

# Αρχιτεκτονική Υπολογιστών

## Κεφάλαιο 1

### Αφηρημένες έννοιες και τεχνολογία υπολογιστών

[Έχει χρησιμοποιηθεί υλικό από τις διαφάνειες  
*Computer Organization and Design, 4th Edition,*  
Patterson & Hennessy, © 2008, MK]

## Η επανάσταση των υπολογιστών

§1.1 Εισαγωγή

- Που οφείλεται η πρόοδος των υπολογιστών;
  - Στην ραγδαία εξέλιξη της τεχνολογίας των ολοκληρωμένων κυκλωμάτων
    - Νόμος του Moore (Moore's Law)
- Γνωρίζετε πόσα τρανζίστορ κατασκευάζονται ετησίως μέσα σε ολοκληρωμένα κυκλώματα;
  - Το 2006, για πρώτη φορά στην ανθρώπινη ιστορία:
    - Παρήχθησαν περισσότερα τρανζίστορ (και με μικρότερο κόστος) από κόκκους ρυζιού !!!

## Εφαρμογές της υπολογιστικής

- Στο πρόσφατο παρελθόν ήταν «επιστημονική φαντασία» αλλά πλέον είναι υλοποιήσιμες
  - Υπολογιστές σε αυτοκίνητα
  - Κινητά τηλέφωνα
  - Ανάλυση του ανθρώπινου γονιδιώματος (DNA)
  - Παγκόσμιος ιστός
  - Μηχανές αναζήτησης
- Εφαρμογές «επιστημονικής φαντασίας» του μέλλοντος
  - Για σκεφτείτε ...

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 3

## Κατηγορίες υπολογιστών

- Επιτραπέζιοι υπολογιστές (desktop computers)
  - Γνωστοί και ως προσωπικοί υπολογιστές (personal computers)
  - Στόχος: κόστος/απόδοση
    - Έμφαση στις υψηλές επιδόσεις για ένα χρήστη
  - Από τις μεγαλύτερες αγορές
  - Οδηγεί την εξέλιξη στην τεχνολογία των υπολογιστών

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 4

## Διακομιστές (servers)

- Πολλοί χρήστες μαζί
- Υποστήριξη μεγάλου φόρτου εργασίας
  - Παραδείγματα:
    - μια πολύπλοκη επιστημονική ή τεχνική εφαρμογή,
    - διαχείριση πολλών μικρών εργασιών, π.χ. Web server
- Συνήθως είναι προσπελάσιμοι μέσω δικτύου
- Έμφαση στην φερεγγυότητα
  - Μια βλάβη είναι συνήθως πιο δαπανηρή από ό,τι σε ένα PC
- Μεγάλο φάσμα κόστους και δυνατοτήτων
  - Από ένα PC (κόστος 1K \$) έως τους υπερυπολογιστές (supercomputers) (κόστος 10s M \$)

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 5

## Υπερυπολογιστές

- Υψηλή απόδοση και κόστος
  - Αποτελούνται από χιλιάδες επεξεργαστές, terabyte μνήμης και petabyte αποθηκευτικού χώρου
- Εφαρμογές υψηλών απαιτήσεων
  - Επιστημονικούς και τεχνικούς υπολογισμούς, π.χ. η πρόγνωση του καιρού, ο προσδιορισμός της δομής των πρωτεϊνών
- Μικρό κλάσμα της συνολικής αγοράς
- Νο.1 υπερυπολογιστής (πηγή: [www.top500.org](http://www.top500.org), Νοε 2021):
  - SUPERCOMPUTER FUGAKU (Japan)
    - A64FX 48C 2.2GHz
    - 7,630,848 cores
    - Απόδοση (θεωρητική): 537,212 TFlop/s
    - ισχύς 29,899 kW
    - μνήμη 5,087,232 Gb

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 6

## Μεγέθη μνήμης

- Kilobyte –  $2^{10}$  ή 1024 bytes
- Megabyte–  $2^{20}$  ή 1,048,576 bytes
  - κάποιες φορές στρογγυλοποιείται σε  $10^6$  ή 1,000,000 bytes
- Gigabyte –  $2^{30}$  ή 1,073,741,824 bytes
  - στρογγυλοποιείται σε  $10^9$  ή 1,000,000,000 bytes
- Terabyte –  $2^{40}$  ή 1,099,511,627,776 bytes
  - στρογγυλοποιείται σε  $10^{12}$  ή 1,000,000,000,000 bytes
- Petabyte –  $2^{50}$  ή 1024 terabytes
  - στρογγυλοποιείται σε  $10^{15}$  ή 1,000,000,000,000,000 bytes
- Exabyte –  $2^{60}$  ή 1024 petabytes
  - στρογγυλοποιείται σε  $10^{18}$  ή 1,000,000,000,000,000,000 bytes

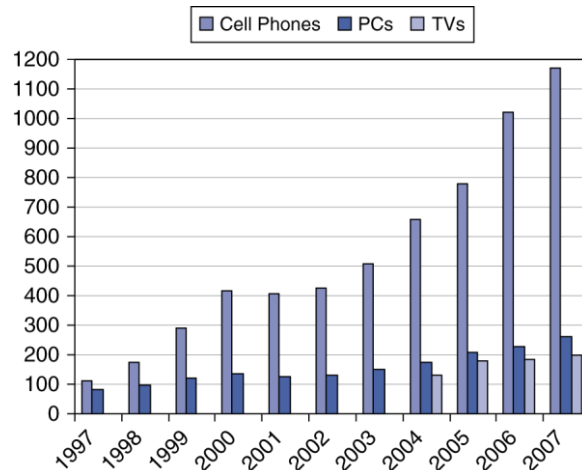
Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 7

## Ενσωματωμένοι υπολογιστές

- Κρυμμένοι ως συστατικά στοιχεία συσκευών
- Παραδείγματα: οι μικροεπεξεργαστές που υπάρχουν:
  - στο αυτοκίνητό σας, σε ένα κινητό τηλέφωνο, σε ένα βιντεοπαιχνίδι ή μια ψηφιακή τηλεόραση, τα δίκτυα επεξεργαστών ενός αεροπλάνου
- Σχεδιασμένοι να εκτελούν μία προκαθορισμένη εφαρμογή
- Μοναδικές απαιτήσεις που συνδυάζουν μια ελάχιστη απόδοση με αυστηρούς περιορισμούς κόστους και ισχύος
  - Για παράδειγμα, σε ένα κινητό τηλέφωνο ο επεξεργαστής πρέπει
    - Να είναι τόσο γρήγορος όσο είναι αναγκαίο για να χειριστεί τις περιορισμένες του λειτουργίες και, να κοστίζει ελάχιστα και να καταναλώνει λίγο
- Λιγότερο ανεκτικοί σε βλάβες
  - Τα αποτελέσματα της βλάβης μπορεί να είναι ενοχλητικά (όταν η νέα σας τηλεόραση δεν παίζει) έως ολέθρια (όταν ο υπολογιστής σε ένα αεροπλάνο ή ένα αυτοκίνητο αστοχήσει)

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 8

## Η αγορά των υπολογιστών



Αύξηση ενσωματωμένων >> Αύξηση των προσωπικών υπολογιστών

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 9

## Τι θα μάθετε

- Πώς τα προγράμματα μεταφράζονται σε γλώσσα του υλικού
  - και πώς το υλικό τα εκτελεί
- Ποια είναι η διασύνδεση μεταξύ λογισμικού και υλικού
  - πώς το λογισμικό καθοδηγεί το υλικό
- Τι καθορίζει την απόδοση ενός προγράμματος
  - και πώς ο προγραμματιστής μπορεί να τη βελτιώσει
- Πώς οι σχεδιαστές υλικού μπορούν να βελτιώσουν την απόδοση

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 10

## Κατανόηση της απόδοσης

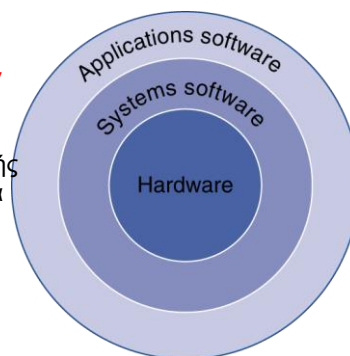
- **Αλγόριθμος**
  - Καθορίζει τον αριθμό των λειτουργιών που εκτελούνται (εντολές πηγαίου κώδικα και λειτουργίες εισόδου/εξόδου)
    - Θέματα απόδοσης αλγορίθμων καλύπτονται σε άλλα μαθήματα
- **Γλώσσα προγραμματισμού, μεταγλωττιστής, αρχιτεκτονική**
  - Καθορίζουν τον αριθμό των εντολών του υπολογιστή για κάθε εντολή πηγαίου κώδικα
- **Επεξεργαστής και σύστημα μνήμης**
  - Καθορίζουν πόσο γρήγορα εκτελούνται οι εντολές
- **Σύστημα εισόδου/εξόδου (υλικό και λειτουργικό σύστημα)**
  - Καθορίζει πόσο γρήγορα εκτελούνται οι λειτουργίες εισόδου/εξόδου

