



THE UNIVERSITY *of* EDINBURGH

Analogue IC Design

FET Theory and Operations

Sep – Dec 2022

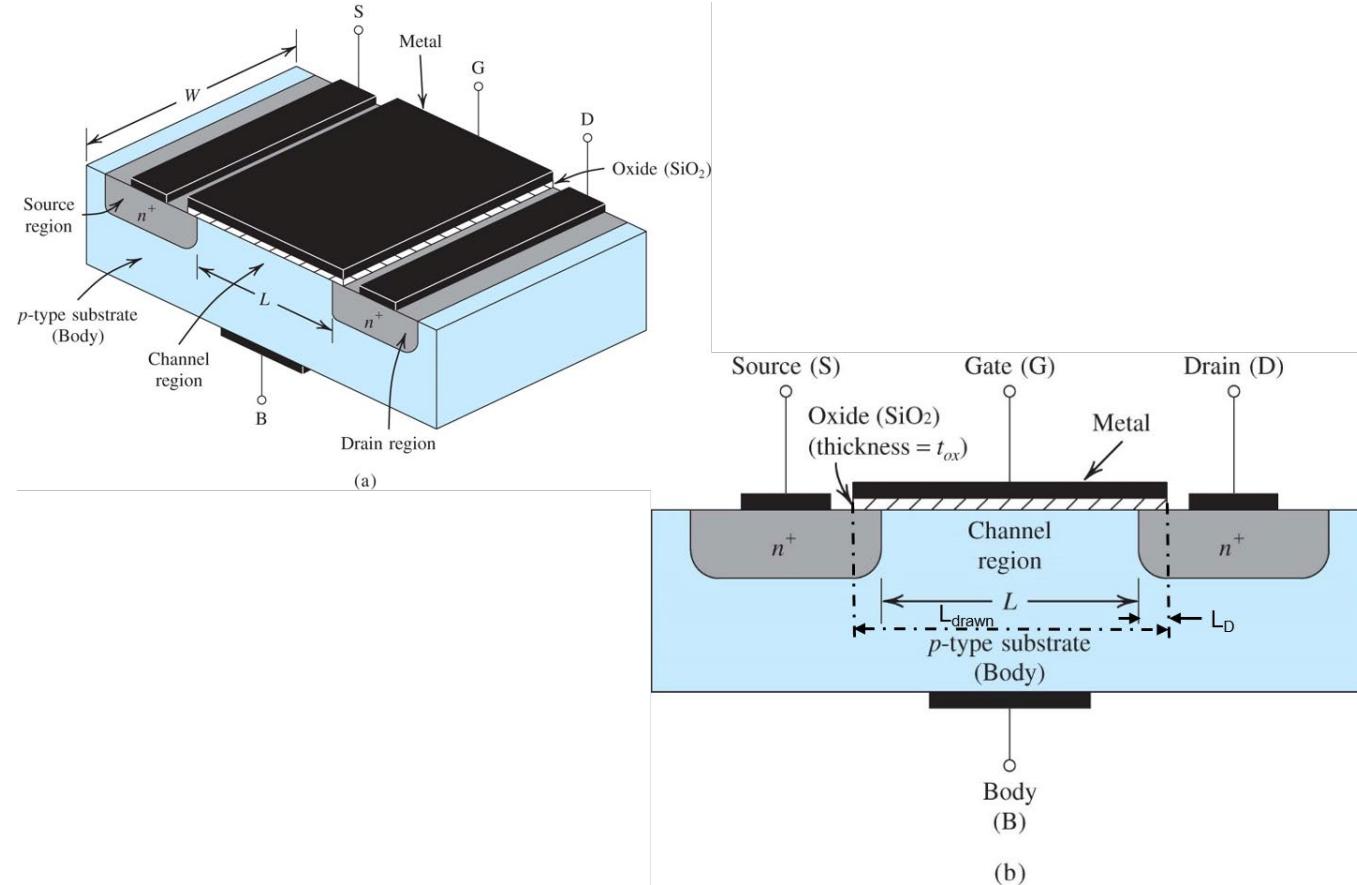
Dr Danial Chitnis

d.chitnis@ed.ac.uk

Basics: FET structure and Physical Operation

- This background topic, ‘FET structure and Physical Operation’, is a pre-requisite for the Analogue IC Design Course. It is expected that you have previously studied the basic structure and physical operation of FETs.
- These slides are thus a revision material on the structure and basic operation of a FET. It is based on chapters 5 & 6 of the text book: *‘Microelectronic Circuits’ by Sedra and Smith, 7th edition.*

Basics: FET structure and Physical Operation



- Figure 3: Physical structure of the enhancement-type NMOS transistor: (a) perspective view; (b) cross section

Physical structure of the enhancement-type NMOS transistor

- Dimension of the Gate along the Source-Drain path is called the Length, (L). The width, (W), is perpendicular to the length.
- During fabrication, the Source and Drain ‘side-diffuse’ by length LD . Thus, the actual length between the Source and the Drain is slightly reduced.
- For clarity: $L = L_{\text{drawn}} - 2LD$. L_{drawn} is the dimension drawn in the transistor layout.
- With zero voltage at the gate, back-to-back diodes exist in series between drain and source. These diodes prevent current conduction from drain to source when voltage V_{DS} is applied. Thus, the path between drain and source has very high resistance (of the order of $10^{12}\Omega$).
- If a positive voltage, V_{GS} applied to the gate and Source, Body and Drain all grounded (see Figure 4):

Physical structure of the enhancement-type NMOS transistor

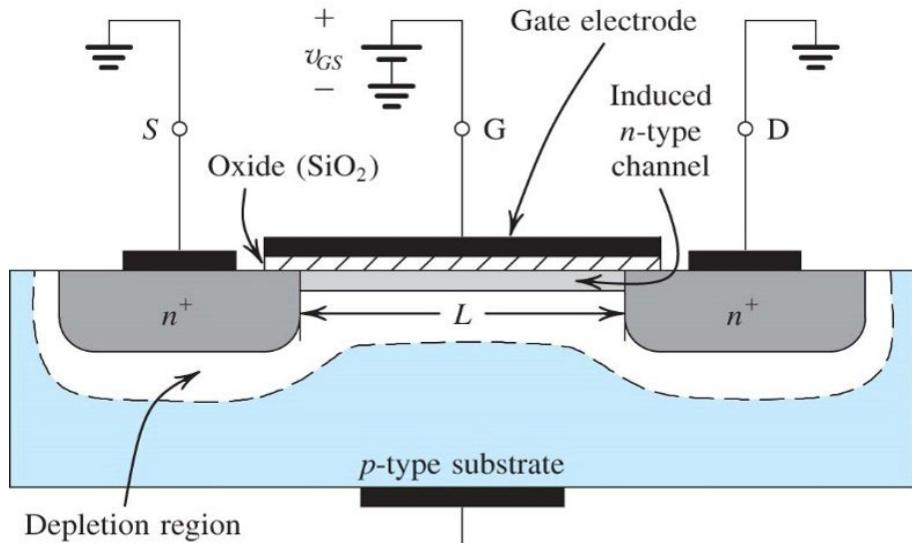


Figure 4: Enhancement-type NMOS transistor with a positive V_{GS} .

- The positive voltage V_{GS} on the gate causes, in the first instance, the free holes to be repelled from the region of the substrate under the gate i.e the channel region.
- The repelled holes are pushed downward into the substrate, leaving behind a carrier-depletion region. The depletion region is populated by the bound negative charge associated with the acceptor atoms. These charges are “uncovered” because the neutralizing holes have been pushed downward into the substrate.



- Also, the positive gate voltage attracts electrons from the n source and drain regions into the channel region.
- When sufficient number of electrons accumulate near the surface of the substrate under the gate, an n-region is in effect created, connecting the source and the drain regions.
- When a positive voltage is now applied between the drain and source, current flows through this induced n region. The mobile electrons carrying the current.
- The “induced” n region form a channel for current to flow from the drain to source (electrons flow from source to drain). The resulting device is called n-channel MOSFET or NMOS.
- The channel is created by inverting the substrate surface from p-type to n-type. So, the induced channel is also called inversion layer.

Physical structure of the enhancement-type NMOS transistor

- The value of V_{GS} that is sufficient to form a conducting channel is called “threshold voltage”, V_t .
- For NMOS, V_t is positive and negative for PMOS.
- PMOS can be explained similarly but with n-type substrate and p+ sources and drain.
- V_t for n-channel FET is typically between 0.3 V and 1.0 V. The exact value depends on device fabrication process.
- The Gate-SiO₂-Channel combination forms a parallel-plate capacitor. An electric field hence develops in the vertical direction.
- This electric field controls the amount of charge in the channel and hence, it determines the channel conductivity and the amount of current that will flow through the channel when V_{DS} is applied. This gives rise to the field-effect transistor (FET).

Physical structure of the enhancement-type NMOS transistor

- To form the channel, the voltage across the parallel-plate capacitor described above must be greater than V_t .
- When $V_{DS} = 0$, the voltage at every point along the channel is zero. Then the voltage across the oxide (SiO_2) is then V_{GS} .
- The excess of V_{GS} over V_t is the “effective voltage or the overdrive voltage”, V_{ov} which determines the charge and conductivity of the channel.

$$V_{ov} = V_{GS} - V_t$$

- Quantity of charge in the channel: Recall that,

$$Q = CV;$$

And

$$C = \epsilon A/d,$$

where C is the capacitance, A is the area and d is the distance

- Here, $A = W \times L$, where W is the width of channel and L is the length of the channel

Physical structure of the enhancement-type NMOS transistor

- Capacitance of Gate-SiO₂-Channel combination is:

$$C = \frac{\varepsilon_{\text{ox}}WL}{t_{\text{ox}}} = C_{\text{ox}}WL$$

where $C_{\text{ox}} = \varepsilon_{\text{ox}}/t_{\text{ox}}$, t_{ox} is the thickness of the oxide, ε_{ox} is the permittivity of SiO₂ with a value of $3.9 \times \varepsilon_0$ and $\varepsilon_0 = 8.854 \times 10^{-12} \text{ F/m}$.

- Therefore,

$$Q = C_{\text{ox}} \times W \times L \times V_{\text{ov}}$$

where V_{ov} is the effective voltage across the capacitor with $V_{\text{DS}} = 0$.

- Note that as V_{ov} is increased, the quantity of charge in the channel increases proportionately.

