Digital Integrated Circuits
(83-313)

Lecture 3:

MOSFET Modeling

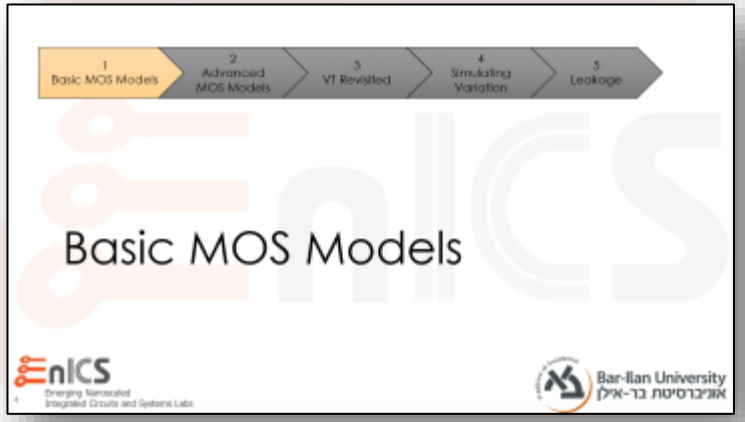
27 April 2020

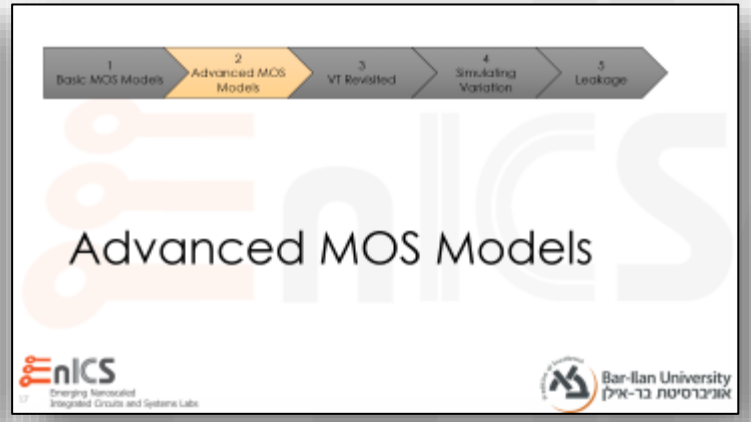
Emerging Nanoscaled
Integrated Circuits and Systems Labs

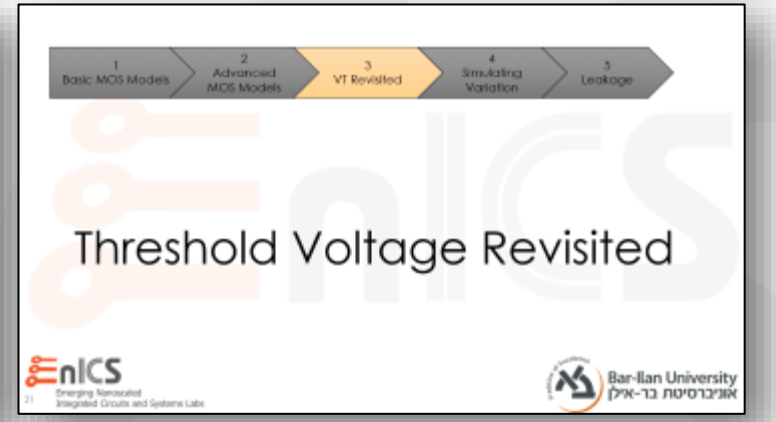
Bar-Ilan University
אוניברסיטת בר-אילן

Disclaimer: This course was prepared, in its entirety, by Adam Teman. Many materials were copied from sources freely available on the internet. When possible, these sources have been cited; however, some references may have been cited incorrectly or overlooked. If you feel that a picture, graph, or code example has been copied from you and either needs to be cited or removed, please feel free to email adam.teman@biu.ac.il and I will address this as soon as possible.

Lecture Content







© Adam Teman, 2020

Basic MOS Models

TCAD vs. Compact Models

- **Technology CAD (TCAD) is a simulation environment for accurately simulating device behavior:**
  - Provide a process "recipe" and device layout
  - Produce IV or CV curves through device simulator
  - Used to predict device and process physics
  - Takes 1hr-1day per IV curve and 100s MB RAM per transistor
- **Compact models (a.k.a. SPICE models or ECAD) are simple models used for circuit simulation**
  - Provide a set of equations that SPICE uses to calculate IV or CV curves
  - Should take <100us per IV curve and a few KB per transistor
  - Usually extracted empirically from measurements

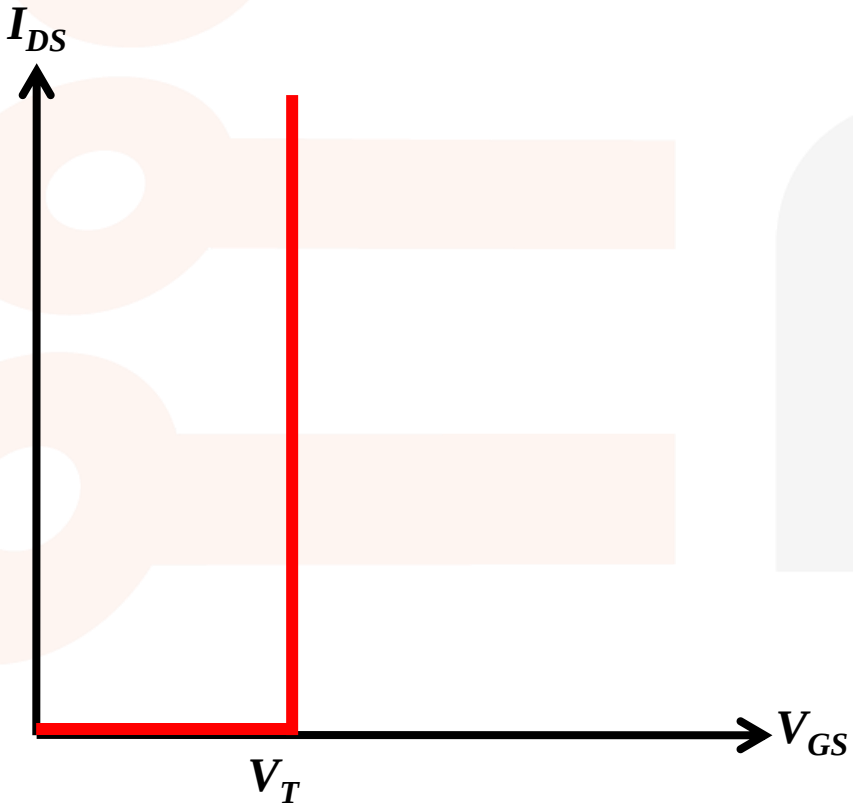
TCAD is too **slow** and **memory hungry** to be used for circuit simulation!

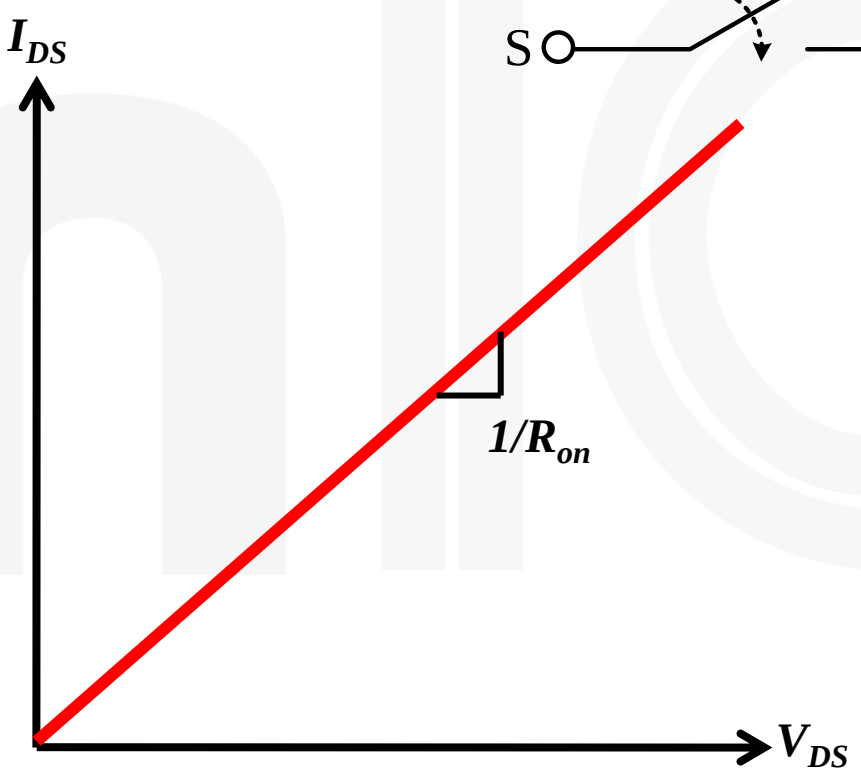
SPICE uses ***compact models*** for calculating device behavior

© Adam Teman, 2020

Switch Model

- The most simple MOSFET model is the **Switch Model**.

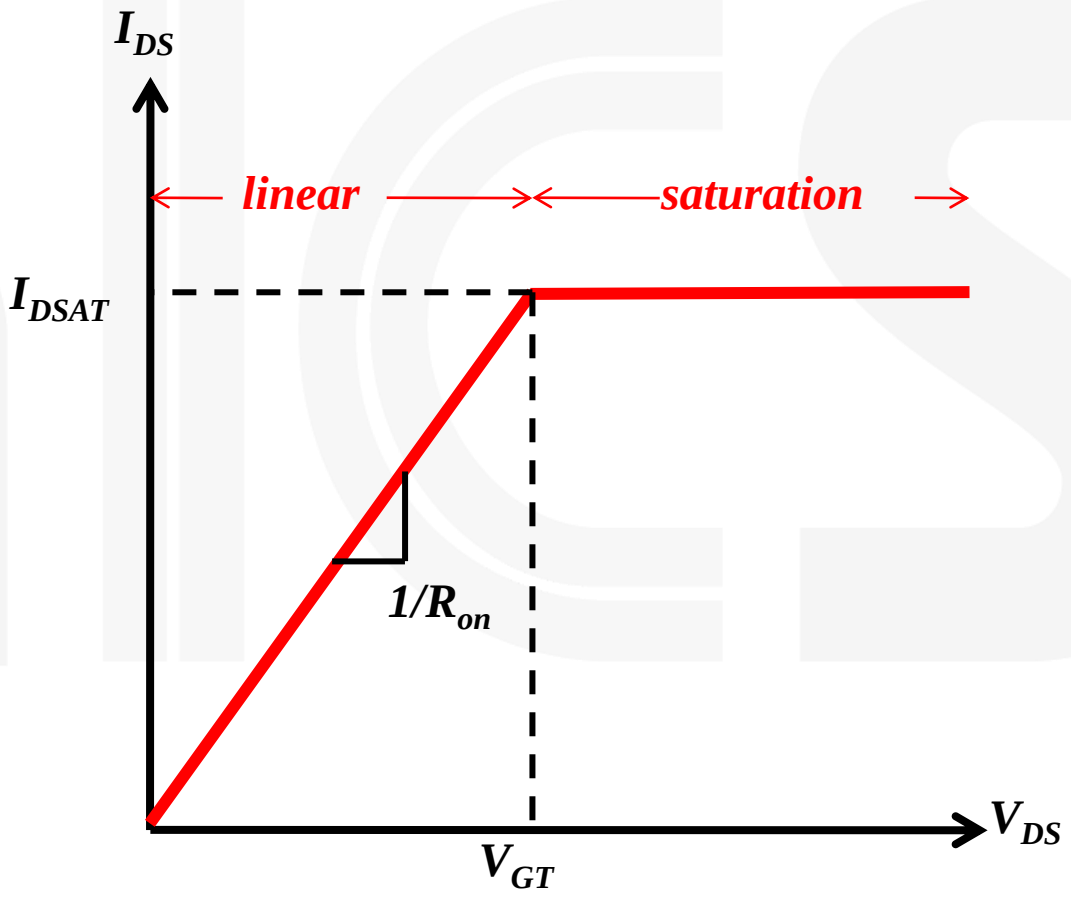




© Adam Teman, 2020

The Piece-Wise Linear Model

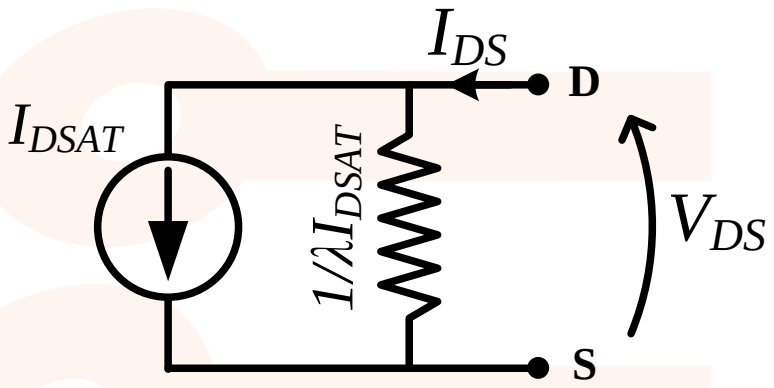
- As we know, when the channel pinches off, the current saturates.
- This can be depicted with the simple Piece-Wise Linear Switch Model

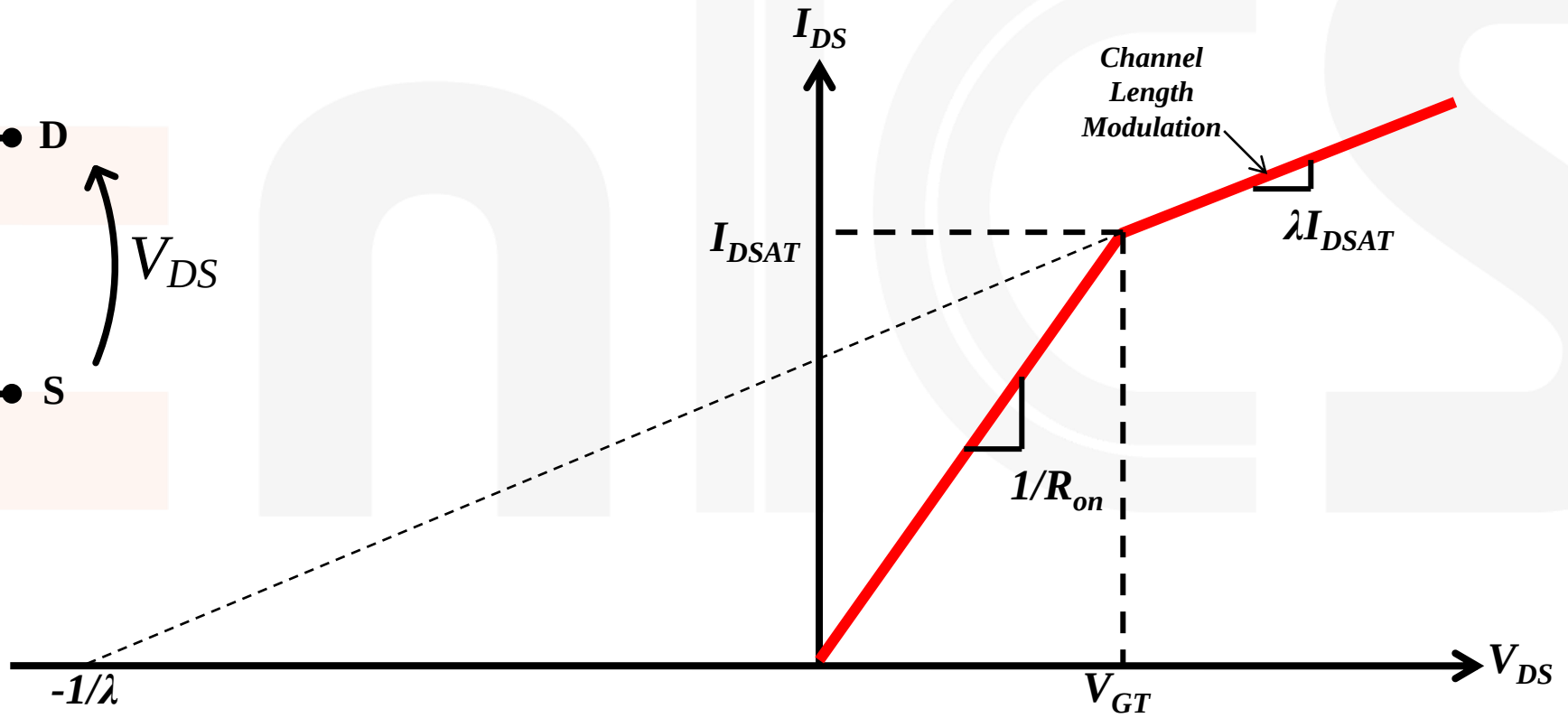


© Adam Teman, 2020

Adding Channel Length Modulation

- **Channel Length Modulation** modeled as a finite output resistance, causes a saturation current dependence on V_{DS}.





© Adam Teman, 2020

Square Law (Shockley) Model

- To get a more accurate model, we already are familiar with the Shockley or Square Law Model.
- Current is just charge times velocity, so at any point, x, along the channel:

$$I_D(x) = -\nu(x)Q(x)Wdx$$

- We found that charge can be approximated as:

$$Q(x) = -C_{ox}\left[V_{GS} - V_{CS}(x) - V_T\right]$$

- And the velocity is the mobility times the electrical field:

$$\nu(x) = -\mu E(x) = \mu_n \frac{dV}{dx}$$

© Adam Teman, 2020

Square Law (Shockley) Model

- **So we get:**

$$I_D dx = \mu_n C_{ox} W \left(V_{GS} - V - V_T \right) dV$$

- **And integrating from source to drain, we get**

$$I_{DS} = \int_0^L I_D dx = \int_0^{V_{DS}} \mu_n C_{ox} W \left(V_{GS} - V - V_T \right) dV = \mu_n C_{ox} \frac{W}{L} V_{DS} \left(V_{GS} - V_T - \frac{1}{2} V_{DS} \right)$$

- **At pinch-off ($V_{DS}=V_{GS}-V_T$), the voltage over the channel is constant, so we get:**

$$I_{DSAT} = \mu_n C_{ox} \frac{W}{2L} \left(V_{GS} - V_T \right)^2$$

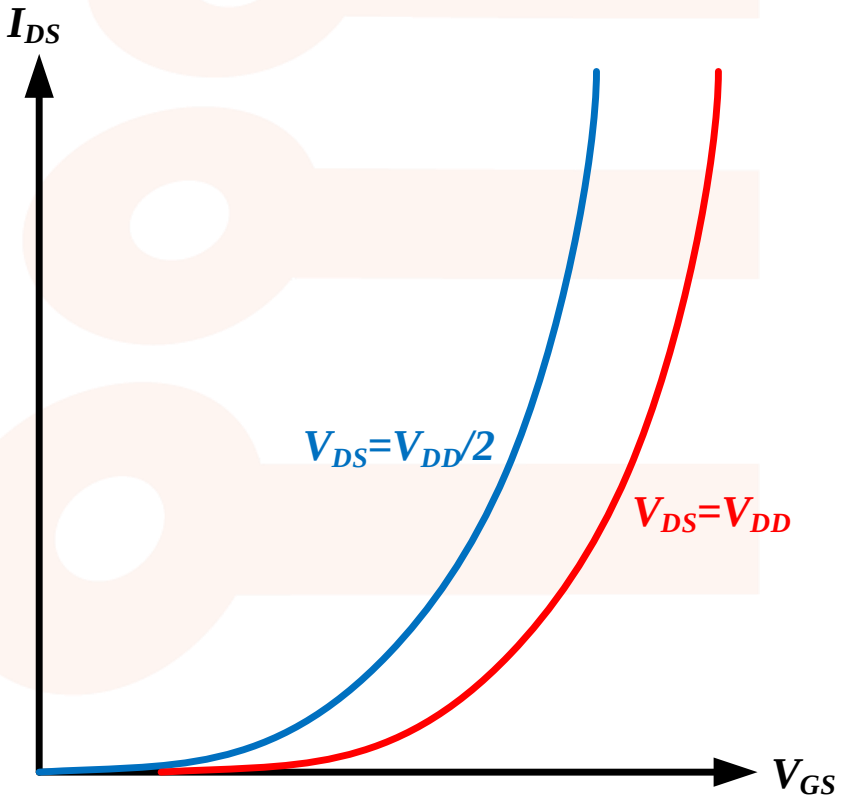
- **This is where the "Square-Law" name comes from.**