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 11

## Κάτω από το πρόγραμμά σας

### λογισμικό εφαρμογών (application software)

Μια τυπική εφαρμογή, όπως ένας επεξεργαστής κειμένου ή ένα σύστημα βάσης δεδομένων



### λογισμικό συστημάτων (systems software)

Λογισμικό που παρέχει υπηρεσίες συστήματος, όπως τα λειτουργικά συστήματα, οι μεταγλωττιστές, και οι συμβολομεταφραστές (assemblers)

- Το λογισμικό εφαρμογών γράφεται σε γλώσσα υψηλού επιπέδου
- Το υλικό μπορεί να εκτελέσει απλές εντολές χαμηλού επιπέδου
  - Η μετάβαση περιλαμβάνει πολλά επίπεδα λογισμικού που μεταφράζουν λειτουργίες υψηλού επιπέδου σε απλές εντολές του υπολογιστή

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 12

## Λογισμικό συστημάτων

- **Λειτουργικό σύστημα (operating system)**
  - Πρόγραμμα επίβλεψης που διαχειρίζεται τους πόρους ενός υπολογιστή και διασυνδέει τα προγράμματα χρήστη με το υλικό
  - Βασικές λειτουργίες:
    - Διαχείριση λειτουργιών εισόδου/εξόδου
    - Κατανομή μνήμης
    - Κοινή χρήση του επεξεργαστή από πολλές εφαρμογές
  - Παραδείγματα σύγχρονων λειτουργικών συστημάτων:
    - Windows, Linux, MacOS
- **Μεταγλωττιστής (compiler)**
  - Πρόγραμμα που μεταφράζει εντολές μιας γλώσσας υψηλού επιπέδου σε εντολές συμβολικής γλώσσας (assembly language)
- **Συμβολομεταφραστής (assembler)**
  - Πρόγραμμα που μεταφράζει μια συμβολική έκδοση των εντολών στη δυαδική έκδοσή τους

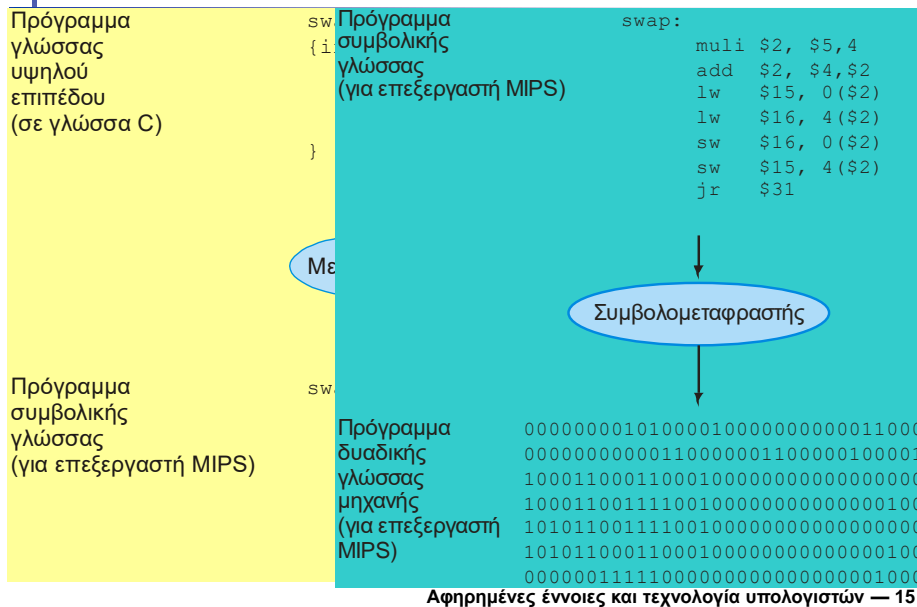
Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 13

## Επίπεδα προγραμμάτων

- **Γλώσσα μηχανής (machine language)**
  - Εντολές (instructions) : συλλογές δυαδικών ψηφίων (bit) κατανοητές από τον υπολογιστή
    - Π.χ. Η ακολουθία των bit `1000110010100000` δηλώνει σε έναν υπολογιστή να εκτελέσει την πρόσθεση `A + B`
- **Συμβολική γλώσσα (assembly language)**
  - Μια συμβολική αναπαράσταση των εντολών της μηχανής
    - Π.χ. Η εντολή `add A, B` σημαίνει «εκτέλεσε την πρόσθεση `A+B`» (το ισοδύναμο δηλαδή της `1000110010100000`)
- **Γλώσσα προγραμματισμού υψηλού επιπέδου (high-level programming language)**
  - Επίπεδο αφαίρεσης πιο κοντά στο πεδίο του προβλήματος
  - Αποτελείται από λέξεις και αλγεβρικές παραστάσεις
    - Π.χ `A + B` σημαίνει `add A, B` (σε συμβολική γλώσσα)

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 14

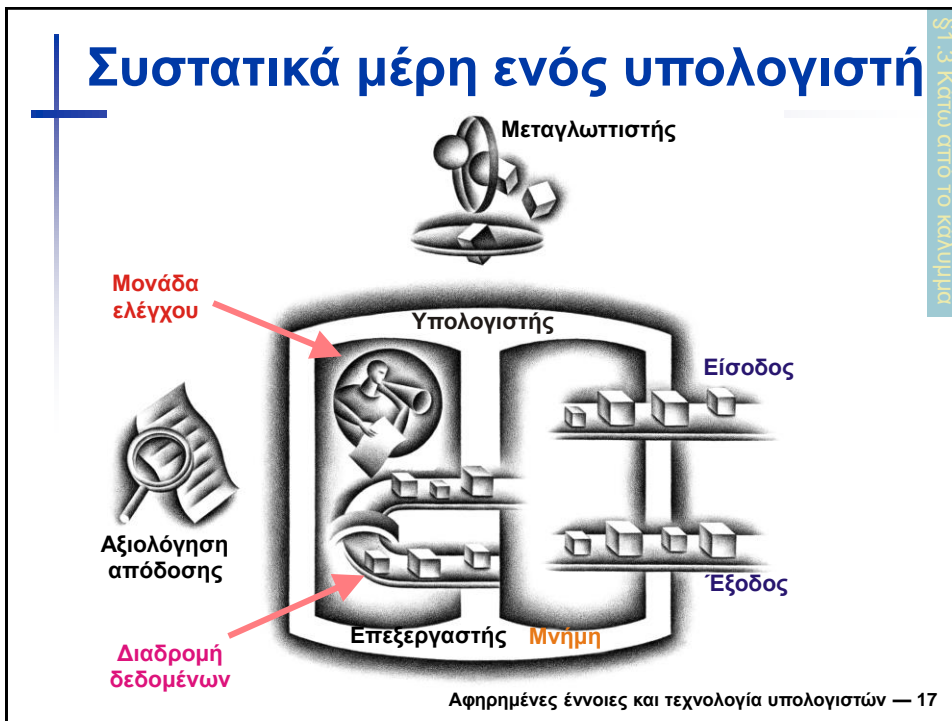
## Μεταγλώττιση, συμβολομετάφραση



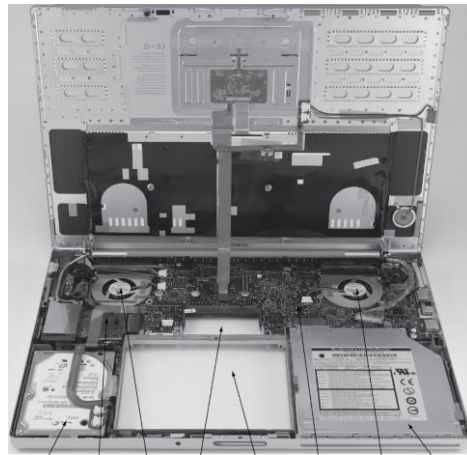
## Γιατί προγραμματίζουμε σε HLL;

- Επιτρέπουν να σκεπτόμαστε σε μια πιο φυσική γλώσσα
  - Τα προγράμματα μοιάζουν περισσότερο με κείμενα παρά με πίνακες από δυσνόητα σύμβολα
- Επιτρέπουν το σχεδιασμό γλωσσών με βάση την εφαρμογή
  - Π.χ. η Fortran σχεδιάστηκε για επιστημονικούς υπολογισμούς
- Βελτιώνουν την παραγωγικότητα του προγραμματιστή
  - Η λακωνικότητα είναι πάντα πλεονέκτημα
- Επιτρέπουν στα προγράμματα να είναι ανεξάρτητα από συγκεκριμένο υπολογιστή
  - οι μεταγλωττιστές και οι συμβολομεταφραστές μπορούν να μεταφράσουν προγράμματα γλωσσών υψηλού επιπέδου στη γλώσσα μηχανής οποιουδήποτε υπολογιστή