FET Operation with Small VDS

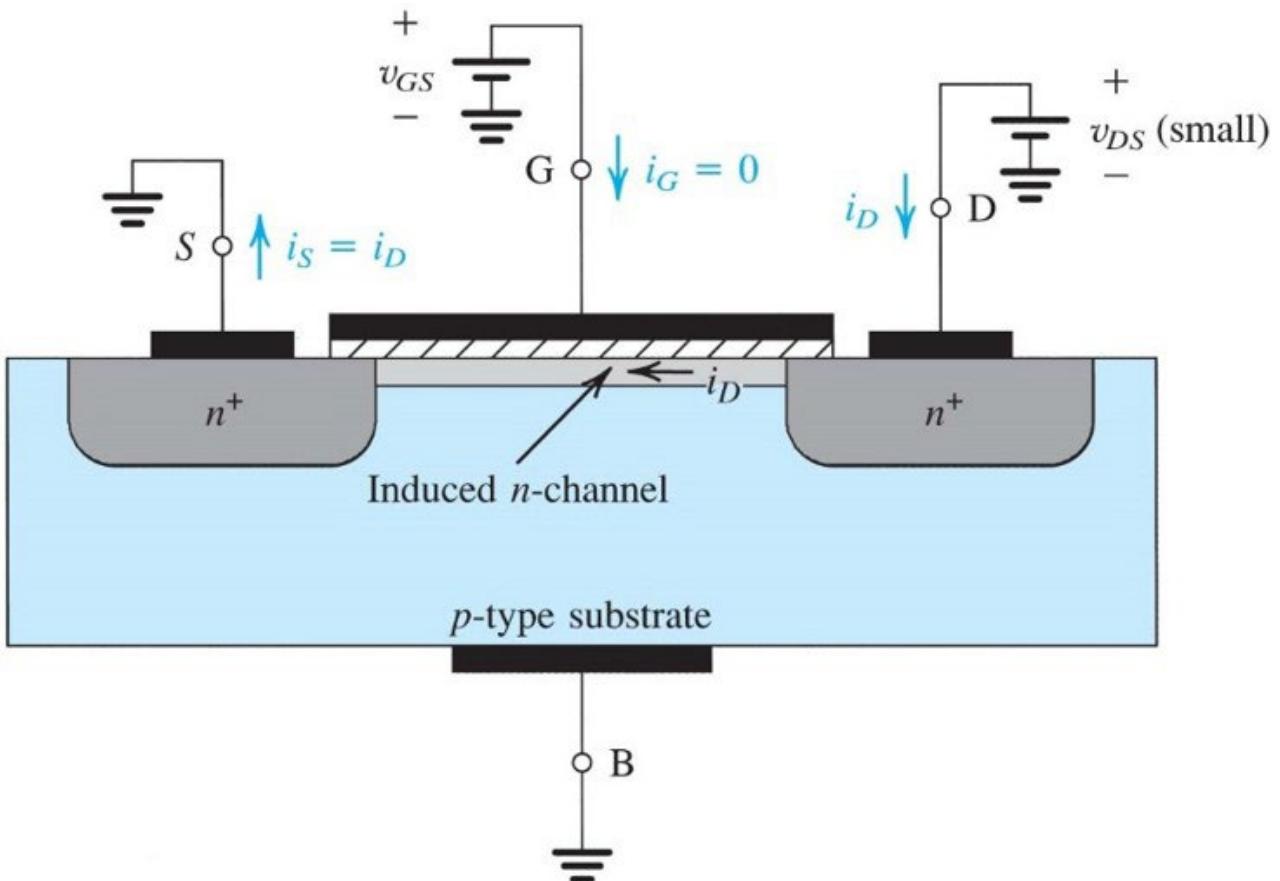


Figure 5: An NMOS transistor with $V_{GS} > V_t$ and with a small V_{DS} applied.

Small VDS

- With the channel already induced, consider a small voltage V_{DS} , (i.e. $V_{DS} \leq 50$ mV), applied between the drain and the source.
- The small V_{DS} causes current i_D to flow through the n-channel.
- i_D is carried by free electrons travelling from source to drain. Thus, current flows from drain to source.
- Since V_{DS} is small, it can be assumed that the voltage between the gate and various points along the channel remains approximately constant and equal to the value at the source end. Thus, the effective voltage across the G-SiO₂ remains V_{ov} .
- To estimate the value of i_D : Recall that

$$Q = C_{ox} W L V_{ov}$$

- So, charge per unit length is:

$$\left| \frac{Q}{L} \right| = C_{ox} W V_{ov}$$

Small VDS

- Voltage V_{DS} establishes an electric field, E across the length of the channel,

$$E = \frac{V_{DS}}{L}$$

- This electric field in turns causes the channel electrons to drift towards the drain with a velocity given by:

$$\text{Electric field drift velocity} = \mu_n |E| = \mu_n \frac{V_{DS}}{L}$$

μ_n is the electron mobility. $i = \frac{\partial Q}{\partial t} = \frac{\partial Q}{\partial L} \times \frac{\partial L}{\partial t}$

- Thus, i_D can be found by multiplying charge per length by electron drift velocity.

$$i_D = \left[\mu_n C_{ox} \left(\frac{W}{L} \right) V_{ov} \right] V_{DS}$$

Small VDS

- Note that:

$$V_{ov} = V_{GS} - V_t$$

- Thus:

$$i_D = \left[\mu_n C_{ox} \left(\frac{W}{L} \right) (V_{GS} - V_t) \right] V_{DS}$$

- So for small V_{DS} , the channel behaves as a linear resistance whose value is controlled by V_{ov} .
- The conductance, g_{DS} , of the channel is thus:

$$g_{DS} = \frac{i_D}{V_{DS}} = \mu_n C_{ox} \left(\frac{W}{L} \right) V_{ov} = \mu_n C_{ox} \left(\frac{W}{L} \right) (V_{GS} - V_t) \quad K'_n = \mu_n C_{ox}$$

where K'_n is the process transconductance parameter. Its value is dependent on the MOSFET fabrication technology.

Small VDS

$$K_n = K'_n \frac{W}{L} = \mu_n C_{\text{ox}} \frac{W}{L}$$

where K_n is the MOSFET transconductance parameter.

- With V_{DS} kept small, the MOSFET behaves as a linear resistance r_{DS} whose value is governed by V_{GS} .

$$r_{\text{DS}} = \frac{1}{g_{\text{DS}}} = \frac{1}{\mu_n C_{\text{ox}} \left(\frac{W}{L}\right) V_{\text{ov}}}$$

- Note that for MOSFET to conduct, a channel has to be induced. Increasing the V_{GS} above V_t enhances the channel hence the name “enhanced-mode operation” and enhancement type MOSFET. Current that leaves the source, i_S is equal to the current that enters the drain, i_D , so gate current $i_G = 0$.

FET Operation with VDS no longer small

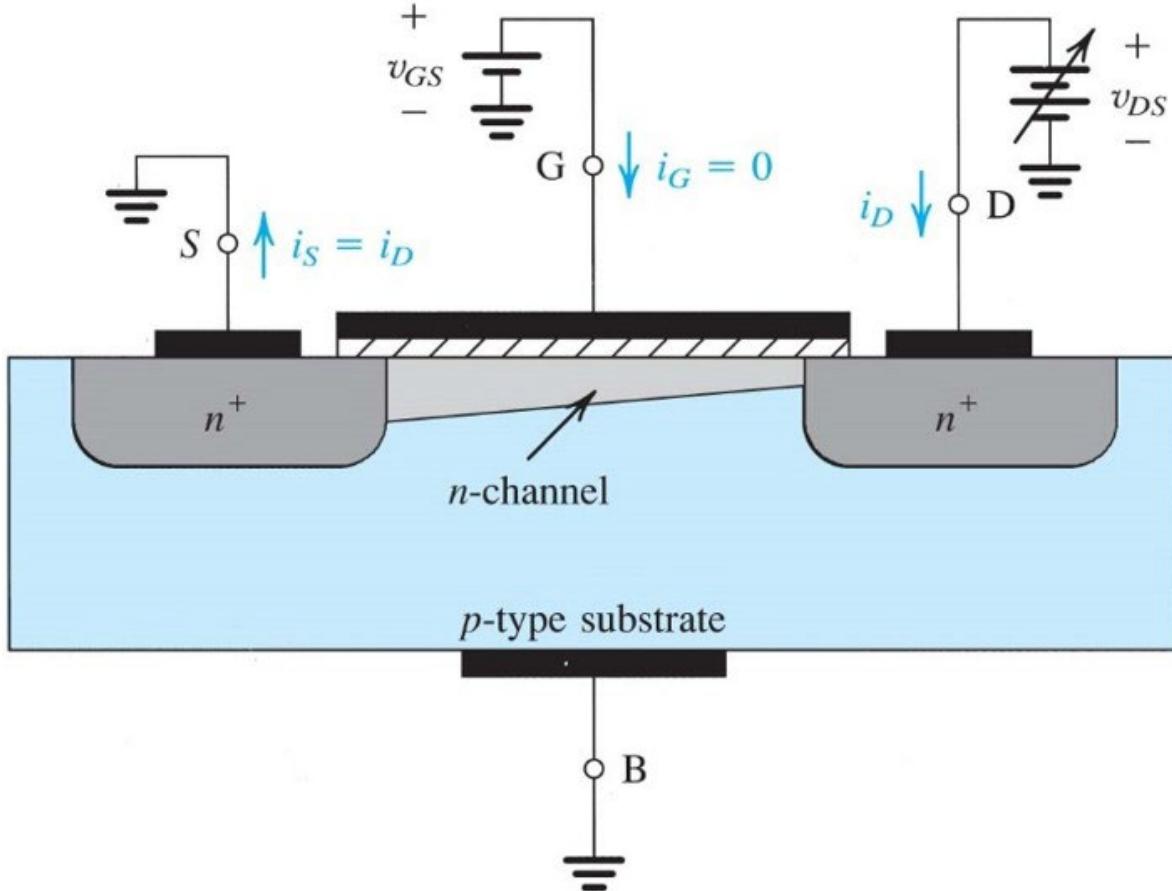


Figure 6: Operation of the enhancement NMOS transistor as V_{DS} is increased. The induced channel acquires a tapered shape, and its resistance increases as V_{DS} is increased. Here, V_{GS} is kept constant at a value $> V_t$; $V_{GS} = V_t + V_{ov}$.

VDS no longer small

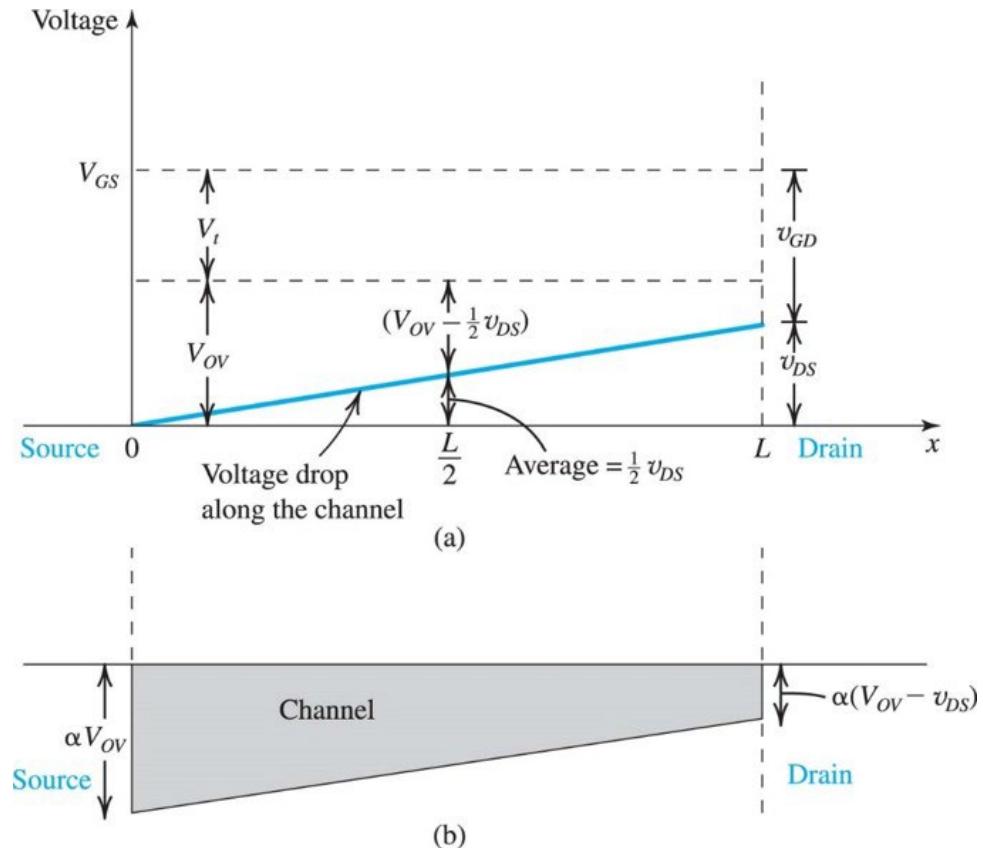


Figure 7: (a) For a MOSFET with $V_{GS} = V_t + V_{OV}$, application of V_{DS} causes the voltage drop along the channel to vary linearly, with an average value of $0.5 \times V_{DS}$ at the midpoint. Since $V_{GD} > V_t$, the channel still exists at the drain end (b) The channel shape corresponding to the situation in (a).

