COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

ELEN E3106/4106 Lecture 23

MOSFETs Part III

Outline

- Long vs. short channel MOSFETs
- Wrapping up basic quantitative current-voltage model
- High-k dielectrics
- MOSFETs in ICs & Secondary Effects (Tunneling, Scaling, Leakage, DIBL)

Assignments:

Homework 9 due Thursday Dec. 12th by 5pm

Extra credit opportunity during in-class review session on Thurs. Dec 5th

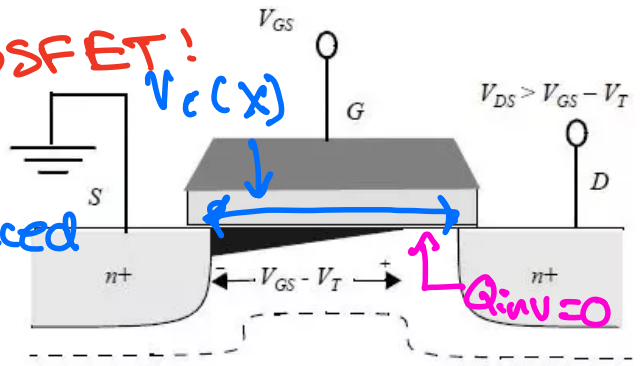
Final Exam Thurs. Dec. 19th 4-7pm

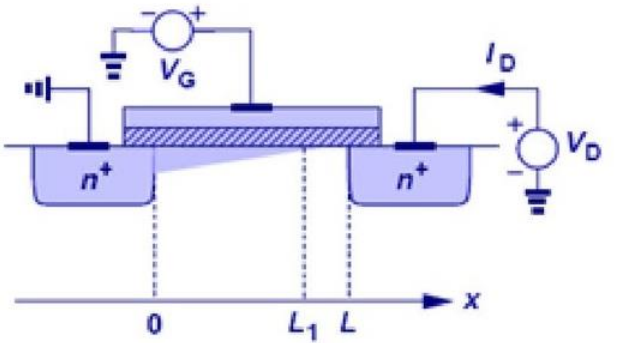
Electric Field in the Pinch-Off Region

- The channel potential V_C is always equal to $V_{d,sat} = V_{gs} - V_t$ at the pinch-off, where $Q_{inv} =$ 0
- Channel narrowing: The channel cross-sectional area is greatly reduced in the pinch-off region!
- In the pinch-off region, most of the applied V_{ds} drops across the very small depleted region near the drain, because the resistance of this region is much higher compared to the rest of the channel
- Large voltage drop over small distance = high E-field
- How do carriers travel across the pinched-off region?
  - Due to the high E-field, e- in channel are pulled into pinch-off region and dragged across

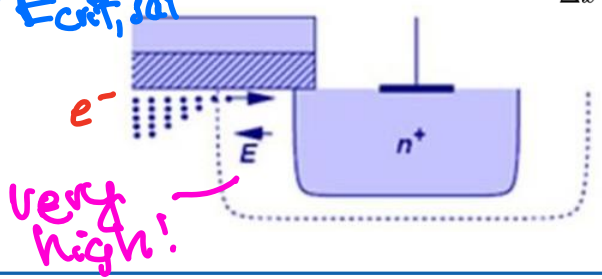
$$E_{pinch-off} = \frac{V_{ds} - (V_{gs} - V_t)}{L - L_1}$$

  - They *usually* travel at drift saturation velocity, v_{sat} if $E_{pinch-off} \geq E_{crit,sat}$
- The pinch-off region is just depletion, not inverted
- Recall, depletion regions lacks mobile carriers
- Probability of recombination for e- in the depletion region is very small since region is depleted of h+, so we can neglect recombination





$$E = -\frac{\Delta V}{\Delta x}$$



Sources: C. Liu EE105 Fall 2007 Slides, Quora

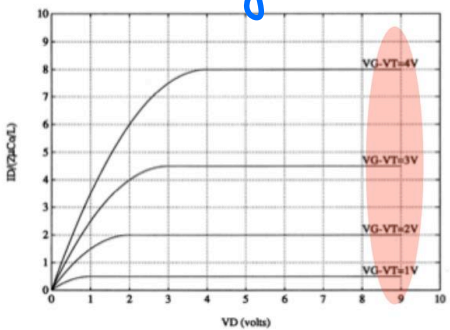
Long vs. Short Channel MOSFET

- So far, our saturation current analysis has assumed our MOSFETs are long channel
- But for short channel lengths, carries travel at v_{sat} over most of the channel!
- Therefore, the saturation drain current doesn't increase quadratically ($V_{gs} - V_t$) like in long channels

Square Law

$$I_{DSAT} \propto (V_{GS} - V_T)^2$$

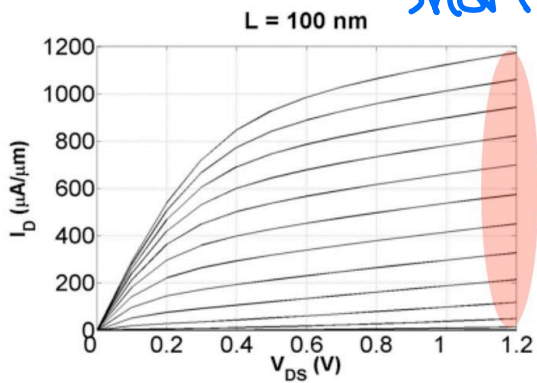
Long



Velocity saturated

$$I_{DSAT} \propto (V_{GS} - V_T)$$

short



- Modern MOSFETs tend to be short channel (L < 35 nm). Lower resistance (R $\propto$ L), higher conductivity
- Downside? Short channel devices can only block lower source-drain voltages