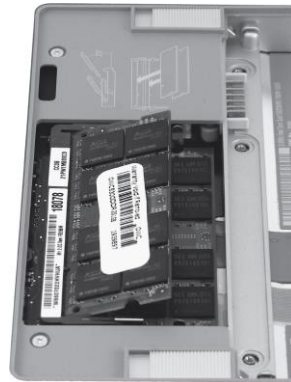




## Ανοίγοντας το κουτί



Hard drive Processor Fan with cover Spot for memory DIMMs Spot for battery Motherboard Fan with cover DVD drive



Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 19

## Βασική ορολογία

- **Μητρική κάρτα (motherboard)**
  - Μια πλαστική κάρτα που περιέχει ολοκληρωμένα κυκλώματα, μεταξύ των οποίων ο επεξεργαστής, η μνήμη, και συνδέσεις για τις συσκευές I/O, όπως δίκτυα και δίσκοι
- **Ολοκληρωμένο κύκλωμα (integrated circuit)**
  - Ονομάζεται επίσης τσιπ (chip). Μια συσκευή που συνδυάζει δεκάδες έως δισεκατομμύρια τρανζίστορ
- **Μνήμη (memory)**
  - Η περιοχή αποθήκευσης στην οποία διατηρούνται τα προγράμματα όταν εκτελούνται, και η οποία περιέχει τα δεδομένα που χρειάζονται τα προγράμματα αυτά

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 20

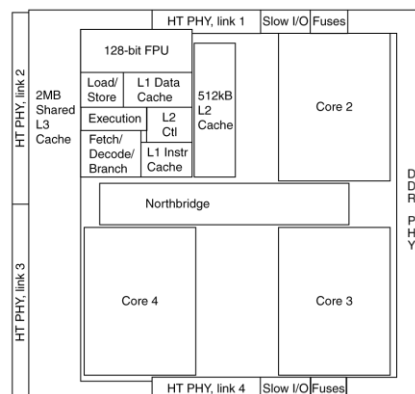
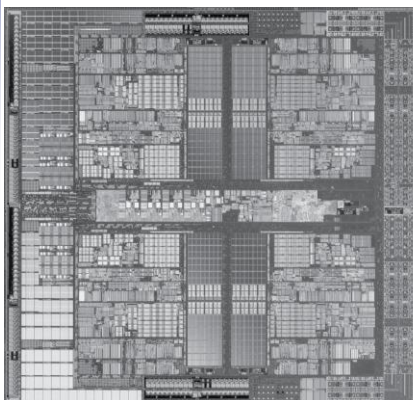
## Μέσα στον επεξεργαστή (CPU)

- **Κεντρική μονάδα επεξεργασίας (central processor unit - CPU)**
  - Ονομάζεται και επεξεργαστής. Το ενεργό τμήμα του υπολογιστή που διαβάζει από τη μνήμη, εκτελεί τις εντολές, ενεργοποιεί τις συσκευές I/O, κοκ.
- **Διαδρομή δεδομένων (datapath)**
  - Το τμήμα τού επεξεργαστή που εκτελεί τις πράξεις
- **Μονάδα ελέγχου (control unit)**
  - Το τμήμα του επεξεργαστή που ελέγχει τη διαδρομή δεδομένων, τη μνήμη, και τις συσκευές I/O σύμφωνα με τις εντολές του προγράμματος

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 21

## Μέσα στον επεξεργαστή

- AMD Barcelona: 4 πυρήνες



Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 22

## Αποθήκευση δεδομένων

- Κατηγορίες μνημών ανάλογα με το αν διατηρούν ή όχι τα δεδομένα
- **Πτητική μνήμη (volatile memory)**
  - Διατηρεί τα δεδομένα μόνον όταν τροφοδοτείται με ηλεκτρικό ρεύμα
  - Κύρια μνήμη (DRAM)
- **Μη πτητική μνήμη (nonvolatile memory)**
  - Διατηρεί τα δεδομένα ακόμη και σε απουσία ηλεκτρικού ρεύματος
  - Μαγνητικοί δίσκοι
  - Μνήμες flash
  - Οπτικοί δίσκοι (CDROM, DVD)



Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 23

## Αποθήκευση δεδομένων

- Κατηγορίες μνημών ανάλογα με τη χρήση τους
- **Πρωτεύουσα μνήμη (primary memory)**
  - Ονομάζεται επίσης κύρια μνήμη (main memory)
  - Χρησιμοποιείται για να διατηρεί προγράμματα ενώ εκτελούνται
  - Πτητική μνήμη
  - Στους σημερινούς υπολογιστές αποτελείται από DRAM
- **Δευτερεύουσα μνήμη (secondary memory)**
  - Χρησιμοποιείται για την αποθήκευση προγραμμάτων και δεδομένων μεταξύ διαδοχικών τους ενεργοποιήσεων
  - Μη πτητική μνήμη
  - Στους σημερινούς υπολογιστές αποτελείται από μαγνητικούς δίσκους

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 24

## Κύρια μνήμη και δίσκος

- Οι δίσκοι είναι **μη πτητικοί** επειδή είναι μαγνητικοί
- Οι δίσκοι έχουν **μεγαλύτερο χρόνο προσπέλασης** επειδή είναι μηχανικές συσκευές
- Οι δίσκοι είναι **φθηνότεροι** ανά μονάδα αποθήκευσης επειδή έχουν πολύ μεγάλη χωρητικότητα αποθήκευσης με μέτριο κόστος

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 25

## Τεχνολογίες κατασκευής

- **Τρανζίστορ (transistor)**
  - Ένας ανοιχτός/κλειστός διακόπτης που ελέγχεται από ένα ηλεκτρικό σήμα
- **Πολύ μεγάλης κλίμακας ολοκληρωμένο κύκλωμα (very large scale integrated - VLSI circuit)**
  - Μια συσκευή που περιέχει εκατοντάδες χιλιάδες έως εκατομμύρια τρανζίστορ
- **Νόμος του Moore (Moore's law) :**
  - η χωρητικότητα σε τρανζίστορ ενός ολοκληρωμένου κυκλώματος διπλασιάζεται κάθε 18-24 μήνες
  - πρόβλεψη για την χωρητικότητα των ολοκληρωμένων κυκλωμάτων που έγινε από τον Gordon Moore, έναν από τους ιδρυτές της Intel, τη δεκαετία του '60

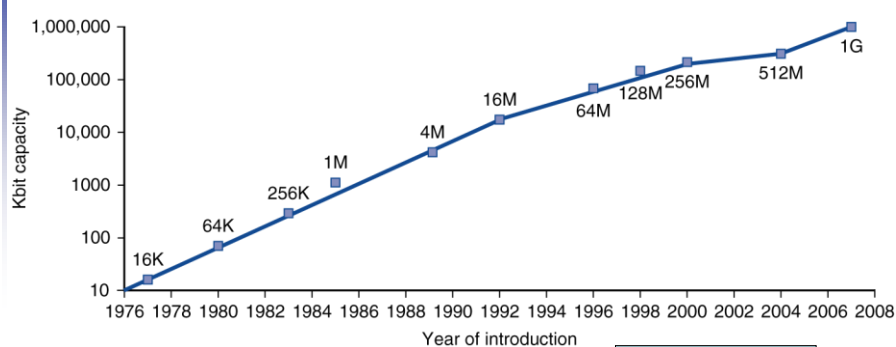
Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 26

## Απόδοση κυκλωμάτων

Έτος	Τεχνολογία που χρησιμοποιείται στους υπολογιστές	Σχετική απόδοση/ κόστος
1951	Λυχνία κενού	1
1965	Τρανζίστορ	35
1975	Integrated circuit (IC)	900
1995	Very large scale integrated circuit (VLSI)	2,400,000
2005	Ultra large scale integrated circuit (ULSI)	6,200,000,000

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 27

## Εξέλιξη της τεχνολογίας



- Η τεχνολογία συνεχίζει να εξελίσσεται
  - Αυξάνεται η χωρητικότητα και η απόδοση
  - Μειώνεται το κόστος

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 28

## ITRS Roadmap

Έτος	2006	2008	2010	2012	2015	2020
Τεχνολογία (nm)	65	45	32	22	11	6
Χωρητικότητα (BT)	4	8	16	32	64	128

