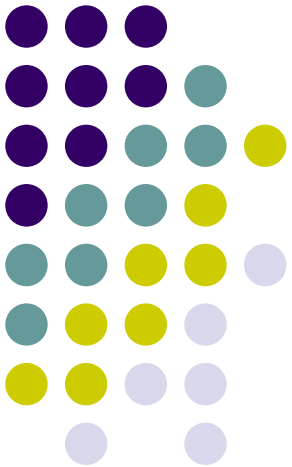
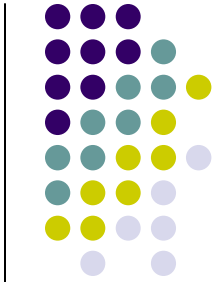
CS256
Applied Theory of Computation

VLSI Model I

John E Savage



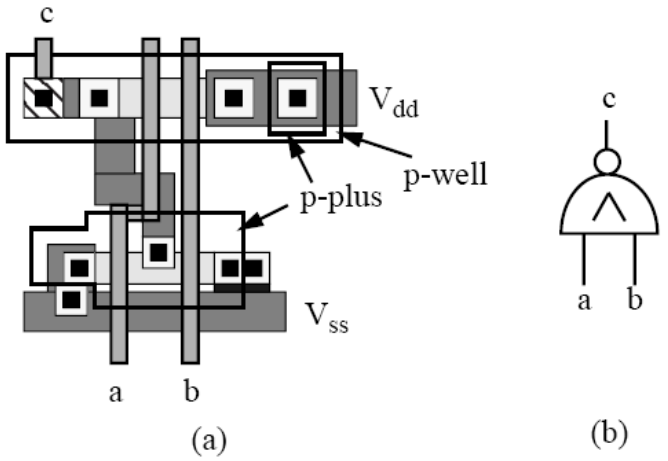
Overview



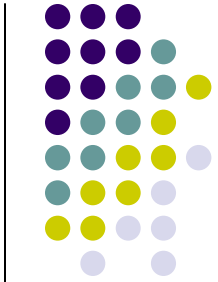
- History of the VLSI revolution
- The VLSI model
  - Physical, architectural, functional and computational models
- The H-tree
- Parallel prefix computations on trees
- The Cube Connected Cycles architecture
  - Efficient use of area and time

The VLSI Revolution

- **Time Line**:
  - 1947 - transistor invented
  - 1958-9 - integrated circuit invented
  - 1970's - small CPU realized on a chip
  - 2000's - > 100 million transistors on a chip
- Gates constructed of overlaid rectangular sections of materials.



The VLSI Physical Model



- Circuits grown on a semiconductor substrate.
- Wires and transistors formed by depositing wires and doping regions of the substrate.
- **Lithography**: Use of light to define rectangular regions on the surface of a chip.
- Regions exposed to light can be removed by an etchant. Exposed areas (or complements) can be doped or covered with metals.

The VLSI Architectural Model – Chips Realize FSMs

- Wires are rectilinear.
- Wires have bounded width and separation λ.
- Gates can be binary or non-binary.
- Gates/memory cells occupy area proportional to λ^2.
- I/O pads occupy area proportional to λ^2.
- Wires on at most $\nu \geq 1$ levels. Gates on one level.
- Time for signal to travel wire of length l is constant.
- (Time in *diffusion model* is proportional to l^2.)

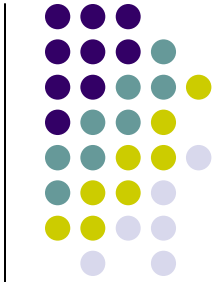
The VLSI Functional and I/O Models

- Finite functions computed.

  $f : B^n \rightarrow B^m$ where B is any set, usually {0,1}.

- I/O Assumptions:
  - Inputs provided at times and places on the chip that are data-independent.
  - Each input is supplied once (*semellective alg.*) or provided multiple times (*multilective alg.*).
  - Outputs produced once at times and places on the chip that are data-independent.