Sources: M. Lundstrom ECE 305 Slides

Channel Length Modulation

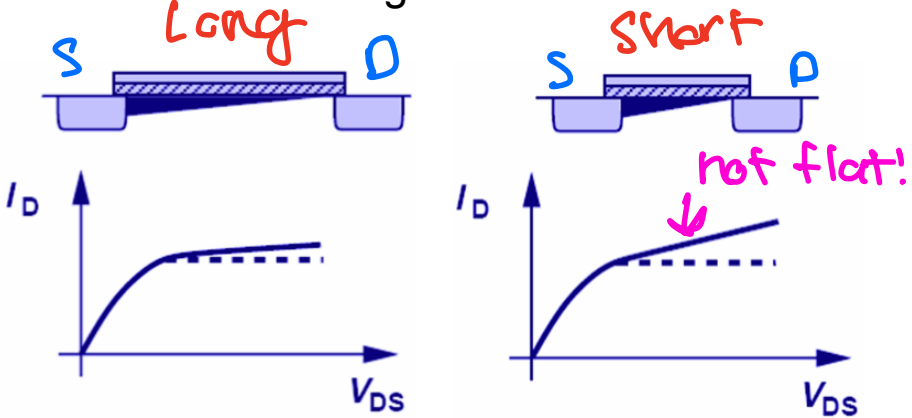
- As we have seen, the pinch-off point moves towards the source as V_{ds} increases
- Therefore, the length of the inversion-layer channel becomes shorter
- I_d will increase (slightly) with increasing V_{ds} in saturation due to channel length modulation! Not desired!

$$I_{Dsat} \propto \frac{1}{L-\Delta L} \cong \frac{1}{L}\left(1+\frac{\Delta L}{L}\right)$$

$$\Delta L \propto \left(V_{DS}-V_{DSsat}\right)$$

$$I_{D,sat} = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}\left(V_{GS}-V_{TH}\right)^2\left[1+\lambda\left(V_{DS}-V_{D,sat}\right)\right]$$

λ: channel length modulation coefficient



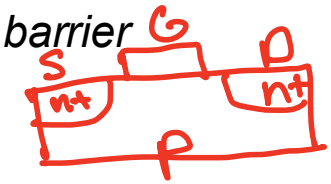
- The effect of channel length modulation is less for a long channel MOSFET than for a short channel MOSFET

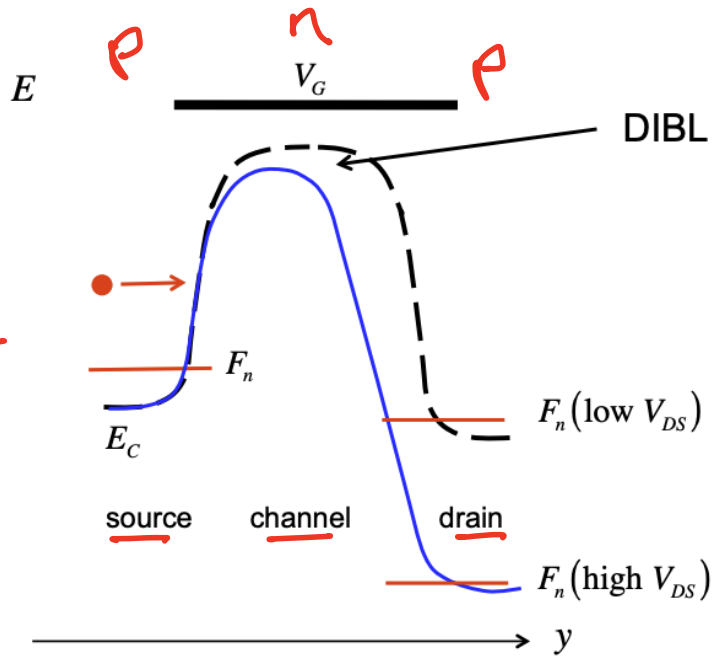
$$\lambda \propto \frac{1}{L} \Rightarrow \text{short channel MOSFET has larger } \lambda$$

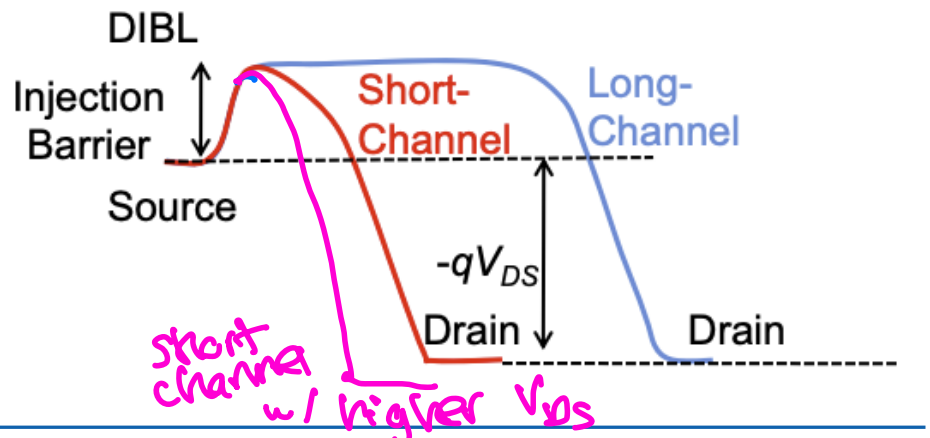
Sources: https://inst.eecs.berkeley.edu/~ee105/sp08/lectures/lecture17.pdf

Drain-Induced Barrier Lowering (DIBL)