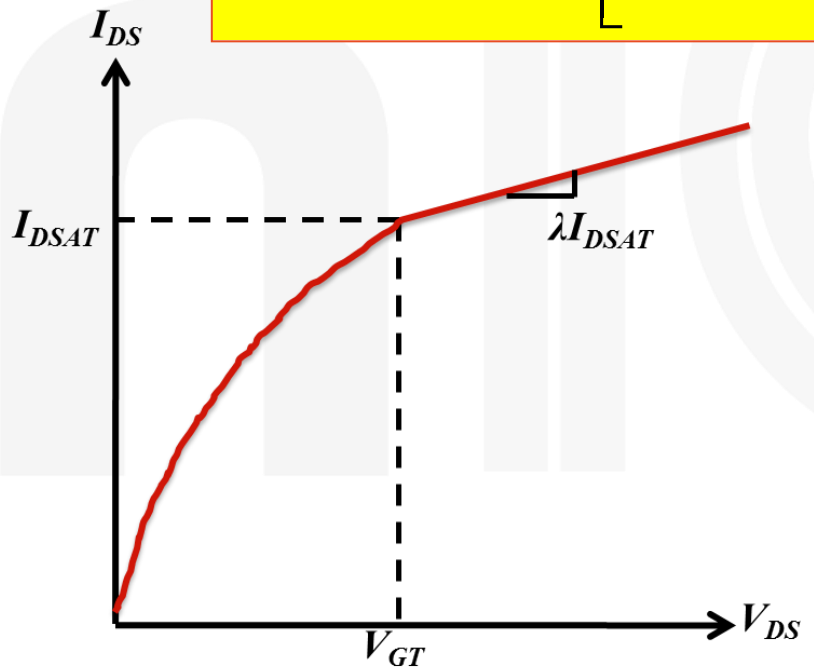
© Adam Teman, 2020

Square Law (Shockley) Model

- **Replacing V_{DS} with $V_{DSeff}=\min(V_{GS}-V_T, V_{DS})$ we get:**

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

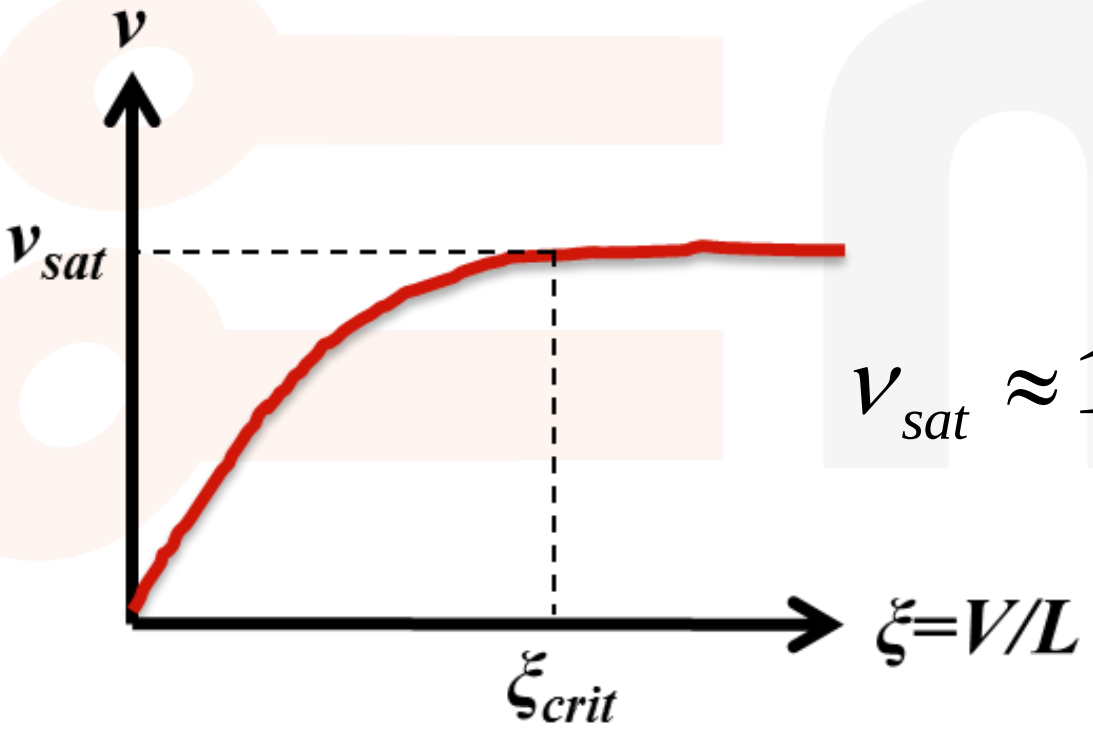




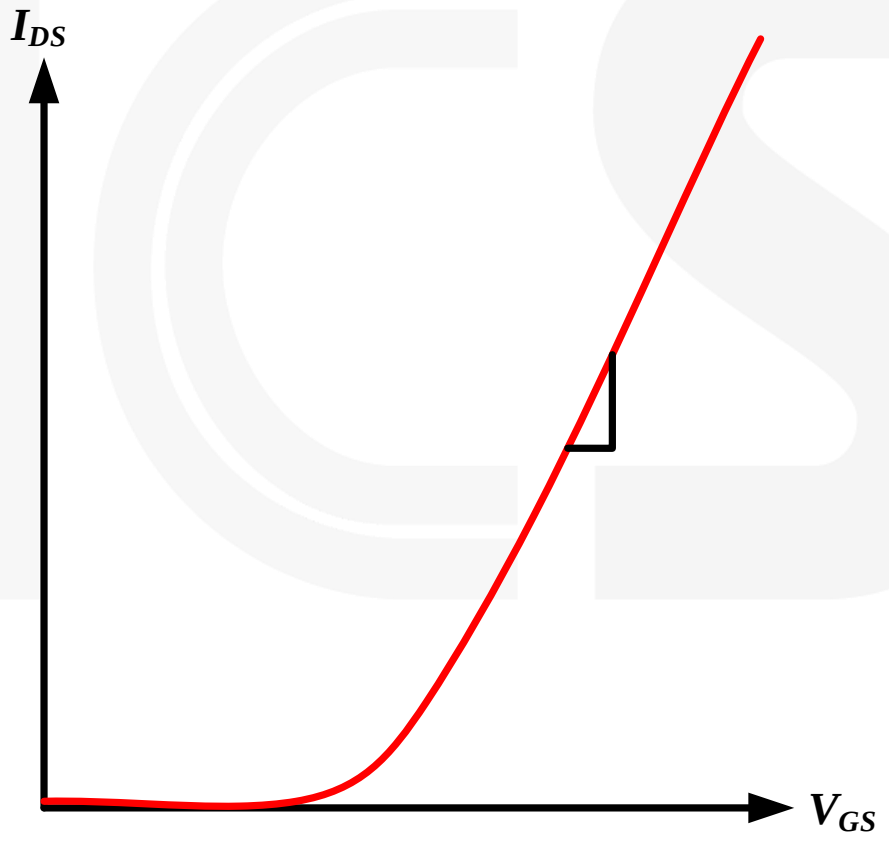
© Adam Teman, 2020

The Velocity Saturation Model

- However, when looking at a short channel device, we see a linear dependence on V_{GS}.
- This can be attributed to Velocity Saturation.



$$v_{sat} \approx 10^5 \, m/s$$

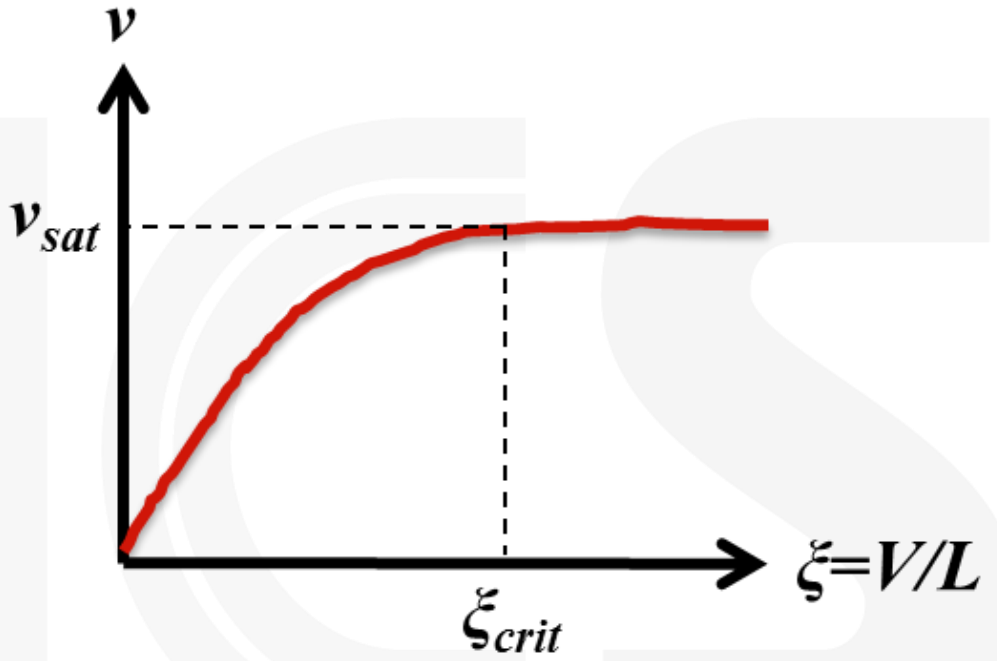


© Adam Teman, 2020

The Velocity Saturation Model

- **A good approximation of the mobility curve is:**

$$v = \begin{cases} \dfrac{\mu\xi}{1+\xi/\xi_{crit}} & \xi < \xi_{crit} \\ v_{sat} & \xi > \xi_{crit} \end{cases}$$



- **For continuity:**

$$\xi_{crit} = \frac{2v_{sat}}{\mu}$$

- **After integration, we get:**

$$I_{DS} = \frac{\mu_n C_{ox}}{1+V_{DS}/\xi_{crit}L}\frac{W}{L}\left[(V_{GS}-V_T)V_{DS}-\frac{V_{DS}^2}{2}\right]$$

© Adam Teman, 2020

The Velocity Saturation Model

- This is hard to use, but we can reach an important conclusion.
- We found that:

$$I_{DS} = \frac{\mu_n C_{ox}}{1 + \frac{V_{DS}}{\xi_{crit} L}} \frac{W}{L}\left[(V_{GS} - V_T)V_{DS} - \frac{V_{DS}^2}{2}\right]$$

- And we know that for a velocity saturated device:

$$I_{DS} = WC_{ox}(V_{GS} - V_{DSAT} - V_T)v_{sat}$$

- Equating, we get:

$$V_{DSAT} = \frac{(V_{GS} - V_T)\xi_{crit} L}{(V_{GS} - V_T) + \xi_{crit} L}$$

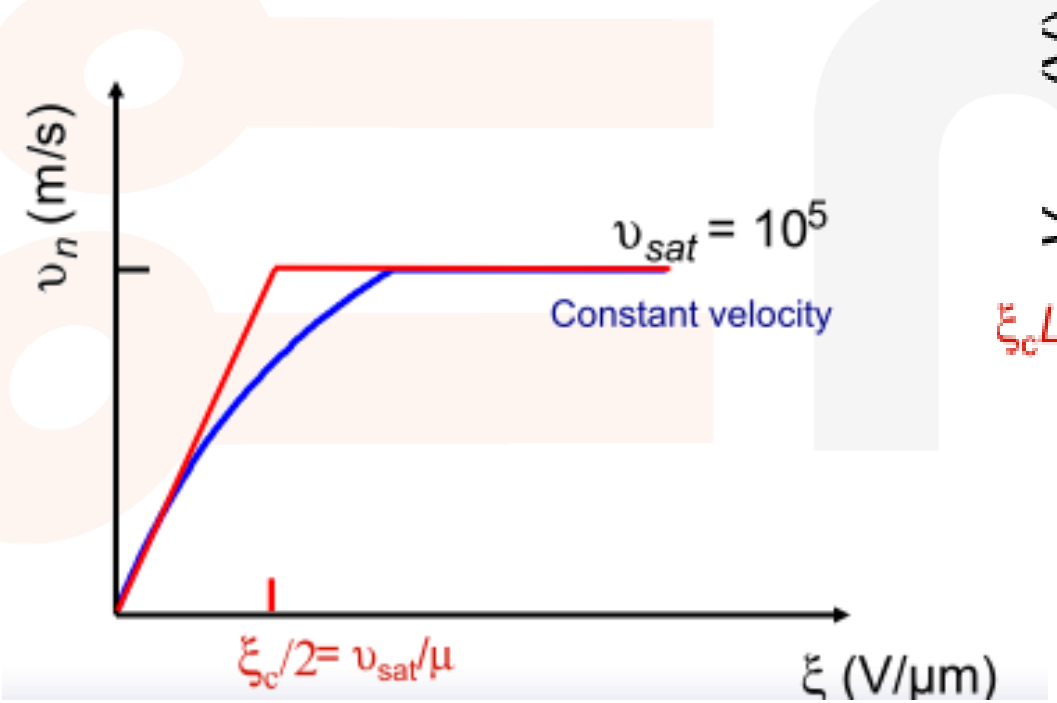
$$V_{DSAT}(\xi_{crit} L >> V_{GT}) = V_{GS} - V_T \Rightarrow pinch\,off$$

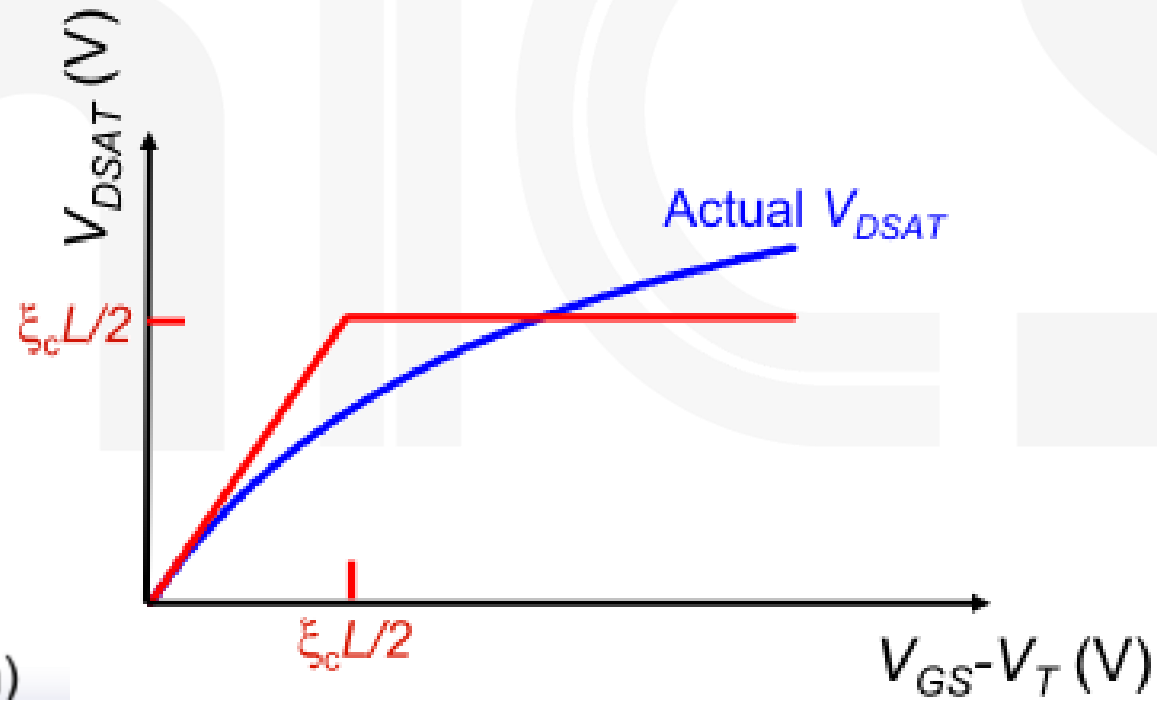
$$V_{DSAT}(\xi_{crit} L << V_{GT}) = \xi_{crit} L \Rightarrow vel\,sat$$

© Adam Teman, 2020

The Unified Model for Hand Analysis

- **A few simple estimations will make the V-Sat model more user-friendly:**
  - The mobility is piecewise linear, saturating at $\xi > \xi_{crit}/2$
  - V_{DSAT} is piecewise linear, saturating at $V_{DSAT} = \xi_{crit} L/2$, when $V_{GT} > \xi_{crit} L/2$





© Adam Teman, 2020

The Unified Model for Hand Analysis

- **This brings us to the Unified Model:**

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

$$V_{DSeff} = \min(V_{GS} - V_T, V_{DS}, V_{DSAT})$$

$$V_{DSAT} = \frac{\xi_{crit} L}{2} \qquad \xi_{crit} = \frac{2v_{sat}}{\mu}$$

© Adam Teman, 2020

Advanced MOS Models

Emerging Nanoscaled Integrated Circuits and Systems Labs

Bar-Ilan University
אוניברסיטת בר-אילן

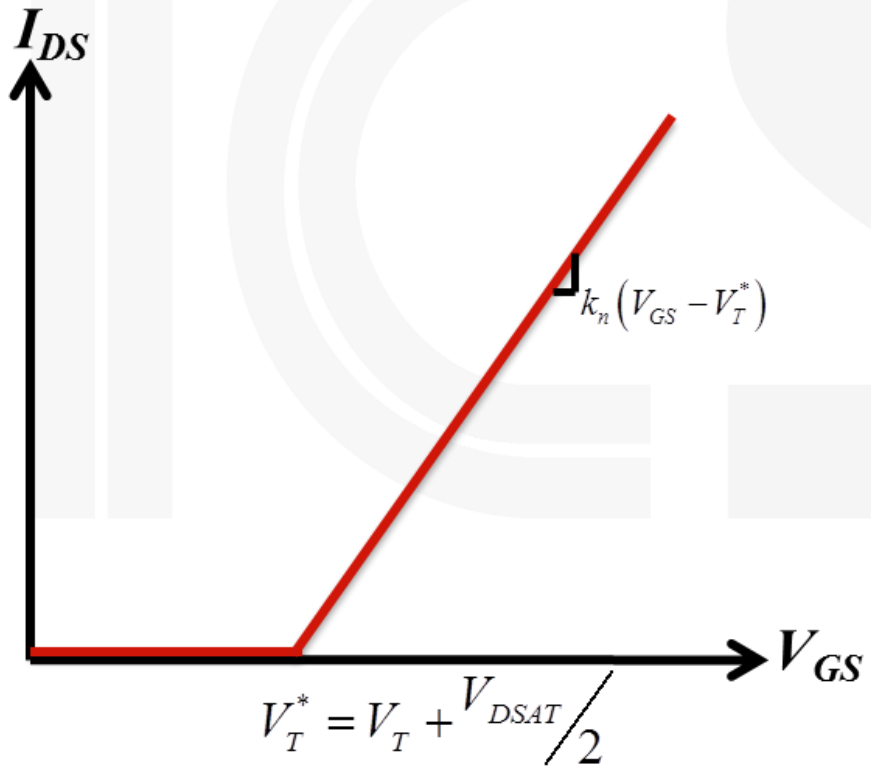
VT* Model

- Sometimes we want to use a really simple model.
- We can assume that if the transistor is on, it's velocity saturated.

$$I_{DS} = k_n\left[(V_{GS}-V_T)V_{DSAT} - \frac{V_{DSAT}^{\ 2}}{2}\right] =$$

$$= k_n\left[\left(V_{GS} - \underbrace{\left(V_T + \frac{V_{DSAT}}{2}\right)}_{V_T^*}\right)V_{DSAT}\right]$$

$$I_{DS} = \begin{cases} 0 & V_{GS} < V_T^* \\ k_n\left(V_{GS}-V_T^*\right)V_{DSAT} & V_{GS} > V_T^* \end{cases}$$



