Υπερβαθμωτή (superscalar) Οργάνωση Υπολογιστών

Περιορισμοί των βαθμωτών αρχιτεκτονικών

1. Μέγιστο throughput: 1 εντολή/κύκλο ρολογιού (IPC≤1)
2. Υποχρεωτική ροή όλων των (διαφορετικών) τύπων εντολών μέσα από κοινή σωλήνωση
3. Εισαγωγή καθυστερήσεων σε ολόκληρη την ακολουθία εκτέλεσης λόγω stalls μίας εντολής (οι απόλυτα βαθμωτές αρχιτεκτονικές πραγματοποιούν εν σειρά (in-order) εκτέλεση των εντολών)

Πώς μπορούν να ξεπεραστούν οι περιορισμοί;

1. Εκτέλεση πολλαπλών εντολών ανά κύκλο μηχανής (παράλληλη εκτέλεση)

   → υπερβαθμωτές αρχιτεκτονικές

2. Ενσωμάτωση διαφορετικών αγωγών ροής δεδομένων, ο καθένας με όμοιες (πολλαπλή εμφάνιση του ίδιου τύπου) ή και ετερογενείς λειτουργικές μονάδες

   → multicycle operations

3. Δυνατότητα εκτέλεσης εκτός σειράς (out-of-order) των εντολών

   → δυναμικές αρχιτεκτονικές

Παραλληλισμός Επιπέδου Εντολών
ILP: Instruction-Level Parallelism

- Ο ILP είναι ένα μέτρο του βαθμού των πραγματικών εξαρτήσεων δεδομένων που υφίστανται ανάμεσα στις εντολές

- Average ILP = #instructions / #cycles required

  code1: ILP = 1

  οι εντολές πρέπει να εκτελεστούν σειριακά

  code2: ILP = 3

  οι εντολές μπορούν να εκτελεστούν παράλληλα

```
code1:   r1 ← r2 + 1
         r3 ← r1 / 17
         r4 ← r0 - r3
```

```
code2:   r1 ← r2 + 1
         r3 ← r9 / 17
         r4 ← r0 - r10
```

ILP parameters (Jouppi and Wall, 1989)

- Operation Latency (OL)

Number of machine cycles until a result is available for use by a subsequent instruction

- Machine Parallelism (MP)

Max number of simultaneously executing instructions the machine can support

- Issue Latency (IL)

Number of machine cycles required between issuing two consecutive instructions

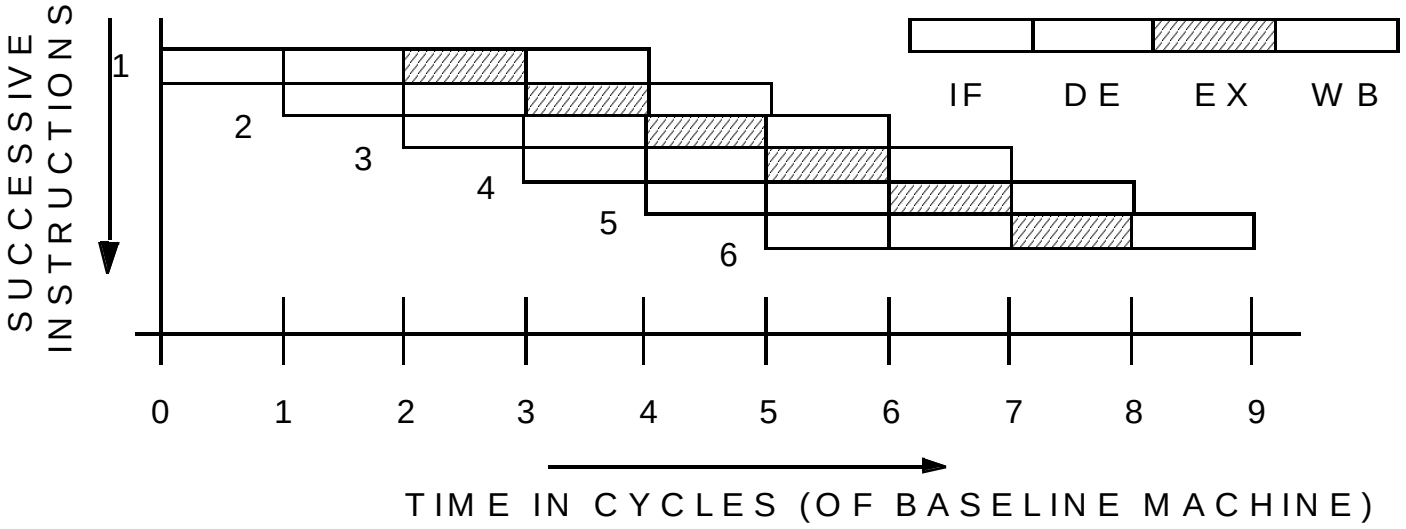
- Issue Parallelism (IP)

Max number of instructions that can be issued in every cycle

Κατηγοριοποίηση επεξεργαστών με βάση τον ILP

[Jouppi, DECWRL 1991]