- The concept that the drain can lower the source barrier and reduce V_t is called *drain-induced barrier lowering* or DIBL
- Classically, V_t should be independent of V_{ds}
- But as the drain voltage increases, the reverse bias on the body-drain p-n junction increases the conduction band edge is pulled down, and the depletion region widens
- V_t will decrease with increasing drain bias!
- I_d will increase with increasing drain bias!
- DIBL is a larger concern in short channel devices
- At extremely short lengths, the gate can entirely fail to turn the device off due to DIBL





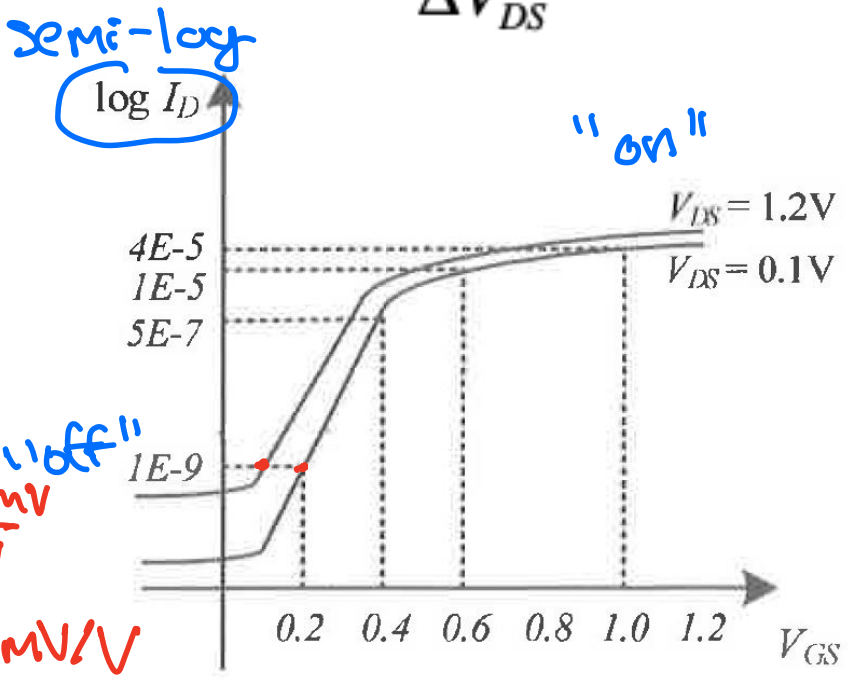


Sources: M. Lundstrom ECE 305 Slides, https://inst.eecs.berkeley.edu/~ee105/sp08/lectures/lecture17.pdf

Metric: DIBL

- Units: mV/V
- We take the horizontal shift in the sub-threshold characteristics on the $\log(I_d) - V_{gs}$ transfer curve (in milivolts) and divide by the change in V_{ds}
- Make sure to select a region of the plot where the current is exponential with gate voltage (linear on the log plot) where the low V_{ds} and high V_{ds} characteristics are parallel
- What's the DIBL of this MOSFET?

$$DIBL \equiv \frac{\Delta V_{GS}}{\Delta V_{DS}} (\text{mV/V})$$

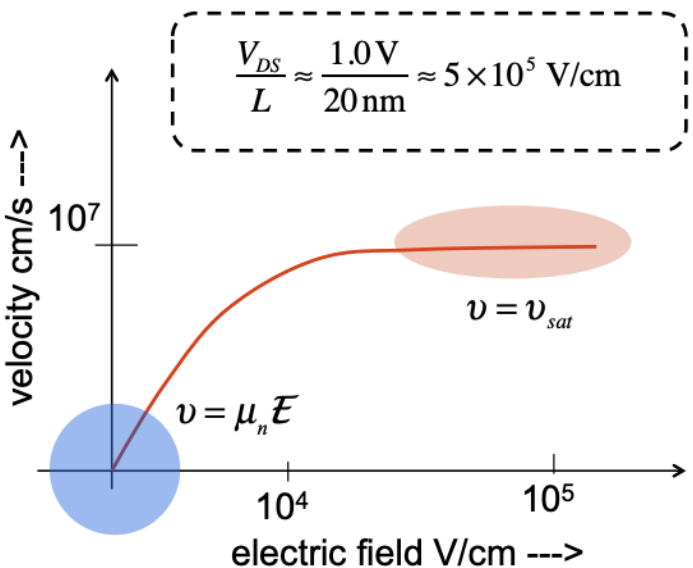


$$DIBL = \frac{\Delta V_{GS}}{\Delta V_{DS}} = \frac{0.2 - 0.1}{1.2 - 0.1} = \frac{0.1}{1.1\text{V}} = \frac{100\text{mV}}{1.1\text{V}} \approx 91\ \text{mV/V}$$

Sources: M. Lundstrom ECE 305 Slides, https://inst.eecs.berkeley.edu/~ee105/sp08/lectures/lecture17.pdf

Short Channel: Velocity Saturation

- In state-of-the-art MOSFETs, the channel is very short (<100 nm)
- The lateral electric field across the entire channel is very high such that carriers reach saturation velocity, v_{sat}
- Ex. Modern devices $L = 35\,nm$, so at $V_{ds} = 1$ V, $E \approx 3 \times 10^5$ V/cm > E_{crit}(Si)
- Recall: The E-field value at which the carriers reach v_{sat} is called E_{crit}



$$v_{sat} = \begin{cases} 8\times10^6 \text{ cm/s for electrons in Si} \\ 6\times10^6 \text{ cm/s for holes in Si} \end{cases}$$

$$\begin{cases} \text{NMOS: } \mu_n \approx 250 \text{ cm}^2/\text{V-s} \Rightarrow E_{sat} \approx 30{,}000 \text{ V/cm} \\ \text{PMOS: } \mu_n \approx 80 \text{ cm}^2/\text{V-s} \Rightarrow E_{sat} \approx 80{,}000 \text{ V/cm} \end{cases}$$

