LECTURE #14  NOVEMBER 7

- **TODAY:** Review Exam; Effects of Scaling
  **GLEN:** Cell Library
- **NOTE:** Project Assign #2 due Thurs. Please hand in Xerox copy so I can keep your det. prod. desc. for reference.

Time is marching on. Anyone who hasn't got a project idea by now should make an appt. to see me - I'll help select an idea.

By the way: Good news RE Project Set Imp.

- **NOTE:** On Nov 21 (Tues) Dick Lyon of Xerox Park will pres a seminar on VLSI impl. of speech processing functions.

He will pres. some basic building blocks: adders, multiplier, memory subsystems, etc. of which one can build a signal proc. system. He will then discuss an example project chip now in design: A Word-All Digital BandPass Filterbank which will fit on one chip even in prod. technology.

He will give a pres. of the ov. sys. des of an isolated utterance recognition sys. which should fit on one or a few chips within a few years. (For voice input to com-pou.)

- **NOTE:** Calvoc Mead of Caltech will present a seminar on highly concurrent systems (Chap 8 mail!) on Tuesday Dec 5.

[Try to have Levitt, Todd, take exam during class.]
• **Let's Review The Exam:** Hand-Out Graded Exams
  
  • The mean = 80, median = 86
  Highest grade = 96 (2 people: Guy Steele, Moshe Ban)
  
  • 11 between 90-96
  9 between 80-89
  10 below 80

  Most did very well. Consider over 80 as a good grade.

  Refer to exam sheets:

  I graded fairly hard. A few who have done very well on HW did poorly. Maybe had a sloppy. Don't be discouraged. But done who did...

Problem 1 (a) \(C_L = 0.8 \text{ pf}\)

\[C_g(\text{min}) = 4 \times 10^{-4} \text{ pf} \times (6 \times 6) \mu m^2 = 0.0144 \text{ pf}\]

\[Y = \frac{C_L}{C_g} = \frac{0.8}{0.0144} = 55.6\]

For min delay, ratio of successive sizes (f) = e.

\[f^N = Y, \quad N = \ln_e Y = \ln(55.6) = 4.017\]

\[\therefore N = 4\] stages

Grading: Calc. err: \(n = 4\); lnerr: form \(n = 8\)

(b) Trickier than it looks: Array contains 256 pairs of inverters.

[3 SLIDES]

i.e. 16 rows of 16 cells \(\times 2\) invs. per cell (4 per cell pair)

• However, only half can be on at one time maximum.

• First of each pair is approx. 5 \(\mu\)sec long when "on".

• Second is \(6/16 \approx "2"\) sec.

• So worst case is when all "1st" inverters "on".

• In this case, \(I_{total} = 256 \times \frac{5}{5 \times 10^{-4}} = 25.6 \text{ ma}\)

• Wires can carry max 1mA/mm² but aren't 1mm thick, so:

\[\therefore \text{ wires} \geq 25.6 \mu m\] wide

Grading: Calc. errors -3 to -4; Inv. pl: Assume: -4 to -6; Forget pull-down R
Problem 2: Most people did O.K. Most PLA's matched the logic equations used.

Errors were in manipulation of logic equations, or in not getting minimum # p-terms.

Coding: -4 to -6 for log. eqn error(s)
-3 for not minimizing # p-terms
-3 for PLA coding error

Problem 3:

(a) Yellow implant must be 1/2 \( \lambda \) from neighboring euh. mode transistor gate region. Slide

Some said metal not over cuts, but look closely, O.K.

(b) Several Major Errors: Slide

> Ratio should be 8:1, not 4:1.
> No implant over pullup to make dept. mode.
> Metal over Bitty contact is 2\( \lambda \) not 3\( \lambda \) from GND METAL.
> Some DIFFS ARE 2\( \lambda \) Sep. vs. not 3\( \lambda \).

(c) Examples of real improv. in desc. layouts over what C1F can do:

> Dig at corners of boxes (not C1F)
> Direct iteration (as in inform. desc. long in check)
> Pass parameters to called symbol: Use for example: scaling
> Use #'s other than wire indices, allowing easy
> spef of sizes in \( \lambda \).
> Call symbols by name, inst. of #.

; etc.
Problem 4: \( T \rightarrow \text{TFP} \rightarrow Q \)

Many solutions: A frequently used PLA!

