Switching (AC) Characteristics of MOS Inverters

Prof. MacDonald
MOS Inverters

- Performance is inversely proportional to delay
- Delay is time to raise (lower) voltage at nodes
  - node voltage is changed by charging (discharging) load cap
  - more current means more charge transported over time

\[ Q = I \cdot t = C \cdot V \]

\[ t_{\text{delay}} = \frac{Q}{I} = \frac{C \cdot V}{I} \]
MOS Inverters

wire cap particularly bad when driving a load far away.
MOS Inverters

Lumped cap

\[ C_L = C_{gdn} + C_{gdp} + C_{dbn} + C_{dbp} + C_w + C_g \]
MOS Inverters – discharge delay

\[ C_L = C_{gdn} + C_{gdp} + C_{dbn} + C_{dbp} + C_w + C_g \]
MOS Inverters – charge delay

Lumped cap
\[ C_L = C_{gd} + C_{gp} + C_{db} + C_{dp} + C_w + C_g \]

0V

\[ \text{time} \]
Defined twice – once for a falling output and once for a rising output. The propagation delay is the delay from the input crossing the 50% point of Vdd to the resulting output signal crossing of the 50% point. $T_{plh} = $ Rising propagation delay $T_{phl} = $ Falling propagation delay
The rise time is the time for the signal to cross from 10% to 90% of Vdd. The fall time is the time for the signal to cross from 90% to 10% of Vdd.

If an inverter is driven by a signal with a really slow rise or fall time, the delay through the inverter is aggravated and since the inverter is in the transition region longer, a lot of short circuit current can be generated.
If excessive rise or fall times exist, fix them by cranking up drive source or decreasing the load.

Increasing drive strength usually means widening transistors.

Decreasing the load usually means splitting up load with buffers.
Simplest approach is to use average current and average capacitance models to calculate propagation delays for both edges.

\[ \tau_{plh} = \frac{C_{load} \cdot \Delta V_{hl}}{I_{avghl}} \]

\[ \tau_{phl} = \frac{C_{load} \cdot \Delta V_{lh}}{I_{avghl}} \]
MOS Inverters – fall delay

\[ V_{\text{out}}(t) = V_{dd} \cdot e^{-\frac{t}{R_nC_l}} \]
MOS Inverters – rise delay

\[ V_{out}(t) = V_{dd} \cdot \left(1 - e^{-\frac{t}{R_pC_l}}\right) \]
Combating delays

- **Reduce Capacitive load**
  - drive fewer gates – buffer tree
  - drive smaller gates (less gate capacitance) in subsequent stage
  - drive closer gates (less distance means less interconnect load)

- **Increase Drive current**
  - reduce Vt – not really an option for circuit designers
  - reduce L’ s – most transistors are minimum sized for area
  - increase V_{dd} – can’t because of gate oxide integrity
  - increase W_{eff} – main weapon of circuit designer

- **Reduce wire lengths for long wires (more later…)**
Delay vs. Width

- Diminishing returns because of increased junction cap with larger transistors.
- Also will add to load of previous stage and slow total circuit path.
Power*Delay vs Width
Interconnect Delays

- As technology scales...
  - devices tend to get faster
  - interconnects tend to get slower
- Resistance goes up with each shrink
  - motivation for new metals (aluminum to copper transition)
- Capacitance goes up with each shrink
  - motivation for low K dielectrics
  - cross-coupling between parallel-running signals
  - slower
  - noisy
- Inductance is generally ignored for on-chip simulation
Wire dimensions

Substrate (ground plane)
Parasitic capacitance

Capacitance per unit length of wire to supply planes or other fixed or non-active signals

Low K dielectric helps to reduce this cap.

Substrate (ground plane)
Cross-coupling capacitance

As spacing between lines decreases, coupling cap between the signals increases.

Not a big problem from one level to another because lines run orthogonal (i.e. metal 1 and metal 2 signals).

However, for lines on same metal that run long distances in parallel, this can cause significant problems and is the subject of current research efforts on design automation.

Substrate (ground plane)
Cross-coupling capacitance

Consider a simple 3 bit bus running long distance.

The first impression is that the coupling cap seen by the inner line is 2X because of the “sandwich” effect.

Now consider if the bus carried the value 3’b010 and then switched the next cycle to 3’b101. The voltage swing relative to the inner line would be 2 x Vdd so the effective capacitance would not 2x but 4x greater.
Cross Coupling Capacitance

Coupling noise spike

Noise induced delay
Capacitor Divider Review
Coupling Analysis

\[ V_{\text{victim}} = \frac{C_{\text{coupling}} \cdot V_{dd}}{C_{\text{coupling}} + C_{\text{good}}} \]
Minimizing Coupling Capacitance

- Wire spreaders are tools that search through a routed design and find places where signals can be spread.
- Noise sensitive signals (i.e. clock signal) can be shielded by running fixed signals (i.e. gnd, vdd) between clock and other signals.
- Technologies are being developed that raise the permittivity of the inter layer dielectric.
  - problems persist with this new materials
  - thermal cycling the material causes ruptures due to differences in the thermal expansion coefficient.
Wire Spreading Example

Before

After
Shielding Signals

Coupling capacitance goes down with a 1/T relationship.