For $L = 0.1\ \mu$m

$$\begin{cases} V_{D,sat} = 0.3 \text{ V for NMOS} \\ V_{D,sat} = 0.8 \text{ V for PMOS} \end{cases}$$

Sources: C. Hu *Modern Semiconductor Devices for Integrated Circuits*, M. Lundstrom ECE 305 Slides

Short Channel: Current in the Saturation Region (High V_{ds})

- The impact of saturation velocity leads to the velocity-saturated drain current equation for short channel devices in saturation mode
- Recall that $I_D = WQ_{inv}v$
- If V_{ds} ___ $E_{sat}L$, the carrier velocity saturates and hence the drain current fully saturates:

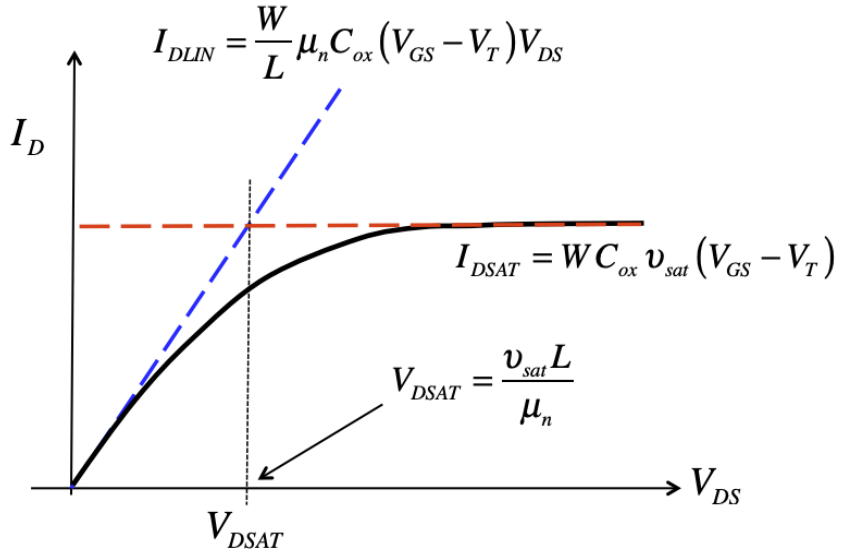
$$I_D = WC_{ox}\upsilon_{sat}\left(V_{GS} - V_T\right) = WC_{ox}v_{sat}V_{d,sat}$$

- **So, for short channel devices, $I_{d,sat}$ $\propto$ ___ rather than ___ in long channel devices**
- Also note $I_{d,sat}$ is not directly dependent on ___, but is dependent on ___ !

Sources: C. Hu *Modern Semiconductor Devices for Integrated Circuits*, M. Lundstrom ECE 305 Slides

Current-Voltage Model Across the Full Range

- Putting together our piecewise approximation of the current, we now have a model for the I-V in a **short channel** MOSFET



$$I_D/W = -Q_n(V_{GS})\langle \upsilon(V_{DS})\rangle$$

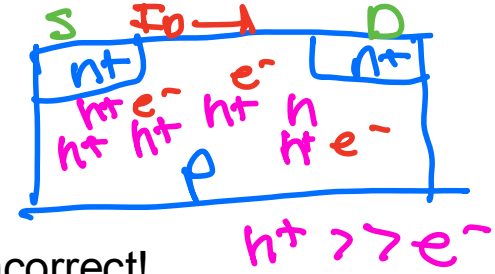
$V_{GS} \geq V_T$: $Q_n(V_{GS}) = -C_{ox}(V_{GS} - V_T)$ $\quad$ $V_{DS} \leq V_{DSAT}$: $\langle \upsilon(V_{DS})\rangle = \left(\mu_n \frac{V_{DS}}{L}\right)$

$V_{GS} < V_T$: $Q_n(V_{GS}) = 0$ $\quad$ $V_{DS} > V_{DSAT}$: $\langle \upsilon(V_{DS})\rangle = \upsilon_{sat}$

- Note: for long channel MOSFETs, $I_{DSAT} = \frac{W}{2L}\mu_n C_{ox}(V_{GS} - V_T)^2$
- You may notice we will not get a smooth curve with this model. Typically, we empirically adjust the fit from high to low V_{ds} with an extra parameter β

Sources: M. Lundstrom ECE 305 Slides

Subthreshold Current – "Off" is not totally "Off"



- What about cut-off mode (off state)?
- So far, we have assumed $I_d = 0$ in the cut-off region ($V_{gs} <$ V_t). This is incorrect!
- Below threshold, the inversion electron concentration (n_s) is small but nonetheless can allow a small current to flow between the source and the drain
- As V_{gs} is reduced below V_t, the potential barrier from the source into the channel is increased
- Therefore, I_d becomes limited by carrier diffusion through the channel, rather than drift
- The subthreshold current decreases exponentially with linearly decreasing V_{gs}/m

$$I_D = \mu_{eff} C_{ox} \frac{W}{L}(m-1)\left(\frac{kT}{q}\right)^2 e^{q(V_G - V_T)/mkT}\left(1 - e^{-qV_{DS}/kT}\right)$$

