Chapter 24

Sorting networks

NEW CS 473: Theory II, Fall 2015
November 19, 2015

24.1 Model of Computation

24.1.0.1 Model of Computation

(A) Q: Perform a computational task considerably faster by using a different architecture? Yep.
(B) ***Spaghetti sort***!

24.1.0.2 Spaghetti

Pastafarianism

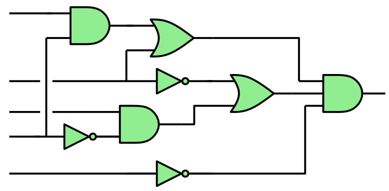
The spaghetti tree hoax was a three-minute hoax report broadcast on April Fools' Day 1957 by the BBC current-affairs programme Panorama, purportedly showing a family in southern Switzerland harvesting spaghetti from the family "spaghetti tree". At the time spaghetti was relatively little-known in the UK, so that many Britons were unaware that spaghetti is made from wheat flour and water; a number of viewers afterwards contacted the BBC for advice on growing their own spaghetti trees. Decades later CNN called this broadcast "the biggest hoax that any reputable news establishment ever pulled."

24.1.0.3 Spaghetti sort

(A) Input: $S = \{s_1, \ldots, s_n\} \subseteq [1, 2]$.
(B) Have much Spaghetti (this are longish and very narrow tubes of pasta).
(C) cut ith piece to be of length s_i, for $i = 1, \ldots, n$.
(D) take all these pieces of pasta in your hand..
(E) make them stand up vertically, with their bottom end lying on a horizontal surface
(F) lower your handle till it hit the first (i.e., tallest) piece of pasta.
(G) Take it out, measure it height, write down its number
(H) and continue in this fashion till done.
(I) Linear time sorting algorithm.
(J) ...but sorting takes $\Omega(n \log n)$ time.

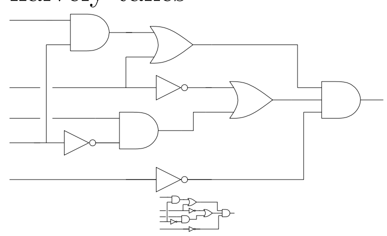
24.1.0.4 What is going on?

(A) Faster algorithm achieved by changing the computation model.
(B) allowed new “strange” operations
(cutting a piece of pasta into a certain length, picking the longest one in constant time, and measuring the length of a pasta piece in constant time)
(C) Using these operations we can sort in linear time.
(D) So, are there other useful computation models?

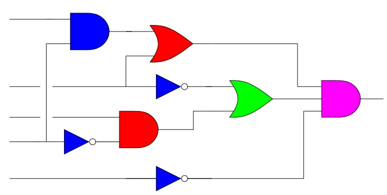


24.1.0.5 Circuits are fast...

(A) Computing the following circuit naively takes



8 units of time.
(B) Use parallelism!

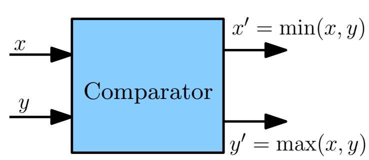


4 time units!
(C) Circuits are really parallel...
(D) Sorting numbers with circuits?
(E) Q: Can sort in *sublinear* time by allowing parallel comparisons?

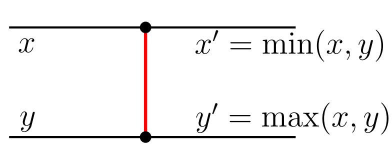
24.2 Sorting with a circuit – a naive solution

