

Semiconductor Technology for Integrated Circuit Front Ends

Giovanni Anelli

CERN - European Organization for Nuclear Research
Physics Department
Microelectronics Group
CH-1211 Geneva 23 – Switzerland
giovanni.anelli@cern.ch



Outline

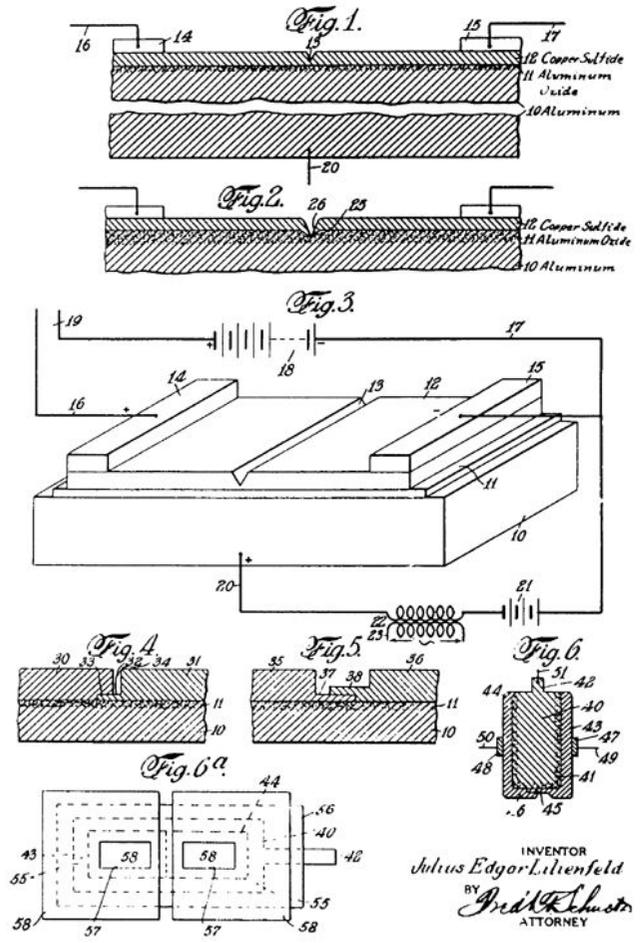
- Operation and characteristics of MOS and Bipolar transistors
- Sub-micron CMOS and BiCMOS technologies
- Feature size scaling
- Radiation effects and reliability
- Mixed-signal circuits



- Operation and characteristics of MOS and Bipolar transistors
 - MOS transistor equations and characteristics
 - MOS transistor small-signal equivalent circuit
 - Bipolar transistor equations and characteristics
 - Bipolar transistor small-signal equivalent circuit
- Sub-micron CMOS and BiCMOS technologies
- Feature size scaling
- Radiation effects and reliability
- Mixed-signal circuits

The MOSFET

March 7, 1933. J. E. LILIENFELD 1,900,018
 DEVICE FOR CONTROLLING ELECTRIC CURRENT
 Filed March 28, 1928 3 Sheets-Sheet 1

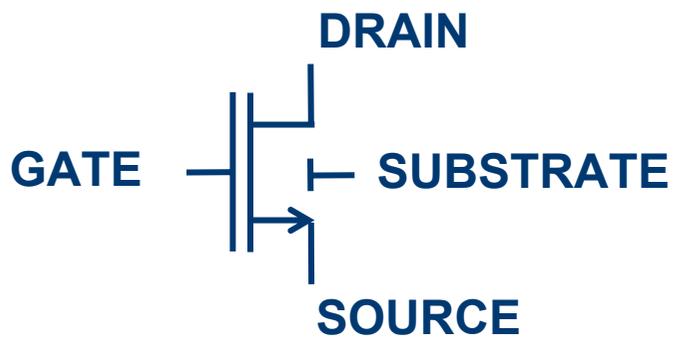
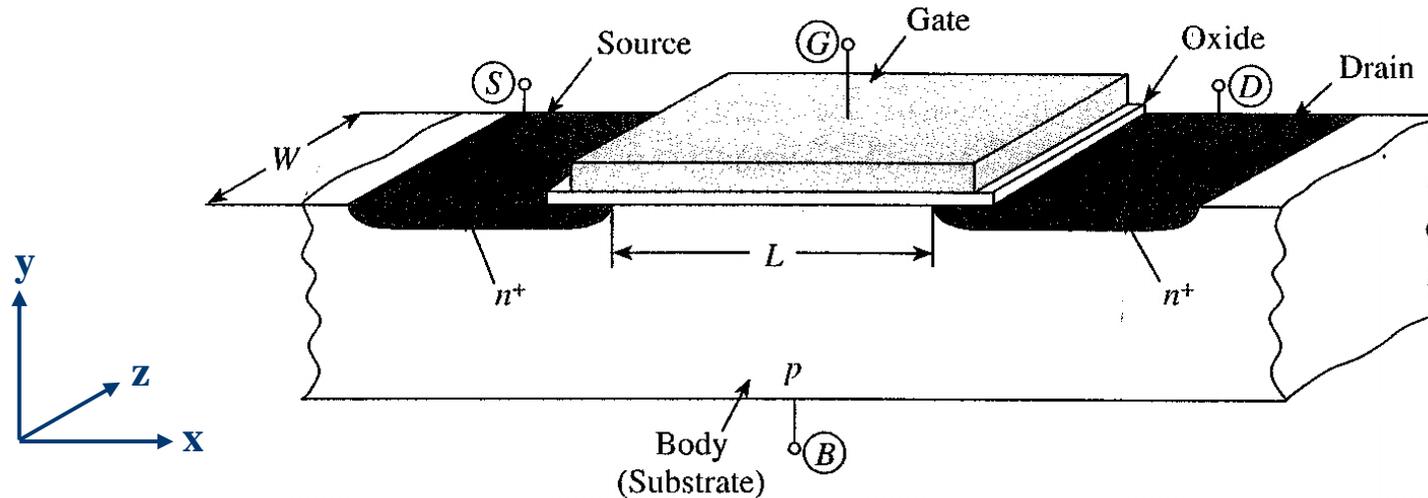


The basic idea behind Field Effect Transistors (FETs) was first patented by J. Lilienfeld in 1930 for the MESFET (Metal Semiconductor FET) and in 1933 for the MOSFET (Metal Oxide Semiconductor FET). But we had to wait until 1960 to have a technology capable to make a working device.



The first working MOSFET was made in 1960 at the Bell Laboratories by D. Kahng and M. M. Atalla.

The (N)-MOS transistor



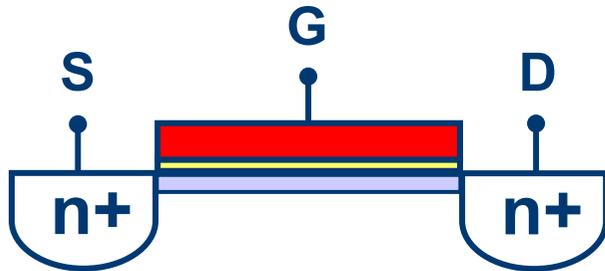
What is a MOS transistor?

Analog circuits: amplifier (V to I)

Digital circuits: switch

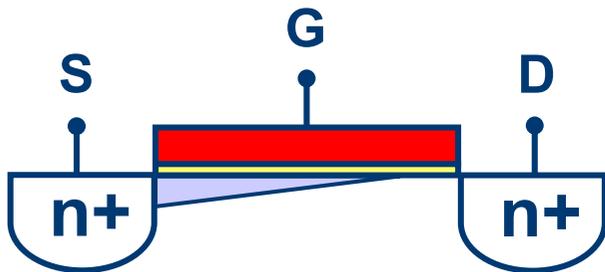
Y. Tsididis, *Operation and Modeling of The MOS Transistor*, 2nd edition, McGraw-Hill, 1999.

Linear and Saturation regions



LINEAR REGION (Low V_{DS}):

Electrons are attracted to the $\text{SiO}_2 - \text{Si}$ interface. A conductive channel is created between source and drain. We have a Voltage Controlled Resistor (VCR).

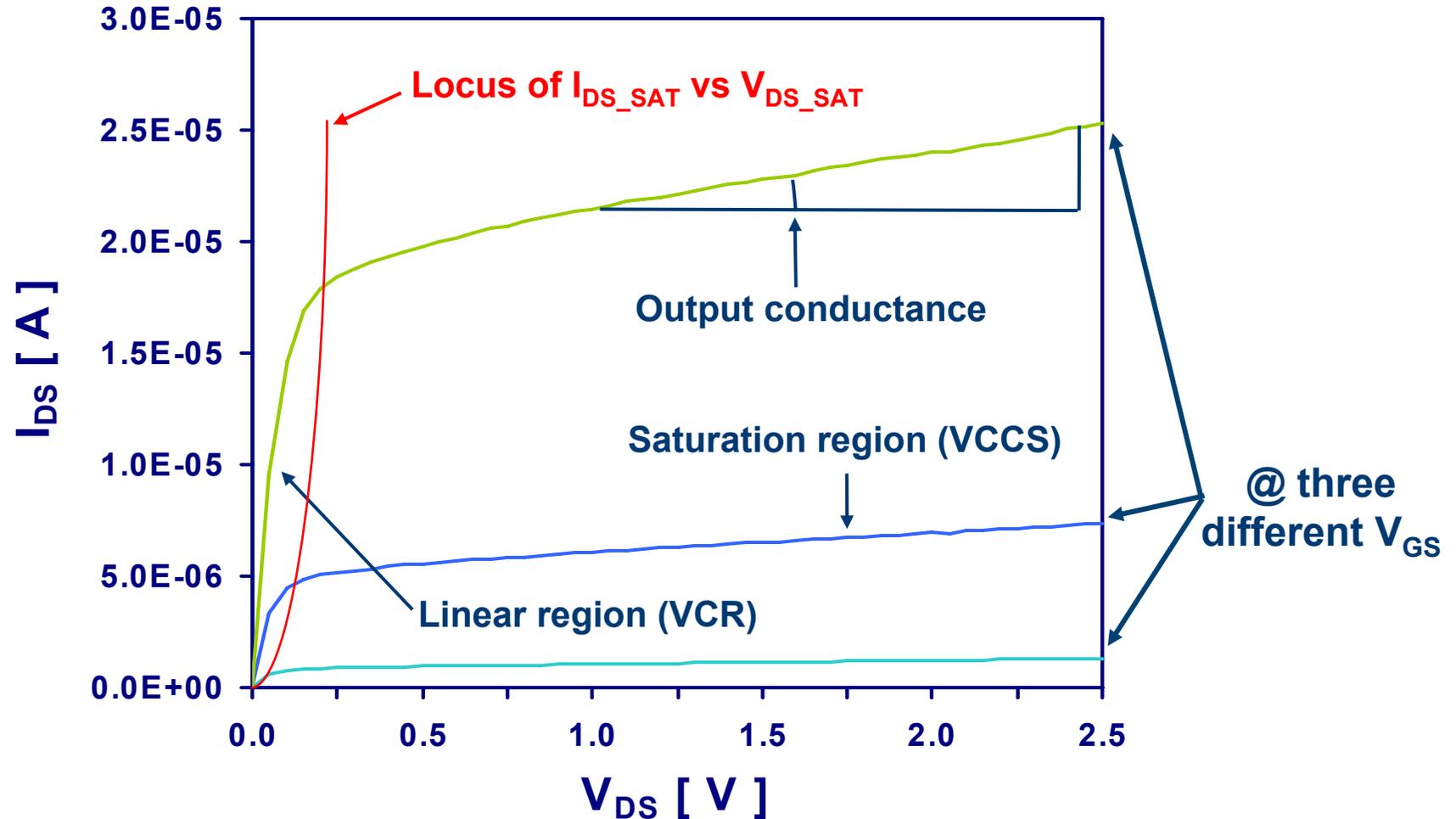


SATURATION REGION (High V_{DS}):

When the drain voltage is high enough the electrons near the drain are insufficiently attracted by the gate, and the channel is pinched off. We have a Voltage Controlled Current Source (VCCS).

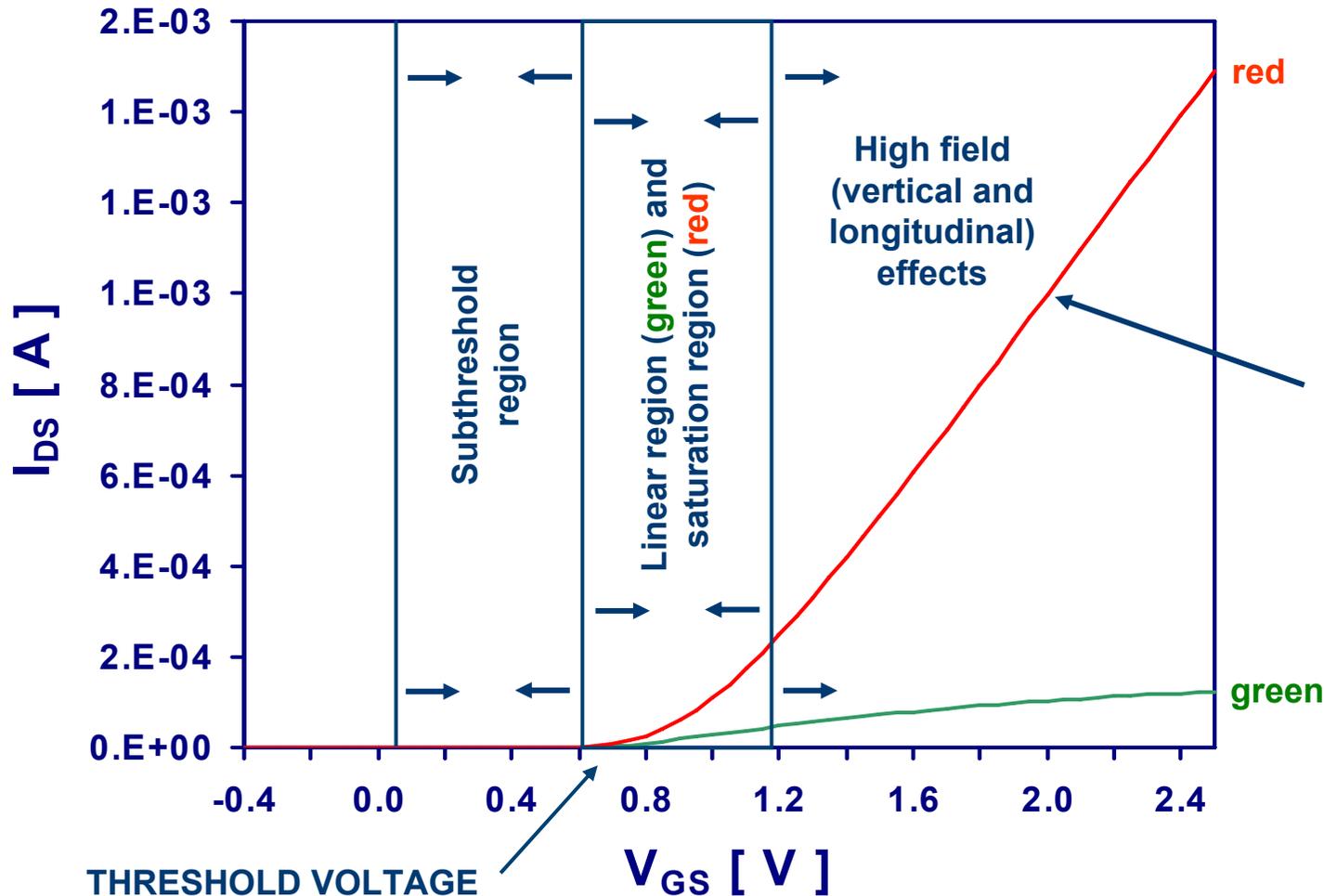
Drain current vs Drain voltage

This is a real device measurement !



Drain current vs Gate voltage

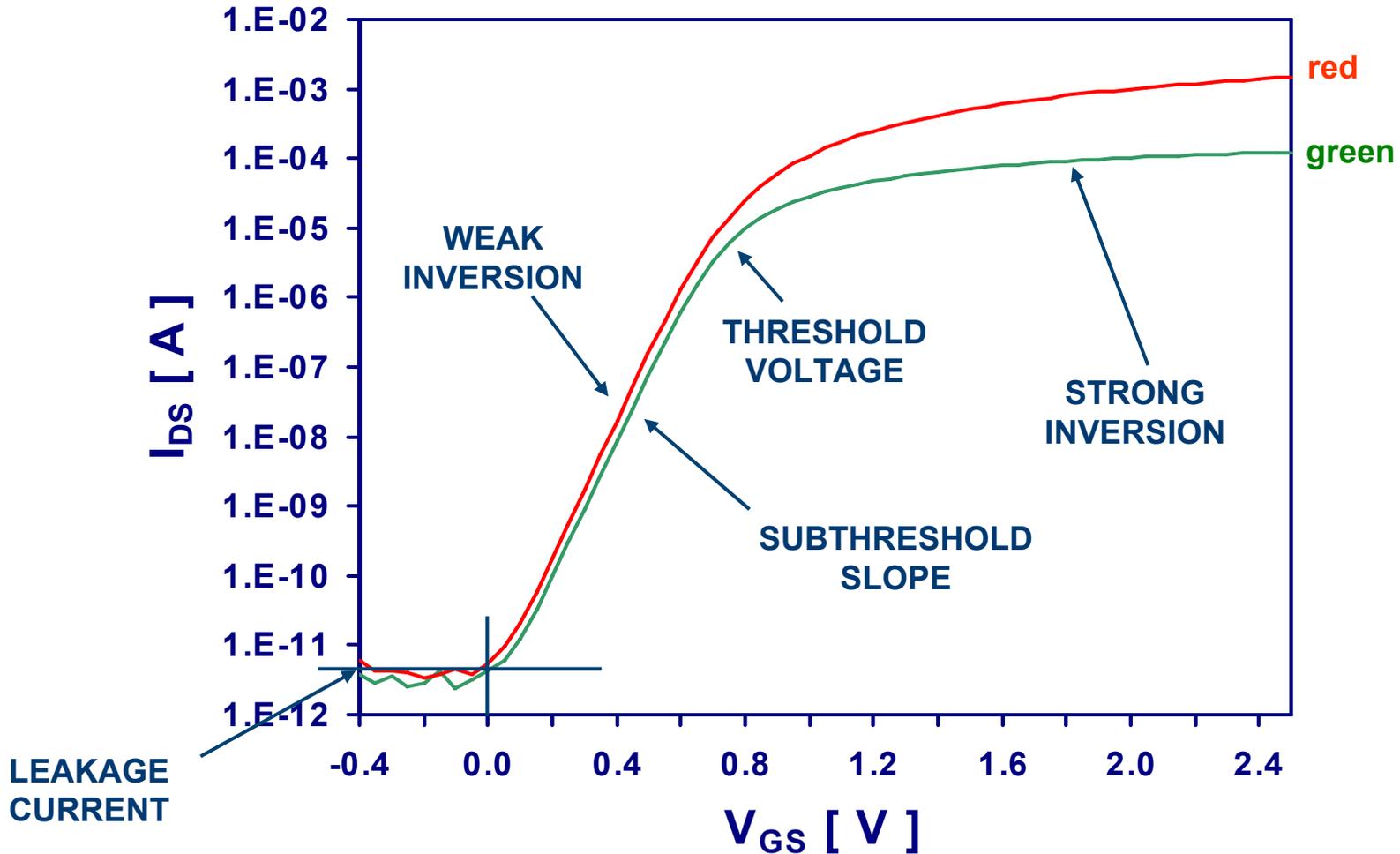
This is also a measurement, same device.



The SLOPE of this plot is called **Transconductance**, and is a very important parameter for analog design (is the “gain” of the V-to-I amplifier).

Log(I_{DS}) vs V_{GS}

Exactly same measurement as before, but semi log scale



Equations: strong inversion

LINEAR REGION:

$$V_{DS} < \frac{V_{GS} - V_T}{n} = V_{DS_SAT} \quad I_{DS} = \beta \left(V_{GS} - V_T - \frac{nV_{DS}}{2} \right) V_{DS}$$

$$\text{Transconductance: } g_m = \frac{\partial I_{DS}}{\partial V_{GS}} = \beta V_{DS}$$

SATURATION REGION:

$$V_{DS} > \frac{V_{GS} - V_T}{n} = V_{DS_SAT} \quad I_{DS} = \frac{\beta}{2n} (V_{GS} - V_T)^2$$

$$\text{Transconductance: } g_m = \frac{\partial I_{DS}}{\partial V_{GS}} = \frac{\beta}{n} (V_{GS} - V_T) = \sqrt{2 \frac{\beta}{n} I_{DS}}$$

$$n = \frac{g_m + g_{mb}}{g_m} \approx 1.x$$

$$g_{mb} = \frac{\partial I_{DS}}{\partial V_{BS}}$$

$$\beta = \mu C_{ox} \frac{W}{L}$$

$$C_{ox} = \frac{\epsilon_{SiO_2}}{t_{ox}}$$

Equations: weak inversion

$$I_{DS} = I_{D0} \frac{W}{L} e^{\frac{V_{GS}}{n\phi_t}} (1 - e^{-\frac{V_{DS}}{n\phi_t}})$$

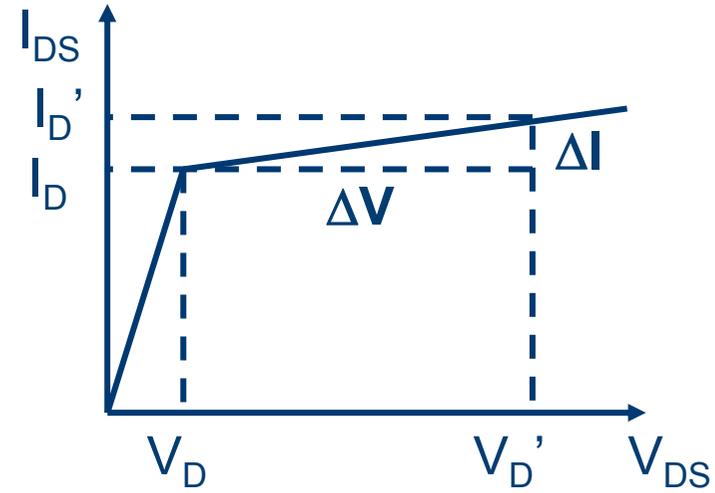
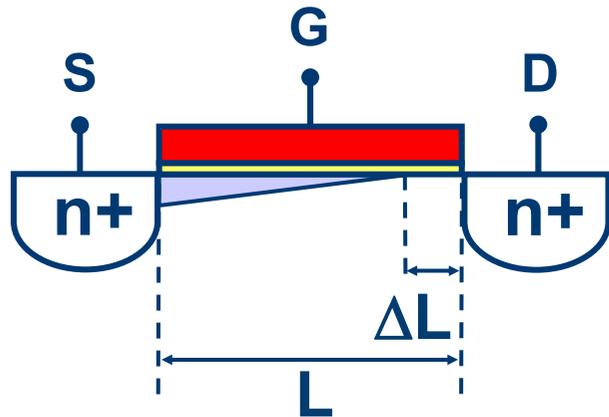
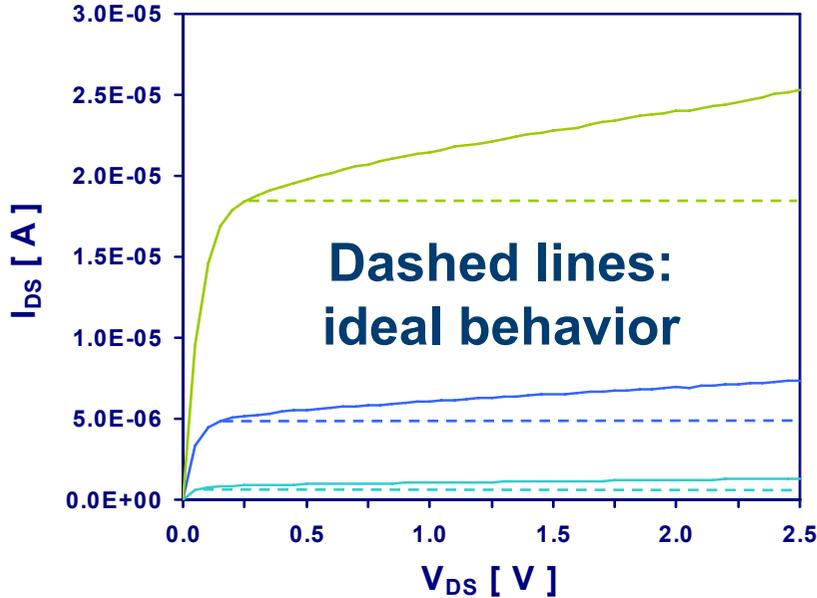
If $V_{DS} > 4n\phi_t$ **then the drain current does not depend on**
 V_{DS} **any longer (saturation)**

$$I_{DS} = I_{D0} \frac{W}{L} e^{\frac{V_{GS}}{n\phi_t}} \quad g_m = \frac{\partial I_{DS}}{\partial V_{GS}} = \frac{I_{DS}}{n\phi_t}$$

Almost like a bipolar transistor (see later)!

$$\phi_t = \frac{kT}{q} \approx \mathbf{25 \text{ mV @ 300 K}}$$

Output conductance



$$G_{\text{out}} = \frac{\Delta I}{\Delta V} = \frac{I_D}{\Delta V} \cdot \frac{\Delta L}{L - \Delta L}$$

The non-zero output conductance is related to a phenomenon called *channel length modulation*

Output conductance / resistance

Drain-to-source current in saturation

$$I_{DS} = \frac{\beta}{2n} (V_{GS} - V_T)^2 (1 + \lambda V_{DS}) = I_{DS_SAT} \cdot (1 + \lambda V_{DS})$$

Output conductance

$$g_{out} = g_{ds} = \frac{\partial I_{DS}}{\partial V_{DS}} = \lambda \cdot I_{DS_SAT}$$

Output resistance

$$r_0 = \frac{1}{g_{ds}} = \frac{1}{\lambda \cdot I_{DS_SAT}} = \frac{V_E \cdot L}{I_{DS_SAT}}$$

$$\lambda = \frac{1}{V_{DS}} \cdot \frac{\Delta L}{L - \Delta L} \quad \text{where } \Delta L = f(V_{DS}, N_{Doping})$$

Remember: λ is proportional to $1/L$

Equations: addendum

Bulk effect

$$\Delta V_T = \gamma \cdot \left(\sqrt{V_{sb} + \phi_{Si}} - \sqrt{\phi_{Si}} \right)$$

$$\gamma = \frac{\sqrt{2q\epsilon_{Si}N_a}}{C_{ox}}$$

**Source parasitic
resistance**

$$g'_m = \frac{g_m}{1 + g_m R_s}$$

**Vertical electric
field effect**

$$\mu = \frac{\mu_0}{1 + \theta (V_{GS} - V_T)}$$

**Maximum
frequency**

$$f_{max} = \frac{1}{2\pi} \frac{g_m}{C_{gs}} = \frac{1}{2\pi} \frac{\mu}{nL^2} (V_{GS} - V_T) \quad \text{in s.i.}$$