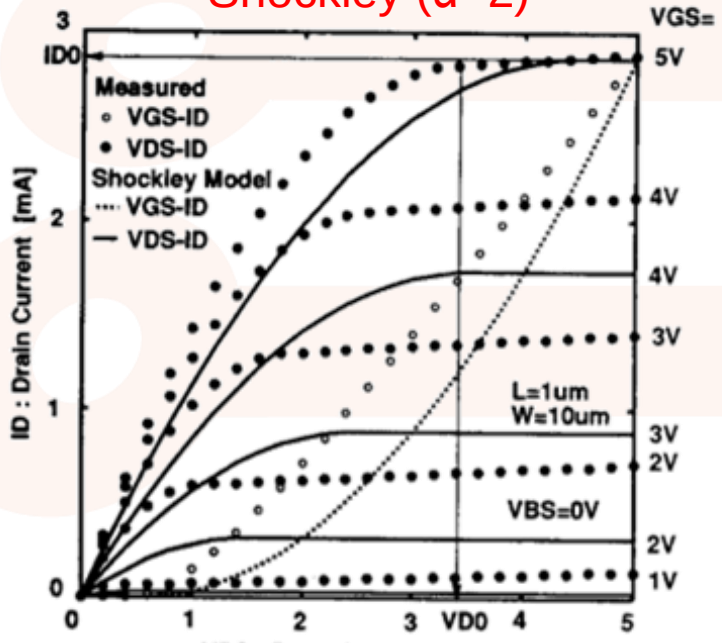
© Adam Teman, 2020

The Alpha Power Law Model

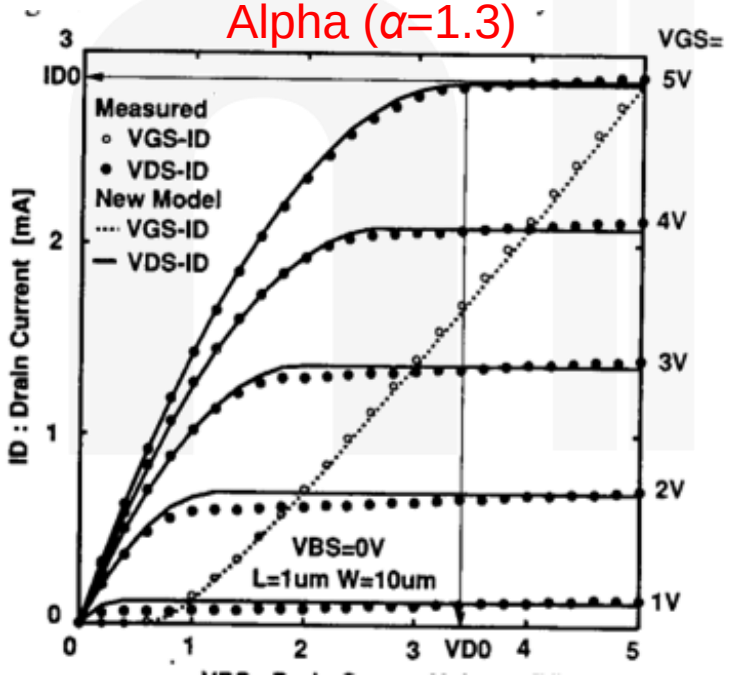
- **Sakurai found that by changing the exponent of the square law, a better fit can be found with simple calculations.**

$$I_{DSAT}(\alpha) = \mu_n C_{ox} \frac{W}{2L}(V_{GS} - V_T)^{\alpha}$$

Shockley (α=2)



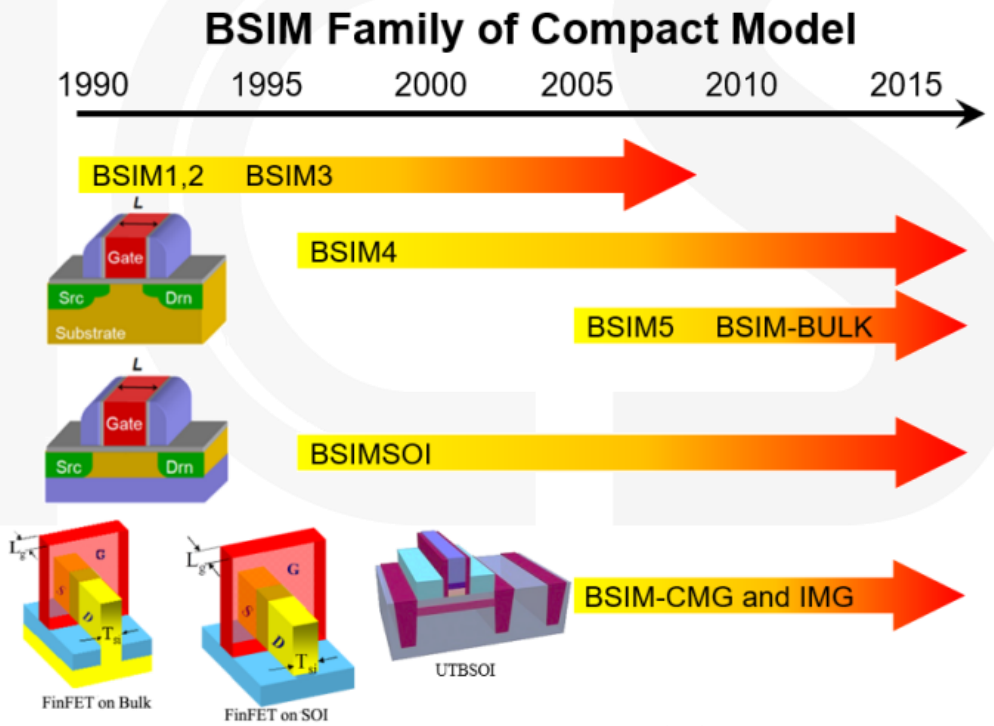
Alpha (α=1.3)



Iqbal Teman, 2020

BSIM and Newer Models

- **BSIM (**Berkeley Short-channel IGFET Model**) is the primary compact model family used for SPICE simulation for the last three decades.**
- **These model use hundreds of parameters to achieve a good fit.**
- **BSIM4 is the main model for bulk CMOS**
  - Takes into account most physical effects as well as many fitting parameters.
- **The Compact Model Coalition (CMC) chooses, maintains and promotes new models**
  - Additional models include EKV, PSP, and models for non-MOS devices.



Source: Yen Ki Lin, UC Berkeley

© Ad 2020

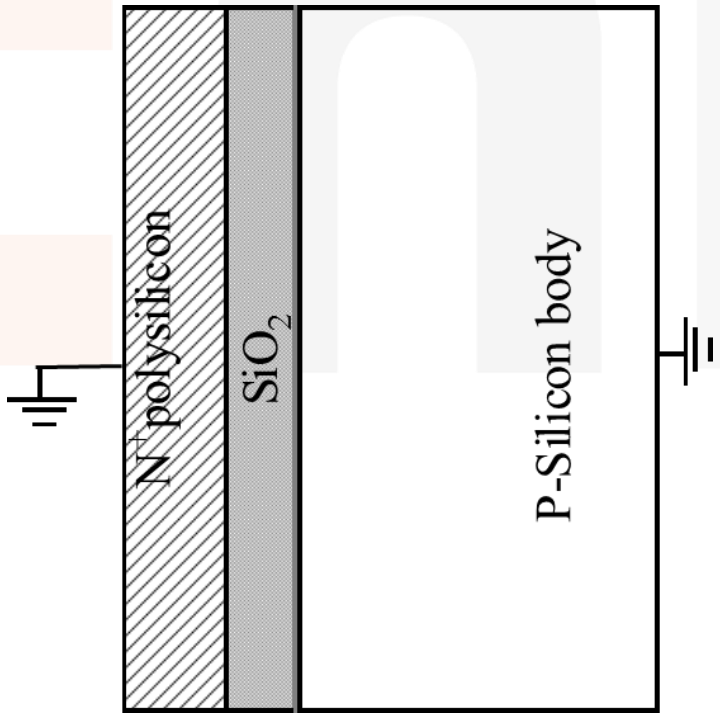
Threshold Voltage Revisited

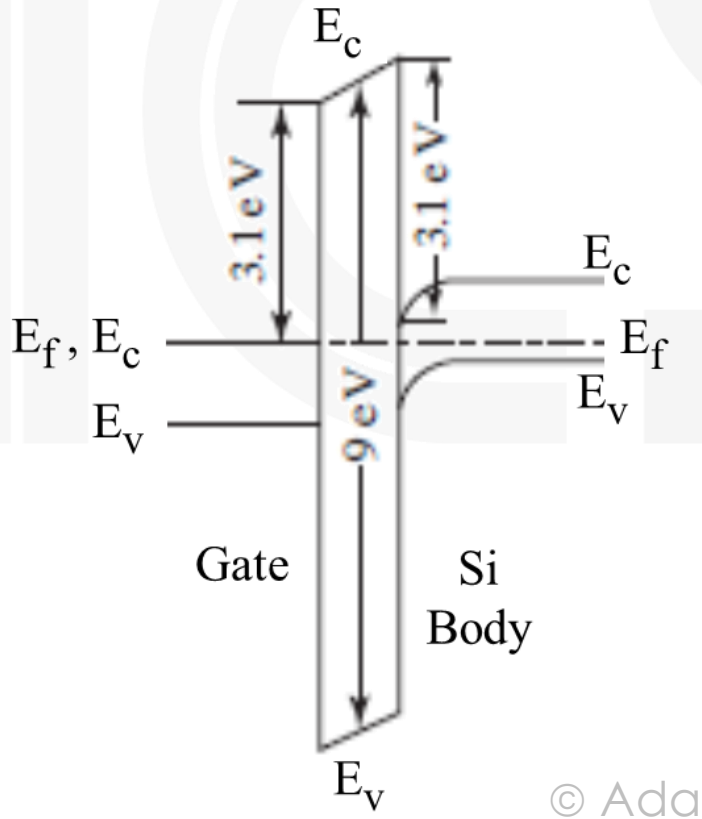
Emerging Nanoscaled Integrated Circuits and Systems Labs

Bar-Ilan University
אוניברסיטת בר-אילן

Energy Band Diagrams

- To understand the threshold voltage and other secondary effects of the MOS device, we often use energy band diagrams.
- The first approach is looking in from the gate:

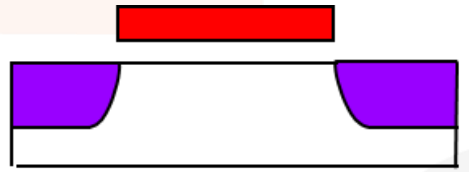


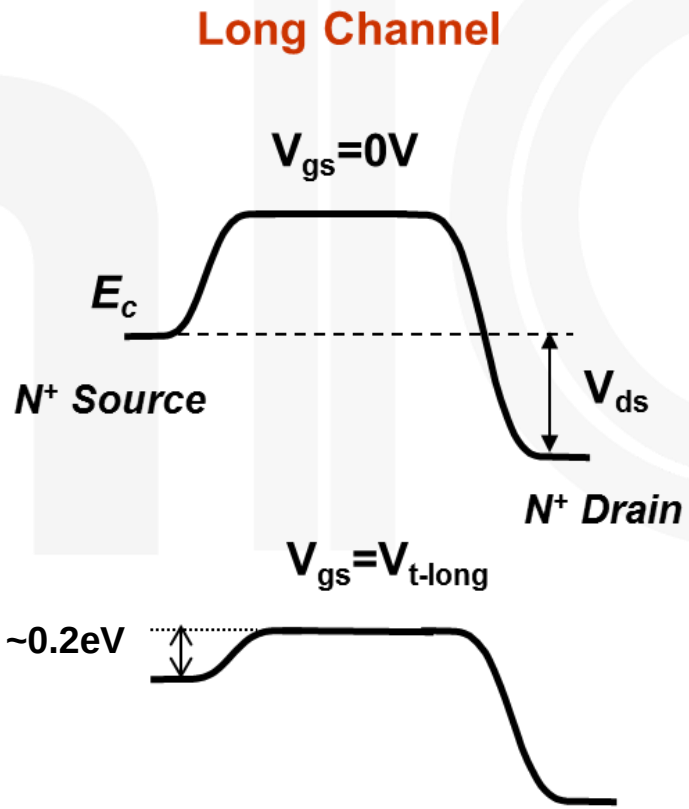


© Adam Teman, 2020

Energy Band Diagrams

- **The second approach is looking <u>from the source to the drain</u>.**





© Adam Teman, 2020

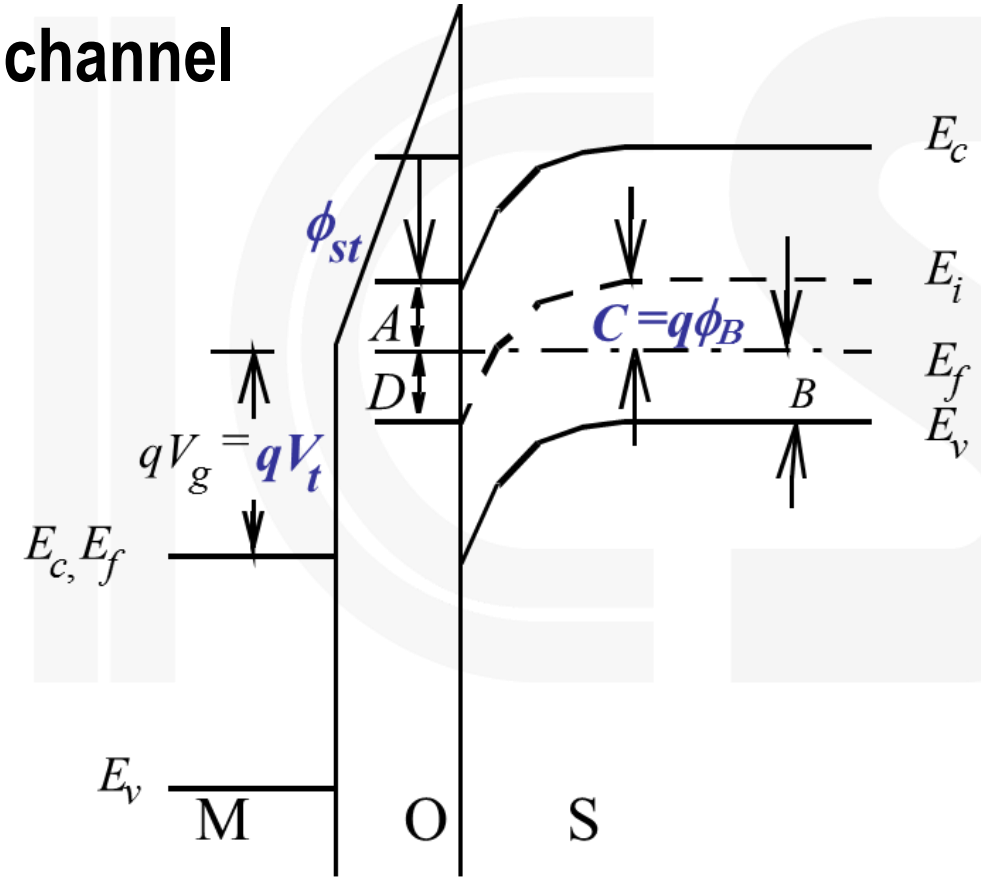
Threshold Voltage - Basic Theory

- **The basic definition of threshold voltage is the gate voltage (V_G) required to invert the channel**

$$V_{T0} = \Phi_{MS} - 2\Phi_F - \frac{Q_{OX}}{C_{ox}} - \frac{Q_{dep}}{C_{ox}} - \frac{Q_I}{C_{ox}}$$

$$Q_{dep} = \sqrt{2qN_A\varepsilon_{si}\left|-2\Phi_F\right|}$$

$$\Phi_F = -\phi_T \ln\frac{N_A}{n_i} \qquad \phi_T \equiv \frac{kT}{q}$$



© Adam Teman, 2020

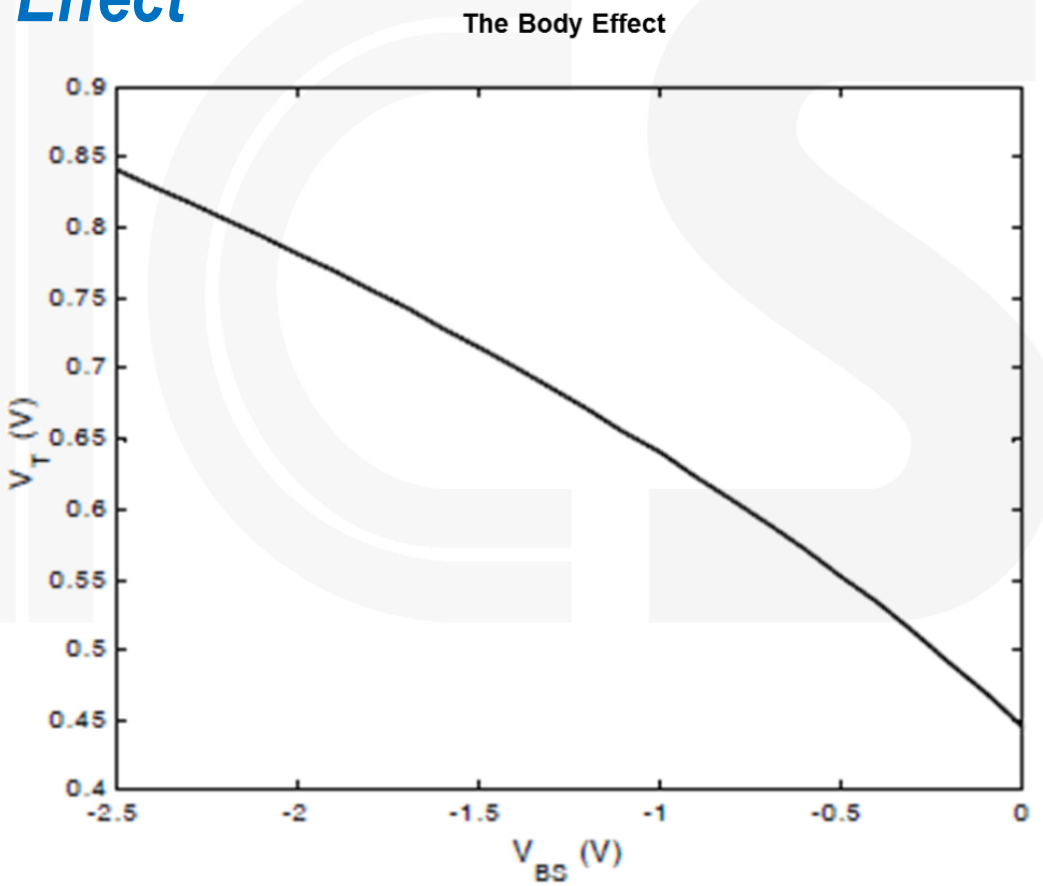
Body Effect

- The appearance of a voltage difference between the source and body (V_{SB}) is known as "*The Body Effect*"
- This can be modeled by the additional charge that needs to be depleted.

$$Q_{dep} = \sqrt{2qN_A\varepsilon_{si}\left(\left|-2\Phi_F + V_{SB}\right|\right)}$$

$$V_T = V_{T0} + \gamma\left(\sqrt{\left|-2\Phi_F + V_{SB}\right|} - \sqrt{\left|-2\Phi_F\right|}\right)$$

$$V_{T0} \equiv \Phi_{MS} - 2\Phi_F - \frac{Q_{ox}}{C_{ox}} - \frac{Q_{dep0}}{C_{ox}} \qquad \gamma = \frac{\sqrt{2q\varepsilon_{si}N_A}}{C_{ox}}$$