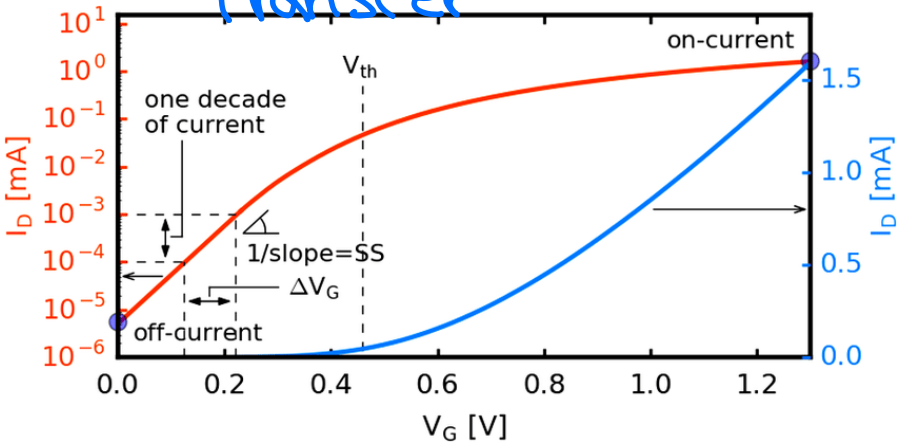
- Where m is an integer
- The subthreshold current is the main contributor to the current in the off-state, I_{off}!

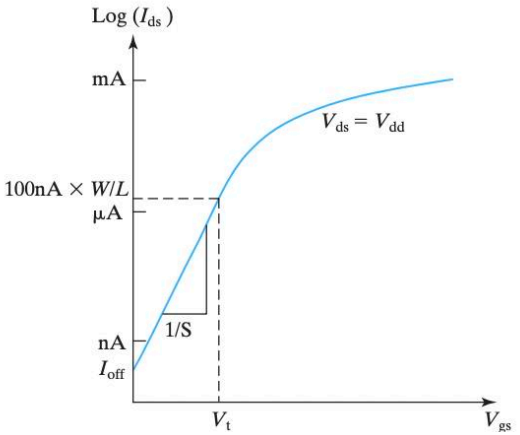
Sources: C. Hu *Modern Semiconductor Devices for Integrated Circuits*, M. Lundstrom ECE 305 Slides

Metric: Subthreshold slope

- If we plot the semi-log of $I_d - V_{gs}$, we should get a linear behavior in the subthreshold regime (cut-off)
- The inverse slope of this line is known as subthreshold slope or swing, *S* or *SS*
- Typical value: 70 -100 mV. We want it as small as possible. Why?
- A change in the input V_{gs} of 70 mV will change the output I_d by an order of magnitude (factor of 10)!
- Clearly, the smaller the value of S, the better the transistor is as a switch (voltage-controlled current source)! I_D
- Key point: A small value of S means a small change in the input bias V_{GS} can modulate the output current I_D considerably

Transfer





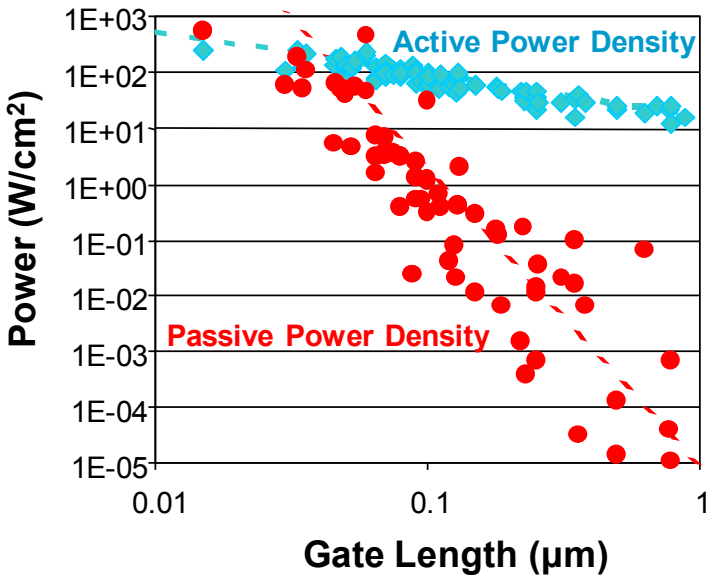
$$S \equiv \left(\frac{d(\log_{10} I_D)}{dV_{GS}} \right)^{-1} = \frac{kT}{q} \ln(10)\left(1 + \frac{C_d}{C_{ox}}\right)$$

Sources: C. Hu *Modern Semiconductor Devices for Integrated Circuits*, DOI:10.1007/s00502-017-0534-y

Metric: Off-Current and Static Power Dissipation

$$P_{static} = V_{dd} I_{off}$$

- I_{off} is the I_d measured at $V_{gs} = 0$ and $V_{ds} = V_{dd}$, the supply voltage
- It is important to keep I_{off} very small in order to minimize the static power that a circuit consumes when it is in the standby mode
- In modern MOSFETs, $I_{off} \sim I_{on}/1000$
- Let's say I_{off} is a modest 100 nA per transistor
- A cell-phone chip containing one hundred million transistors would consume 10 A even in standby!
- The battery would be drained in minutes without receiving or transmitting any calls
- See previous slide for determining off-current from transfer curve!