Πηγή: International Technology Roadmap for Semiconductors (ITRS), 2008

- Τεχνολογία των επεξεργαστών υψηλής απόδοσης
- «Ανέκδοτα» για την τεχνολογία των 45nm
  - 30 εκατομμύρια τρανζίστορ μπορούν να χωρέσουν στη μύτη μιας βελόνας
  - Περισσότερα από 2000 στο πάχος μιας ανθρώπινης τρίχας
  - Εάν οι τιμές των αυτοκινήτων έπεφταν με τον ίδιο ρυθμό που πέφτουν οι τιμές των τρανζίστορ από το 1968, ένα νέο αυτοκίνητο θα έπρεπε σήμερα να κοστίζει 1 λεπτό

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 29

## Απόδοση υπολογιστή

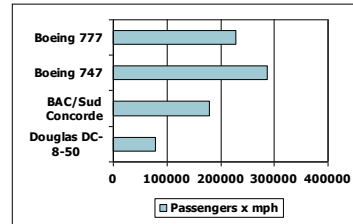
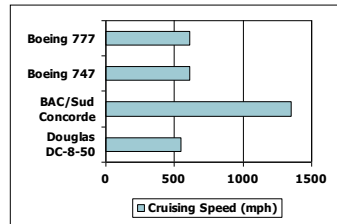
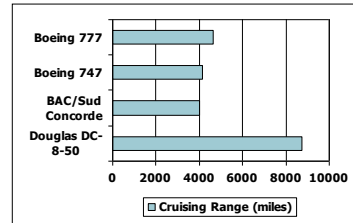
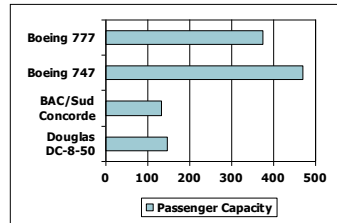
§1.4 Απόδοση

- Πως υπολογίζεται η απόδοση ενός υπολογιστή ;
- Ποιοι παράγοντες του υλικού επιδρούν στην συνολική απόδοση του συστήματος;
  - Πόσο αποδοτικά το πρόγραμμα χρησιμοποιεί τις εντολές του υπολογιστή;
  - Πόσο αποδοτικά το υλικό υλοποιεί τις εντολές;
  - Πόσο αποδοτικά είναι τα συστήματα μνήμης και εισόδου/εξόδου;

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 30

## Ορισμός απόδοσης

- Τι εννοούμε όταν λέμε ότι ένας υπολογιστής έχει καλύτερη απόδοση από έναν άλλο;
  - Παράδειγμα:** αναλογία με την απόδοση επιβατικών αεροσκαφών



Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 31

## Τι μετράμε;

- Πως θα συγκρίνουμε δυο διαφορετικά PCs;
  - Θα εκτελέσουμε το ίδιο πρόγραμμα και στα δυο για να δούμε ποιο θα το εκτελέσει ταχύτερα
- Αυτό που μετράμε είναι ο **χρόνος απόκρισης (response time)** ή αλλιώς **χρόνος εκτέλεσης (execution time)**
  - Ο συνολικός χρόνος που απαιτεί ο υπολογιστής για να ολοκληρώσει μια εργασία, μεταξύ των οποίων:
    - η προσπέλαση δίσκου,
    - η προσπέλαση μνήμης,
    - οι ενέργειες εισόδου/εξόδου,
    - η επιβάρυνση του λειτουργικού συστήματος,
    - ο χρόνος εκτέλεσης της CPU, κοκ.

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 32



## Τι μετράμε;

- Σε ένα κέντρο δεδομένων με πολλούς διακομιστές (servers) να εκτελούν τις εργασίες πολλών χρηστών, ποιος υπολογιστής θα λέγαμε ότι είναι ταχύτερος ;
  - αυτός που ολοκληρώνει τις περισσότερες εργασίες κατά τη διάρκεια της ημέρας
- Αυτό που μετράμε είναι η **διεκπεραιωτική ικανότητα (throughput)**
  - συνολική ποσότητα εργασίας που ολοκληρώνεται σε δεδομένο χρόνο

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 33

## Χρόνος και διεκπερ. ικανότητα

- Οι επόμενες αλλαγές σε ένα υπολογιστικό σύστημα αυξάνουν την διεκπεραιωτική ικανότητα, μειώνουν το χρόνο απόκρισης, ή και τα δύο;
  - Αντικατάσταση του ενός επεξεργαστή με μια ταχύτερη έκδοση
  - Προσθήκη επιπλέον επεξεργαστών σε ένα σύστημα που χρησιμοποιεί πολλούς επεξεργαστές για διαφορετικές εργασίες — π.χ. αναζήτηση στον διαδίκτυο
- Θα επικεντρωθούμε στο χρόνο εκτέλεσης

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 34

## Απόδοση και χρόνος εκτέλεσης

- Για τη μεγιστοποίηση της απόδοσης, πρέπει να ελαχιστοποιήσουμε το χρόνο εκτέλεσης
- Σχέση της απόδοσης με το χρόνο εκτέλεσης για έναν υπολογιστή X

$$\text{Απόδοση}_X = \frac{1}{\text{Χρόνος εκτέλεσης}_X}$$

## Σχετική απόδοση

- Άρα μεταξύ υπολογιστών X και Y
  - Απόδοση<sub>X</sub> > Απόδοση<sub>Y</sub>
- σημαίνει
  - Χρόνος εκτέλεσης<sub>Y</sub> > Χρόνος εκτέλεσης<sub>X</sub>
- Επίσης «ο X είναι n φορές ταχύτερος από τον Y» σημαίνει:

$$\frac{\text{Απόδοση}_X}{\text{Απόδοση}_Y} = n \qquad \frac{\text{Απόδοση}_X}{\text{Απόδοση}_Y} = \frac{\text{Χρόνος εκτέλεσης}_Y}{\text{Χρόνος εκτέλεσης}_X} = n$$

## Μέτρηση χρόνου εκτέλεσης

- **Παρελθών χρόνος (elapsed time)**
  - Συνολικός χρόνος απόκρισης που περιλαμβάνει
    - Επεξεργασία, I/O, επιβάρυνση του λειτουργικού, ανενεργός χρόνος
  - Καθορίζει την απόδοση του συστήματος
- Ένας υπολογιστής μπορεί να εκτελεί ταυτόχρονα πολλά προγράμματα
  - Πρέπει να ξεχωρίσουμε τον παρελθόντα χρόνο από το χρόνο που ο επεξεργαστής εργάζεται για το συγκεκριμένο πρόγραμμα
- **Χρόνος εκτέλεσης CPU (CPU execution time) ή απλώς χρόνος CPU (CPU time)**
  - ο πραγματικός χρόνος που καταναλώνει η CPU για μια εργασία
    - δεν περιλαμβάνει το χρόνο που δαπανάται στην αναμονή για είσοδο/έξοδο ή στην εκτέλεση άλλων προγραμμάτων, κα.
  - Ο χρόνος απόκρισης που αντιλαμβάνεται ο χρήστης θα είναι ο παρελθών χρόνος του προγράμματος και όχι ο χρόνος CPU

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 37

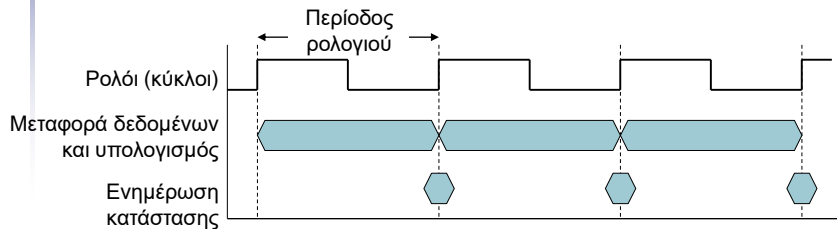
## Μέτρηση χρόνου εκτέλεσης

- Ο χρόνος CPU διαιρείται σε:
  - **χρόνο CPU χρήστη (user CPU time)**: Ο χρόνος CPU που δαπανάται για το ίδιο το πρόγραμμα
  - **χρόνο CPU συστήματος (system CPU time)**: Ο χρόνος CPU που δαπανάται από το λειτουργικό σύστημα στην εκτέλεση εργασιών για λογαριασμό του προγράμματος
- Η διάκριση μεταξύ χρόνου CPU χρήστη και συστήματος είναι δύσκολη

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 38

## Ρολόι της CPU

- Η λειτουργία των ψηφιακών κυκλωμάτων ελέγχεται από ένα ρολόι σταθερού ρυθμού



- Περίοδος ρολογιού: διάρκεια ενός κύκλου
  - π.χ.,  $250\text{ps} = 0.25\text{ns} = 250 \times 10^{-12}\text{s}$
- Συχνότητα ρολογιού (ρυθμός): κύκλοι ανά sec
  - π.χ.,  $4.0\text{GHz} = 4000\text{MHz} = 4.0 \times 10^9\text{Hz}$