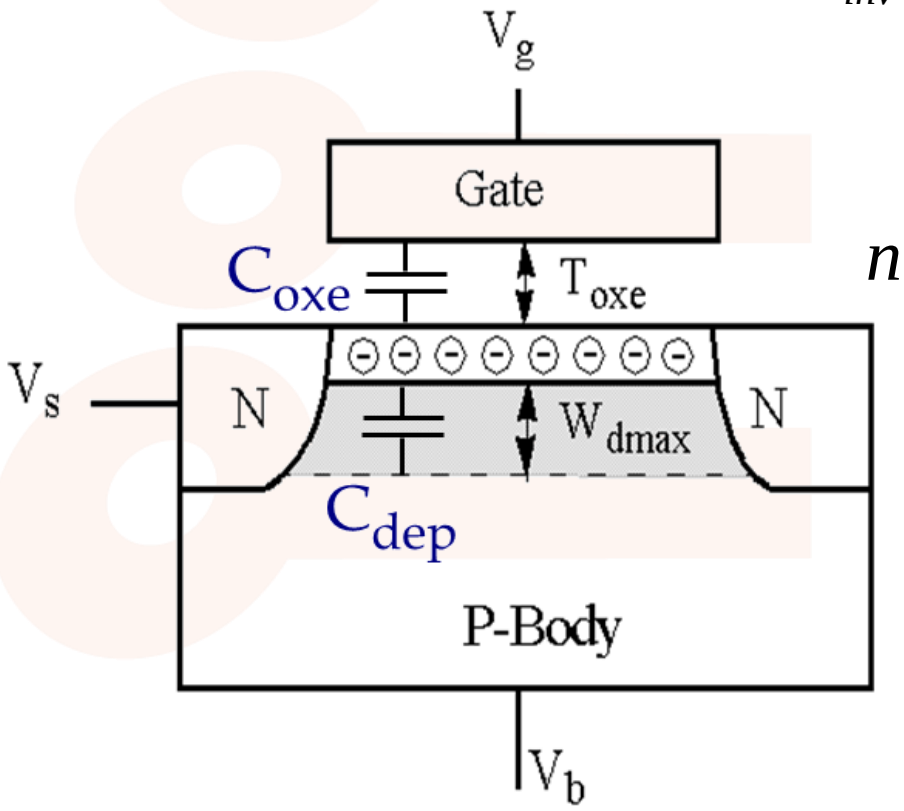
© Adam Teman, 2020

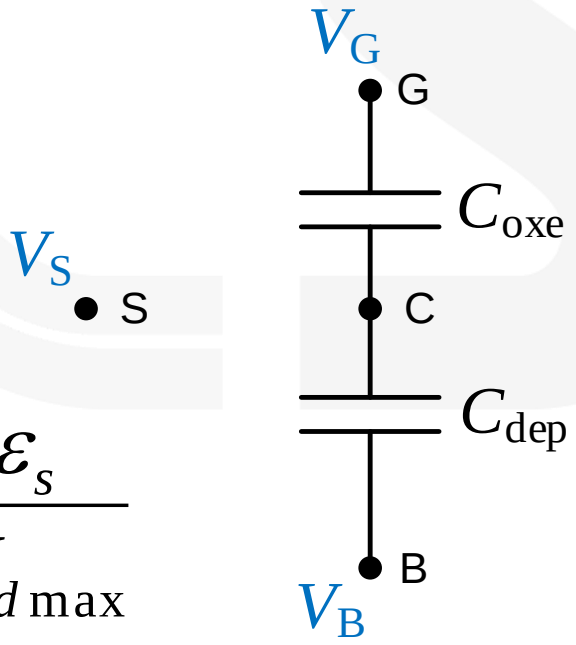
Modern Body Effect

- **A different approach is to look at the capacitive voltage divider between the gate and body (C_{GB})**

$$Q_{inv} = -C_{oxe}\left(V_{GS} - V_{CS} - V_{T0}\right) + C_{dep}\left(V_{SB} + V_{CS}\right)$$
$$= -C_{oxe}\left(V_{GS} - nV_{CS} - V_{T0}\right)$$

$$n \triangleq 1 + \frac{C_{dep}}{C_{oxe}} = 1 + \frac{3T_{oxe}}{W_{d\max}}$$





$$C_{dep} = \frac{\varepsilon_s}{W_{d\max}}$$

© Adam Teman, 2020

Modern Body Effect

- This can be shown to redefine V_T as:

$$V_T(V_{SB}) = V_{T0} + \frac{C_{dep}}{C_{oxe}} V_{SB}$$

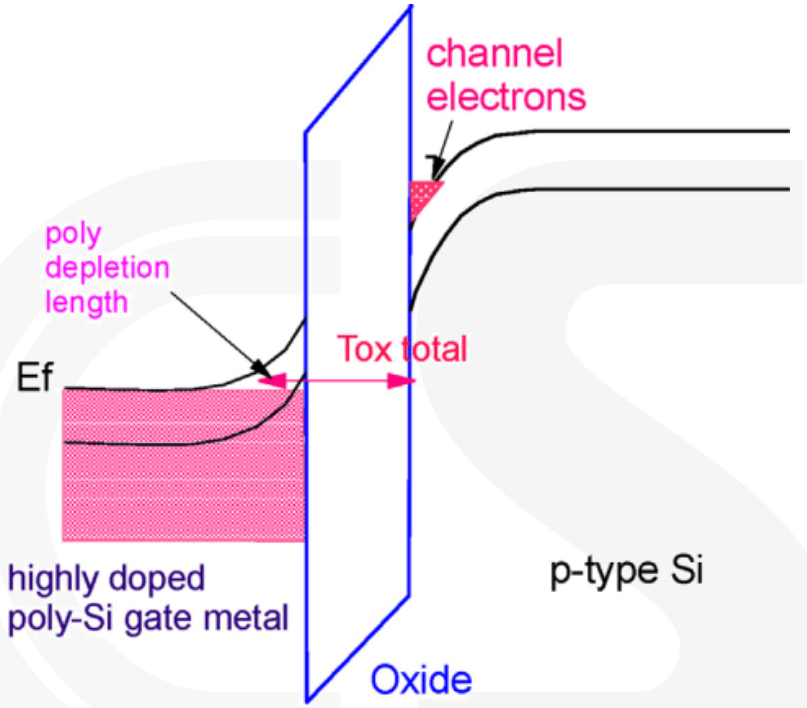
- In modern technologies, C_{dep}/C_{oxe} is a constant, so V_T is *linearly dependent* on V_{SB}!

© Adam Teman, 2020

Poly Depletion and Channel Depth

The threshold voltage is affected by two additional factors that we have disregarded until now:

- **Polysilicon Depletion**
  - Since polysilicon is, itself, a semiconductor, the depletion layer into the poly effectively increases the oxide thickness.
- **Channel Depth**
  - Since the channel is not a 2-dimensional line along the surface, the oxide thickness is essentially increased.



$$n \triangleq 1 + \frac{C_{dep}}{C_{oxe}} = 1 + \frac{3T_{oxe}}{W_{d\max}}$$

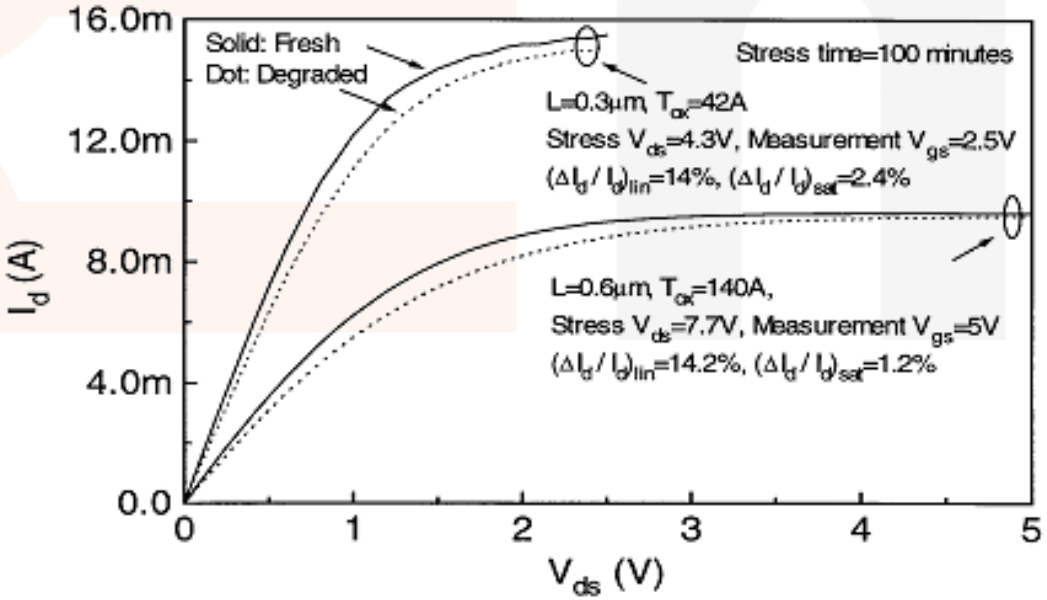
© Adam Teman, 2020

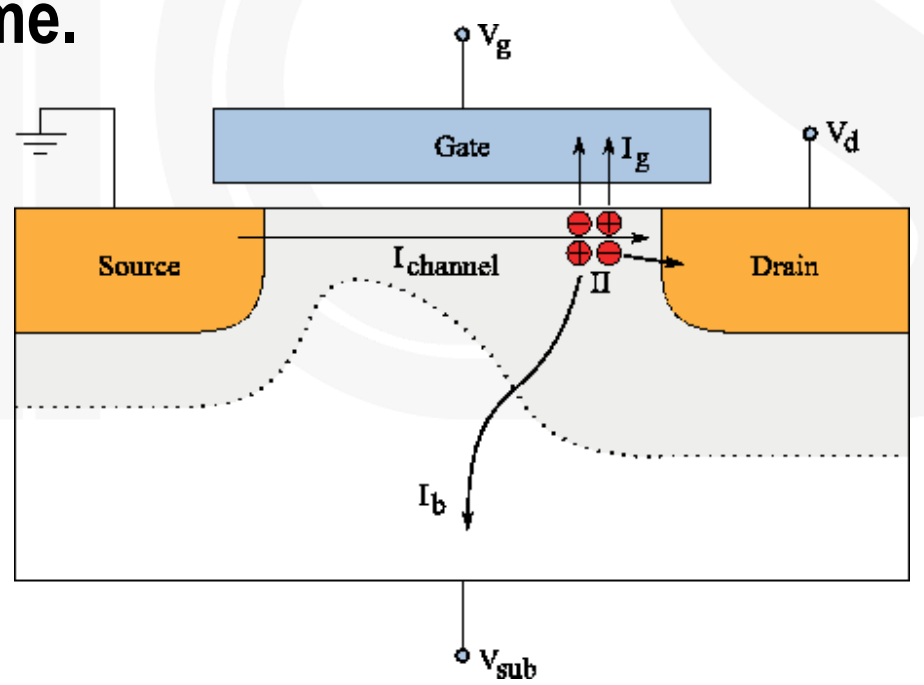
Hot Carrier Effects

- **Electrons can get so *fast* that they can tunnel into the gate oxide and increase the threshold voltage.**

$$V_{T0} = \Phi_{MS} - 2\Phi_F - \frac{Q_{OX}}{C_{ox}} - \frac{Q_{dep}}{C_{ox}} - \frac{Q_I}{C_{ox}}$$

- **This is a reliability issue as it happens over time.**

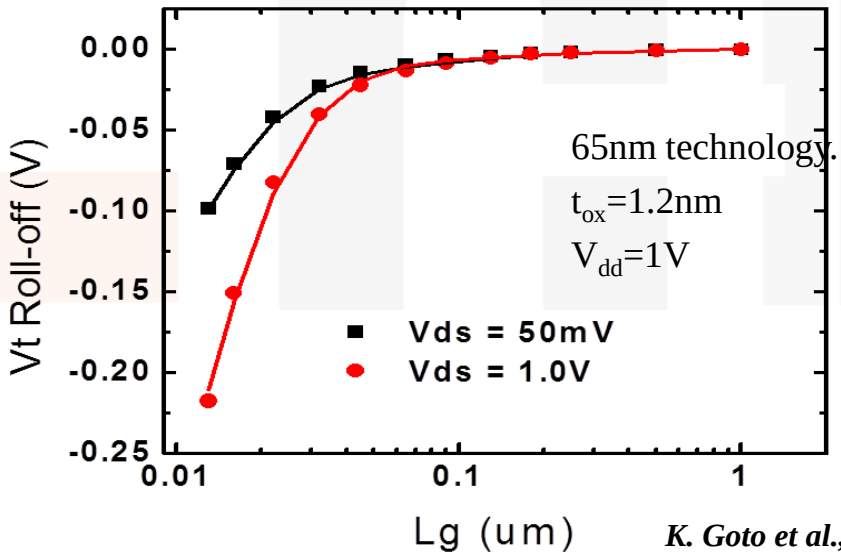




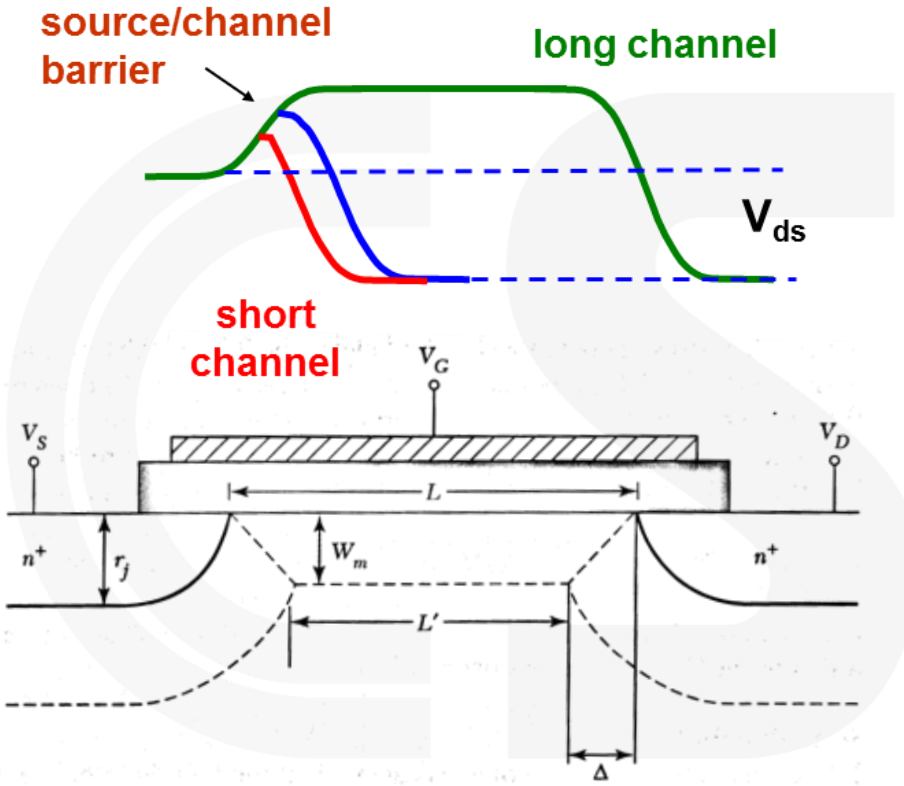
© Adam Teman, 2020

V_T Roll Off (Short Channel Effect)

- As channel length is reduced, effective channel length is reduced by depletion regions.
- A trapezoid is created under the gate, dividing the channel into the region controlled by the gates and by the drain.
- In essence, V_T is reduced.



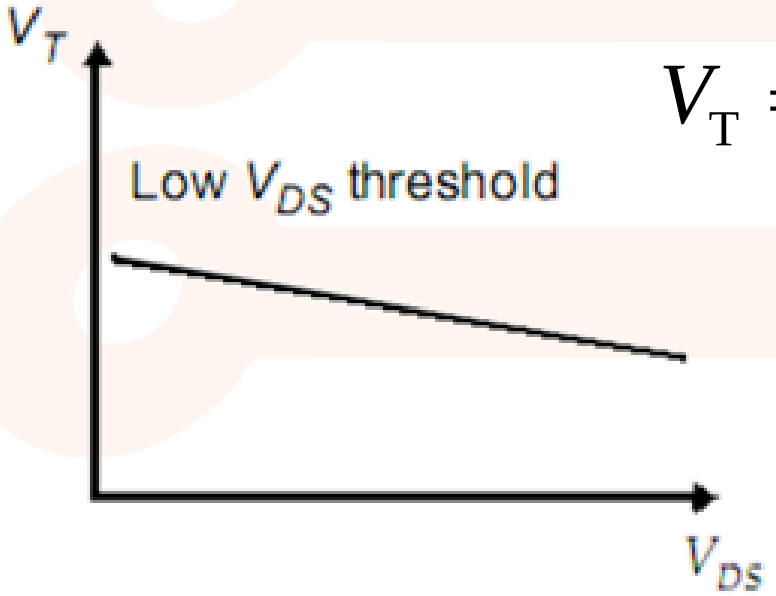
K. Goto et al., IEDM 2003



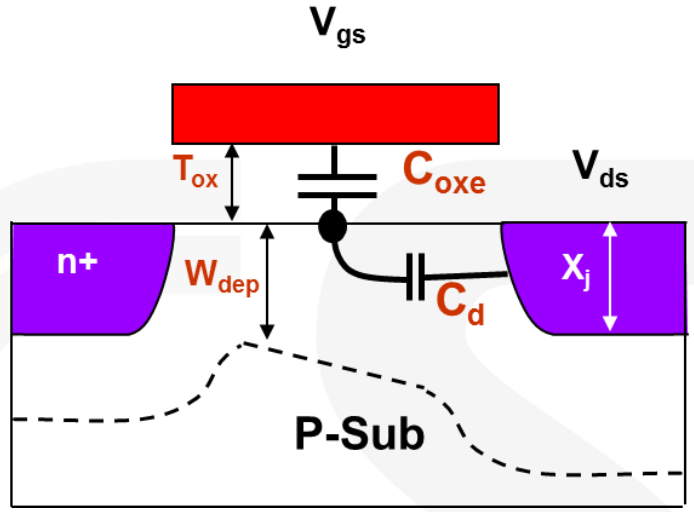
© Adam Teman, 2020

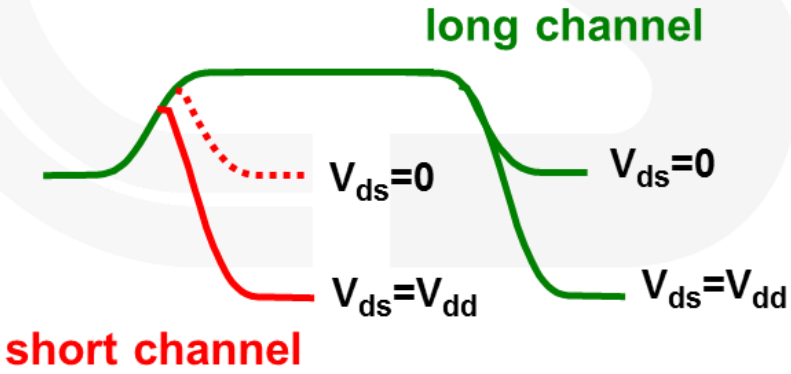
DIBL (Drain Induced Barrier Lowering)

- In short channels, the barrier of the channel is essentially lowered, as the drain causes the energy band to drop closer to the source.
- This is exponentially dependent on V_{DS}.



