Δίκτυα Ταξινόμησης

(CLR κεφάλαιο 28)

Στην ενότητα αυτή θα μελετηθούν τα εξής θέματα:

Δίκτυα σύγκρισης, δίκτυα ταξινόμησης

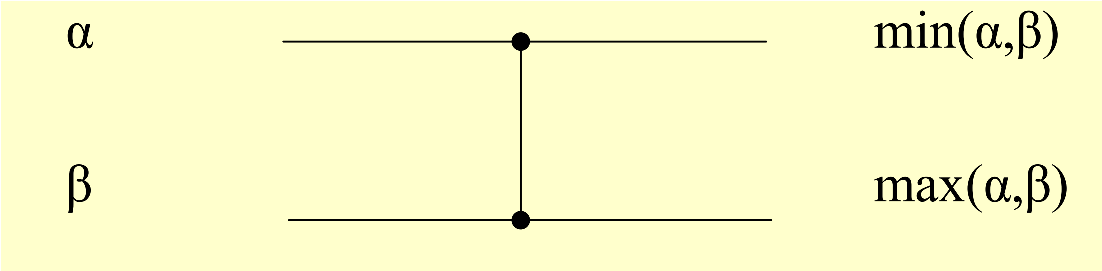
Αρχή 0-1

Διτονική ταξινόμηση

Δίκτυο Ταξινόμησης

- Μοντέλο στο οποίο
  1. η μόνη δυνατή πράξη είναι η σύγκριση.
  2. πολλές συγκρίσεις μπορούν να γίνουν ταυτόχρονα.

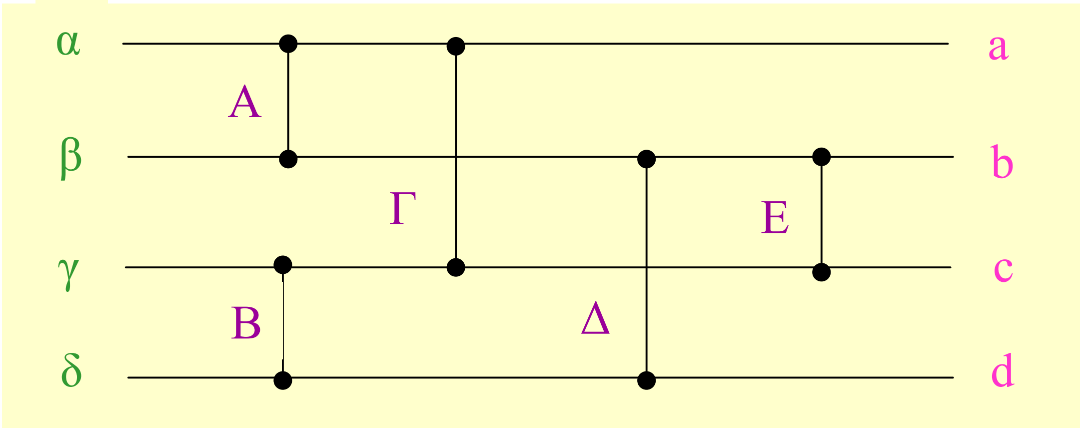
- ***Συγκριτής:*** δέχεται δύο δεδομένα εισόδου α, β και επιστρέφει στην "πάνω" του έξοδο το μικρότερο και στην "κάτω" το μεγαλύτερο



- ***Δίκτυο σύγκρισης***: ένα δίκτυο από συγκριτές.

- ***Δίκτυο ταξινόμησης***: ένα δίκτυο σύγκρισης το οποίο ταξινομεί το δεδομένο εισόδου του.

