ECE 225
High-Speed Digital IC Design

Lecture 8

Prof. Kaustav Banerjee
Electrical and Computer Engineering

E-mail: kaustav@ece.ucsb.edu
Designing Combinational Logic Circuits
Improved Loads

Adaptive Load
**Improved Loads (2)**

- **Differential Cascode Voltage Switch Logic (DCVSL)**

- **Completely eliminates static currents**
- **Provides rail to rail swing**
- **Differential Gate**: each input is provided in a complementary format
- **Feedback Mechanism**: ensures that the load device is turned off when not needed
- **PDNs are mutually exclusive**: when PDN1 is ON, PDN2 is OFF and vice versa
- **If Out=1, Out=0, PDN1=1 causes Out node to be pulled down with some contention between M1 and PDN1**
- **As Out=Vdd-Vtp, M2 turns on and starts charging Out to Vdd and M1 is turned off enabling Out to discharge completely**
**DCVSL Example (1)**

![XOR-NXOR gate diagram]

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Out=XOR</th>
<th>Out=XNOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: The circuit is still Ratioed!
DCVSL Example (2)

As $Out=Vdd-Vtp$, $Out$ starts to charge up to $Vdd$

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>$Out=\text{NAND}$</th>
<th>$Out=\text{AND}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

$Out=A\cdot B$

$\overline{Out}=A\cdot B$

AND-NAND gate

Note: The circuit is still Ratioed!
Transient Response: AND/NAND
DCVSL Gate

Delay from Input (A, B) to Out (AB) is 197 ps and to Out (A.B) is 321 ps

In comparison a static CMOS AND gate: NAND+INV has a delay of 200 ps
Design Issues with DCVSL Gates

- **Pros:**
  - Provides differential (or complementary outputs)
  - Reduces number of required gates (for complex function) by almost a factor of 2!
  - Number of gates in the critical timing path is often reduced
  - Prevents some time differential problems introduced by the use of additional inverters (when both a signal and its complement are needed simultaneously)

- **Cons:**
  - Increased design complexity
  - Doubles the number of interconnects
  - High dynamic power dissipation: more power dissipation due to cross-over (short-circuit) currents during the transition
The Wire

schematics

physical
Interconnect delay has become the dominant factor determining chip performance……..
IC performance is being dominated by interconnects....

Modern Interconnect
Impact of Interconnect Parasitics

- Interconnect parasitics
  - affect performance and power consumption
  - affect reliability

- Classes of parasitics
  - Capacitive
  - Resistive
  - Inductive
**Wire Resistance**

\[ R = \frac{\rho L}{H W} \]

Sheet Resistance \( R_0 \)

\[ R_1 \equiv R_2 \]
## Interconnect Resistance

<table>
<thead>
<tr>
<th>Material</th>
<th>$\rho$ ((\Omega\cdot\text{m}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silver (Ag)</td>
<td>$1.6 \times 10^{-8}$</td>
</tr>
<tr>
<td>Copper (Cu)</td>
<td>$1.7 \times 10^{-8}$</td>
</tr>
<tr>
<td>Gold (Au)</td>
<td>$2.2 \times 10^{-8}$</td>
</tr>
<tr>
<td>Aluminum (Al)</td>
<td>$2.7 \times 10^{-8}$</td>
</tr>
<tr>
<td>Tungsten (W)</td>
<td>$5.5 \times 10^{-8}$</td>
</tr>
</tbody>
</table>
Dealing with Resistance

- Selective Technology Scaling
- Use Better Interconnect Materials
  - reduce average wire-length
  - e.g. copper, silicides
- More Interconnect Layers
  - reduce average wire-length
## Sheet Resistance

<table>
<thead>
<tr>
<th>Material</th>
<th>Sheet Resistance (Ω/□)</th>
</tr>
</thead>
<tbody>
<tr>
<td>n- or p-well diffusion</td>
<td>1000 – 1500</td>
</tr>
<tr>
<td>$n^+, p^+$ diffusion</td>
<td>50 – 150</td>
</tr>
<tr>
<td>$n^+, p^+$ diffusion with silicide</td>
<td>3 – 5</td>
</tr>
<tr>
<td>$n^+, p^+$ polysilicon</td>
<td>150 – 200</td>
</tr>
<tr>
<td>$n^+, p^+$ polysilicon with silicide</td>
<td>4 – 5</td>
</tr>
<tr>
<td>Aluminum</td>
<td>0.05 – 0.1</td>
</tr>
</tbody>
</table>
Example: Intel 0.25 micron Process