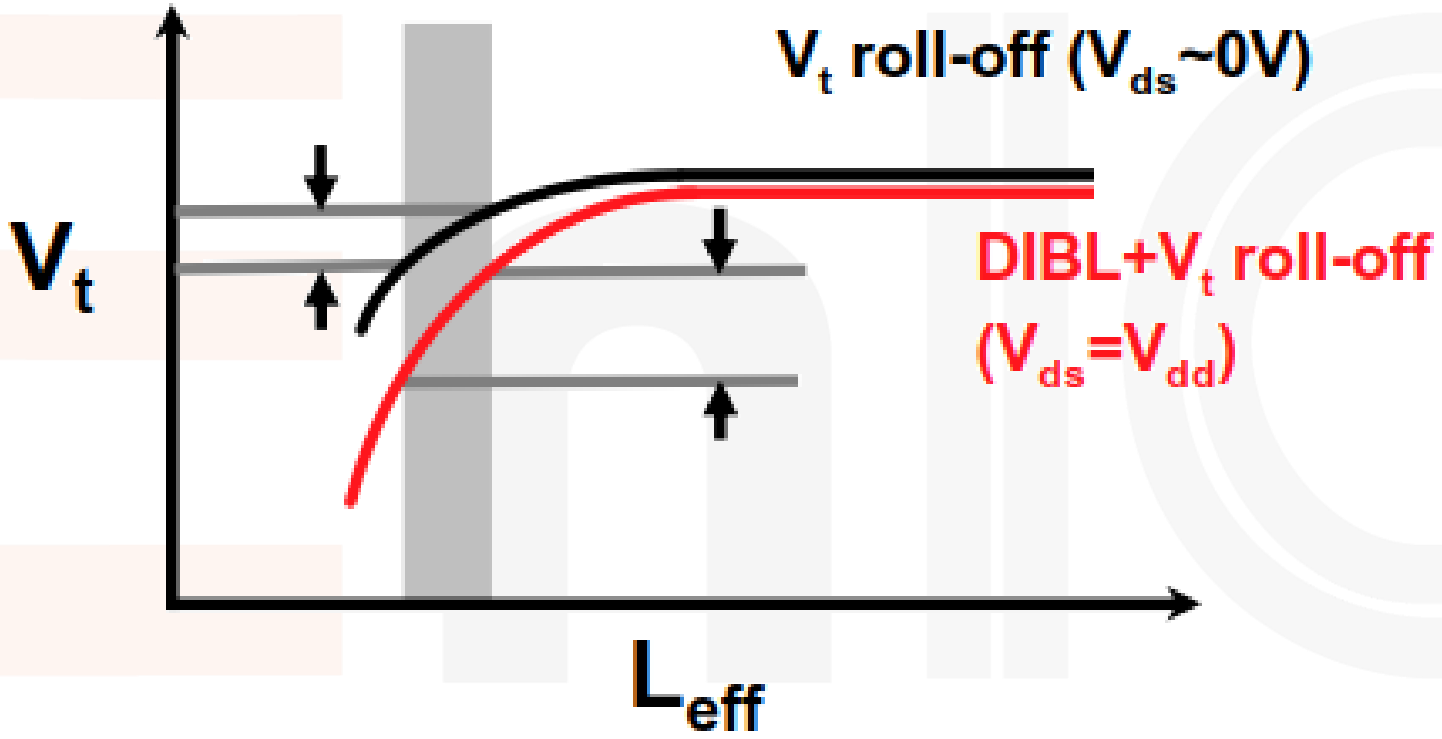
$$V_T = V_{T,long} - (V_{DS} + 0.4) \cdot \frac{C_d}{C_{oxe}}$$





© Adam Teman, 2020

Roll Off / DIBL combined

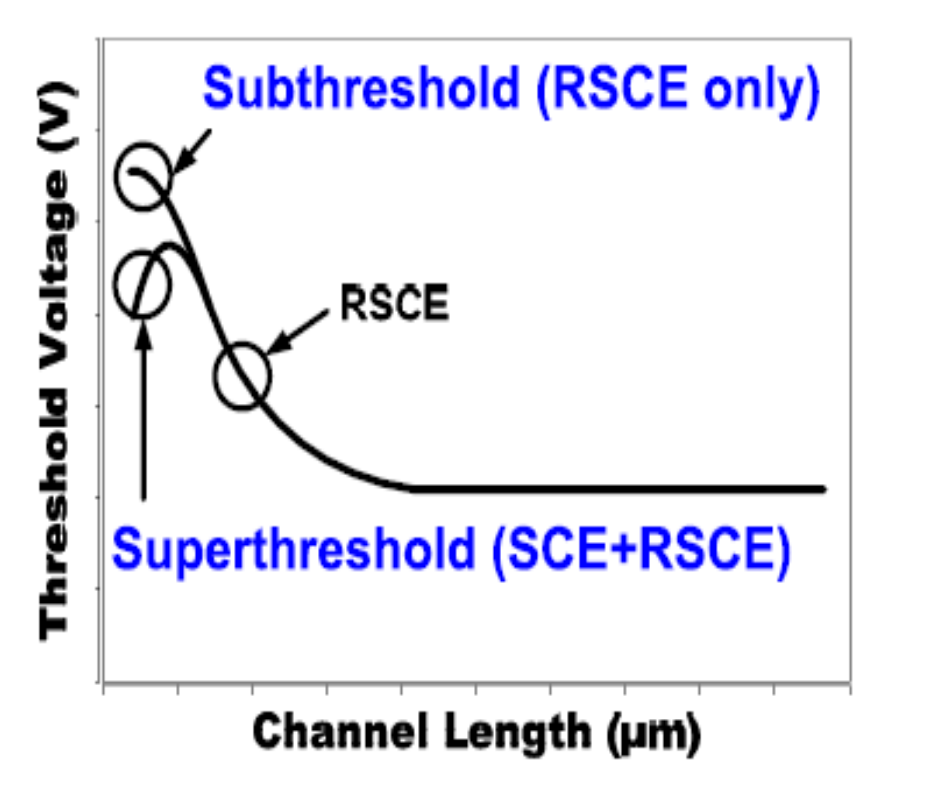


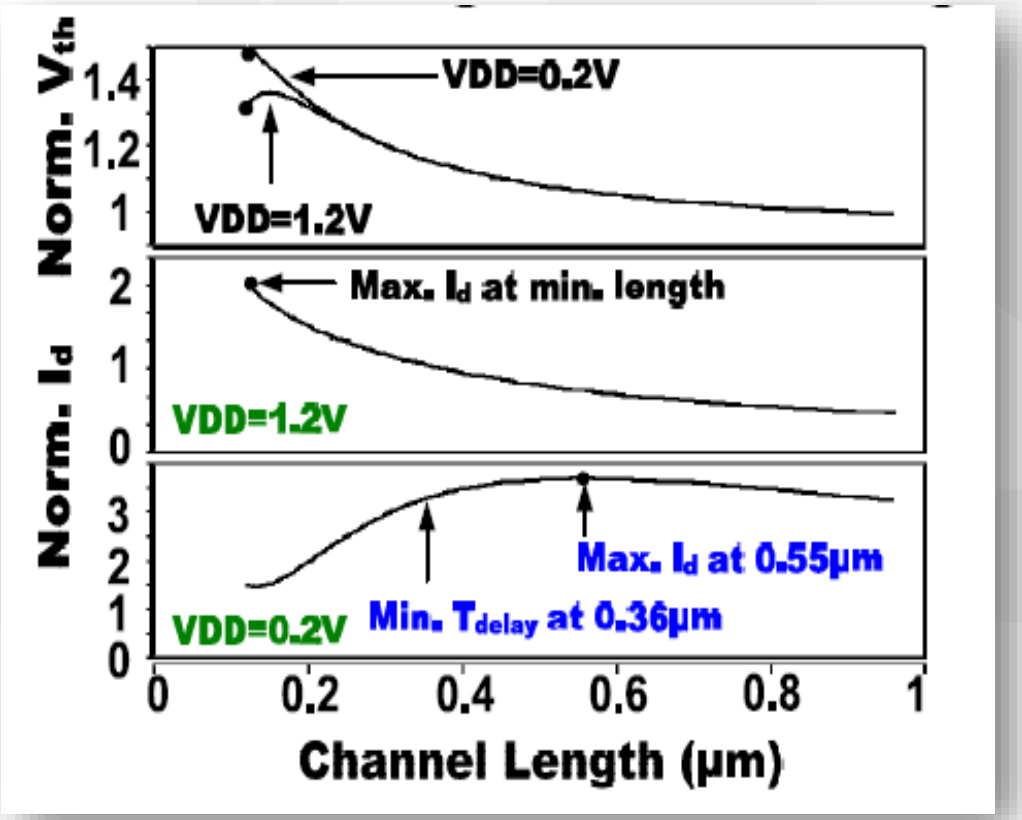
Source: Chris Kim, U. Minn

© Adam Teman, 2020

Reverse Short Channel Effect (RSCE)

- V_T actually ***increases*** at channel lengths a bit higher than minimum…





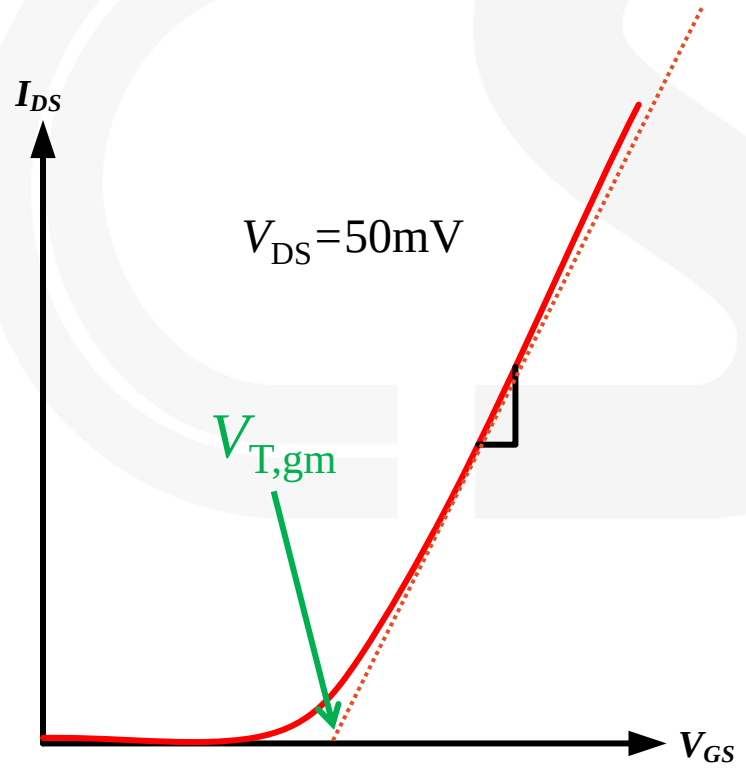
Adam Teman, 2020

How to Measure VT

- There are various ways to measure V_T
- One classic way takes a small V_{DS} and sweeps V_{GS}.

$$I_d = k\left[(V_{GS} - V_T)V_{DS} - \cancel{0.5V_{DS}^2}\right] \propto V_{GS} - V_T$$

- So we can find the V_{GS} at which the linear part crosses I_{ds}=0.

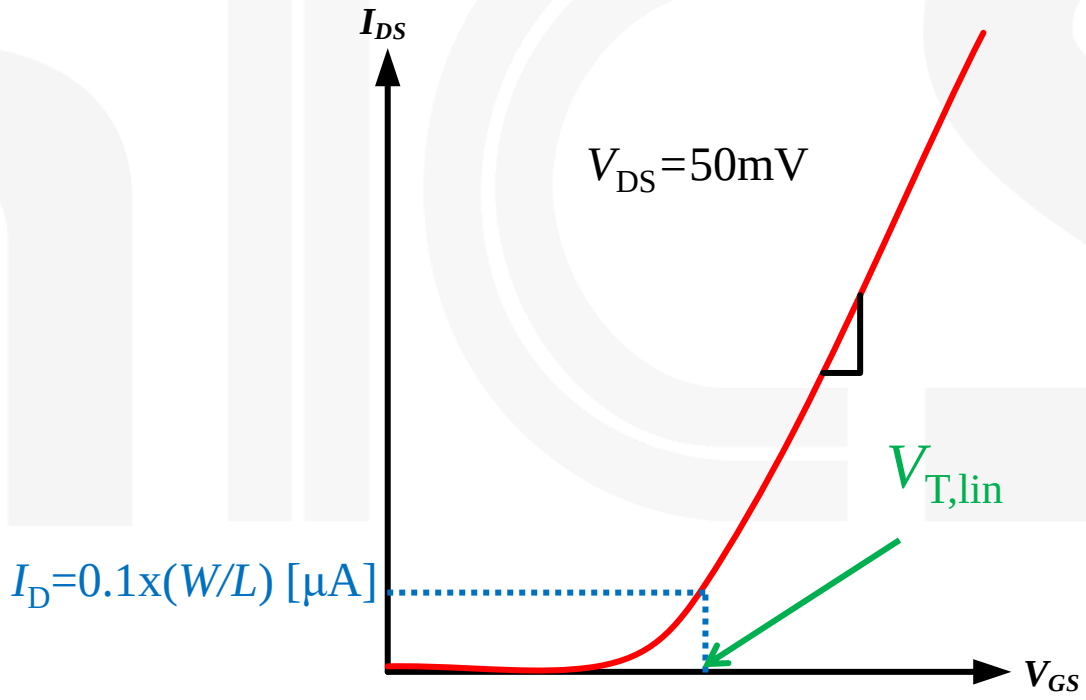


© Ada 2020

How to Measure VT

- **One of the more common ways is to find the V_{GS} at which I_{DS}=100 nA x W/L.**

- **For $V_{T,lin}$, set a low V_{DS} (V_{DS}=50mV)**
- **For $V_{T,sat}$ set a high V_{DS} (V_{DS}=V_{DD})**



© Ad 2020

Note about Simulation

- **The first step of all Spice (Spectre) simulations is a DC Operating Point calculation.**

© Adam Teman, 2020

Simulation tip: OP and MP in Spectre

- So we saw that the threshold voltage is dependent on the operating point.
- How do you know what the the V_T of a transistor is in a given simulation?
- To find V_{T0}, use the "**MP**" option.

- To find $V_{T,lin}$, $V_{T,sat}$, $V_{T,gm}$, use the "**OP**" option.

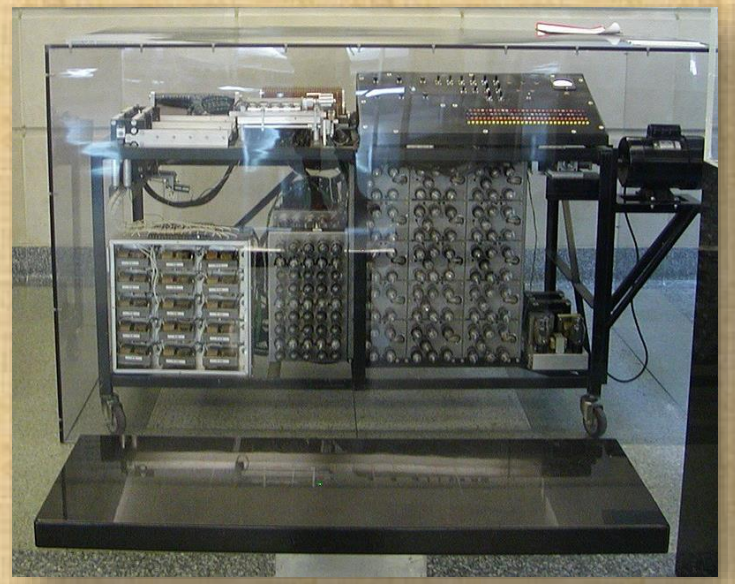
- How would you go about plotting roll-off (SCE) and RSCE?

© Adam Teman, 2020

The Computer Hall of Fame

- **We generally consider the ENIAC to be the first computer, but the official first fully electronic computer was the**

ACE



Source: wikipedia

- *Atansoff-Berry Computer*

  Conceived in 1937, operational in 1942

- Built at Iowa State University by Prof. John Atansoff and his student Clifford Berry.
- Not programmable nor Turing-complete, but included binary arithmetic and electronic switching elements.
- A patent dispute over the first electronic computer was settled in 1973, when the patent of the ENIAC was invalidated.

Source: wikipedia

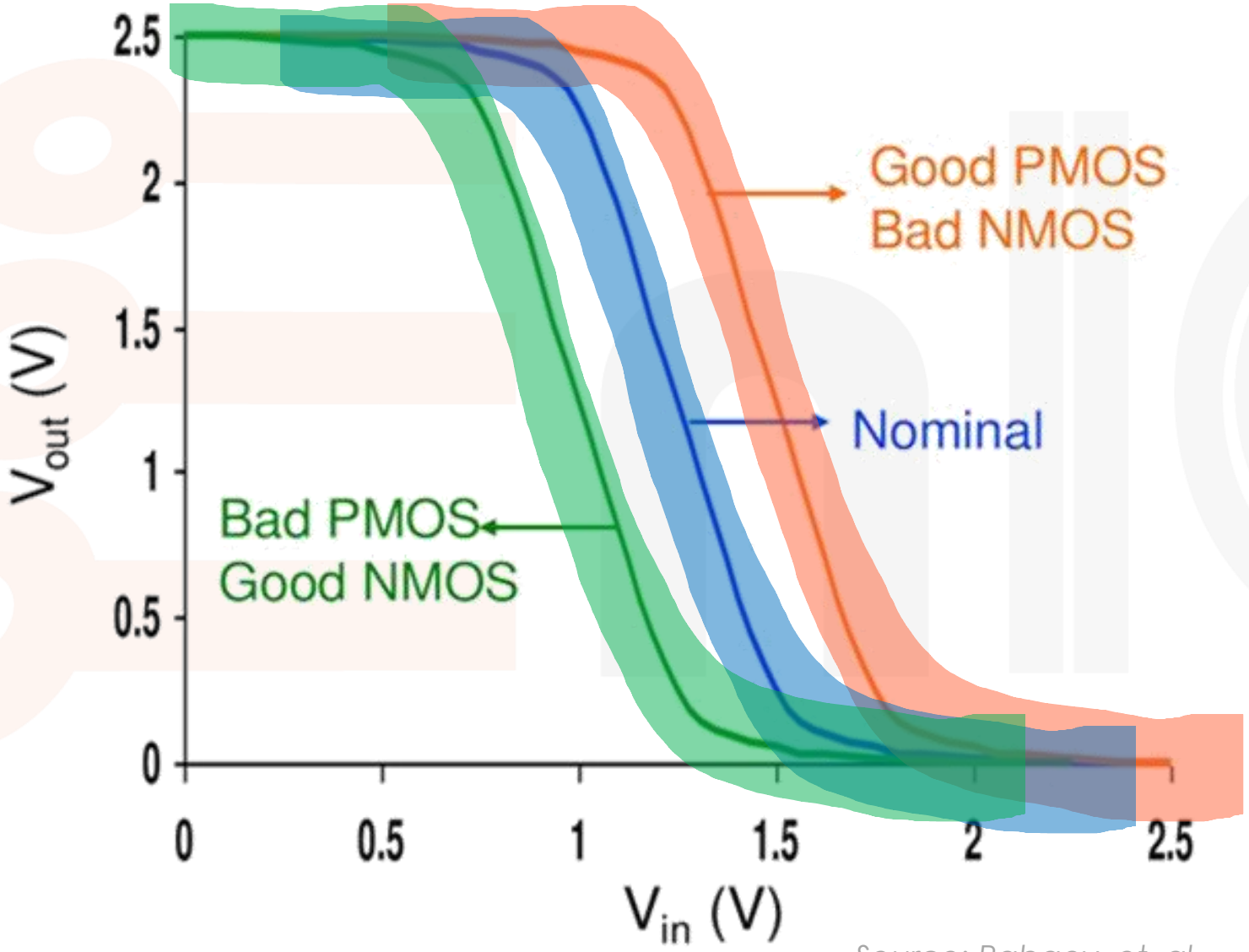
Simulating Variation

Process Corners and Monte Carlo Simulation

Emerging Nanoscaled
Integrated Circuits and Systems Labs

Tradition of Excellence
Bar-Ilan University
אוניברסיטת בר-אילן

Reminder: Impact of Process Variations



How do we take this into consideration during simulation?

Source: Rabaey, et. al.