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 39

## Χρόνος CPU

- Συσχέτιση χρόνου CPU με κύκλους ρολογιού :

Χρόνος εκτέλεσης CPU για ένα πρόγραμμα =  $\frac{\text{Κύκλοι ρολογιού CPU για ένα πρόγραμμα}}{\text{Ρυθμός ρολογιού}}$

$$\text{Χρόνος εκτέλεσης CPU για ένα πρόγραμμα} = \frac{\text{Κύκλοι ρολογιού CPU για ένα πρόγραμμα}}{\text{Ρυθμός ρολογιού}}$$

- Άρα ο σχεδιαστής ενός επεξεργαστή μπορεί να βελτιώσει την απόδοση μειώνοντας :
  - είτε την περίοδο του ρολογιού
  - είτε τον αριθμό των κύκλων ρολογιού του προγράμματος
  - ο σχεδιαστής πρέπει να βρει τον καλύτερο συμβιβασμό μεταξύ της ταχύτητας του ρολογιού και του αριθμού των κύκλων

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 40

## Παράδειγμα 1: ρυθμός ρολογιού

- Υπολογιστής A: χρόνος εκτέλεσης προγράμματος 10 sec, ρολόι 4 GHz
- Κατασκευή ενός υπολογιστή B, που θα εκτελεί το πρόγραμμα σε 6 sec
- Σχεδιαστική αλλαγή επιφέρει σημαντική αύξηση στο ρυθμό ρολογιού, αλλά επηρεάζει την υπόλοιπη σχεδίαση της CPU:
  - Ο υπολογιστής B απαιτεί 1,2 φορές τους κύκλους ρολογιού που απαιτεί ο υπολογιστής A
- Ποιο ρυθμό ρολογιού πρέπει να έχει ο υπολογιστής B;

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 41

## Παράδειγμα 1: απάντηση

- Ο αριθμός κύκλων ρολογιού στον A είναι:

$$\text{Χρόνος CPU}_A = \frac{\text{Κύκλοι ρολογιού CPU}_A}{\text{Ρυθμός ρολογιού}_A} \quad 10 \text{ δευτερόλεπτα} = \frac{\text{Κύκλοι ρολογιού CPU}_A}{4 \times 10^9 \frac{\text{κύκλοι}}{\text{δευτερόλεπτο}}}$$

$$\text{Κύκλοι ρολογιού CPU}_A = 10 \text{ δευτερόλεπτα} \times 4 \times 10^9 \frac{\text{κύκλοι}}{\text{δευτερόλεπτο}} = 40 \times 10^9 \text{ κύκλοι}$$

- Ο χρόνος CPU του προγράμματος στον B:

$$\text{Χρόνος CPU}_B = \frac{1,2 \times \text{Κύκλοι ρολογιού CPU}_A}{\text{Ρυθμός ρολογιού}_B} \quad 6 \text{ δευτερόλεπτα} = \frac{1,2 \times 40 \times 10^9 \text{ κύκλοι}}{\text{Ρυθμός ρολογιού}_B}$$

$$\text{Ρυθμός ρολογιού}_B = \frac{1,2 \times 40 \times 10^9 \text{ κύκλοι}}{6 \text{ δευτερόλεπτα}} = \frac{8 \times 10^9 \text{ κύκλοι}}{\text{δευτερόλεπτο}} = 8 \text{ GHz}$$

- Άρα ο B πρέπει να έχει διπλάσιο ρυθμό από τον A

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 42

## Αριθμός εντολών και CPI

- Ως τώρα δεν αναφέραμε τον αριθμό των εντολών του προγράμματος
- Ωστόσο, ο χρόνος εκτέλεσης εξαρτάται από τον **αριθμό των εντολών (instruction count)** που εκτελεί ο επεξεργαστής
  - Καθορίζεται από το πρόγραμμα, το σύνολο εντολών και το μεταγλωττιστή
- Average **cycles per instruction (CPI)**
  - Μέσος αριθμός κύκλων ρολογιού ανά εντολή για ένα πρόγραμμα
  - Καθορίζεται από το υλικό της CPU
  - Διαφορετικές εντολές μπορεί να έχουν διαφορετικό CPI

$\text{Clock Cycles} = \text{Instruction Count} \times \text{Cycles per Instruction}$

$\text{CPU Time} = \text{Instruction Count} \times \text{CPI} \times \text{Clock Cycle Time}$

$$= \frac{\text{Instruction Count} \times \text{CPI}}{\text{Clock Rate}}$$

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 43

## Παράδειγμα 2: CPI

- Υποθέστε δύο υλοποιήσεις της ίδιας αρχιτεκτονικής συνόλου εντολών
- Υπολογιστής A: χρόνο κύκλου ρολογιού 250 ps και CPI 2,0 για κάποιο πρόγραμμα
- Υπολογιστής B: χρόνο κύκλου ρολογιού 500 ps και CPI 1,2 για το ίδιο πρόγραμμα
- Ποιος υπολογιστής είναι ταχύτερος γι' αυτό το πρόγραμμα και πόσο;

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 44

## Παράδειγμα 2: απάντηση

- Και ο A και ο B εκτελούν το ίδιο αριθμό εντολών  $I$  για ένα πρόγραμμα
- Οι κύκλοι ρολογιού για κάθε υπολογιστή είναι:
  - Κύκλοι ρολογιού  $CPU_A = I \times 2,0$
  - Κύκλοι ρολογιού  $CPU_B = I \times 1,2$
- Άρα, ο χρόνος CPU για τους δύο υπολογιστές:
  - Χρόνος  $CPU_A = \text{Κύκλοι ρολογιού } CPU_A \times \text{Χρόνος κύκλου ρολογιού}_A = I \times 2,0 \times 250 \text{ ps} = 500 \times I \text{ ps}$
  - Χρόνος  $CPU_B = I \times 1,2 \times 500 \text{ ps} = 600 \times I \text{ ps}$

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 45

## Παράδειγμα 2: απάντηση

- Προφανώς ο A είναι ταχύτερος αφού:
  - Χρόνος  $CPU_A = 500 \times I \text{ ps}$
  - Χρόνος  $CPU_B = 600 \times I \text{ ps}$
- Πόσο;

$$\frac{\text{Απόδοση } CPU_A}{\text{Απόδοση } CPU_B} = \frac{\text{Χρόνος εκτέλεσης}_B}{\text{Χρόνος εκτέλεσης}_A} = \frac{600 \times I \text{ ps}}{500 \times I \text{ ps}} = 1,2$$

- Άρα, ο A είναι 1,2 φορές ταχύτερος του B για το συγκεκριμένο πρόγραμμα.

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 46

## Βασική εξίσωση απόδοσης

- Και οι τρεις παράγοντες μετέχουν τώρα:

Χρόνος CPU = Πλήθος εντολών × CPI × Χρόνος κύκλου ρολογιού

$$\text{Χρόνος CPU} = \frac{\text{Πλήθος εντολών} \times \text{CPI}}{\text{Ρυθμός ρολογιού}}$$

- Μπορούμε να χρησιμοποιήσουμε τους τύπους
  - για να συγκρίνουμε δύο διαφορετικές υλοποιήσεις ή
  - για να αξιολογήσουμε μια σχεδιαστική επιλογή αν γνωρίζουμε την επίδρασή της στις 3 παραμέτρους

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 47

## Μέτρηση των παραγόντων

Χρόνος CPU = Πλήθος εντολών × CPI × Χρόνος κύκλου ρολογιού

- Πώς μπορούμε να μετρήσουμε την τιμή αυτών των παραγόντων;
  - Χρόνος εκτέλεσης CPU : εκτελώντας το πρόγραμμα
  - Χρόνος κύκλου ρολογιού : στην τεκμηρίωση ενός υπολογιστή
  - Πλήθος εντολών : με εργαλεία λογισμικού που αναλύουν την εκτέλεση ή χρησιμοποιώντας έναν προσομοιωτή της αρχιτεκτονικής
- Το πλήθος των εντολών εξαρτάται από την αρχιτεκτονική
  - δεν εξαρτάται από τη συγκεκριμένη υλοποίηση
  - μπορούμε να μετρήσουμε το πλήθος εντολών χωρίς να γνωρίζουμε όλες τις λεπτομέρειες της υλοποίησης
- Το CPI εξαρτάται από μια μεγάλη ποικιλία σχεδιαστικών λεπτομερειών, π.χ. το σύστημα μνήμης, η δομή του επεξεργαστή, το μίγμα των τύπων εντολών που εκτελούνται σε μια εφαρμογή, κα.
  - Το CPI ποικίλει ανά εφαρμογή, όπως επίσης και μεταξύ υλοποιήσεων με το ίδιο σύνολο εντολών