Equations: velocity saturation

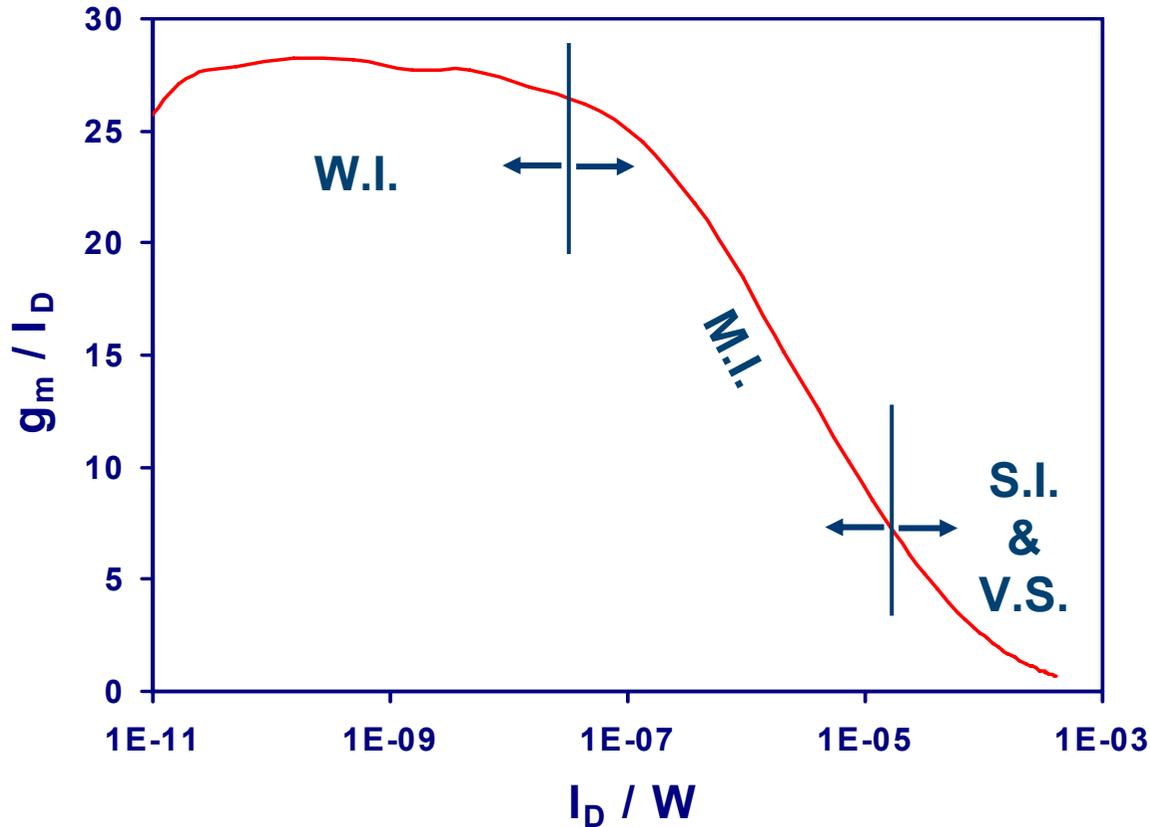
For low values of the longitudinal electric field, the velocity of the carriers increases proportionally to the electric field (and the proportionality constant is the mobility). For high values of the longitudinal electric field (3 V/ μm for electrons and 10 V/ μm for holes) the velocity of the carriers saturates.

$$I_{\text{DS_v.s.}} = WC_{\text{ox}} v_{\text{sat}} (V_{\text{GS}} - V_{\text{T}}) \quad \text{with} \quad v_{\text{sat}} = 10^7 \frac{\text{cm}}{\text{s}}$$

$$g_{\text{m_v.s.}} = WC_{\text{ox}} v_{\text{sat}}$$

$$f_{\text{max_v.s.}} = \frac{1}{2\pi} \frac{g_{\text{m}}}{C_{\text{gs}}} = \frac{1}{2\pi} \frac{v_{\text{sat}}}{L}$$

g_m / I_D vs $\log(I_D / W)$



Weak Inversion (W.I.)

$$g_m = \frac{I_{DS}}{n\phi_t} \rightarrow \frac{g_m}{I_D} = \frac{1}{n\phi_t}$$

Strong Inversion (S.I.)

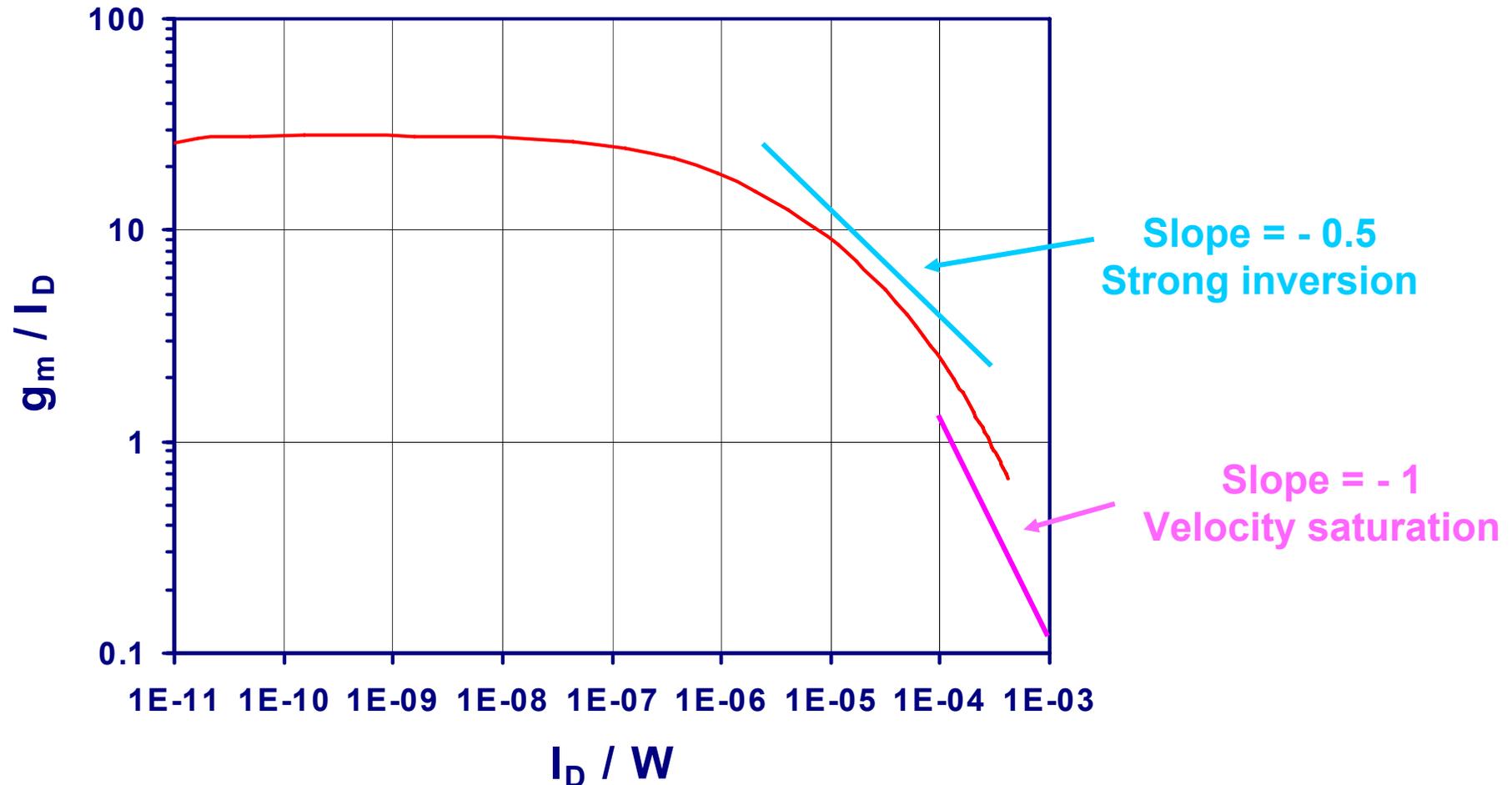
$$g_m = \sqrt{2 \frac{\beta}{n} I_{DS}} \rightarrow \frac{g_m}{I_D} = \sqrt{2 \frac{\beta}{n} \frac{1}{I_{DS}}}$$

Velocity Saturation (V.S.)

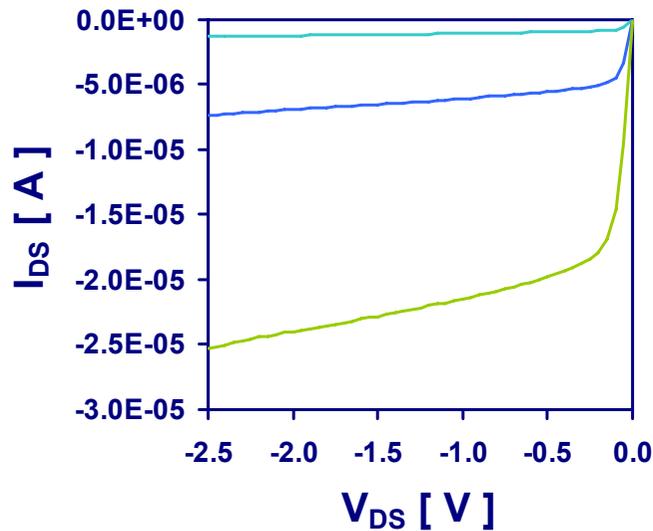
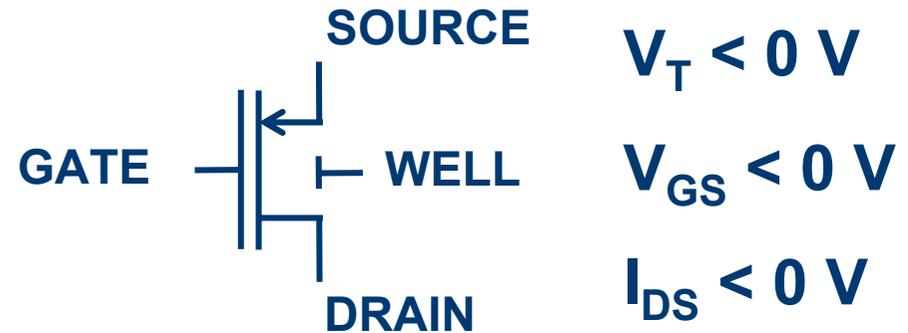
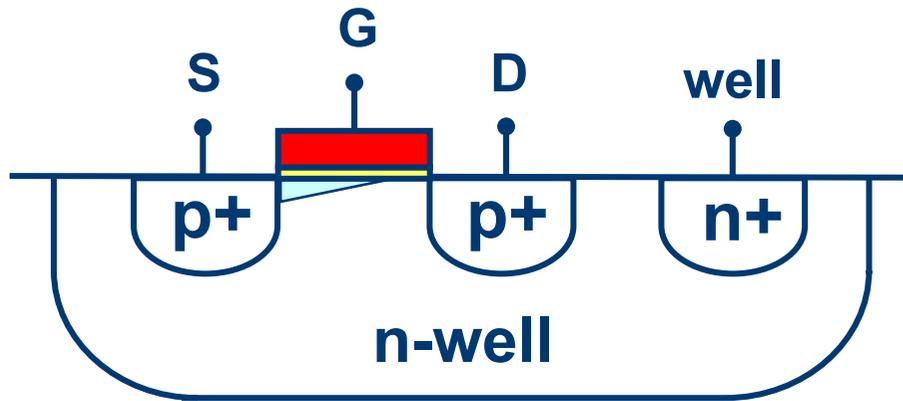
$$g_m = WC_{ox} v_{sat} \rightarrow \frac{g_m}{I_{DS}} = \frac{WC_{ox} v_{sat}}{I_{DS}}$$

Moderate Inversion (M.I.): No Simple Equations

Log(g_m / I_D) vs log(I_D / W)



The poor PMOS transistor

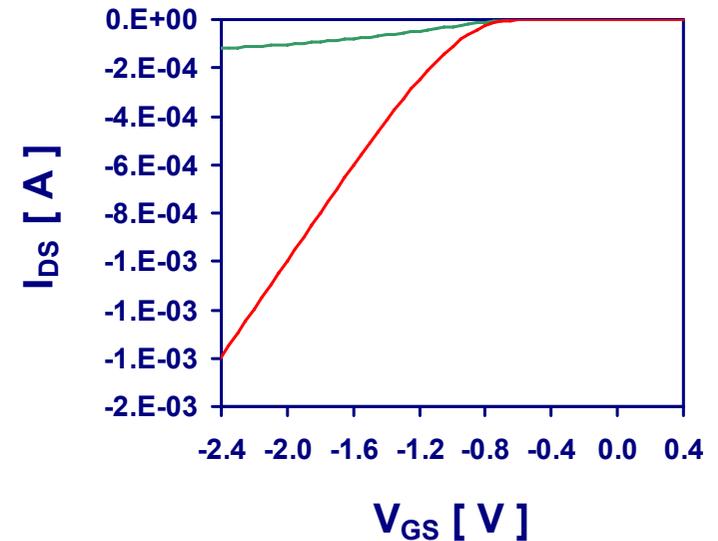


$$I_{DS} = -\frac{\beta}{2n} (V_{GS} - V_T)^2$$

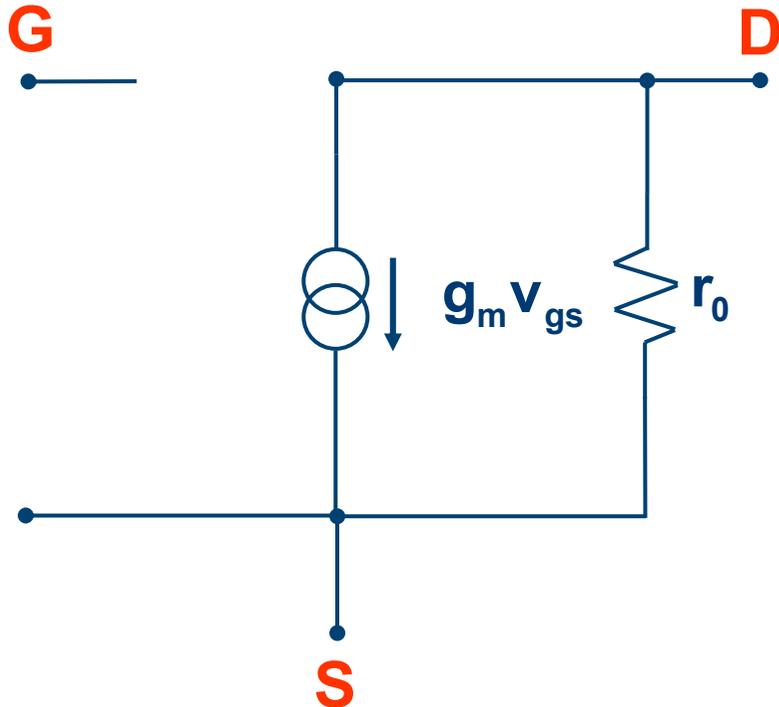
$$g_m = \frac{\partial I_{DS}}{\partial V_{GS}} =$$

$$= -\frac{\beta}{n} (V_{GS} - V_T) =$$

$$= \sqrt{-2 \frac{\beta}{n} I_{DS}}$$



Small-signal equivalent circuit



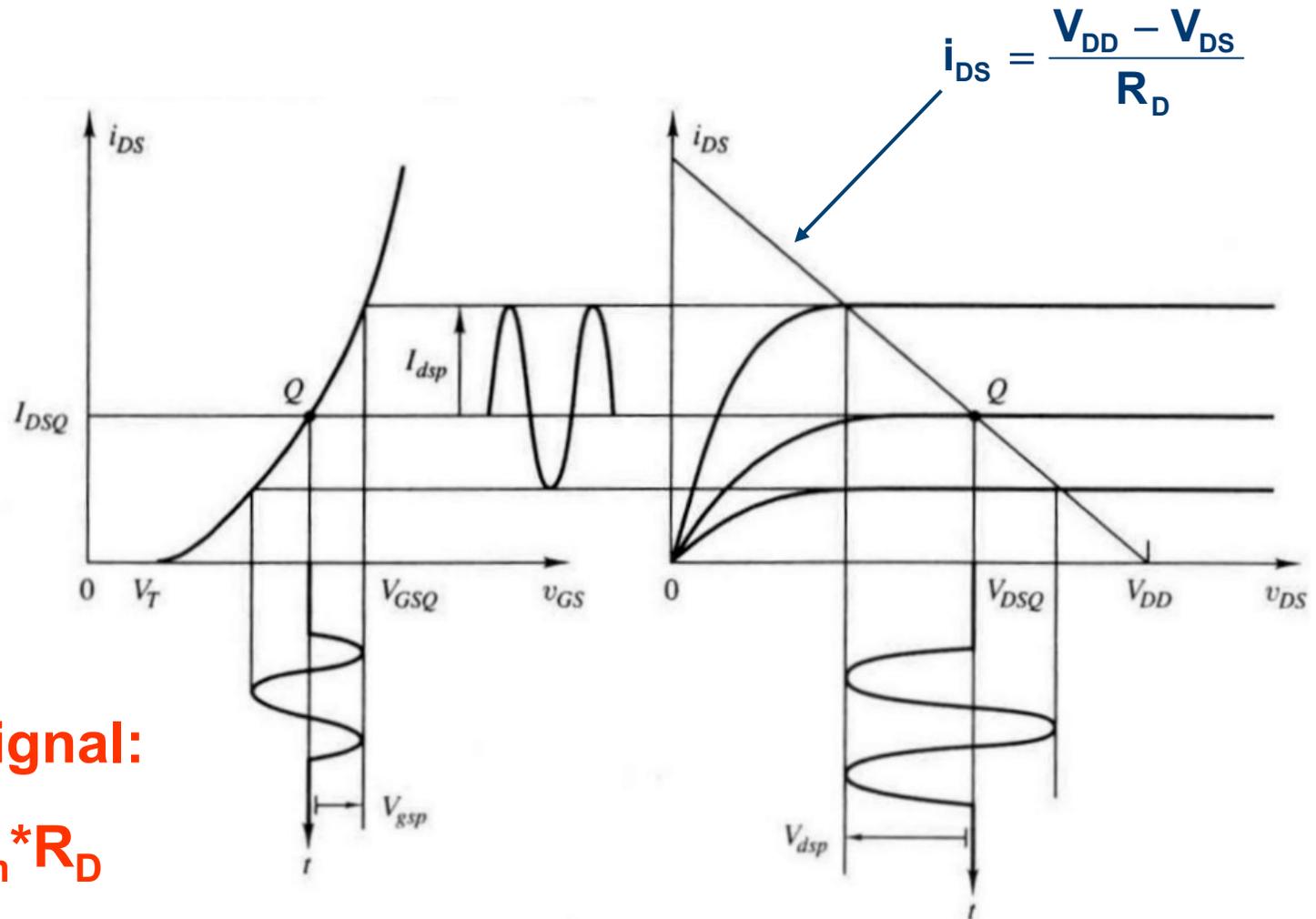
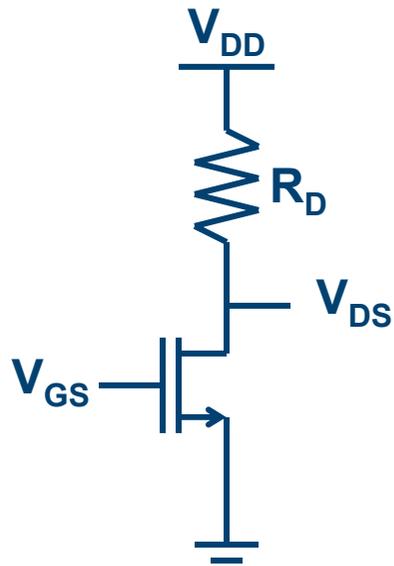
**Valid only in
saturation and at
very low
frequencies**

No bulk effect

$$I_{DS} = \frac{\beta}{2n} (V_{GS} - V_T)^2 (1 + \lambda V_{DS}) \longrightarrow \text{This equation fixes the bias point}$$

$$i_{ds} = g_m \cdot v_{gs} \longrightarrow \text{This equation defines the small signal behavior}$$

A very simple amplifier

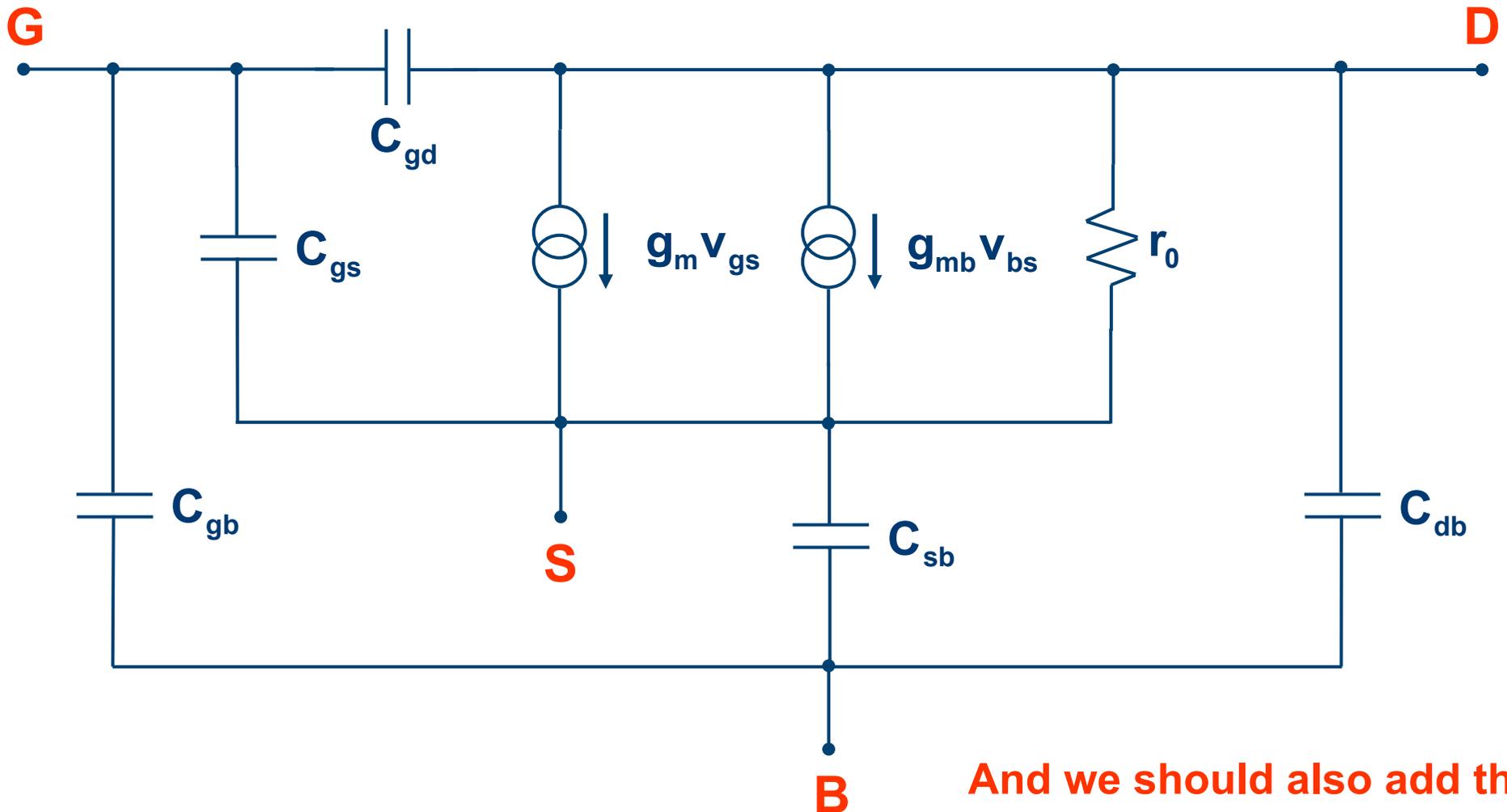


For a small signal:

$$v_{ds} = -v_{gs} * g_m * R_D$$

K. R. Laker and W. M. C. Sansen, *Design of Analog Integrated Circuits and Systems*, McGraw-Hill, 1994.

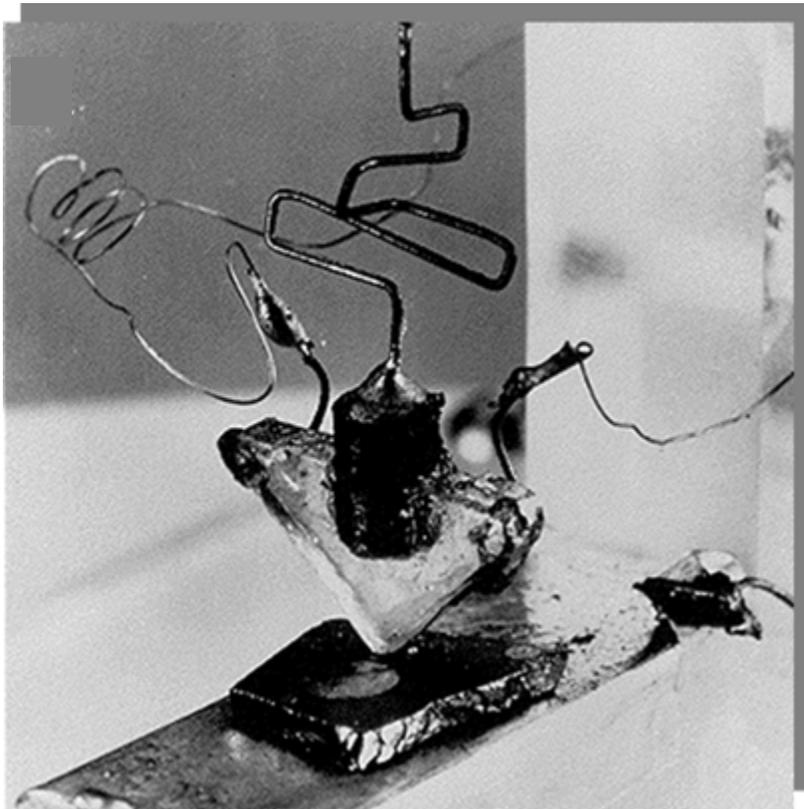
A better equivalent circuit



And we should also add the series resistances...