5 metal layers
Ti/Al - Cu/Ti/TiN
Polysilicon dielectric

<table>
<thead>
<tr>
<th>LAYER</th>
<th>PITCH</th>
<th>THICK</th>
<th>A.R.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isolation</td>
<td>0.67</td>
<td>0.40</td>
<td>-</td>
</tr>
<tr>
<td>Polysilicon</td>
<td>0.64</td>
<td>0.25</td>
<td>-</td>
</tr>
<tr>
<td>Metal 1</td>
<td>0.64</td>
<td>0.48</td>
<td>1.5</td>
</tr>
<tr>
<td>Metal 2</td>
<td>0.93</td>
<td>0.90</td>
<td>1.9</td>
</tr>
<tr>
<td>Metal 3</td>
<td>0.93</td>
<td>0.90</td>
<td>1.9</td>
</tr>
<tr>
<td>Metal 4</td>
<td>1.60</td>
<td>1.33</td>
<td>1.7</td>
</tr>
<tr>
<td>Metal 5</td>
<td>2.56</td>
<td>1.90</td>
<td>1.5</td>
</tr>
</tbody>
</table>

Layer pitch, thickness and aspect ratio $=h/w$
Interconnect (RC) Delay

On-Chip VLSI interconnects can be modeled as RC elements

• R is the wire resistance =

\[ \rho \frac{L}{A} = \rho \frac{L}{w \cdot h} \]

- \( \rho \) is the resistivity of the metal
- \( L \) is the wire length
- \( A \) is the cross sectional area = \( wh \) (\( w \) is the width and \( h \) is the height of the wire)

• C is the wire capacitance. For a parallel plate capacitor

\[ C = \varepsilon_d \frac{A}{d}, \ \varepsilon_d = k \varepsilon_0 \]

- \( A \) is the area of the plate
- \( d \) is the distance between the plates
- \( k \) is the dielectric constant of the insulating material between the plates
- \( \varepsilon_0 \) is the permittivity of free space
If we can reduce both $R$ and $C$, we can reduce wire delay.....

Will better materials like copper and low-k dielectrics solve the interconnect problem?

Cu has lower resistivity than Al, and is more robust (reliable) than Al.......
Changing Interconnect Materials

- Replace Al wires by Cu wires
- Resistivity of Al = 2.65 $\mu\Omega$-cm
  @ Room Temp. (20-25 °C)
- Resistivity of Cu = 1.67 $\mu\Omega$-cm
  @ Room Temp. (20-25 °C)
- Reduction in R is not even a factor of 2.....
Frequency Dependence of $R$

- At high enough frequencies, $R$ can increase due to **skin effect**
- Current flow constrained in a thin layer along the surface of the conductor: $R$ increases
- The current falls off exponentially with depth into the interconnect
- **Skin depth**, $\delta$ is defined as the depth at which current falls off to a value of $1/e$ of its nominal value

$$\delta = \sqrt{\frac{\rho}{\pi f \mu}}$$

$\mu$ is the permeability of the surrounding dielectric

Resistance per unit length:

$$r(f) = \frac{\sqrt{\pi f \mu \rho}}{2(H+W)}, \text{ for } W, H >> \delta$$
Skin Effect

- Onset of skin effect: at $f = f_s$, $\delta = 1/2 \times$ smallest dimension

\[ f_s = \frac{4\rho}{\pi \mu (\min(W,H))^2} \]

- For Cu wires:

<table>
<thead>
<tr>
<th>Freq</th>
<th>1 GHZ</th>
<th>5 GHZ</th>
<th>10 GHZ</th>
<th>15 GHZ</th>
<th>20 GHZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skin depth ($\mu m$)</td>
<td>2.08</td>
<td>0.93</td>
<td>0.66</td>
<td>0.53</td>
<td>0.46</td>
</tr>
</tbody>
</table>
Capacitance: The Parallel Plate Model

Valid when, $W \gg t_{di}$

$$C_{int} = \frac{\varepsilon_{di}}{t_{di}} WL$$
## Permittivity

<table>
<thead>
<tr>
<th>Material</th>
<th>( \varepsilon_r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free space</td>
<td>1</td>
</tr>
<tr>
<td>Aerogels</td>
<td>( \sim1.5 )</td>
</tr>
<tr>
<td>Polyimides (organic)</td>
<td>3-4</td>
</tr>
<tr>
<td>Silicon dioxide</td>
<td>3.9</td>
</tr>
<tr>
<td>Glass-epoxy (PC board)</td>
<td>5</td>
</tr>
<tr>
<td>Silicon Nitride (( \text{Si}_3\text{N}_4 ))</td>
<td>7.5</td>
</tr>
<tr>
<td>Alumina (package)</td>
<td>9.5</td>
</tr>
<tr>
<td>Silicon</td>
<td>11.7</td>
</tr>
</tbody>
</table>
Fringing Capacitance Model

\[ c_{wire} = c_{pp} + c_{fringe} = \frac{w \varepsilon_{di}}{t_{di}} + \frac{2 \pi \varepsilon_{di}}{\log(t_{di}/H)} \]

- Fringing Cap.
- Parallel plate Cap.
Fringing versus Parallel Plate

Saturates to \(~1\text{pF/cm due to fringing fields}\)

(from [Bakoglu89])
Adjacent wires are orthogonal....why?