© Adam Teman, 2020

Remove the Rust: Probability Basics

- **Properties of Random Variables**
  - The probability distribution function (PDF) $f(x)$ specifies the probability that a value of a continuous random variable X falls in a particular interval:

$$P\left[a < X \leq b\right] = \int_a^b f\left(x\right) dx$$

  - The cumulative distribution function (CDF) $F(x)$ specifies the probability that X is less than some value x:

$$F\left(x\right) = P\left[X < x\right] = \int_{-\infty}^{x} f\left(u\right) du \qquad f\left(x\right) = \frac{d}{dx} F\left(x\right)$$

  - The mean (μ) and variance (σ^2) are defined as:

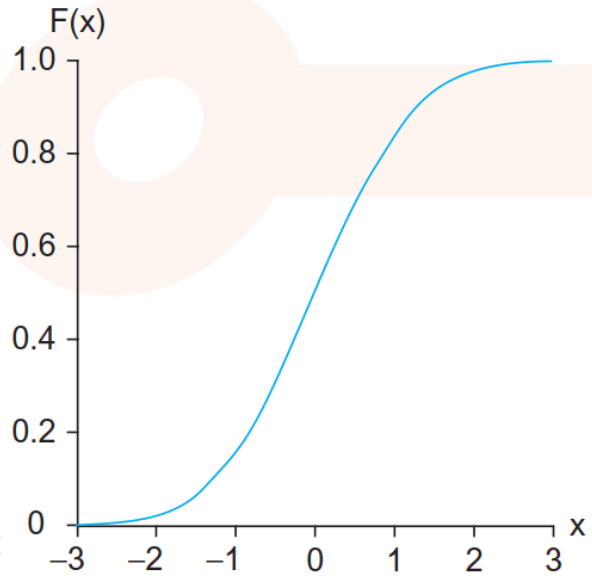
$$\mu\left(X\right) = \bar{X} = E\left[X\right] = \int_{-\infty}^{\infty} x \cdot f\left(x\right) dx \qquad \sigma^2\left(X\right) = E\left[\left(x - \bar{X}\right)^2\right] = \int_{-\infty}^{\infty} \left(x - \bar{X}\right)^2 f\left(x\right) dx$$

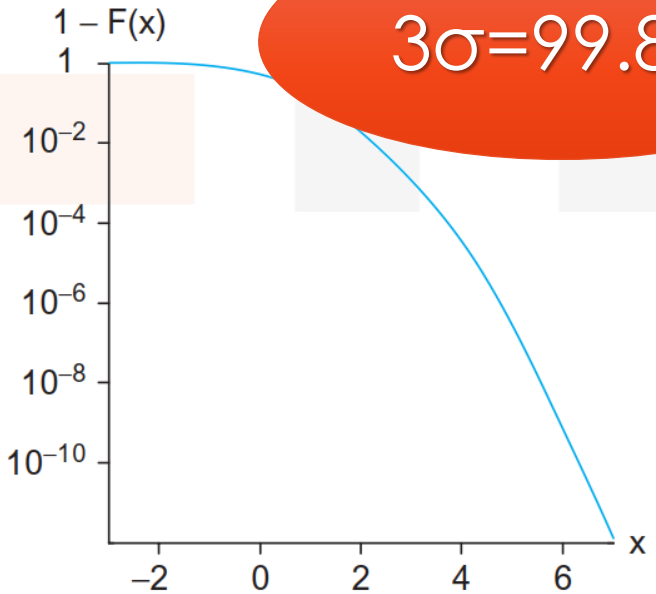
© Adam Teman, 2020

Remove the Rust: Probability Basics

- **Normal Random Variables**
  - A normal (Gaussian) random variable, shifted to have a zero mean ($\mu=0$) and a normalized standard variation ($\sigma^2=1$) has:

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} \qquad F(x)=\frac{1}{2}\left[1+\operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right]$$





3σ=99.8%

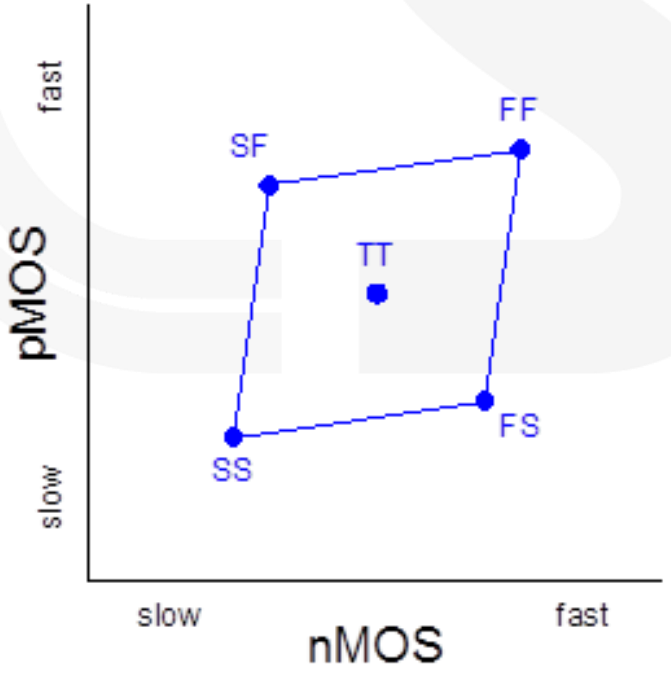
	F(x)	1 – F(x)
	0.8413	1.59×10^{-1}
	0.9772	2.28×10^{-2}
3	0.998650	1.35×10^{-3}
4	0.9999683	3.17×10^{-5}
5	0.999999713	2.87×10^{-7}
6	0.999999999013	9.87×10^{-10}

6σ = 1 in a billion

© Adam Teman, 2020

Global Variation Modeling: Process Corners

- **Global Variation assumption:**
  - If a certain process step is *skewed*, the entire chip is affected equivalently.
  - We will define "*corner cases*" of fabrication, i.e., 3σ from the mean.
  - We also assume the **voltage** and **temperature** are globally affected.
- **Devices are modeled for *fast*, *slow* and *nominal* corners.**
  - Changes in V_T, W, L, t_{ox}
- **Devices are tested at various *temperatures***
  - Temperature affects mobility and V_T.
  - Typically 0°C – 85°C or -40°C – 125°C
- **Devices are tested at various supply voltages**
  - Higher voltages cause increased currents
  - Typically $\pm 10\% V_{DD}$



© Adam Teman, 2020

Process Corners

- **What are the PVT (=Process, Voltage, Temperature) settings for each simulation corner?**

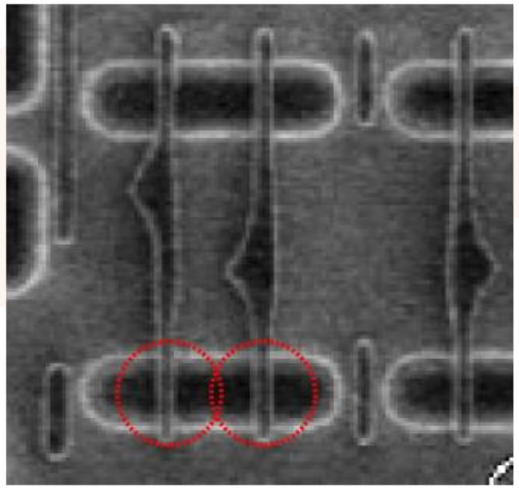
Corner	V_T	L_{eff}	t_{ox}	V_{DD}	T
Fast	↓	↓	↓	↑	↓
Typical	X	X	X	X	X
Slow	↑	↑	↑	↓	↑

© Adam Teman, 2020

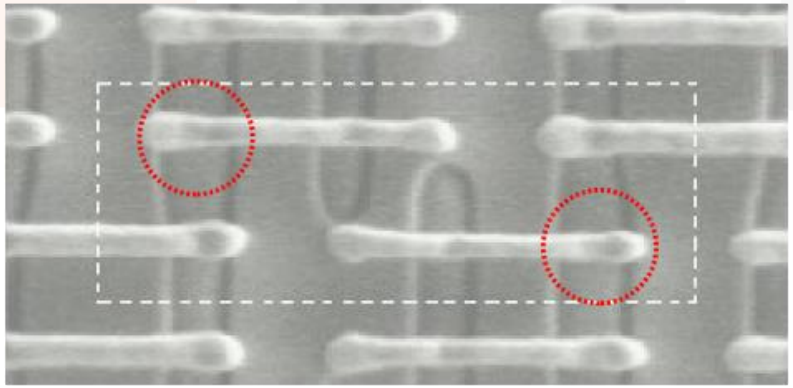
What about Local Variation?

- **Often there are *too many* parameters to think about and setting a specific corner case is insufficient.**
- **For example: Pelgrom's Law**
  - V_T variance is inversely proportional to transistor area

$$\sigma(V_T) = \frac{K}{\sqrt{W \cdot L}}$$

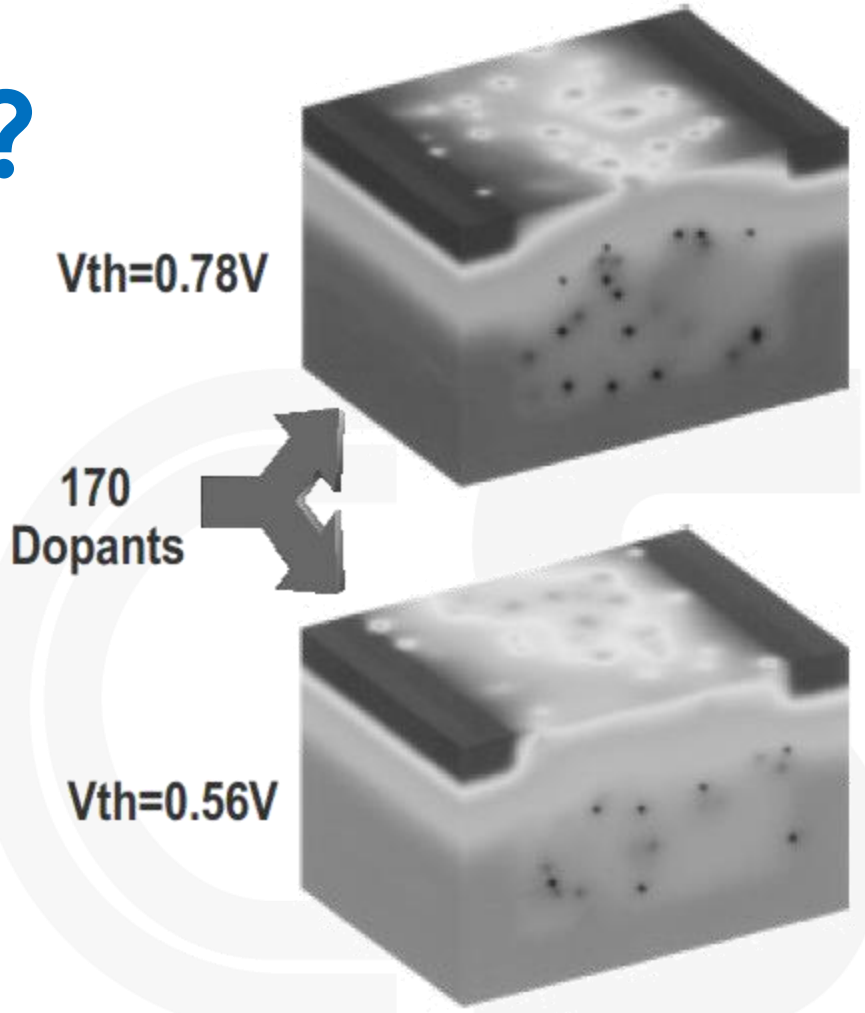


65nm CMOS NAND cell



Intel 65nm 6T SRAM cell

Source: Stanford, EE380



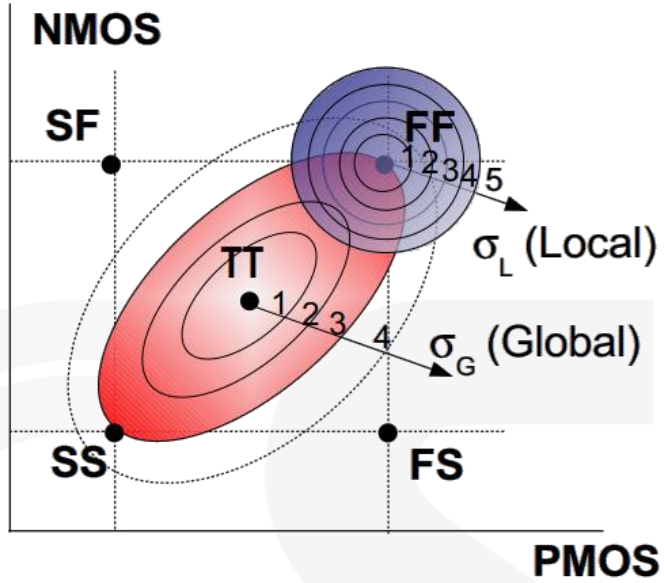
Source: Asenov, IEEE TED 1998

How do we deal with this?

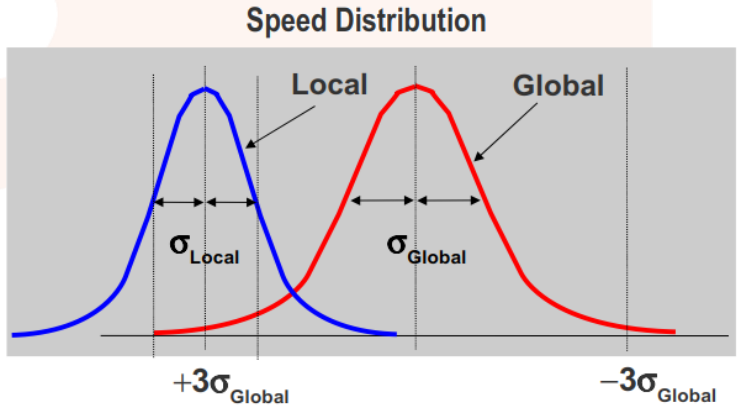
© Adam Teman, 2020

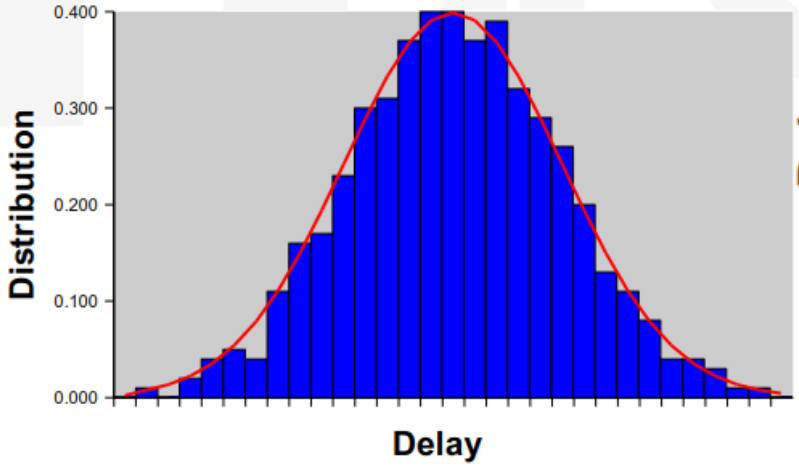
Monte Carlo Simulation

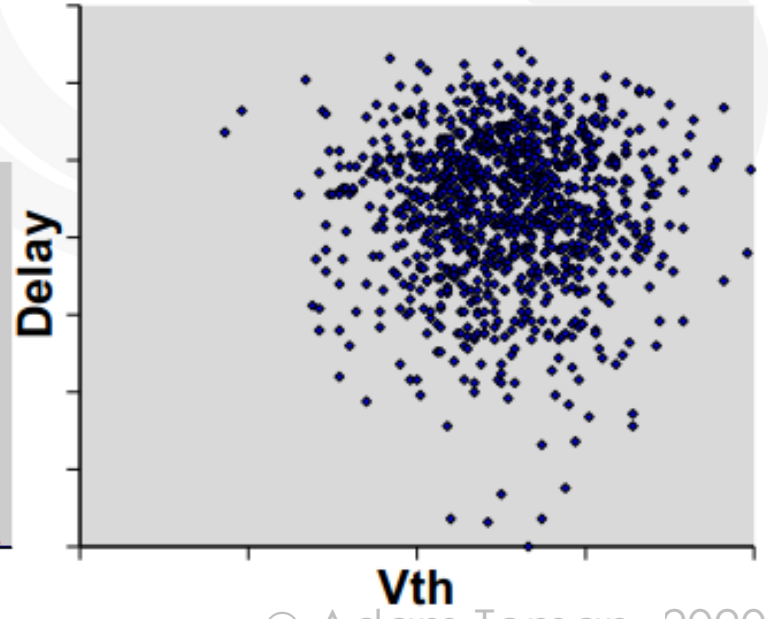
- **The basic approach is to "roll the dice" for each parameter and run a simulation.**
  - These are called ***Monte Carlo*** Statistical Simulations.
- **The result is a distribution plot of design constraints, e.g., delay or noise margin**
- **Both Global and Local Variations can be taken into consideration.**



Source: Stanford, EE380







© Adam Teman, 2020

Simulation questions

- How do we plot threshold voltage variation?

© Adam Teman, 2020

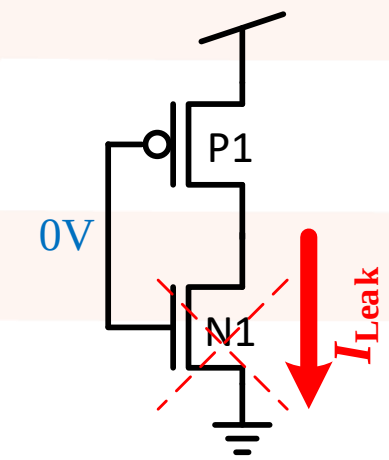
Leakage in NanoScaled Transistors

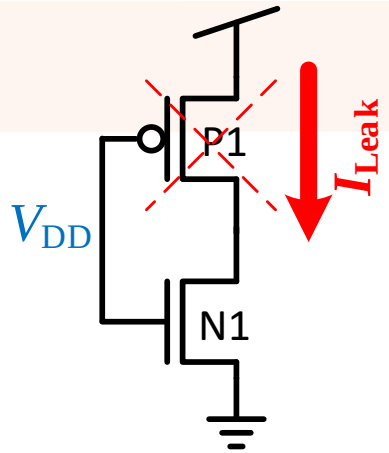
Emerging Nanoscaled Integrated Circuits and Systems Labs

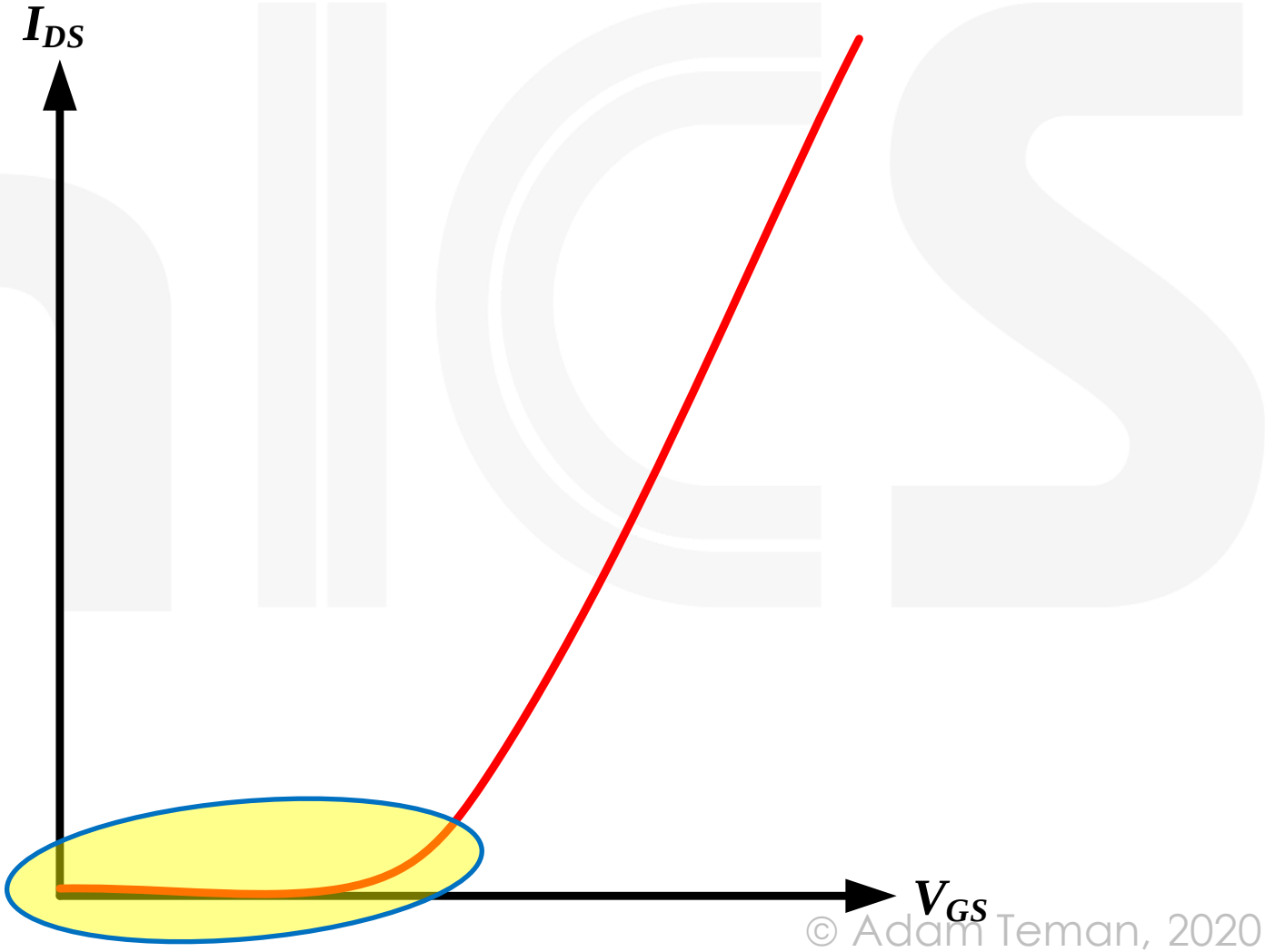
Tradition of Excellence
Bar-Ilan University
אוניברסיטת בר-אילן