Δίκτυα Ταξινόμησης



- Το πιο πάνω δίκτυο σύγκρισης αποτελείται από 5 συγκριτές, τους Α, Β, Γ, Δ, Ε. Παρατηρούμε ότι το δεδομένο εξόδου του είναι το δεδομένο εισόδου ταξινομημένο.
- Συγκεκριμένα, το ελάχιστο των α, β, βγαίνει στην πάνω έξοδο του Α και το ελάχιστο των γ, δ, βγαίνει στην πάνω έξοδο του Β, οι δύο αυτό τιμές συγκρίνονται από τον Γ, ο οποίος επιστρέφει στην έξοδο a, την ελάχιστη (εκ των α,β,γ,δ) τιμή.
- Παρόμοια το συνολικό μέγιστο βγαίνει στην έξοδο d, ενώ τα ενδιάμεσα στοιχεία βγαίνουν ταξινομημένα στις γραμμές εξόδου b και c.
- Παρατήρηση: οι συγκριτές Α και Β μπορούν να παράξουν τα αποτελέσματα τους ταυτόχρονα (παράλληλα). Το ίδιο ισχύει για τους συγκριτές Γ και Δ.

Ορισμοί

Ορίζουμε:

- *Βάθος* ενός δικτύου το μήκος ενός μακρύτερου μονοπατιού από μια γραμμή εισόδου σε μια γραμμή εξόδου.

  (Στο παράδειγμα, το βάθος είναι 3)

- *Μέγεθος* ενός δικτύου: ο αριθμός συγκριτών που περιέχει.

  (Στο παράδειγμα το μέγεθος είναι 5)

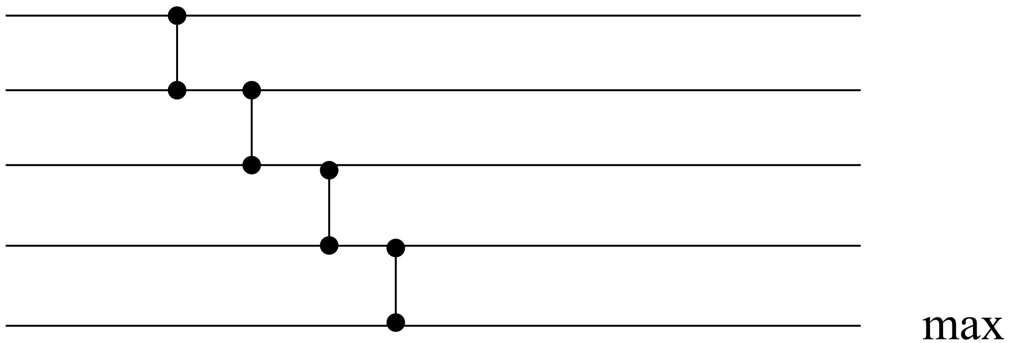
- *Χρόνος εκτέλεσης* ενός δικτύου σύγκρισης είναι ο χρόνος που χρειάζεται για να λάβουν όλες οι γραμμές εξόδου τις τιμές τους εφόσον όλες οι γραμμές εισόδου έχουν λάβει τα δεδομένα εισόδου.

- *Ο χρόνος εκτέλεσης ενός δικτύου σύγκρισης δίνεται από το βάθος του.*

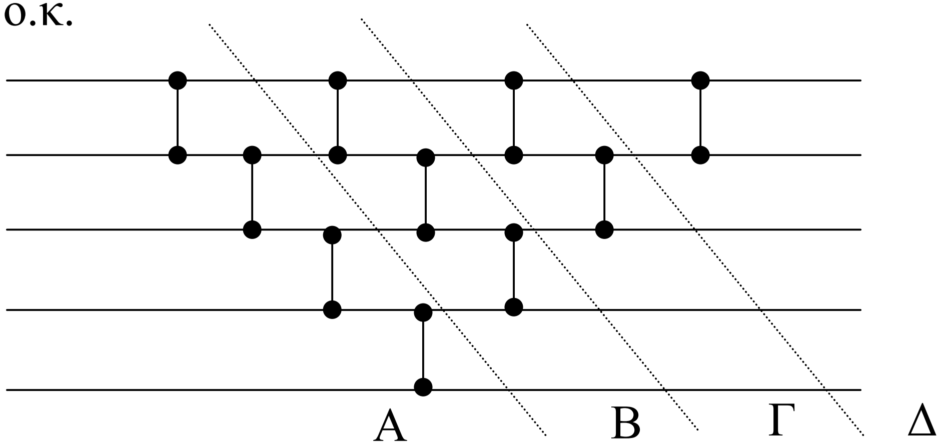
Παράδειγμα

Δίκτυο ταξινόμησης βασισμένο στον αλγόριθμο Selection Sort.