VDS no longer small

- Let VGS be held constant but higher than V_t .
- As VDS increases, it means that the voltage with respect to the source increases from zero to VDS as we travel along the channel from source to drain.
- So the voltage between the G and points along the channel decreases from $V_{GS} = V_t + V_{ov}$ at the source to:

$$V_{GD} = V_{GS} - V_{DS} = V_t + V_{ov} - V_{DS} \text{ at the drain end.}$$

- The channel, which depends on the voltage between gate and points along the channel, is now no longer uniform. The channel will take a tapered shape.
- As VDS increases, the channel becomes more tapered and its resistance increases correspondingly.

VDS no longer small

- So

$$i_D = K'_n \left(\frac{W}{L} \right) \left(V_{ov} - \frac{1}{2} V_{DS} \right) V_{DS}$$

$$i_D = K'_n \left(\frac{W}{L} \right) \left(V_{ov} V_{DS} - \frac{1}{2} V_{DS}^2 \right)$$

- iD has a non-linear relationship with VDS.

For $V_{DS} \geq V_{OV}$ - Channel pinch-off and current saturation

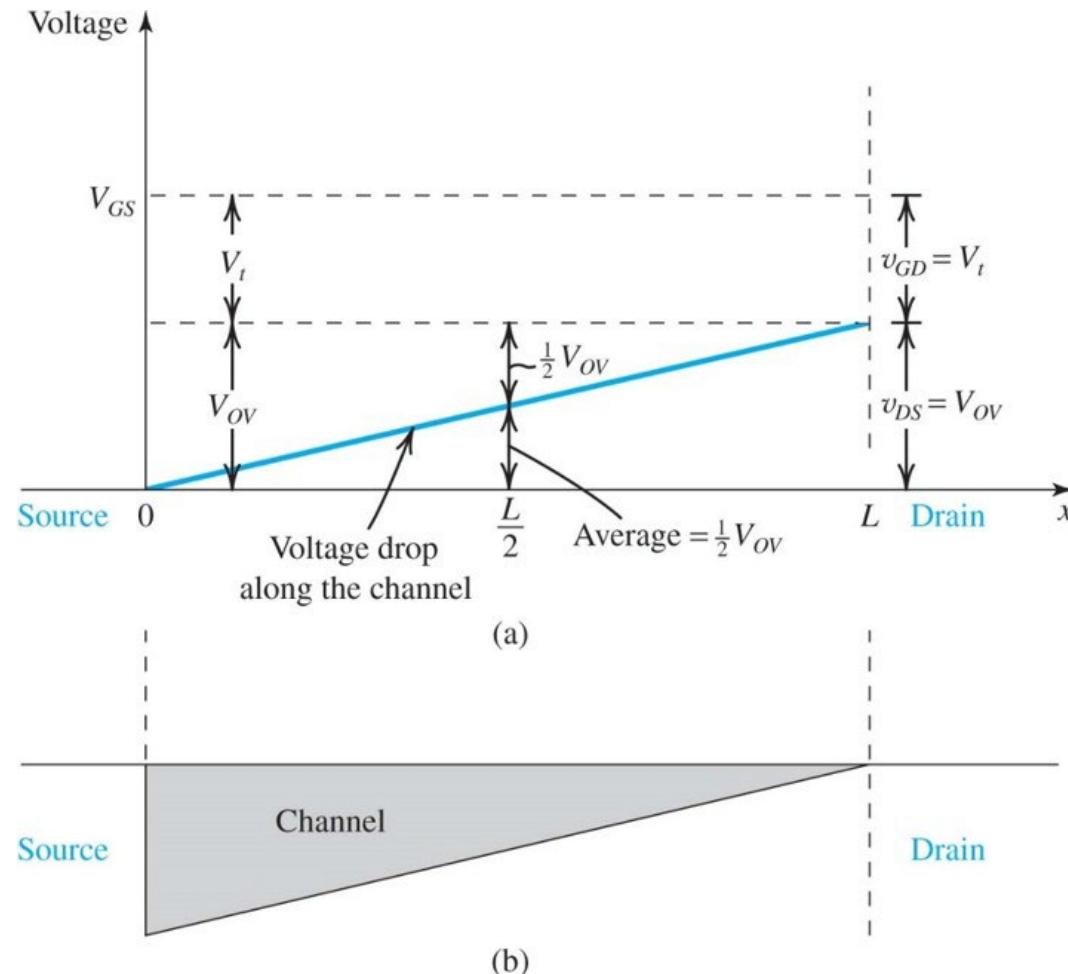


Figure 8: Operation of MOSFET with $V_{GS} = V_t + V_{OV}$, V_{DS} is increased to V_{OV} .

For $V_{DS} \geq V_{ov}$ - Channel pinch-off and current saturation

- As V_{DS} increases, the channel becomes tapered and when $V_{DS} = V_{ov}$ (i.e $V_{GD} = V_t$), the channel depth at the drain end reduces to zero. This is the condition referred to as channel pinch-off.
- Increasing V_{DS} beyond this value, (i.e $V_{DS} \geq V_{ov}$) has no effect on the channel shape and charge. And the current through the channel remains constant at the value reached for $V_{DS} = V_{ov}$.
- The MOSFET is now said to be in saturation mode.
- Thus, the drain current saturates at the value when $V_{DS} = V_{ov}$.

For $V_{DS} \geq V_{ov}$ - Channel pinch-off and current saturation

- Therefore

$$i_{D(sat)} = \frac{1}{2} K'_n \left(\frac{W}{L} \right) V_{DS}^2$$

where $i_{D(sat)}$ is i_D in saturation mode.

- The voltage V_{DS} at which saturation occurs

$$V_{DS(sat)} = V_{ov} = V_{GS} - V_t$$

For $V_{DS} \geq V_{OV}$ - Channel pinch-off and current saturation

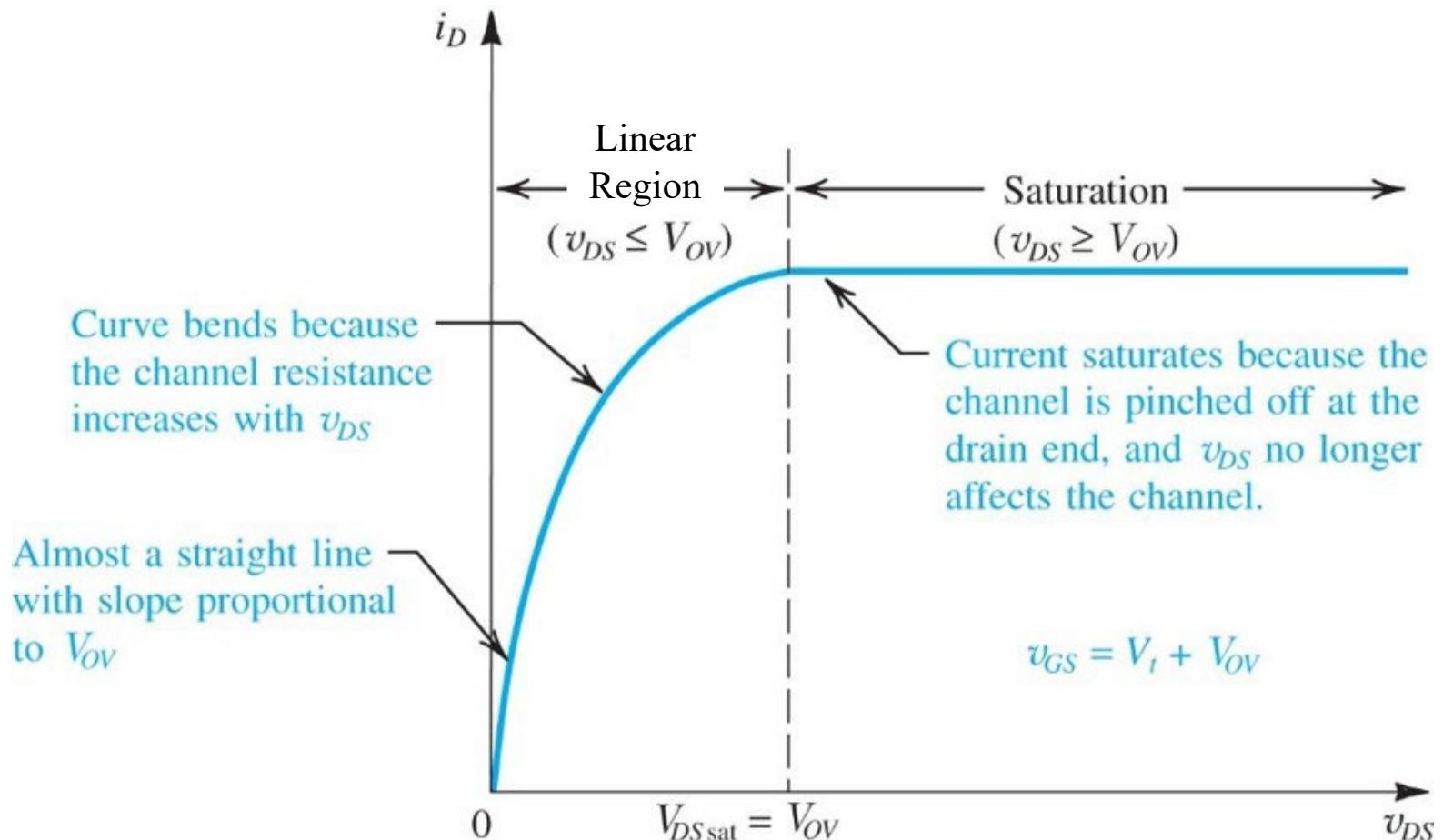


Figure 9: The drain current i_D versus the drain-to-source voltage V_{DS} for an enhancement-type NMOS transistor operated with $V_{GS} = V_t + V_{OV}$.

For $V_{DS} \geq V_{OV}$ - Channel pinch-off and current saturation

- Note that the channel pinch-off does not mean channel blockage. Current continues to flow through the pinched-off channel. Electrons that reach the drain end of the channel are accelerated through the depletion region that exists there and into the drain terminal.

PMOS and Complementary MOS (CMOS)

- The description of the physical operation of the PMOS follows that of CMOS except for the sign reversal of all voltages.
- The advantage of NMOS over PMOS reason for this is that electron mobility μ_n is higher by a factor of 2 to 4 than the hole mobility μ_p . This implies that NMOS electrons have greater gains and speed of operation than PMOS devices.
- Technology is now in place to fabricate both NMOS and PMOS transistors on the same chip. This is called complementary MOS (CMOS). CMOS is currently the dominant IC technology.
- In CMOS, the NMOS transistor is implemented directly in the p-type substrate.