Performance of VLSI Chips

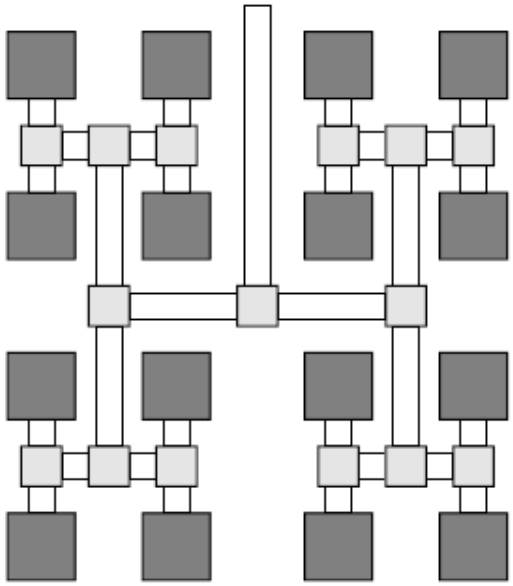


- Wires occupy area – the new ingredient.
- Model assumes unit time for data to move through gates or across wires!

- VLSI performance measures:
  - Chip area = A
    - Chip failure probability rises quickly with area.
  - Computation time = T

Chip Layout – The H Tree

- Trees are very important. The binary H-tree uses area wisely. Each leaf and interior node is square. Leaves have area b^2 and interior nodes have area c^2, $c \leq b$.
- Let H_k be an H-tree with $4k$ leaves.
- Let $S(k)$ denote length of one side of H_k.



Chip Layout – The H Tree

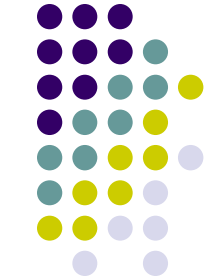
- Then,

$$S(1) = 2b + c$$

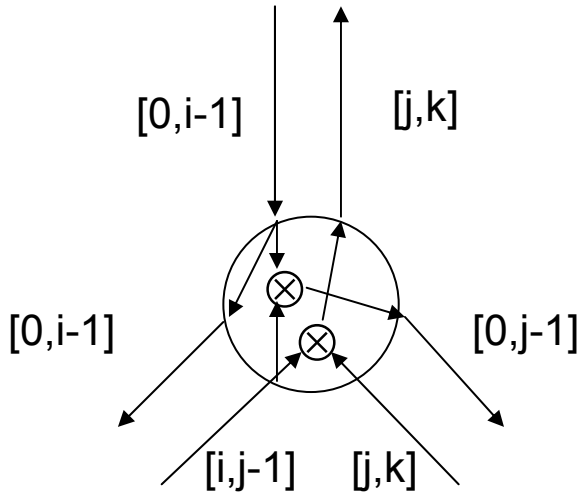
$$S(k) = 2S(k-1) + c = (b+c)\,2^k - c$$

  - Area $A(k)$ of the H-tree H_k is $A(k) \leq (b+c)^2\,4^k$.
  - $S(k) = O(\sqrt{A(k)})$
- If a binary tree has its n leaves placed on a convex boundary, it requires area $\Omega(n \log n)$.
  - How would you show this?

