6.896 Theory of Parallel Hardware

Administrivia «Mention webpage bug»
Overview «Discuss models»

Systolic computation

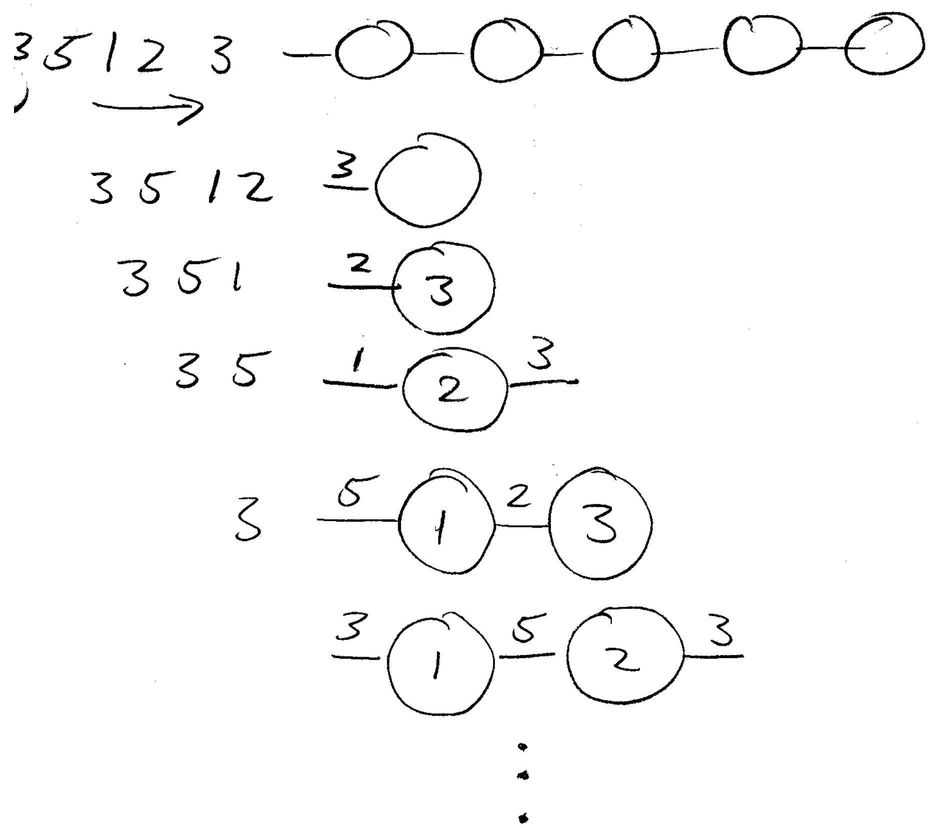
E.g. linear array I/O –O–O–O–O– ... –O

"Fixed-connection" network
1. Underlying graph fixed.
2. Local communication only
3. I/O location restricted.

At each step of a globally synchronous clock, each processor
1. receives inputs from neighbors (or I/O)
2. inspects local memory
3. performs local computation
4. updates local memory
5. generates outputs for neighbors

Example: Sorting

- Accept left input
- Compare input to stored value
- Store smaller value
- Output bigger # to right.



Correctness: induction

N inputs. How many steps? $2N = \Theta(N)$

«Discuss outputting of values.»

Total time $= 3N$

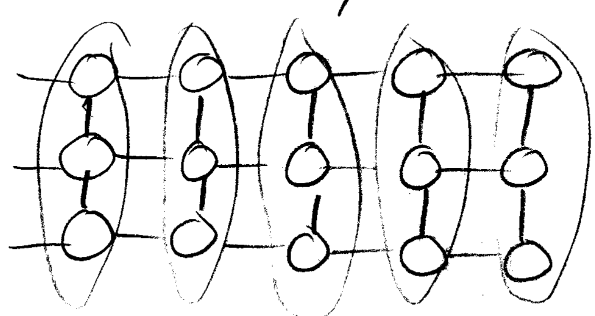
<u>Sorting in the bit model</u> (vs. word model)

- One processor per bit.