Leakage in Nanoscaled Transistors

- **Transistors that are supposed to be off actually leak!**

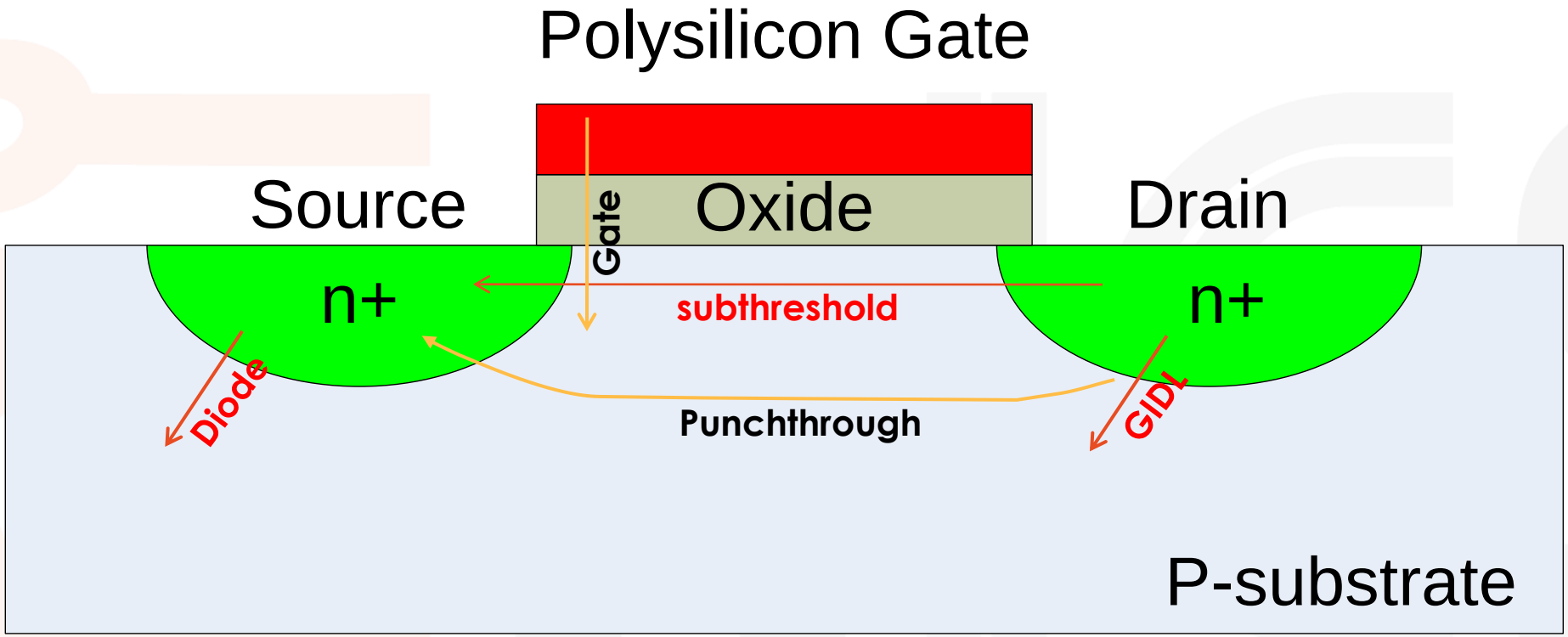






© Adam Teman, 2020

Main Types of Leakage



$$I_{sub} \propto e^{V_{GS}-V_T}$$

$$I_{gate} \propto V_{GC}, \; 1/t_{ox}$$

$$I_{GIDL} \propto V_{DG}$$

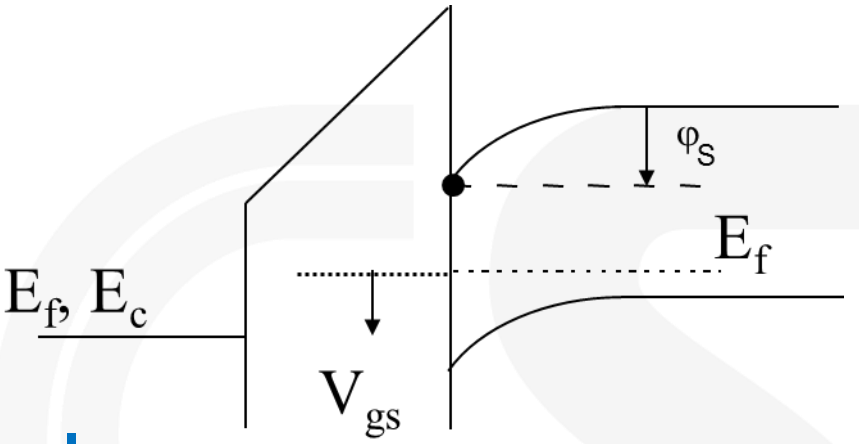
$$DIBL \propto e^{V_{DS}}$$

© Adam Teman, 2020

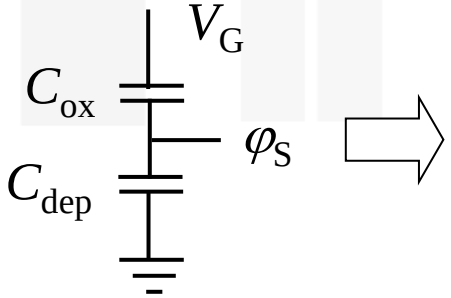
Subthreshold Leakage

- When $V_{GS} < V_T$, there is still a finite carrier concentration at the surface:

$$n_s \propto e^{\varphi_s / \phi_T} \Rightarrow I_{sub} \propto e^{V_{GS} - V_T / n\phi_T}$$



- As we saw with the body effect, due to bulk to channel capacitance, the surface voltage isn't only controlled by the gate:



$$n \equiv 1 + \frac{C_{dep}}{C_{oxe}}$$

$$\frac{d\varphi_s}{dV_G} = \frac{C_{oxe}}{C_{oxe} + C_{dep}} = \frac{1}{n}$$

© Adam Teman, 2020

Subthreshold Leakage

- **Let's make things easier:**
  - Remember that: $$I_{DS}(\text{sub}) = Const \frac{W}{L} \cdot e^{V_{GS}-V_T / n\phi_T}$$
  - And we now defined the threshold voltage according to current: $$I_{DS}(V_{GS}=V_T) = 100\text{nA} \cdot \frac{W}{L}$$
  - So the boundary condition requires: $$I_{DS}(\text{sub})\big|_{V_{GS}=V_T} = Const \frac{W}{L} \cdot e^{0/n\phi_T} = Const \frac{W}{L} = 100\text{nA} \frac{W}{L}$$
  - And we can now calculate subthreshold current as: $$I_{sub}[nA] = 100 \frac{W}{L} e^{V_{GS}-V_T / n\phi_T}$$

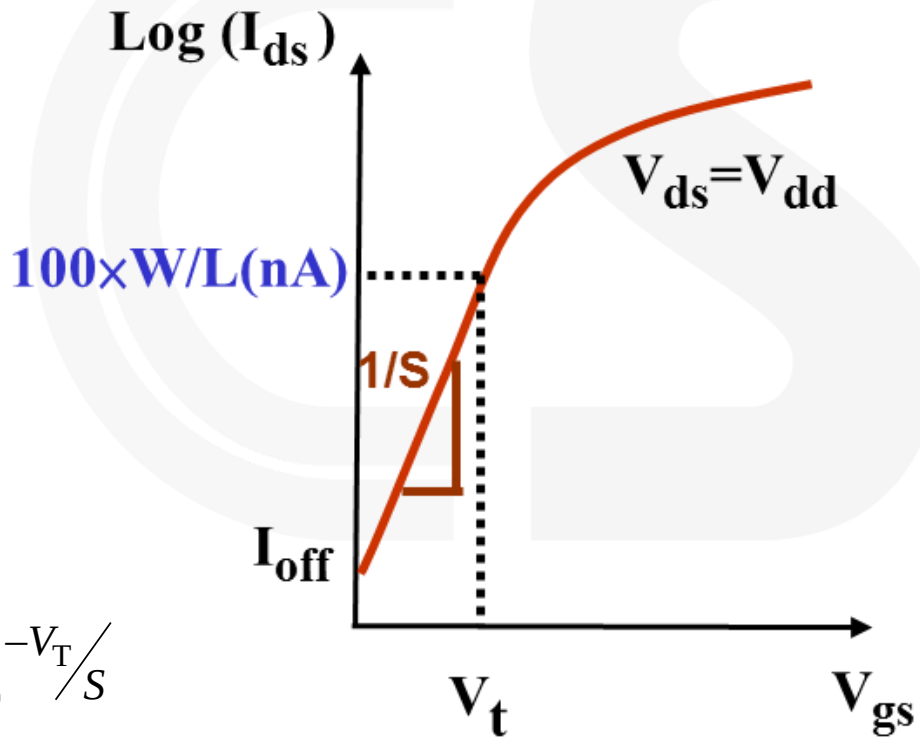
© Adam Teman, 2020

Subthreshold Leakage

- **An even easier way is to look at the plot of $\log(I_{DS})=f(V_{GS})$:**
  - The slope of this curve is called the "Subthreshold Slope"
  - The inverse of this slope is known as the "Subthreshold Swing" (S):

$$S \equiv \ln(10)\frac{kT}{q}\left(1+\frac{C_{dep}}{C_{ox}}\right) = 2.3 \cdot n \cdot \phi_T$$

$$I_{sub}\left[nA\right] = 100\frac{W}{L}e^{V_{GS}-V_T/n\phi_T} = 100\frac{W}{L}10^{V_{GS}-V_T/S}$$



- **And I_{off}, which is defined as the current when $V_{GS}=0$ is:**

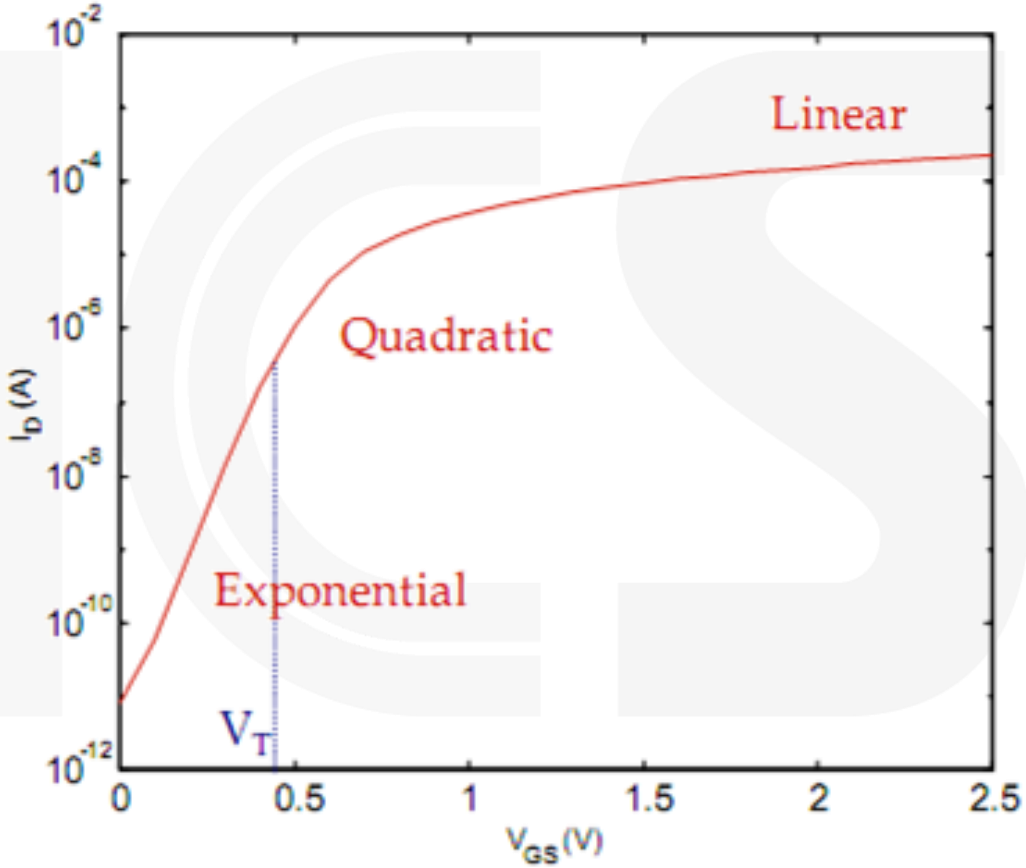
$$I_{off}\left[nA\right] = 100\frac{W}{L}10^{-V_T/S}$$

© Adam Teman, 2020

Subthreshold Leakage

- **So subthreshold leakage is:**
  - Exponentially dependent on V_{GS}.
  - Exponentially dependent on V_T.
- S **is the subthreshold swing coefficient.**
  - Optimally, S_{opt}=60 mV/dec
  - Realistically $S \approx 100$ mV/dec

$$S \equiv n\phi_T \ln 10 \geq 0.06 \qquad n \equiv 1 + \frac{C_{dep}}{C_{oxe}}$$



© Adam Teman, 2020

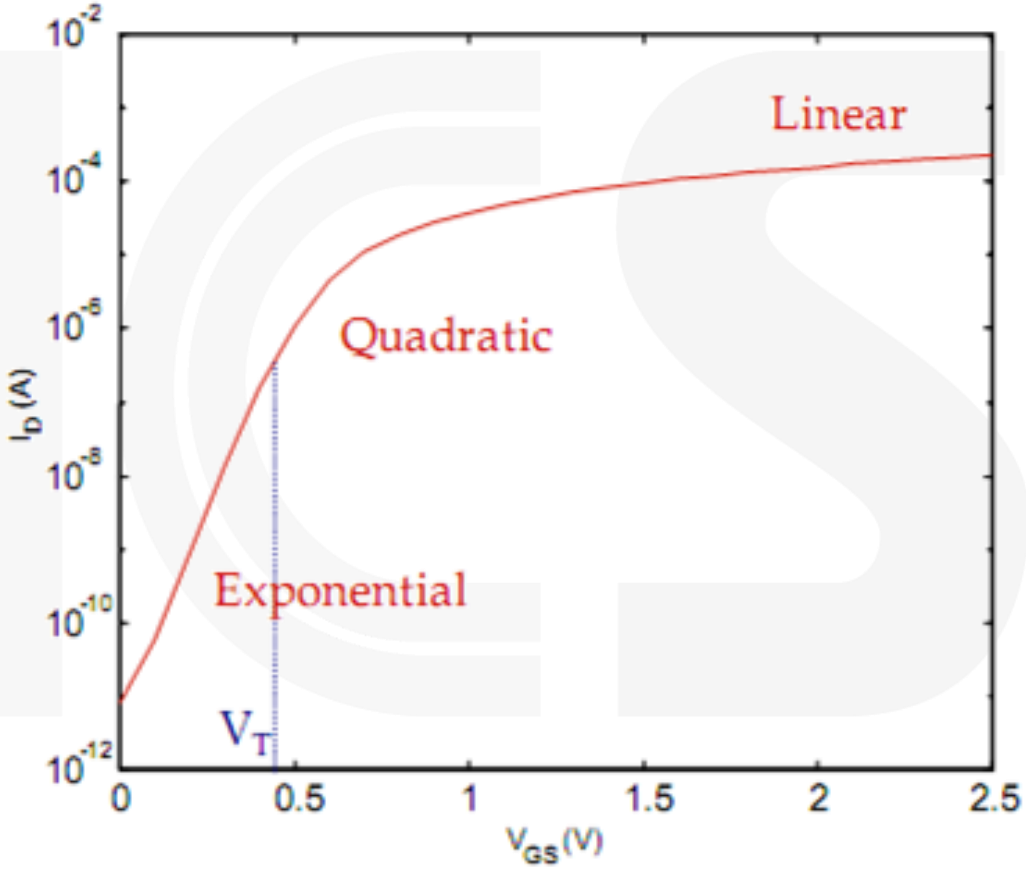
Subthreshold Leakage

Example:

- **We want to design a transistor with:**

$$\frac{I_{on}}{I_{off}} \geq 10^4 \qquad S = 60\frac{\text{mV}}{\text{dec}}$$

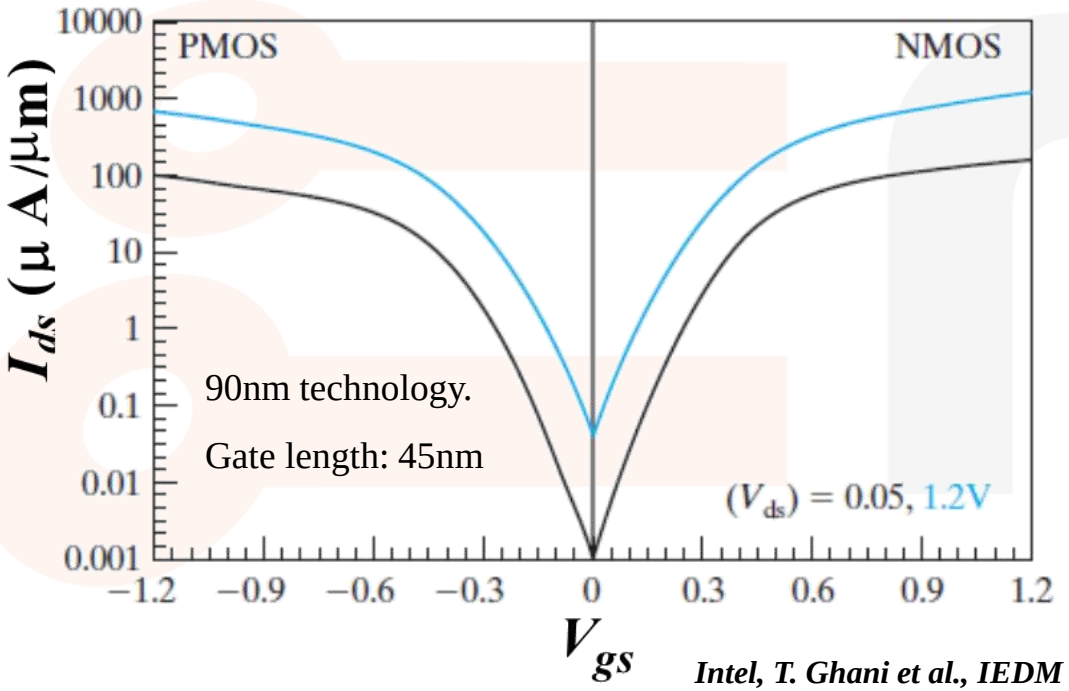
- **What is the minimum V_T?**



© Adam Teman, 2020

Impact of DIBL

- **DIBL causes an additional exponential increase in subthreshold leakage with V_{DS}.**



Intel, T. Ghani et al., IEDM 2003

$$I_{sub} = I_0 e^{\left(\frac{V_{GS}-V_T}{n\phi_T}\right)} \times \left(1 - e^{\frac{-V_{DS}}{\phi_T}}\right) \times e^{\frac{\eta V_{DS}}{n\phi_T}}$$

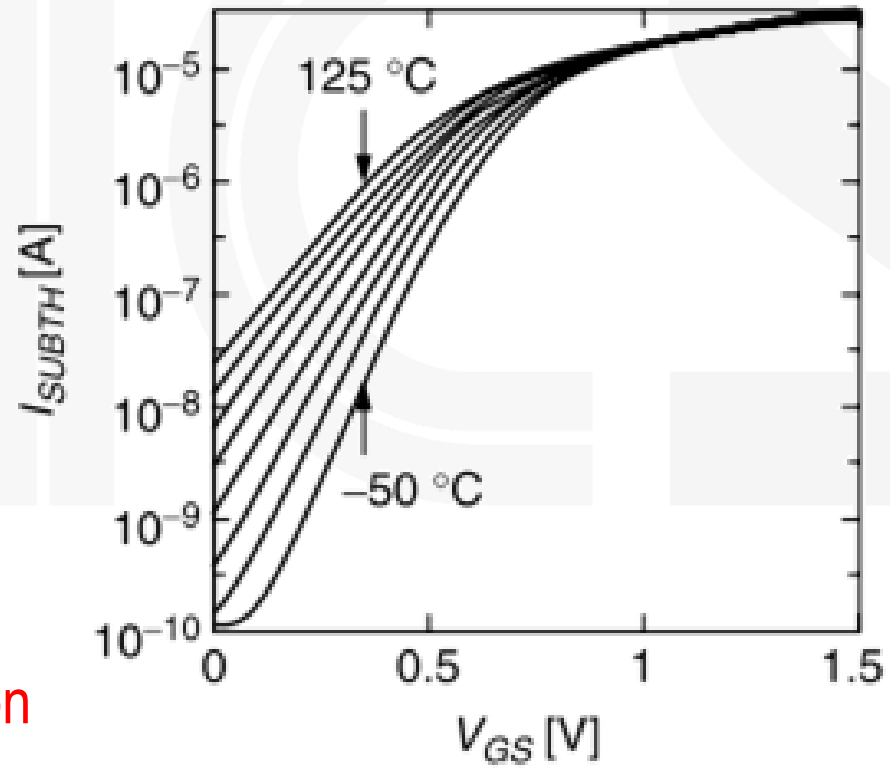
© Adam Teman, 2020

Subthreshold Dependence on Temperature

- This is rather complex, as mobility degrades with temperature and other device values (such as flatband voltage) are temperature dependent.
- Altogether, subthreshold leakage rises exponentially with temperature*.

$$I_{\text{sub}} \propto e^{V_{GS}-V_T \big/ \eta\phi_T}$$

$$\phi_T \equiv kT \big/ q$$



* Without considering temperature inversion

© Adam Teman, 2020

Temperature Inversion

- **Classic approach to temperature effect on delay:**

$$\left.\begin{array}{l} I \propto \mu \\ \mu \frac{1}{\propto} T \end{array}\right\} \rightarrow t_{pd} \propto T$$

So speed decreases with temperature

$$\left.\begin{array}{l} V_T \frac{1}{\propto} T \\ t_{pd} \propto V_T \end{array}\right\} \rightarrow t_{pd} \frac{1}{\propto} T$$

So speed actually increases with temperature!

