5. CMOS Gates: DC and Transient Behavior

J. A. Abraham

Department of Electrical and Computer Engineering
The University of Texas at Austin

VLSI Design
Fall 2015

September 14, 2015
Topics

- DC Response
- Logic Levels and Noise Margins
- Transient Response
- Delay Estimation
Transistor Behavior

Behavior in different situations (increase, decrease, or not change).

1. If the width of a transistor increases, the current will
2. If the length of a transistor increases, the current will
3. If the supply voltage of a chip increases, the maximum transistor current will
4. If the width of a transistor increases, its gate capacitance will
5. If the length of a transistor increases, its gate capacitance will
6. If the supply voltage of a chip increases, the gate capacitance of each transistor will
Behavior in different situations (increase, decrease, or not change).

1. If the width of a transistor increases, the current will increase.
2. If the length of a transistor increases, the current will decrease.
3. If the supply voltage of a chip increases, the maximum transistor current will increase.
4. If the width of a transistor increases, its gate capacitance will increase.
5. If the length of a transistor increases, its gate capacitance will increase.
6. If the supply voltage of a chip increases, the gate capacitance of each transistor will not change.
**DC Response: \( V_{out} \) vs. \( V_{in} \) for a Gate**

Study the response of Inverters

- When \( V_{in} = 0 \) \( \Rightarrow \) \( V_{out} = V_{DD} \)
- When \( V_{in} = V_{DD} \) \( \Rightarrow \) \( V_{out} = 0 \)
- In between, \( V_{out} \) depends on transistor size and current
- By KCL, current must be such that \( I_{dsn} = |I_{dsp}| \)
- We could solve equations, but graphical solution gives more insight

![Schematic of an inverter circuit](image)
Current through transistor depends on the region of operation

- Need to identify for what $V_{in}$ and $V_{out}$ are nMOS and pMOS in Cutoff, Linear or Saturation

### nMOS Operation

<table>
<thead>
<tr>
<th>Cutoff</th>
<th>Linear</th>
<th>Saturated</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{gsn} &lt; V_{tn}$</td>
<td>$V_{gsn} &gt; V_{tn}$</td>
<td>$V_{gsn} &gt; V_{tn}$</td>
</tr>
<tr>
<td>$V_{in} &lt; V_{tn}$</td>
<td>$V_{in} &gt; V_{tn}$</td>
<td>$V_{in} &gt; V_{tn}$</td>
</tr>
<tr>
<td></td>
<td>$V_{dsn} &lt; V_{gsn} - V_{tn}$</td>
<td>$V_{dsn} &gt; V_{gsn} - V_{tn}$</td>
</tr>
<tr>
<td></td>
<td>$V_{out} &lt; V_{in} - V_{tn}$</td>
<td>$V_{out} &gt; V_{in} - V_{tn}$</td>
</tr>
</tbody>
</table>

\[
V_{gsn} = V_{in} \\
V_{dsn} = V_{out}
\]
## pMOS Operation

<table>
<thead>
<tr>
<th>Cutoff</th>
<th>Linear</th>
<th>Saturated</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{gsp} &gt; V_{tp}$</td>
<td>$V_{gsp} &lt; V_{tp}$</td>
<td>$V_{gsp} &lt; V_{tp}$</td>
</tr>
<tr>
<td>$V_{in} &gt; V_{DD} + V_{tp}$</td>
<td>$V_{in} &lt; V_{DD} + V_{tp}$</td>
<td>$V_{in} &lt; V_{DD} + V_{tp}$</td>
</tr>
<tr>
<td>$V_{dsp} &gt; V_{gsp} - V_{tp}$</td>
<td>$V_{dsp} &lt; V_{gsp} - V_{tp}$</td>
<td>$V_{dsp} &lt; V_{gsp} - V_{tp}$</td>
</tr>
<tr>
<td>$V_{out} &gt; V_{in} - V_{tp}$</td>
<td>$V_{out} &lt; V_{in} - V_{tp}$</td>
<td>$V_{out} &lt; V_{in} - V_{tp}$</td>
</tr>
</tbody>
</table>

\[
V_{gsp} = V_{in} - V_{DD}
\]

\[
V_{dsp} = V_{out} - V_{DD}
\]

\[
V_{tp} < 0
\]
Make pMOS wider than nMOS such that $\beta_n = \beta_p$

$$\beta = \mu C_{ox} \frac{W}{L}$$
Current vs. $V_{out}$, $V_{in}$
Load Line Analysis