PMOS and Complementary MOS (CMOS)

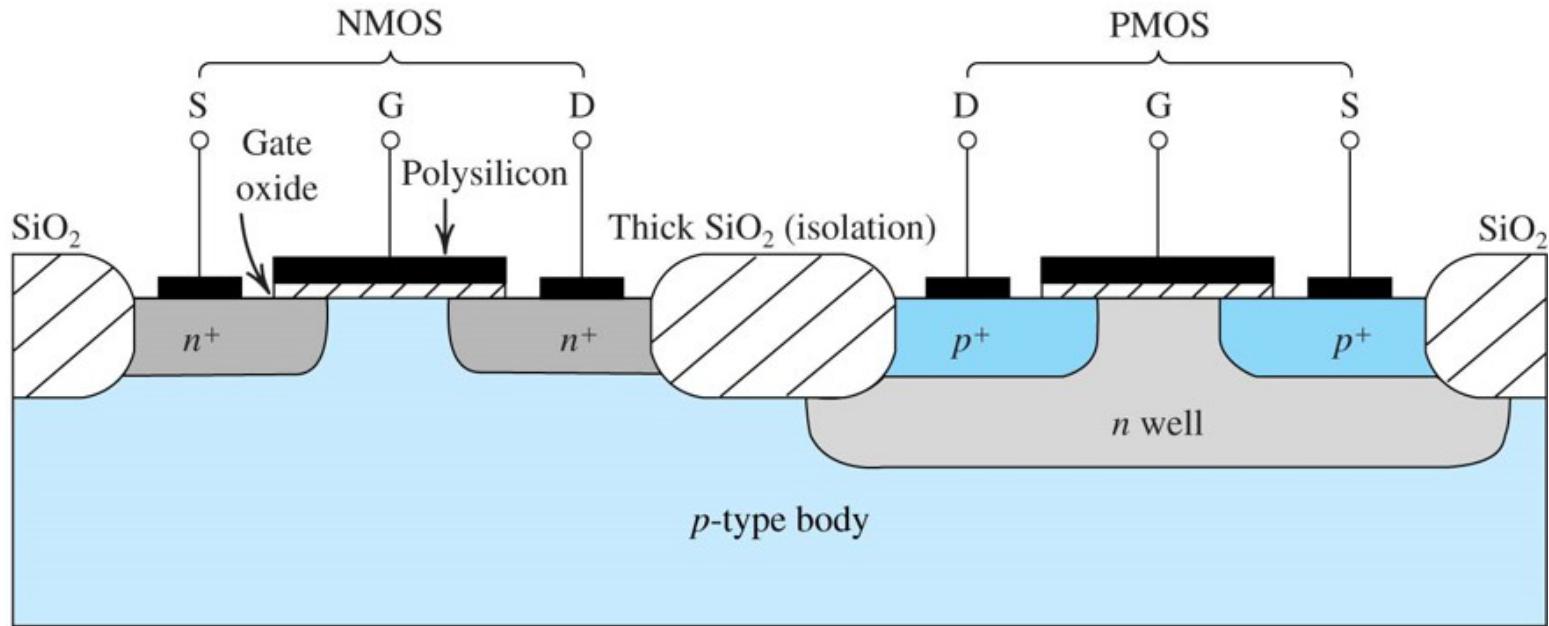


Figure 10: Cross section of a CMOS integrated circuit. Note that the PMOS transistor is formed in a separate n-type region, known as an n well. Another arrangement is also possible in which an n-type substrate (body) is used and the n device is formed in a p well. Not shown are the connections made to the p-type body and to the n well; the latter functions as the body terminal for the p-channel device.

PMOS and Complementary MOS (CMOS)

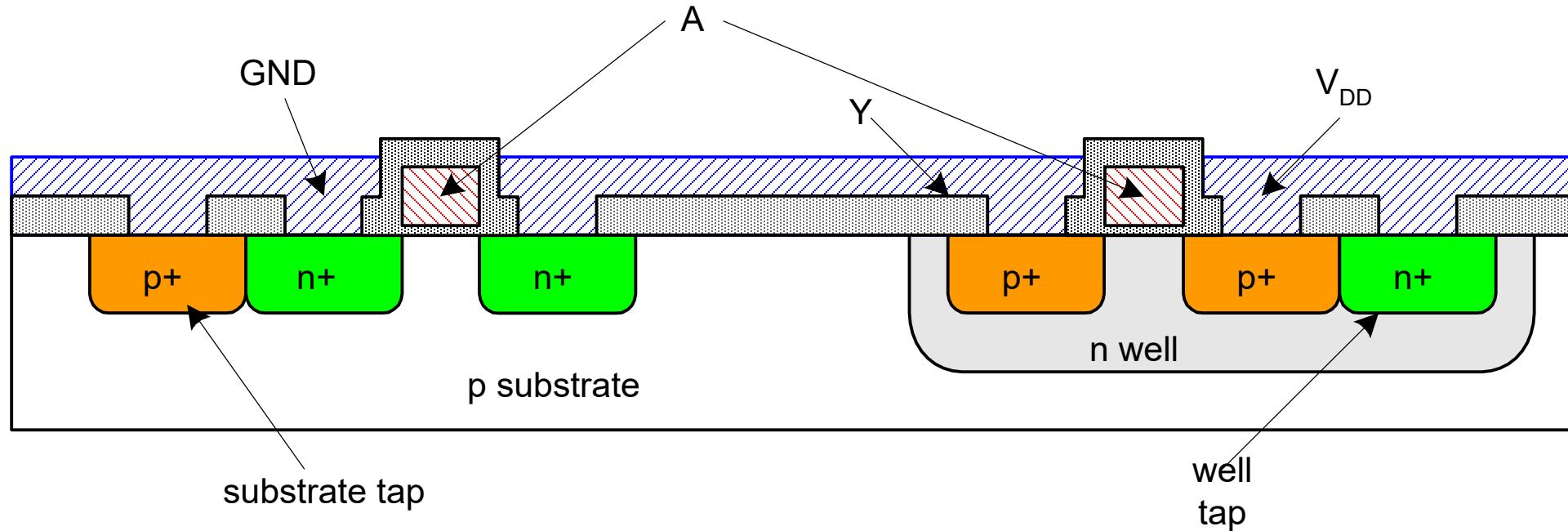
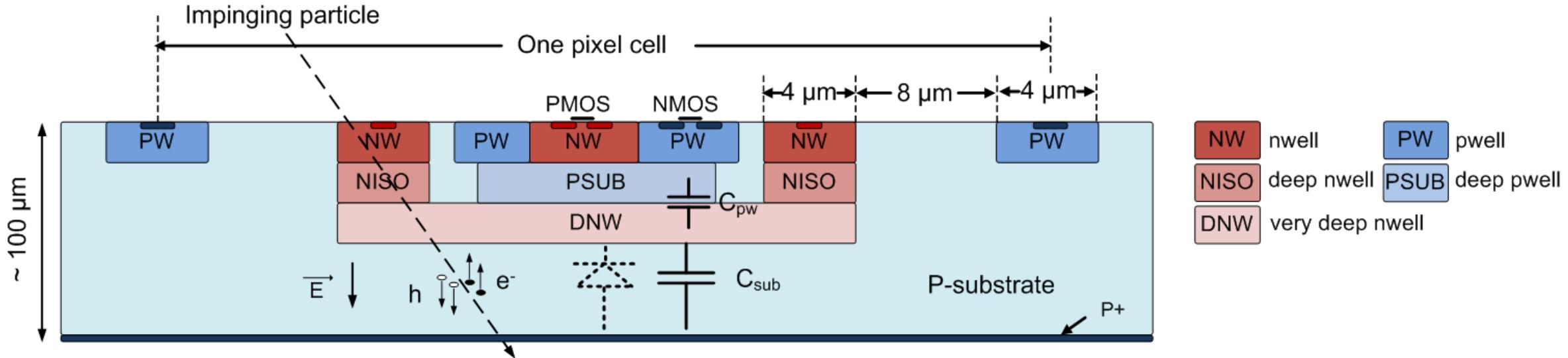


Figure 10: Cross section of a CMOS integrated circuit. Note that the PMOS transistor is formed in a separate n-type region, known as an n well. Another arrangement is also possible in which an n-type substrate (body) is used and the n device is formed in a p well. Not shown are the connections made to the p-type body and to the n well; the latter functions as the body terminal for the p-channel device.

PMOS and Complementary MOS (CMOS)

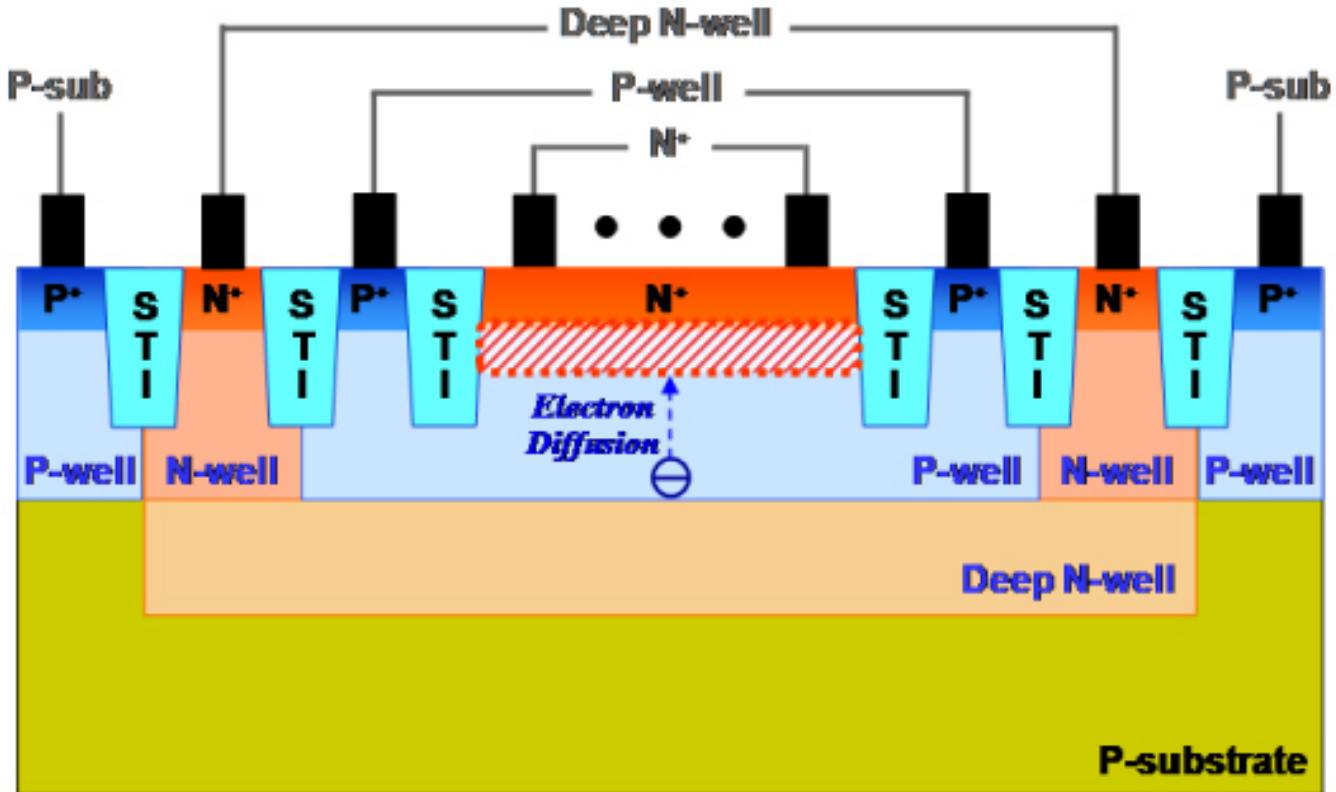
- The PMOS transistor is fabricated in a specially created n- region called n-well.
- Multiple isolated n-wells can be fabricated on a single circuit and can be connected to different potentials in different ways.
- NMOS and PMOS are isolated from each other by a thick region of oxide (insulator).
- In CMOS, (for an n-well process), the p-bulk (substrate) is common throughout the integrated circuit. The bulk is usually connected to the most negative supply, (V_{ss}). This ensures that the Source/Drain junction diodes of the NMOS are always reverse-biased.