24.2.0.1 Sorting with a circuit – a naive solution

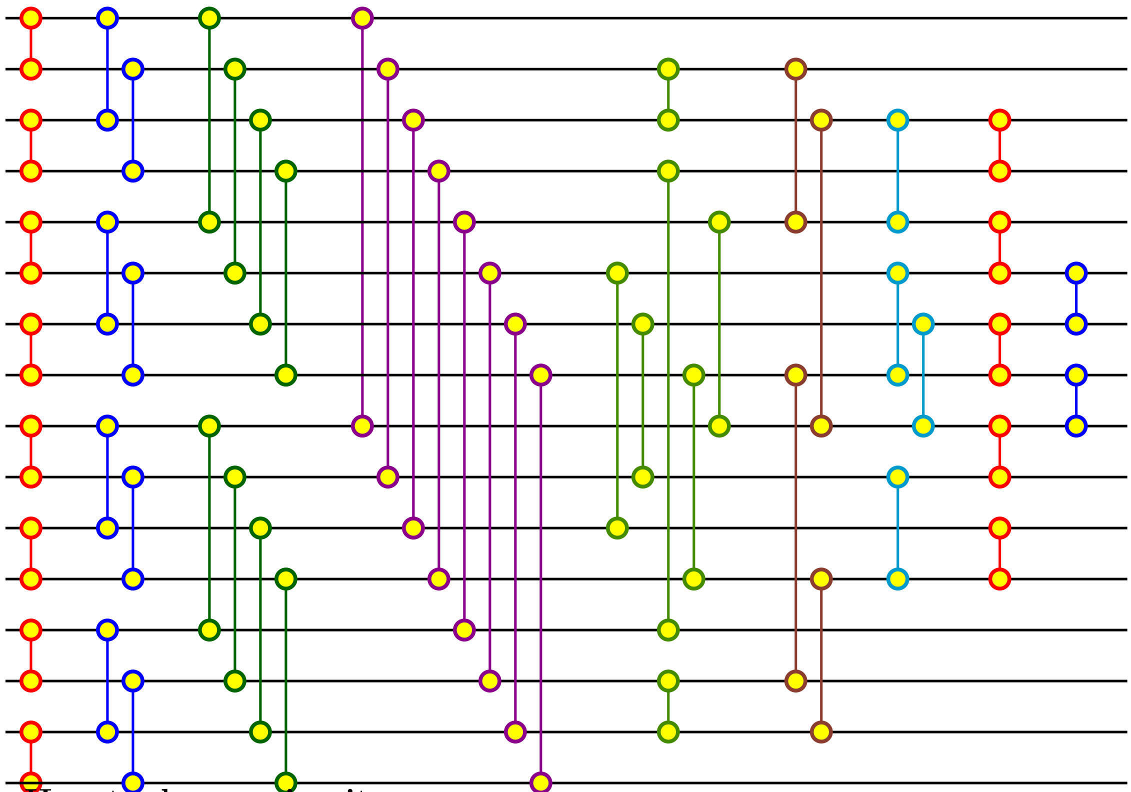
(A) ***comparator*** gate:



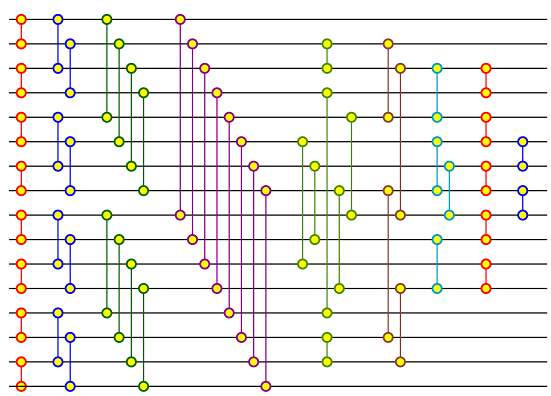
(B) Draw it as:



24.2.0.2 Sorting network - an example



24.2.0.3 How to draw a circuit...



(A) **wires**: horizontal lines
(B) **gates**: vertical segments (i.e., gates) connecting lines.
(C) Inputs arrive the wires from left.
(D) Output on the right side of wires.
(E) largest number is output on the bottom line.
(F) Sorting algorithms $\implies$ sorting circuits.

24.2.1 Definitions

24.2.1.1 Definitions

Definition 24.2.1. A ***comparison network*** is a DAG, with n inputs and n outputs, where each gate has two inputs and two outputs.

Definition 24.2.2. ***depth*** of a wire is 0 at input. For gate with two inputs of depth d_1 and d_2 the depth on the output wire is $1 + \max(d_1, d_2)$.

depth of comparison network is maximum depth of an output wire.

Definition 24.2.3. ***sorting network***: comparison network such that for any input, the output is monotonically sorted.

