Using MOS Models

C.K. Ken Yang
UCLA
yangck@ucla.edu
Courtesy of MAH
Overview

• **Reading**
  – Rabaey 5.4
  – W&H 4.2

• **Background**
  – In the past two lectures we have reviewed the iV and CV curves for MOS devices, both nMOS and pMOS. The results are fairly interesting, and can be used to explain a number of strange results that we found with our RC models. The iV models can be used to predict how delay with change with supply voltage, and why the logical effort of gates is a smaller number then our simple RC models predict. In fact we can use these models to explain how input slope affect gate delay which is part of the logical effort dilemma.
Review: Transistor as a Switched Current

\[
\text{delay} = C \Delta V / I_{\text{avg}}
\]

\[
\zeta = \lambda L_{\text{min}}
\]

\[
I_{\text{eff}} \approx \frac{1}{2}(I_{\text{max}} + I_{\text{mid}}) \approx I_{ds} \left( V_{gs} = V_{dd} ; V_{ds} = \frac{3}{4} V_{dd} \right)
\]

\[
I_{\text{eff}} = \frac{W}{L_{\text{eff}}} \mu C_{ox} \left( \frac{1}{2} \right) \left( V_{gs} - V_{th} \right)^2
\]

\[
L_{\text{eff}}^* = L_{\text{elec}} \left( 1 - \frac{\lambda L_{\text{min}} \left( \frac{3}{4} V_{dd} - V_{dsat} \right) }{L_{\text{elec}}} \right)
\]
iV Results

• Model(s) allow us to answer a number of questions
  – Performance scaling with voltage
  – Performance scaling with technology
  – LE of stacked transistors
  – Input slope affect on delay
  – Voltage gain issues
    • Relates to noise margins
  – Injected noise in coupling situations
• In the next few slides, we will focus on delay issues
  – Delay depends on $i_{dsat}$, $C_{load}$, $V_{dd}$
  – Look at $i_{dsat}$ first
Limitation of Simple RC Model

- Impact of velocity saturated and CLM
  - Current is not quadratic so the time-domain waveform is going to be different.
  - Shape of I-V curve is not the same when two transistors are placed in series.

- Input slope
  - Delay of a gate is a function of the delay of the driver gate
  - As gate ‘A’ slows down, the delay of gate ‘B’ gets larger
  - Previous analysis ignored this effect
Peak Saturation Current

• We had a couple of models for $i_{dsat}$
  – Full velocity saturation model with mobility reduction
  – Alpha power law model
• First, review how good these models are.
  – Full-velocity saturation model is physically intuitive.
  – But it is quite complex if we account for the vertical field dependence of $E_c$ and $\mu_{eff}$.
  – We can look at what happens if we ignore gate field effect
    • Set mobility to be a constant (not function of $V_{gs}$)
    • Use the high-field mobility for electron (NMOS) and holes (PMOS). since we are dealing with digital signals.
Equations

\[ \mu_e := 270 \frac{\text{cm}^2}{\text{V} \cdot \text{s}} \quad \mu_h := 90 \frac{\text{cm}^2}{\text{V} \cdot \text{s}} \quad v_{\text{sat}} := 8 \cdot 10^6 \frac{\text{cm}}{\text{s}} \quad E_{\text{ce}} := \frac{2 \cdot v_{\text{sat}}}{\mu_e} \quad E_{\text{ch}} := \frac{2 \cdot v_{\text{sat}}}{\mu_h} \]

\[ E_{\text{ce}} = 5.926 \frac{V}{\mu\text{m}} \quad E_{\text{ch}} = 17.778 \frac{V}{\mu\text{m}} \]

\[ i_{\text{dsNsat}}(V_{\text{gs}}, L) := \frac{W \cdot v_{\text{sat}} \cdot C_{\text{ox}} (V_{\text{gs}} - V_{\text{th}})^2}{(V_{\text{gs}} - V_{\text{th}}) + E_{\text{ce}} \cdot L} \quad i_{\text{dsPsat}}(V_{\text{gs}}, L) := \frac{W \cdot v_{\text{sat}} \cdot C_{\text{ox}} (V_{\text{gs}} - V_{\text{th}})^2}{(V_{\text{gs}} - V_{\text{th}}) + E_{\text{ch}} \cdot L} \]