The First Transistor

First Point Contact Transistor



**Bardeen and Brattain, 1947
Bell Laboratories**



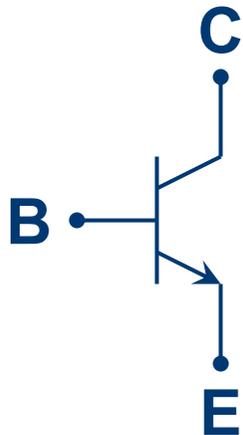
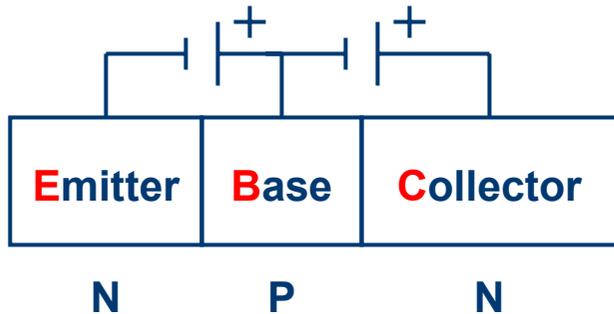


The Nobel Prize in Physics 1956

"for their researches on semiconductors and their discovery of the transistor effect"

		
William Bradford Shockley	John Bardeen	Walter Houser Brattain
🏆 1/3 of the prize	🏆 1/3 of the prize	🏆 1/3 of the prize
USA	USA	USA
Semiconductor Laboratory of Beckman Instruments, Inc. Mountain View, CA, USA	University of Illinois Urbana, IL, USA	Bell Telephone Laboratories Murray Hill, NJ, USA
b. 1910 (in London, United Kingdom) d. 1989	b. 1908 d. 1991	b. 1902 d. 1987

The Bipolar Junction Transistor (BJT)

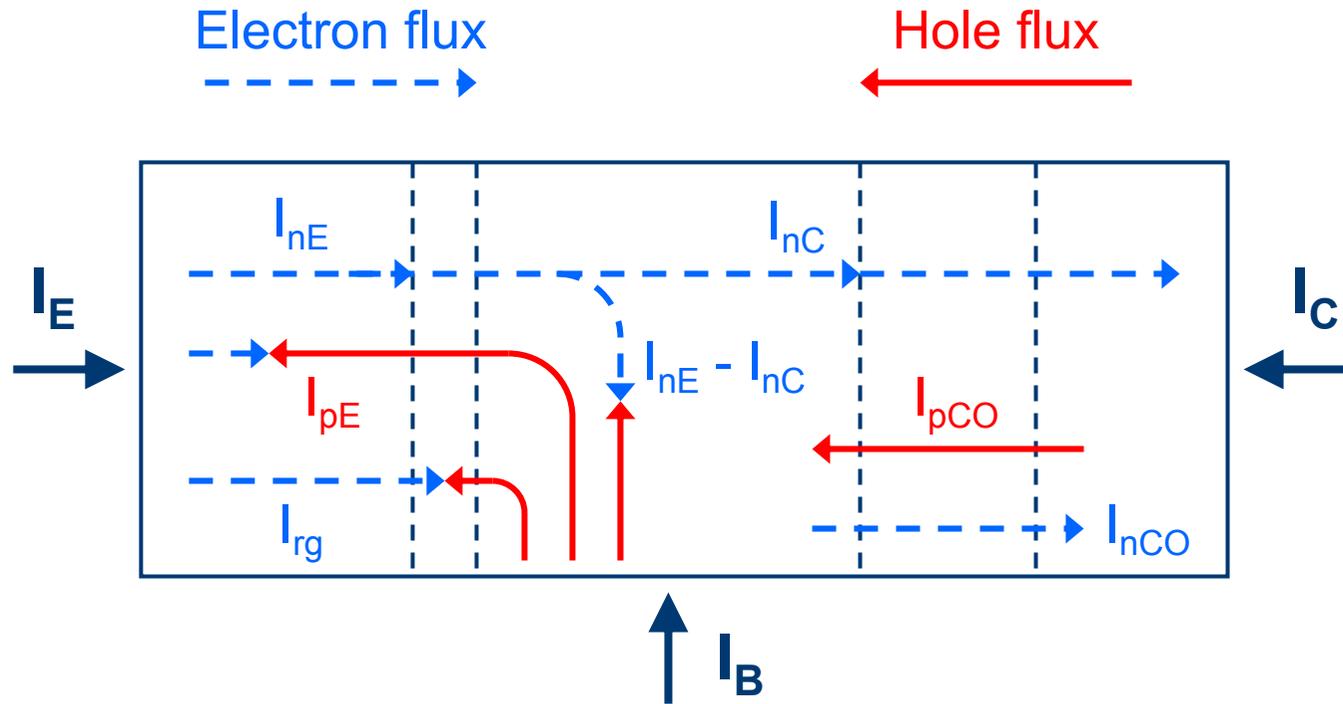


In normal operation condition (also called active region) the Emitter – Base diode is forward biased and the Base – Collector diode is reverse biased. Electrons are injected from the Emitter into the Base. They diffuse then towards the Collector, where they are collected.

At the same time holes are injected from the base into the emitter, but in a much smaller number since the base is normally much less doped than the emitter.

A fraction of the electrons injected into the base will not reach the collector because it will recombine with holes in the base. To make this a negligible effect the electrons lifetime must be long so that their diffusion length is much larger than the base width.

Currents in a bipolar transistor



$$-I_E = I_{pE} + I_{nE} + I_{rg}$$

$$I_B = I_{pE} + I_{rg} + (I_{nE} - I_{nC}) - I_{CO}$$

$$I_C = I_{nC} + I_{CO}$$

$$I_E + I_B + I_C = 0$$

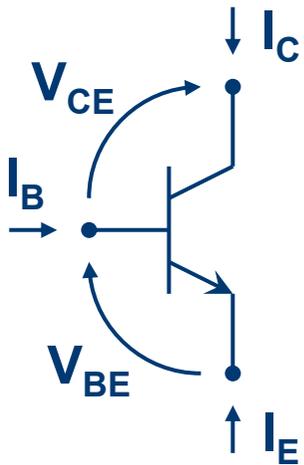
I_{nE} : electron current injected into the base

I_{nC} : electron current collected by the collector

I_{rg} : emitter space-charge-layer recombination current

I_{nCO} and I_{pCO} : collector reverse saturation currents ($I_{nCO} + I_{pCO} = I_{CO}$)

Currents in a bipolar transistor



Common base current gain

$$\alpha = \frac{I_C - I_{CO}}{-I_E} = \frac{I_{nC}}{I_{nE} + I_{pE} + I_{rg}} = \gamma \cdot \beta_T$$

Emitter injection efficiency

$$\gamma = \frac{I_{nE}}{I_{nE} + I_{pE} + I_{rg}}$$

Base transport factor

$$\beta_T = \frac{I_{nC}}{I_{nE}}$$

γ , β_T and therefore α are normally very close to 1 (but not 1)

$$I_C = \frac{\alpha}{1 - \alpha} \cdot I_B + \frac{I_{CO}}{1 - \alpha} = h_{FE} \cdot I_B + I_{CEO}$$

h_{FE} is the common emitter current gain (sometimes also called β).

If α is 0.99 h_{FE} is 99!

Current vs voltage equations

$$I_{nE} = qA_{EB} D_n \frac{n_i^2}{N_{aB} W_B} e^{\frac{V_{BE}}{\phi_t}}$$

$$I_{pE} = qA_{EB} D_p \frac{n_i^2}{N_{dE} W_E} e^{\frac{V_{BE}}{\phi_t}}$$

$$\gamma = \frac{1}{1 + \frac{D_p N_{aB} W_B}{D_n N_{dE} W_E}}$$

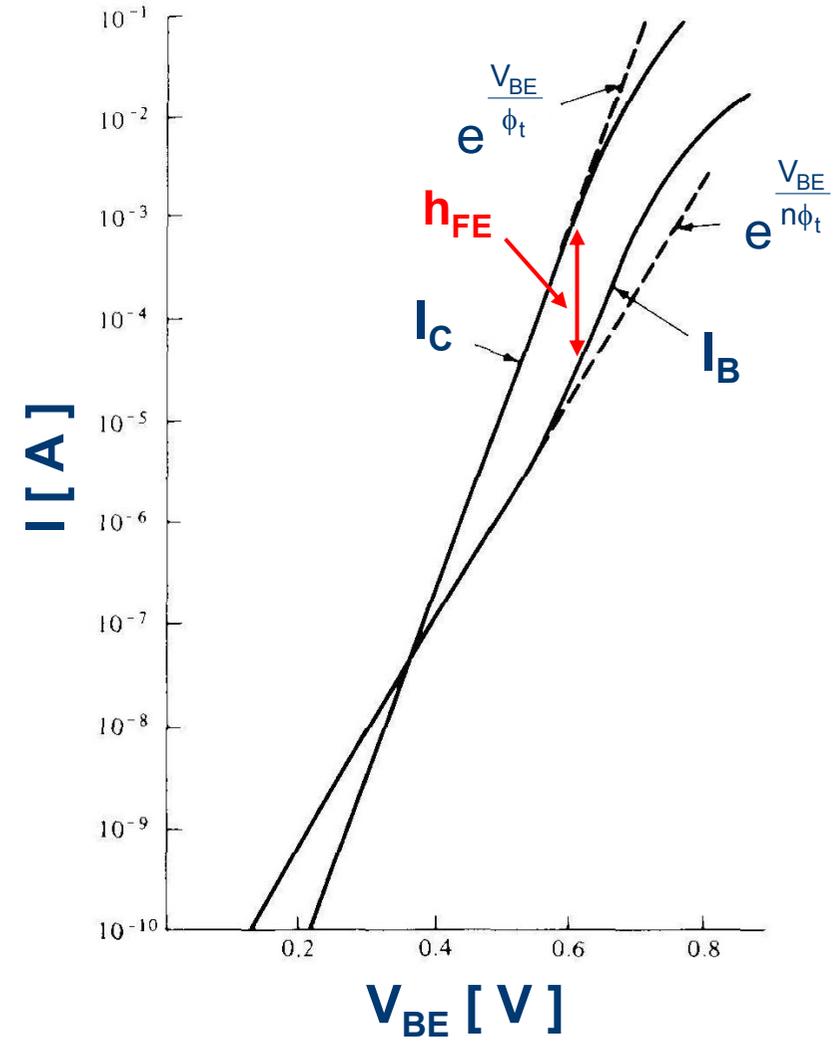
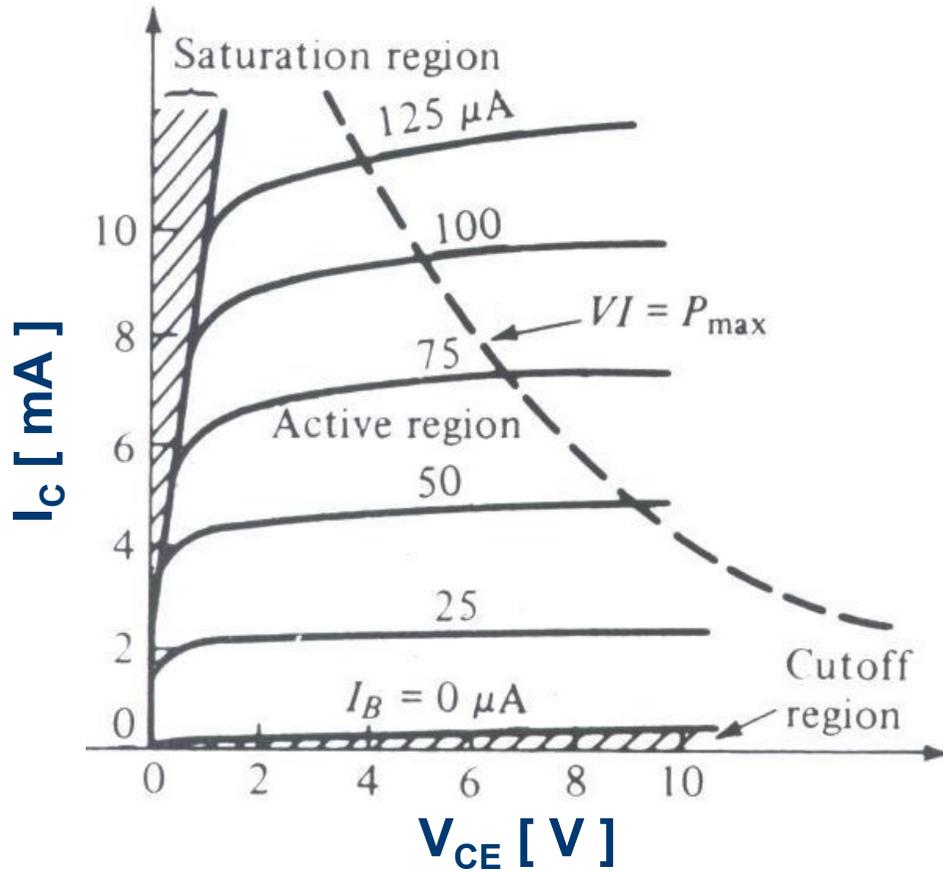
A_{EB} : E-B junction area
 D : diffusion constant ($=\mu\phi_t$)
 n_i : intrinsic carrier conc.
 N : dopant concentration
 W : E or B width
 ϕ_t : thermal voltage

$$I_{nB} = I_{nE} - I_{nC} = \frac{qA_{EB} W_B n_{po}}{2\tau_n} e^{\frac{V_{BE}}{\phi_t}} \rightarrow \beta_T = 1 - \frac{W_B^2}{2L_n^2} \quad L_n = \sqrt{D_n \tau_n}$$

n_{po} : equilibrium electron density in the base ($=n_i^2/N_{aB}$)
 τ_n : electron lifetime in the base
 L_n : diffusion length
 n : number between 1 and 2

$$I_{rg} \propto e^{\frac{V_{BE}}{2\phi_t}} \rightarrow I_B \propto e^{\frac{V_{BE}}{n\phi_t}}$$

Bipolar transistor characteristics



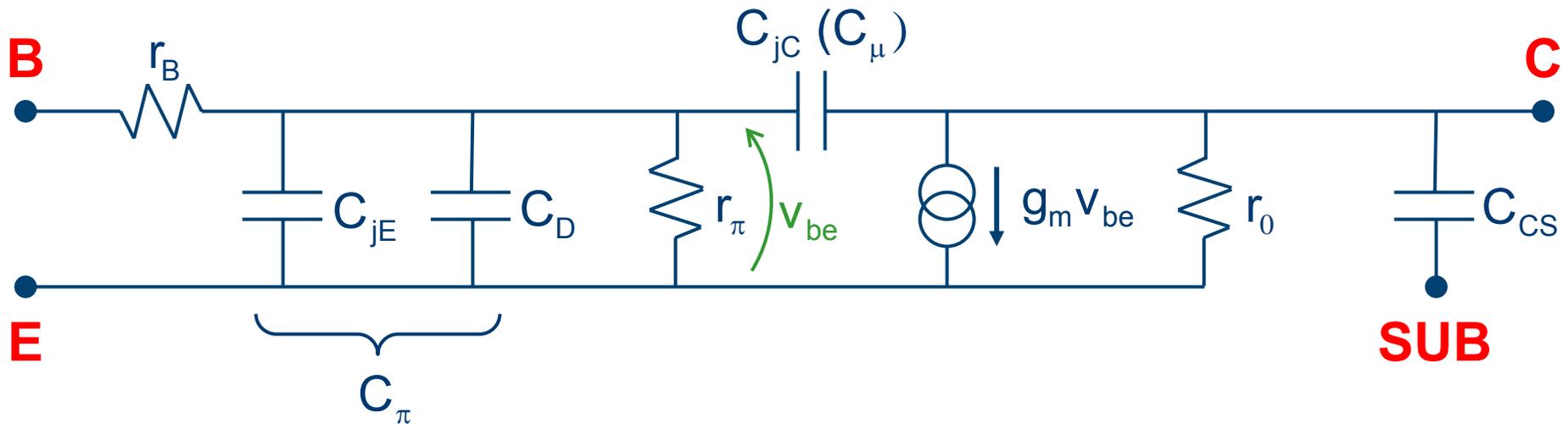
E. S. Yang, *Microelectronic Devices*, McGraw-Hill, 1988.

A “few” important equations

Collector current	$I_C \approx I_{nE} = I_S e^{\frac{V_{BE}}{\phi_t}}$	$I_S = qA_{EB} D_n \frac{n_i^2}{N_{aB} W_B}$
Transconductance	$g_m = \frac{\partial I_C}{\partial V_{BE}} = \frac{I_C}{\phi_t}$	$\frac{g_m}{I_C} = \frac{1}{\phi_t} = 40 @ 300 \text{ K}$
Input resistance	$r_\pi = \frac{h_{FE}}{g_m}$	
Output resistance	$r_o = \frac{V_E}{I_C}$	$V_E = \frac{qA_{BC} N_{aB} W_B}{C_{jC}}$
Max frequency	$\omega_T = \frac{2D_n}{W_B^2}$	
Diffusion capacitance	$C_D = \frac{\partial Q_B}{\partial V_{BE}} = g_m \frac{W_B^2}{2D_n} = \frac{g_m}{\omega_T}$	

Small-signal equivalent circuit

The hybrid- π model



r_B is the base resistance, which is layout dependent

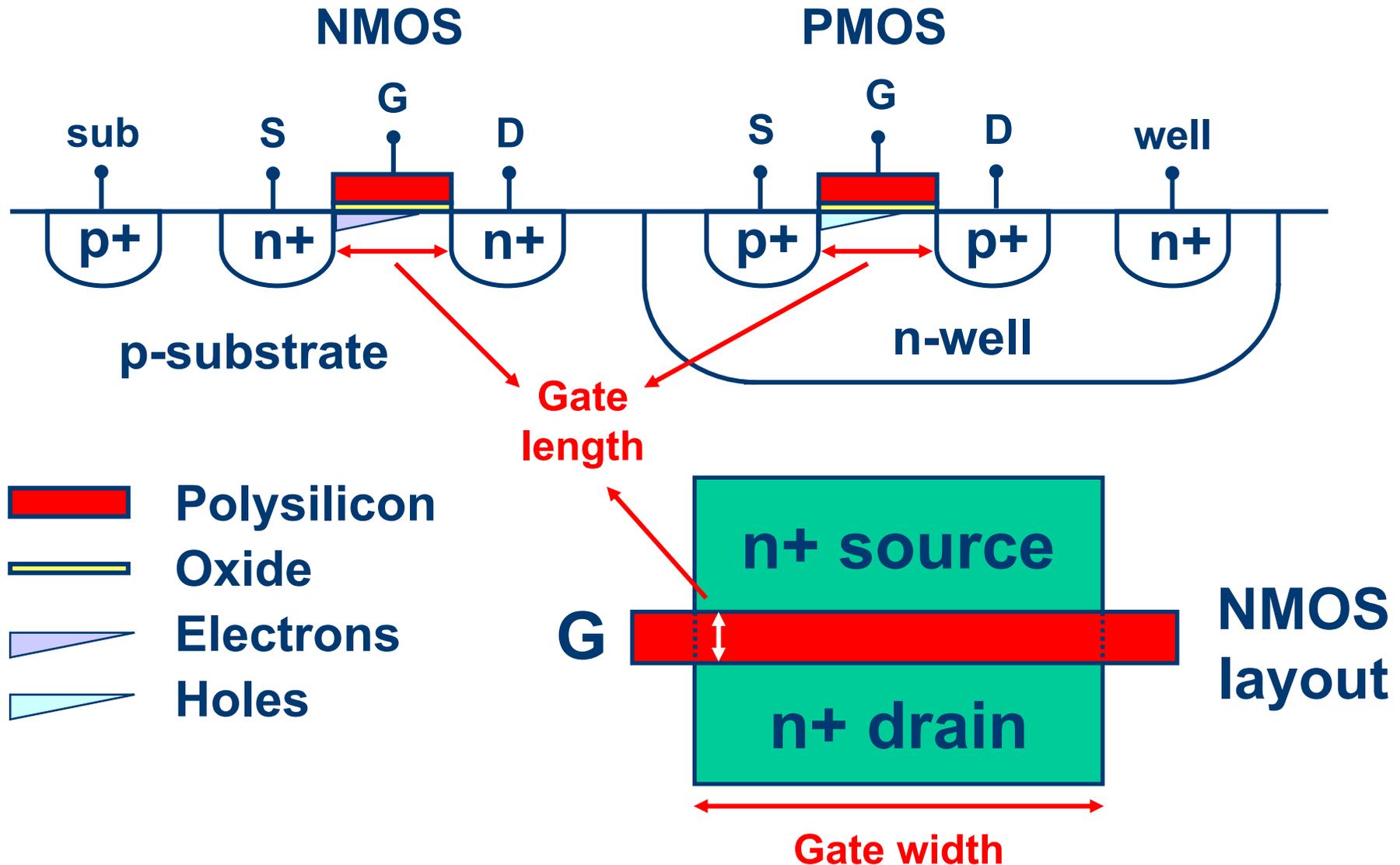
Comparison MOS - Bipolar

	MOS Transistor	Bipolar Transistor
I_{IN}	0	I_C / h_{FE}
R_{IN}	∞	$r_\pi + r_B$
$V_{DSsat} / V_{CEactive}$	$(V_{GS} - V_T) / n$	few ϕ_t
g_m / I	w.i. - $1 / n\phi_t$	$1 / \phi_t$
	s.i. - $2 / (V_{GS} - V_T) = \sqrt{2\beta / (nI_{DS})}$	
	v.s. - $WC_{OX}v_{sat} / I_{DS} = 1 / (V_{GS} - V_T)$	
Design variables	$W / L, V_{GS} - V_T$	ϕ_t
I range	~ 2 decades	> 5 decades
Max f_T @ low I	g_m / C_{gs}	g_m / C_D
Max f_T @ high I	v_{sat} / L_{eff}	v_{sat} / W_B
Thermal Noise	$4kT (0.7/g_m + R_G)$	$4kT (0.5/g_m + r_B)$
1/f noise	Higher for MOS than for bipolar transistors	

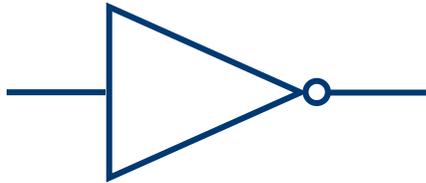


- Operation and characteristics of MOS and Bipolar transistors
- Sub-micron CMOS and BiCMOS technologies
 - CMOS technology
 - Integrated Bipolar Transistors
 - Technology price comparison
- Feature size scaling
- Radiation effects and reliability
- Mixed-signal circuits

Complementary MOS technology



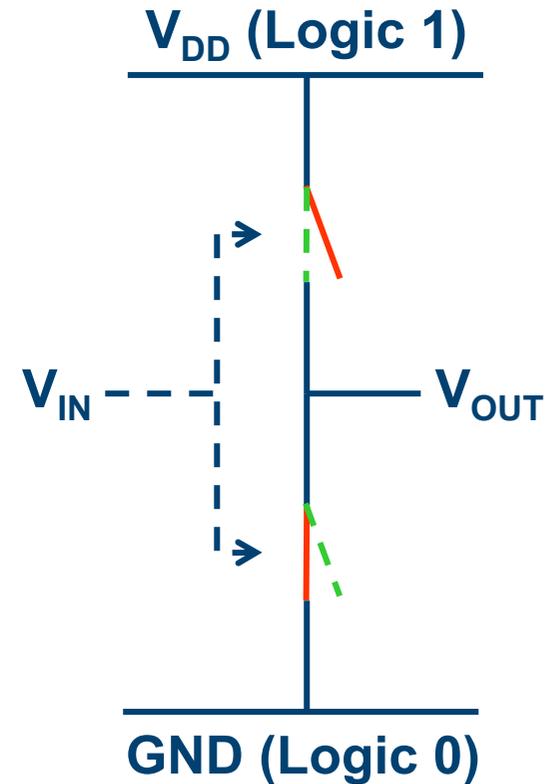
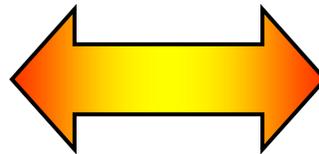
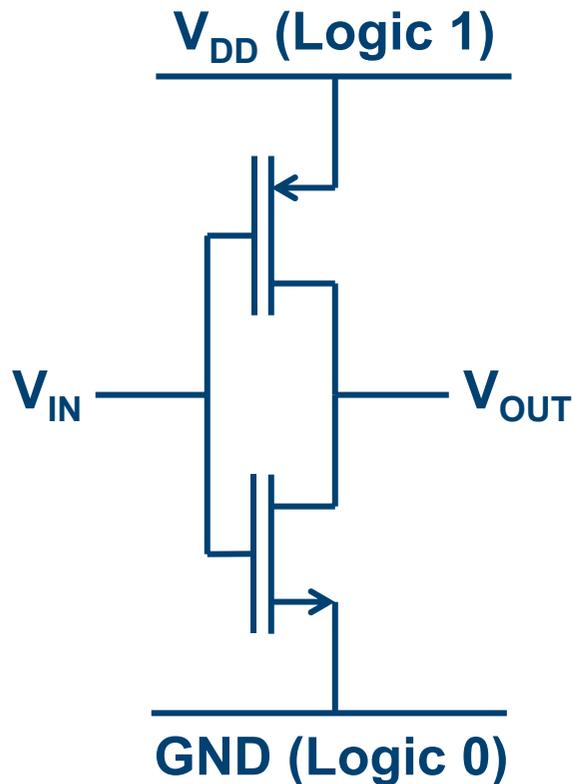
Why is CMOS logic so attractive?



