αποτυχίας της κρυφής μνήμης). Γι' αυτό το λόγο οι εγγραφές μπορούν να εκτελεστούν διαδοχικά η μία μετά την άλλη, μία για κάθε κύκλο ρολογιού γιατί η ΚΜΕ δεν χρειάζεται να περιμένει για τον έλεγχο ετικέτας πριν την εγγραφή. Οι εγγραφές δεν παίζουν ρόλο στη σωλήνωση αφού ήδη λειτουργούν παράλληλα με τον έλεγχο της ετικέτας.



Σχήμα 4.3.4 – Η οργάνωση υλικού για σωληνωμένες εγγραφές Είναι πιθανόν να βρεθούν τα ζητούμενα δεδομένα στον καταχωρητή εγγραφών. Σε αυτή την περίπτωση, είτε ο καταχωρητής εγγραφής προσφέρει τα νέα δεδομένα ή ολοκληρώνει τη λειτουργία του και στη συνέχεια τα νέα δεδομένα διαβάζονται από την μνήμη.



Παράδειγμα

Ας υποθέσουμε ότι ένα σύστημα μνήμης χρειάζεται 40 κύκλους ρολογιού για πλεονάζοντα χώρο (αχρησιμοποίητος χώρος ενός φορέα δεδομένων μετά την μορφοποίησή του) και μεταφέρει 16 bytes κάθε 2 κύκλους ρολογιού. Επομένως το σύστημα μπορεί να παρέχει 16 bytes σε 42 κύκλους ρολογιού, 32 bytes σε 44 κύκλους ρολογιού κ.τ.λ. Ποιο μέγεθος από τα μπλοκ της κρυφής μνήμης παρέχει το μικρότερο μέσο χρόνο πρόσβασης για κάθε κρυφή μνήμη του παρακάτω πίνακα:

Μέγεθοςμπλοκ	Μέγεθος Κρυφής Μνήμης				
	1 K	4 K	16 K	64 K	256 K
16	15.05%	8.57%	3.94%	2.04%	1.09%
32	13.34%	7.24%	2.87%	1.35%	0.70%
64	13.76%	7.00%	2.64%	1.06%	0.51%
128	16.64%	7.78%	2.77%	1.02%	0.49%
256	22.01%	9.51%	3.29%	1.15%	0.49%

Απάντηση:

Γνωρίζουμε ότι ο μέσος χρόνος προσπέλασης δίνεται από τον τύπο:

$$\text{Μέσος χρόνος προσπέλασης μνήμη} = \text{Χρόνος επιτυχίας} + \text{Ρυθμός αποτυχίας} \times \text{Ποινή αποτυχίας}$$

Εάν υποθέσουμε ότι ο χρόνος επιτυχίας της κρυφής μνήμης είναι ένας κύκλος ρολογιού ανεξάρτητα από το μέγεθος του μπλοκ, τότε ο χρόνος πρόσβασης για ένα μπλοκ μεγέθους 16 – bytes σε μια κρυφή μνήμη μεγέθους 1 KB, θα είναι:

$$\begin{aligned} \text{Μέσος χρόνος προσπέλασης μνήμη} &= 1 + (15.05\% \times 42) = 7.321 \text{ κύκλοι ρολογιού} \\ \text{Χρόνος επιτυχίας} &= 1 \text{ κύκλος ρολογιού} \\ \text{Ρυθμός αποτυχίας} &= 15.05\% \\ \text{Ποινή αποτυχίας} &= 42 \text{ κύκλοι ρολογιού} \end{aligned}$$

Ενώ για ένα μπλοκ μεγέθους 256– bytes σε μία κρυφή μνήμη μεγέθους 256 KB ο Μέσος χρόνος προσπέλασης μνήμη είναι :

$$\begin{aligned} \text{Μέσος χρόνος προσπέλασης μνήμη} &= 1 + (0.49\% \times 72) = 1.353 \text{ κύκλοι ρολογιού} \\ \text{Χρόνος επιτυχίας} &= 1 \text{ κύκλος ρολογιού} \\ \text{Ρυθμός αποτυχίας} &= 0.49\% \\ \text{Ποινή αποτυχίας} &= 72 \text{ κύκλοι ρολογιού} \end{aligned}$$

Ο παρακάτω πίνακας δείχνει το μέσο χρόνο προσπέλασης μνήμη για όλα τα μεγέθη μπλοκ της κρυφής μνήμης. Οι καταχωρήσεις με τα τονισμένα μαύρα γράμματα δείχνουν τον μικρότερο χρόνο πρόσβασης και έτσι βρίσκουμε το αντίστοιχο μέγεθος του μπλοκ για κάθε μέγεθος της κρυφής μνήμης. Είναι 32 bytes για 1 KB, 4 KB κρυφές μνήμες και 64 bytes για τις μεγαλύτερες κρυφές μνήμες. Αυτά τα μεγέθη είναι και τα ποιο διαδεδομένα μεγέθη μπλοκ για επεξεργαστές κρυφής μνήμης.