PMOS and Complementary MOS (CMOS)



Radiation hard DMAPS pixel sensors in 150 nm CMOS technology for operation at LHC - Barbero, M. et al - arXiv:1911.01119

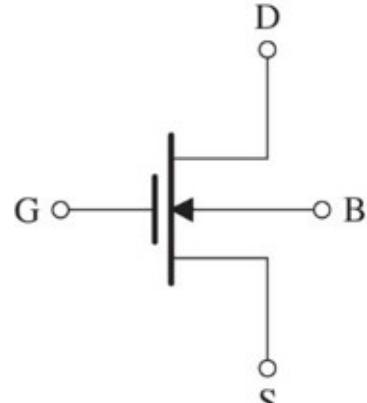
PMOS and Complementary MOS (CMOS)



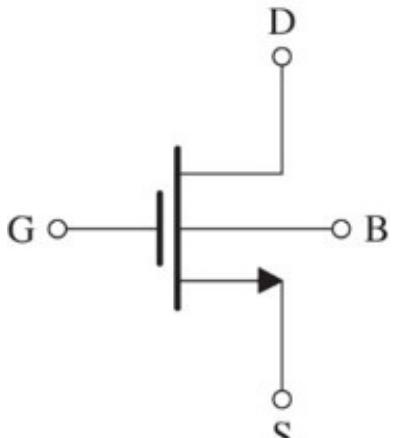
Myung-Jae Lee, Woo-Young Choi, "A silicon avalanche photodetector fabricated with standard CMOS technology with over 1 THz gain-bandwidth product," Opt. Express **18**, 24189-24194 (2010);
<https://www.osapublishing.org/oe/abstract.cfm?uri=oe-18-23-24189>

Circuit symbol

- MOSFET is a symmetrical device, but it is often useful in circuit design to designate one terminal as source and the other as drain.
- Note that in an NMOS, the drain is always positive relative to the source.



(a)



(b)

Circuit symbol

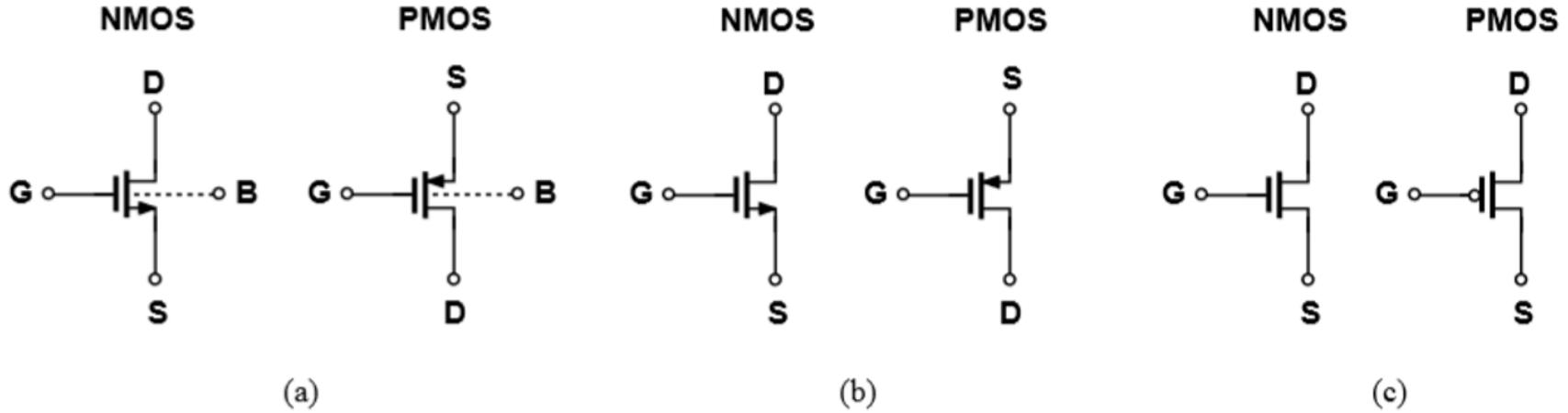
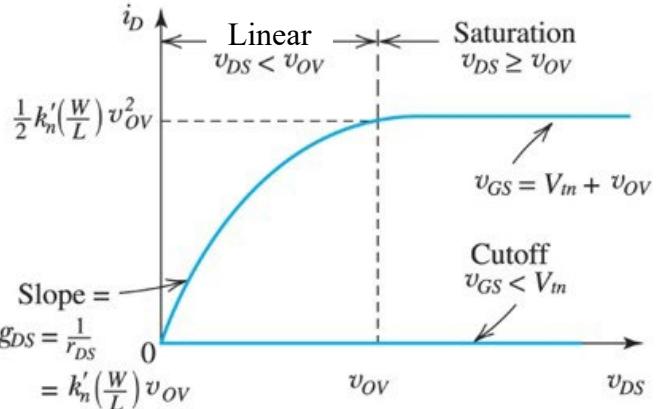
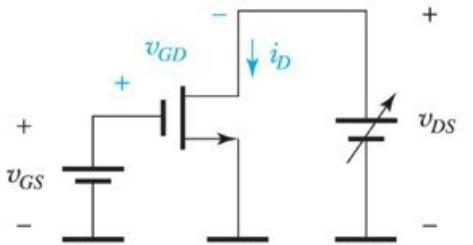


Figure 13: Modified circuit symbols with an arrowhead on the source terminal to distinguish it from the drain and to indicate device polarity.

Substrate is denoted by B (bulk). PMOS source is positioned on top since it has a higher potential than the gate. Most circuits have NMOS and PMOS bulk tied to ground and VDD, respectively, so we tend to omit the connections (b,c). Digital circuits tend to incorporate switch symbols (c).

iD – VDS characteristics

Table 5.1 Regions of Operation of the Enhancement NMOS Transistor



- $v_{GS} < V_{tn}$: no channel; transistor in cutoff; $i_D = 0$
- $v_{GS} = V_{tn} + v_{OV}$: a channel is induced; transistor operates in the triode region or the saturation region depending on whether the channel is continuous or pinched off at the drain end;

Linear Region

Continuous channel, obtained by:

$$v_{GD} > V_{tn}$$

or equivalently:

$$v_{DS} < v_{OV}$$

Then,

$$i_D = k'_n \left(\frac{W}{L} \right) \left[(v_{GS} - V_{tn}) v_{DS} - \frac{1}{2} v_{DS}^2 \right]$$

or equivalently,

$$i_D = k'_n \left(\frac{W}{L} \right) \left(v_{OV} - \frac{1}{2} v_{DS} \right) v_{DS}$$

Saturation Region

Pinched-off channel, obtained by:

$$v_{GD} \leq V_{tn}$$

or equivalently:

$$v_{DS} \geq v_{OV}$$

Then

$$i_D = \frac{1}{2} k'_n \left(\frac{W}{L} \right) (v_{GS} - V_{tn})^2$$

or equivalently,

$$i_D = \frac{1}{2} k'_n \left(\frac{W}{L} \right) v_{OV}^2$$

Contents

- Three possible regions of operations are: the cut-off region, the linear region (~~triode~~), and the saturation region.
- Cut-off and linear regions are useful if MOSFET is used as a switch.
- Saturation region is useful for operating the MOSFET as an amplifier.
- In liner region, $V_{DS} < V_{ov}$;

$$i_D = K'_n \left(\frac{W}{L} \right) \left(V_{ov} V_{DS} - \frac{1}{2} V_{DS}^2 \right)$$

- Saturation sets in when $V_{DS} = V_{ov}$. Therefore, at the boundary between linear and the saturation region.

$$i_D = \frac{1}{2} K'_n \left(\frac{W}{L} \right) V_{DS}^2$$

iD – VDS characteristics

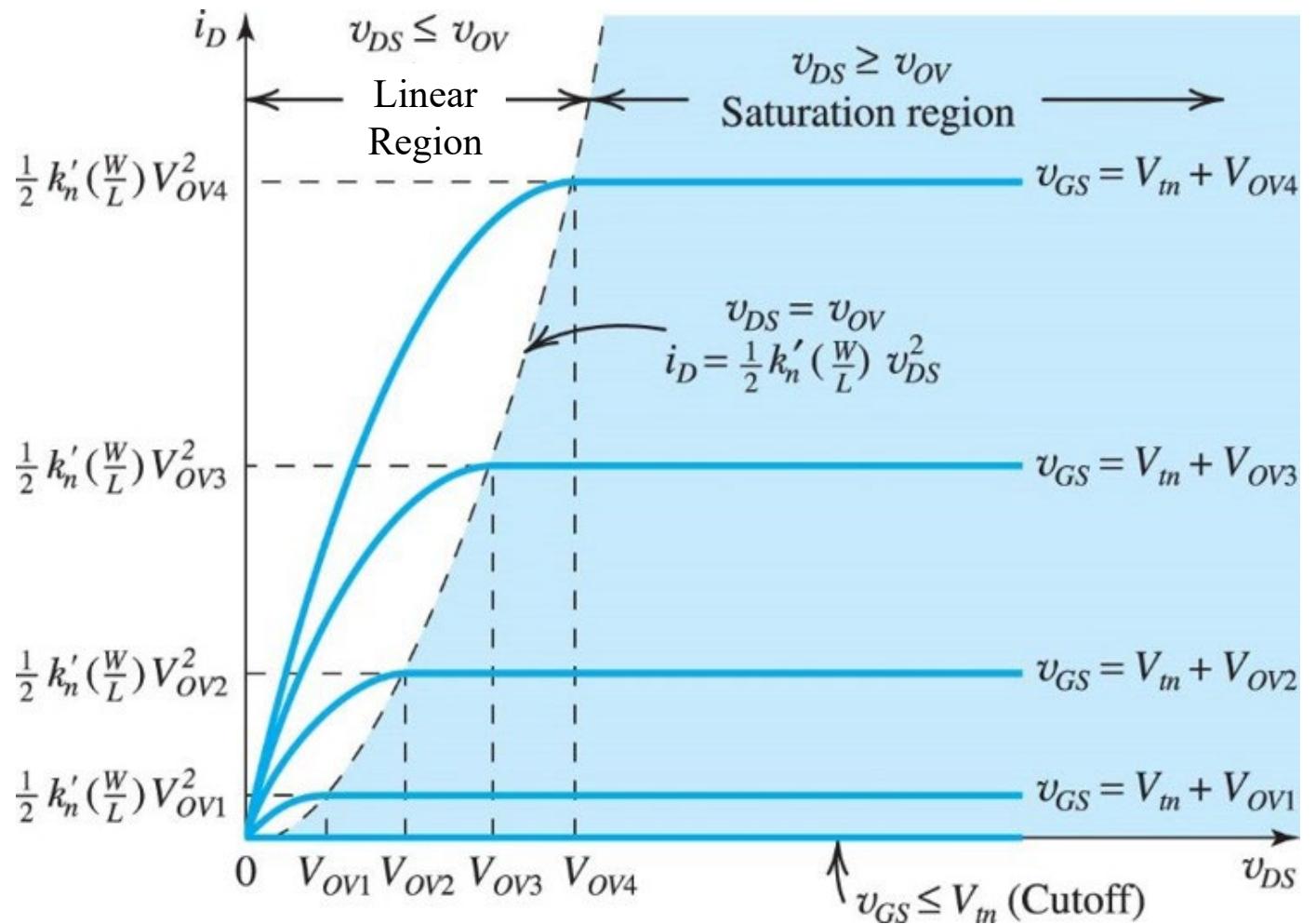


Figure 14: The iD – VDS characteristics for an enhancement-type NMOS transistor.