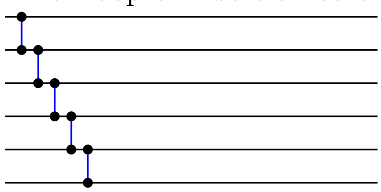
size: sorting network is number of gates.

running time of sorting network is its depth.

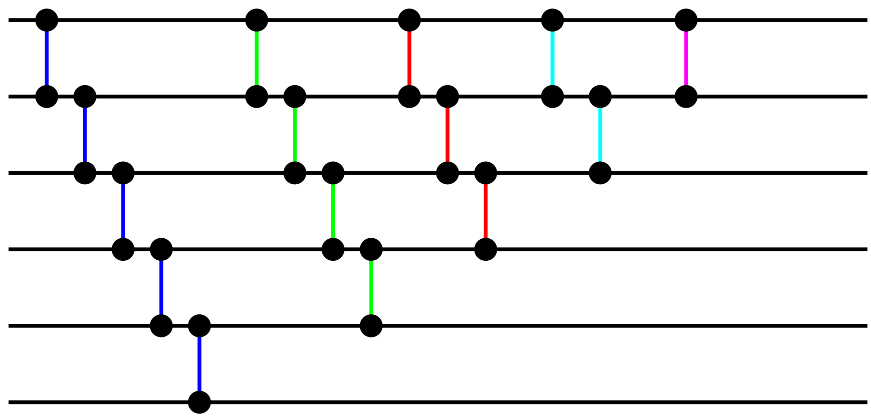
24.2.2 Sorting network based on insertion sort

24.2.2.1 Sorting network based on insertion sort

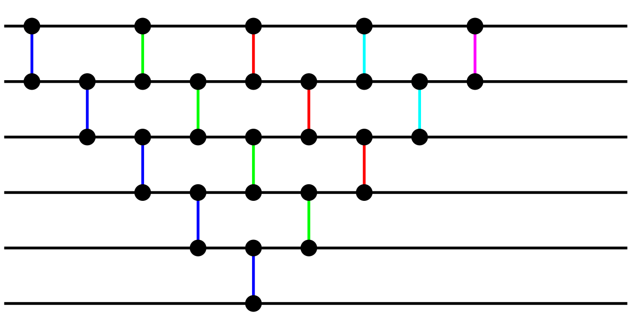
(A) Inner loop of insertion sort is:



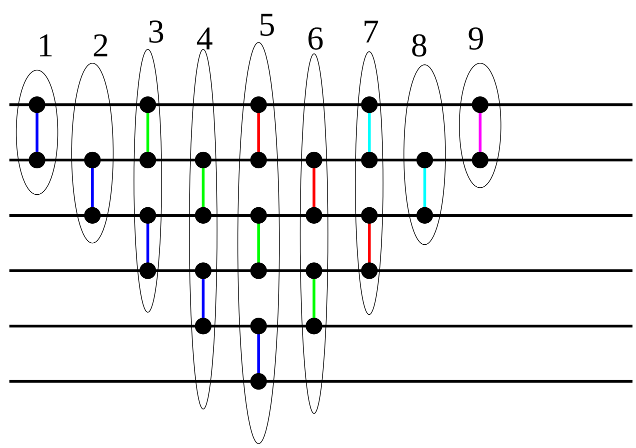
(B) Insertion sort as a network:



24.2.2.2 Sorting network based on insertion sort



(i)