Μέγεθος μπλοκ	Ποινή αποτυχίας	Μέγεθος Κρυφής Μνήμης				
		1K	4K	16K	64K	256K
16	42	7,321	4,599	2,655	1,85	1,458
32	44	6,780	4,186	2,263	1,59	1,308
64	48	7,605	4,360	2,267	1,50	1,245
128	56	10,318	5,357	2,551	1,57	1,274
256	72	16,847	7,847	3,369	1,82	1,353



ΔΡΑΣΤΗΡΙΟΤΗΤΑ 8

Έστω ότι έχουμε ένα ρυθμό αποτυχίας εντολών για το πρόγραμμα δοκιμής gcc 5% και ρυθμό αποτυχίας δεδομένων 10%. Αν το υπολογιστικό σύστημα έχει CPI=4 χωρίς καθυστερήσεις στην μνήμη και η ποινή αποτυχίας είναι 12 κύκλους για όλες τις αποτυχίες, να ορίσετε πόσο γρηγορότερα θα «τρέξει» το υπολογιστικό σύστημα με τέλεια μνήμη και χωρίς αποτυχίες.



ΑΠΑΝΤΗΣΗ ΔΡΑΣΤΗΡΙΟΤΗΤΑΣ 8

Η απόδοση της κρυφής μνήμης δίνεται από τον εξής τύπο :

Χρόνος ΚΜΕ = IC x (CPI_{Εκτέλεσης} + Κύκλοι ρολογιού καθυστέρησης μνήμης ανά εντολή) x χρόνος κύκλου ρολογιού .

Ο αριθμός των κύκλων ρολογιού αποτυχίας μνήμης για εντολές είναι:

Αριθμός κύκλων ρολογιού αποτυχίας μνήμης για εντολές = IC x 5% x 12 = 0,6 x IC

Ενώ για δεδομένα είναι :

Αριθμός κύκλων ρολογιού αποτυχίας μνήμης για δεδομένα = IC x 33% x 10% x 12
= 0,4 x IC

Γνωρίζουμε ότι η συχνότητα φόρτωσης και αποθήκευσης είναι 33%. Έτσι ο συνολικός αριθμός των κύκλων καθυστέρησης μνήμης είναι:

Αριθμός των κύκλων καθυστέρησης μνήμης ολικός = 0,6 x IC + 0,4 x IC = 1,0 x IC

Αυτό σημαίνει ένας κύκλος καθυστέρησης ανά εντολή . Τελικά :

Συνολικά CPI_{καθυστέρησης} = CPI_{ιδανική} + CPI = 4 + 1 = 5.

Αν δεν υπάρξει αλλαγή στον αριθμό των εντολών ή στο ρυθμό του ρολογιού, τότε έχουμε:

$$\frac{\text{Χρόνος ΚΜΕ}_{\text{με καθυστερήσεις}}}{\text{Χρόνος ΚΜΕ}_{\text{με ιδανική κρυφή μνήμη}}} = \frac{\text{IC} \times \text{CPI}_{\text{καθυστέρησης}} \times \text{κύκλοι ρολογιού}}{\text{IC} \times \text{CPI}_{\text{ιδανική κρυφή μνήμη}} \times \text{κύκλοι ρολογιού}} = \frac{5}{4}$$

$$\text{Χρόνος ΚΜΕ}_{\text{με ιδανική κρυφή μνήμη}} = \frac{\text{IC} \times \text{CPI}_{\text{ιδανική κρυφή μνήμη}} \times \text{κύκλοι ρολογιού}}{5}$$

Ανακεφαλαιώνοντας λοιπόν...



Η απόδοση της κρυφής μνήμης αυξάνεται αν επιτευχθεί η μείωση των αποτυχιών κρυφής μνήμης, της ποινής αποτυχίας της κρυφής μνήμης ή του χρόνου επιτυχίας, αφού τα τρία αυτά μεγέθη αθροιστικά δίνουν και τον μέσο χρόνο πρόσβασης στη μνήμη.



Από τις 7 τεχνικές μείωσης του ρυθμού αποτυχίας ξεχωρίζουμε εκείνη με τις υψηλότερες συσχετιστικότητες και τη κρυφή μνήμη θύμα. Όσο πηγαίνουμε σε υψηλότερες συσχετιστικότητες αυξάνεται ο κύκλος ρολογιού, ενώ στην κρυφή μνήμη θύμα, τοποθετούνται τα μπλοκ που απορρίπτονται από την κρυφή μνήμη σε μία αποτυχία.



Οι αποτυχίες κρυφής μνήμης χωρίζονται σε τρεις κατηγορίες: αναγκαστικές, χωρητικότητας και σύγκρουσης. Η μείωσή τους προσανατολίζεται στις τρεις αυτές κατευθύνσεις με δυσκολότερη περίπτωση εκείνη της χωρητικότητας.



Για τη μείωση της ποινής αποτυχίας δίνεται προτεραιότητα στις αποτυχίες ανάγνωσης έναντι στις αποτυχίες εγγραφής ή χρησιμοποιείται ένας καταχωρητής εγγραφής. Μία άλλη τεχνική που χρησιμοποιείται για τη μείωση της ποινής αποτυχίας είναι η χρήση κρυφής μνήμης 2^{ου} επιπέδου.



Για τη μείωση του χρόνου επιτυχίας χρησιμοποιούνται είτε μικρές και απλές μνήμες, είτε εικονικές κρυφές μνήμες, είτε σωλήνωση εγγραφών.