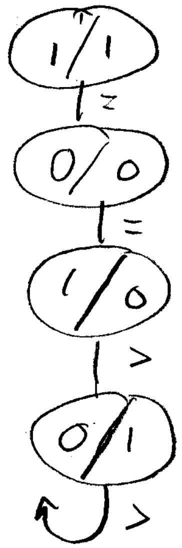
N k-bit #'s

3	5	1	2	3
0	1	0	0	0
1	0	0	1	1
1	1	1	0	1

N × k array

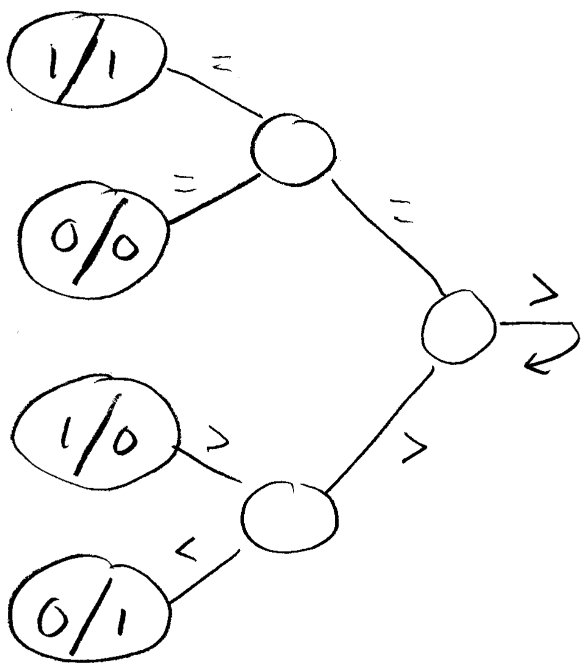


Comparing 2 k-bit words



$\Theta(k)$ steps to compute
Sort in $\Theta(Nk)$ time

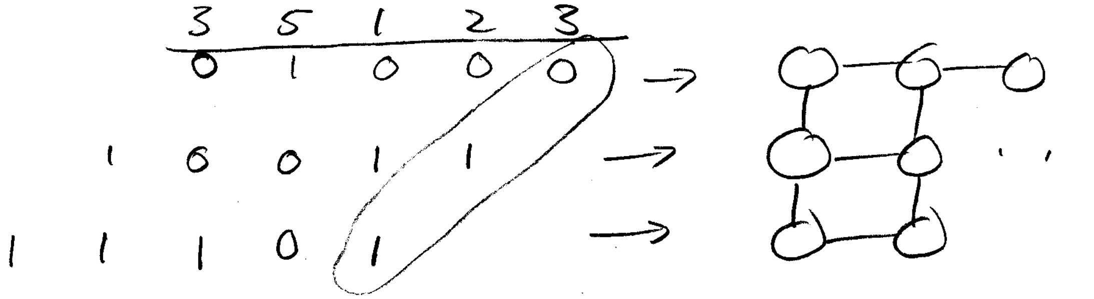
Faster comparison - binary tree



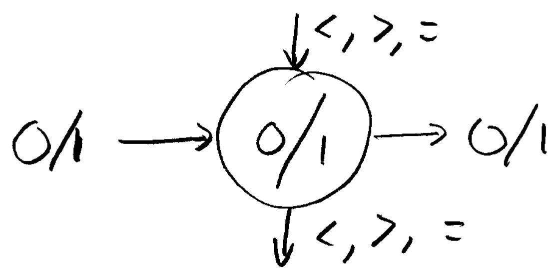
$\Theta(\lg k)$ compare
Sort in $\Theta(N \lg k)$
$\lg = \log_2$

Pipelining

- compare while sorting
- stagger bits of input.