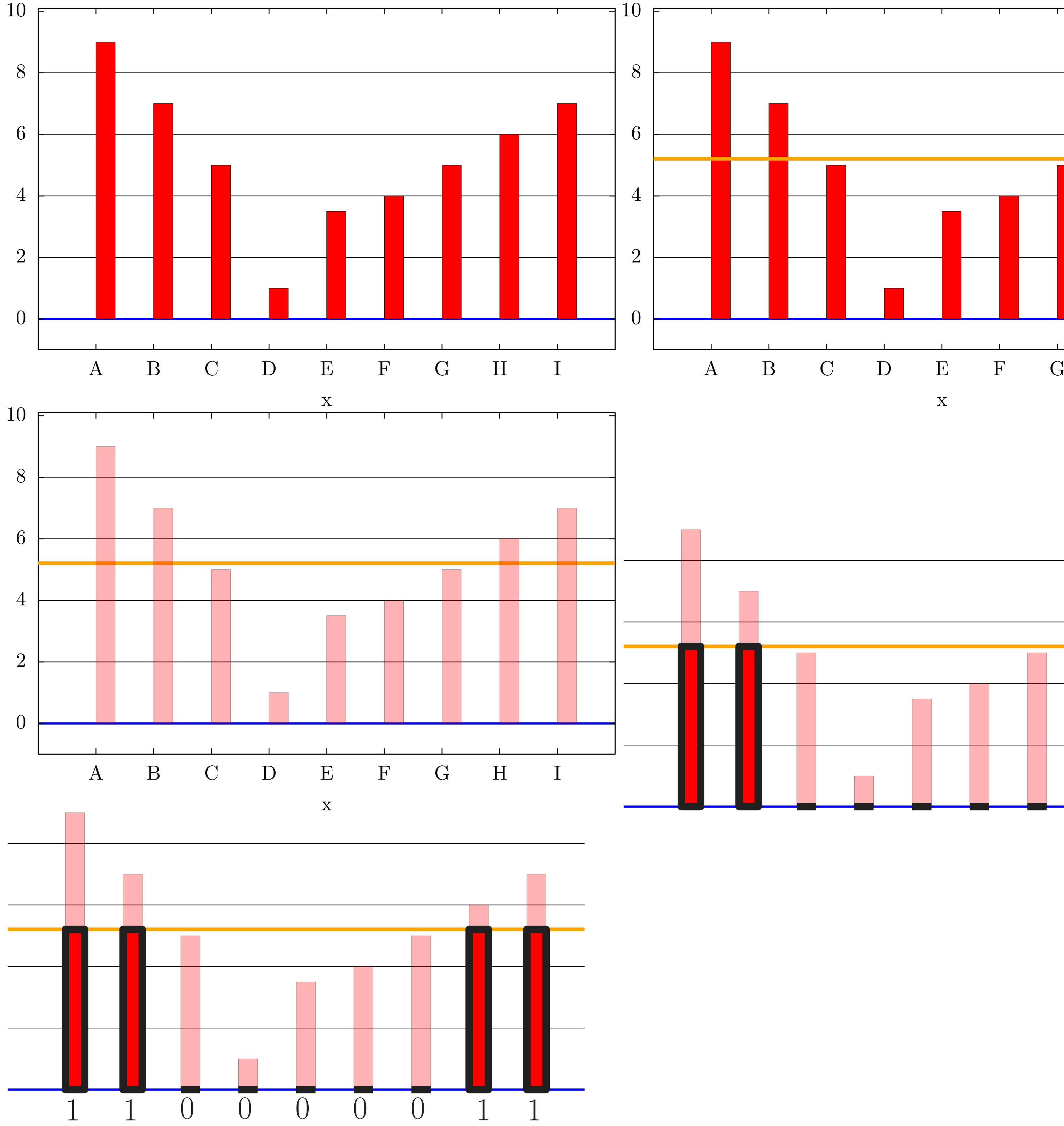


(ii)

Lemma 24.2.4. *The sorting network based on insertion sort has $O(n^2)$ gates, and requires $2n-1$ time units to sort n numbers.*

24.3 The Zero-One Principle

24.3.0.1 Converting a sequence into a binary sequence



24.3.1 The zero-one principle

24.3.1.1 The Zero-One Principle

Definition 24.3.1. ***zero-one principle*** states that if a comparison network sort correctly all binary inputs ($\forall$ input is 0 or 1) then it sorts correctly all inputs (input is real number).

Need to prove the zero-one principle.

Lemma 24.3.2. *A comparison network transforms input sequence*

$$a = \langle a_1, a_2, \ldots, a_n \rangle \implies b = \langle b_1, b_2, \ldots, b_n \rangle$$

Then for any monotonically increasing function f, the network transforms

$$f(a) = \Big\langle f(a_1), \ldots, f(a_n) \Big\rangle \implies f(b) = \Big\langle f(b_1), \ldots, f(b_n) \Big\rangle$$

