Lecture 10: Wires

Slides courtesy of Deming Chen

Slides based on the initial set from David Harris
Outline

- Introduction
- Interconnect Modeling
  - Wire Resistance
  - Wire Capacitance
- Wire RC Delay
- Crosstalk
- Wire Engineering
- Repeaters

- Readings: 6.1-6.2.2; 6.3.1-6.3.3; 6.4.1-6.4.2
Introduction

- Chips are mostly made of wires called *interconnect*
  - In stick diagram, wires set size
  - Transistors are little things under the wires
  - Many layers of wires
- Wires are as important as transistors
  - Speed
  - Power
  - Noise
- Alternating layers run orthogonally
Wire Geometry

- Pitch = w + s
- Aspect ratio: AR = t/w
  - Old processes had AR << 1
  - Modern processes have AR ≈ 2
    • Pack in many skinny wires
Layer Stack

- AMI 0.6 μm process has 3 metal layers
  - M1 for within-cell routing
  - M2 for vertical routing between cells
  - M3 for horizontal routing between cells
- Modern processes use 6-10+ metal layers
  - M1: thin, narrow (< 3λ)
    - High density cells
  - Mid layers
    - Thicker and wider, (density vs. speed)
  - Top layers: thickest
    - For $V_{DD}$, GND, clk
Example

Intel 90 nm Stack

[Thompson02]

Intel 45 nm Stack

[Moon08]
Interconnect Modeling

- Current in a wire is analogous to current in a pipe
  - Resistance: narrow size impedes flow
  - Capacitance: trough under the leaky pipe must fill first
  - Inductance: paddle wheel inertia opposes changes in flow rate
    - Negligible for most wires
Wires are a distributed system
- Approximate with lumped element models

- 3-segment $\pi$-model is accurate to 3% in simulation
- L-model needs 100 segments for same accuracy!
- Use single segment $\pi$-model for Elmore delay
Wire Resistance

- \( \rho = \text{resistivity} \ (\Omega \cdot \text{m}) \)

- \( R = \)

- \( R_{\square} = \text{sheet resistance} \ (\Omega/\square) \)
  - \( \square \) is a dimensionless unit(!)

- Count number of squares
  - \( R = R_{\square} \times \text{(# of squares)} \)

1 Rectangular Block
\( R = R_{\square} (L/W) \ \Omega \)

4 Rectangular Blocks
\( R = R_{\square} (2L/2W) \ \Omega \)
Choice of Metals

- Until 180 nm generation, most wires were aluminum
- Contemporary processes normally use copper
  - Cu atoms diffuse into silicon and damage FETs
  - Must be surrounded by a diffusion barrier

<table>
<thead>
<tr>
<th>Metal</th>
<th>Bulk resistivity ((\mu\Omega \cdot \text{cm}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silver (Ag)</td>
<td>1.6</td>
</tr>
<tr>
<td>Copper (Cu)</td>
<td>1.7</td>
</tr>
<tr>
<td>Gold (Au)</td>
<td>2.2</td>
</tr>
<tr>
<td>Aluminum (Al)</td>
<td>2.8</td>
</tr>
<tr>
<td>Tungsten (W)</td>
<td>5.3</td>
</tr>
<tr>
<td>Titanium (Ti)</td>
<td>43.0</td>
</tr>
</tbody>
</table>
Contacts Resistance

- Contacts and vias also have 2-20 Ω.
- Use many contacts for lower R.
  - Many small contacts for current crowding around periphery.
Copper Issues

- Copper wires diffusion barrier has high resistance
- Copper is also prone to *dishing* during polishing
- Effective resistance is higher

\[
R = \frac{\rho}{(t - t_{\text{dish}} - t_{\text{barrier}})} \frac{l}{(w - 2t_{\text{barrier}})}
\]
Example

- Compute the sheet resistance of a 0.22 \( \mu \text{m} \) thick Cu wire in a 65 nm process. The resistivity of thin film Cu is 2.2 \( \times 10^{-8} \) \( \Omega \cdot \text{m} \). Ignore dishing.