- Βασικό συγκρότημα: το δίκτυο που βρίσκει το μέγιστο:



- Το δίκτυο συμπληρώνεται αναδρομικά: βρες το μέγιστο από τα υπόλοιπα στοιχεία, κ.ο.κ.



Παράδειγμα

- Ποιο είναι το βάθος του δικτύου για ταξινόμηση n στοιχείων;

- Φαίνεται να είναι $\Theta(n^2)$, υπάρχουν

  $n\text{-}1 + n\text{-}2 + \ldots + 1 \in \Theta(n^2)$ συγκριτές.

- Όμως, λόγω του ότι συγκρίσεις μπορούν να γίνουν παράλληλα (pipelining) το βάθος είναι 2n-3 (για τη φάση Α χρειάζεται χρόνος n-1, αλλά κάθε μια από τις επόμενες n-2 φάσεις απαιτεί μια ακόμα μονάδα χρόνου εφόσον, όλες εκτός από την τελευταία σύγκριση, μπορούν να γίνονται ταυτόχρονα με συγκρίσεις της "προηγούμενης" φάσης.

- Συνεπώς το πιο πάνω δίκτυο ταξινόμησης είναι ταχύτερο από κάθε ακολουθιακό αλγόριθμο που βασίζεται σε συγκρίσεις ($\Omega(n \lg n)$)

Κάτω φράγμα παράλληλης ταξινόμησης

- *Κάθε δίκτυο ταξινόμησης αντιστοιχεί σε ένα παράλληλο αλγόριθμο που πραγματοποιεί ταυτόχρονα πολλές συγκρίσεις.*

- *Πόσο γρήγορα μπορούμε να ταξινομούμε παράλληλα;* (Πόσο είναι το ελάχιστο δυνατό βάθος ενός δικτύου ταξινόμησης;)

- Όλες οι συγκρίσεις που γίνονται από συγκριτές σε ένα δίκτυο ταξινόμησης μπορούν να προσομοιωθούν σειριακά. Άρα πρέπει να υπάρχουν $\Omega(n \lg n)$ συγκριτές.

- Σε κάθε "στρώμα" του δικτύου μπορούν να γίνουν το πολύ n/2 συγκρίσεις. *Έτσι πρέπει να υπάρχουν $\Omega(\lg n)$ στρώματα.*

Αρχή 0-1

- Πόσο γρήγορα μπορούμε να ελέγξουμε αλγοριθμικά αν ένα δίκτυο σύγκρισης είναι δίκτυο ταξινόμησης;
- Να ελέγξουμε για κάθε μια από τις n! μεταθέσεις n αριθμών αν το δίκτυο επιστρέφει στις εξόδους του την ακολουθία ταξινομημένη.
- Χρονική πολυπλοκότητα: Ω(n!)
- Μπορεί να γίνει ο έλεγχος πιο γρήγορα;

ΘΕΩΡΗΜΑ (Αρχή 0-1)

Αν ένα δίκτυο σύγκρισης με n εισόδους ταξινομεί κάθε μια από τις ακολουθίες από 0 και 1, τότε είναι δίκτυο ταξινόμησης (δηλαδή, ταξινομεί οποιαδήποτε ακολουθία εισόδου).

Αρχή 0-1

ΑΠΟΔΕΙΞΗ

- Έστω ότι η f είναι μια μονοτονικά αύξουσα συνάρτηση, δηλαδή, αν $x \le y$, τότε $f(x) \le f(y)$.

- Τότε

$$\min(f(x),f(y)) = f(\min(x,y)) \text{ και } \max(f(x),f(y)) = f(\max(x,y))$$

ΛΗΜΜΑ Αν ένα δίκτυο απεικονίζει την ακολουθία εισόδου $a_1, a_2, \ldots, a_n$ στην ακολουθία εξόδου $b_1, b_2, \ldots, b_n$, τότε απεικονίζει την ακολουθία εισόδου $f(a_1), f(a_2), \ldots, f(a_n)$ στην ακολουθία εξόδου $f(b_1), f(b_2), \ldots, f(b_n)$.