24.3.1.2 Proof

(A) Induction on number of comparators.
(B) Consider a comparator with inputs x and y, and outputs $x' = \min(x, y)$ and $y' = \max(x, y)$.
(C) If $f(x) = f(y)$ then the claim trivially holds.
(D) If $f(x) < f(y)$ then clearly

$$\begin{aligned} \max(f(x), f(y)) &= f(\max(x, y)) \text{ and} \\ \min(f(x), f(y)) &= f(\min(x, y)), \end{aligned}$$

since $f(\cdot)$ is monotonically increasing.
(E) $\langle x, y \rangle$, for $x < y$, we have output $\langle x, y \rangle$.
(F) Input: $\langle f(x), f(y) \rangle \implies$ output is $\langle f(x), f(y) \rangle$.
(G) Similarly, if $x > y$, the output is $\langle y, x \rangle$. In this case, for the input $\langle f(x), f(y) \rangle$ the output is $\langle f(y), f(x) \rangle$. This establish the claim for a single comparator.

24.3.1.3 Proof continued

(A) Claim: if a wire carry a value a_i, when the sorting network get input $a_1, \ldots, a_n$, then for input $f(a_1), \ldots, f(a_n)$ this wire would carry the value $f(a_i)$.
(B) Proof by induction on the depth on the wire at each point.
(C) If point has depth 0, then its input and claim trivially hold.
(D) Assume holds for all points in circuit of depth $\leq qi$, and consider a point p on a wire of depth $i + 1$.
(E) G: gate which this wire is an output of.
(F) By induction, claim holds for inputs of G.
Now, the claim holds for the gate G itself.
Apply above single gate proof for G.
$\implies$ claim holds at p. ■

24.3.1.4 Sorting correctly binary sequences implies real sorting

24.3.1.5 0/1 sorting implies real sorting

Theorem 24.3.3. *If a comparison network with n inputs sorts all 2^n binary strings of length n correctly, then it sorts all sequences correctly.*

24.3.1.6 Proof: $0/1$ sorting implies real sorting

(A) Assume for contradiction that fails for input $a_1, \ldots, a_n$. Let $b_1, \ldots b_n$ be the output sequence for this input.
(B) Let $a_i < a_k$ be the two numbers that are output in incorrect order (i.e. a_k appears before a_i in output).
(C) $f(x) = \begin{cases} 0 & x \leq a_i \\ 1 & x > a_i. \end{cases}$
(D) By lemma for input $\langle f(a_1), \ldots, f(a_n) \rangle$,
circuit would output $\langle f(b_1), \ldots, f(b_n) \rangle$.
(E) This sequence looks like: 000..0????$f(a_k)$????$f(a_i)$??1111
(F) but $f(a_i) = 0$ and $f(a_j) = 1$. Namely, the output is a sequence of the form ????1????0????, which is not sorted.
(G) bin. input $\langle f(b_1), \ldots, f(b_n) \rangle$ sorting net' fails. A contradiction. ■

24.4 A bitonic sorting network

24.4.0.1 Bitonic sorting network

Definition 24.4.1. A ***bitonic sequence*** is a sequence which is first increasing and then decreasing, or can be circularly shifted to become so.

example The sequences $(1, 2, 3, \pi, 4, 5, 4, 3, 2, 1)$ and $(4, 5, 4, 3, 2, 1, 1, 2, 3)$ are bitonic, while the sequence $(1, 2, 1, 2)$ is not bitonic.

24.4.0.2 Binary bitonic sequences

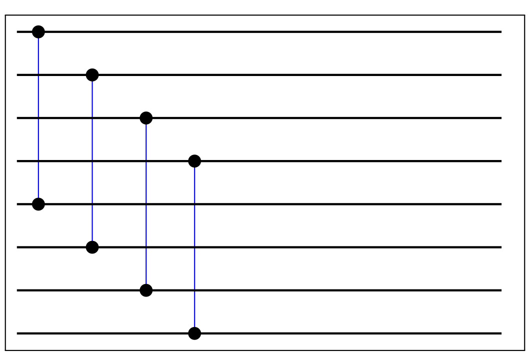
Observation 24.4.2. *binary bitonic sequence is either of the form $0^i 1^j 0^k$ or of the form $1^i 0^j 1^k$, where 0^i (resp, 1^i) denote a sequence of i zeros (resp., ones).*

24.4.0.3 Bitonic sorting network

Definition 24.4.3. A ***bitonic sorter*** is a comparison network that sorts all bitonic sequences correctly.