iD – VGS characteristics

- Consider MOSFET in the saturation region. That is $V_{DS} = V_{ov}$, and

$$i_D = \frac{1}{2} K'_n \left(\frac{W}{L} \right) (V_{GS} - V_t)^2$$

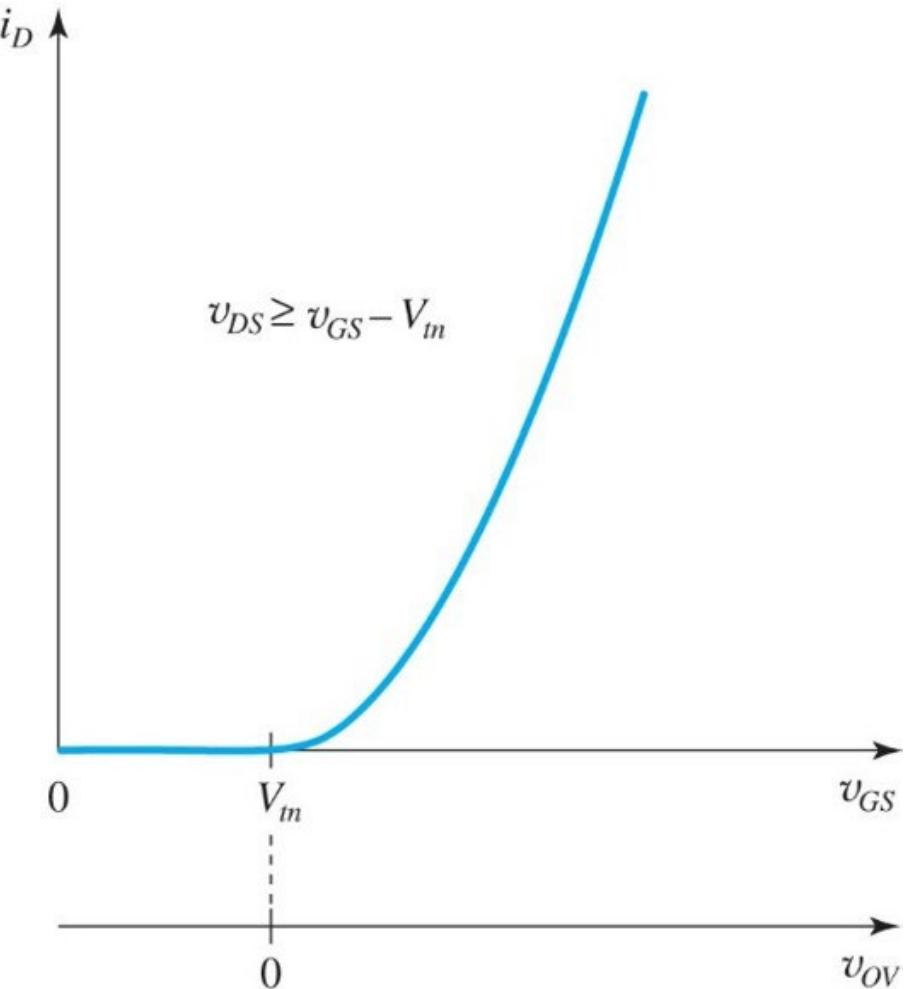
$$i_D = \frac{1}{2} K'_n \left(\frac{W}{L} \right) V_{ov}^2$$

and $V_{GS} - V_t = V_{ov}$.

- Note that in the saturation region, the current is at maximum and does not depend on V_{DS} for as long the output resistance is infinite i.e $r_o = \infty$.
- So in saturation, MOSFET operates as a voltage-controlled current source i.e i_D is controlled by V_{ov} or V_{GS} .

iD – VGS characteristics

Figure 15: The iD – VDS characteristic of an NMOS transistor operating in the saturation region. The iD – VDS characteristic can be obtained by simply relabelling the horizontal axis, that is, shifting the origin to the point The VGS = Vtn.



iD – VGS characteristics

- Effect of finite output resistance: Above description assumes that iD is independent of VDS . This implies output resistance is infinite. This assumption is based on the premise that once the channel is pinched-off at the drain end, further increases in VDS have no effect on the channel shape.
- But in practice, increasing VDS beyond Vov does affect the channel. As VDS is increased, the channel pinch-off point is slightly moved away from the drain towards the source.
- This causes the effective channel length to reduce from L to $(L - \Delta L)$. This phenomenon is called "channel-length modulation".
- Since $iD \propto (1/\text{channel length})$. Therefore, iD increases with VDS .

iD – VGS characteristics

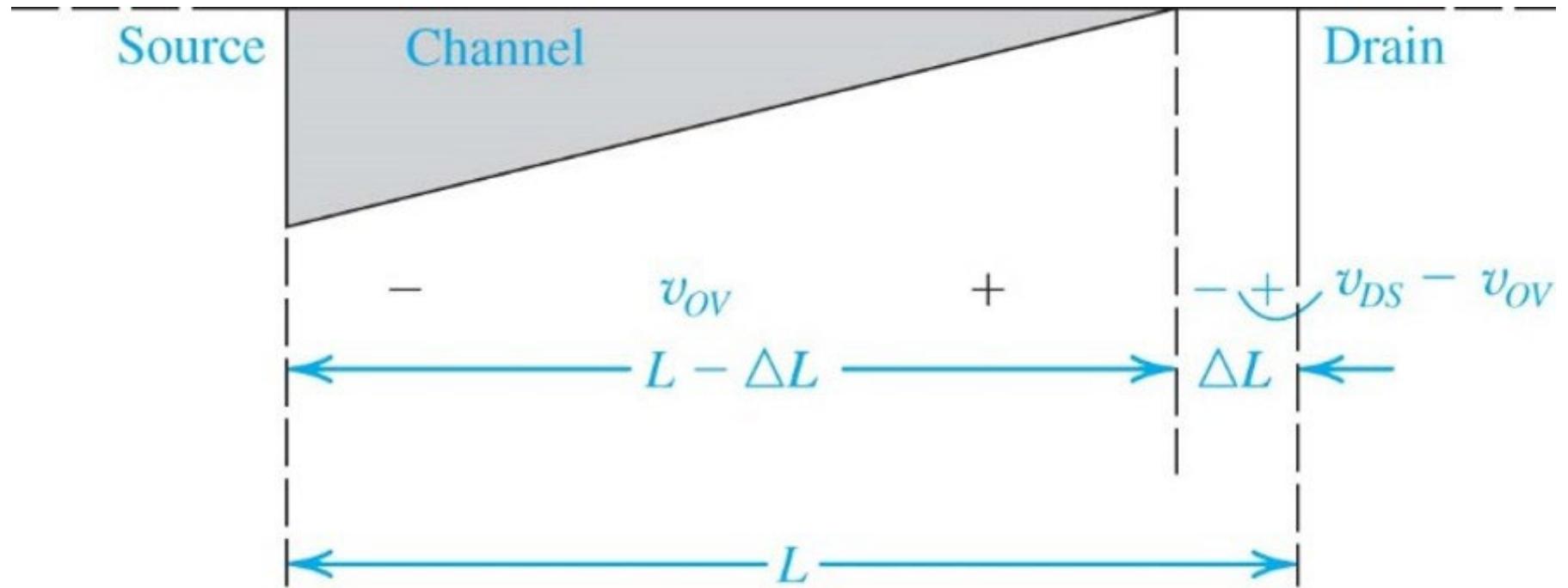


Figure 16: Increasing V_{DS} beyond V_{DSsat} causes the channel pinch-off point to move slightly away from the drain, thus reducing the effective channel length (by ΔL).

iD – VGS characteristics

- So with “channel length modulation”:

$$i_D = \frac{1}{2} K'_n \left(\frac{W}{L} \right) (V_{GS} - V_t)^2 (1 + \lambda V_{DS})$$

where λ is the device parameter with unit V^{-1} . λ depends on both the process technology used to fabricate the device and on the channel length, L.

$$\lambda \propto \frac{1}{L}$$

- With “channel length modulation”, the output resistance is no longer infinite. It is given as:

$$r_o \equiv \left[\frac{\partial i_D}{\partial V_{DS}} \right]^{-1} \Bigg|_{V_{GS} \text{ constant}}$$