\[ \mu_e(V_{\text{gs}}) := \frac{540}{1 + \left[ \frac{V_{\text{gs}} + V_{\text{th}}}{t_{\text{ox}} \cdot 0.54 \cdot 10^3 \frac{V}{\mu\text{m}}} \right]^{1.85} \frac{\text{cm}^2}{\text{V} \cdot \text{s}}} \]

\[ k := 7.28 \frac{\text{A}}{\text{m}} \]

\[ i_{\text{dsNalpha}}(V_{\text{gs}}, L) := k \cdot W \cdot \left[ \frac{C_{\text{ox}}}{1 \cdot \frac{F}{\text{m}^2}} \right]^{0.8} \left[ \frac{(V_{\text{gs}} - V_{\text{th}})}{1 \cdot \text{V}} \right]^{1.25} \]

\[ i_{\text{dsNfull}}(V_{\text{gs}}, L) := \frac{W \cdot v_{\text{sat}} \cdot C_{\text{ox}} (V_{\text{gs}} - V_{\text{th}})^2}{(V_{\text{gs}} - V_{\text{th}}) + E_{\text{ce}} (V_{\text{gs}}) \cdot L} \]

\[ V_{\text{gs}} := V_{\text{th}} \cdot \frac{V_{\text{dd}}}{N} + V_{\text{th}} \cdot V_{\text{dd}} \]
Current Comparison

\[ \begin{align*}
&i_{dsN10 - i, i + 1} \\
&20i_{dsNsat(V_{gs}, 0.4 \mu m)} \\
&20i_{dsNfull(V_{gs}, 0.4 \mu m)} \\
&20i_{dsNalpha(V_{gs}, 0.4 \mu m)} \\
\end{align*} \]
Model Comparison

- Most of the models are quite good
  - Alpha power and full velocity model match current well
  - Alpha power model has a fit constant for setting the current
  - Constant mobility model is off at low $V_{gs}$
    - Expected since it underestimates the current at low $V_{dd}$
    - Mobility is larger at these operating conditions
- Use simplest model for task
  - Often that is the alpha power model
  - For fixed operating voltage, simple fixed mobility velocity saturation model works well
Aside: A Quick Look at Voltage Scaling

- Using the alpha power model
  - Delay should scale as $C*V_{dd}/(V_{dd}-V_{th})^{1.25}$
  - Note that the important effect is not the 1.25 power, but rather that the scaling is $V/(V-x)$. As $V/x$ ratio gets smaller the circuit gets slower.
Velocity Saturation and CLM
Impacts of Velocity Saturation

• Changes the i-V curve so impacts how transistor turns on to discharge the load

• When you have two transistors in series (with same width)
  – Lateral field is smaller, since V is dropped across two transistors
  – Changes the degree of velocity saturation of each device.

• Velocity saturation current model can be very intuitively useful
  
  \[ i_{dsat} = \mu_{eff} C_{ox} \frac{W}{2L} \frac{(V_g - V_{th})^2}{E_{cL} + 1} \]

  – Saturation current is the relationship between \( V_{gs} - V_t \) and \( E_{cL} \).
  
  • With \( V_{gs} - V_t \ll E_{cL} \), the equation simplifies to the normal equation.
  
  – Alpha power model would need to be refit to account for this.
Inverter Driving a Capacitor

- Assume an infinite input slope
  
  **NMOS**
  - Let $L=0.4\,\mu\text{m}$ (for a 0.35$\mu\text{m}$ technology)
  - $E_c=6\text{V}/\mu\text{m}$, $V_{dd}=3.3\text{V}$
  - $V_{DSAT} = (2.6)2.4/(2.6+2.4) = 1.25 < V_{dd}/2$
  - So current is $\sim$ constant for the entire transition