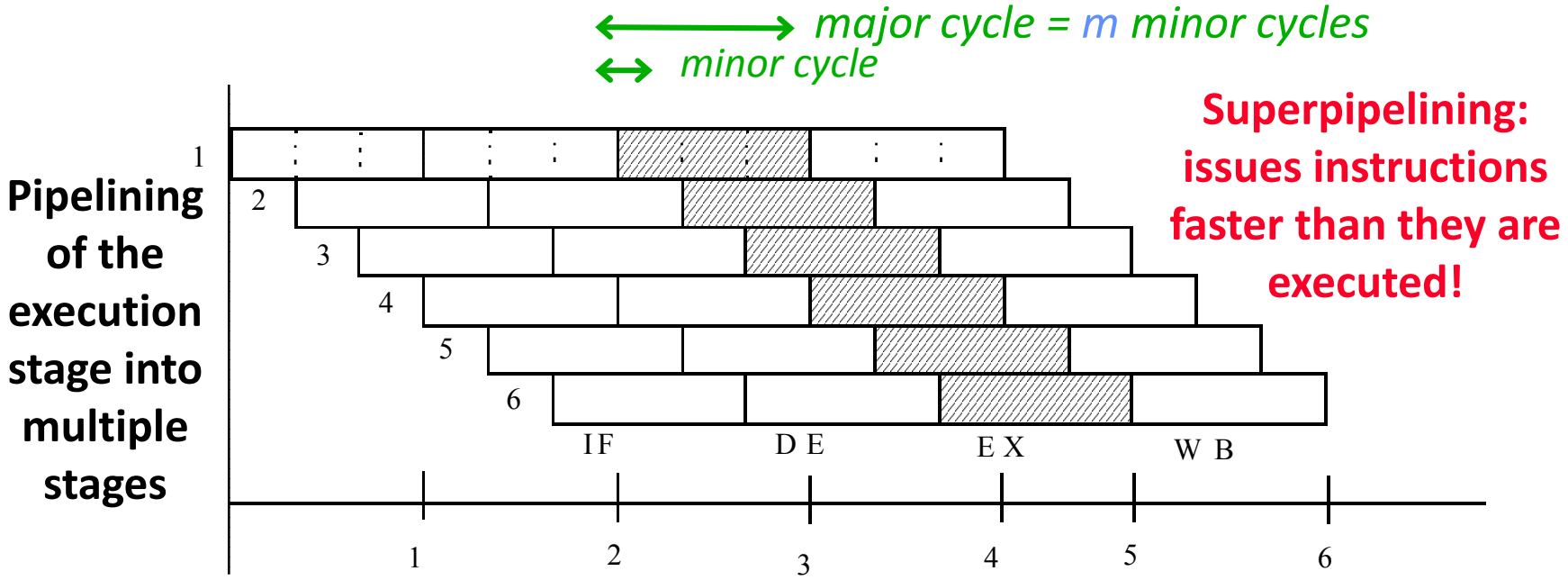
- Baseline Scalar pipeline (π.χ. κλασικό 5-stage MIPS)
  - Παραλληλισμός διανομής IP (Issue Parallelism) = 1 εντολή/κύκλο
  - IL (Issue Latency) = 1 cycle
  - MP (Machine Parallelism) = k (k stages in the pipeline)
  - OL (operation latency) = 1 cycle
  - Μέγιστο IPC = 1 εντολή/κύκλο



Κατηγοριοποίηση επεξεργαστών με βάση τον ILP

[Jouppi, DECWRL 1991]

- Superpipelined: κύκλος ρολογιού = 1/m του baseline
  - Issue Parallelism IP = 1 εντολή / minor κύκλο
  - Operation Latency OL = 1 major cycle = m minor κύκλοι
  - Issue Latency IL = 1 minor cycle
  - MP = m x k
  - Μέγιστο IPC = m εντολές / major κύκλο (m x speedup?)

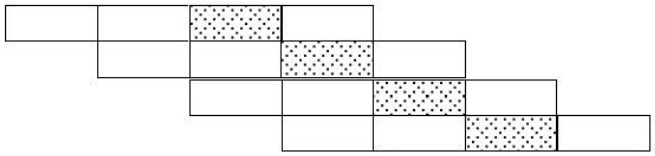


Superpipelining

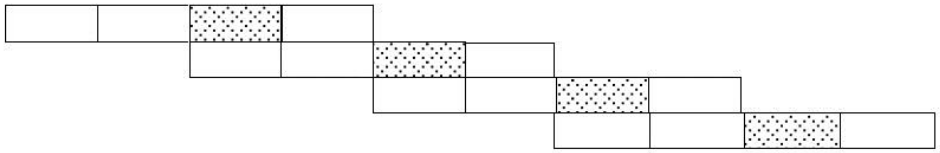
"Superpipelining is a new and special term meaning pipelining. The prefix is attached to increase the probability of funding for research proposals. There is no theoretical basis distinguishing superpipelining from pipelining. Etymology of the term is probably similar to the derivation of the now-common terms, methodology and functionality as pompous substitutes for method and function. The novelty of the term superpipelining lies in its reliance on a prefix rather than a suffix for the pompous extension of the root word."