ΑΠΟΔΕΙΞΗ Με επαγωγή στο βάθος του δικτύου.

- Υποθέτουμε για να φτάσουμε σε αντίφαση, ότι για κάποιο δίκτυο σύγκρισης με n εισόδους που ταξινομεί κάθε μια από τις ακολουθίες από 0 και 1, υπάρχει ακολουθία εισόδου $a_1, a_2, \ldots, a_n$ τέτοια ώστε $a_i < a_j$, αλλά το a_i είναι πιο κάτω από το a_j στην ακολουθία εξόδου.

Αρχή 0-1

- Ορίζουμε τη συνάρτηση

$$f(x) = \begin{cases} 0, & x \le a_i \\ 1, & x > a_i \end{cases}$$

- Η f είναι μονοτονικά αύξουσα συνάρτηση.

- Από το Λήμμα, το δίκτυο απεικονίζει την ακολουθία εισόδου $f(a_1)$, $f(a_2),\ldots, f(a_n)$ στην ακολουθία εξόδου με το $f(a_i)$ πιο κάτω από το $f(a_j)$. Από την υπόθεση, η ακολουθία εξόδου είναι ταξινομημένη, δηλαδή $f(a_j) \le f(a_i)$. Από τον ορισμό της f, $f(a_j)=1$, $f(a_i)=0$. Αντίφαση.

Κατασκευή ενός δικτύου ταξινόμησης

<u>Βήμα 1 της κατασκευής</u>

- Μια ακολουθία ονομάζεται διτονική, αν
  1. είτε αρχικά είναι μονοτονικά αύξουσα και στη συνέχεια είναι μονοτονικά φθίνουσα,
  2. είτε αρχικά είναι μονοτονικά φθίνουσα και στη συνέχεια είναι μονοτονικά αύξουσα.

  π.χ. $\langle 1,2,3,5,4,2 \rangle$, $\langle 9,5,3,1,2,4,7 \rangle$, $\langle 1,2,3 \rangle$.

- Διτονικές ακολουθίες από 0 και 1, έχουν μια από τις μορφές 0…01…10…0 ή 1…10…01…1 .

- Γραφικά

0	**1**	**0**

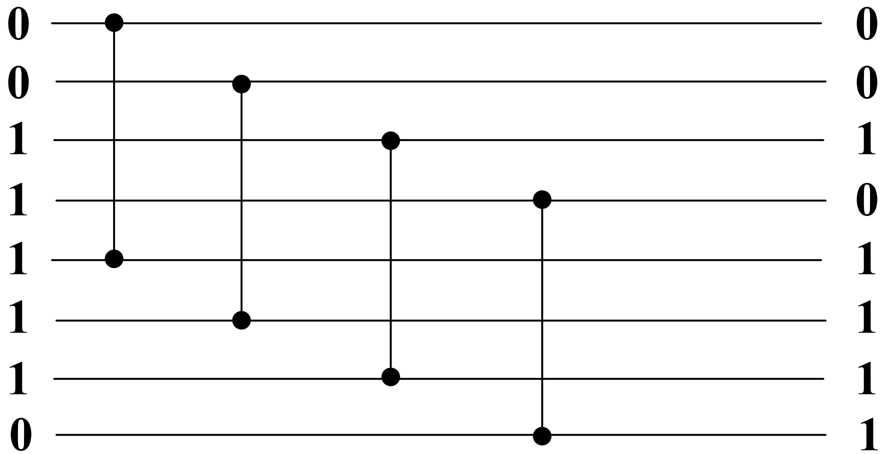
ή

1	**0**	**1**

- Θα κατασκευάσουμε ένα δίκτυο διτονικής ταξινόμησης, δηλαδή ένα δίκτυο το οποίο ταξινομεί κάθε διτονική ακολουθία από 0 και 1.

Δίκτυο του ημι-καθαριστή

- Βασικό δομικό στοιχείο ενός δικτύου διτονικής ταξινόμησης είναι το δίκτυο του *ημι-καθαριστή*.