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 48



## Βασική εξίσωση απόδοσης

$$\text{Χρόνος} = \frac{\text{Δευτερόλεπτα}}{\text{Πρόγραμμα}} = \frac{\text{Εντολές}}{\text{Πρόγραμμα}} \times \frac{\text{Κύκλοι ρολογιού}}{\text{Εντολή}} \times \frac{\text{Δευτερόλεπτα}}{\text{Κύκλος ρολογιού}}$$

- Το μόνο αξιόπιστο μέτρο είναι ο χρόνος
  - Η αλλαγή του συνόλου εντολών για τη μείωση του πλήθους τους μπορεί να οδηγήσει σε μια οργάνωση με πιο αργό κύκλο ρολογιού
    - Άρα, μπορεί να μην βελτιώσει την απόδοση
  - Το CPI εξαρτάται από τον τύπο των εντολών που εκτελούνται
    - Άρα, ο κώδικας που εκτελεί το μικρότερο πλήθος εντολών μπορεί να μην είναι ο ταχύτερος

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 49

## Κύκλοι ρολογιού και CPI

- Μπορούμε να υπολογίσουμε τους κύκλους ρολογιού CPU μετρώντας:
  - τους διαφορετικούς τύπους εντολών και
  - τα ξεχωριστά τους πλήθη κύκλων ρολογιού

$$\text{Κύκλοι ρολογιού CPU} = \sum_{i=1}^n (\text{CPI}_i \times C_i)$$

- $\text{CPI}_i$  : CPI της κατηγορίας  $i$  και
- $C_i$  : πλήθος εντολών της κατηγορίας  $i$

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 50

## Συστατικά υλικού/λογισμικού

Από ποια συστατικά εξαρτάται η απόδοση ενός προγράμματος και τι επηρεάζουν:

- **Αλγόριθμος**
  - Πλήθος εντολών, πιθανόν το CPI
- **Γλώσσα προγραμματισμού**
  - Πλήθος εντολών, CPI
- **Μεταγλωττιστής**
  - Πλήθος εντολών, CPI
- **Αρχιτεκτονική συνόλου εντολών**
  - Πλήθος εντολών, ρυθμό ρολογιού, CPI

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 51

## Συστατικά υλικού/λογισμικού

- Με ποιον τρόπο επηρεάζουν την απόδοση:
- **Αλγόριθμος**
  - Καθορίζει τον αριθμό εντολών του πηγαίου προγράμματος που εκτελούνται και, συνεπώς, τον αριθμό των εκτελούμενων εντολών του επεξεργαστή
  - Μπορεί να επηρεάσει επίσης το CPI, ευνοώντας πιο αργές ή πιο γρήγορες εντολές.
    - Για παράδειγμα, αν ο αλγόριθμος χρησιμοποιεί περισσότερες πράξεις (λειτουργίες) κινητής υποδιαστολής, συνήθως έχει υψηλότερο CPI
- **Γλώσσα προγραμματισμού**
  - Επηρεάζει το πλήθος εντολών, αφού οι εντολές της γλώσσας μεταφράζονται σε εντολές του επεξεργαστή, κάτι που καθορίζει το πλήθος εντολών
  - Μπορεί επίσης να επηρεάσει το CPI λόγω των χαρακτηριστικών της

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 52

## Συστατικά υλικού/λογισμικού

- **Μεταγλωττιστής**
  - Αποφασίζει τη μετάφραση των εντολών της πηγαίας γλώσσας σε εντολές του υπολογιστή
    - Επομένως, η αποδοτικότητα του μεταγλωττιστή επηρεάζει τόσο το πλήθος των εντολών όσο και τους μέσους κύκλους ανά εντολή.
- **Αρχιτεκτονική συνόλου εντολών**
  - Επηρεάζει και τις τρεις πλευρές της απόδοσης της CPU
    - τις εντολές που χρειάζονται για μια συνάρτηση,
    - το κόστος σε κύκλους της κάθε εντολής, και
    - το συνολικό ρυθμό ρολογιού του επεξεργαστή.

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 53

## Παράδειγμα 3: Σύγκριση κώδικα

- Ο σχεδιαστής ενός μεταγλωττιστή προσπαθεί να διαλέξει ανάμεσα σε δύο ακολουθίες κώδικα για έναν υπολογιστή.
- Οι σχεδιαστές του υλικού παρέχουν τα εξής στοιχεία:

	CPI γι' αυτή την κατηγορία εντολών		
	A	B	C
CPI	1	2	3

- Για μια συγκεκριμένη εντολή γλώσσας υψηλού επιπέδου, ο συγγραφέας του μεταγλωττιστή λαμβάνει υπόψη του δύο ακολουθίες κώδικα που απαιτούν τα παρακάτω πλήθη εντολών:

Ακολουθία κώδικα	Πλήθος εντολών για την κατηγορία εντολών		
	A	B	C
1	2	1	2
2	4	1	1

- Ποια ακολουθία κώδικα εκτελεί τις περισσότερες εντολές; Ποια θα είναι η ταχύτερη; Ποιο είναι το CPI για κάθε ακολουθία;

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 54

## Παράδειγμα 3: απάντηση

- Η ακολουθία 1 εκτελεί  $2+1+2 = 5$  εντολές
- Η ακολουθία 2 εκτελεί  $4+1+1 = 6$  εντολές
- Για να βρούμε τους κύκλους ρολογιού για κάθε ακολουθία :

$$\text{Κύκλοι ρολογιού CPU} = \sum_{i=1}^n (\text{CPI}_i \times C_i)$$

Κύκλοι ρολογιού  $\text{CPU}_1 =$

$$(2 \times 1) + (1 \times 2) + (2 \times 3) = 2 + 2 + 6 = 10 \text{ κύκλοι}$$

Κύκλοι ρολογιού  $\text{CPU}_2 =$

$$(4 \times 1) + (1 \times 2) + (1 \times 3) = 4 + 2 + 3 = 9 \text{ κύκλοι}$$

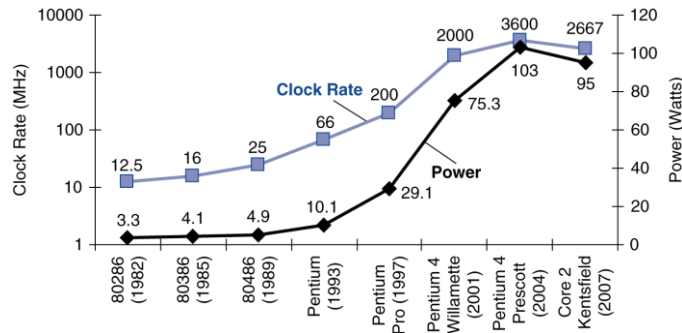
## Παράδειγμα 3: απάντηση

- Αφού
  - Κύκλοι ρολογιού  $\text{CPU}_1 = 10 \text{ κύκλοι}$
  - Κύκλοι ρολογιού  $\text{CPU}_2 = 9 \text{ κύκλοι}$
- Η ακολουθία 2 είναι ταχύτερη παρόλο που εκτελεί μία εντολή παραπάνω (6 έναντι 5)
- Για το CPI

$$\text{CPI}_1 = \frac{\text{Κύκλοι ρολογιού CPU}_1}{\text{Πλήθος εντολών}_1} = \frac{10}{5} = 2$$

$$\text{CPI}_2 = \frac{\text{Κύκλοι ρολογιού CPU}_2}{\text{Πλήθος εντολών}_2} = \frac{9}{6} = 1,5$$

# Ισχύς



- Σε τεχνολογία ολοκληρωμένων κυκλωμάτων CMOS

$$\text{Power} = \text{Capacitive load} \times \text{Voltage}^2 \times \text{Frequency}$$