- Some based on XOR:

- Others:

  Based on directly selecting inverted or non-inverted feedback based on value of \( T \)

(a)

(b)

(c) Jim Cherry:

(d) Jim Frankel:

(e) Geddy Young:
WHAT HAPPENS AS WE MAKE THINGS SMALLER?

IS. AS \( \lambda \rightarrow \infty \) IN FUTURE

THE EFFECTS OF SCALING DOWN DIMENSIONS OF INT. SYSTEMS:

SUPPOSE WE USE A NEW PROCESS IN FUTURE, SO THAT
ALL DIMENSIONS \text{PARALLEL TO SURFACE OR VERT}\ ARE
REDUCED BY A FACTOR \( \lambda \). WHAT WILL BE EFFECTS ON
ELECTRICAL PERFORMANCE OF DEVICES?

- Distance divided by \( \lambda \):
  \[ W' = \frac{W}{\lambda} \quad ; \quad L' = \frac{L}{\lambda} \quad ; \quad D' = \frac{D}{\lambda} \]

- Also, to keep all Electric Fields same as before,
  we reduce VDD and set all thresholds to be reduced
  by same factor \( \lambda \).

- This is a particularly simple form of scaling to analyze.
  The actual forms used will differ as we approach
  extremely small device sizes, but \( \lambda < \frac{1}{2} \) for
  MOS down to or under 1 \( \mu \text{m} \) feature sizes \( (\lambda < \frac{1}{2} \mu\text{m}) \).

- Now what happens to \( T' \), to \( C_g \), to \( I_{\text{ds}} \), etc.?

- Remember Eqn 1 (CH. 1): \( T = L^2 / \mu \text{Vd} \)
  \[ \frac{T'}{T} = \frac{(L/\lambda)^2}{(V/\lambda^2) \cdot L^2} = \frac{1}{\lambda} \]
  \[ T' = T / \lambda \]

- Transit time of MOS devices, i.e. circuit speeds all scale down linearly with \( \lambda \)

- Gate Capacitance:
  \[ C_g = \varepsilon W L / D \]
  \[ \frac{C_g'}{C_g} = \left( \frac{W}{\lambda} \right) \cdot \left( \frac{D}{\lambda} \right) \cdot \frac{1}{W L / D} = \frac{1}{\lambda} \]
  \[ C_g' = \frac{C_g}{\lambda} \]
- Note, however, that capacitances in general scale up by $\alpha$ if given in terms of $C/\mu m^2$. Since for given absolute $W \times L$, $D$ gets thinner.

- **Note:** Resistances in general scale up by $\alpha$ since lines get thinner vertically. However, the $R/\ell$ of transistors remains $\sim$ same.

- Because of this, and because of other problems with POLY (if clumps of oxides too big $R$ gets large and $V_{th}$ gets worse as $\ell$ gets smaller), the ratio of POLY $\ell$ to FET $\ell$ may get worse as $\ell$ gets smaller.

- **From Eqn 3: (Ch. 1)** 
  \[ I_{ds} = \frac{M \cdot W \cdot (V_{gs} - V_{th})}{L \cdot D} \]
  
  So, \[ I_{ds} \propto \frac{W \cdot V^2}{L \cdot D} \]
  \[ \frac{I'}{I} = \frac{W \cdot V^2}{L \cdot D} \cdot \frac{1}{\ell^2} = \frac{1}{W \cdot V^2 \cdot \ell^2} \]
  \[ \frac{I_{ds}}{\ell^2} = \frac{I_{ds}}{\alpha^2} \]

- So, current per device goes down by $\frac{1}{\alpha^2}$

- However, # devices per unit area goes up by $\alpha^2$

- **Current Density over the chip (if uniform design or repeated designs...etc.) goes up linearly with $\alpha$.**

This is another problem: We may need METAL wires which are proportionally wider compared to previous designs as we scale down other features. If the wires are near the $1 mA/\mu m^2$ current limit.

One solution: Increase $H/w$ of wires by $\alpha^2$ as we scale down. Some of this can be done, but impractical to do completely. So some wires get wider!
Scaling of Power Density:

- **DC Power Dissipation**: Per device $P_{dc} = I \cdot V$

  Since $I' = I/\alpha$, $V' = V/\alpha$, $P_{dc} = \frac{P_{dc}}{\alpha^2}$ per device.