Ex: see IBM journal of Research & Dev.
http://www.research.ibm.com/journal/rd/504/tocpdf.html

Sources: E. Pop ECE 340 Slides, M. Lundstrom ECE 305 Slides

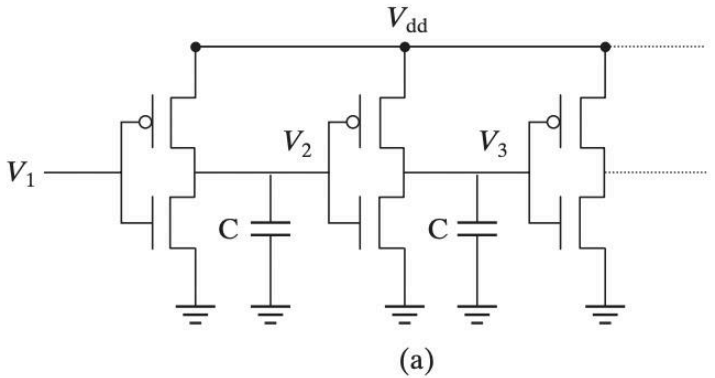
Inverter Speed and the Importance of I_{on}

- In a CMOS inverter chain, the propagation delay is the average of the delays of "pull-up" (rising V1 pulls down the output, V2) and "pull-down" (falling V2 pulls up the output V3)

$$\tau_d \approx \frac{CV_{dd}}{4}\left(\frac{1}{I_{onN}} + \frac{1}{I_{onP}}\right)$$

minimize / maximize

$$I_{on} \equiv I_{dsat}\Big|_{maximum\ |V_{gs}|}$$



- Clearly, in order to maximize circuit speed (the delay may be interpreted as RC), we need to maximize I_{on}!

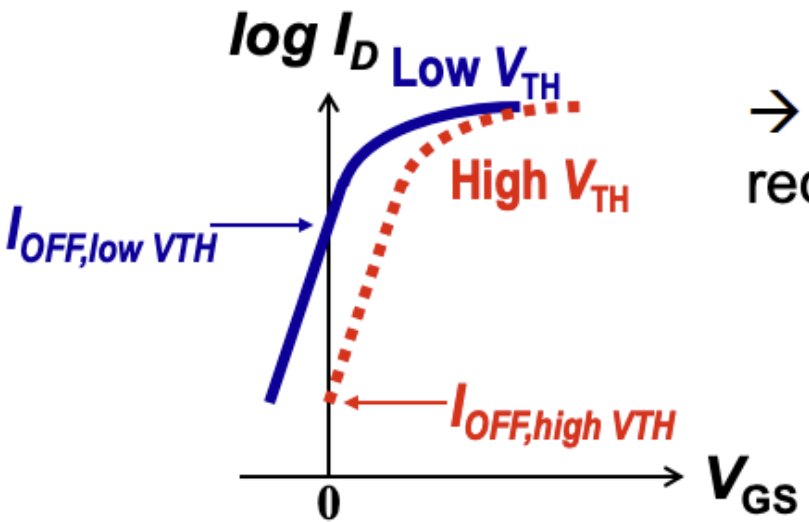
Sources: C. Hu *Modern Semiconductor Devices for Integrated Circuits*

Threshold Voltage Design Trade-Offs

- Low V_t is desirable for high on-state current, I_{on}:

$$I_{dsat} \propto (V_{DD} - V_t)^{\eta} \text{ for } 1 < \eta < 2$$

- But high V_t is needed for low off-state current, ~~I_{off}~~! This is not desirable because a large V_t (reduced $I_{on} = I_{dsat}$) degrades the circuit speed



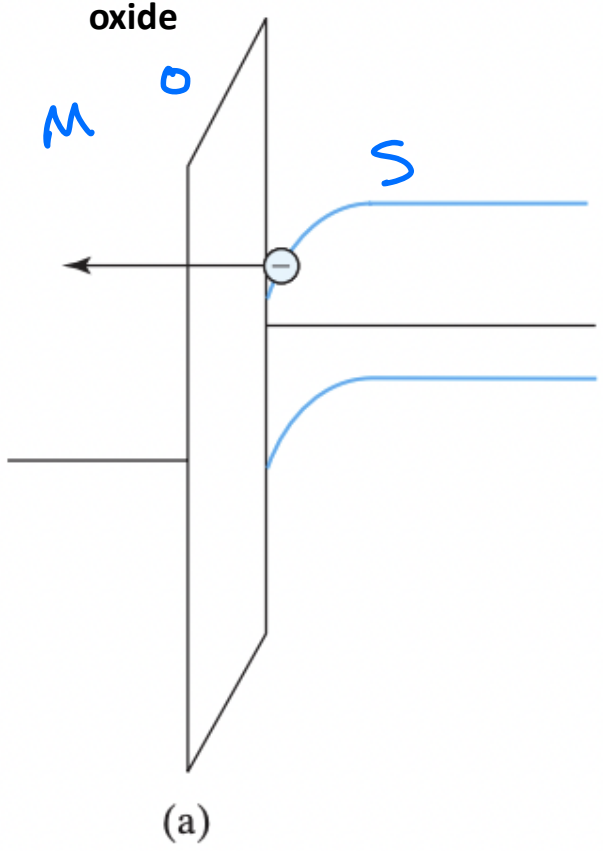
→ V_{TH} cannot be reduced aggressively.

Sources: White EECS42 Week 9a slides

Gate Tunneling Leakage

- An alternative way to reduce I_{off} (instead of altering V_t) is to reduce the subthreshold swing
- We can do this by increasing C_{ox} i.e., using a thin oxide
- But, when we make the SiO_2 too thin, carriers can tunnel through, causing gate leakage current!
- Oxide breakdown is another limiting factor: thin oxide -> large E. field -> material breakdown

Energy band diagram in inversion showing electron tunneling path through the gate oxide