\[ R_{\square} = \cdot \]

- Find the total resistance if the wire is 0.125 \( \mu \text{m} \) wide and 1 mm long. Ignore the barrier layer.

\[ R = \cdot \]
Wire Capacitance

- Wire has capacitance per unit length
  - To neighbors
  - To layers above and below
- \( C_{\text{total}} = C_{\text{top}} + C_{\text{bot}} + 2C_{\text{adj}} \)
Capacitance Trends

- Parallel plate equation: \( C = \varepsilon_{ox} A/d \)
  - Wires are not parallel plates, but obey trends
  - Increasing area \((W, t)\) increases capacitance
  - Increasing distance \((s, h)\) decreases capacitance

- Dielectric constant
  - \( \varepsilon_{ox} = k\varepsilon_0 \)
    - \( \varepsilon_0 = 8.85 \times 10^{-14} \text{ F/cm} \)
    - \( k = 3.9 \) for SiO\(_2\)

- Processes are starting to use low-\(k\) dielectrics
  - \( k \approx 3 \) (or less) as dielectrics use air pockets
Capacitance Formula

- Capacitance of a line without neighbors can be approximated as
  \[
  C_{tot} = \varepsilon_{ox} l \left[ \frac{w}{h} + 0.77 + 1.06 \left( \frac{w}{h} \right)^{0.25} + 1.06 \left( \frac{t}{h} \right)^{0.5} \right]
  \]

- This empirical formula is accurate to 6% for AR < 3.3
M2 Capacitance Data

- Typical dense wires have ~ 0.2 fF/μm
  - Compare to 1-2 fF/μm for gate capacitance
Diffusion & Polysilicon

- Diffusion capacitance is very high (1-2 fF/μm)
  - Comparable to gate capacitance
  - Diffusion also has high resistance
  - Avoid using diffusion *runners* for wires!
- Polysilicon has lower C but high R
  - Use for transistor gates
  - Occasionally for very short wires between gates
Estimate the delay of a 10x inverter driving a 2x inverter at the end of the 1 mm wire. Assume wire capacitance is 0.2 fF/μm and that a unit-sized inverter has $R = 10 \, \text{KΩ}$ and $C = 0.1 \, \text{fF}$.

$$t_{pd} = (1000 \, \text{Ω})(100 \, \text{fF}) + (1000 + 800 \, \text{Ω})(100 + 0.6 \, \text{fF}) = 281 \, \text{ps}$$
Wire Energy

- Estimate the energy per unit length to send a bit of information (one rising and one falling transition) in a CMOS process.

\[ E = \frac{(0.2 \text{ pF/mm})(1.0 \text{ V})^2}{\text{bit/mm}} = 0.2 \text{ pJ/bit/mm} = 0.2 \text{ mW/Gbps/mm} \]
Crosstalk

- A capacitor does not like to change its voltage instantaneously.
- A wire has high capacitance to its neighbor.
  - When the neighbor switches from 1->0 or 0->1, the wire tends to switch too.
  - Called capacitive coupling or crosstalk.
- Crosstalk effects
  - Noise on nonswitching wires
  - Increased delay on switching wires
Crosstalk Delay

- Assume layers above and below on average are quiet
  - Second terminal of capacitor can be ignored
  - Model as $C_{\text{gnd}} = C_{\text{top}} + C_{\text{bot}}$
- Effective $C_{\text{adj}}$ depends on behavior of neighbors
  - **Miller effect**

<table>
<thead>
<tr>
<th>B</th>
<th>$\Delta V$</th>
<th>$C_{\text{eff}(A)}$</th>
<th>MCF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>$0$</td>
<td>$0$</td>
<td></td>
</tr>
<tr>
<td>Switching with A</td>
<td>$2V_{\text{DD}}$</td>
<td>$C_{\text{gnd}} + C_{\text{adj}}$</td>
<td></td>
</tr>
<tr>
<td>Switching opposite A</td>
<td>$2V_{\text{DD}}$</td>
<td>$C_{\text{gnd}} + C_{\text{adj}}$</td>
<td></td>
</tr>
</tbody>
</table>
Crosstalk Noise