THE INVERTER

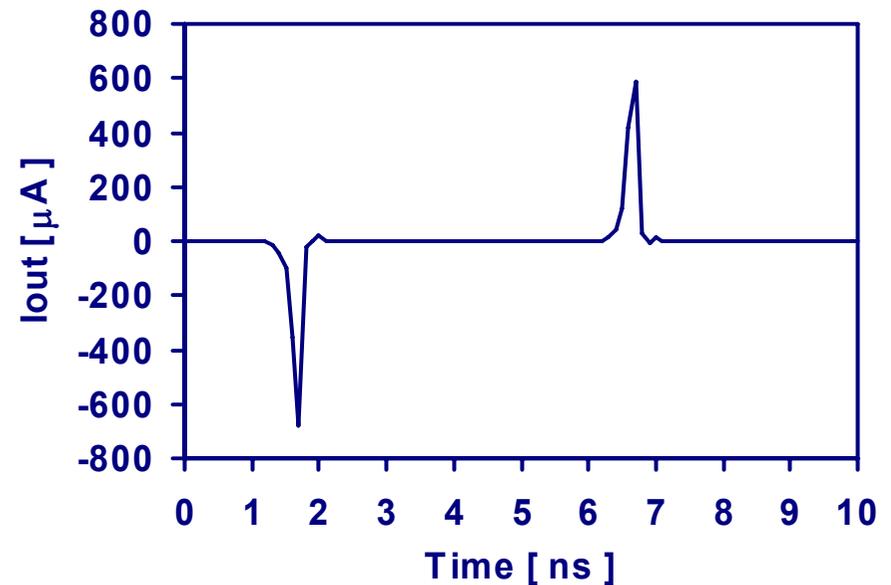
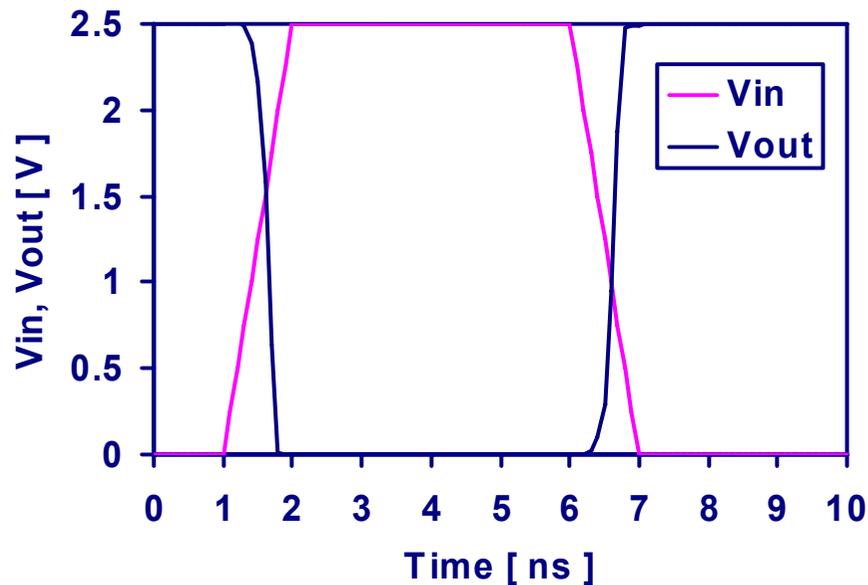
Truth table

IN	OUT
0	1
1	0

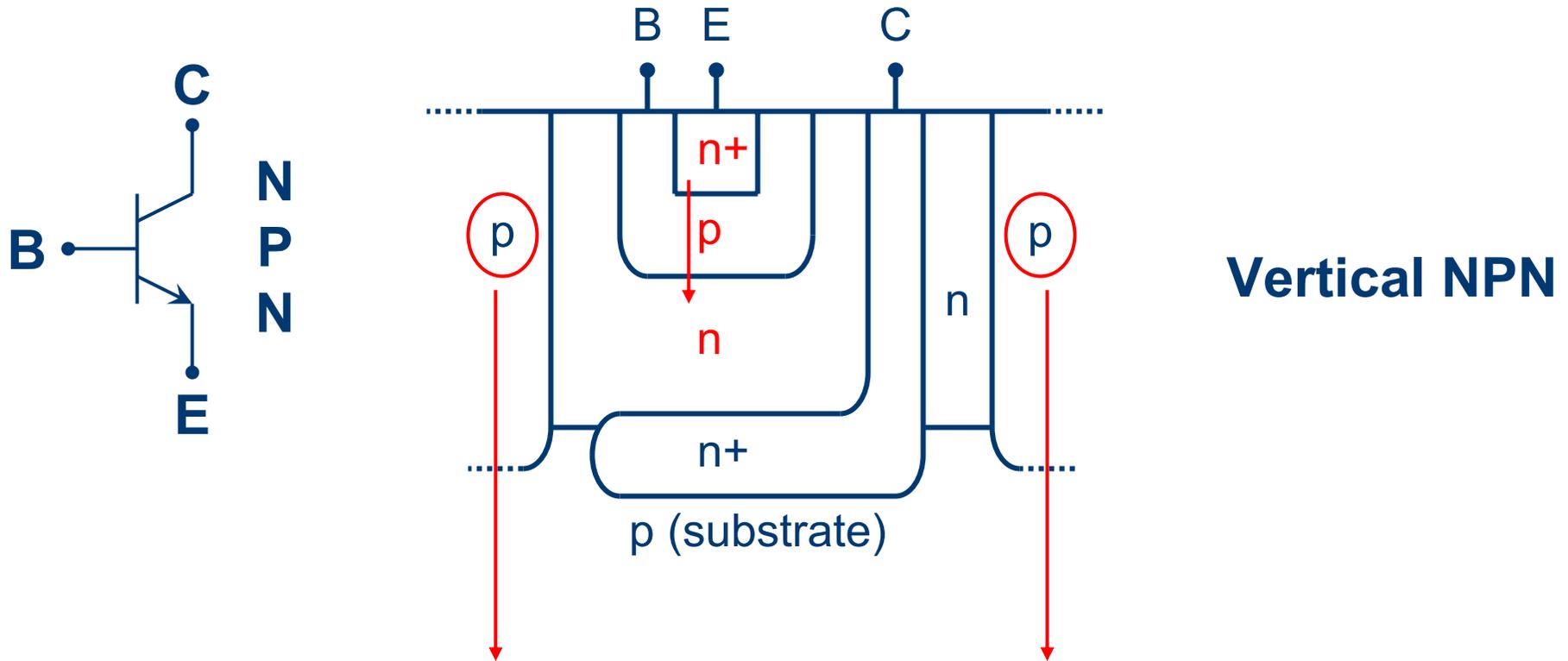


Dynamic current in an inverter

Simulation of a chain of two inverters in a 0.25 μm CMOS technology

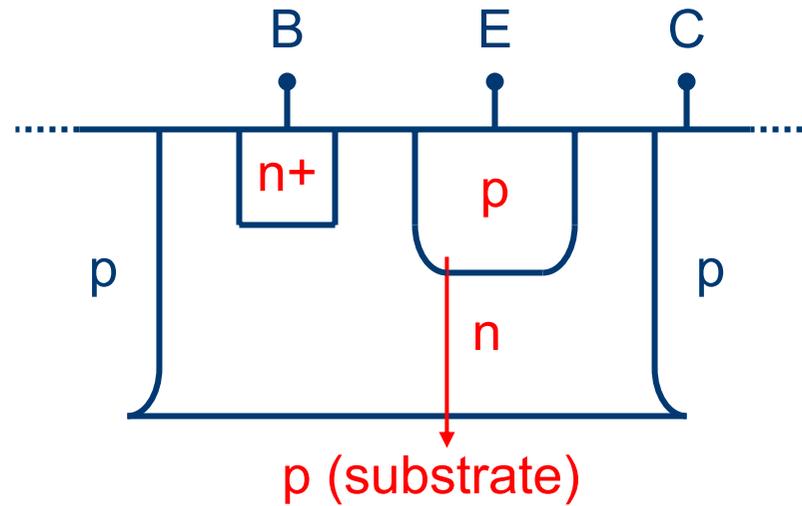
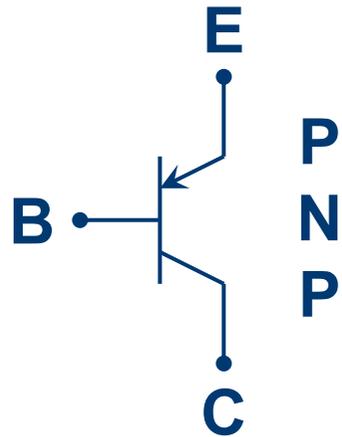


Integrated Bipolar Transistors

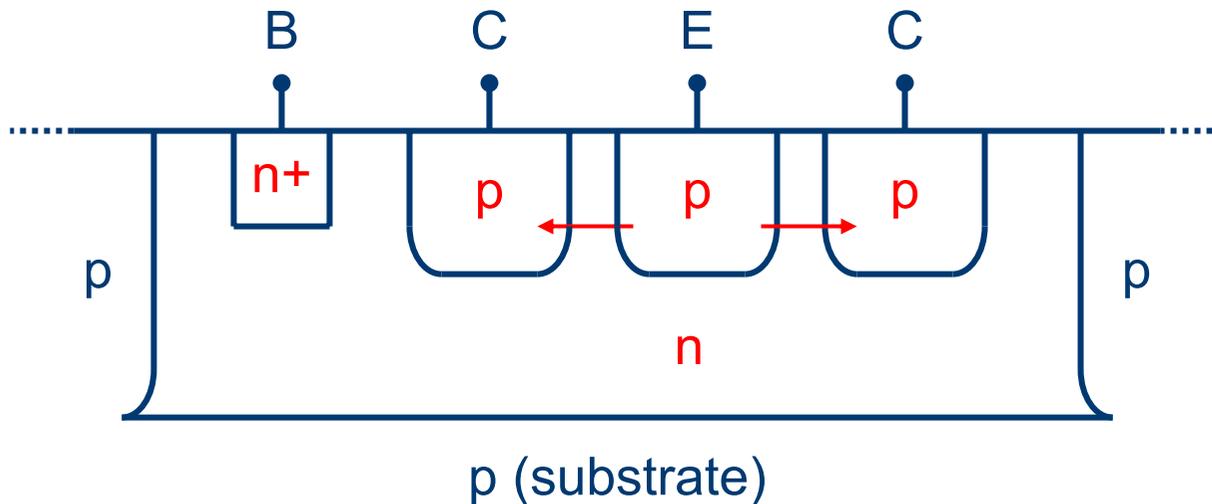


p-type pockets to isolate devices one from another

Integrated Bipolar Transistors



**Vertical PNP
(Common collector!)**



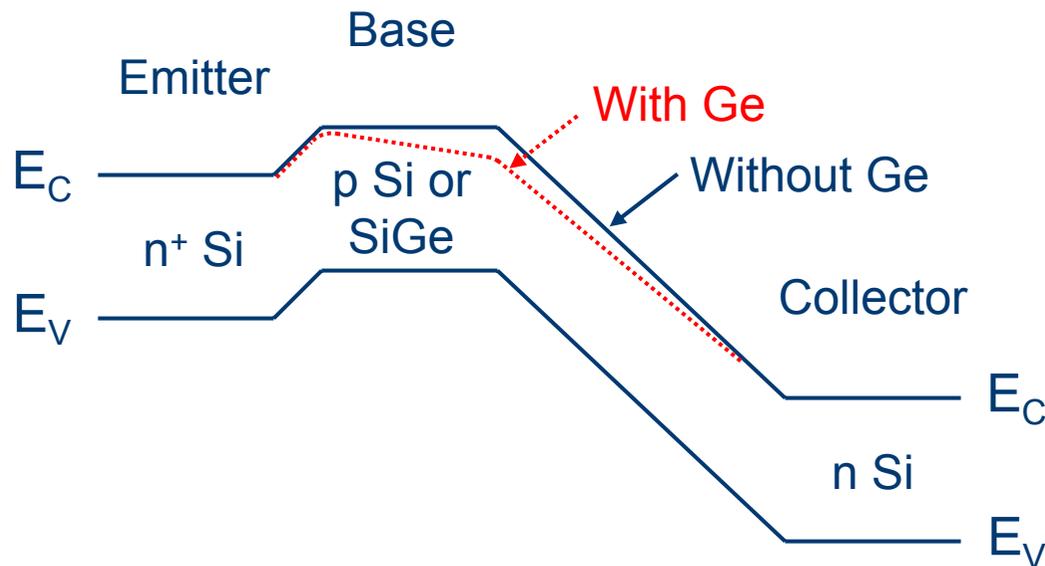
Lateral PNP

Main innovations in modern BJTs

- **Deep trench isolations:** reduce significantly the device horizontal dimensions. p-type diffusions were normally used in the past, but they are as wide as they are deep
- **Polysilicon emitter:** compared to a diffused emitter, it gives smaller emitter junction depth and has a lower thermal cycle
- **Self-aligned polysilicon base contact:** reduces dimensions and allows thin-base transistors to be made more easily
- **SiGe graded-base-bandgap transistors (also referred to as Heterojunction Bipolar Transistors – HBTs):** Germanium is incorporated into the base region of a standard Si bipolar transistor. This is normally done with the ultrahigh-vacuum chemical vapor deposition epitaxial growth technique. The many advantages are discussed in the next slides

SiGe bipolar transistors

The bandgap of Germanium (0.66 eV) is smaller than that of Silicon (1.22 eV). Incorporating Ge into the base region of a Si bipolar transistor modifies the bandgap in the base region. Ge concentration is normally higher closer to the collector, so that an electric field is created in the base which helps the electrons drifting from Emitter to Collector. This has an important impact on the collector current and does not change the base current.



A 100 meV change of the bandgap across a base width of 100 nm gives a 10 kV/cm field

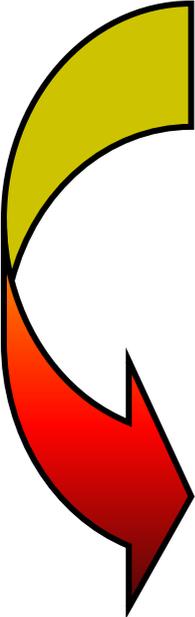
SiGe bipolar transistors

The effects of base-bandgap grading obtained by the inclusion of Ge in the base are all beneficial:

- **The collector current increases while the base current does not change → the current gain increases (x4)**
- **The Early voltage increases (x12)**
- **The base transit time decreases (x2.5), and therefore the speed increases**

(The factors in parenthesis refer to a total base bandgap narrowing of 100 meV)

Why is CMOS so widespread?

- 
- The IC market is driven by digital circuits (memories, microprocessors, ...)
 - Bipolar logic and NMOS - only logic had a too high power consumption per gate
 - Progress in the manufacturing technology made CMOS technologies a reality
 - Modern CMOS technologies offer excellent performance (especially for digital): high speed, low power consumption, VLSI, relatively low cost, high yield

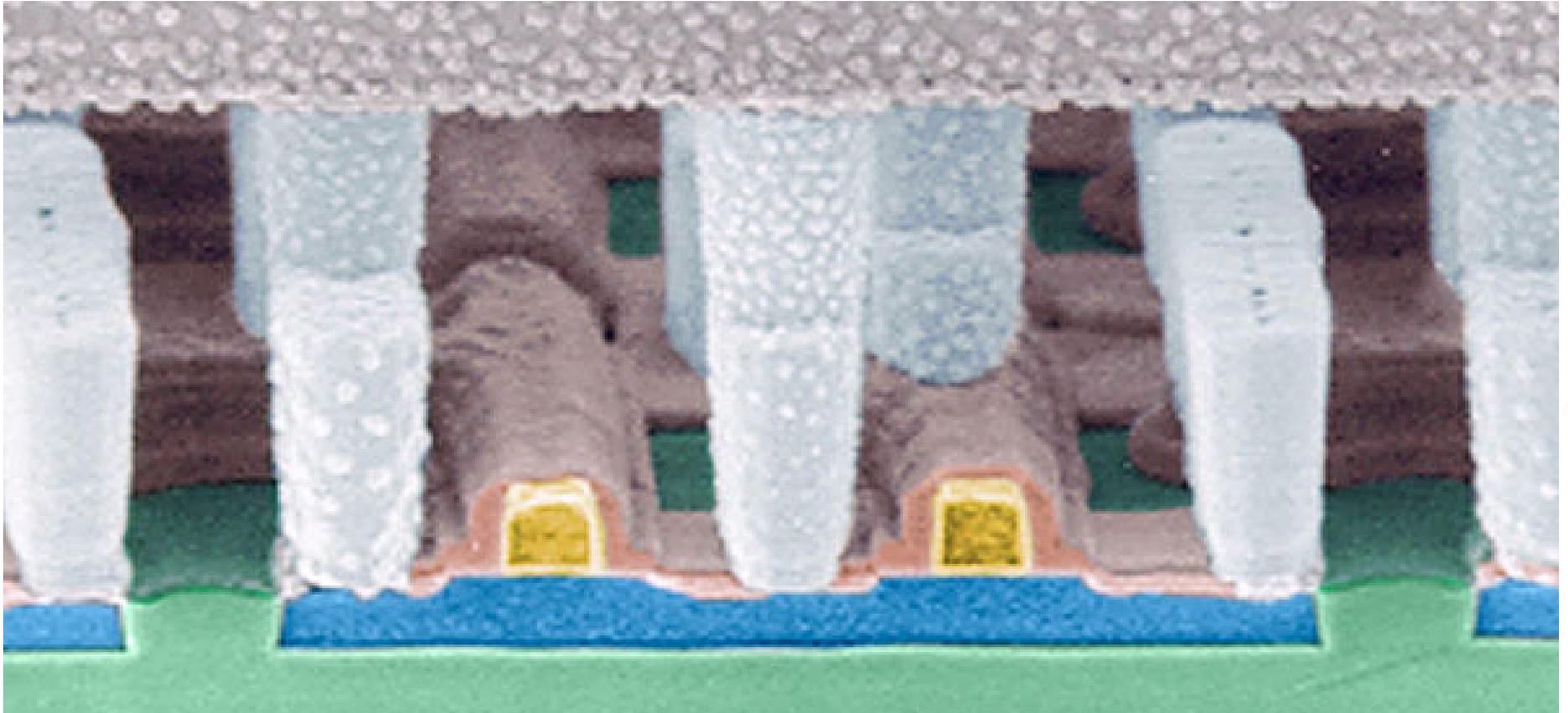
CMOS technologies occupies the biggest portion of the IC market

BiCMOS technologies offer, in addition, high performance npn bipolar transistors, but are more expensive

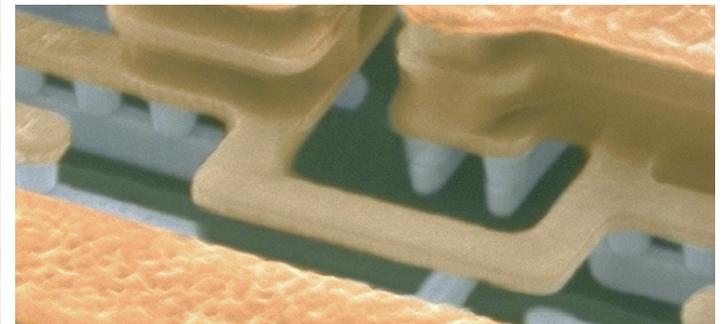
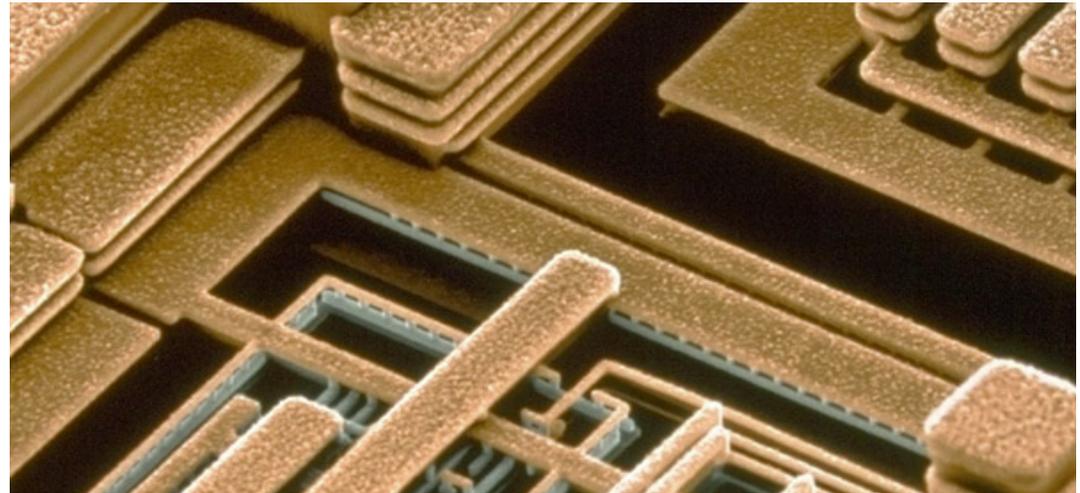
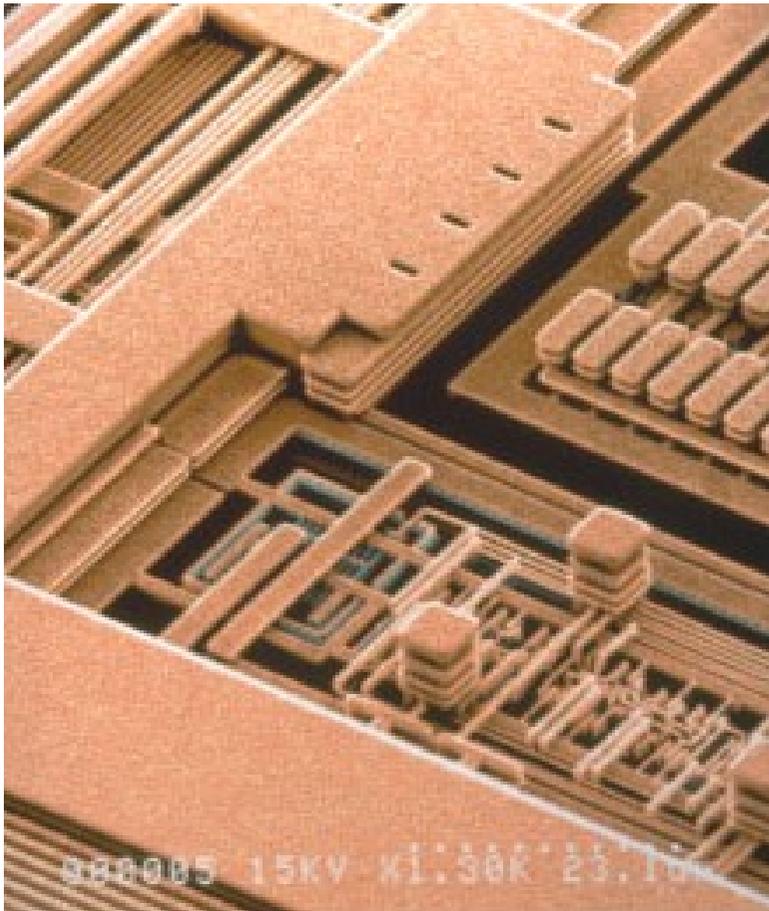
How much does all this cost?

Vendor	Technology	Number of Metal Levels	Price for 10 mm ² [k\$]	Price for 25 mm ² [k\$]	Price for 50 mm ² [k\$]
<i>MOSIS</i>					
IBM	0.13 μm CMOS MixMode	8	57.5	135.1	250.1
IBM	0.13 μm SiGe BiCMOS	7	165	412.5	825
IBM	0.18 μm CMOS MixMode	6	36	70	126.8
IBM	0.18 μm SiGe BiCMOS	7	56	89	144
IBM	0.25 μm CMOS	5	21	47.2	77.2
<i>EUROPRACTICE</i>					
UMC	0.13 μm CMOS	8	Min price given for 25 mm ²	55.4	110.8
UMC	0.18 μm CMOS	6		30.4	60.8
UMC	0.25 μm CMOS	5		16.2	32.4
IHP	0.25 μm SiGe BiCMOS	4	30 - 91	75 - 227.5	150 - 455

SOI CMOS from IBM



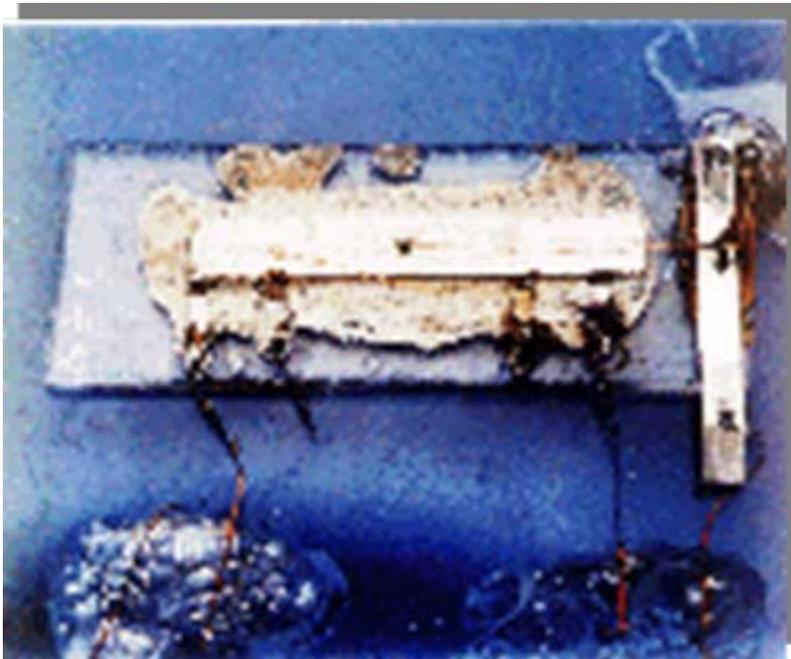
Metalization examples (IBM)



- Operation and characteristics of MOS and Bipolar transistors
- Sub-micron CMOS and BiCMOS technologies
- **Feature size scaling**
 - A bit of history
 - A look into the future
 - CMOS scaling: how does it work?
 - Scaling of Bipolar Transistors
 - Scaling impact on CMOS analog circuits
- Radiation effects and reliability
- Mixed-signal circuits

The first IC

First integrated circuit



Jack S. Kilby, 1958
Texas Instruments



The Nobel Prize in Physics 2000

"for basic work on information and communication technology"

"for developing semiconductor heterostructures used in high-speed- and opto-electronics"

"for his part in the invention of the integrated circuit"



Zhores I. Alferov

🕒 1/4 of the prize

Russia

A.F. Ioffe Physico-Technical Institute
St. Petersburg,
Russia

b. 1930



Herbert Kroemer

🕒 1/4 of the prize

Federal Republic of
Germany

University of
California
Santa Barbara, CA,
USA

b. 1928



Jack S. Kilby

🕒 1/2 of the prize

USA

Texas Instruments
Dallas, TX, USA

b. 1923
d. 2005

Moore's law



1965: Number of Integrated Circuit components will double every year

G. E. Moore, "Cramming More Components onto Integrated Circuits", *Electronics*, vol. 38, no. 8, 1965.

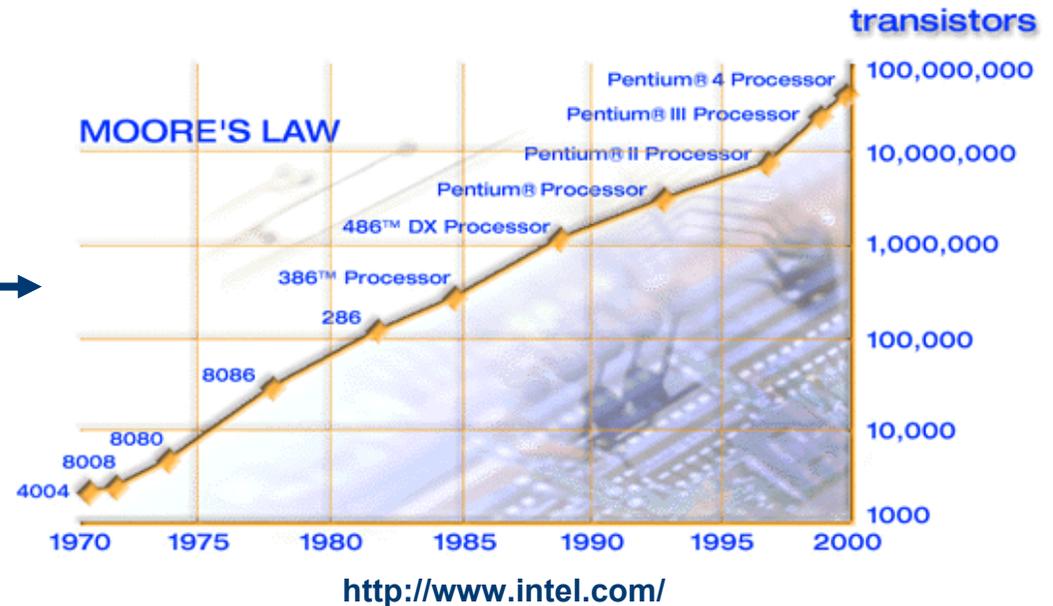
1975: Number of Integrated Circuit components will double every 18 months

G. E. Moore, "Progress in Digital Integrated Electronics", *Technical Digest of the IEEE IEDM 1975*.

1996: The definition of "Moore's Law" has come to refer to almost anything related to the semiconductor industry that when plotted on semi-log paper approximates a straight line. I don't want to do anything to restrict this definition. - G. E. Moore, 8/7/1996

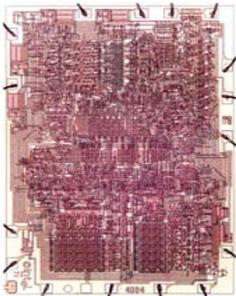
P. K. Bondyopadhyay, "Moore's Law Governs the Silicon Revolution", *Proc. of the IEEE*, vol. 86, no. 1, Jan. 1998, pp. 78-81.