ΛΗΜΜΑ

- Έστω ότι η είσοδος ενός ημικαθαριστή είναι διτονική δυαδική ακολουθία. Τότε η έξοδος έχει τις εξής ιδιότητες:
    1. και η πάνω και η κάτω ημιακολουθίες είναι διτονικές
    2. κάθε στοιχείο στην πάνω ημιακολουθία είναι μικρότερο ή ίσο από κάθε στοιχείο στην κάτω ημιακολουθία
    3. είτε η πάνω είτε η κάτω ημιακολουθία είναι "καθαρή" δηλαδή αποτελείται από μόνο 0 ή μόνο 1.

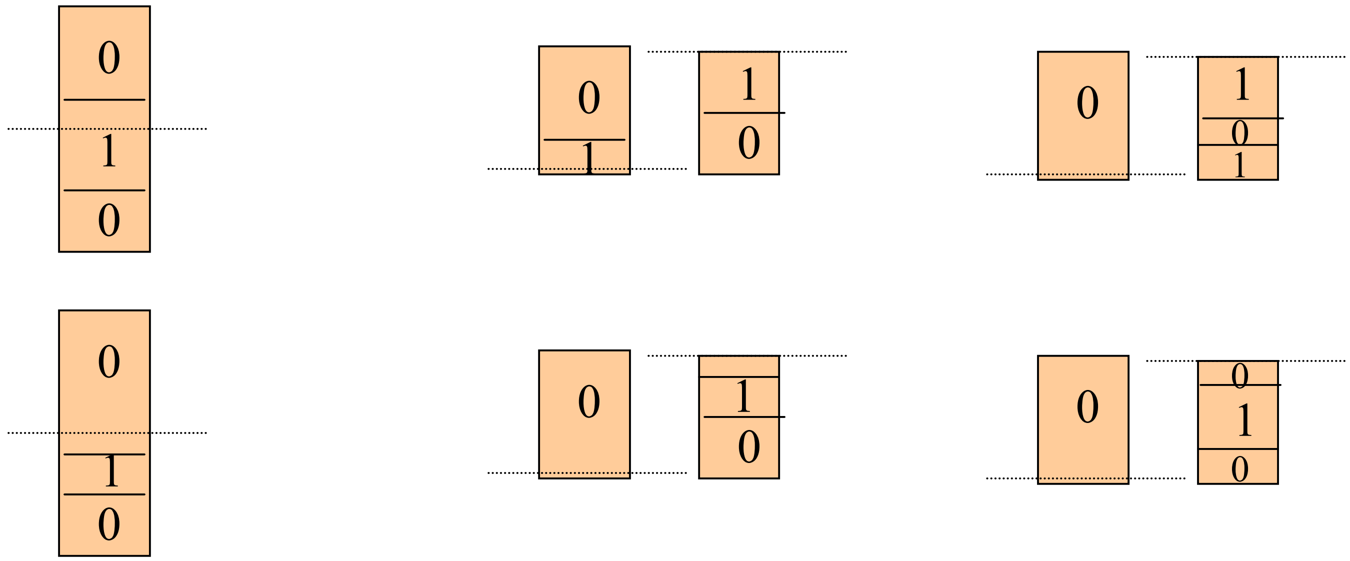
Απόδειξη

- Εξετάζουμε την περίπτωση που η είσοδος έχει τη μορφή

0	**1**	**0**

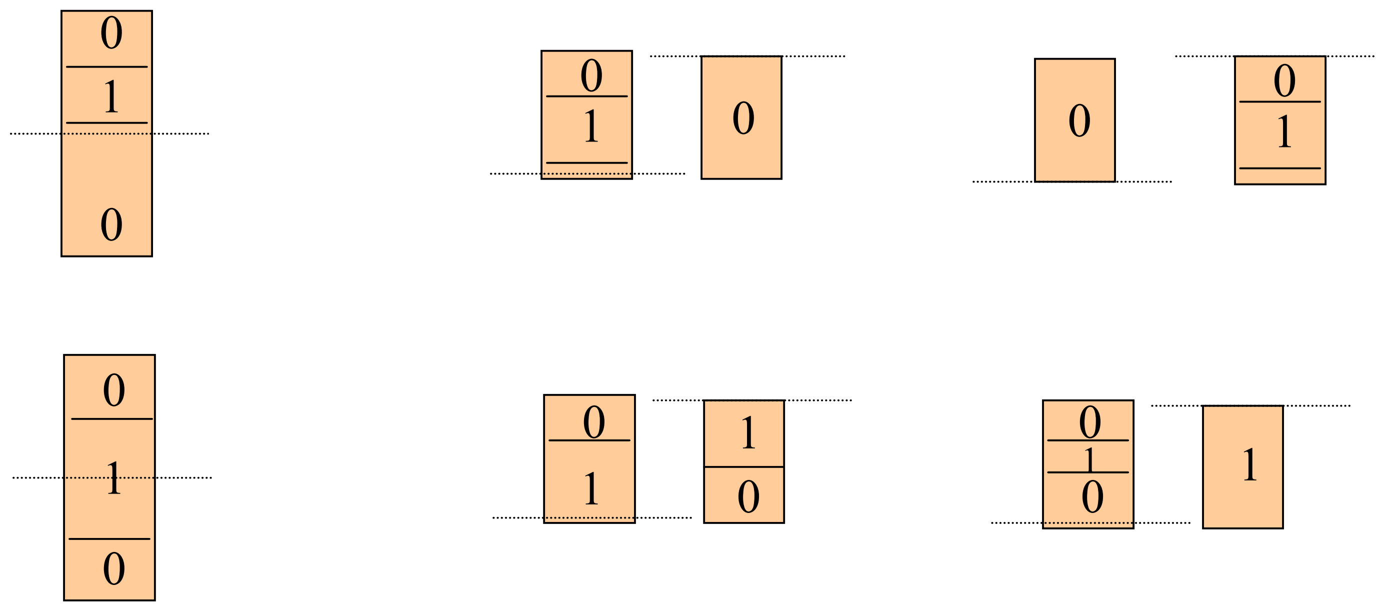
- Υπάρχουν οι εξής περιπτώσεις