- **PMOS**
  - $V_{DSAT} = (2.6)7.2/(2.6+7.2) = 1.91 \sim V_{dd}/2$
  - Less constant current (but more so with smaller channel lengths)

$$V_{dsat} = \frac{(V_{gs} - V_{th})E_cL}{(V_{gs} - V_{th}) + E_cL}$$

$$i_{dsat} = WV_{sat}C_{ox} \frac{(V_g - V_{th})^2}{(V_{gs} - V_{th}) + E_cL}$$

$$= \mu_{eff}C_{ox} \frac{W}{2L} \frac{(V_g - V_{th})^2}{(V_{gs} - V_{th})} / E_cL + 1$$
Resistor Comparison (NMOS)

- CV/I an excellent approximation $t/\tau$
  - Especially with velocity saturation

0.18µm technology

Courtesy of MG Johnson
Resistor Comparison (PMOS)

0.18 μm technology

Courtesy of MG Johnson
Logical effort and Elmore assumes that network of transistors is equivalent to network of resistors.
  - Parallel halves the resistance, series doubles the resistance.

So $I_{DS1} = 2I_{DS2}$
  - $R_2C$ is doubled (hence twice the delay)
  - True for quadratic model...
  - Velocity saturation causes the model to be incorrect.
nMOS Series Stacks

• Can do this using the velocity saturated model

\[ i_{dsat} = \mu_{\text{eff}} C_{\text{ox}} \frac{W}{2L_{\text{elec}}} \frac{(V_g - V_{th})^2}{(V_{gs} - V_{th})/E_c L_{\text{elec}} + 1} \]

• Adjust \( L_{\text{elec}} = m \cdot L_{\text{min}} \)
  - 2-stack = changing L by 2x, does not \( 1/2 \) the current
    • \( E_{ce} \) is around 6V/\( \mu \)m for 3.3V, so \( EL_{\text{min}} = 2.4 \)
    • \( L_{\text{elec}} = 2L_{\text{min}} = 0.8 \mu \)m changes this to 4.8
    • Current changes from 1/5.1 to 1/7.5
    • Only need to make the transistors 1.5x wider!
  - 3 Transistor stack is only \( 1/2 \) the current
    • Need to make the transistors only 2x wider
pMOS Series Stacks

- Velocity saturation is a much smaller effect here
- \( E_{ch} \) is much larger
  - \( 18V/\mu\text{m} \), so \( E_cL = \ldots \) \( V \)
- Two series devices are 10/17
  - Need to make devices \( \ldots \) times wider for same current
- Three series devices are 10/24
  - Need to make devices \( \ldots \) times as wide for same current
- In general
  - Increase the size of an m-series stack by a factor, \( X(m) \)

\[
X(m) = \frac{(V_{gs} - V_{th}) + m \cdot E_c L_{min}}{\left(V_{gs} - V_{th}\right) + E_c L_{min}}
\]
Accounting for CLM

• Replace $L_{\text{elec}}$ with the $L_{\text{eff}}^*$

$$L_{\text{eff}}^* = L_{\text{elec}} \left( 1 - \frac{\lambda L_{\text{min}} \left( \frac{3}{4} V_{dd} - V_{dsat} \right)}{L_{\text{elec}}} \right)$$

$$i_{\text{dsat}} = \mu_{\text{eff}} C_{\text{ox}} \frac{W}{2L_{\text{eff}}^*} \left( \frac{V_{g} - V_{\text{th}}}{V_{gs} - V_{\text{th}}} \right)^2 \sqrt{E_{c}L_{\text{eff}}^* + 1}$$

- With $\lambda=0.05$, $L_{\text{eff}}^* = 0.37\mu m$ with $L_{\text{elec}} = L_{\text{min}}$
  • $i_{\text{dsmax}}$ is about 7% higher than $i_{\text{dsat}}$ (w/o CLM)
  • $I_{\text{dsavg}} = 1.037 i_{\text{dsat}}$ (w/o CLM)
- With $\lambda=0.05$, $L_{\text{eff}}^* = 0.78\mu m$ with $L_{\text{elec}} = 2L_{\text{min}}$
  • $I_{\text{dsavg}} = 1.015 i_{\text{dsat}}$ (w/o CLM)