- Nick Tredennick, 1991

Superpipelining: Hype vs. Reality

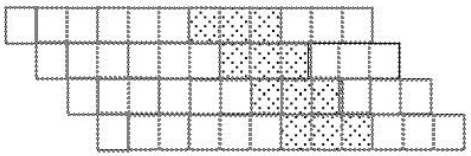


baseline



underpipelined

η ταχύτητα διανομής των εντολών δεν ακολουθεί το ρυθμό επεξεργασίας τους



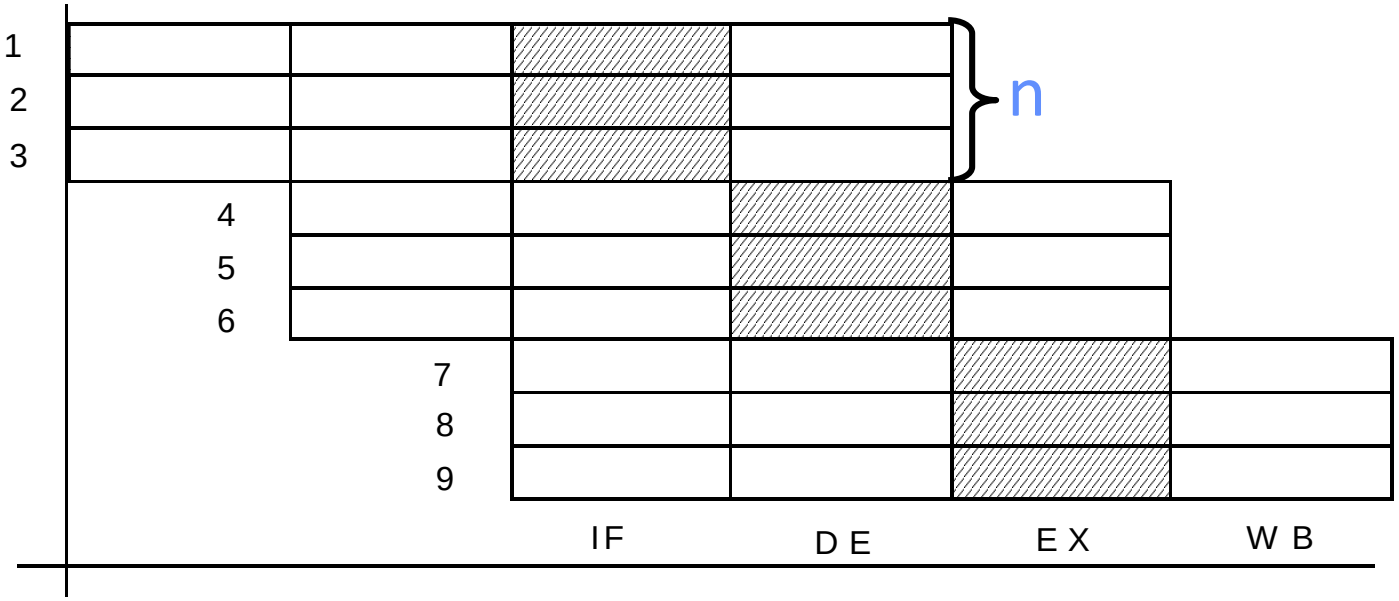
superpipelined

τα αποτελέσματα μιας εντολής δεν είναι διαθέσιμα στις επόμενες m-1 διαδοχικές εντολές

Κατηγοριοποίηση επεξεργαστών με βάση τον ILP

[Jouppi, DECWRL 1991]

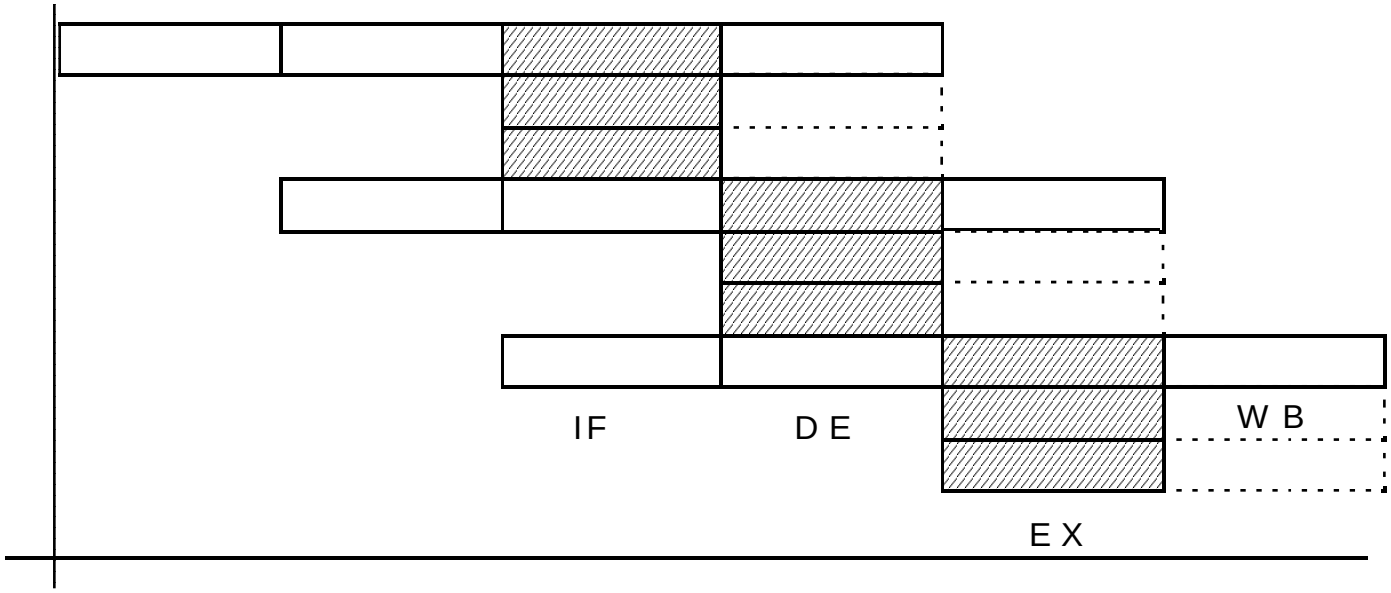
- Superscalar:
  - Παραλληλισμός διανομής = IP = n εντολές / κύκλο
  - Καθυστέρηση λειτουργίας = OP = 1 κύκλος
  - Μέγιστο IPC = n εντολές / κύκλο (n x speedup?)



Κατηγοριοποίηση επεξεργαστών με βάση τον ILP

[Jouppi, DECWRL 1991]

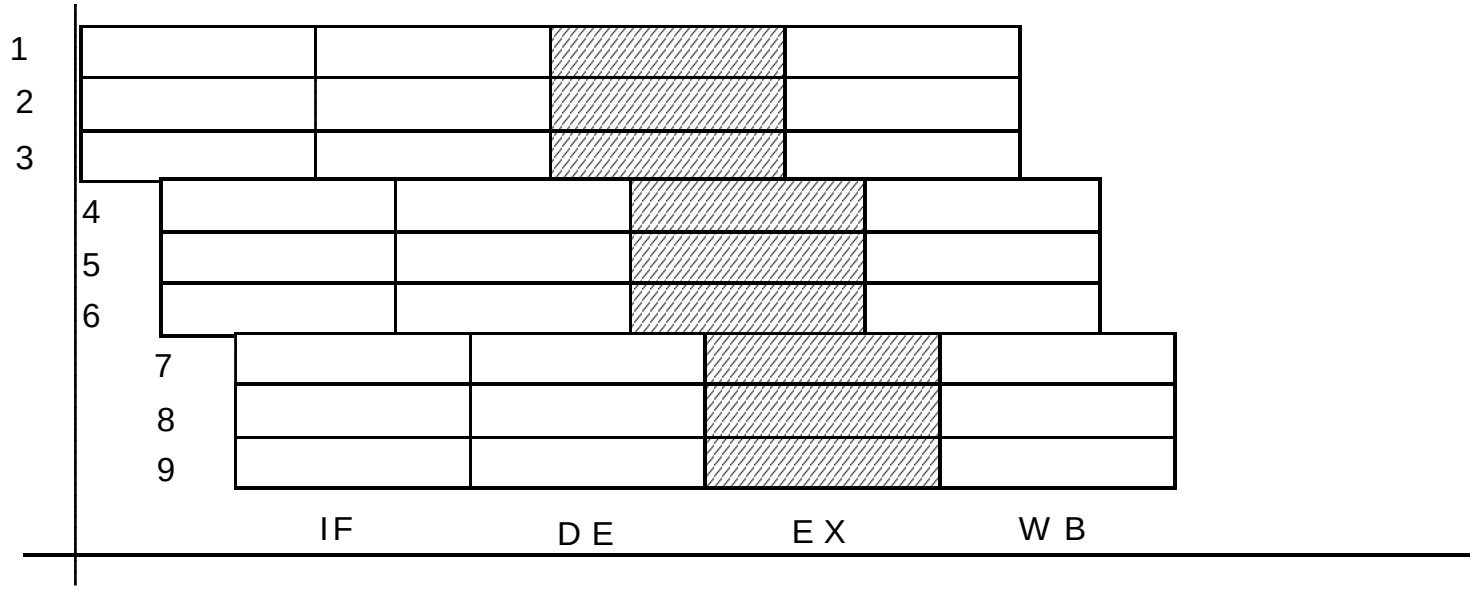
- VLIW: Very Long Instruction Word
  - Παραλληλισμός διανομής = IP = n εντολές / κύκλο
  - Καθυστέρηση λειτουργίας = OP = 1 κύκλος
  - Μέγιστο IPC = n εντολές / κύκλος = 1 VLIW / κύκλο