ακολουθία εισόδου | πάνω και κάτω υποακολουθίες | μετά τη σύγκριση



Απόδειξη

ακολουθία εισόδου | πάνω και κάτω υποακολουθίες | μετά τη σύγκριση

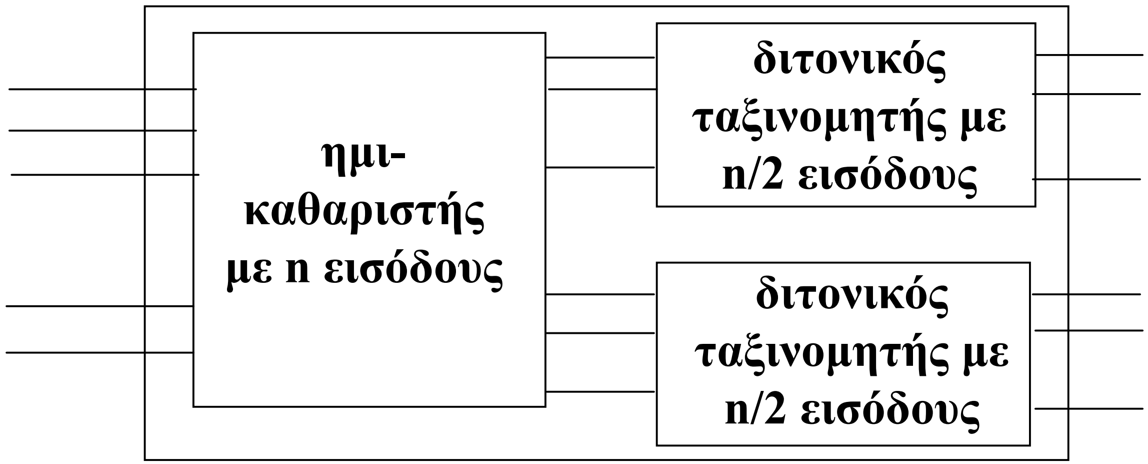


Όμοια εξετάζεται και η περίπτωση που η είσοδος έχει τη μορφή

1	**0**	**1**

Διτονική ταξινόμηση

- Άρα ένα *δίκτυο διτονικής ταξινόμησης* μπορεί να κατασκευαστεί αναδρομικά με τη χρήση
  1. ενός ημικαθαριστή (ο οποίος θα μοιράσει το δεδομένο εισόδου σε δύο υποακολουθίες για τις οποίες ισχύει ότι κάθε στοιχείο στην πάνω ημιακολουθία είναι το μικρότερο ή ίσο με κάθε στοιχείο στην κάτω ημιακολουθία) και στη συνέχεια