Each processor:



Time = $\Theta(N+k)$ bit steps.
Can we do better on $N \times k$ array?

Lower bounds

1. I/O bandwidth
   Nk bits to input at k places $\Rightarrow \Omega(N)$ steps.

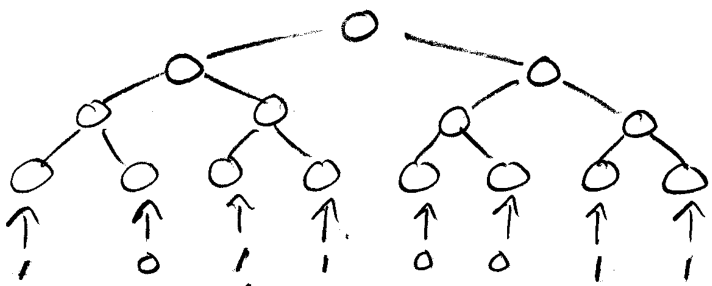
2. Network diameter
   $\Omega(N+k)$

3. Communication bandwidth (bisection width).

$$T \geq \frac{\#\text{ bits crossing cut}}{\text{size of cut}}$$

$$T \geq \frac{\Theta(Nk)}{\Theta(k)} = \Theta(N).$$

<u>Problem</u>. N 1-bit #'s input at N leaves of complete binary tree. Time to sort?

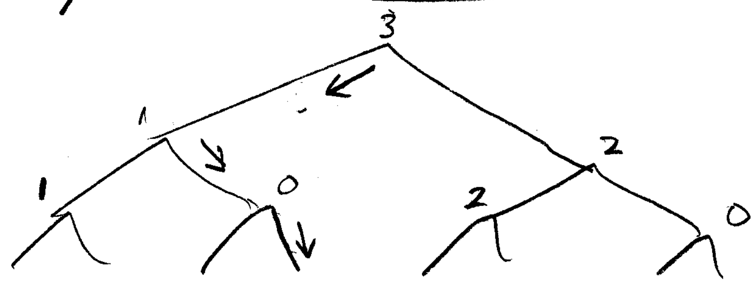


I/O : $T \geq N/N = 1$.
Diam : $T \geq 2 \lg N$
BW : $T \geq \Theta(N)/1 = \Theta(N)$

∴ Sorting N 1-bit #'s takes $\Omega(N)$ time on CBT.

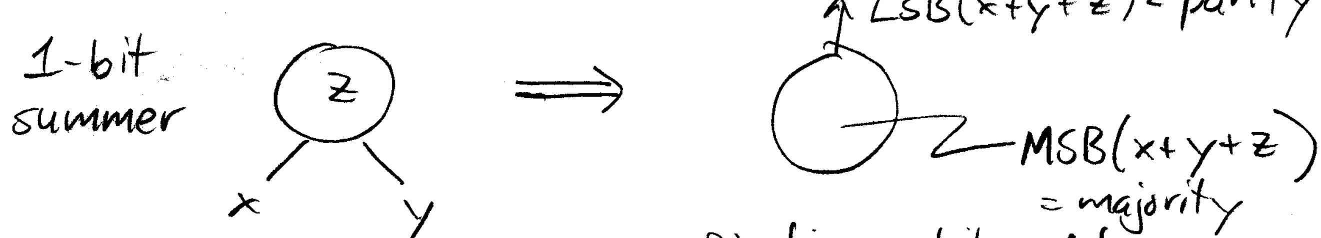
Wrong! Can sort in $O(\lg N)$ time!

<u>Idea</u>: Only need to <u>count</u> # 0's.



Input	1	0	1	1	0	0	1	1
Output	0	0	0	1	1	1	1	1

Sum 0's upward, select downward.



Pipeline in bit model.

Q. Why doesn't BW lower bound hold?
A. Didn't really need to ship $\Theta(N)$ bits across bisection. Could encode info more compactly.
Careful!