Κατηγοριοποίηση επεξεργαστών με βάση τον ILP

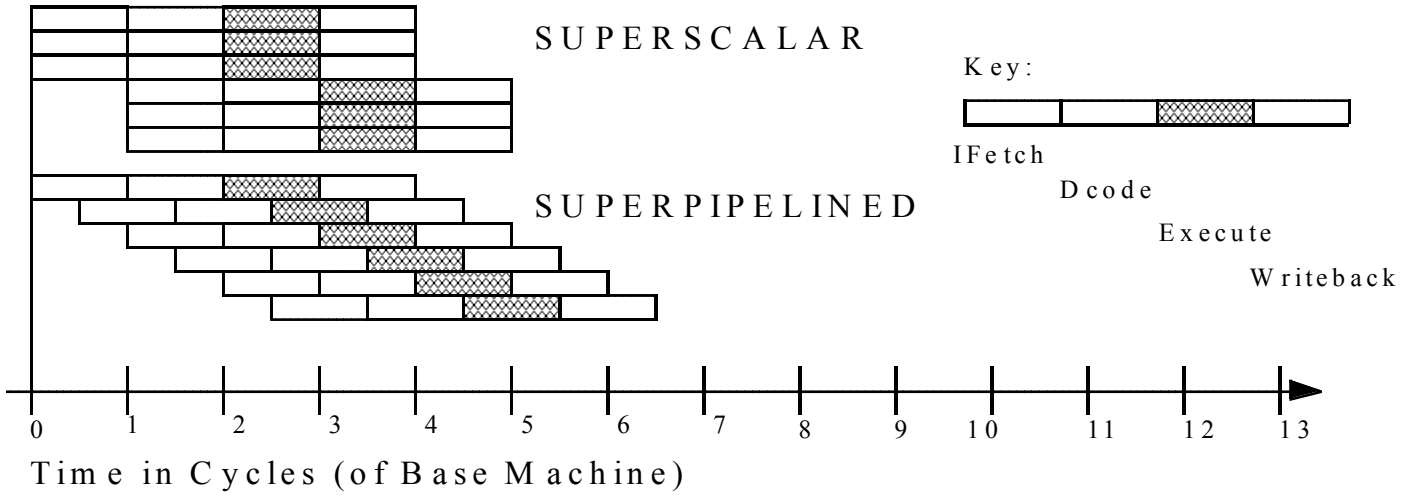
[Jouppi, DECWRL 1991]

- Superpipelined-Superscalar
  - Παραλληλισμός διανομής = IP = n εντολές / minor κύκλο
  - Καθυστέρηση λειτουργίας = OP = m minor κύκλοι
  - Μέγιστο IPC = n x m εντολές / major κύκλο