  2. δύο δικτύων διτονικής ταξινόμησης μισού μεγέθους για ταξινόμηση των δύο ημιακολουθιών.



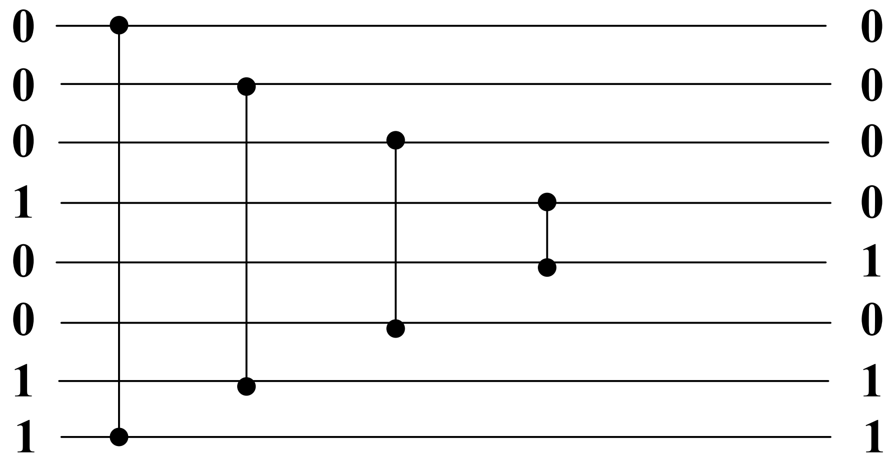
Βάθος

$$D(n) = D(n/2) + 1$$
$$= D(n/4) + 1 + 1$$
$$= \ldots = \lg n$$

Δεύτερη φάση της κατασκευής

- Θα κατασκευάσουμε ένα *δίκτυο συγχώνευσης*, δηλαδή ένα δίκτυο το οποίο με δεδομένο δύο ταξινομημένες ημιακολουθίες τις συγχωνεύει επιστρέφοντας την ακολουθία ταξινομημένη.

- Βασικό δομικό στοιχείο ενός δικτύου διτονικής ταξινόμησης είναι το δίκτυο του *τροποποιημένου ημι-καθαριστή*.