×30

5V → 1V

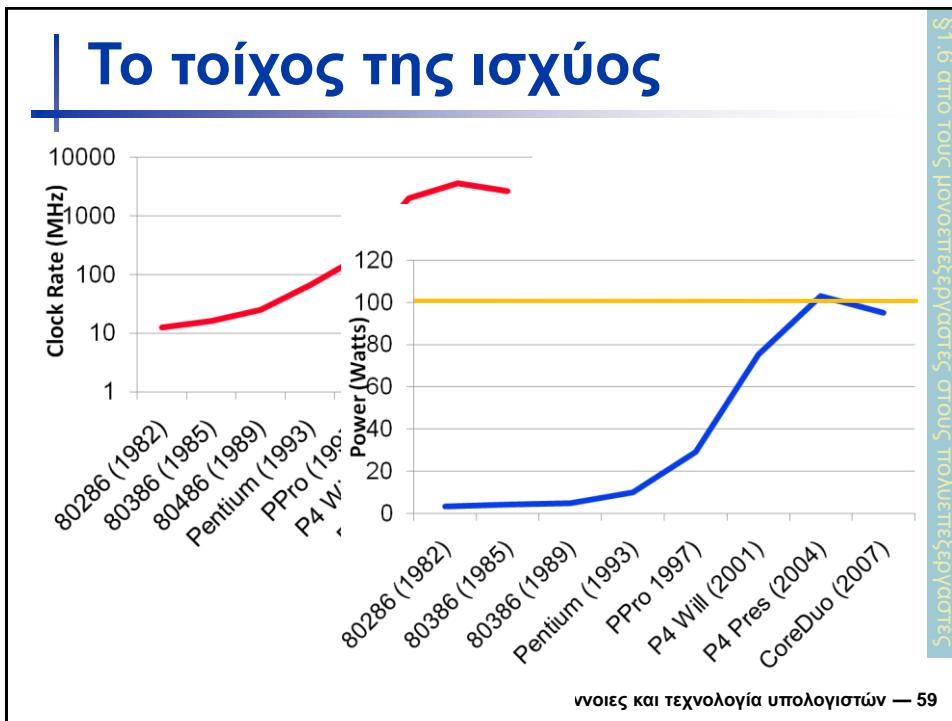
×1000

# Μείωση της κατανάλωσης

- Υποθέστε ότι μια νέα CPU έχει
  - 15% μείωση του χωρητικού φορτίου
  - 15% μείωση τάσης και 15% συχνότητας

$$\frac{P_{\text{new}}}{P_{\text{old}}} = \frac{C_{\text{old}} \times 0.85 \times (V_{\text{old}} \times 0.85)^2 \times F_{\text{old}} \times 0.85}{C_{\text{old}} \times V_{\text{old}}^2 \times F_{\text{old}}} = 0.85^4 = 0.52$$

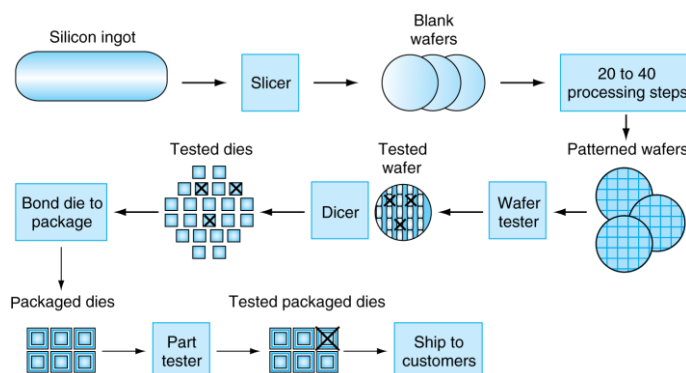
- Το τείχος της ισχύος
  - Δεν μπορούμε να μειώσουμε περαιτέρω την τάση
  - Δεν μπορούμε να απάγουμε τη θερμότητα
- Πώς αλλιώς μπορούμε να βελτιώσουμε την απόδοση



## Πολυεπεξεργαστές

- Επεξεργαστές πολλαπλών πυρήνων (multicore microprocessors)
  - Περισσότεροι επεξεργαστές σε ένα τσιπ
- Απαιτούν παράλληλο προγραμματισμό
  - Σε σύγκριση με την παραλληλία επιπέδου εντολής
    - Το υλικό εκτελεί πολλαπλές εντολές ταυτόχρονα
    - Χωρίς να το γνωρίζει ο προγραμματιστής
  - Δυσκολίες
    - Προγραμματισμός με στόχο την απόδοση
    - Εξισορρόπηση φορτίου
    - Βελτιστοποίηση επικοινωνίας και συγχρονισμού

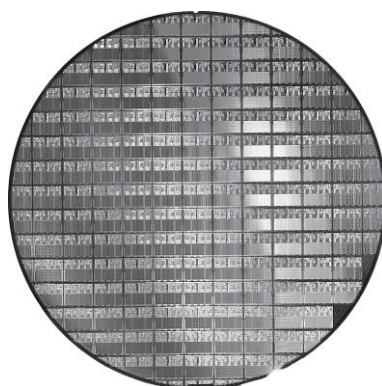
## Κατασκευή IC



- **Εσοδεία (yield):** ΠΟΣΟΣΤΟ ΤΩΝ ΤΣΙΠ ΑΝΑ ΠΛΑΚΙΔΙΟ ΠΟΥ ΔΟΥΛΕΥΟΥΝ ΣΩΣΤΑ

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 61

## Πλακίδιο AMD Opteron X2



- X2: πλακίδιο 300mm, 117 τσιπ, τεχνολογία 90nm
- X4: τεχνολογία 45nm technology

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 62

## Αξιολόγηση της Απόδοσης

- **Φορτίο εργασίας (workload)**
  - Σύνολο προγραμμάτων που εκτελούνται σε έναν υπολογιστή
    - είτε πραγματική συλλογή εφαρμογών είτε κατασκευασμένα προγράμματα
    - ένα φορτίο εργασίας καθορίζει τόσο τα προγράμματα όσο και τις σχετικές συχνότητες
- Για την αξιολόγηση της απόδοσης δύο υπολογιστικών συστημάτων, πρέπει να συγκρίνουμε το χρόνο εκτέλεσης του φορτίου εργασίας
  - Αυτό δεν είναι εύκολο να γίνει από τους χρήστες
- Συνήθως η αξιολόγηση του υπολογιστή γίνεται με τη χρήση ενός **συνόλου μετροπρογραμμάτων (benchmarks)**
  - Ειδικά επιλεγμένα προγράμματα για να μετρούν την απόδοση
  - Σχηματίζουν ένα φορτίο εργασίας ώστε να προσεγγίσουν το πραγματικό φορτίο εργασίας.

## Μετροπρογράμματα

- Οι τεχνικές βελτίωσης της απόδοσης βασίζονται στα χαρακτηριστικά των μετροπρογραμμάτων:
  - Η χρήση μετροπρογραμμάτων των οποίων η απόδοση εξαρτάται από πολύ μικρά τμήματα κώδικα ενθαρρύνει βελτιστοποιήσεις είτε στην αρχιτεκτονική είτε στο μεταγλωττιστή, που στοχεύουν σε αυτά τα τμήματα
    - Επομένως, οι εταιρείες μπορεί να εισάγουν αλλαγές που έχουν στόχο συγκεκριμένα μετροπρογράμματα
    - Στην αναζήτηση της δημιουργίας εξαιρετικά βελτιστοποιημένου κώδικα για μετροπρογράμματα, οι μηχανικοί μπορεί να εισάγουν λανθασμένες βελτιστοποιήσεις



## Μετροπρογράμματα

- Διαφορετικές κατηγορίες και εφαρμογές υπολογιστών απαιτούν διαφορετικούς τύπους μετροπρογραμμάτων:
  - Για επιτραπέζιους υπολογιστές:  
τα πιο συνηθισμένα μετροπρογράμματα επικεντρώνονται σε ένα συγκεκριμένο έργο, π.χ. αναπαραγωγή DVD, γραφικά για παιχνίδια
  - Για διακομιστές ανάλογα με τη φύση της εφαρμογής :
    - Για επιστημονικούς διακομιστές: τυπικά μετροπρογράμματα με επιστημονικές εφαρμογές προσανατολισμένα στη CPU
    - Για άλλα περιβάλλοντα διακομιστών: μετροπρογράμματα διακομιστών Ιστού, διακομιστών αρχείων, και βάσεων δεδομένων.
  - Για ενσωματωμένους υπολογιστές:  
συχνά οι πελάτες χρησιμοποιούν τη δική τους συγκεκριμένη ενσωματωμένη εφαρμογή

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 65

## Μετροπρογράμματα SPEC

- **SPEC (System Performance Evaluation Cooperative - Οργανισμός Αξιολόγησης Απόδοσης Συστημάτων)**
  - Παρέχει διαφορετικά σύνολα μετροπρογραμμάτων σχεδιασμένα
  - Χρηματοδοτείται από έναν αριθμό κατασκευαστών υπολογιστών
- Τα σύνολα μετροπρογραμμάτων SPEC περιλαμβάνουν :
  - προγράμματα για την απόδοση της CPU, γραφικά, υπολογιστική υψηλής απόδοσης, αντικειμενοστρεφή υπολογιστική, εφαρμογές Java, μοντέλα πελάτη-διακομιστή, συστήματα αλληλογραφίας, συστήματα αρχείων, διακομιστές Ιστού, κτλ.