- Crosstalk causes noise on nonswitching wires
- If victim is floating:
  - model as capacitive voltage divider

\[ \Delta V_{\text{victim}} = \frac{C_{\text{adj}}}{C_{\text{gnd-v}} + C_{\text{adj}}} \Delta V_{\text{aggressor}} \]
Driven Victims

- Usually victim is driven by a gate that fights noise
  - Noise depends on relative resistances
  - Victim driver is in linear region, agg. in saturation
  - If sizes are same, $R_{\text{aggressor}} = 2-4 \times R_{\text{victim}}$

\[
\Delta V_{\text{victim}} = \frac{C_{\text{adj}}}{C_{\text{gnd-v}} + C_{\text{adj}}} \frac{1}{1 + k} \Delta V_{\text{aggressor}}
\]

\[
k = \frac{\tau_{\text{aggressor}}}{\tau_{\text{victim}}} = \frac{R_{\text{aggressor}} \left( C_{\text{gnd-a}} + C_{\text{adj}} \right)}{R_{\text{victim}} \left( C_{\text{gnd-v}} + C_{\text{adj}} \right)}
\]
Coupling Waveforms

- Simulated coupling for $C_{\text{adj}} = C_{\text{victim}}$

![Graph showing coupling waveforms with different driver sizes and their effects on the victim signal.](image-url)
Noise Implications

- *So what* if we have noise?
- If the noise is less than the noise margin, nothing happens
- Static CMOS logic will eventually settle to correct output even if disturbed by large noise spikes
  - But glitches cause extra delay
  - Also cause extra power from false transitions
- Dynamic logic never recovers from glitches
- Memories and other sensitive circuits also can produce the wrong answer
Wire Engineering

- Goal: achieve delay, area, power goals with acceptable noise
- Degrees of freedom:
  - Width
  - Spacing
  - Layer
  - Shielding

\[
\text{Delay (ns)}: \frac{RC}{2}
\]

\[
\text{Wire Spacing (nm)}
\]

\[
\text{Coupling: } \frac{2C_{adj}}{2C_{adj} + C_{gnd}}
\]
Repeaters

- R and C are proportional to wire length: $l$
- RC delay is proportional to $l^2$
  - Unacceptably great for long wires
- Break long wires into N shorter segments
  - Drive each one with an inverter or buffer

![Diagram of repeaters](image-url)
Repeater Design

- How many repeaters should we use?
- How large should each one be?
- Equivalent Circuit
  - Wire length $l/N$
    - Wire Capacitance $C_w * l/N$, Resistance $R_w * l/N$
  - Inverter width $W$ (nMOS = $W$, pMOS = $2W$)
    - Gate Capacitance $C' * W$, Resistance $R/W$
Repeater Results

- Write equation for Elmore Delay
  - Differentiate with respect to \( W \) and \( N \)
  - Set equal to 0, solve

\[
\frac{l}{N} = \sqrt{\frac{2RC'}{R_wC_w}}
\]

\[
\frac{t_{pd}}{l} = \left(2 + \sqrt{2}\right) \sqrt{\frac{RC'R_wC_w}{l}}
\]

\[
W = \sqrt{\frac{RC_w}{R_wC'}}
\]

\(~40\ \text{ps/mm}\)

in 65 nm process
Repeater Energy

- Energy / length $\approx 1.87C_w V_{DD}^2$
  - 87% premium over unrepeated wires
  - The extra power is consumed in the large repeaters

- If the repeaters are downsized for minimum EDP:
  - Energy premium is only 30%
  - Delay increases by 14% from min delay
Summary

- For modern chips, wire delay, power consumption, and reliability issues can be a big concern
  - Need accurate modeling and sufficient optimization
  - Hard to have early estimation without a layout
  - Layout-driven synthesis techniques

- Next lecture
  - Adders
  - Readings: 11.1-11.2.2.8