EE141-Fall 2012 Digital Integrated Circuits

Lecture 11
MOS Capacitance
and Delay

Announcements

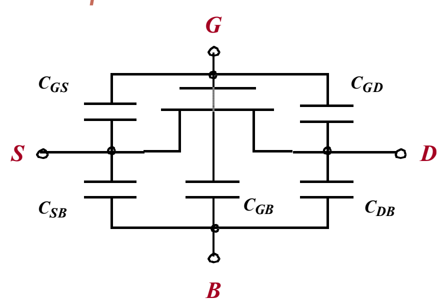
- No labs this week
  - Labs restart next week
- Midterm #1 Thurs. Oct. 4th, 6:30-8:00pm
  - Exam is open notes, book, calculators, etc.
  - Covers up to lecture 10

Class Material

- Last lecture
  - Using the MOS model: Inverter VTC
- Today's lecture
  - MOS Capacitance
  - Using the MOS Model: Delay
- Reading (3.3.2, 5.4.2)

MOS Capacitance

MOS Capacitances



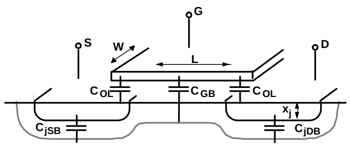
Gate Capacitance

- Capacitance (per area) from gate across the oxide is $W \cdot L \cdot C_{ox}$, where $C_{ox} = \varepsilon_{ox}/t_{ox}$
  - But channel isn't really a terminal in my MOS transistor model…

Gate Capacitance

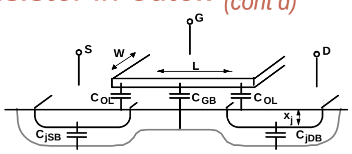
- Distribution between terminals is complex
  - Capacitance is really distributed
    - Useful models lump it to the terminals
  - Several operating regions:
    - Way off, off, transistor linear, transistor saturated

Transistor In Cutoff



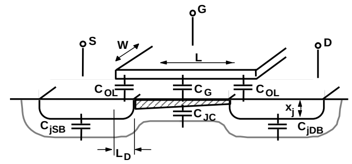
- When the transistor is off, no carriers in channel to form the other side of the capacitor.
  - Substrate acts as the other capacitor terminal
  - Capacitance becomes series combination of gate oxide and depletion capacitance

Transistor In Cutoff *(cont'd)*



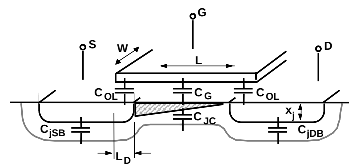
- When $|V_{GS}| < |V_T|$, total C_{GCB} much smaller than $W \cdot L \cdot C_{ox}$
  - Usually just approximate with $C_{GCB} = 0$ in this region.
- (If V_{GS} is "very" negative (for NMOS), depletion region shrinks and C_{GCB} goes back to $\sim W \cdot L \cdot C_{ox}$)

Transistor in Linear Region



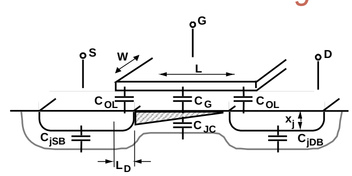
- Channel is formed and acts as the other terminal
  - C_{GCB} drops to zero (shielded by channel)
- Model by splitting oxide cap equally between source and drain
  - Changing either voltage changes the channel charge

Transistor in Saturation Region



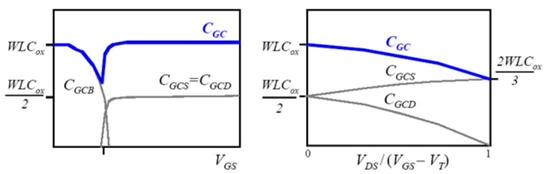
- Changing source voltage doesn't change V_{GC} uniformly
  - E.g. V_{GC} at pinch off point still V_{TH}
- Bottom line: $C_{GCS} \approx 2/3 \cdot W \cdot L \cdot C_{ox}$

Transistor in Saturation Region *(cont'd)*



- Drain voltage no longer affects channel charge
  - Set by source and V_{DS_sat}
- If change in charge is 0, $C_{GCD} = 0$

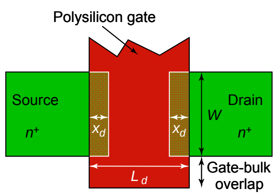
Gate Capacitance



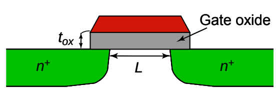
C_{gate} vs. V_{GS} (with $V_{DS} = 0$)