An example:
Intel's Microprocessors

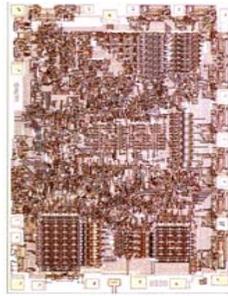


The intel. Microprocessors

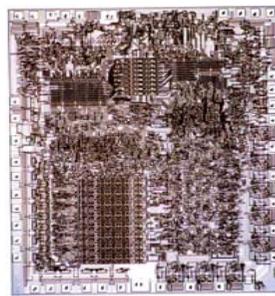
4004
11 / 1971
2300
10 μm
108 KHz



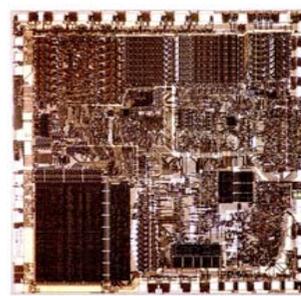
8008
04 / 1972
3500
10 μm
200 KHz



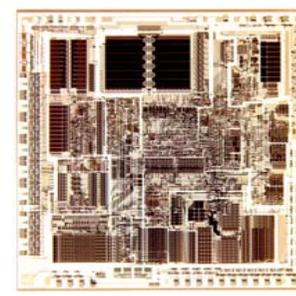
8080
04 / 1974
4500
6 μm
2 MHz



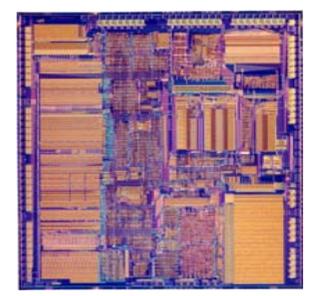
8088
06 / 1979
29000
3 μm
8 MHz



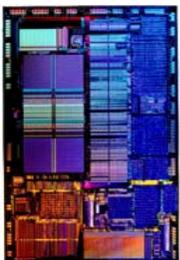
80286
02 / 1982
134000
1.5 μm
12 MHz



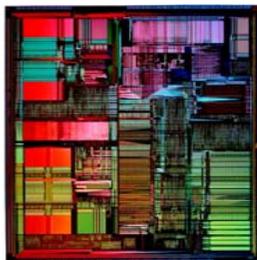
Intel386™
10 / 1985
275000
1 μm
16 MHz



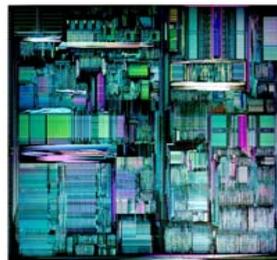
Intel486™ DX
04 / 1989
1.2 M
1 μm
25 MHz



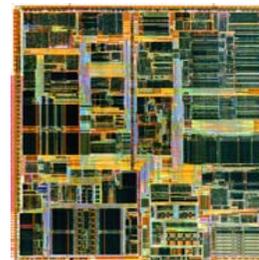
Pentium®
03 / 1993
3.1 M
0.8 μm
66 MHz



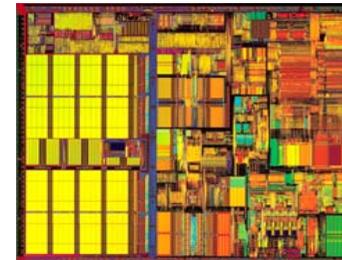
Pentium® Pro
11 / 1995
5.5 M
0.6 μm
150 MHz



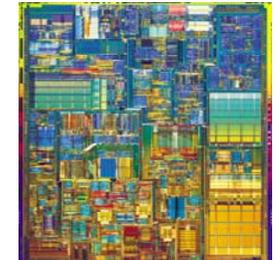
Pentium® II
05 / 1997
7.5 M
0.35 μm
233 MHz



Pentium® III
02 / 1999
9.5 M
0.25 μm
500 MHz



Pentium® 4
11 / 2000
42 M
0.18 μm
1.5 GHz



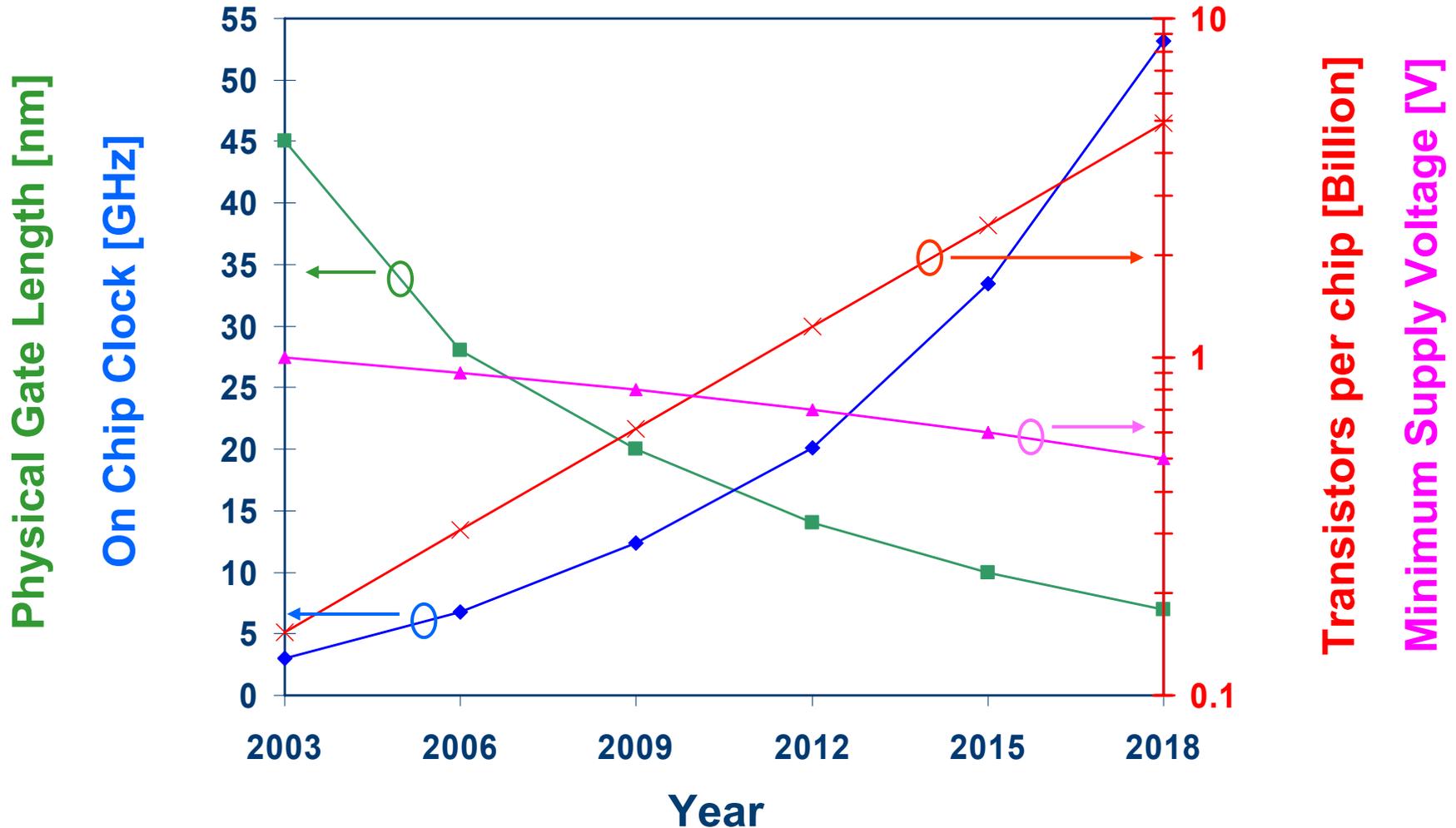
<http://www.intel.com/>

The Roadmap History

- **The National Technology Roadmap for Semiconductors (NTRS):**
Sponsored by the Semiconductor Industry Association (SIA)
Edited in 1992, 1994 and 1997
- **The International Technology Roadmap for Semiconductors (ITRS):**
Sponsored by:
 - **Semiconductor Industry Association**
 - **European Semiconductor Industry Association**
 - **Korea Semiconductor Industry Association**
 - **Japan Electronics and Information Technology Industries Association**
 - **Taiwan Semiconductor Industry Association****Edited in 1998 (Update), 1999, 2000 (Update), 2001, 2002 (Update), 2003, 2004 (Update), 2005**

These documents always contained a 15-year outlook of the semiconductor industry major trends

Future perspectives

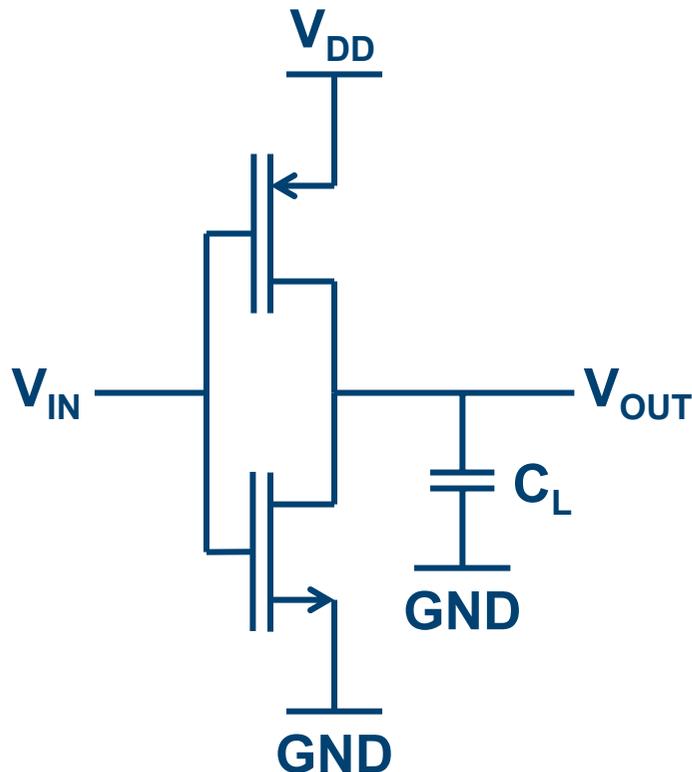


Data taken from the International Technology Roadmap for Semiconductors (2004 Update)

Why CMOS scaling?

Scaling improves density, speed and power consumption of CMOS digital circuits

Example: CMOS inverter



$$P_{\text{static}} = I_{\text{leakage}} \cdot V_{\text{DD}}$$

$$P_{\text{dynamic}} = C_L \cdot V_{\text{DD}}^2 \cdot f$$

$$\text{PDP} = C_L \cdot V_{\text{DD}}^2$$

Power-delay product

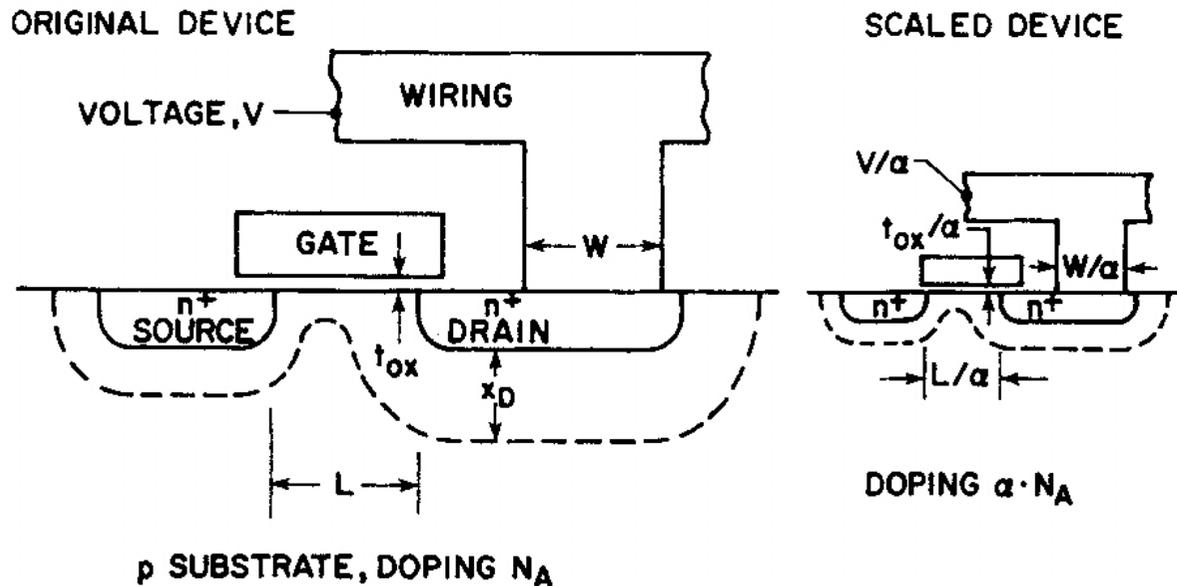
t_{ox} ↓

V_{DD} ↓

C_L ↓

Constant field scaling

The aim of scaling is to reduce the device dimensions (to improve the circuit performance) without introducing effects which could disturb the good operation of the device.



$$x_d = \sqrt{\frac{2\epsilon_{Si}}{qN_A}} \sqrt{\phi_{bi} + V}$$

$$\alpha > 1$$

Constant field scaling

Summary of the scaling factors for several quantities

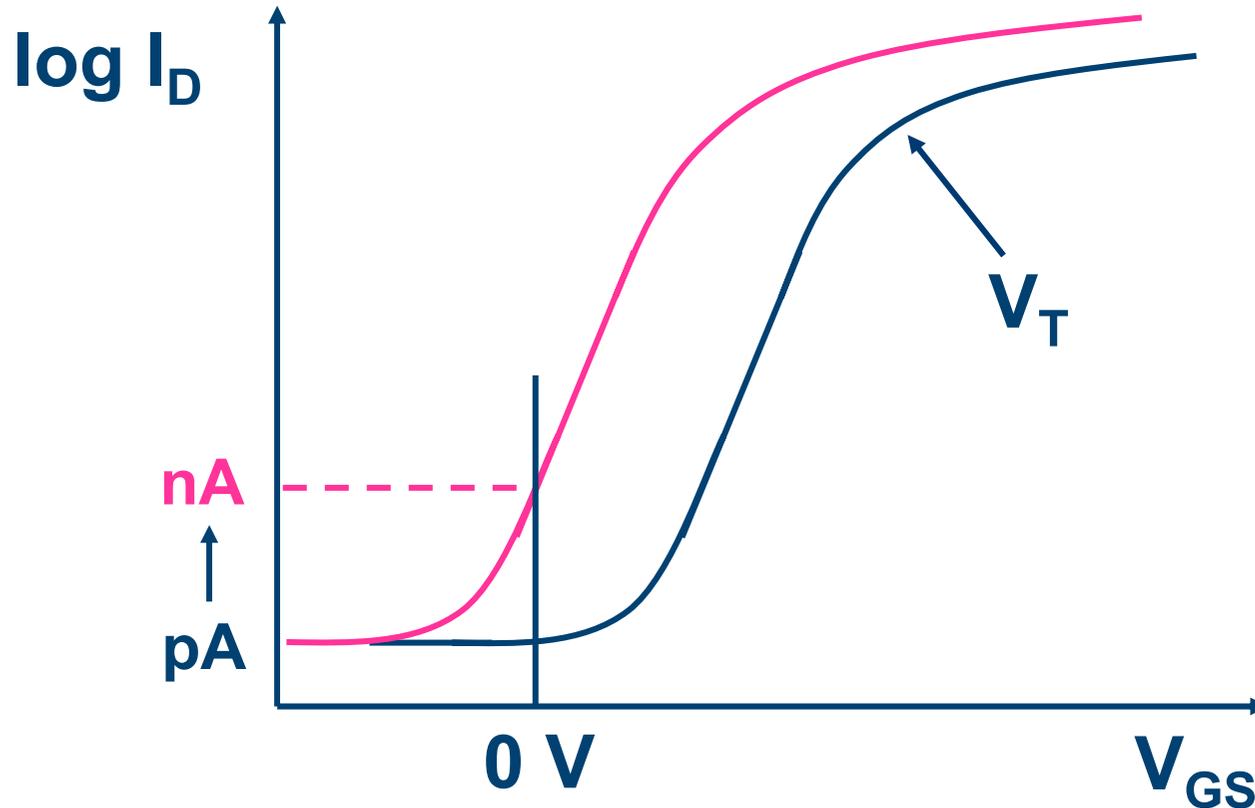
Quantity	Scaling Factor	Quantity	Scaling factor
Device dimensions (L, W, t_{ox} , x_D)	$1/\alpha$	Capacitances	$1/\alpha$
Area	$1/\alpha^2$	Capacitances per unit area	α
Devices per unit of chip area (density)	α^2	Charges	$1/\alpha^2$
Doping concentration (N_A)	α	Charges per unit area	1
Bias voltages and V_T	$1/\alpha$	Electric field intensity	1
Bias currents	$1/\alpha$	Body effect coefficient (γ)	$1/\sqrt{\alpha}$
Power dissipation for a given circuit	$1/\alpha^2$	Transistor transit time (τ)	$1/\alpha$
Power dissipation per unit of chip area	1	Transistor power-delay product	$1/\alpha^3$

$$\alpha > 1$$

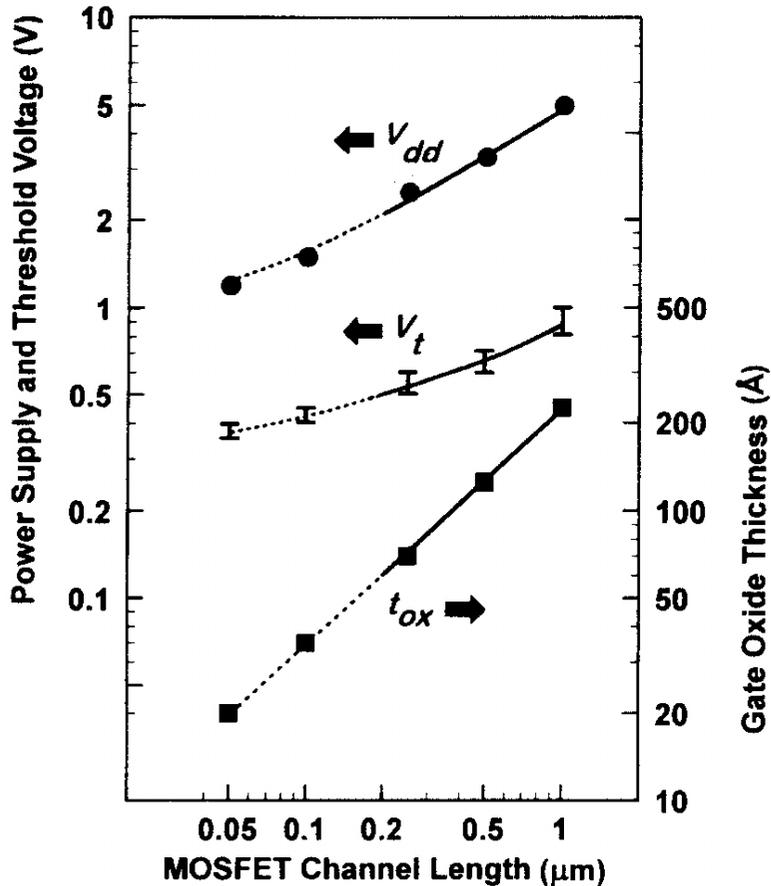
$$C_{ox} = \frac{\epsilon_{SiO_2}}{t_{ox}}$$

Constant field scaling problem

Subthreshold slope and width of the moderate inversion region do not scale. This can have a devastating impact on the static power consumption of a digital circuit.



Generalized scaling



- The dimensions in the device scale as in the *constant field* scaling
- V_{dd} scales to have reasonable electric fields in the device, but slower than t_{ox} , to have an useful voltage swing for the signals
- The doping levels are adjusted to have the correct depletion region widths
- To limit the subthreshold currents, V_T scales more slowly than V_{dd}

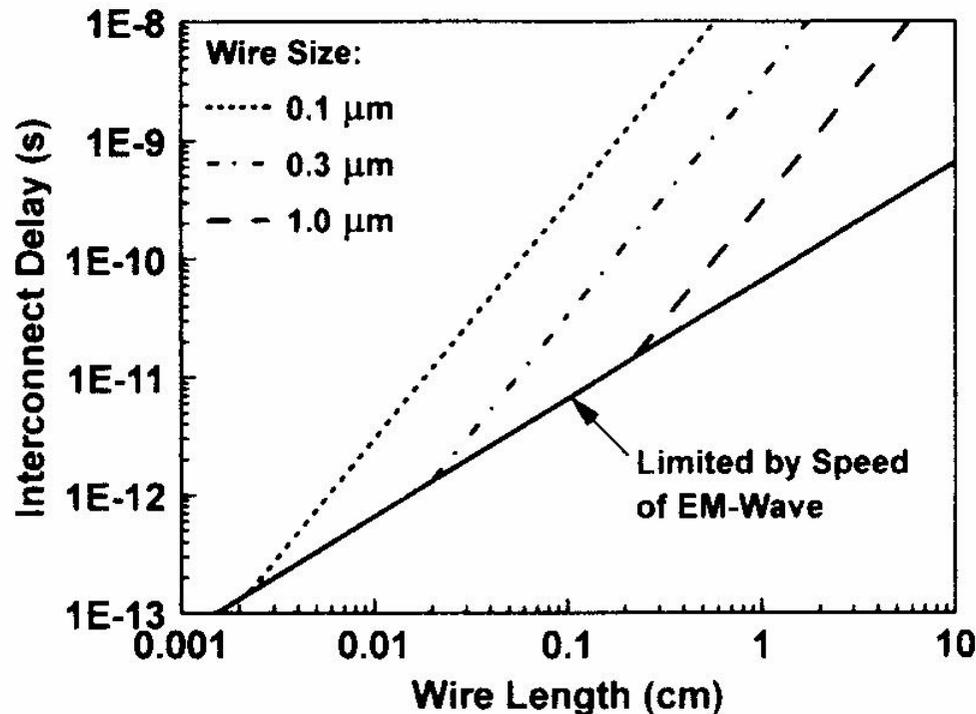
Y. Taur et al., "CMOS Scaling into the Nanometer Regime", *Proc. of the IEEE*, vol. 85, no. 4, Apr. 1997, pp. 486-504.

Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge University Press, 1998, p. 186.

Scaling of interconnections

An accurate scaling of the interconnections is needed as well, so that we can profit at the circuit level of the improvements made at the device level.

Interconnections are becoming more and more important in modern technologies because the delay they introduce is becoming comparable with the switching time of the digital circuits.



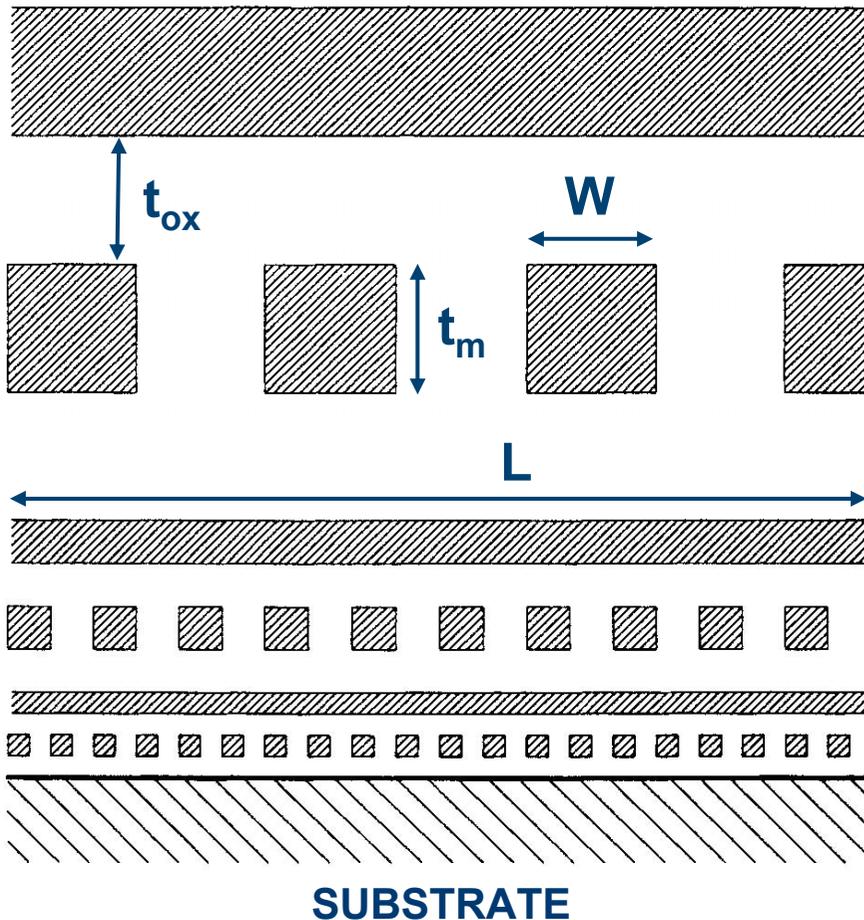
**Wires
with
square
section**

Y. Taur et al., "CMOS Scaling into the Nanometer Regime", *Proceedings of the IEEE*, vol. 85, no. 4, Apr. 1997, pp. 486-504.