Prefix Computations on a Tree

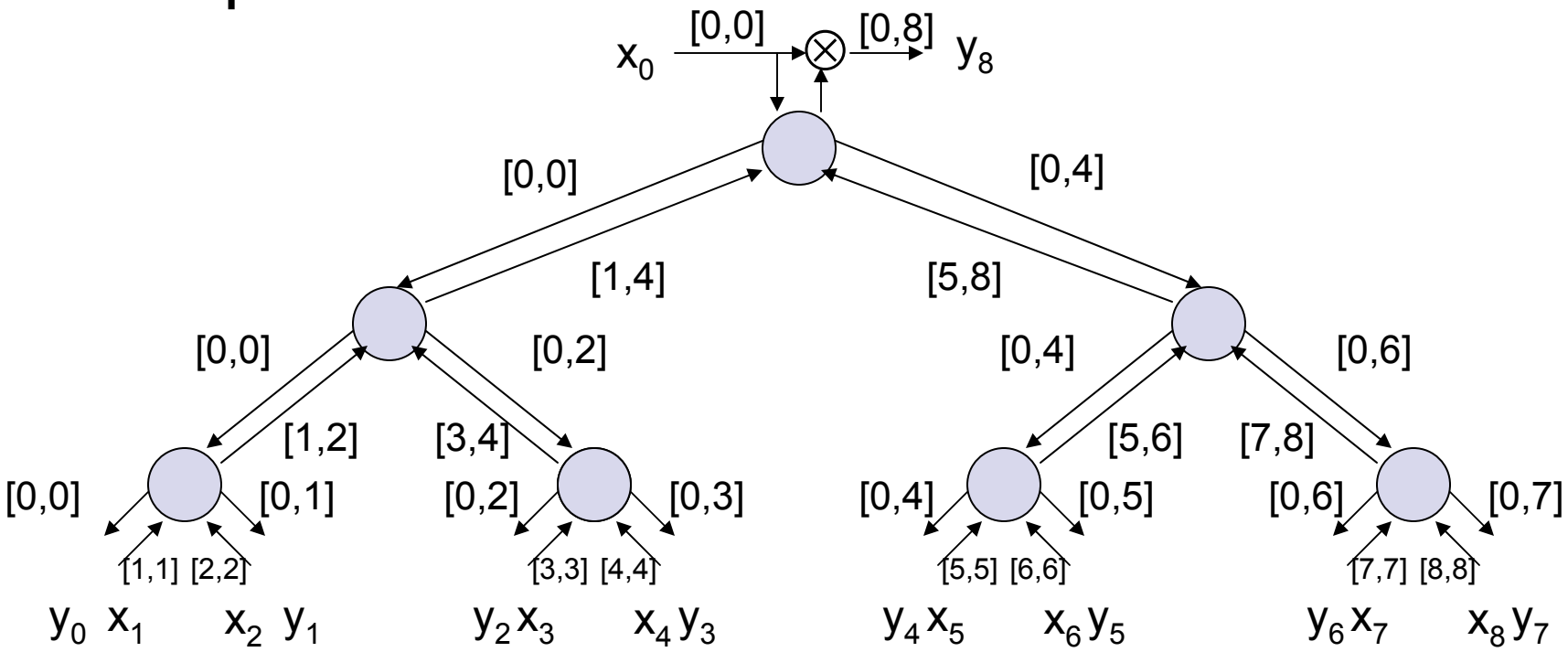


- Denote $x_i \otimes x_{i+1} \otimes x_j = [i, j]$ where $\otimes$ is an associative operator.
- Consider the following widget:



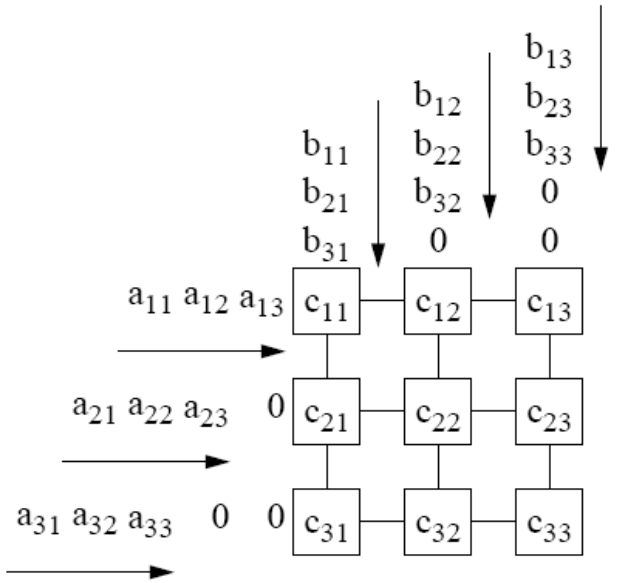
Prefix Computation on a Tree

- A binary tree of this widget performs a prefix computation.



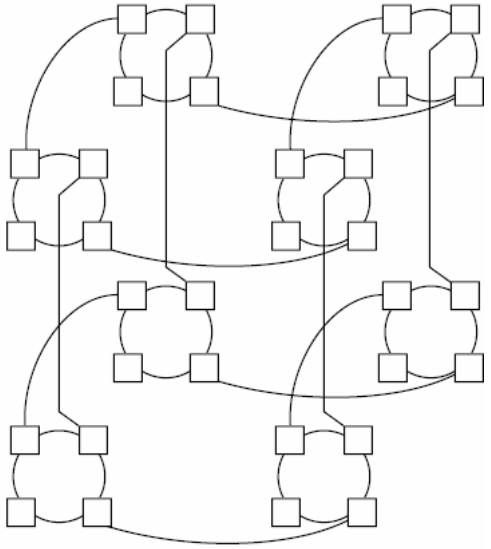
VLSI Architectures

- Tree-based computations
  - Matrix-vector multiplication
  - Prefix computations
- Mesh-based computation

