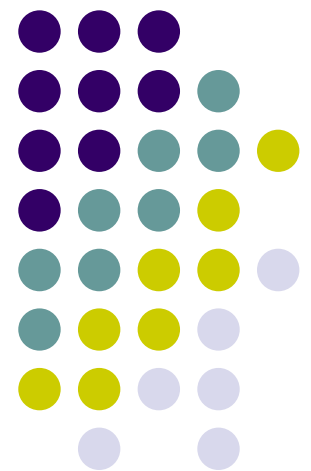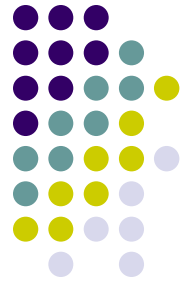# CS159 Introduction to Computational Complexity

The VLSI Model II

The VLSI Model

- **Architectural Model**:
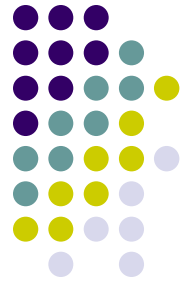 - Chips realize FSMs
  - Wires are rectilinear.
  - Wires have bounded width and separation λ.
  - Gates can be binary or non-binary.
  - Gates/memory cells occupy area proportional to λ^2.
  - I/O pads also occupy area proportional to λ^2.
  - Wires on at most v ≥ 1 levels. Gates on one level.
  - Time for signal to travel wire of length l is constant.
 - (Time for *diffusion model* is proportional to l^2!)

The VLSI Model

- **Performance Measures**:
  - Chip area A - manufacturing cost increases with A
 - Number of steps T - reflects computational cost

The VLSI Model

- **Functional Model**: Finite functions computed
  $f : B^n \to B^m$ where B is any set, usually $\{0,1\}$.

- **Algorithmic Model**: Limitations on I/O
  - 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 are produced once at times and places on the chip that are data-independent.
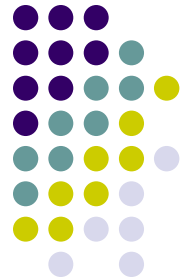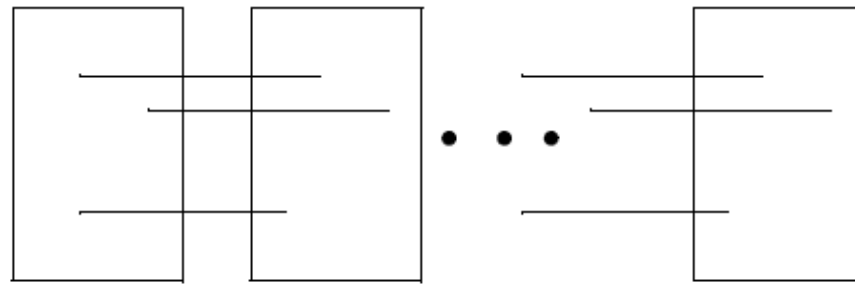
Tradeoffs Via Planar Circuits

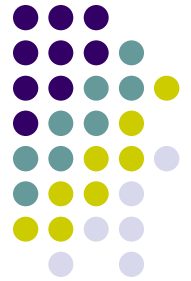- We simulate chip (an FSM) in two ways to create a planar circuits.
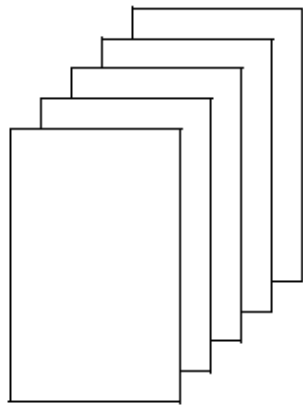
(a)   (b)

- Simulation (a) gives a planar circuit of size $O(AT^2)$.
- Simulation (b) gives size $O(A^2T)$.