$C_{\text{intra}}$ is the dominating capacitance......why?
Usually interconnect layers alternate wire direction
Impact of Interwire Capacitance

For a fixed dielectric and wire thickness

- Reduce wire width, w
- Reduce spacing (=w)
- $C_{total} = C_{ground} + C_{interwire}$
- $C_{ground} = C_{p-p} + C_{fringing}$

Interwire cap. becomes important in a multi-level structure

(from [Bakoglu89])

$W/H = 1.75$
## Wiring Capacitances (0.25 μm CMOS)

<table>
<thead>
<tr>
<th></th>
<th>Field</th>
<th>Active</th>
<th>Poly</th>
<th>Al1</th>
<th>Al2</th>
<th>Al3</th>
<th>Al4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poly</td>
<td>88</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al1</td>
<td>30</td>
<td>41</td>
<td>57</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>47</td>
<td>54</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al2</td>
<td>13</td>
<td>15</td>
<td>17</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>27</td>
<td>29</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al3</td>
<td>8.9</td>
<td>9.4</td>
<td>10</td>
<td>15</td>
<td>41</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>27</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al4</td>
<td>6.5</td>
<td>6.8</td>
<td>7</td>
<td>8.9</td>
<td>15</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>15</td>
<td>18</td>
<td>27</td>
<td>45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al5</td>
<td>5.2</td>
<td>5.4</td>
<td>5.4</td>
<td>6.6</td>
<td>9.1</td>
<td>14</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>14</td>
<td>19</td>
<td>27</td>
<td>52</td>
</tr>
</tbody>
</table>

Table rows represent the top plate of the capacitor and the columns represent the bottom plate.

The wire at level 5 is twice as thick and is embedded in a dielectric of higher permittivity.
Interconnect Modeling
Capacitances of Driver-Wire-Load

Simplified Model
Wire Models

All-inclusive model

Capacitance-only
The Lumped Capacitor Model

Assuming no voltage drops along the wire…
i.e., wire is equipotential

If \( r \) is small and switching frequencies are not too high…

Advantage: Operation can be described by ordinary differential equation
The Lumped Capacitor Model

Operation of a simple RC circuit:

\[ C_{lumped} \frac{dV_{out}}{dt} + \frac{V_{out} - V_{in}}{R_{driver}} = 0 \]

If a step input is applied from 0 to \( V \):

\[ V_{out}(t) = \left(1 - e^{-t/\tau}\right)V \]

where \( \tau = R_{driver} C_{lumped} \): time constant of network

The time to reach 50% point can be calculated as:

\[ t = \ln(2) \tau = 0.69 \tau \]

If \( R_{driver} = 10K\Omega \) and \( C_{lumped} = 11 \text{ pF} \):

\[ t = 76 \text{ ns} \ (a \ very \ large \ number \ for \ high - performance \ digital \ circuits) \]

Driver modeled as a voltage source and a source resistance \( (R_{driver}) \), \( C_{lumped} = L \times C_{wire} \)

Only impact on performance is due to the loading effect of the capacitor on the driving gate.
The Lumped RC-Model

- On-chip metal wires at the semi-global and global tiers can be several millimeters long and can have significant resistance
- The equipotential assumption is no longer valid
- Need an RC model
- A simple approach: lump wire resistance of each segment into a single R and the global capacitance into a single C
  - This is called a lumped RC model....pessimistic and inaccurate for long wires
  - Should use a distributed RC model for such long wires

Then what is the use of the lumped RC model?
The Lumped RC-Model

• The distributed RC model is complex: involves partial differential equations, closed form solution does not exist

• However, the behavior of a distributed RC line can be adequately modeled by a simple RC network

• It is more convenient to reduce any driver-wire-load networks to an equivalent RC network and predict its first order response

• However, unlike the single R-C network analyzed earlier that had a single time constant (network pole), deriving the correct waveforms for a complex network becomes intractable

  • Need to solve a set of ordinary differential equations with many time constants (poles and zeroes)….