Transistors

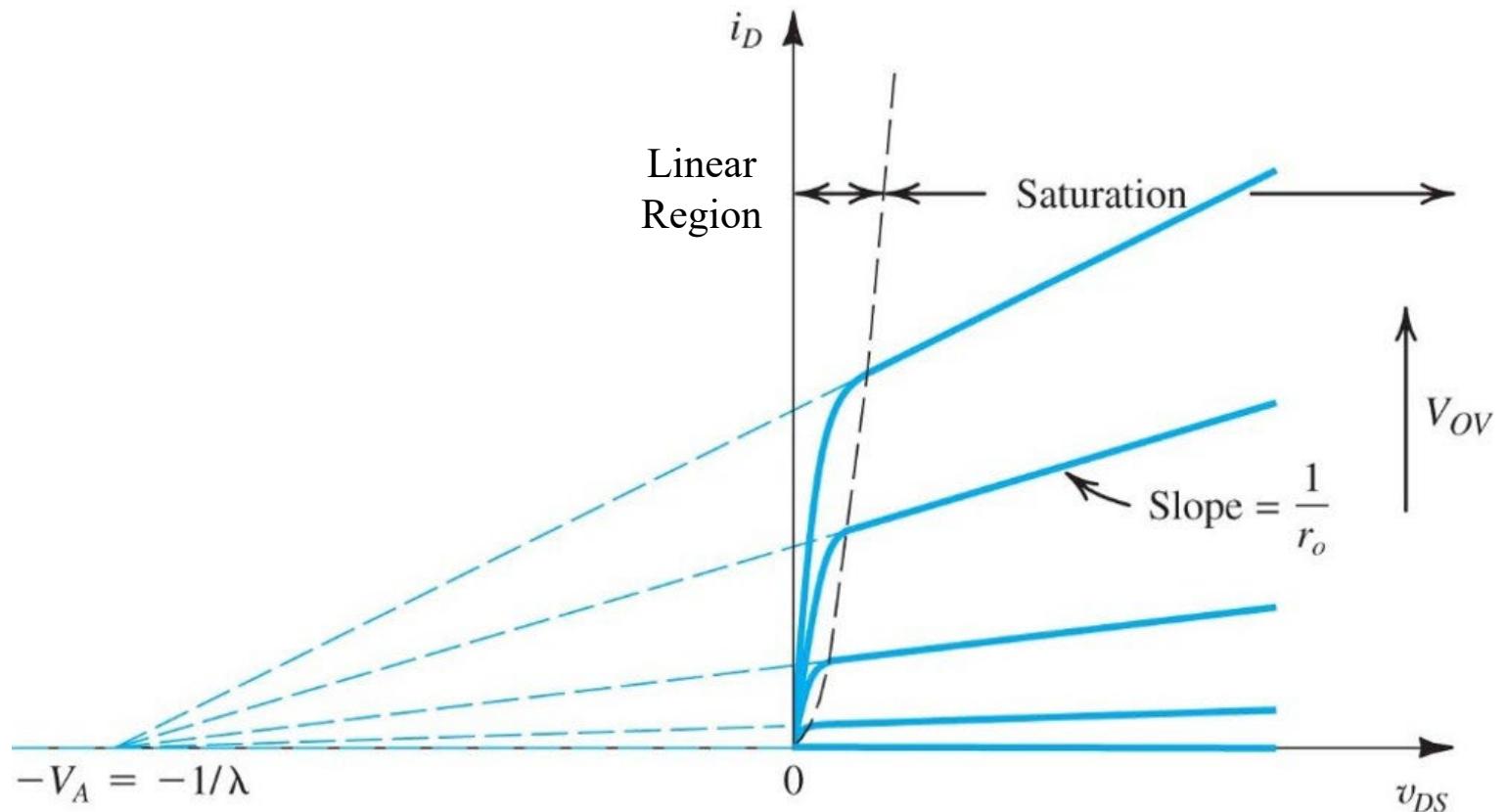


Figure 17: Effect of V_{DS} on i_D in the saturation region. The MOSFET parameter V_A depends on the process technology and, for a given process, is proportional to the channel length L .

Transistors

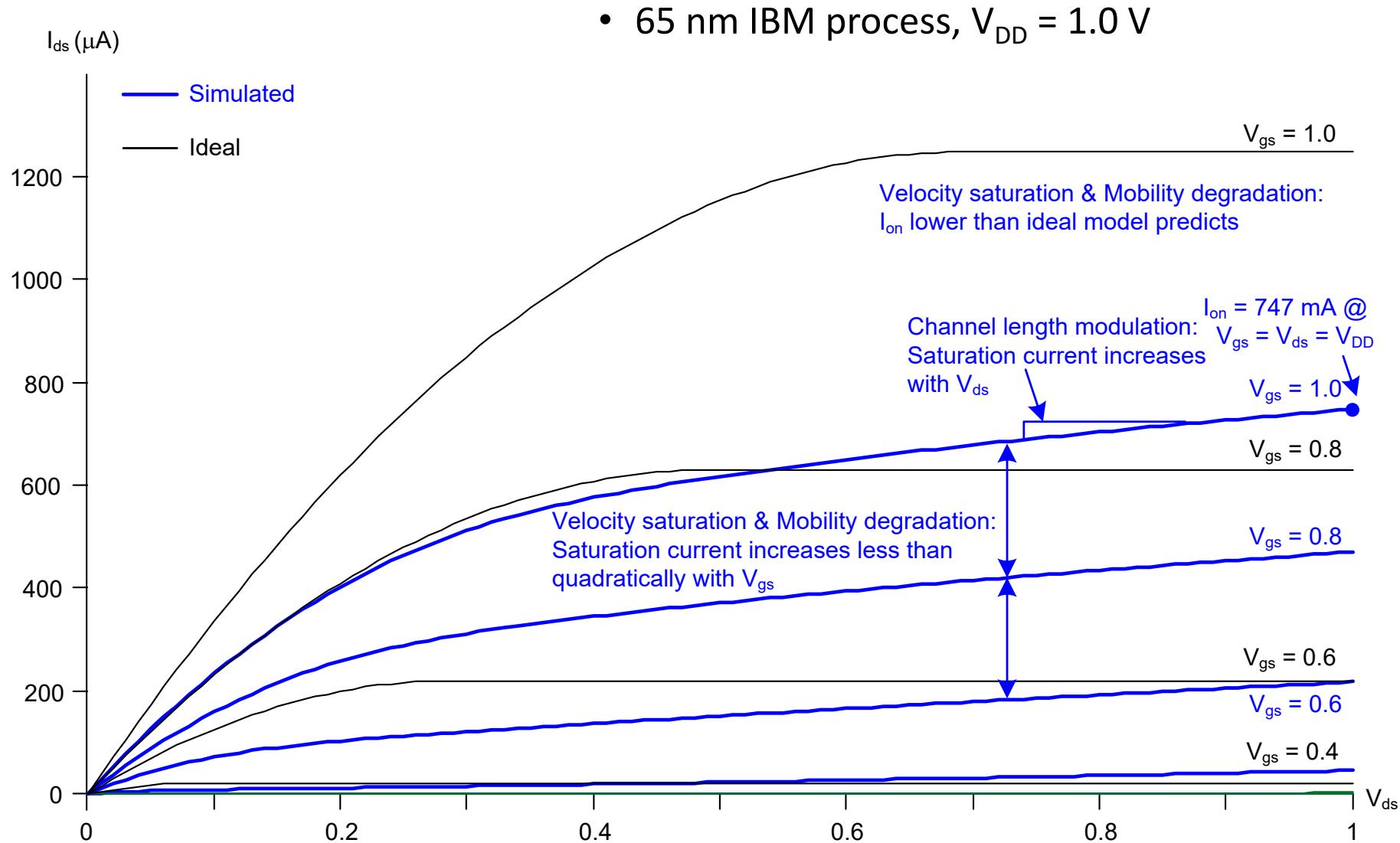
- In modern processes there is no longer a “metal” oxide in the gate.
- Hence from now on we call it FET (Field Effect Transistor)
- nFET, pFET, FinFET, etc.

FET I-V Summary

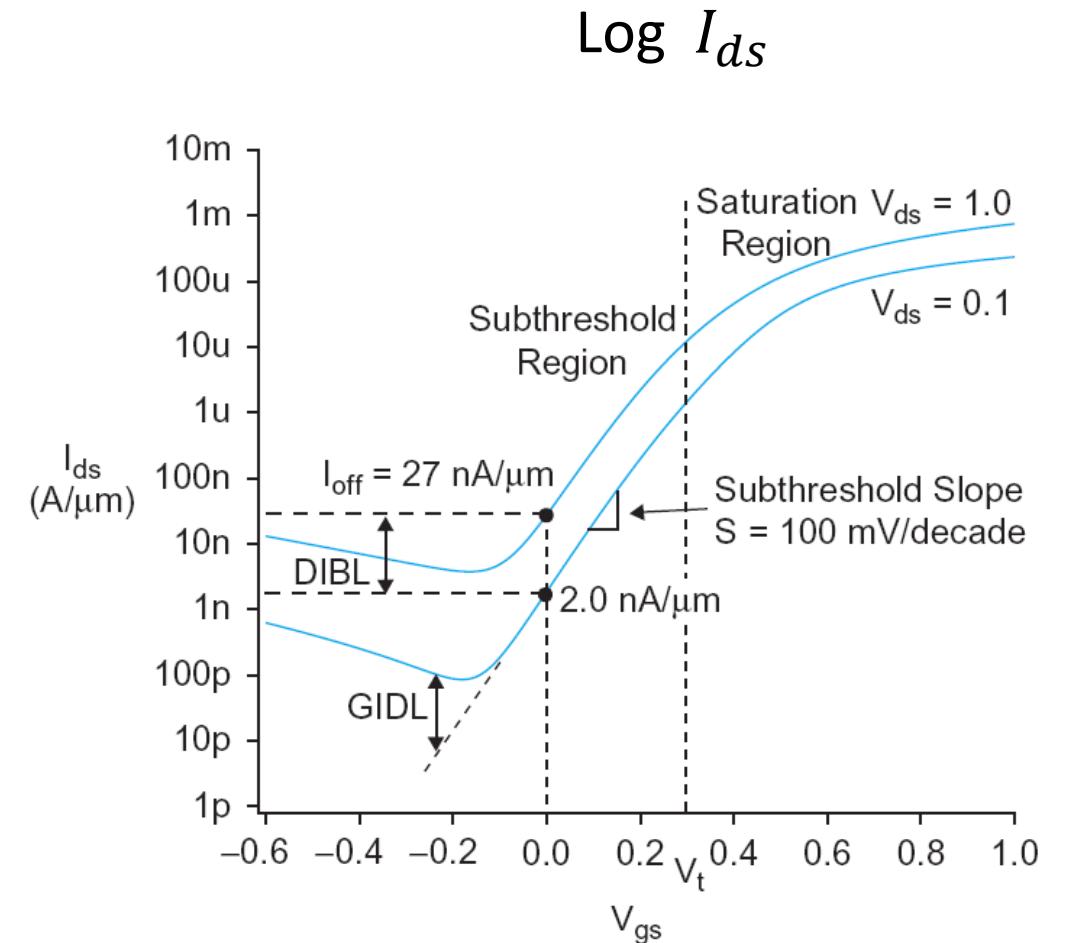
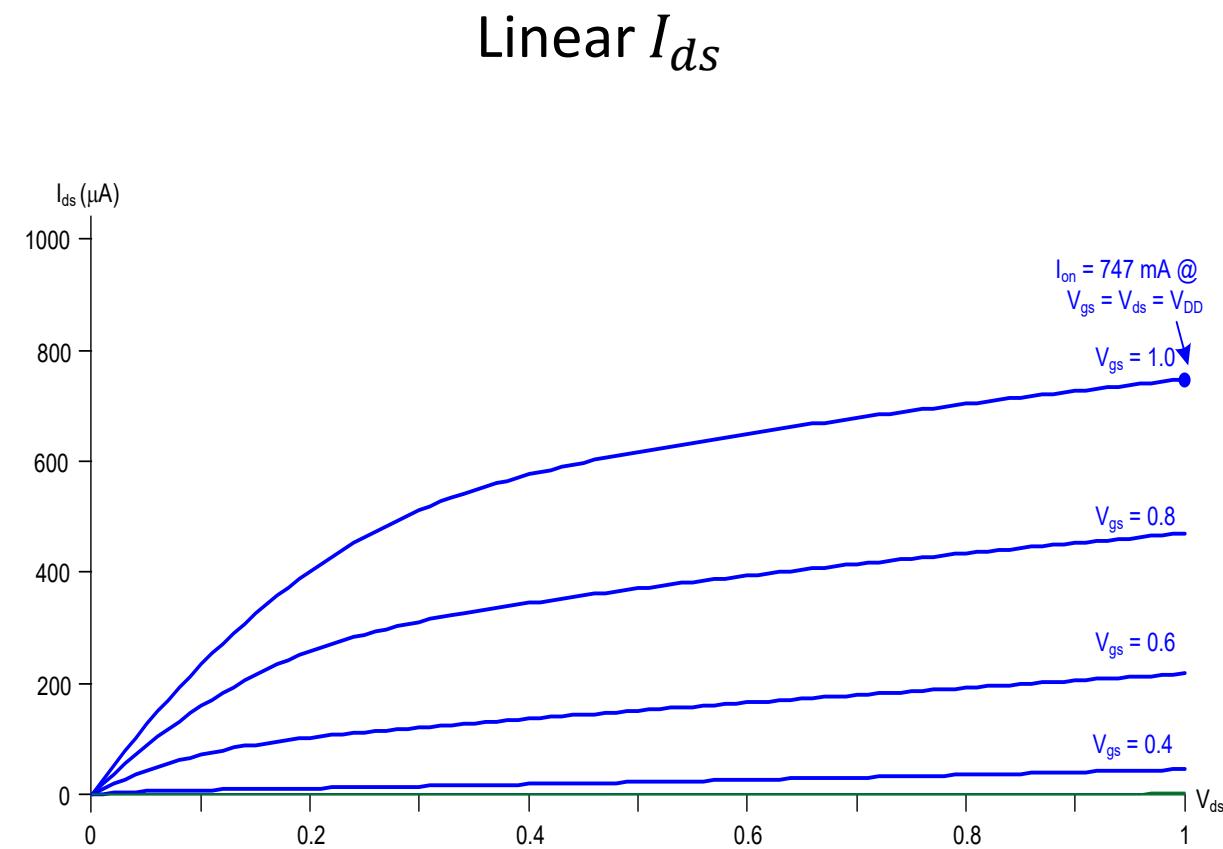
- Shockley 1st order transistor models (for nFET):

$$I_{ds} = \begin{cases} 0 & V_{gs} < V_t \quad \text{cutoff} \\ \beta \left(V_{gs} - V_t - \frac{V_{ds}}{2} \right) V_{ds} & V_{ds} < V_{gs} - V_t \quad \text{linear} \\ \frac{\beta}{2} (V_{gs} - V_t)^2 & V_{ds} > V_{gs} - V_t \quad \text{saturation} \end{cases}$$