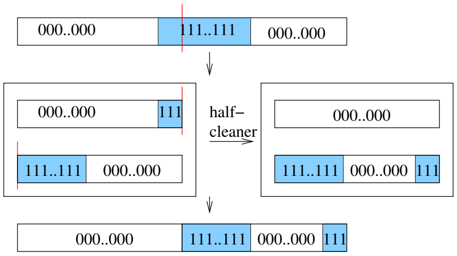
24.4.0.4 Half cleaner...

Definition 24.4.4. ***half-cleaner***: a comparison network, connecting line i with line $i + n/2$.



Half-Cleaner$[n]$ denote half-cleaner with n inputs.
Depth of **Half-Cleaner**$[n]$ is one.

24.4.0.5 Half cleaner on bitonic sequence...



(A) What a half-cleaner do to an input which is a (binary) bitonic sequence?
(B) In example... left half size is clean and all equal to 0.
(C) Right side of the output is bitonic.
(D) Specifically, one can prove by simple (but tedious) case analysis that the following lemma holds.

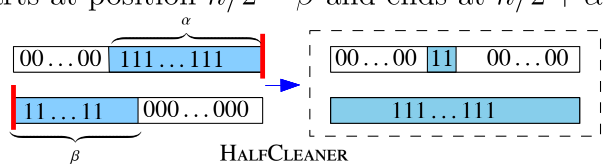
24.4.0.6 Half cleaner half sorts a bitonic sequence...

Lemma 24.4.5. *If the input to a half-cleaner (of size n) is a binary bitonic sequence then for the output sequence we have that*

(i) the elements in the top half are smaller than the elements in bottom half, and
(ii) one of the halves is clean, and the other is bitonic.

24.4.0.7 Proof

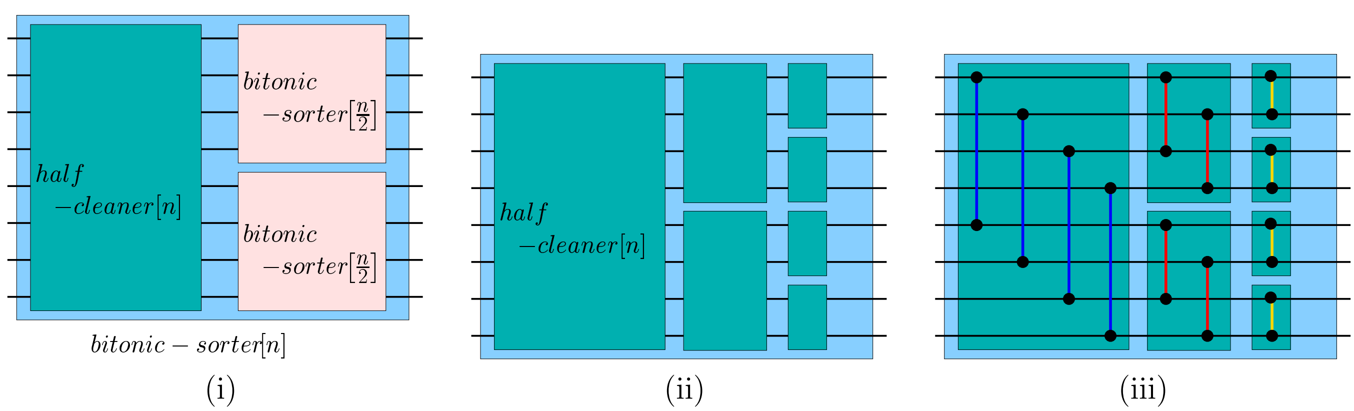
Proof: If the sequence is of the form $0^i 1^j 0^k$ and the block of ones is completely on the left side (i.e., its part of the first $n/2$ bits) or the right side, the claim trivially holds. So, assume that the block of ones starts at position $n/2 - \beta$ and ends at $n/2 + \alpha$.



If $n/2 - \alpha \geq \beta$ then this is exactly the case depicted above and claim holds. If $n/2 - \alpha < \beta$ then the second half is going to be all ones, as depicted on the right. Implying the claim for this case.