- **BUT!**
  - V_T decreases by as much as -3mV/°C
  - The point of *temperature inversion* is the *voltage* at which speed increases with temperature ($\sim V_{DD}$=1V)!

© Adam Teman, 2020

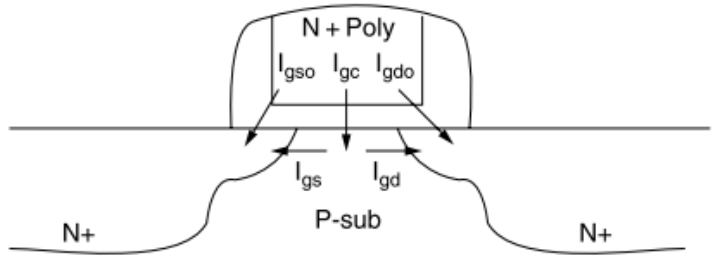
Gate Leakage

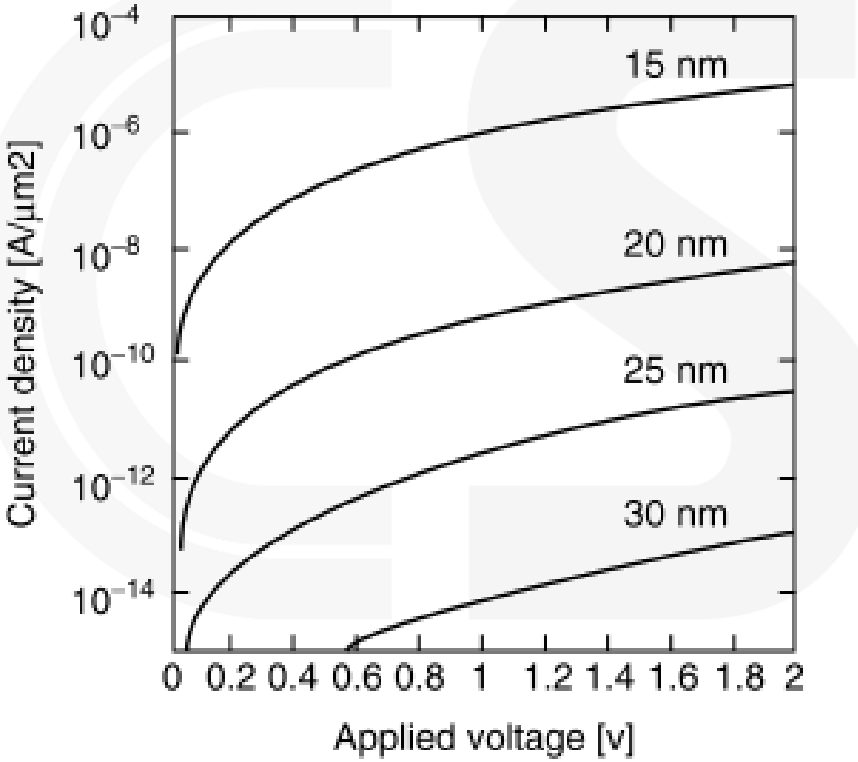
- **Two mechanisms:**
  - Direct tunneling (dominant)
  - Fowler Nordheim tunneling
- **Exponentially Dependent on:**
  - Gate Voltage (V_G)
  - Oxide Thickness (t_{ox}).

$$I_{gate} = AE_{ox}^2 e^{-B/E_{ox}}$$

$$E_{ox} = \frac{V_{DD} - V_T}{t_{ox}}$$

- **Non-dependent on temperature.**
- **Much stronger in nMOS than pMOS (higher barrier for holes)**
- **Minimum t_{ox}=1.2nm!!!**

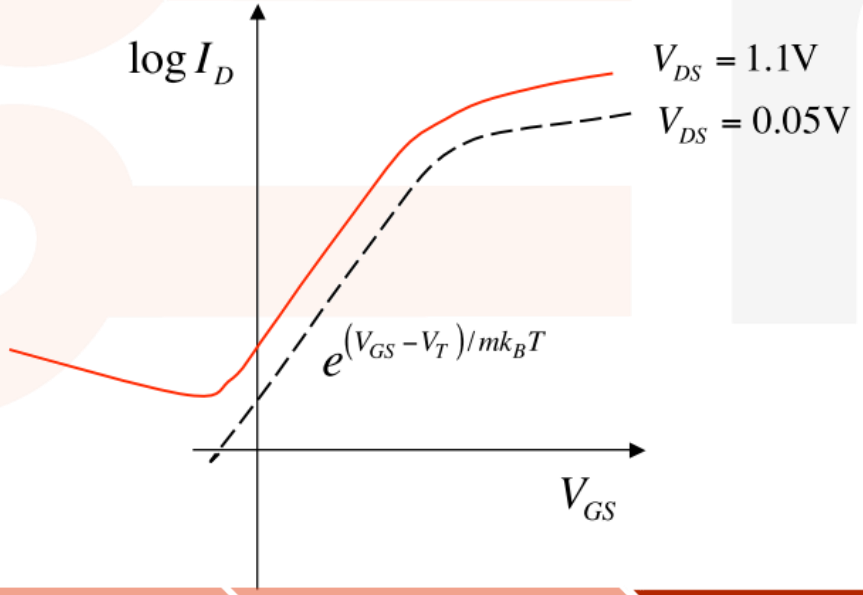


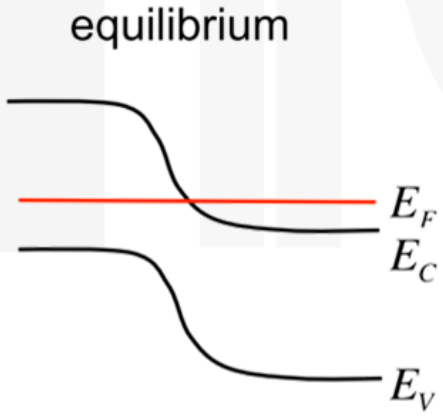


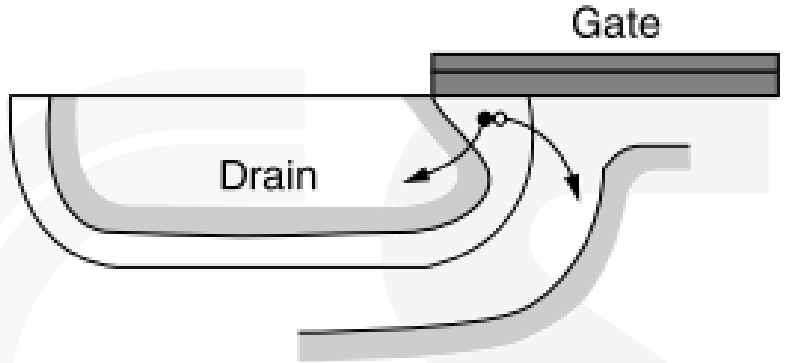
© Adam Teman, 2020

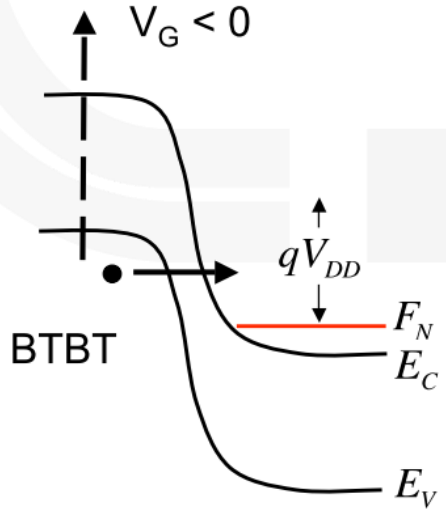
Gate Induced Drain Leakage

- GIDL current flows from the drain to the substrate.
- Caused by high electric field under the gate/drain overlap, causing *e-h pair* creation.
- Main phenomena is Band-to-Band Tunneling









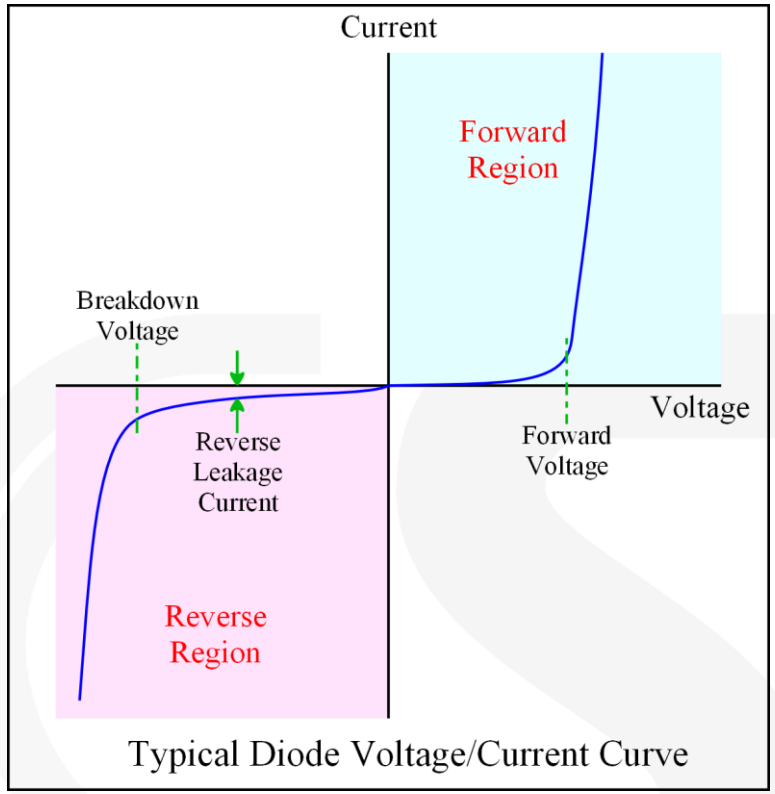
© Adam Teman, 2020

Diode Leakage

- J_S=10-100pA/μm² @25°C **for** 0.25μm **CMOS.**
- J_S **doubles for every** 9°C

$$I_{DL} = J_S \times A$$

- **Much smaller than other leakages in deep sub-micron.**
  - But a bigger factor in low subthreshold leakage processes, like FinFET.

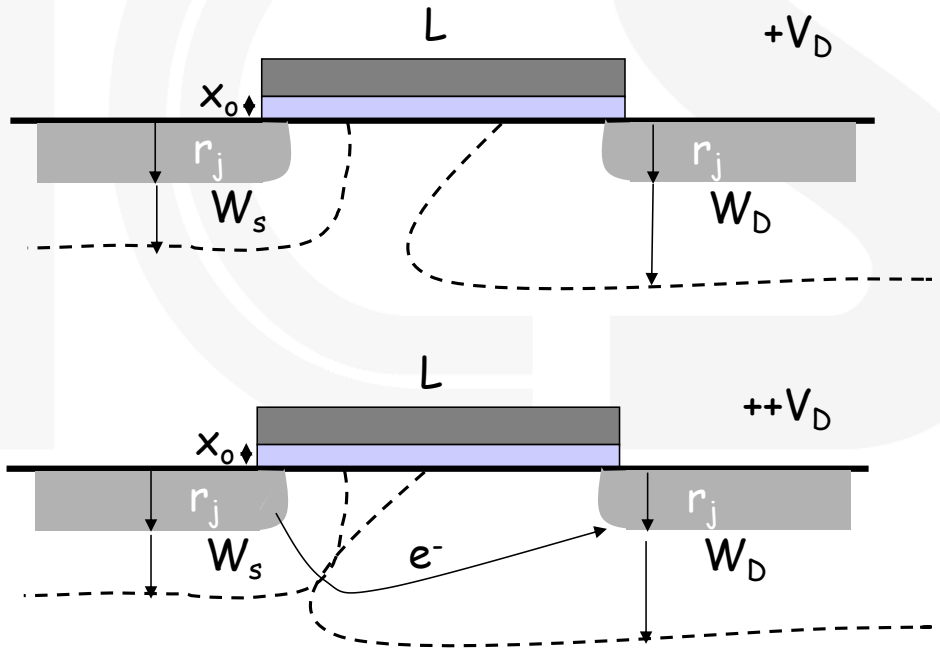


Source: cmicrotek.com

© Adam Teman, 2020

Punchthrough

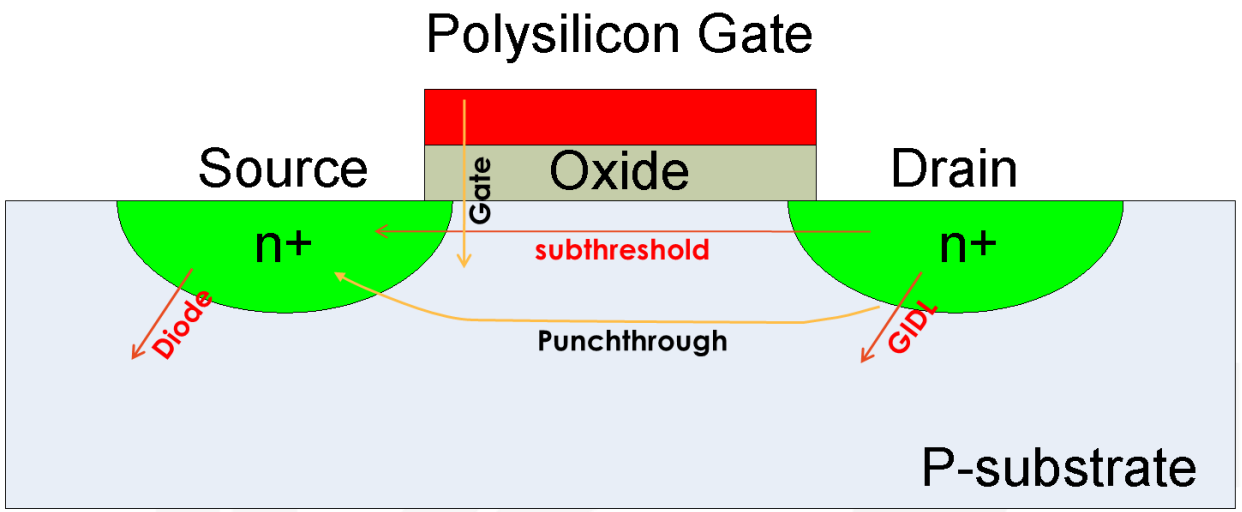
- As V_{DS} grows, so does the drain depletion region, and the channel length decreases.
- In severe cases, the source and drain are connected causing non-controllable leakage current.

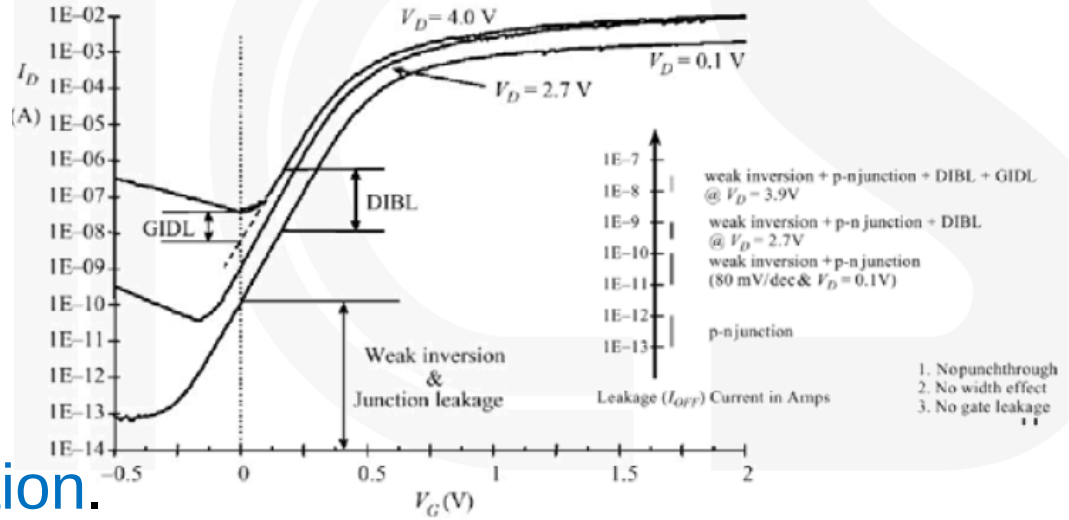


© Adam Teman, 2020

Leakage Summary

- **Subthreshold Leakage:**
  - $I_{DS}>0$ when $V_{GS}<V_T$ due to weak inversion.
  - Grows with V_{GS}, V_{DS}, lower V_T
- **Gate Leakage:**
  - $I_G>0$ due to direct tunneling through the oxide.
  - Grows with V_{GB}, t_{ox}
- **Gate Induced Drain Leakage (GIDL):**
  - $I_{DB}>0$ due to high electric field in the GD overlap region (V_{GD}).
- **Reverse Biased Diode Leakage**
  - I_{SB}, I_{DB} due to diffusion and thermal generation.
- **Punchthrough:**
  - I_{DS} due to drain and source depletion layers touching.





[IEEE press, 2000]

© Adam Teman, 2020

Process Corners - Revisited

But what about Temperature Inversion?

- **Should we now redefine the PVT settings?**

Corner	V_T	L_{eff}	t_{ox}	V_{DD}	T
Fast	↓	↓	↓	↑	↓
Typical	X	X	X	X	X
Slow	↑	↑	↑	↓	↑
Max Leakage	↓	↓	↓↑	↑	↑

© 2020

Further Reading

- J. Rabaey, “Digital Integrated Circuits” 2003, Chapters 2.5, 3.3-3.5
- Weste, Harris “CMOS VLSI Design”, Chapter 7
- C. Hu, “Modern Semiconductor Devices for Integrated Circuits”, 2010, Chapters 4-7
- Tzividis, et al. “Operation and Modeling of MOS Transistor”Chapters 1-5
- E. Alon, Berkeley EE-141, Lecture 9 (Fall 2009)
- M. Alam, Purdue ECE-606 – lectures 32-38 (2009) nanohub.org
- A. B. Bhattacharyya “Compact MOSFET models for VLSI design”, 2009,
- T. Sakurai, “Alpha Power-Law MOS Model” – JSSC Newsletter Oct 2004
- Managing Process Variation in Intel’s 45nm CMOS Technology, Intel Technology Journal, 2008
- Berkeley “BSIM 4.6.4 User’s Manual”
  http://www-device.eecs.berkeley.edu/~bsim/BSIM4/BSIM464/BSIM464_Manual.pdf

© Adam Teman, 2020