Ideal vs. Simulated nFET I-V Plot



ON and OFF Current



Channel Length Modulation

$$I_{ds} = \frac{\beta}{2} (V_{gs} - V_t)^2 (1 + \lambda V_{ds})$$

- λ = *channel length modulation coefficient*
 - not feature size
 - Empirically fit to I-V characteristics

Body Effect

- Body is a fourth transistor terminal
- V_{sb} affects the charge required to invert the channel
 - Increasing V_s or decreasing V_b increases V_t

$$V_t = V_{t0} + \gamma \left(\sqrt{\phi_s + V_{sb}} - \sqrt{\phi_s} \right)$$

- V_{t0} = nominal threshold voltage
- ϕ_s = *surface potential* at threshold

$$\phi_s = 2v_T \ln \frac{N_A}{n_i}$$

- Depends on doping level N_A
- And intrinsic carrier concentration n_i
- γ = *body effect coefficient*

$$\gamma = \frac{t_{ox}}{\epsilon_{ox}} \sqrt{2q\epsilon_{si}N_A} = \frac{\sqrt{2q\epsilon_{si}N_A}}{C_{ox}}$$

Variation

- Process
 - Threshold
 - Channel length
 - Interconnect dimensions
- Environment
 - Voltage
 - Temperature
- Aging/Wearout

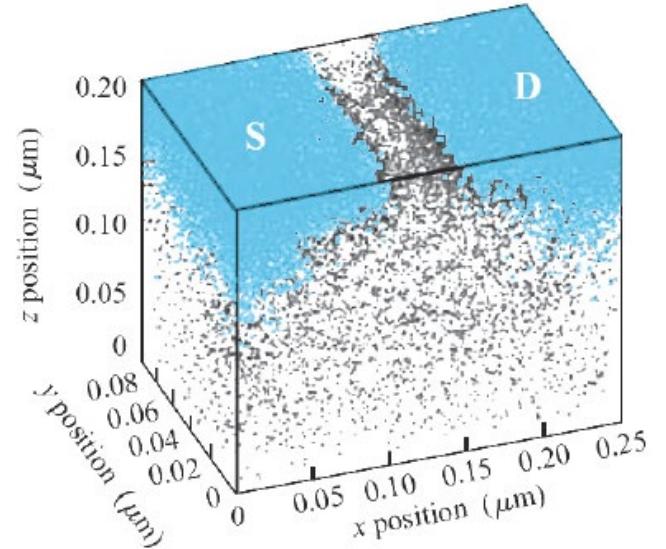
Process Variation

- Threshold Voltage
 - Depends on placement of dopants in channel
 - Standard deviation inversely proportional to channel area

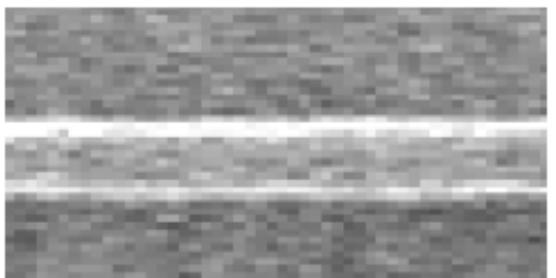
$$\sigma_{V_t} = \frac{t_{\text{ox}}}{\varepsilon_{\text{ox}}} \frac{\sqrt[4]{q^3 \varepsilon_{\text{si}} \phi_b N_a}}{\sqrt{2LW}} = \frac{A_{V_t}}{\sqrt{LW}}$$

- Channel Length
 - Systematic *across-chip linewidth variation* (ACLV)
 - Random line edge roughness (LER)

- Interconnect
 - Etching variations affect w, s, h



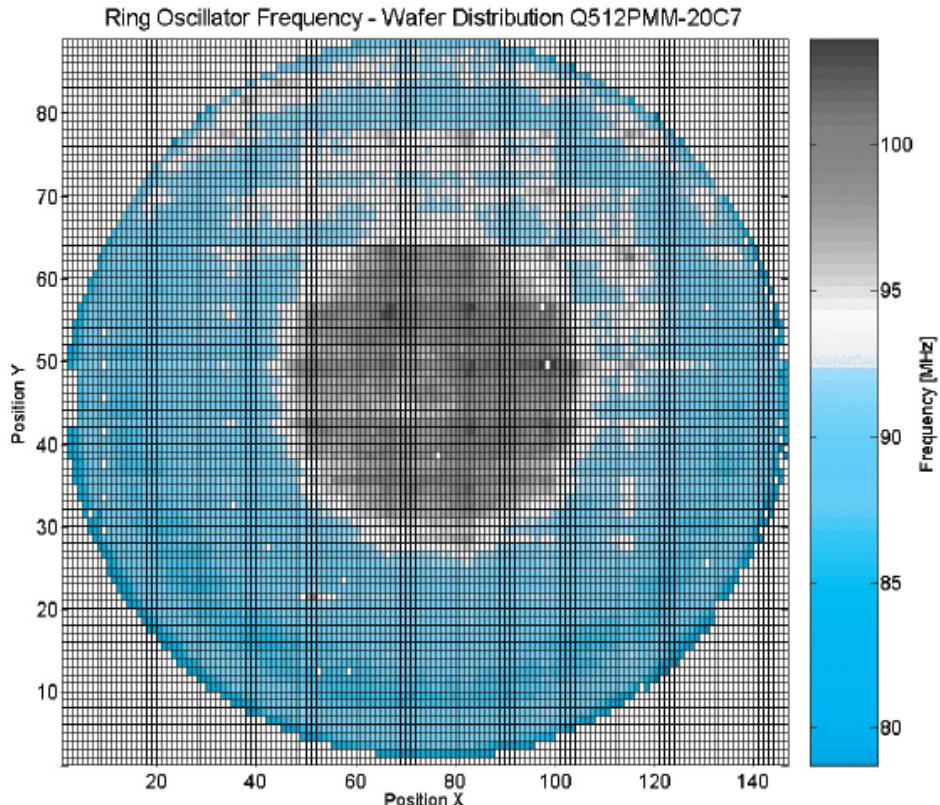
[Bernstein06]



Courtesy Texas Instruments

Spatial Distribution

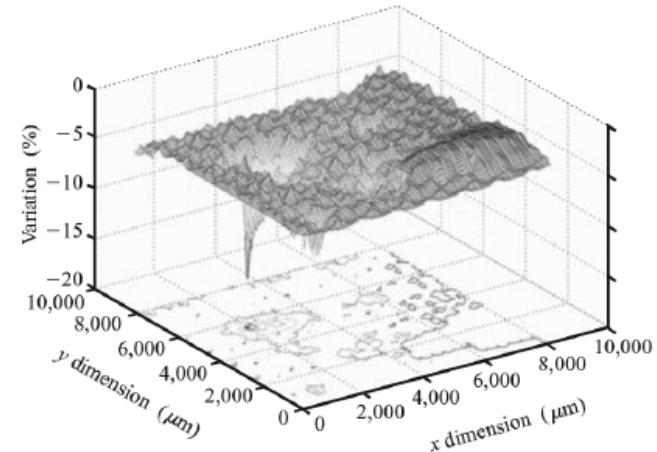
- Variations show spatial correlation
 - *Lot-to-lot (L2L)*
 - *Wafer-to-wafer (W2W)*
 - *Die-to-die (D2D)/inter-die*
 - *Within-die (WID)/intradie*
- Closer transistors match better



Courtesy M. Pelgrom

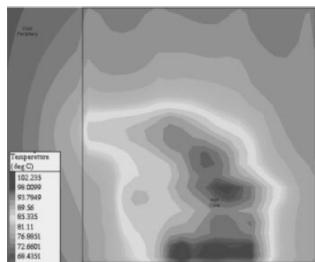
Environmental Variation

- Voltage
 - V_{DD} is usually designed $+/- 10\%$
 - Regulator error
 - On-chip droop from switching activity
- Temperature
 - Ambient temperature ranges
 - On-die temperature elevated by chip power consumption



Courtesy IBM

Standard	Minimum	Maximum
Commercial	0 °C	70 °C
Industrial	-40 °C	85 °C
Military	-55 °C	125 °C



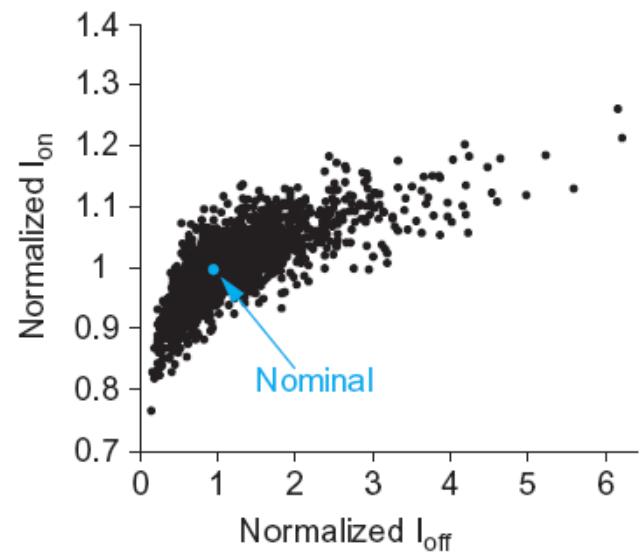
[Harris01b]

Aging

- Transistors change over time as they wear out
 - Hot carriers, trapped charges
 - Negative bias temperature instability
 - Time-dependent dielectric breakdown
- Causes threshold voltage changes

Monte Carlo Simulation

- As process variation increases, the worst-case corners become too pessimistic for practical design
- Monte Carlo: repeated simulations with parameters randomly varied each time
- Look at scatter plot of results to predict yield
- E.g., impact of V_t variation
 - ON-current
 - leakage



EEsim practice

- RC pulsed circuit transient simulation
- I-V curve nFET 90nm
- I-V curve nFET 16nm