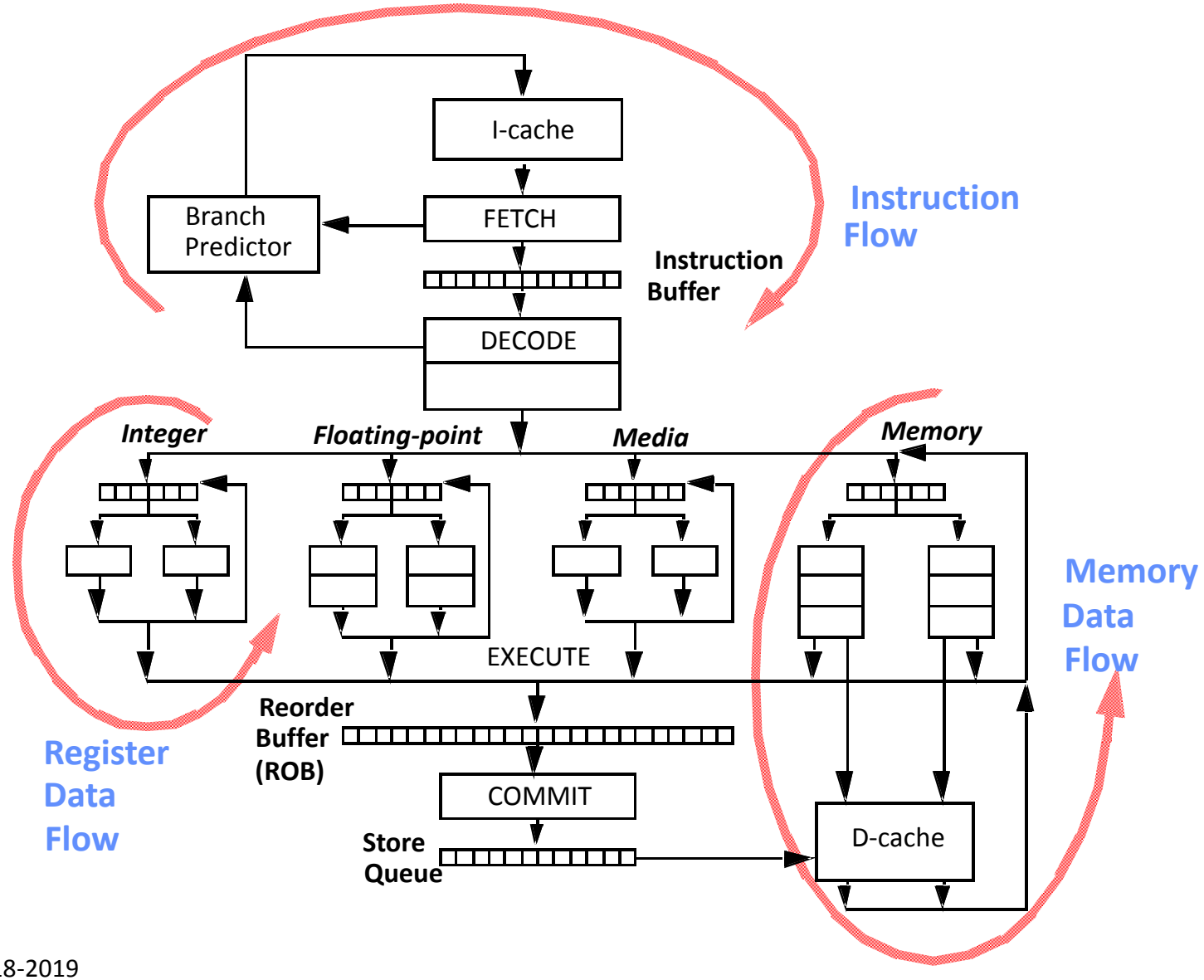


Superscalar vs. Superpipelined

- Περίπου ισοδύναμη επίδοση
  - Αν n = m τότε και οι δύο έχουν περίπου το ίδιο IPC
  - Παραλληλισμός στο «χώρο» vs. παραλληλισμός στον χρόνο



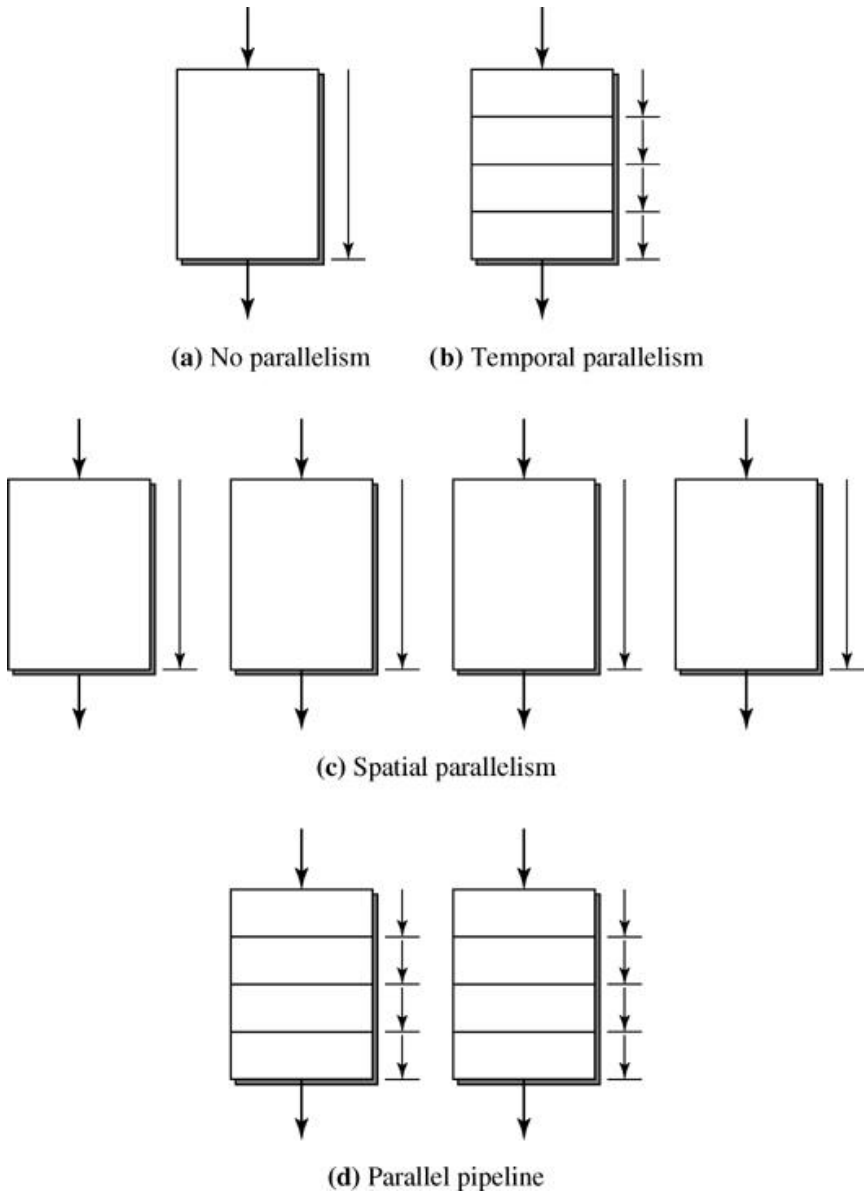
Μοντέλο ροών στους Superscalars



Παράλληλες αρχιτεκτονικές αγωγού

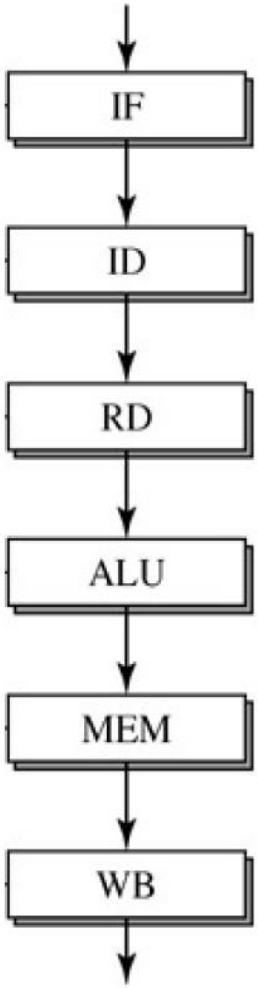
- Βαθμός παραλληλισμού μηχανήματος: ο μέγιστος αριθμός εντολών που μπορούν ταυτόχρονα να είναι σε εξέλιξη

- Σε ένα μια βαθμωτή αρχιτεκτονική ισούται με τον αριθμό σταδίων του pipeline (pipeline depth)