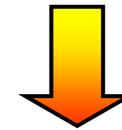
T. N. Theis, "The future of interconnection technology", *IBM Journal of Research and Development*, vol. 44, no. 3, May 2000, pp. 379-390.

“Reverse” scaling

The scaling method is different from the one applied to devices



If W , L , t_m and t_{ox} are decreased by α



- Current density increases by α
- R increases by α , C decreases by α
- RC (delay) does not scale!!!

In practice, wires dimensions are reduced only for local interconnections (but not t_m). At the chip scale, t_m and t_{ox} are increased (reverse scaling).

Bipolar transistors scaling

Scaling has been much less emphasized for bipolar than for MOS transistors, but reducing the lateral and vertical dimensions of a BJT has a beneficial impact on its performance.

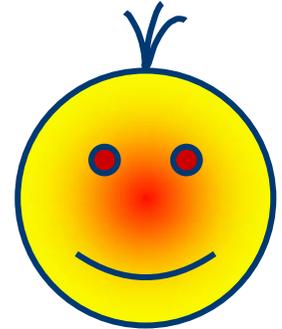
Reducing the lateral dimensions increases the density and reduces the parasitic resistances and capacitances (→ speed increases).

Reducing the vertical dimensions (in particular the base width / depth W_B) increases the collector current density, the current gain and the speed.

Scaling must be performed adjusting doping levels and bias voltages accordingly (breakdown voltages!).

Scaling impact on CMOS analog ICs

t_{ox} scales \longrightarrow for the same device dimensions



- Threshold voltage matching improves

$$\sigma_{\Delta V_{th}} = \frac{\text{Const} \cdot t_{ox}}{\sqrt{WL}}$$

- 1/f noise decreases

$$\frac{\overline{v_{in_1/f}^2}}{\Delta f} = \frac{K_a}{C_{ox}^2 WL} \frac{1}{f^\alpha}$$

- Transconductance increases (same current)

$$g_m = \sqrt{\frac{2}{n} \mu C_{ox} \frac{W}{L} I_{DS}}$$

- White noise decreases

Scaling impact on CMOS analog ICs

- **New noise mechanisms**
- **Modeling difficulties**
- **Lack of devices for analog design**
- **Reduced signal swing (new architectures needed)**
- **Substrate noise in mixed-signal circuits**
- **Velocity saturation. Critical field: 3 V/ μm for electrons,
10 V/ μm for holes**



$$g_{m_vel.sat.} = WC_{ox} v_{sat}$$

From weak inversion to velocity saturation

Weak inversion (w.i.)

$$I_{DS_w.i.} = I_{D0} \frac{W}{L} e^{\frac{V_{GS}}{n\phi_t}}$$

$$g_{m_w.i.} = \frac{I_{DS}}{n\phi_t}$$

Strong inversion (s.i.)

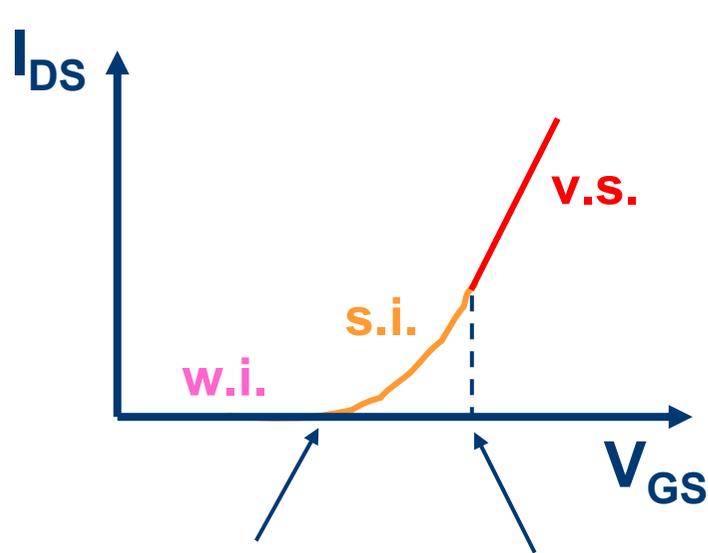
$$I_{DS_s.i.} = \frac{\beta}{2n} (V_{GS} - V_T)^2$$

$$g_{m_s.i.} = \sqrt{2 \frac{\beta}{n} I_{DS}}$$

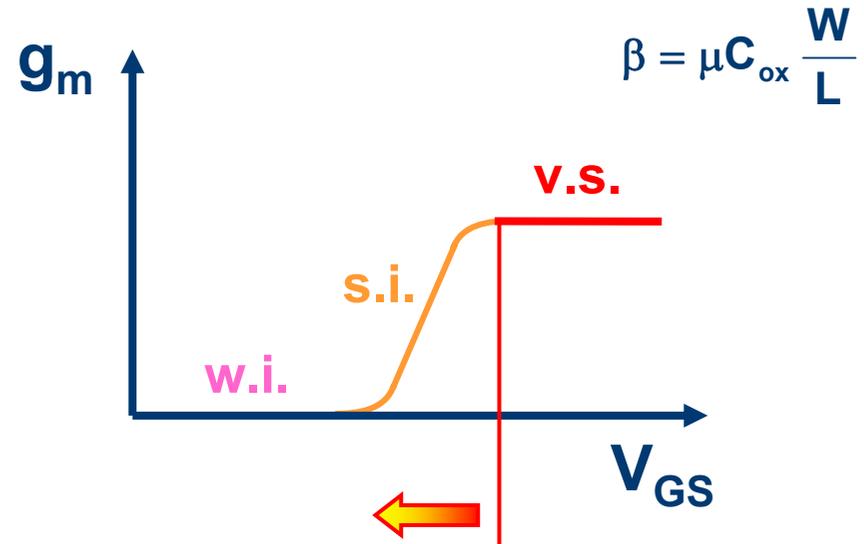
Velocity saturation (v.s.)

$$I_{DS_v.s.} = WC_{ox} v_{sat} (V_{GS} - V_T)$$

$$g_{m_v.s.} = WC_{ox} v_{sat}$$



$$V_{w.i._to_s.i.} = 2n\phi_t \quad V_{s.i._to_v.s.} = 2nL \frac{v_{sat}}{\mu}$$



$V_{s.i._to_v.s.}$ decreases with scaling!!!

$$\beta = \mu C_{ox} \frac{W}{L}$$

Scaling impact on power, speed, SNR

$$PWR = I_{DS} \cdot V_{DD}$$

$$\overline{v_{n_white}^2} = 4kTn\gamma \frac{1}{g_m}$$

$$SNR_w = \frac{V_{DD}}{\sqrt{\overline{v_{n_white}^2}}}$$

Assuming constant field scaling and strong inversion:

W	L	β	I_{DS}	PWR	$C_{ox} * W * L$	Q	$\Delta t = Q/I$	$\sqrt{\overline{v_{n_white}^2}}$	SNR_w
$1/\alpha$	$1/\alpha$	α	$1/\alpha$	$1/\alpha^2$	$1/\alpha$	$1/\alpha^2$	$1/\alpha$	1	$1/\alpha$
1	$1/\alpha$	α^2	1	$1/\alpha$	1	$1/\alpha$	$1/\alpha$	$1/\alpha^{1/2}$	$1/\alpha^{1/2}$
$1/\alpha$	1	1	$1/\alpha^2$	$1/\alpha^3$	1	$1/\alpha$	α	$\alpha^{1/2}$	$1/\alpha^{3/2}$
1	1	α	$1/\alpha$	$1/\alpha^2$	α	1	α	1	$1/\alpha$
α	$1/\alpha$	α^3	α	1	α	1	$1/\alpha$	$1/\alpha$	1

To maintain the same SNR we do not gain in Power !!!

Outline

- Operation and characteristics of MOS and Bipolar transistors
- Sub-micron CMOS and BiCMOS technologies
- Feature size scaling
- **Radiation effects and reliability**
 - Introduction
 - Total Ionizing Dose (TID) effects
 - Displacement Damage
 - Single Event Effects (SEE)
 - Solutions to the problems induced by radiation in CMOS technologies
 - Reliability issues
- Mixed-signal circuits



Interaction radiation-matter

The two most important phenomena to be considered are ionization and nuclear displacement.

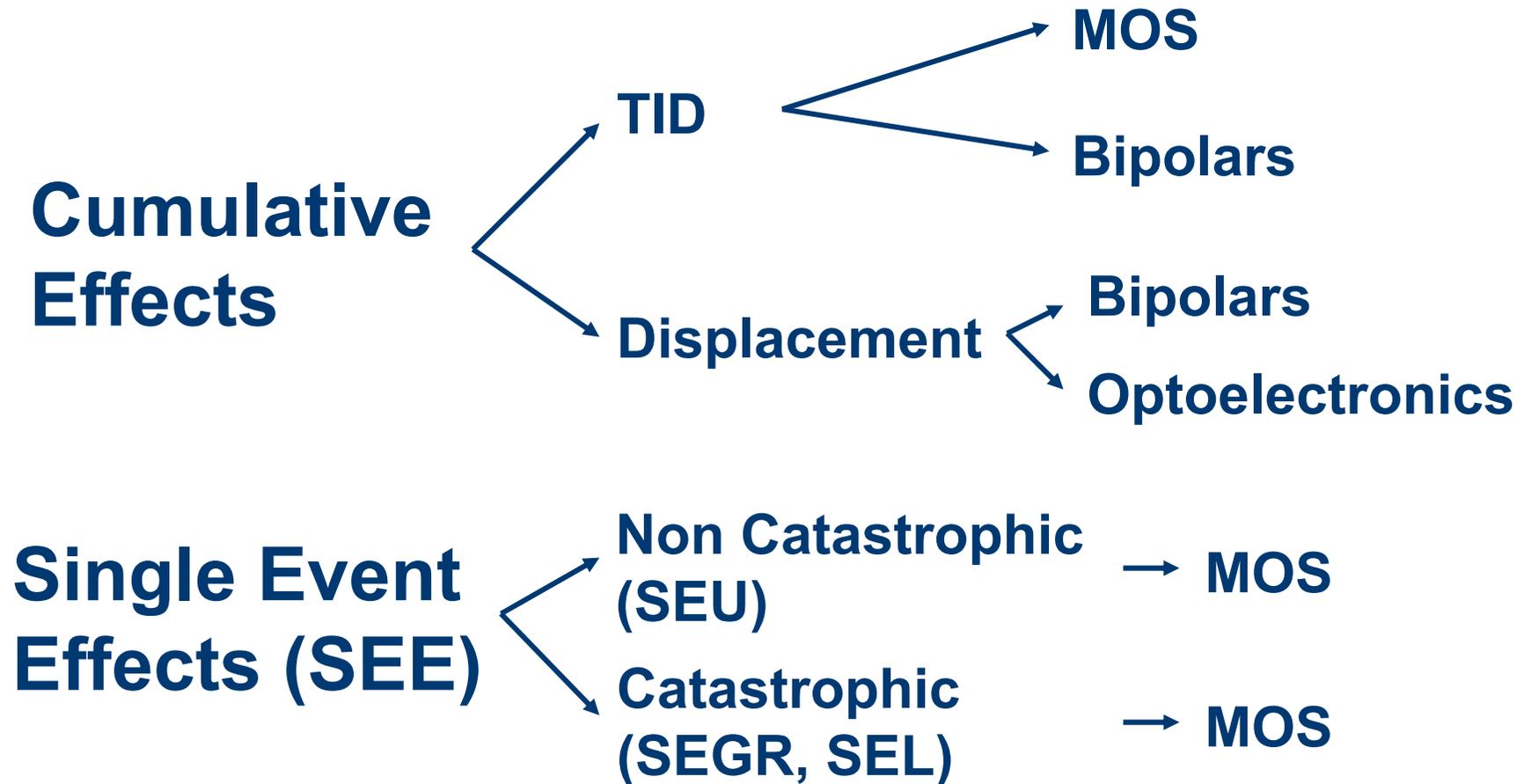
Neutrons give origin mainly to nuclear displacement.

Photons give ionization.

Charged hadrons and heavy ions give both at the same time.

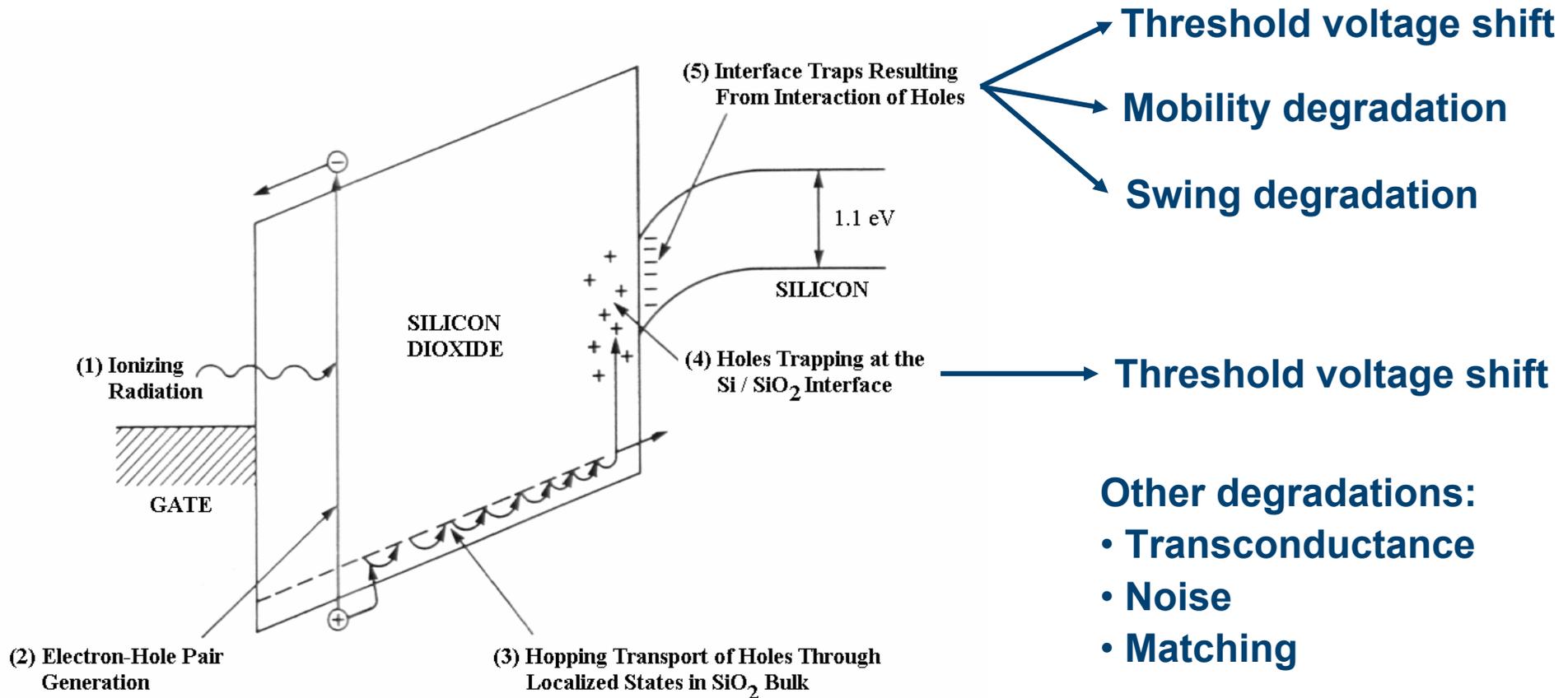
For ionization we talk about **Total Ionizing Dose (TID)**, for nuclear displacement about **Fluence**

Interaction radiation - ICs



Ionizing particles through a MOST

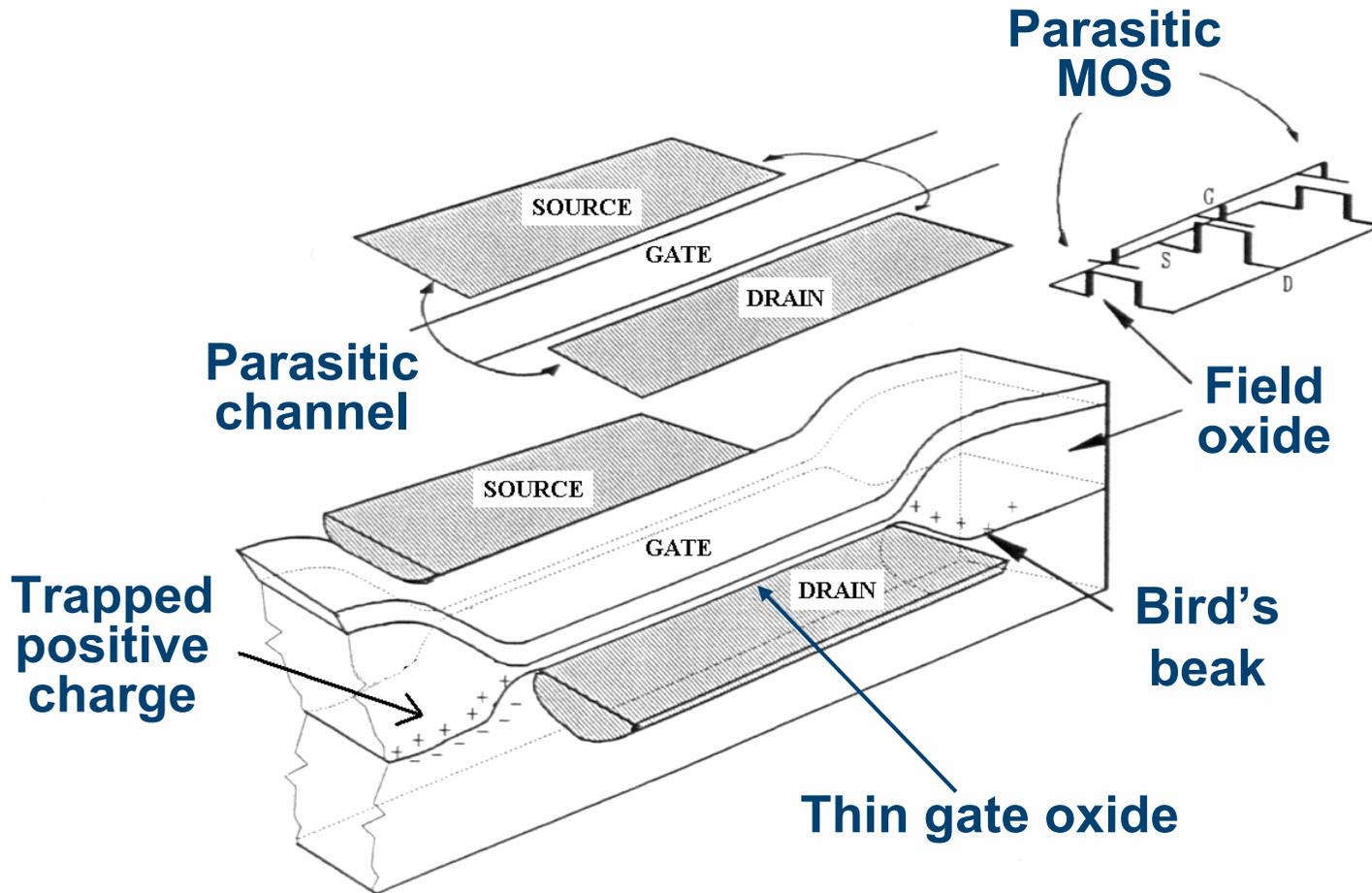
Problems due to ionization in the thin gate oxide



F. B. McLean and T. R. Oldham, Harry Diamond Laboratories Technical Report, No. HDL-TR-2129, September 1987.

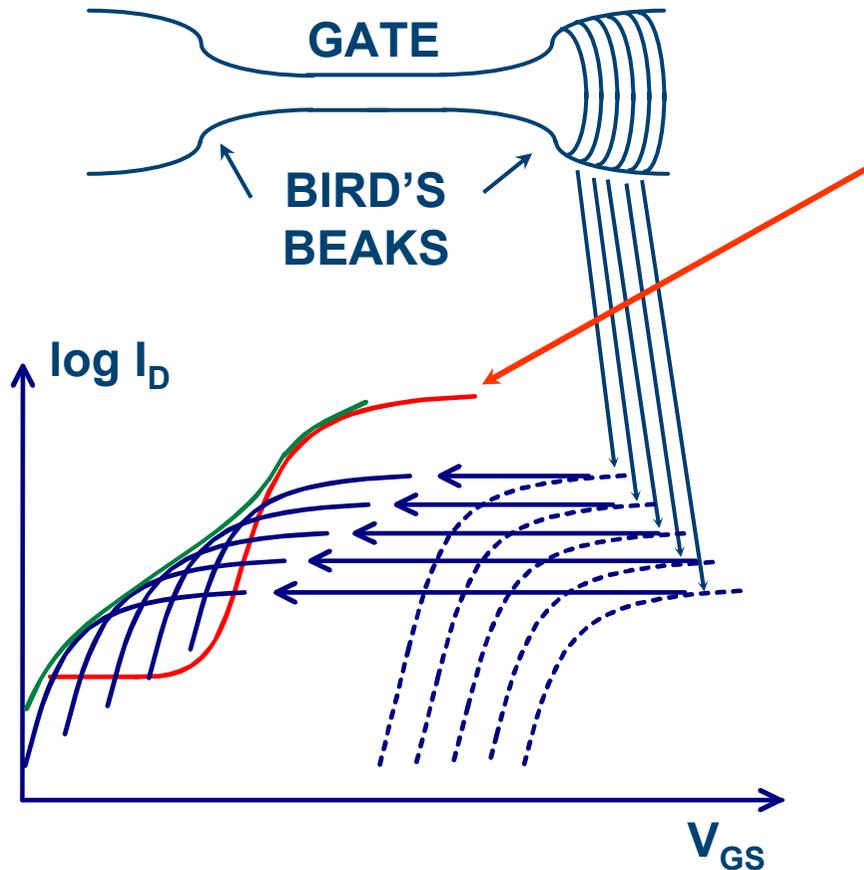
Transistor level leakage (NMOS)

Problem due to ionization in the thick field (lateral) oxide

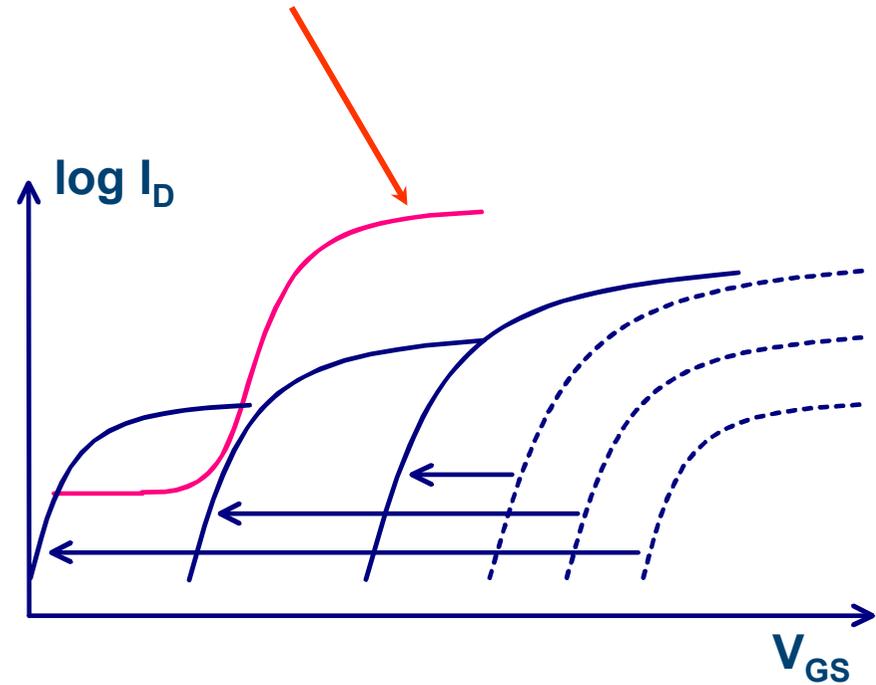


R. Gaillard, J.-L. Leray, O. Musseau et al., "Techniques de durcissement des composants, circuits, et systèmes électroniques", Notes of the Short Course of the 3rd European Conference on Radiation and its Effects on Components and Systems, Arcachon (France), Sept. 1995.