Tradeoffs Via Planar Circuits

- The size of the smallest planar circuit, $C_p(f)$, for a function f computed by the chip of area A in T steps is no larger than either bound.

$$C_p(f) = O(A^2 T) \text{ and } C_p(f) = O(AT^2)$$

- We use the **planar separator theorem** to derive lower bounds on the planar circuit size, $C_p(f)$, of the function f. If $C_p(f)$ is large, so are both $A^2 T$ and AT^2.
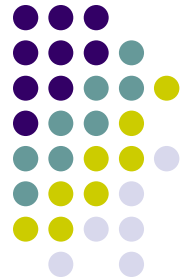
First Computational Inequality
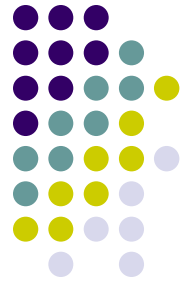
To show $C_p(f) = O(AT^2)$, stack T copies of chip one above the other; replace each memory cell by a wire and pass a wire from its output on one stack to its input on the next. Produce a planar circuit by shrinking all inputs, wires, and gates to zero width.
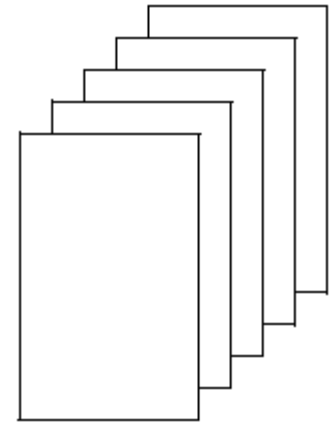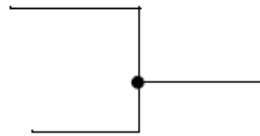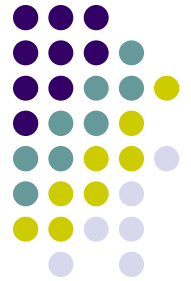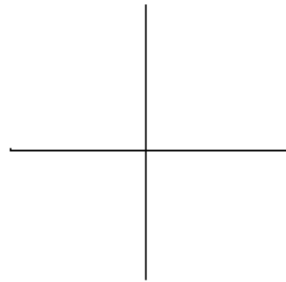
First Computational Inequality

- Displace T copies of the chip infinitesimally from one another to the northeast. When seen from above, this exposes crossing (rectilinear) wires as well as individual gates and inputs. The number of gates and inputs is T times the number on one chip. How many crossing wires result from displacement?
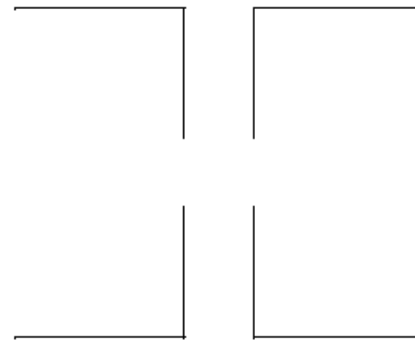
First Computational Inequality