• Effect can be more with 90nm technology or beyond (~15% increase for $i_{\text{dsmax}}$)
Logical Effort Comparison

• Accounting for vel-sat (assume the inverter P:N = 2)
  – 2-input NAND
    • PMOS=2, NMOS=1.48 (instead of size 2)
    • $g_{\text{NAND}_2} = 3.48/3 = 1.16$ (instead of 1.33)
  – 3-input NAND
    • PMOS=2, NMOS=1.96 (instead of size 3)
    • $g_{\text{NAND}_3} = 4/3 = 1.33$ (instead of 1.67)

• The current is less than the quadratic model but improves the logical effort.

• Accounting for CLM
  – Inverter’s average current is slightly higher b/c of CLM
  – 2-input NAND
    • PMOS=2, NMOS=1.51 (a factor of $(1+(0.037-0.015)$ higher)
Unequal Stack Sizing

- Not common (due to area)
- Roughly compensate by calculating the correct $L_{elec}$.
  - Device closest to the output (M1) is saturated
  - Equivalent length of a W1 device
    - $L_{elec} = L_{min}(1+W_1/W_2)$
    - Can use the $L_{elec}$ for CLM as well.
- Example: $W_1=1$, $W_2=2$
  - $L_{elec} = L_{min} \times _______
Input Slope Effect
Input Slope

- Delay of gate depends on the speed of input
  - If input is very fast, transistor will be fully on before the output starts to change.
    - Get the smallest delay possible
  - If input is slow, then the transistor will have its gate at less than Vdd, and the current will be less than expected.
    - The delay will increase
- Not purely dependent on input.
  - If input is slow but output is slower, then transistor will still be fully on before output starts to change.
  - Function of the ratio of input to output slope
Inverter Model

- Difference in nMOS and pMOS currents drive the output

\[ C \frac{d}{dt} V_{out} = i = i_{pMOS} - i_{nMOS} \]

- Devices basically acts like a voltage controlled current source.
- If the input voltage is changing, we can apply the nonlinear i-V equations to solve the equation.

- Too difficult.
  - Make some simplifications.

\[ V_{DD} \]

\[ V_{in} \]

\[ C_L \]
Input Slope Effect – Simplified Calculation

- Assume
  - Gate switches when input=Vdd/2
  - Linear current to voltage.
- Solve for the $t_{\text{PHL, ramp}}$
  \[
  \int_{0}^{t_{r}/2} i_{\max} \frac{t}{t_{r}/2} \, dt + \int_{t_{r}/2}^{t_{\text{PHL, ramp}}} i_{\max} \, dt = - \int_{V_{DD}/2}^{0} C_{L} \, dV_{\text{out}}
  \]
  
  \[t_{\text{PHL, ramp}} = \frac{t_{r}}{4} + \frac{C_{L} \cdot V_{DD} / 2}{i_{\max}} = \frac{t_{r}}{4} + \frac{C_{L} \cdot V_{DD} / 2}{I_{DS}(\text{sat})}
  \]
  - Delay is increased by $t_{r}/4$
  - With $t_{r} \sim 2t_{\text{d, step}}$ ($t_{\text{d, step}}$ = step input delay of input gate).
Input Slope Effect - Simulation

- Sweep input signal slope.
  - $t_d(\text{step}) = \text{delay of inverter generating the input (from a step)}$
  - $t_d(\text{ramp}) = \text{delay of inverter (with sloped input)}$
- Result is less than estimated.
  - Equation is a slight overestimate.
  - Result is quite linear.
- Question our assumptions
  - Assume $i_L=0$ until $V_{IN}=V_{dd}/2$
  - Pessimistic for delay

![Graph showing input signal delay vs. time with labeled equation $t_d(\text{ramp}) = t_d(\text{step}) + 0.5(t_d(\text{step}) \text{ of driving stage})$.]
Inverter Drive Curves

- Look at a contour plot of output current vs. $V_{in}$, and $V_{out}$

$\text{i}_{out} < 0$ (decreases)

$\text{i}_{out} > 0$ (increases)

$\text{i}_{out} = 0$
Different (Optimistic) Assumption

- Assume that $i_L$ starts to discharge/charge $C_L$ at $V_{TH^*}$
  - This is optimistic
- Repeat the integration or do this graphically.
- Behavior
  - Initial discharge ramp is quadratic.
  - Once, $i_{OUT}=i_{MAX}$, discharge linearly

Courtesy of MG Johnson
Input Slope Effect – Graphical Analysis

• Now, assume a step input that causes the output to transition to $V_{\text{OUT}} = V_{dd}/2$ at the same time.