Sources: C. Hu *Modern Semiconductor Devices for Integrated Circuits,*

High-K Dielectric

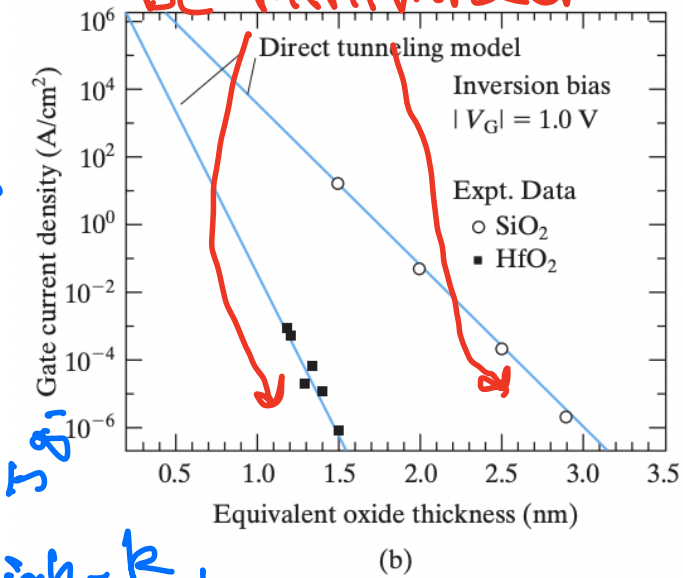
$C_{ox} = \frac{\epsilon_{HfO_2}\epsilon_0}{6nm} = \frac{\epsilon_{Si}\epsilon_0}{1\,nm}$

- Hence, engineers have developed $high-k\ dielectrics$ to replace SiO_2 (note $k = \epsilon_{relative}$)
  - Ex: HfO_2, ZrO_2, Al_2O_3
- These high-k dialectics have larger relative permittivity, meaning they can be made ~~thinner~~ thicker than SiO_2 to produce a given C_{ox}
- $\epsilon(HfO_2) = 24$, ~6x larger than $\epsilon(SiO_2) \simeq 3.9$
- A 6 nm thick HfO_2 film is equivalent to 1 nm thick SiO_2 in the sense that both films produce the same C_{ox}
- We say that this HfO_2 film has an equivalent oxide thickness or *EOT* of 1 nm
- The HfO_2 film presents a much thicker (albeit lower) tunneling barrier to the carriers
- The consequence is that the leakage current through HfO_2 is several orders of magnitude smaller than that through SiO_2!
- Disadvantages? Manufacturing is difficult for high-k, chemical reactions between films and Si substrate

We want to maximize C_{ox} to lower S to lower I_{off} without causing tunneling due to too thin dielectric!

$$EOT = t_{high-k}\left(\frac{k_{SiO_2}}{k_{high-k}}\right)$$

We want I_G to be minimized



Sources: C. Hu *Modern Semiconductor Devices for Integrated Circuits*

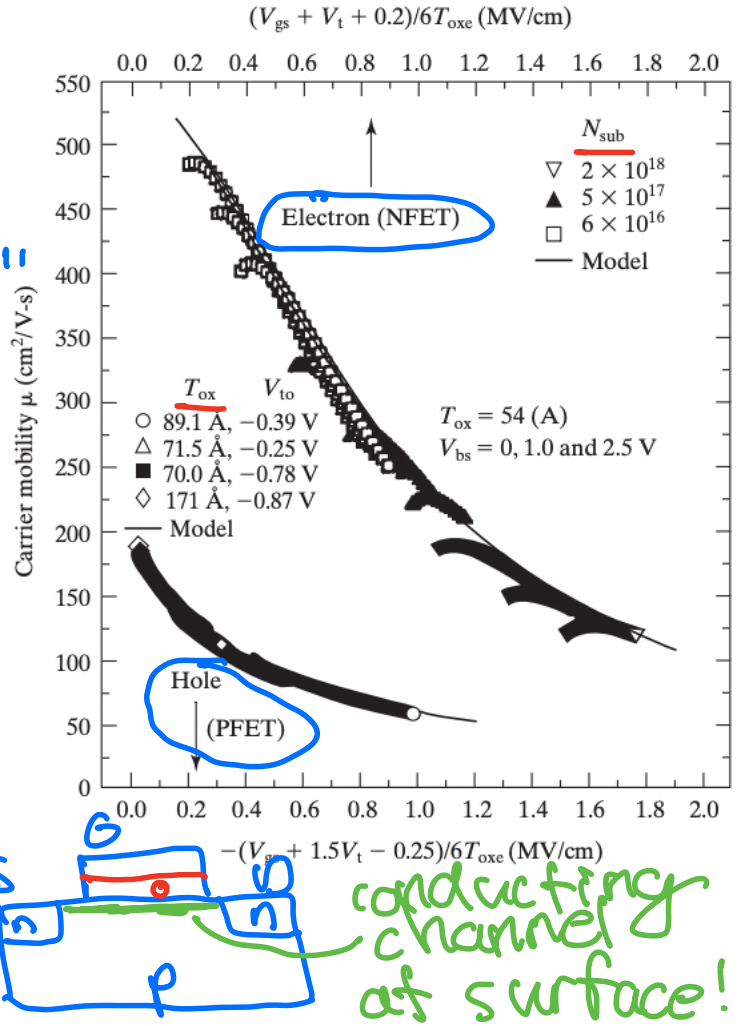
Effective Mobility in the Channel

- What is the true mobility in our MOSFET channel?
- Can we look it up in the bulk silicon charts? No!

We are looking at the semi "surface"

- Scattering mechanisms affect the mobility in the channel (which is very close to the interface between two dissimilar materials!)
  - Charged impurity (Coulomb) scattering
  - Lattice vibration (phonon) scattering
  - Surface roughness scattering!