Τροποποιημένος ημικαθαριστής

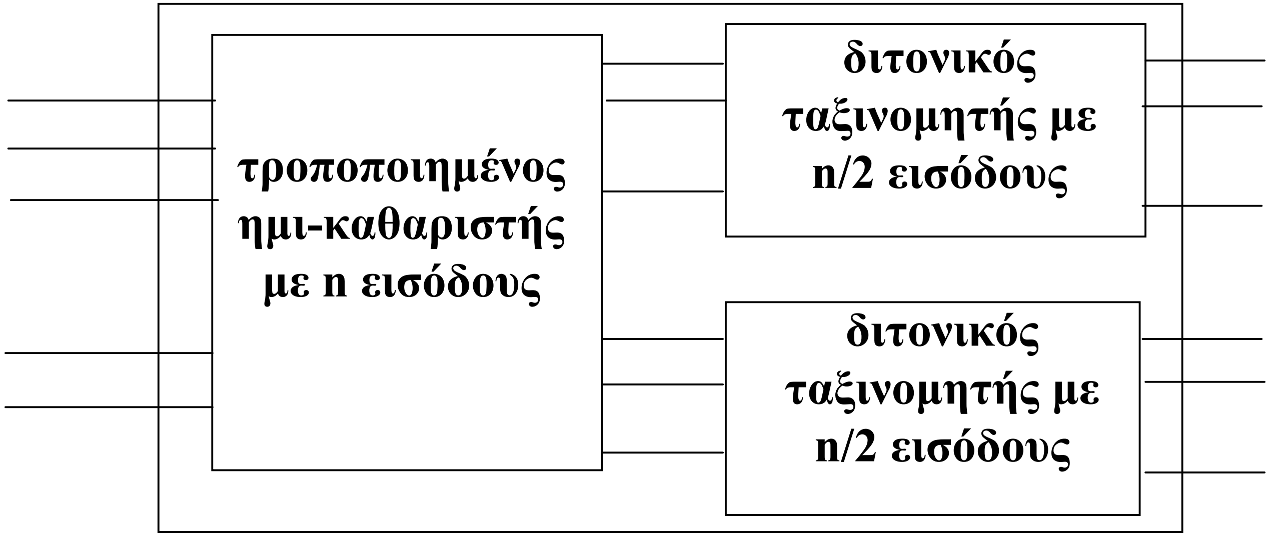
ΛΗΜΜΑ

Με δεδομένο εισόδου δύο ταξινομημένες ημι-ακολουθίες, η έξοδος του τροποποιημένου ημι-καθαριστή είναι δύο διτονικές ακολουθίες όπου

1. κάθε στοιχείο στην πάνω ημιακολουθία είναι το πολύ ίσο με κάθε στοιχείο στην κάτω ημιακολουθία και
2. είτε η πάνω είτε η κάτω ημιακολουθία είναι "καθαρή" δηλαδή αποτελείται από μόνο 0 ή μόνο 1.

Δίκτυο συγχώνευσης

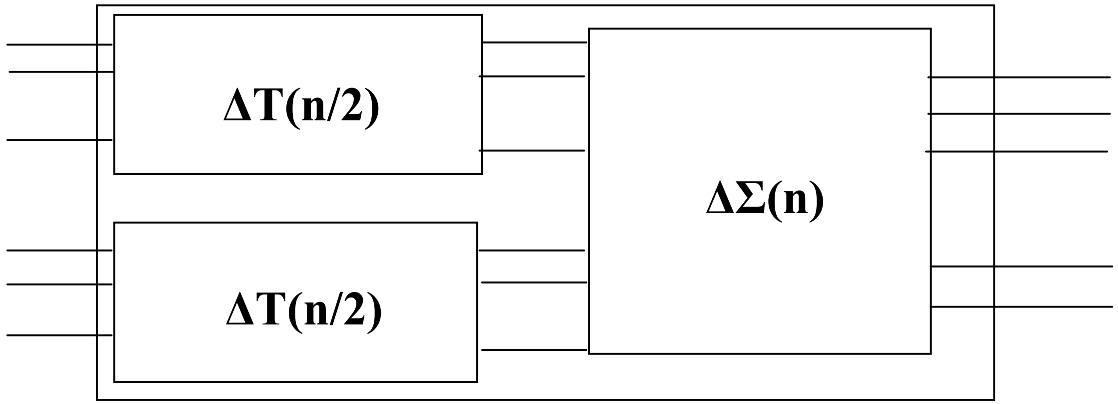
- Άρα ένα δίκτυο συγχώνευσης μπορεί να κατασκευαστεί ως
  1. ένα τροποποιημένος ημικαθαριστής,
  2. ακολουθούμενος από δύο διτονικούς ταξινομητές με n/2 εισόδους



- Βάθος;

Τρίτη φάση της κατασκευής

- Πιο κάτω δίνεται δίκτυο ταξινόμησης για n στοιχεία ΔΤ(n): Αρχικά ταξινομούμε τις πάνω και κάτω ημιακολουθίες με δύο δίκτυα ταξινόμησης με n/2 εισόδους (ΔΤ(n/2)), και στη συνέχεια συγχωνεύουμε τα αποτελέσματα με ένα συγχωνευτή με n εισόδους (ΔΣ(n)).



- Ποιο είναι το βάθος του ΔΤ(n);

$$D(n) = D(n/2) + \lg n$$
$$= D(n/4) + \lg(n/2) + \lg n$$
$$= \ldots \in \Theta(\lg^2 n)$$

ΔΤ(8)