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 66

## SPEC CPU Benchmark

- SPEC CPU2006
  - Παρελθών χρόνος για την εκτέλεση των προγραμμάτων
    - Αμελητέα I/O, εστιάζει στην απόδοση της CPU
  - Κανονικοποίηση του χρόνου εκτέλεσης ως προς μηχανή αναφοράς
    - Spec ratio = Ref time / Exec time
    - Μεγαλύτερο Spec ratio, καλύτερη απόδοση
  - Γεωμετρικός μέσος (geometric mean) των λόγων απόδοσης
    - CINT2006 (ακέραια) και CFP2006 (κινητής υποδιαστολής)

$$\sqrt[n]{\prod_{i=1}^n \text{Execution time ratio}_i}$$

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 67

## CINT2006 για τον Opteron X4 2356

Name	Description	IC×10 <sup>9</sup>	CPI	Tc (ns)	Exec time	Ref time	SPECratio
perl	String processing	2,118	0.75	0.40	637	9,777	15.3
bzip2	Block-sorting compression	2,389	0.85	0.40	817	9,650	11.8
gcc	GNU C Compiler	1,050	1.72	0.47	24	8,050	11.1
mcf	Combinatorial optimization	336	10.00	0.40	1,345	9,120	6.8
go	Go game (AI)	1,658	1.09	0.40	721	10,490	14.6
hmmer	Search gene sequence	2,783	0.80	0.40	890	9,330	10.5
sjeng	Chess game (AI)	2,176	0.96	0.48	37	12,100	14.5
libquantum	Quantum computer simulation	1,623	1.61	0.40	1,047	20,720	19.8
h264avc	Video compression	3,102	0.80	0.40	993	22,130	22.3
omnetpp	Discrete event simulation	587	2.94	0.40	690	6,250	9.1
astar	Games/path finding	1,082	1.79	0.40	773	7,020	9.1
xalancbmk	XML parsing	1,058	2.70	0.40	1,143	6,900	6.0
Geometric mean							11.7

Υψηλό CPI λόγω χαμηλής απόδοσης της cache

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 68

## Παγίδα: νόμος του Amdahl

- Λανθασμένη εντύπωση σχεδιαστών :
  - Η βελτίωση μιας πλευράς ενός υπολογιστή θα αυξήσει την απόδοση κατά ποσότητα ανάλογη του μεγέθους της βελτίωσης
- Παράδειγμα:
  - Θεωρήστε ότι ένα πρόγραμμα εκτελείται σε 100 sec σε έναν υπολογιστή, όπου οι πράξεις πολλαπλασιασμού είναι υπεύθυνες για τα 80 sec αυτού του χρόνου.
  - Πόσο πρέπει να βελτιώσω την ταχύτητα του πολλαπλασιασμού αν θέλω το πρόγραμμά να εκτελείται 5 φορές ταχύτερα;

## Ο νόμος του Amdahl

- Απλή εξίσωση του νόμου του Amdahl:

Χρόνος εκτέλεσης μετά τη βελτίωση =

$$= \left( \frac{\text{Χρόνος εκτέλεσης που επηρεάζεται από τη βελτίωση}}{\text{Ποσότητα βελτίωσης}} + \text{Χρόνος εκτέλεσης που δεν επηρεάστηκε} \right)$$

- **νόμος του Amdahl (Amdahl's law)**
  - η πιθανή αύξηση της απόδοσης με μια συγκεκριμένη βελτίωση περιορίζεται από το βαθμό χρήσης της βελτιωμένης λειτουργίας

## Παράδειγμα

$$\text{Χρόνος εκτέλεσης μετά τη βελτίωση} = \frac{80 \text{ δευτερόλεπτα}}{n} + (100 - 80 \text{ δευτερόλεπτα})$$

$$20 \text{ δευτερόλεπτα} = \frac{80 \text{ δευτερόλεπτα}}{n} + 20 \text{ δευτερόλεπτα}$$

$$0 = \frac{80 \text{ δευτερόλεπτα}}{n} \text{ ???}$$

- Ο νόμος του Amdahl χρησιμοποιείται για να εκτιμήσουμε βελτιώσεις της απόδοσης όταν γνωρίζουμε το χρόνο που καταναλώνεται σε κάποια λειτουργία
- Συνέπεια του νόμου του Amdahl στη σχεδίαση υλικού :
  - **Κάνε τη συνηθισμένη περίπτωση γρήγορη**

## Πλάνη: χαμηλή ισχύς σε αδράνεια

- X4 SPECpower benchmark
  - Σε 100% φορτίο: 295W
  - Σε 50% φορτίο: 246W (83%)
  - Σε 10% φορτίο: 180W (61%)
- Κέντρο δεδομένων της Google
  - Συνήθως λειτουργεί σε φορτίο 10%–50% load
  - Φορτίο 100% λιγότερο από το 1% του χρόνου
- Επανασχεδίαση του επεξεργαστή για να πετύχουμε κατανάλωση ανάλογη με το φορτίο

## Παγίδα: MIPS ως κριτήριο απόδοσης

- Μια εναλλακτική επιλογή μέτρησης του χρόνου είναι η παράμετρος **εκατομμύρια εντολές ανά δευτερόλεπτο (million instructions per second — MIPS)**
  - Μέτρηση της ταχύτητας εκτέλεσης του προγράμματος, βασισμένη στον αριθμό των εκατομμυρίων των εντολών

$$\text{MIPS} = \frac{\text{πλήθος εντολών}}{\text{χρόνος εκτέλεσης} \times 10^6}$$

## Γιατί είναι παγίδα

- Πλεονεκτήματα
  - Είναι εύκολα κατανοητά, και ταχύτεροι υπολογιστές σημαίνουν μεγαλύτερα MIPS, κάτι που ταιριάζει με τη διαίσθηση
- Μειονεκτήματα στη χρήση των MIPS:
  - Δεν λαμβάνουν υπόψη τους τις δυνατότητες των εντολών
    - Δεν μπορούμε να συγκρίνουμε υπολογιστές με διαφορετικά σύνολα εντολών
  - Διαφέρουν μεταξύ προγραμμάτων στον ίδιο υπολογιστή
    - ένας υπολογιστής δεν μπορεί να έχει μόνο μία κατάταξη ως προς MIPS για όλα τα προγράμματα
  - Μπορούν να μεταβάλλονται αντίστροφα με την απόδοση!
    - Ακολουθεί παράδειγμα

## Παράδειγμα MIPS

- Υπολογιστής με τρεις κατηγορίες εντολών A, B, Γ και μετρήσεις CPI 1, 2, 3, αντίστοιχα
- Κώδικας για το ίδιο πρόγραμμα από δύο διαφορετικούς μεταγλωττιστές:

Κώδικας από	Πλήθος εντολών (σε δις) για κάθε κατηγορία εντολής		
	A	B	Γ
Μεταγλωττιστής 1	5	1	1
Μεταγλωττιστής 2	10	1	1

- Υποθέστε ότι ο ρυθμός ρολογιού είναι 4 GHz
- Ποια ακολουθία κώδικα θα εκτελεστεί γρηγορότερα σύμφωνα με τα MIPS; Σύμφωνα με το χρόνο εκτέλεσης;

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 75

## Παράδειγμα MIPS

- Πρώτα βρίσκουμε το χρόνο εκτέλεσης για τους δύο διαφορετικούς μεταγλωττιστές:

$$\text{Χρόνος εκτέλεσης} = \frac{\text{Κύκλοι ρολογιού CPU}}{\text{Ρυθμός ρολογιού}}$$

$$\text{Κύκλοι ρολογιού CPU} = \sum_{i=1}^n (\text{CPI}_i \times C_i)$$

- Άρα:
  - Κύκλοι ρολογιού CPU<sub>1</sub> = (5×1 + 1×2 + 1×3) × 10<sup>9</sup> = 10 × 10<sup>9</sup>
  - Κύκλοι ρολογιού CPU<sub>2</sub> = (10×1 + 1×2 + 1×3) × 10<sup>9</sup> = 15 × 10<sup>9</sup>

$$\text{Χρόνος εκτέλεσης} = \frac{10 \times 10^9}{4 \times 10^9} = 2,5 \text{ δευτερόλεπτα}$$

- Και:

$$\text{Χρόνος εκτέλεσης} = \frac{15 \times 10^9}{4 \times 10^9} = 3,75 \text{ δευτερόλεπτα}$$

Αφηρημένες έννοιες και τεχνολογία υπολογιστών — 76

## Παράδειγμα MIPS

- Συνεπώς ο κώδικας 1 είναι ταχύτερος με βάση το χρόνο εκτέλεσης
- Οι μετρήσεις MIPS τι δίνουν;

$$\text{MIPS}_1 = \frac{(5 + 1 + 1) \times 10^9}{2,5 \times 10^6} = 2800$$

$$\text{MIPS}_2 = \frac{(10 + 1 + 1) \times 10^9}{3,75 \times 10^6} = 3200$$

- **δηλαδή το αντίθετο αποτέλεσμα!**