- Instead, e- and h+ surface mobilities in the channel are determined by $V_{gs}, V_t, and\ t_{ox}(equivalent)$



Sources: C. Hu *Modern Semiconductor Devices for Integrated Circuits*, E. Pop ECE 340 Slides

MOSFET Speed (Switching)

- Cutoff frequency f_{max} is frequency where MOSFET no longer amplifies input (gate) signal

$$f_{max} = \frac{g_m}{2\pi C_{ox}} = \frac{\mu_{eff}}{2\pi L^2}(V_{GS} - V_T) \propto \frac{1}{L^2}$$

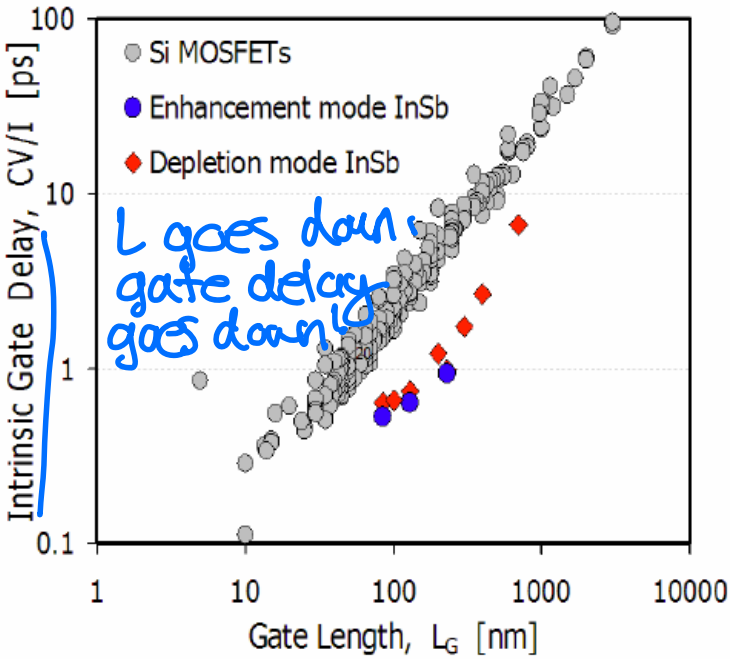
- Obtained by considering high-freq. small-signal model with output shorted, then finding freq. where $|i_{out}/i_{in}| = 1$

No amplification! ($=1$)

- Something we already knew qualitatively → higher MOSFET operating frequency achieved by decreasing channel length, L, increasing mobility, μ_{eff}

- Smaller = faster for devices (though parasitics play a big role in realistic circuits)

series resistance
contact resistance



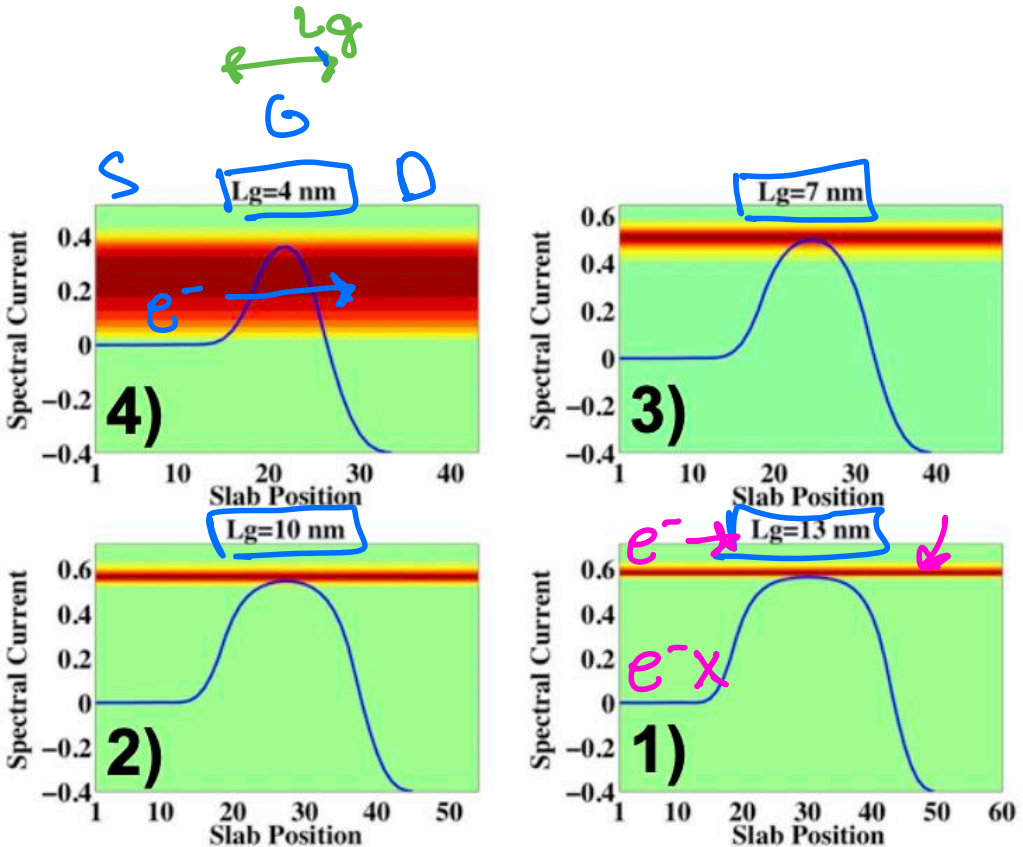
Sources: E. Pop ECE 340 Slides

Quantum Tunneling

- The limit to barrier control in short channel devices (small gate length) is quantum tunneling

- Recall: for thin potential barriers, there's some probability carriers can penetrate the barrier even though they don't have enough kinetic energy to go over the barrier

- Sound familiar? We discussed that a major hurdle to continued gate length scaling is that we are now entering the quantum (tunneling) regime in our first lecture!



from M. Luisier, ETH Zurich / Purdue

Sources: M. Lundstrom ECE 305 Slides