  But the number of devices increases as $\alpha^2$. So, power density (average) over the chip (if regular pattern scaled) remains constant.

- **Switching Power**: The drivers which operate

  pass gates, charging/discharging capacitances, dissipate switching power. We can estimate $N_\text{total}$ roughly with $P_{sw}$ by calculating, estimating $I, V$ during transients. But more directly:

  $P_{sw} \propto \frac{C V^2}{T}$; i.e., $P_{sw}$ is the energy stored on the capacitance of a device divided by the clock period or time between successive charging/discharging.

  Thus, $P_{sw} \propto \frac{N V^3}{D L}$, $P'_{sw} = \frac{P_{sw}}{\alpha^2}$ per device.

  So both dc and sw power remain constant per unit area.

  **Note**: The average dc power for most systems can be approximated by adding total $P_{sw}$ to $1/2$ the dc power resulting if all levels restoring logic pull downs were turned on.

  $[P_{sw} \approx 0.35 \text{ in control drivers to } 0.11 \text{ gates}].$
Rules of Thumb for Power Densities:

While things don't get worse as we scale down, be careful:

- Over areas of dimensions>larger than wafer thickness (i.e. macroscopic dimensions), the following rough rules of thumb apply:

  - < 1 watt/cm² no problem at all
  - 2 watt/cm² I somewhere in here, you begin to
  - 4 watt/cm² need special heat sinking
  - 8 watt/cm² begin to need forced cooling of
    some kind to remove heat.

- As in all thermo problems, all depends on next higher context. One Boltered chip in well heat sinned
  package may do well in still air at 4 w/cm², while
  a whole board (with many chips) all of them would need
  to have plumbing and be room-cooled.

What can I say to summarize: Be careful!

- Calculate power dissipated by large projects or full chips.
  - If > 2 watt/cm² you may get in trouble if you want to use a lot of them.

Ways out: There are often many ways to trade-off
  power vs time.

Remember, by using longer pullups/pulldowns we can keep
  some ratio in shift register, but use less power,
  at price of longer delays.
• **LET'S CALCULATE PERHAPS A WORST CASE FROM AMONG OUR DESIGNS:**

**SHIFT REG IN FIG. 8 b (CH4):** THIS HAS WIDE PULLDOWN, SHORT PULLUP, EVEN THOUGH 8:1.

\[
\text{CELL AREA} = 19\mu m \times 21\mu m = 57\mu m^2 \times 63\mu m^2 = 57 \times 10^{-6} \text{cm}^2 \times 63 \times 10^{-4} \text{cm}^2 = 3.591 \times 10^{-8} \text{cm}^2 = 3.6 \times 10^{-5} \text{cm}^2
\]

\[
\text{Power} = I \cdot V = \frac{5V \cdot 5V}{60K\Omega} = 4.2 \times 10^{-4} \text{Watts}
\]

(Actually the 60kΩ/Ω is for fast tracks; 1Ω is optimistic)

\[
\text{PWR/AREA} = \frac{4.2 \times 10^{-4}}{3.6 \times 10^{-5}} = 12 \text{W/cm}^2. \text{ BUT IN SERIES ONLY 1/2 ON AT A TIME}
\]

\[
\therefore \text{PWR/AREA} \leq 6 \text{ W/cm}^2
\]

So if you filled a chip with these, you'd get in trouble!

• **AS AN ASIDE, ONLY 2K OF THESE PAIRS WOULD FIT ON A 5mm X 5mm CHIP.**

So how do we get 4K, 16K, etc. memories, but stay within power limits?

> We don't use static cells for such memories. We store charge on capacitors. More next week.

• **Also, if want denser cells and have power problems, there are a variety of circuit tricks that can be used, such as "clocked pullups". More next week.**

**In general, within our dig sys, us. no deep meta**

**IN TEXT, WONT HAVE PROBLEMS. REGIONS OF WIRES AND PASS GATES DON'T DIS. DC POWER, J CONTROL. FOR REGIONS OF DENSERsembler. LEVEL REST CELLS.**

**But if make really dig arrays of 5's, stack cells, etc. calculate PWR dens. TO BE SAFE.**

• **IN ANY EVENT: SCALING DOES NOT MAKE THIS WORSE**