A similar analysis holds if the sequence is of the form $1^i 0^j 1^k$. ■

24.4.0.8 Bitonic sorter - sorts bitonic sequences...



(i) recursive construction of **BitonicSorter**$[n]$,
(ii) opening up the recursive construction, and
(iii) the resulting comparison network.

24.4.0.9 Bitonic sorter... the result

Lemma 24.4.6. **BitonicSorter**$[n]$ *sorts bitonic sequences of length* $n = 2^k$*, it uses* $(n/2)k = (n/2)\lg n$ *gates, and it is of depth* $k = \lg n$*.*

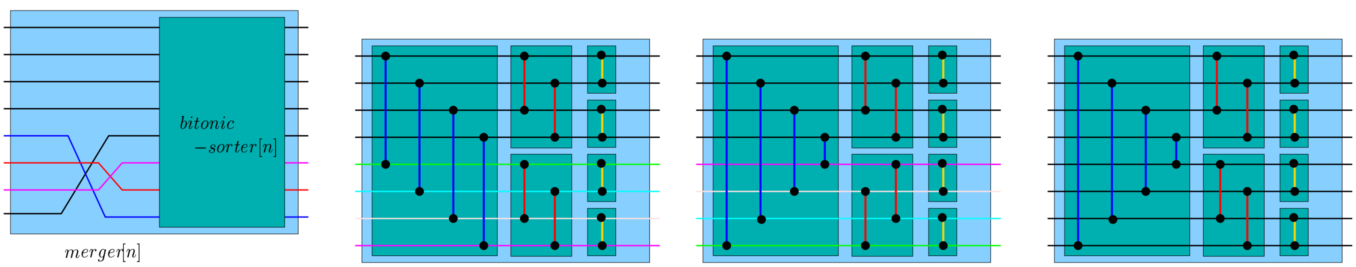
24.4.1 Merging sequence

24.4.1.1 Merging sequence

(A) Merging question: Given two *sorted* sequences of length $n/2$, how do we merge them into a single sorted sequence?
(B) Concatenate the two sequences...
(C) ... second sequence is being flipped (i.e., reversed).
(D) Easy to verify that the resulting sequence is bitonic, and as such we can sort it using the **BitonicSorter**$[n]$.
(E) Given two sorted sequences $a_1 \le a_2 \le \ldots \le a_n$ and $b_1 \le b_2 \le \ldots \le b_n$, observe that the sequence $a_1, a_2, \ldots, a_n, b_n, b_{n-1}, b_{n-2}, \ldots, b_2, b_1$ is bitonic.

24.4.2 $Merger[n]$: Using a bitonic sorter

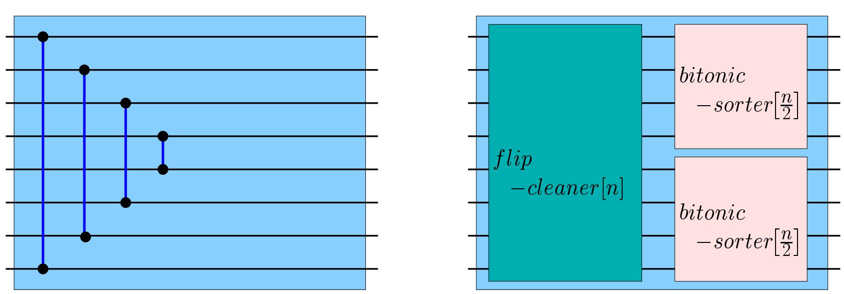
24.4.2.1 Merging two sorted sequences into a sorted sequence



(i) (ii) (iii) (iv)

(i) **Merger** via flipping the lines of bitonic sorter.
(ii) **BitonicSorter**.
(iii) **Merger** after we “physically” flip the lines.
(iv) Equivalent drawing of the resulting **Merger**.

24.4.2.2 Merger$[n]$ described using FlipCleaner



(i) (ii)

(i) **FlipCleaner**$[n]$, and
(ii) **Merger**$[n]$ described using **FlipCleaner**.