Transistor level leakage (NMOS)

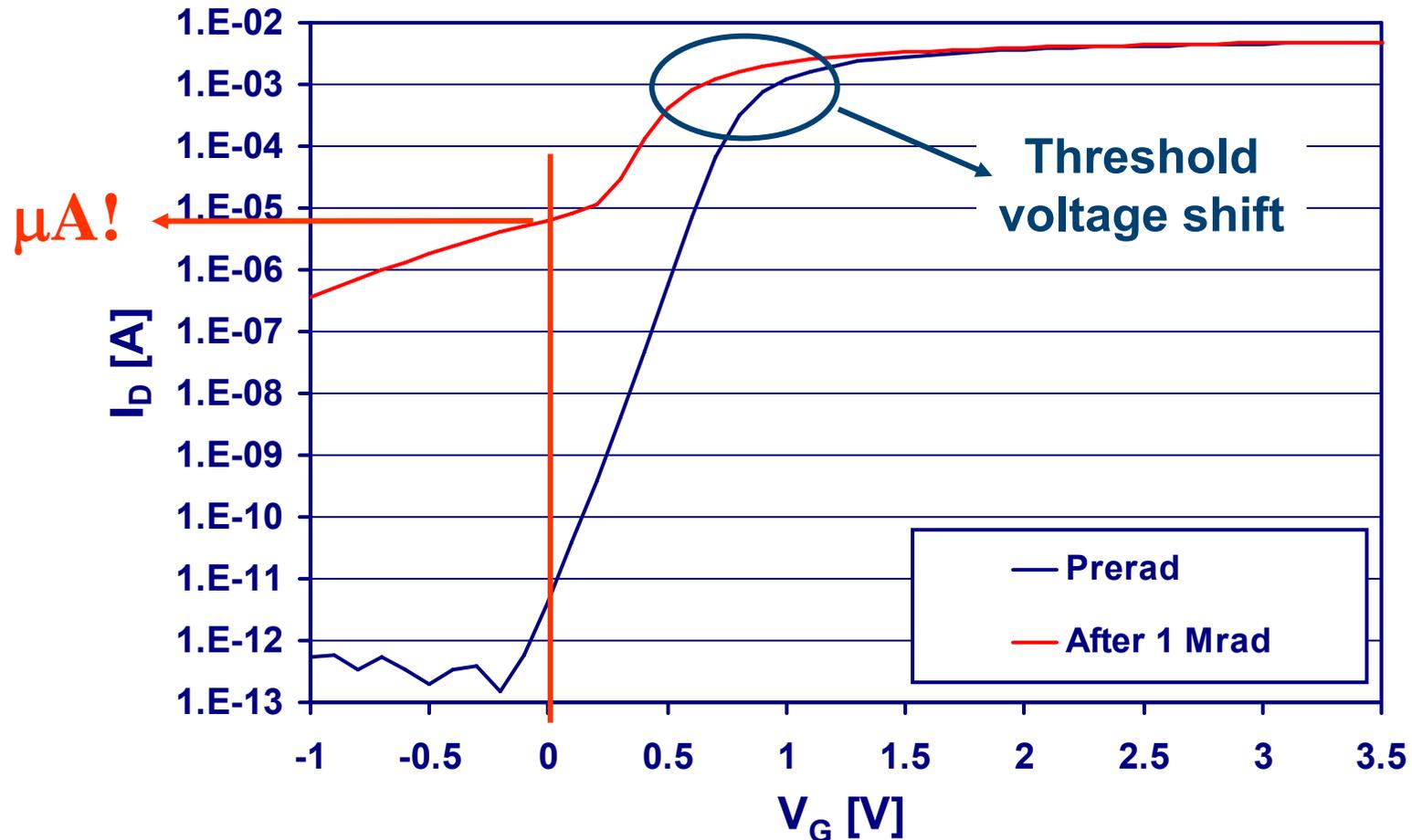


**“CENTRAL”
(MAIN) MOS
TRANSISTOR**



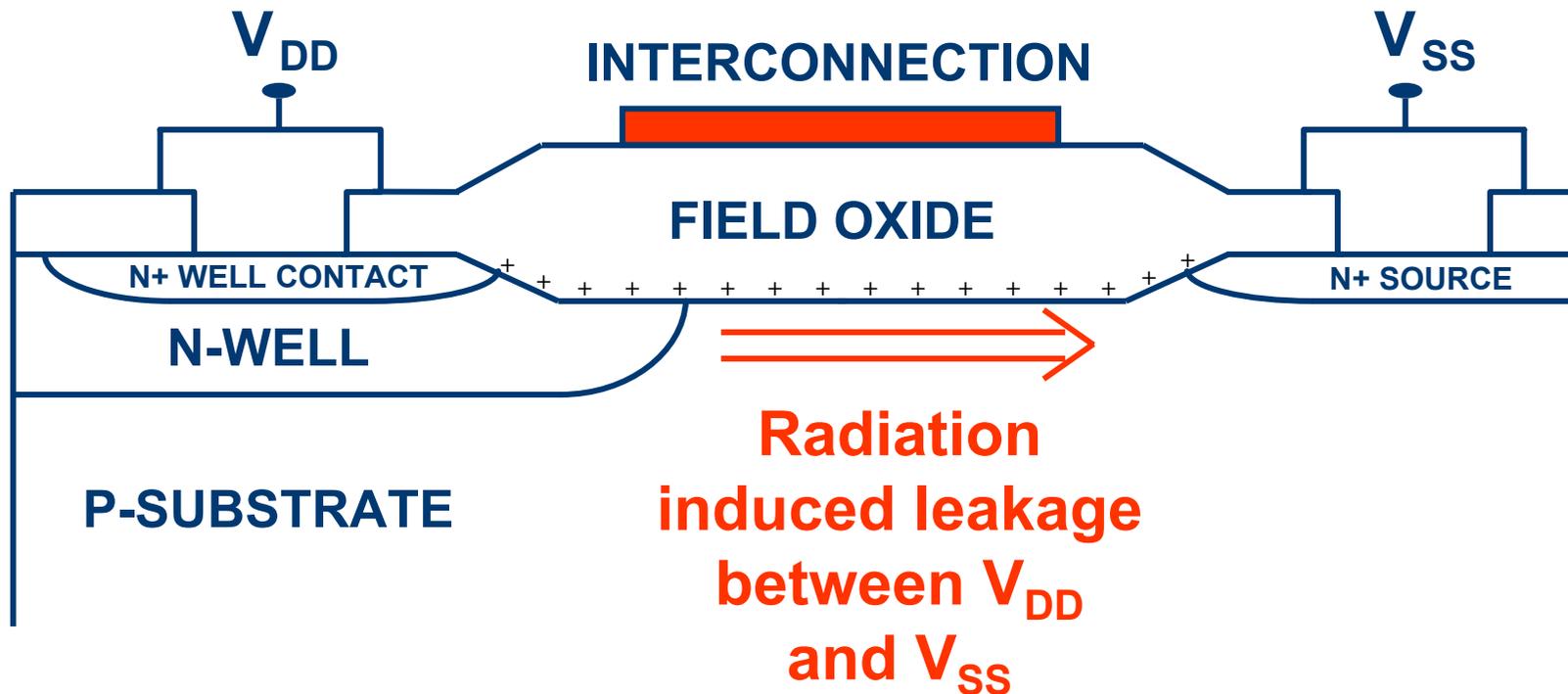
Transistor level leakage: example

NMOS - 0.7 μm technology - $t_{\text{ox}} = 17 \text{ nm}$

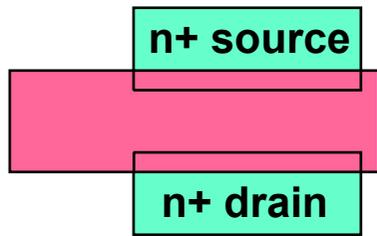


Field oxide leakage

Again problem due to ionization in the thick field (also called lateral or isolation) oxide



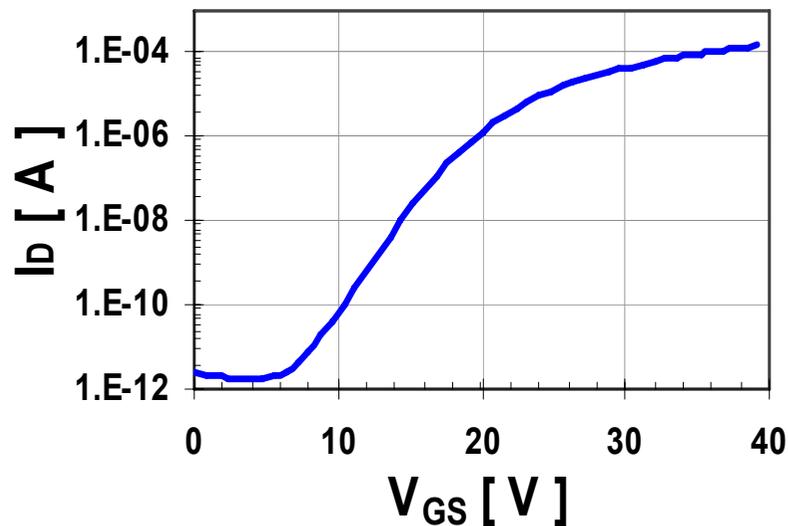
The Field Oxide Transistor



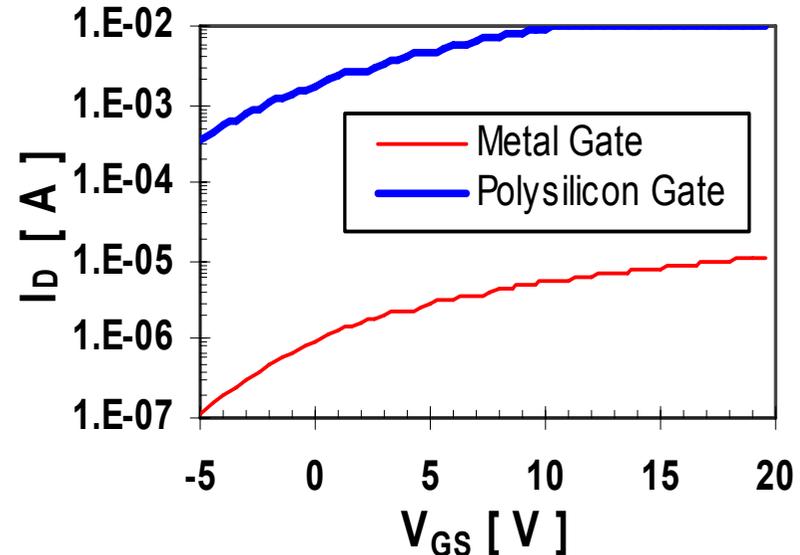
Post-irradiation leakage currents depend on

- Total Dose
- Bias conditions
- Gate Material
- Field oxide quality

PRERAD

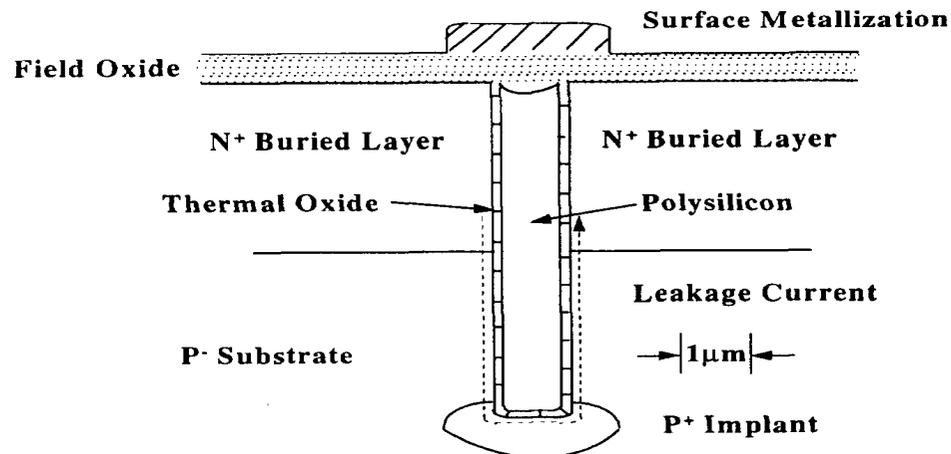
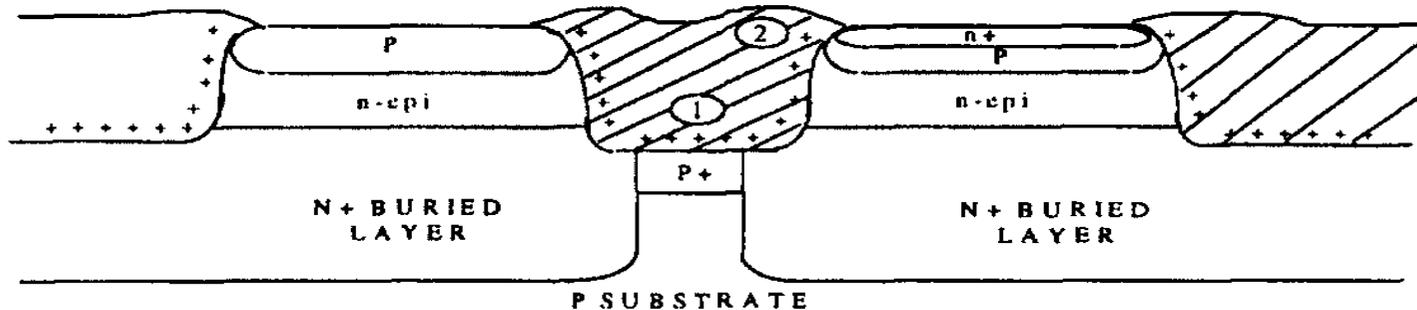


AFTER 1 Mrad



TID damage in bipolar devices

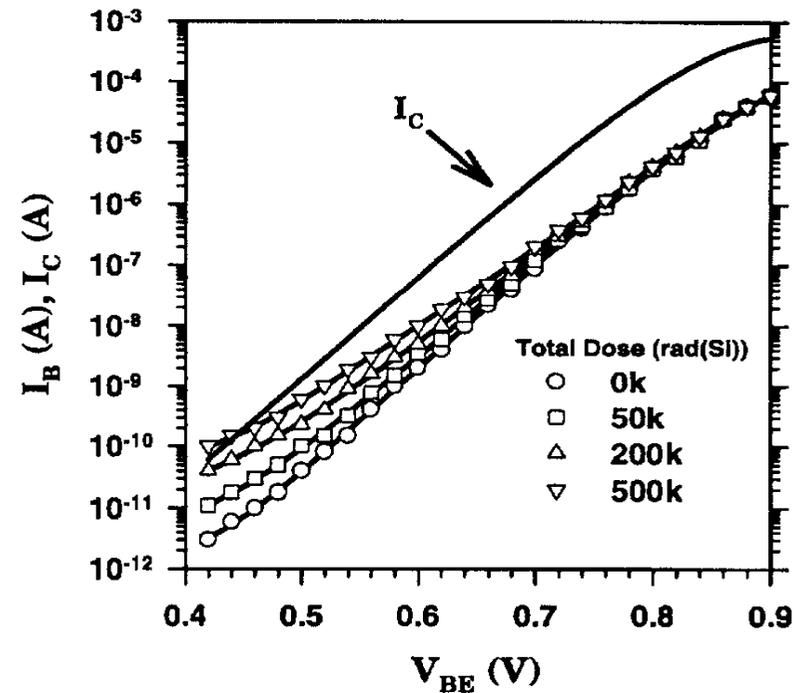
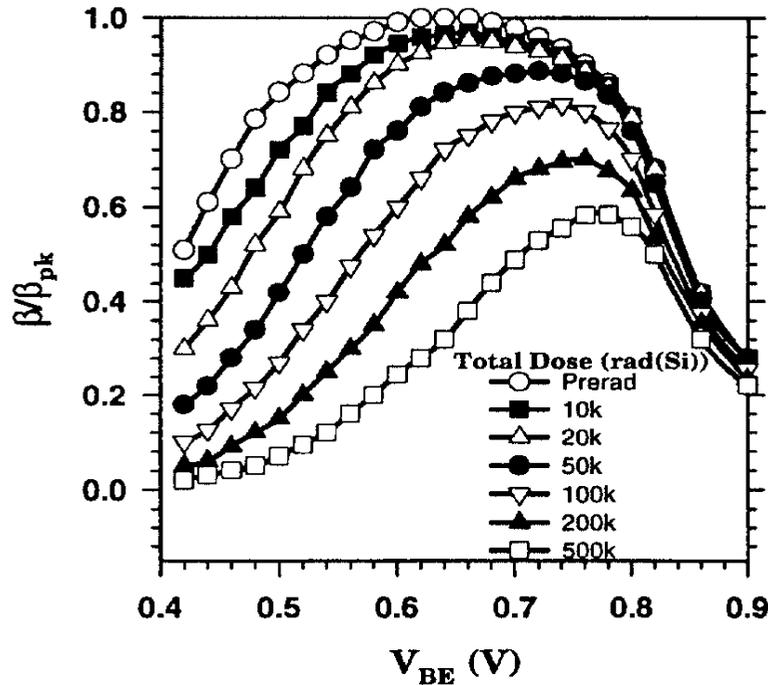
Substrate, sidewall and surface inversion (in oxide-isolated processes)



R.L. Pease et al., IEEE Trans. Nucl. Science. Vol.32, N.6, 1985
E.W. Enlow et al., IEEE Trans. Nucl. Science. Vol.36, N.6, 1989

TID damage in bipolar devices

**Gain degradation:
Increase of the surface component of the base current**



R. N. Nowlin et al., IEEE Trans. Nucl. Science. Vol. 39, N. 6, 1992

TID damage in bipolar devices

Factors affecting TID response of bipolar transistors

- **Transistor polarity**
- **Oxide thickness over base-emitter region**
- **Oxide trap efficiency**
- **Vertical and fringing electric field**
- **Base and Emitter surface concentration**
- **Emitter perimeter-to-area ratio**
- **Transistor geometry (ratio of lateral to vertical current flow)**
- **Injection level**
- **Dose rate**
- **Temperature**

Displacement damage

Displacement increases the recombination of minority carriers in the silicon bulk.

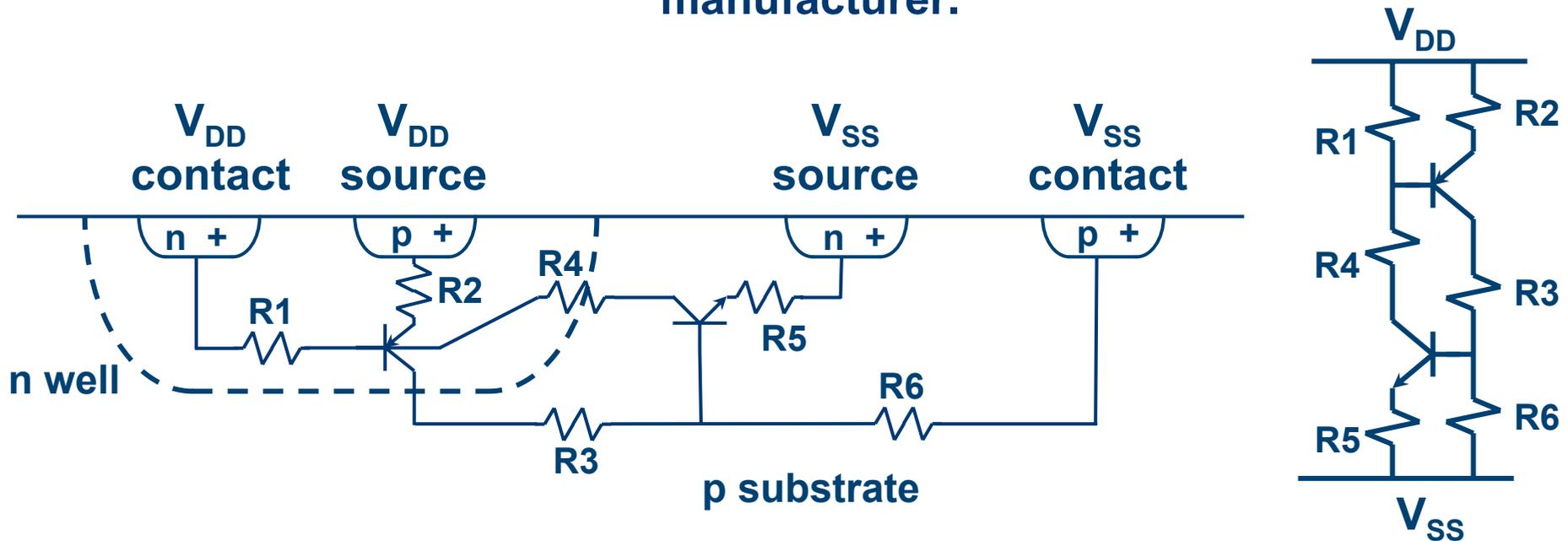
MOS transistors are not sensitive to displacement damage, since they are majority carrier devices and the transistor action takes place close to the SiO_2 -Si interface (not in the silicon bulk).

Bipolar transistors, on the other hand, are sensitive to displacement damage, since they are minority carrier devices and the transistor action takes place in the silicon bulk. The increased recombination in the base increases the base current and reduces the gain.

Scaling of bipolar transistors helps reducing the sensitivity to displacement damage, since it reduces the transistor active volume.

Single Event Latch-up (SEL)

Electrical latch-up in CMOS processes might be initiated by electrical transients on input / output lines, elevated T or improper sequencing of power supply biases. These modes are normally addressed by the manufacturer.

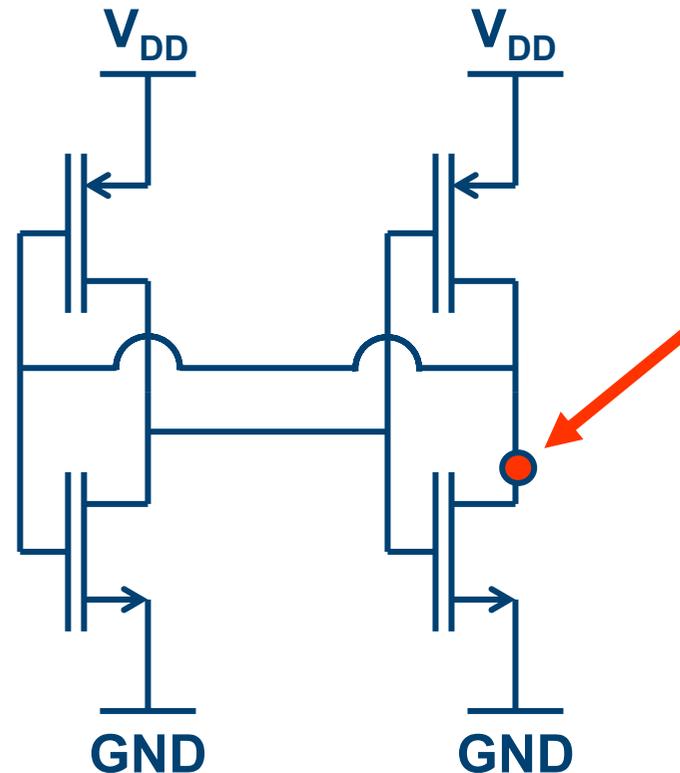
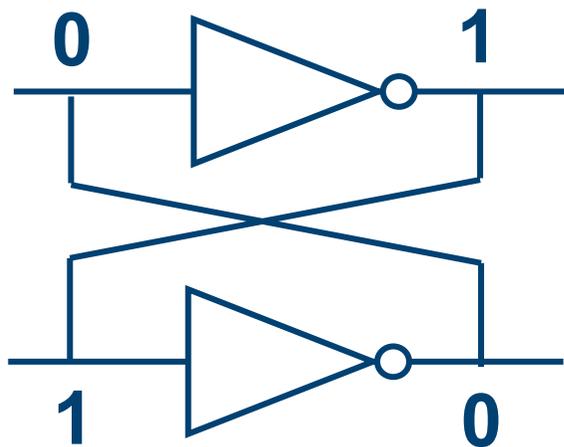


Latch-up can be initiated by ionizing particles (SEL).

If the supply is not cut quickly enough, it can be destructive!

Single Event Upset (SEU)

Static RAM cell



Highly
energetic
particle

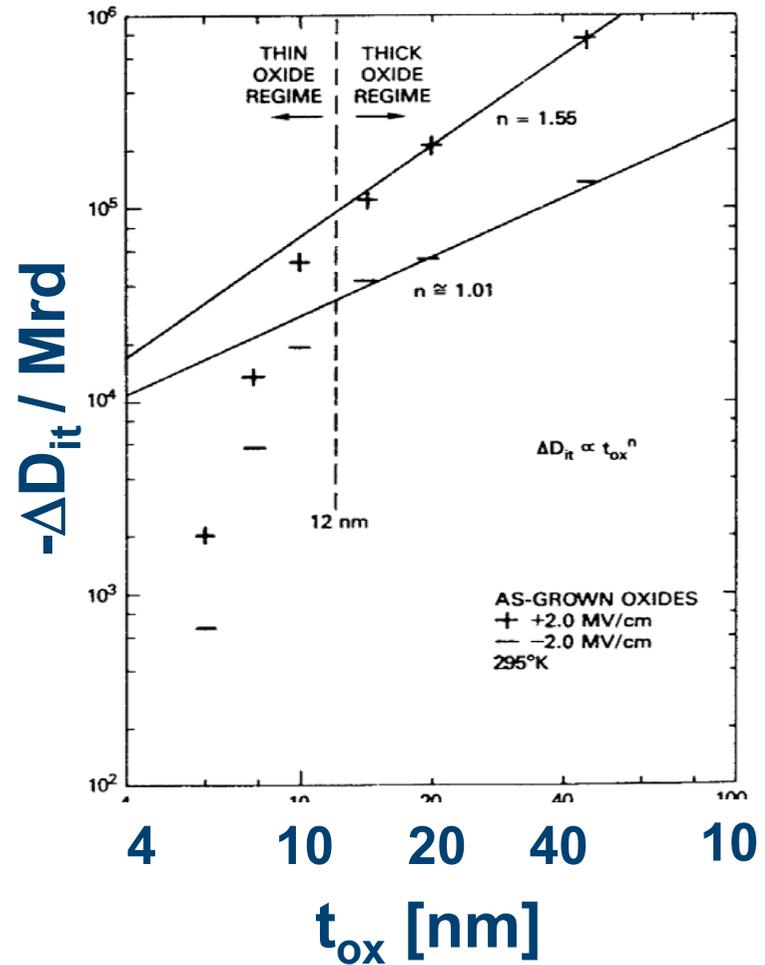
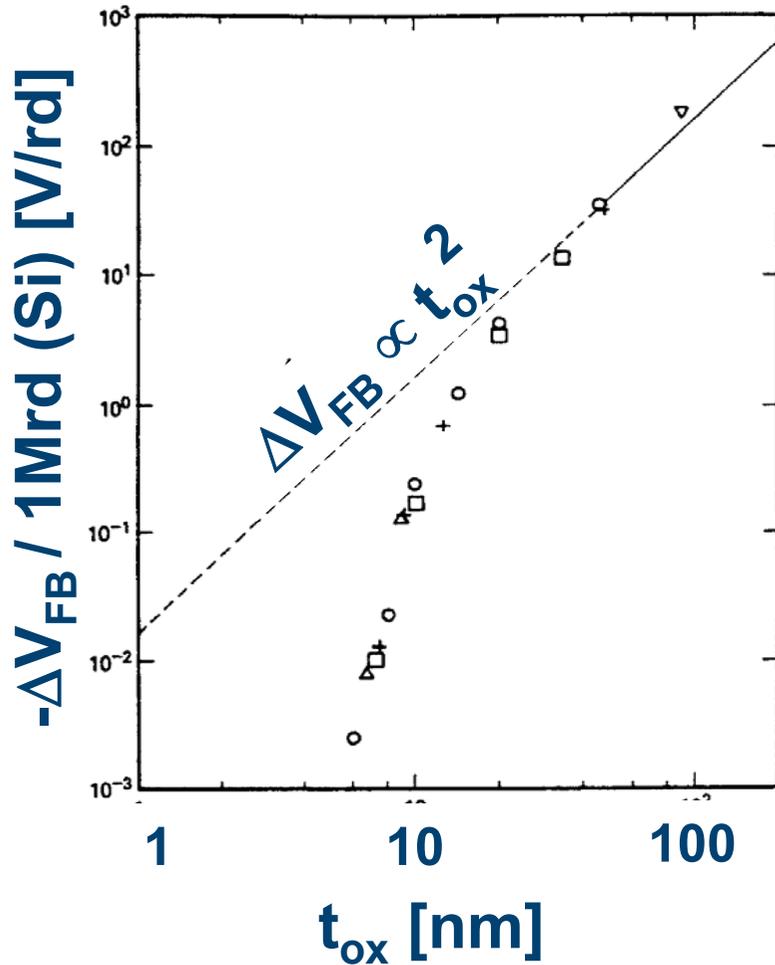
1 → 0

Never destructive, always (very) annoying.