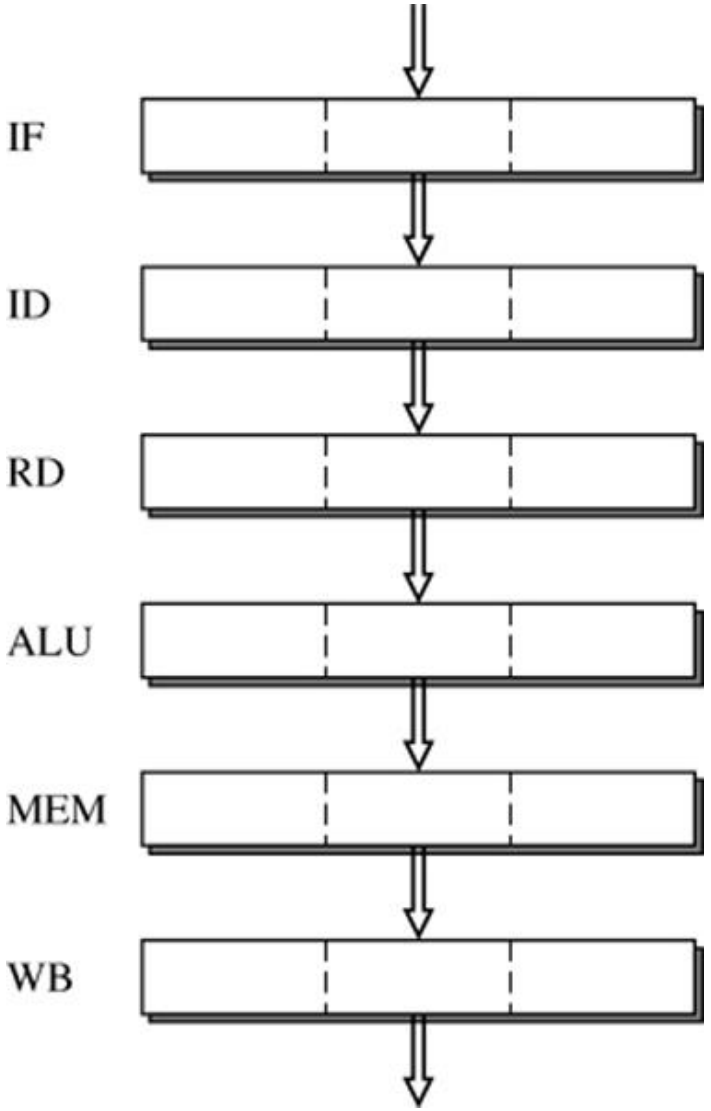
Παράδειγμα βαθμωτής αρχιτεκτονικής αγωγού 6 σταδίων

- IF: instruction fetch
- ID: instruction decode
- RD: register read
- ALU: ALU op/address generation
- MEM: read/write memory
- WB: register write

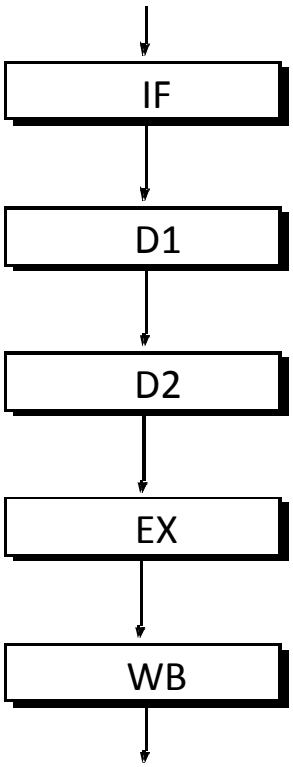


Η ίδια σωλήνωση με πλάτος 3

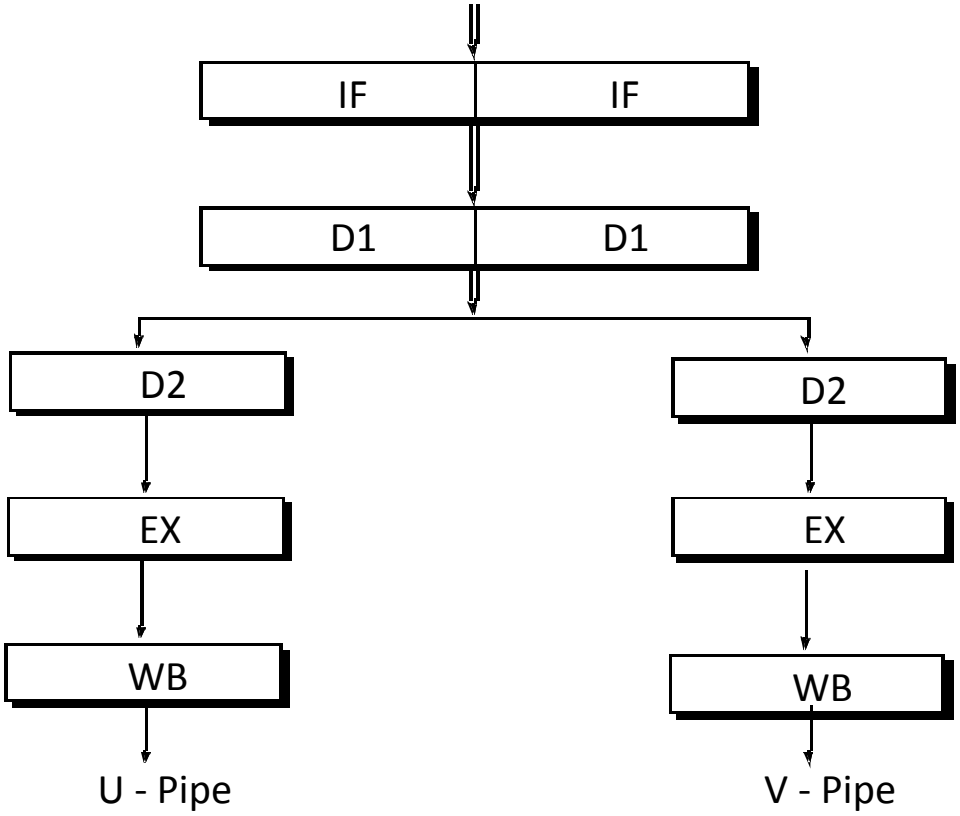
- Πολλαπλά δομικά στοιχεία (functional units) στο hardware
- Αυξάνεται η λογική πολυπλοκότητα των σταδίων του pipeline
- Απαιτούνται πολλαπλές θύρες ανάγνωσης/εγγραφής του register file για την ταυτόχρονη προσπέλαση από όλους τους αγωγούς
- Επιτυγχάνεται στην καλύτερη περίπτωση επιτάχυνση ίση με 3 σε σύγκριση με την αντίστοιχη βαθμωτή σωλήνωση