To find the $V_{out}$ for a given $V_{in}$

- For a given $V_{in}$, plot $I_{dsn}$, $I_{dsp}$ vs. $V_{out}$
- $V_{out}$ must be where $|\text{currents}|$ are equal in the graph below
Transcribe points on to $V_{in}$ vs. $V_{out}$ plot
## Operating Regions

Revisit transistor operating regions

<table>
<thead>
<tr>
<th>Region</th>
<th>nMOS</th>
<th>pMOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Cutoff</td>
<td>Linear</td>
</tr>
<tr>
<td>B</td>
<td>Saturation</td>
<td>Linear</td>
</tr>
<tr>
<td>C</td>
<td>Saturation</td>
<td>Saturation</td>
</tr>
<tr>
<td>D</td>
<td>Linear</td>
<td>Saturation</td>
</tr>
<tr>
<td>E</td>
<td>Linear</td>
<td>Cutoff</td>
</tr>
</tbody>
</table>

![Graph showing transistor operating regions](image-url)
If \( \frac{\beta_p}{\beta_n} \neq 1 \), switching point will move from \( V_{DD}/2 \)

Called **skewed** gate

Analysis of more complex gates

- **Collapse into equivalent inverter**
Noise Margins

How much noise can a gate input see before it does not recognize the input?
Logic Levels

To maximize noise margins

- Select logic levels at unity gain point of DC transfer characteristic


- **DC analysis** gives the $V_{out}$ if $V_{in}$ is constant.
- **Transient analysis** tells us $V_{out}$ as $V_{in}$ changes.
- Input is usually considered to be a step or ramp (from 0 to $V_{DD}$ or vice-versa).

![Transient Response Graph](image.png)
Inverter Step Response

Find the step response of an inverter driving a load capacitance

- $V_{in}(t) = u(t - t_0)V_{DD}$
- $V_{out}(t < t_0) = V_{DD}$
- $\frac{dV_{out}(t)}{dt} = -\frac{I_{dsn}(t)}{C_{load}}$

$I_{dsn}(t) = \begin{cases} 
0 & t \leq t_0 \\
\frac{\beta}{2}(V_{DD} - V)^2 & V_{out} > V_{DD} - V_t \\
\beta \left( V_{DD} - V_t - \frac{V_{out}(t)}{2} \right) V_{out}(t) & V_{out} < V_{DD} - V_t
\end{cases}$
Delay Definitions

- $t_{pdr}$: rising propagation delay
  - From input to rising output crossing $V_{DD}/2$

- $t_{pdf}$: falling propagation delay
  - From input to falling output crossing $V_{DD}/2$

- $t_{pd}$: average propagation delay
  - $t_{pd} = (t_{pdr} + t_{pdf})/2$

- $t_r$: rise time
  - From output crossing $0.2 \ V_{DD}$ to $0.8 \ V_{DD}$

- $t_f$: fall time
  - From output crossing $0.8 \ V_{DD}$ to $0.2 \ V_{DD}$

- $t_{cdr}$: rising contamination delay
  - Minimum time from input to rising output crossing $V_{DD}/2$

- $t_{cdf}$: falling contamination delay
  - Minimum time from input to falling output crossing $V_{DD}/2$

- $t_{cd}$: average contamination delay
  - $t_{cd} = (t_{cdr} + t_{cdf})/2$
Simulated Inverter Delay

- Solving differential equations by hand too hard
- SPICE simulator solves equations numerically
  - Uses more accurate I-V models too!
- But simulations take time to write
We would like to be able to easily estimate delay
- Not as accurate as simulation
- But easier to ask “what if ...”? 

The step response usually looks like a first order RC response with a decaying exponential

Use RC delay models to estimate delay
- $C =$ total capacitance on output node
- Use effective resistance $R$
- So that $t_{pd} = RC$

Characterize transistors by finding their effective $R$
- Depends on average current as gate switches
Example: Sizing 3-Input NAND Gate for Equal Rise and Fall Times

Determine the transistor widths to achieve effective rise and fall resistances (times) equal to that of a unit inverter $R$

Annotate the 3-input NAND gate with gate and diffusion capacitances
Determine the transistor widths to achieve effective rise and fall resistances (times) equal to that of a unit inverter $R$

Annotate the 3-input NAND gate with gate and diffusion capacitances
Example: Sizing Complex Gate

Size the transistors in the circuit below so that it has the same drive strength, in the worst case, as an inverter that has $PW = 5$ and $NW = 3$. Use the smallest widths possible to achieve this ratio.