C_{gate} vs. operating region

Gate Overlap Capacitance



Top view

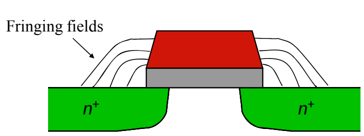


Cross section

$$C_O = C_{ox} \cdot x_d$$

Off/Lin/Sat → $C_{GSO} = C_{GDO} = C_O \cdot W$

Gate Fringe Capacitance

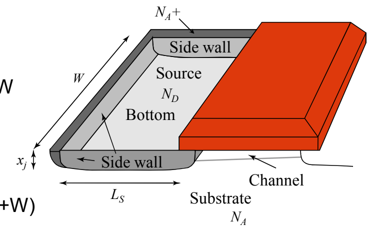


Cross section

- C_{OV} not just from metallurgic overlap – get fringing fields too
- Typical value: ~0.2fF·W(in μm)/edge

Diffusion Capacitance

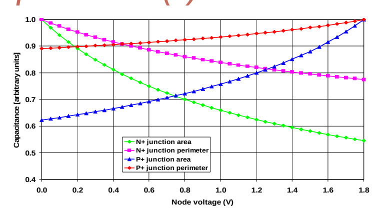
- Bottom
  - Area cap
  - $C_{bottom} = C_j \cdot L_S \cdot W$
- Sidewalls
  - Perimeter cap
  - $C_{sw} = C_{jsw} \cdot (2L_S + W)$



- GateEdge
  - $C_{ge} = C_{jgate} \cdot W$
  - Usually automatically included in the SPICE model

Junction Capacitance (2)

- Junction caps are nonlinear
  - C_J is a function of junction bias



- SPICE model equations:
  - Area C_J = area × C_{J0} / $(1+ |V_{DB}|/\phi_B)^{mj}$
  - Perimeter C_J = perim × C_{JSW} / $(1 + |V_{DB}|/\phi_B)^{mjsw}$
  - Gate edge C_J = W × C_{Jgate} / $(1 + |V_{DB}|/\phi_B)^{mjswg}$
- How do we deal with nonlinear capacitance?

Linearizing the Junction Capacitance

Replace non-linear capacitance by
large-signal equivalent linear capacitance
which displaces equal charge over voltage swing of interest

$$C_{eq} = \frac{\Delta Q_j}{\Delta V_D} = \frac{Q_j(V_{high}) - Q_j(V_{low})}{V_{high} - V_{low}} = K_{eq} C_{j0}$$

$$K_{eq} = \frac{-\phi_0^m}{(V_{high} - V_{low})(1-m)}[(\phi_0 - V_{high})^{1-m} - (\phi_0 - V_{low})^{1-m}]$$

Capacitance Model Summary

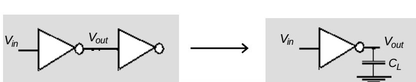
- **Gate-Channel Capacitance**
  - $C_{GC} \approx 0$ ($|V_{GS}| < |V_T|$)
  - $C_{GC} = C_{ox} \cdot W \cdot L_{eff}$ (Linear)
    - 50% G to S, 50% G to D
  - $C_{GC} = (2/3) \cdot C_{ox} \cdot W \cdot L_{eff}$ (Saturation)
    - 100% G to S
- **Gate Overlap Capacitance**
  - $C_{GSO} = C_{GDO} = C_O \cdot W$ (Always)
- **Junction/Diffusion Capacitance**
  - $C_{diff} = C_j \cdot L_S \cdot W + C_{jsw} \cdot (2L_S + W) + C_{jg} W$ (Always)

Capacitances in 0.25 μm CMOS Process

	C_{ox} (fF/µm²)	C_O (fF/µm)	C_j (fF/µm²)	m_j	ϕ_b (V)	C_{jsw} (fF/µm)	m_{jsw}	ϕ_{bsw} (V)
NMOS	6	0.31	2	0.5	0.9	0.28	0.44	0.9
PMOS	6	0.27	1.9	0.48	0.9	0.22	0.32	0.9

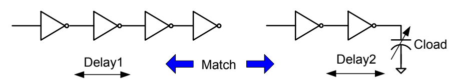
Simplified Model

- Capacitance models important for analysis and intuition
  - But often need something simpler to work with
- Simple switch model:
  - Lump together as effective linear capacitance to (ac) ground
  - In most processes: $C_G = C_D = 1.5 – 2\text{fF} \cdot W(\mu m)$

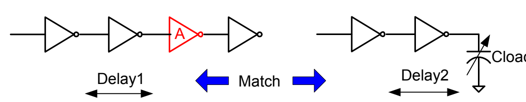


Model Calibration - Capacitance

- Can calculate C_g, C_d based on tech. parameters
  - But these models are simplified too
- Another approach:
  - Tune (e.g., in spice) the linear capacitance until it makes the simplified circuit match the real circuit
  - Matching could be for delay, power, etc.

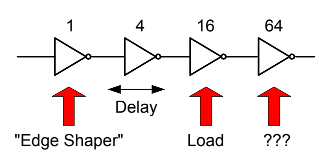


Model Calibration for Delay



- For gate capacitance:
  - Make inverter fanout 4
  - Adjust C_{load} until Delay1 = Delay2
- For diffusion capacitance
  - Replace inverter “A” with a diffusion capacitance load

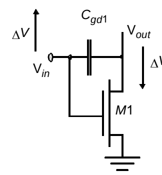
Delay Calibration



- Why did we need that last inverter stage?