Inorder Pipelines



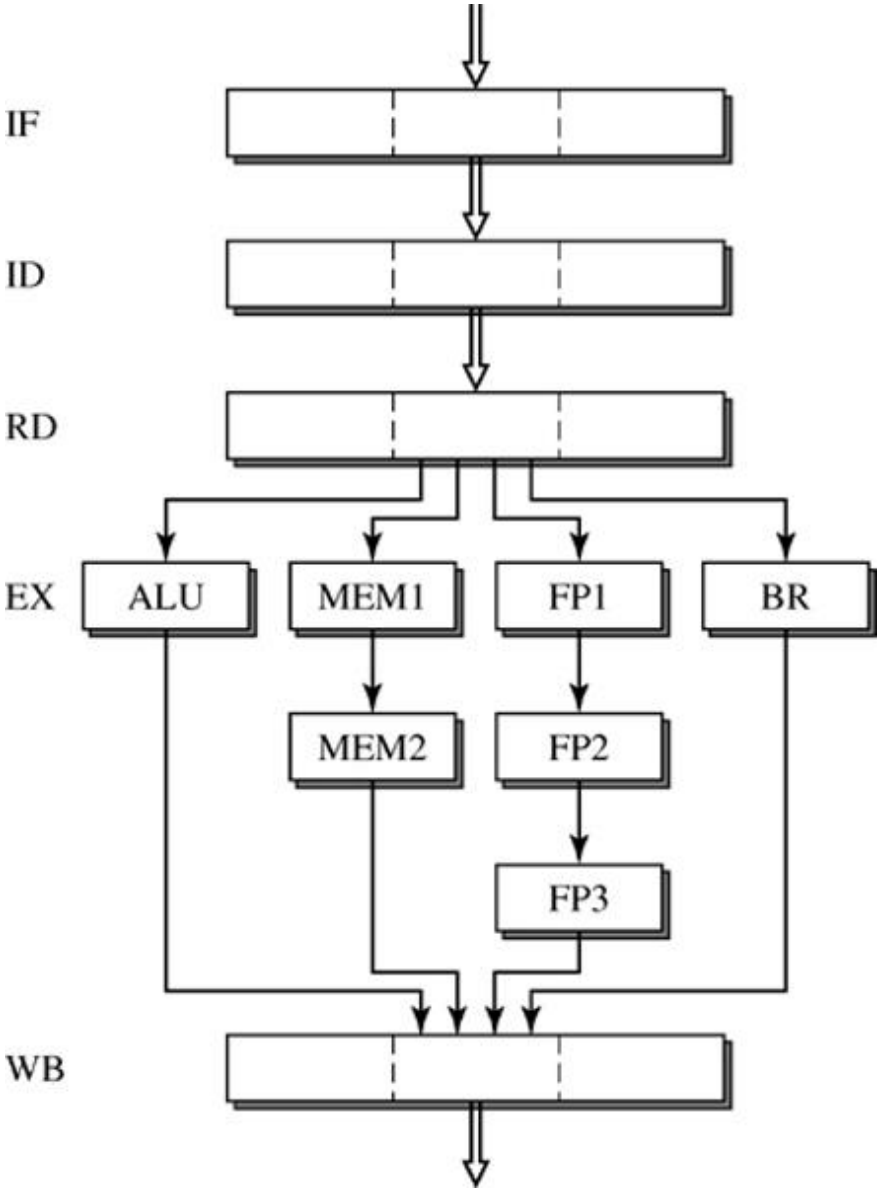
Intel i486



Intel Pentium (2 i486 pipelines)