(a)   (b)   (c)
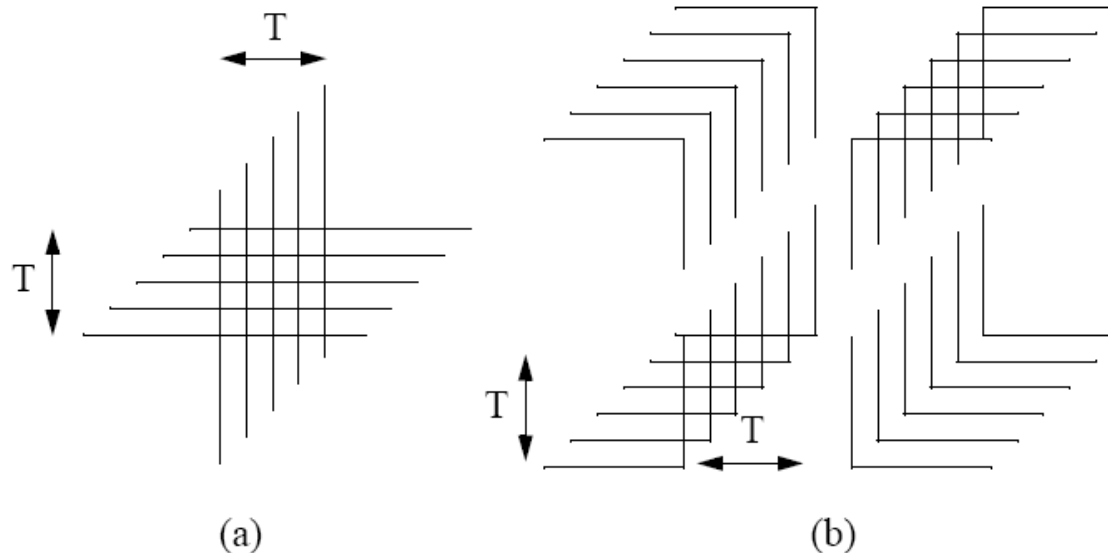
- Shown in a) is effect of gate shrinkage. b) shows crossing wires, and c) shows four ways wires meeting.
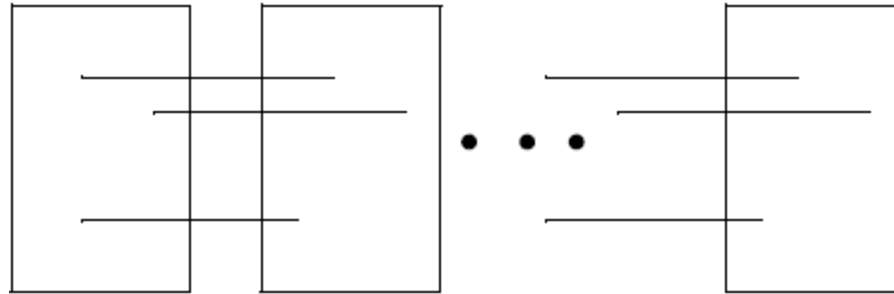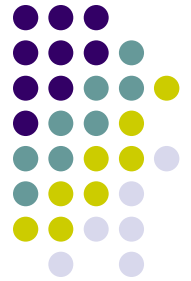
First Computational Inequality

(a)   (b)

- Maximum number of crossings resulting from meeting or crossing wires is $O(T^2)$. Number of original crossings or meetings on chip is at most A/λ^2. Thus, the size of planar circuit in this construction is $O(AT^2/\lambda^2)$.

Second Computational Inequality $C_p(f) = O(A^2T)$

- Stretch each chip enough to pass wires from memory cell outputs on one chip to inputs on another.

- Since a wire from a cell (there are $n_W \leq A/\lambda^2$) may cross every wire in the original circuit, number of crossings in one copy of the chip is increased from the original value ($\leq A/\lambda^2$) by at most n_W^2.

- No. of crossings and gates = $O(AT + A^2T) = O(A^2T)$.

Computational Inequalities

$$C_p(f) = O(min(AT^2, A^2T))$$

- To derive lower bounds on $C_p(f)$ we introduce the **planar separator theorem**. In its simplest form, it states that the vertices in every planar n-vertex graph can be divided into two sets with no edges between them by the removal of $O(\sqrt{n})$ vertices such that each set contains between $n/3$ and $2n/3$ vertices.

- We use the planar separator theorem to show that some functions have a quadratic planar circuit size measured by their number of inputs. The technique used is to show that a lot of information must pass from inputs to outputs.
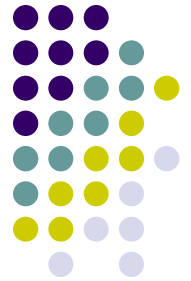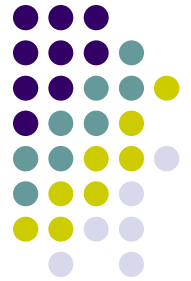
Flow Properties of Functions

Definition $f : A^n \rightarrow A^m$ is (α, n, m, p)-**independent** for $\alpha \geq 1$ and $p \leq m$ if it has a $w(u,v)$-flow satisfying $w(u,v) > (v/\alpha) - 1$ for $n - u + v \leq p$.
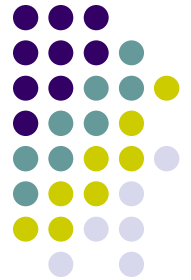
Note: If $|X_1| = u$ and $|Y_1| = v$, $|X_0| + |Y_1| = n - u + v$.