The Miller Effect

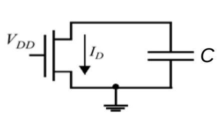
- As V_{in} increases, V_{out} drops
  - Once get into the transition region, gain from V_{in} to V_{out} > 1
- So, C_{gd} experiences voltage swing larger than V_{in}
  - Which means you need to provide more charge
  - Makes C_{gd} look larger than it really is
- Known as the “Miller Effect” in the analog world



CMOS Switching Delay

MOS Transistor as a Switch

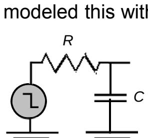
- Discharging a capacitor



$$i_D = i_D(v_{DS})$$

$$i_D = C\frac{dV_{DS}}{dt}$$

- We modeled this with:



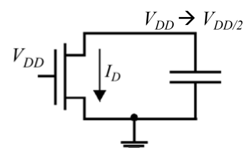
$t_p = \ln(2)\,RC$

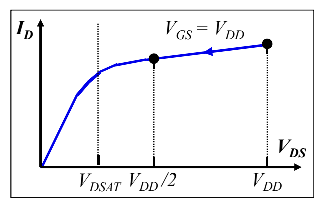
MOS Transistor as a Switch

- Saw that real transistors aren’t exactly resistors
  - Look more like current sources in saturation
- Two questions:
  - Which region of IV curve determines delay?
  - How can that match up with the RC model?

Transistor Driving a Capacitor

- With a step input:

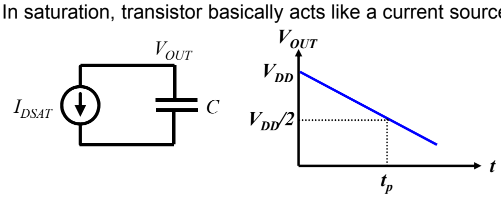




- Transistor is in (velocity) saturation during entire transition from V_{DD} to $V_{DD}/2$

Switching Delay

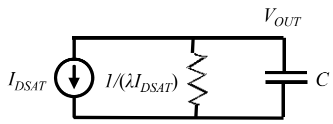
- In saturation, transistor basically acts like a current source:



$V_{OUT} = V_{DD} - (I_{DSAT}/C)t \longrightarrow$ $\boxed{t_p = C(V_{DD}/2)/I_{DSAT}}$

Switching Delay *(with Output Conductance)*

- Including output conductance:

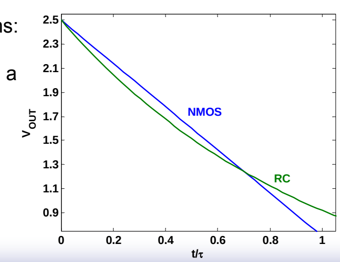


$$V_{OUT} = \left(V_{DD} + \lambda^{-1}\right) e^{-t/(C/\lambda I_{DSAT})} - \lambda^{-1}$$

- For "small" λ: $t_p \approx \dfrac{C\left(V_{DD}/2\right)}{\left(1+\lambda V_{DD}\right) I_{DSAT}}$

RC Model

- Transistor current not linear on V_{OUT} – how is the RC model going to work?
- Look at waveforms:
- Voltage looks like a ramp for RC too



Finding Req

- Match the delay of the RC model with the actual delay:

$$t_p = t_{p,RC}$$

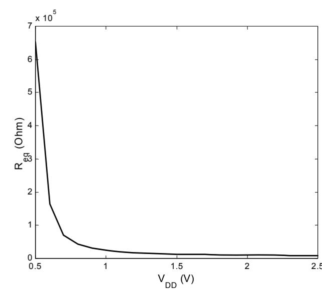
$$\frac{C\left(V_{DD}/2\right)}{\left(1+\lambda V_{DD}\right) I_{DSAT}} = \ln(2) R_{eq} C \longrightarrow R_{eq} = \frac{\left(V_{DD}/2\right)}{\ln(2)\left(1+\lambda V_{DD}\right) I_{DSAT}}$$

- Often just: $R_{eq} \approx \dfrac{1}{2\cdot\ln(2)} \dfrac{V_{DD}}{I_{DSAT}}$

- Note that the book uses a different method and gets $0.75 \cdot V_{DD}/I_{DSAT}$ instead of $\sim 0.72 \cdot V_{DD}/I_{DSAT}$.
  - Why did we do it this way vs. the book's method?

The Book's Method

The Transistor as a Switch



The Transistor as a Switch

Table 3.3 Equivalent resistance R_{eq} (*W/L*= 1) of NMOS and PMOS transistors in 0.25 μm CMOS process (with $L = L_{min}$). For larger devices, divide R_{eq} by *W/L*.

V_{DD} (V)	1	1.5	2	2.5
NMOS (kΩ)	35	19	15	13
PMOS (kΩ)	115	55	38	31