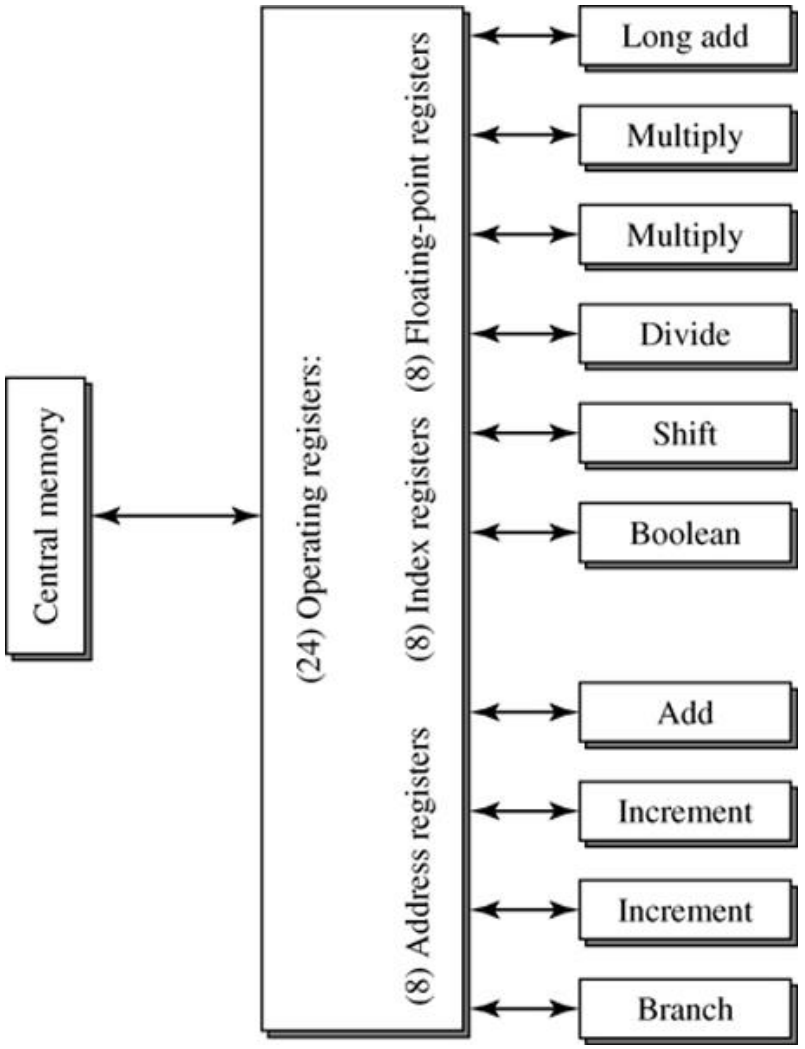
Ετερογενείς υπερβαθμωτές σωληνώσεις

- επέκταση των βαθμωτών σωληνώσεων πολλαπλών κύκλων (π.χ. FP MIPS pipeline) στα στάδια IF, ID, RD,WB
- μετά το RD, κάθε εντολή γίνεται issue μέσα στον αγωγό που αντιστοιχεί στον τύπο της
- Πλεονεκτήματα:
  1. Customization → Efficiency
  2. Αναγκαία καθυστέρηση για κάθε τύπο εντολής
- Προκλήσεις:
  1. Πλήθος των functional units
  2. Τύπος των functional units



CDC 6600-1964

- ο πρώτος υπολογιστής που κατασκευάστηκε στα πρότυπα ενός «υπερ-υπολογιστή»
- 1964 (πριν από τους επεξεργαστές RISC): περιείχε 10 διαφορετικές λειτουργικές μονάδες έξω από τη σωλήνωση, με διαφορετικό latency η κάθε μία
- στόχος η διεκπεραίωση 1 εντολής ανά κύκλο μηχανής
- 8 address registers (18 bits)
- 8 index registers (18 bits)
- 8 fp registers (60 bits)

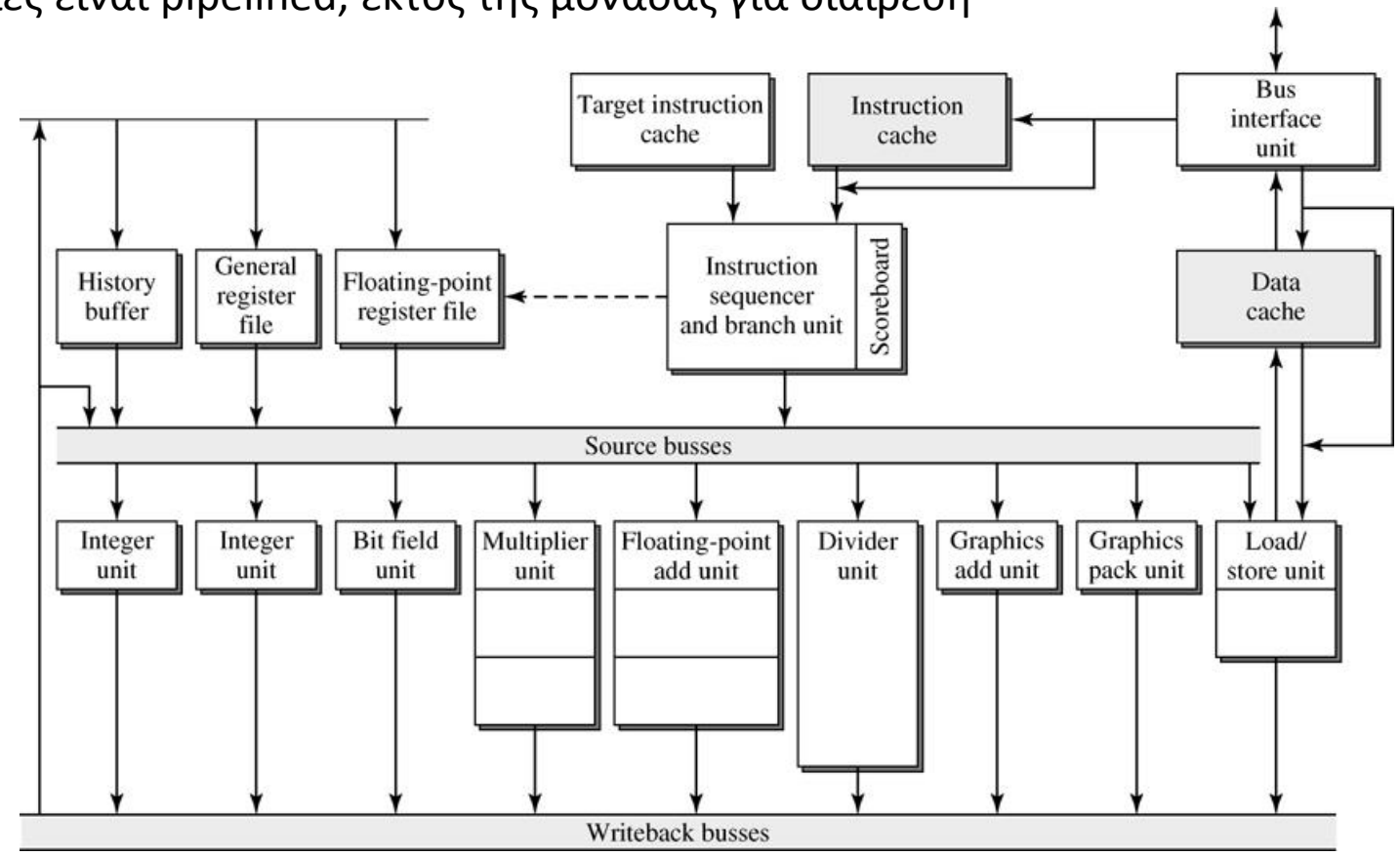


CDC 6600

- 10 functional units/non pipelined/variable exec latency
- 1x Fixed-point adder (18 bits)-3 cycles
- 1 x Floating-point adder (60 bits)
- 2 x Multiply unit (60 bits)-10 cycles
- Divide unit (60 bits)-29 cycles
- Shift unit (60 bits)
- Logical unit (60 bits)
- 2 x increment units
- Branch unit

Motorola 88110-1992

- ένας από τους πλατύτερους αγωγούς (most wider pipelines)
- περιέχει 10 λειτουργικές μονάδες, στην πλειοψηφία τους με latency ενός κύκλου
- όλες είναι pipelined, εκτός της μονάδας για διαίρεση



Source: Diefendorf and Allen (1992)

Ετερογενής σωλήνωση του Power4 (2001)