24.4.2.3 What $Merger[n]$ does...

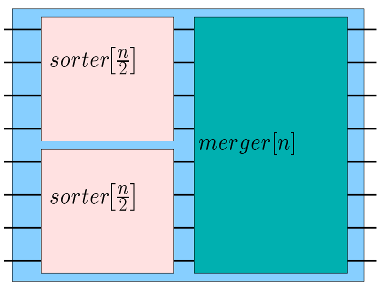
Lemma 24.4.7. *The circuit* **Merger**$[n]$ *gets as input two sorted sequences of length* $n/2 = 2^{k-1}$*, it uses* $(n/2)k = (n/2)\lg n$ *gates, and it is of depth* $k = \lg n$*, and it outputs a sorted sequence.*

24.5 Sorting Network

24.5.1 Sorting Network

24.5.1.1 Finally...

Implement ***merge sort*** using **Merger**$[n]$.

Sorter$[n]$: 

Lemma 24.5.1. *The circuit* **Sorter**$[n]$ *is a sorting network (i.e., it sorts any* n *numbers) using* $G(n) = O(n \log^2 n)$ *gates. It has depth* $O(\log^2 n)$*. Namely,* **Sorter**$[n]$ *sorts* n *numbers in* $O(\log^2 n)$ *time.*

24.5.1.2 Proof

Proof: The number of gates is

$$G(n) = 2G(n/2) + Gates(\textbf{Merger}[n]).$$

Which is $G(n) = 2G(n/2) + O(n \log n) = O(n \log^2 n)$.

As for the depth, we have that $D(n) = D(n/2) + \text{Depth}(\textbf{Merger}[n]) = D(n/2) + O(\log(n))$, and thus $D(n) = O(\log^2 n)$, as claimed. ∎

24.5.1.3 Resulting sorted

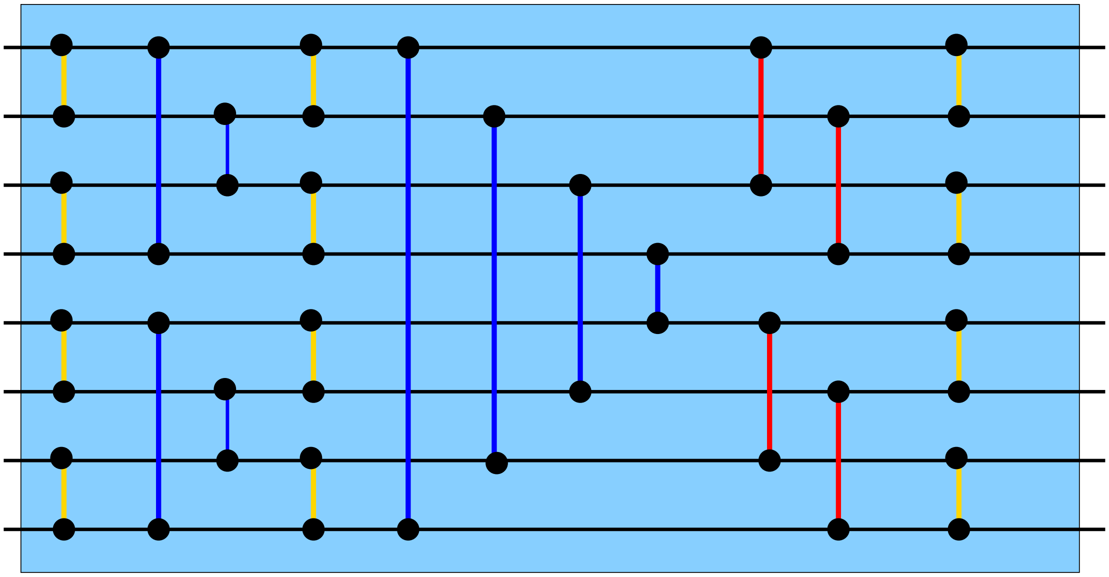


Figure 24.1: **Sorter**[8].

24.6 Faster sorting networks

24.6.0.1 Faster sorting networks

(A) Known: sorting network of logarithmic depth
Ajtai et al. [1983].
(B) Known as the ***AKS sorting network***.
(C) Construction is complicated.
(D) Ajtai et al. [1983] is better than bitonic sort for n larger than 2^{8046}.

Bibliography

M. Ajtai, J. Komlós, and E. Szemerédi. An $O(n \log n)$ sorting network. In *Proc. 15th Annu. ACM Sympos. Theory Comput.* (STOC), pages 1–9, 1983.