Good cap goes up because of shielding.

Substrate (ground plane)
Resistance Estimation

\[ R_{\text{wire}} = \rho \cdot \frac{l}{w \cdot t} = R_{\text{sheet}} \cdot \left( \frac{l}{w} \right) \Omega \]
# Typical Sheet Resistances and Resistivities

<table>
<thead>
<tr>
<th>Material</th>
<th>Resistivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silver (Ag)</td>
<td>1.6x10^{-8}</td>
</tr>
<tr>
<td>Copper (Cu)</td>
<td>1.7x10^{-8}</td>
</tr>
<tr>
<td>Gold (Au)</td>
<td>2.2x10^{-8}</td>
</tr>
<tr>
<td>Aluminum (Al)</td>
<td>2.7x10^{-8}</td>
</tr>
<tr>
<td>Tungsten (W)</td>
<td>5.5x10^{-8}</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Material</th>
<th>Sheet Resistance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well</td>
<td>1000</td>
</tr>
<tr>
<td>Drain/Source</td>
<td>100</td>
</tr>
<tr>
<td>Drain/Source with silicide</td>
<td>10</td>
</tr>
<tr>
<td>Poly</td>
<td>100</td>
</tr>
<tr>
<td>Poly with silicide</td>
<td>5</td>
</tr>
<tr>
<td>Aluminum</td>
<td>0.1</td>
</tr>
</tbody>
</table>
Resistance Effects

\[ V_{out}(t) = V_{dd} \cdot e^{-\frac{t}{(R_{eq} + R_{int})C_l}} \]

<table>
<thead>
<tr>
<th>Req</th>
<th>Distance</th>
<th>Rint</th>
<th>Rtotal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>10</td>
<td>~0</td>
<td>1000</td>
</tr>
<tr>
<td>1000</td>
<td>100</td>
<td>20</td>
<td>1020</td>
</tr>
<tr>
<td>1000</td>
<td>1000</td>
<td>200</td>
<td>1200</td>
</tr>
<tr>
<td>1000</td>
<td>10000</td>
<td>2000</td>
<td>3000</td>
</tr>
<tr>
<td>500</td>
<td>10000</td>
<td>2000</td>
<td>2500</td>
</tr>
</tbody>
</table>
Long Lines and RC Delays

Interconnect RC Delay through a metal line is quadratically related to length. Consequently, it is basically non-existent for lengths less than 100 μ (local region), but grows quickly.

\[ \tau \approx (R \cdot L) \cdot (C \cdot L) = R \cdot C \cdot L^2 \]

\[ L_{crit} = \sqrt{\frac{FO4}{0.38 \cdot R \cdot C}} \]
Long Lines and RC Delays

Buffer can cut down on L and decrease interconnect delay quadratically – of course device delay is inserted but many times the overall delay goes down.

100ps

400ps

100ps

600 pS total
Long Lines and RC Delays

If distance $L$ has 400ps of RC delay, then a distance of $L/2$ will have 100ps of delay - $(L/2)^2$ or $\frac{1}{4}$ of the delay.

550 pS total
Long Lines and RC Delays

If distance $L$ has 400ps of RC delay, then a distance of $L/3$ will have 45ps of delay - $(L/3)^2$ or $1/9$ of the delay.

535 pS total
Note on RC delays and $V_{dd}$

RC values are not affected by $V_{dd}$ values to the first order.

Device delay however is related by the square of the voltage.

\[
\begin{array}{cccccccc}
100\text{ps} & 45\text{ps} & 100\text{ps} & 45\text{ps} & 100\text{ps} & 45\text{ps} & 100\text{ps} \\
V_{dd} = 1.8\text{V} \\
\end{array}
\]

\[
\begin{array}{cccccccc}
400\text{ps} & 45\text{ps} & 400\text{ps} & 45\text{ps} & 400\text{ps} & 45\text{ps} & 400\text{ps} \\
V_{dd} = 0.9\text{V} \\
\end{array}
\]
Interconnect models

First compare time of flight to rise times. If flight isn’t 5x smaller than rise time, you need a sophisticated model like a transmission-line model. Long lines across the chip have long time of flights.

Otherwise, you can get away with using a lumped RC model. Lumped models are normally all that is needed for CMOS.

For 100 microns, \( t = \frac{D}{V} = \frac{1}{1000} / (3\times10^8) = 3 \text{ pS} \)

For 2 cm (across chip) \( t = \frac{D}{V} = 0.02 / 3\times10^8 = 60 \text{ pS} \)

typical rise times will be from 50 to 500 pS
Interconnect models

In this example the time of flight is shorter than rise/fall time.

Most typical example on chips and only requires lumped RC model.
Inverter sizing and Fanout

To drive a huge load with a small inverter we need a string of inverters to “ramp up” the capacitive gain.

If inverter is too small, will have difficult time charging next stage. If inverter is too large, it will overload the previous inverter.

<table>
<thead>
<tr>
<th>Wp</th>
<th>4</th>
<th>12</th>
<th>36</th>
<th>108</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wn</td>
<td>2</td>
<td>6</td>
<td>18</td>
<td>54</td>
</tr>
</tbody>
</table>

Case of huge load (i.e. IO driving off chip loads or clock tree driving 1000s of flip-flops)