Lower Bounds on Planar Circuit Size

If f on n inputs and m outputs is (α, n, m, p)-ind. for $p \leq m$ and $\alpha \geq 1$, there are $> |A|^{(v/\alpha)-1}$ points in the image of its domain when $\leq p - v$ inputs are fixed.

Name	In's (**n**)	Out's (**m**)	Ind. Property
Wrapped conv	$2n$	n	$(2,2n,n,n/2)$
Cyclic shift	$n + \log n$	n	$(2,n+\log n,n,n/2)$
Int. multiply	$2n$	$2n$	$(2,2n,n,n/2)$
n-point FFT	n	n	$(2,2n,n,n/2)$

Area-Time Lower Bounds

Theorem Area A and time T required to compute wrapped convolution and cyclic shift of n inputs, integer multiplication of n-bit integers, or the FFT on n inputs on semellective VLSI chip satisfy following:

$$AT^2, A^2T = \Omega(n^2)$$

AT^2 lower bound can be achieved up to a constant multiplicative factor for each of these functions for $\Omega(\log n) \leq T \leq \sqrt{n}$.
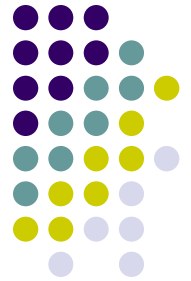
Area-Time Lower Bounds

Proof Lower bound follows from circuit lower bounds.

From previous lecture, for any fully normal algorithm $AT^2 = O(n^2)$ for $\Omega(log\ n) \leq T \leq \sqrt{n}$ on an embedded CCC network. Since cyclic shift and FFT are shown to be fully normal, we have matching upper and lower bounds for them.

From Prob. 12.13 wrapped convolution can be realized with matching bounds on AT^2 over the same range of values for T. Same applies to integer multiplication. (Prob. 12.16.)
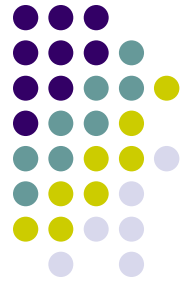
Comments

The planar circuit size for cyclic shifting is $\Omega(n^2)$ but the function can be realized with O($n \log n$) gates.

Since $AT^2 = \Omega(n^2)$, wires occupy more area than gates when T = O($\sqrt{n}/ \log n$).

Matrix multiplication lower bound requires more refined analysis.
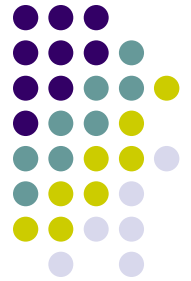
Area-Time Bounds for Matrix Multiplication

Theorem The area A and time T required to multiply two *n* × *n* matrices with a semellective algorithm satisfies the following lower bound:

$$AT^2, A^2T = \Omega(n^4)$$

The AT^2 lower bound can be met to within a constant multiplicative factor.

Area-Time Bounds for Matrix Multiplication

Proof Apply Theorem on p. 6 to matrix multiplication by replacing the number of input variables n by $2n^2$ and the number of output variables m by n^2. The $w(u,v)$-flow function has value

$$w(u,v) = (v-(2n^2-u)^2/4n^2)/2$$

$$w(u,v) \geq (n^2/2)\ (1/(3P) - (3/(2P))^2)$$

The r.h.s. is maximized when P = 14 and the coefficient of n^2 has value greater than 1/163.