  • Or, run time-consuming SPICE simulations for the entire network….
Elmore Delay Comes to the Rescue

**Shared path resistance**: resistance shared among the paths from root node s to nodes k and i

\[ R_{ik} = \sum R_j \implies (R_j \in [\text{path}(s \rightarrow i) \cap \text{path}(s \rightarrow k)]) \]

*Example*: \( R_{i4} = R_1 + R_3 \)

\( R_{i2} = R_1 \)

**RC Tree**: unique resistive paths between s and any node i

Total resistance along this path is called the *path resistance* \( R_{ii} \)

E.g., path resistance between s and node 4:

\[ R_{44} = R_1 + R_3 + R_4 \]
If a step input $V_{in}$ is applied at $t=0$

**Elmore Delay at node $i$:**

$(First\text{-}order\ time\ constant\ of\ the\ network)$

$$\tau_{Di} = \sum_{k=1}^{N} C_{k}R_{ik}$$

$N = \#\ of\ capacitances\ in\ the\ network$
Wire Model

For a non-branched RC chain: \( \tau_{DN} = \sum_{i=1}^{N} C_i R_{ii} \)  
\( t_{Di} = C_i R_1 + C_2 (R_1+R_2)+\ldots+C_i (R_1+R_2+\ldots+R_i) \)

If wire modeled by N equal-length segments \( (R_{seg} = rL/N, \ C_{seg} = cL/N) \)

\[ \tau_{DN} = \left( \frac{L}{N} \right)^2 (r c + 2 r c + \ldots + N r c) = (r c L^2) \frac{N(N+1)}{2N^2} = RC \frac{N+1}{2N} \]

For large values of N:

\[ \tau_{DN} = \frac{RC}{2} = \frac{r c L^2}{2} \]
The **Distributed RC-line**

*(can be approximated by the lumped RC network model)*

\[ rC \frac{\partial V}{\partial t} = \frac{2}{\partial x^2} \]

\[ \tau(V_{out}) = \frac{rcL^2}{2} \]
Step-response of RC wire as a function of time and space

Voltage diffuses (degrades) along the RC wire
**RC-Models**

SPICE Wire (RC) Models:

*accuracy of the model determined by number of stages*

<table>
<thead>
<tr>
<th>Voltage Range</th>
<th>Lumped RC-network</th>
<th>Distributed RC-network</th>
</tr>
</thead>
<tbody>
<tr>
<td>0→50% ($t_p$)</td>
<td>0.69 RC</td>
<td>0.38 RC</td>
</tr>
<tr>
<td>0→63% ($t$)</td>
<td>RC</td>
<td>0.5 RC</td>
</tr>
<tr>
<td>10%→90% ($t_r$)</td>
<td>2.2 RC</td>
<td>0.9 RC</td>
</tr>
</tbody>
</table>

Step Response of Lumped and Distributed RC Networks:
Points of Interest.
Design Rules of Thumb

- rc delays should only be considered when
  \( t_{pRC} \gg t_{pgate} \) of the driving gate
  \[ L_{crit} \gg \sqrt{t_{pgate}/0.38rc} \]

- rc delays should only be considered when the rise (fall) time at the line input is smaller than RC, the rise (fall) time of the line
  \[ t_{rise} < RC \]
  - when not met, the change in the signal is slower than the propagation delay of the wire
Illustration of Rule: Driving an RC-line

\[ \tau_D = R_s C_w + \frac{R_w C_w}{2} = R_s C_w + 0.5 r_w c_w L^2 \]

\[ R_w = r_w L \]
\[ C_w = c_w L \]

\[ t_p = 0.69 R_s C_w + 0.38 R_w C_w \]

Delay due to wire resistance becomes important when
\[ (R_w C_w/2) \geq R_s C_w \text{ or when } L \geq 2R_s/r_w \]
On-Chip Inductance

- \( L \) starts to play a role at high GHz frequencies
- Effect increases for Cu: lower resistivity
- Can increase interconnect delay
- Can cause overshoot and undershoot effects
- Can cause switching noise: \( \text{Ldi/dt} \) voltage drops

\[
Z = R + j \omega L
\]

inductance of a wire surrounded by a uniform dielectric:

\[
c \ell = \varepsilon \mu
\]

- \( c \) = cap. per unit length
- \( I \) = inductance per unit length

*From Maxwell’s Laws:*

\[
v = \frac{1}{\sqrt{c_\ell}} = \frac{1}{\sqrt{\varepsilon \mu}} = \frac{c_0}{\sqrt{\varepsilon_r \mu_r}}
\]

- \( v \) = propagation speed of EM wave
- \( c_0 \) = speed of light in vacuum = 30 cm/ns