Note: if there are multiple paths through a transistor, use the size for the “worst-case” input combination.
Example: Sizing Complex Gate

Size the transistors in the circuit below so that it has the same drive strength, in the worst case, as an inverter that has PW = 5 and NW = 3. Use the smallest widths possible to achieve this ratio.

This solution does NOT use the smallest widths.

Note: if there are multiple paths through a transistor, use the size for the “worst-case” input combination.
Example: Sizing of Complex Gate – Better Solution

Size the transistors in the circuit below so that it has the same drive strength, in the worst case, as an inverter that has PW = 5 and NW = 3. Use the smallest widths possible to achieve this ratio.

Note: if there are multiple paths through a transistor, use the size for the “worst-case” input combination.
Elmore Delay

Finding the delay of ladder networks

- ON transistors look like resistors
- Pullup or pulldown network modeled as RC ladder
- Elmore delay of RC ladder

\[ t_{pd} = \sum_{\text{nodes } i} R_{i \text{-to-source}} C_i \]

\[ = R_1 C_1 + (R_1 + R_2) C_2 + \ldots + (R_1 + R_2 + \ldots + R_N) C_N \]

NOTE: \( C_i \) includes all the “off-path” capacitance on nodes that are connected to node \( i \)
Example: Delay of 2-Input NAND Using Elmore Formulation

Estimate **rising** and falling propagation delays of a 2-input NAND driving \( h \) identical gates

\[
t_{pdr} = (6 + 4h)RC
\]
Example: Delay of 2-Input NAND Using Elmore Formulation

Estimate rising and **falling** propagation delays of a 2-input NAND driving \( h \) identical gates

\[
t_{pdf} = \left(2C\right) \frac{R}{2} + \left[(6 + 4h)C\right] \left(\frac{R}{2} + \frac{R}{2}\right) = (7 + 4h)RC
\]
Calculate the Elmore delay from C to F in the circuit. The widths of the pass transistors are shown, and the inverters have minimum-sized transistors.
Example of Elmore Delay Calculation

Calculate the Elmore delay from C to F in the circuit. The widths of the pass transistors are shown, and the inverters have minimum-sized transistors.

\[
\text{Delay} = \frac{R}{3} 9C + \frac{R}{3} 5C + \left( \frac{R}{3} + \frac{R}{3} \right) 7C + 3RC = 12.33RC
\]
Another Example: Elmore Delay Calculation

Use the Elmore delay approximation to find the *worst-case* rise and fall delays at output F for the following circuit. The gate sizes of the transistors are shown in the figure. Assume NO sharing of diffusion regions, and the worst-case conditions for the initial charge on a node.

Input for worst-case rise delay =

Worst-case rise delay =

Input for worst-case fall delay =

Worst-case fall delay =
Delay with Different Input Sequences

Find the delays for the given input transitions (gate sizes shown in figure)

Assumptions: diffusion capacitance is equal to the gate capacitance, the resistance of an nMOS transistor with unit width is $R$ and the resistance of a pMOS transistor with width 2 is also $R$, and NO sharing of diffusion regions

Off-path capacitances can contribute to delay, and if a node does not need to be charged (or discharged), its capacitance can be ignored

\[
ABCD = 0101 \rightarrow ABCD = 1101
\]

\[
ABCD = 1111 \rightarrow ABCD = 0111
\]

\[
ABCD = 1010 \rightarrow ABCD = 1101
\]
Look at the charges on the nodes at the end of the first input of the sequence; only the capacitances of the nodes which would change with the second vector need to be considered.

\[ ABCD = 0101 \rightarrow\]
\[ ABCD = 1101;\]
Delay = 36RC

\[ ABCD = 1111 \rightarrow\]
\[ ABCD = 0111;\]
Delay = 16RC

\[ ABCD = 1010 \rightarrow\]
\[ ABCD = 1101;\]
Delay = 43RC
Delay has two parts

**Parasitic Delay**
- 6 or 7 RC
- Independent of Load

**Effort Delay**
- 4h RC
- Proportional to load capacitance
Contamination Delay

Minimum (Contamination) Delay

- Best-case (contamination) delay can be substantially less than propagation delay
- Example, If both inputs fall simultaneously
- Important for “hold time” (will see later in the course)

\[ t_{cdr} = (3 + 2h)RC \]
We assumed contacted diffusion on every source/drain.
- Good layout minimizes diffusion area.
- Example, NAND3 layout shares one diffusion contact:
  - Reduces output capacitance by 2C.
  - Merged uncontacted diffusion might help too.

These general observations can be used for initial estimates of area and performance – using tools to extract parasitics will provide more accurate results for a particular technology.