• Step must have occurred when $I_{\text{out}} = I_{\text{max}}/2$
  – Occurs at $(V_{dd} + V_{\text{TH}^*})/2$

• So added delay is $\Delta t$ for the input to change by $V_{\text{TH}^*}/2$

$$t_{d,\text{ramp}} = t_{d,\text{step}} + \frac{V_{\text{TH}^*}}{V_{DD}} t_r = t_{d,\text{step}} + \frac{V_{\text{TH}^*}}{V_{DD}} t_{d,\text{in}}$$

Courtesy of MG Johnson
Input Slope Effect Discussion

• Simple graphical approach for intuition
  – With input current threshold of $V_{TH}^*$
    • $V_{TN}=V_{TP}=0.7$ with $V_{dd}=3.3$ and $t_{\text{delay}} = t_{d\_\text{step}} + \ldots t_{d\_\text{input}}$
  – With input current threshold of $V_{dd}/2$
    • $t_{\text{delay}} = t_{d\_\text{step}} + \ldots t_{d\_\text{input}}$

• A more realistic number is somewhere between the two estimates… closer to $t_{\text{delay}} = t_{d\_\text{step}} + 0.4t_{d\_\text{input}}$
Input Slope Effect Modeling

- Use logical effort methodology instead of dealing with step inputs
  - Input/output has same slope as the baseline for delay
  - Note that $t_{\text{LE}} = (1 + k_{\text{slope}}) t_{\text{step}}$ for equal input and output slope.
  - So $g_{\text{eq_slope}} = (1 + k_{\text{slope}}) g_{\text{step}}$

- Simulation shows very linear fitting for different input slope (and fixed output slope)
Input Slope Effect in Complex Gates

- Which input switches matters
- Top transistor switches
  - Source node rises with the input slope.
  - Current is linearized.
  - Very similar to inverter analysis
- Bottom transistor switches
  - Drain is $V_{dd} - V_{th}$
  - Current saturates with lower $V_{GS}$ (earlier in the transition)
    - $t_r$ is effectively smaller
  - Less input slope dependence.
Input Slope Correction

- Changes for different logic gates
  - And different inputs and Vdd.
  - Result is consistent with our intuition
    - $1/K$ is larger when $V_{dd}/V_{th}$ is smaller
- Incorporate delay equation into LE formalism
  - If $K$'s are roughly similar, the result is greatly simplified.

$$D_{LE,SC} = \sum_{i=1}^{N} \left( g_i \cdot h_i + p_i - \frac{g_i \cdot h_i - g_{i-1} \cdot h_{i-1}}{K_i} \right)$$

$$D_{LE,SC} = D_{LE} - \frac{g_N \cdot h_N - g_{in-driver} \cdot h_{in-driver}}{K}$$
Simulation/Synthesis

- Incorporating ISE in a chain of NAND gates buffering (LEFT)
  - Note that increasing $f$ is better for both power and delay than constant $f$.
- Can improve power and performance of a synthesized adder.

![Graph showing internal energy vs. delay increment and internal power vs. delay normalized to the synthesized adder.](image-url)
Summary

• Our i-V models, while simple, are quite good
  – Gives decent intuition on design and sizing
• Handles several effects and extends existing methodology well
  – Velocity saturation can be incorporated into calculating the logical effort by considering \( (1+V_{gst}/E_{cL}) \)
  – CLM can also be added by using \( L_{eff}^* \)
  – Input slope effect can be added into the LE methodology.
    • As a constant factor when slopes are different.
• We will also apply this model for a more accurate understanding of noise margin and sense-amplifiers