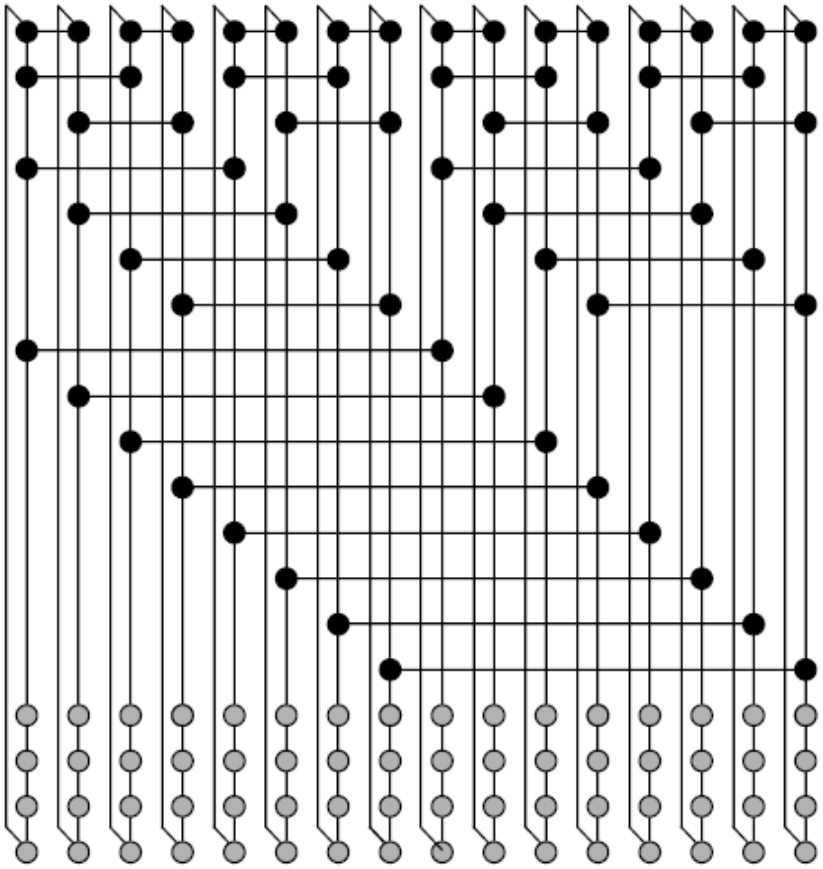


Cube Connected Cycles

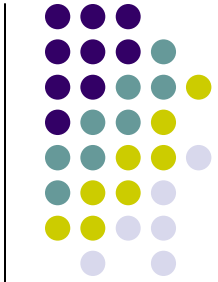
- Consider the (k,d)-CCC network with 2^d cycles, 2^k vertices per cycle. Shown is $k = 2$, $d = 3$. The jth vertex on a cycle connected to jth vertex on neighboring cycle.



Layout of CCC Network When k = 3, d = 4



CCC Layout



- The CCC network has $n = 2^{d+k}$ vertices, $2^k \geq d$, and 2^d columns. Index columns from left by binary tuples starting with (000...0). First (kth) dimension connections are made using one (2^{k-1}) row(s). Then, R = 1 + 2 + ...+ $2^{d-1} = 2^d - 1$ rows are used for connections. $2^k - d$ more rows are used for the $2^k - d$ extra processors on each cycle. Thus, R $\leq 2^d + 2^k - d$ and area A = R$2^d \leq 2^d(2^d + 2^k - d)$. T = O($d + 2^k$) steps for normal alg. to move data across d dimensions and within cycle of length 2^k. When $2^k \approx d$, following theorem holds.

Execution on CCC Layout

Theorem Every fully normal algorithm for a n-processor hypercube can be implemented on a VLSI CCC network with area A and time T satisfying following for $\Omega(\log n) \leq T = O(\sqrt{n})$.

$$AT^2 = O(n^2).$$