Fighting SEU in a chip is a matter of risk management!

TID damage and CMOS scaling

MOS transistor gate oxide (thickness t_{ox})



N. S. Saks et al., *IEEE TNS*, vol. 31, no. 6, Dec. 1984, and vol. 33, no. 6, Dec. 1986.

TID damage and CMOS scaling

**Decreasing t_{ox} we
decrease the
degradation of:**

$$\Delta V_T, \frac{\Delta V_T}{V_T}, \frac{\Delta V_T}{V_{DD}}$$

$$\mu = \frac{\mu_0}{1 + \alpha \cdot (\Delta N_{it})}$$

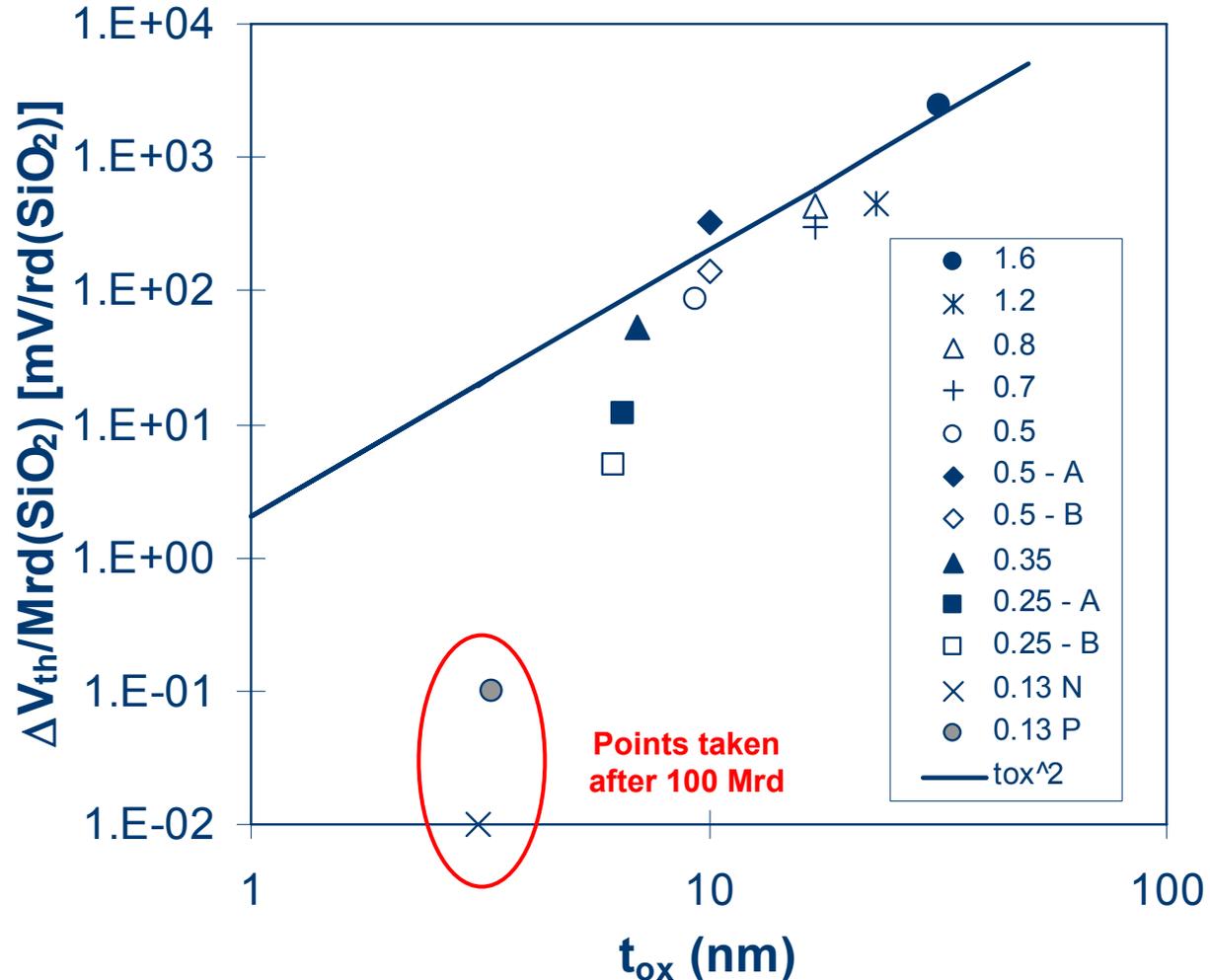
Transconductance

Subthreshold slope

Noise

ΔV_T and CMOS scaling

Measurements done in our group @ CERN: it is really true!!!

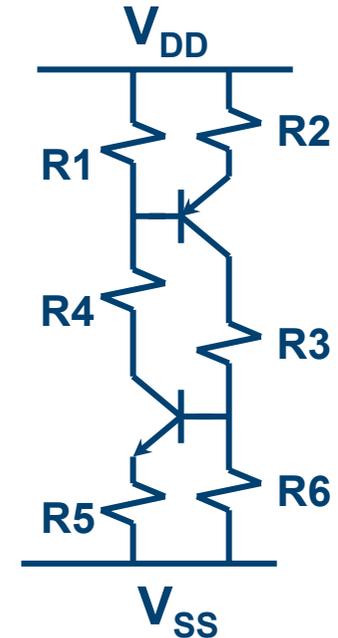


SEL and scaling

Modern CMOS technologies have:

- Retrograde wells
- Thinner epitaxial layers
- Trench isolation
- V_{DD} reduced

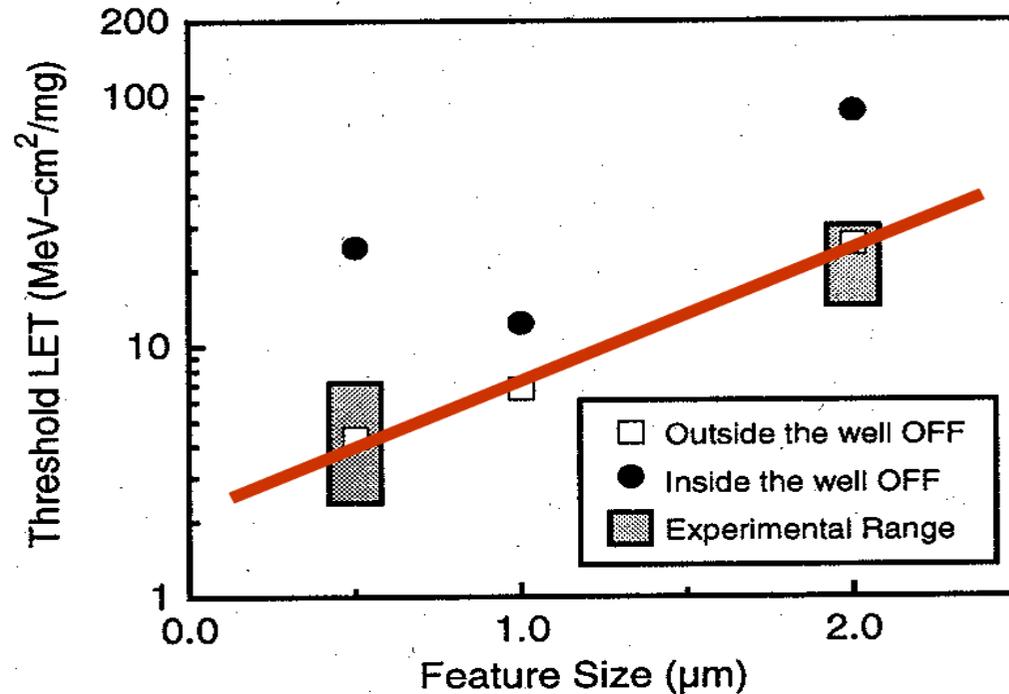
R1 and R6
reduced



**All these issues help in preventing SEL,
but they might not be always effective**

SEU and scaling

- V_{DD} reduced
 - Node C reduced
- BUT**
- Charge collected reduced

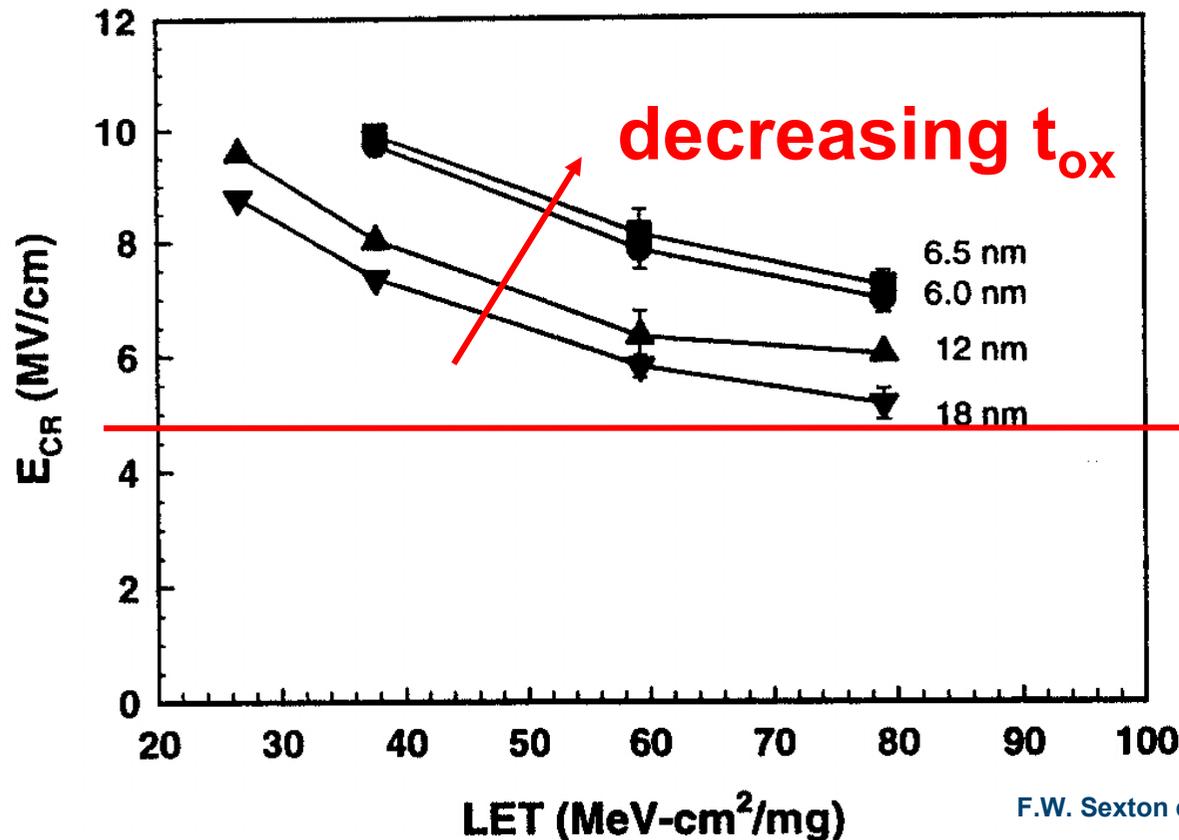


The SEU problem (may) worsen with scaling

SEGR and scaling

SEGR: Single Event Gate Rupture

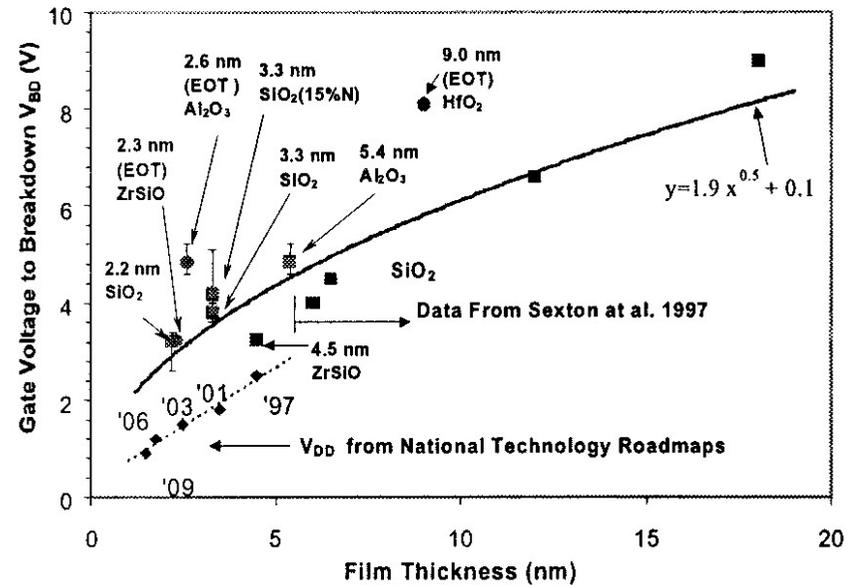
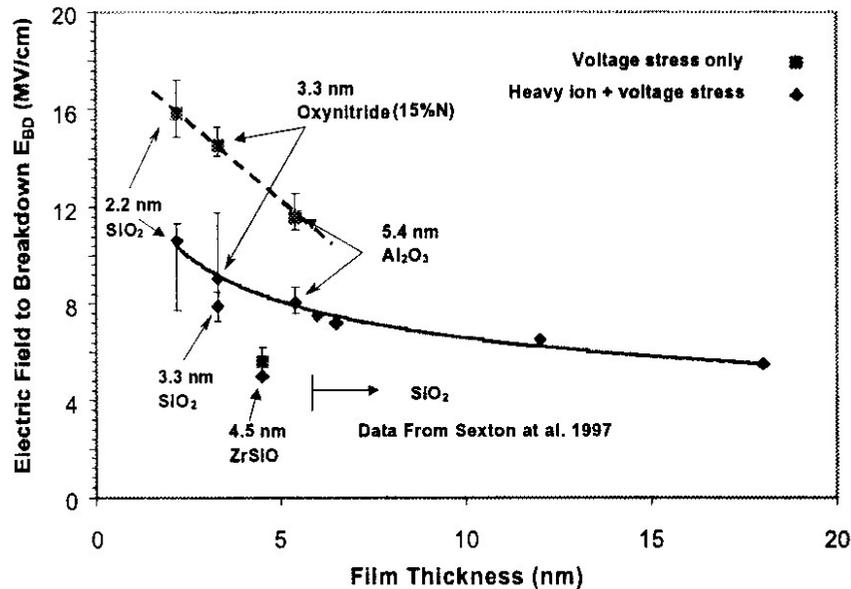
It is caused by a highly ionizing particle going through the gate of a MOS. It is a **destructive** effect (gate rupture).



Maximum electric field for a quarter micron technology

F.W. Sexton et al., *IEEE Transactions on Nuclear Science*, vol. 44, no. 6, December 1997, pp. 2345-2352.

SEGR in ULSI CMOS



SEGR is not a problem even in the most advanced CMOS processes.

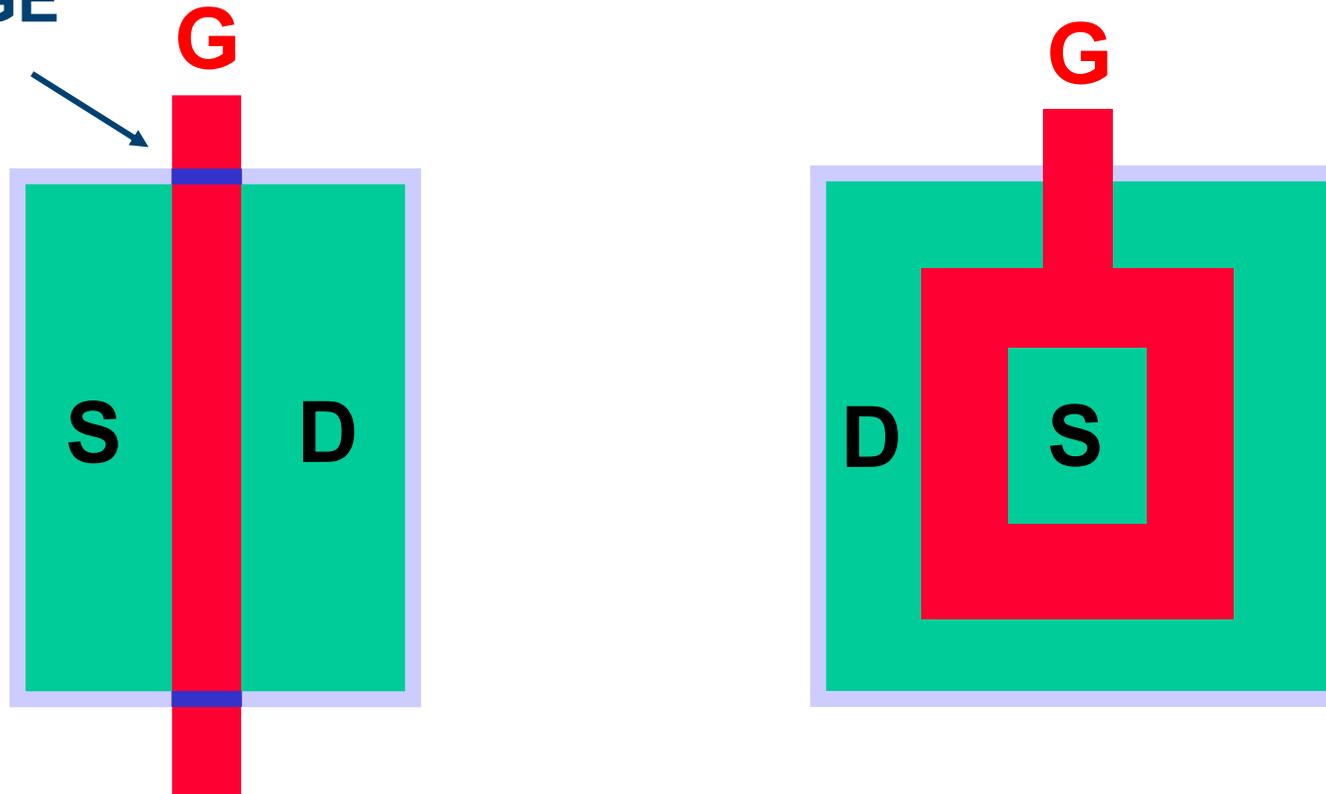
F.W. Massengill et al., *IEEE Transactions on Nuclear Science*, vol. 48, no. 6, December 2001, pp. 1904-1912.

Summary of problems in CMOS

- TID in thin oxide → Mainly V_T shift → Smaller in deep sub- μ CMOS processes.
- TID in Field Oxide → Leakage (in NMOS transistors and between transistors) → Might be better in more advanced technologies, but still a very big issue to be solved.
- SEU → Is it a Problem? This has to be evaluated and, in case it is, solved. Might become worse in deep sub- μ CMOS.
- SEL → Does not seem to be a problem, but care should be taken anyway.
- SEGR → Not a problem in deep sub- μ CMOS.

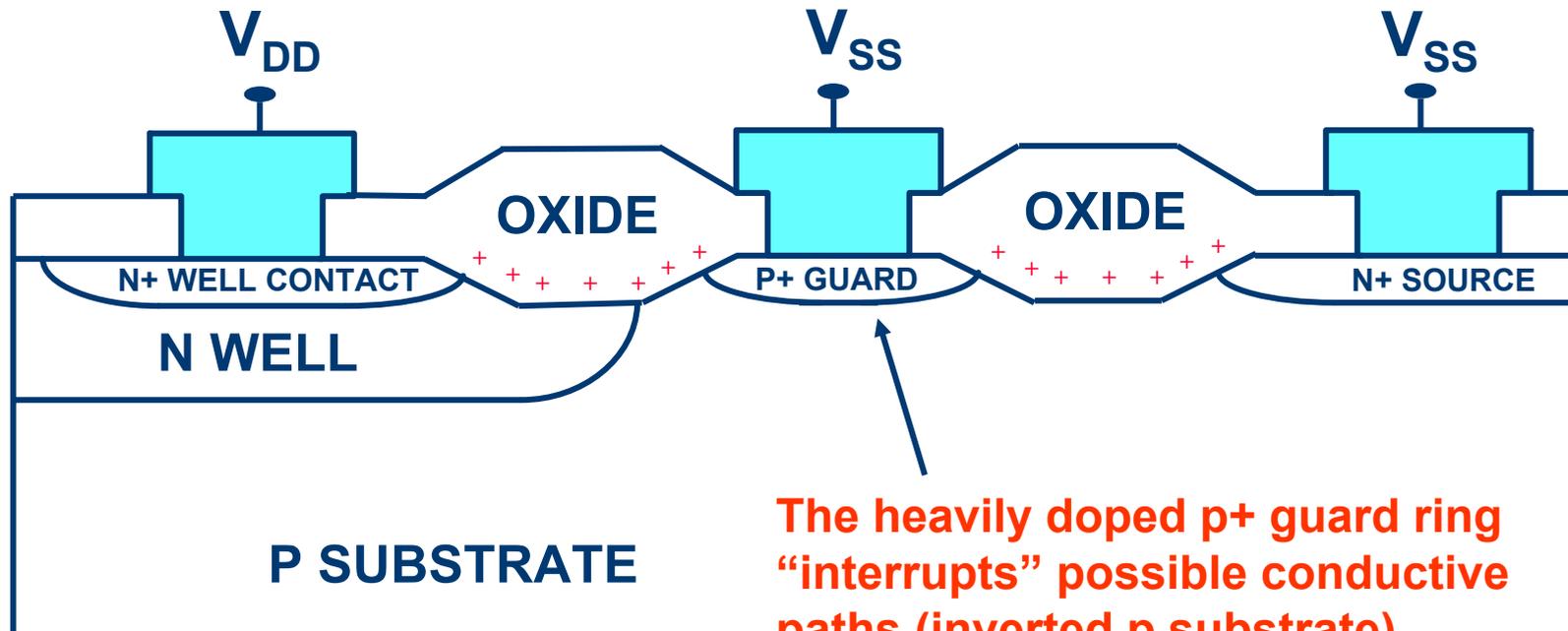
Enclosed Layout Transistor (ELT)

**LEAKAGE
PATH**



ELTs solve the leakage problem in the NMOS transistors
At the circuit level, guard rings are necessary

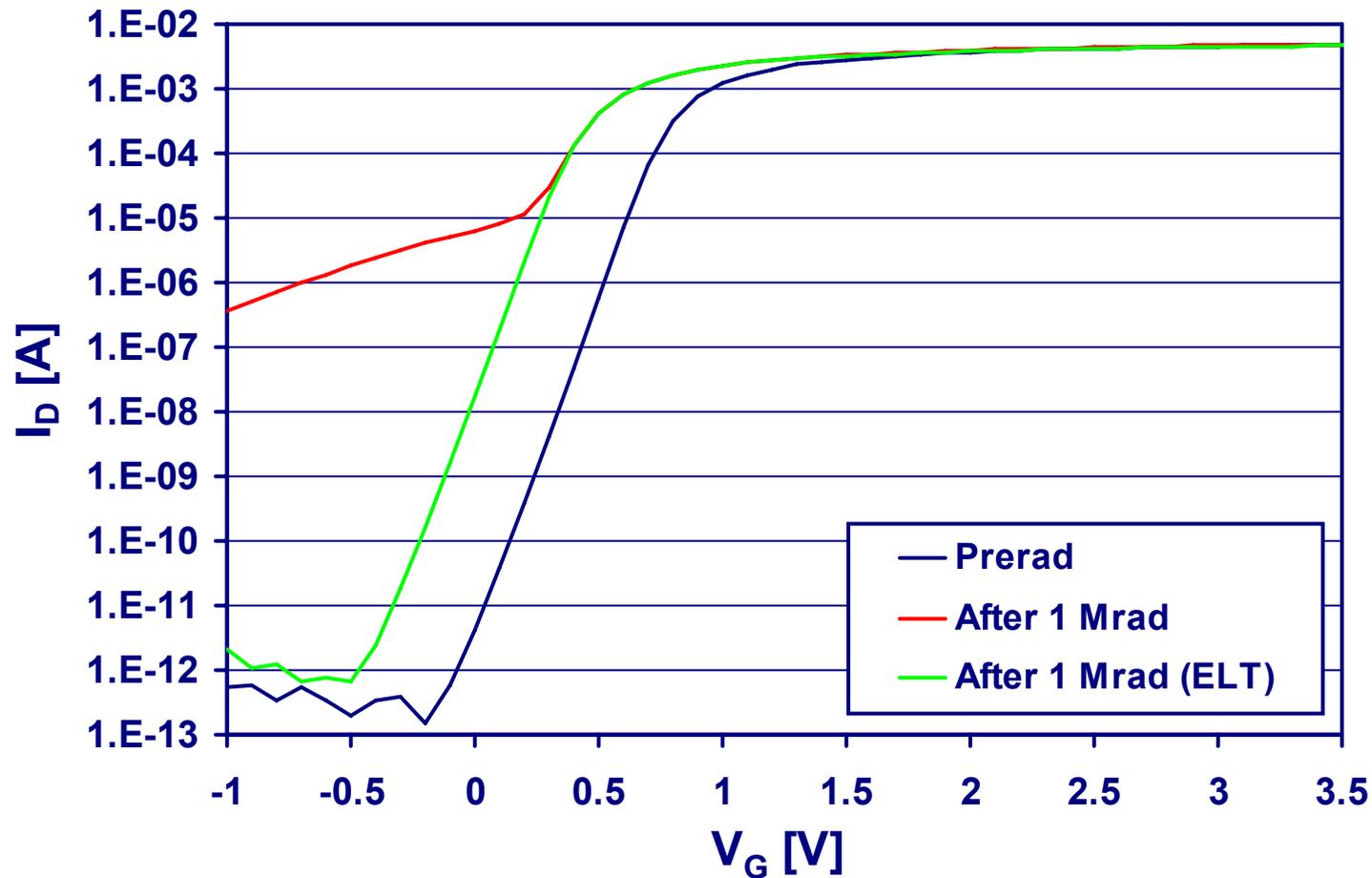
Guard rings



The heavily doped p+ guard ring “interrupts” possible conductive paths (inverted p substrate), preventing inter-device leakage currents

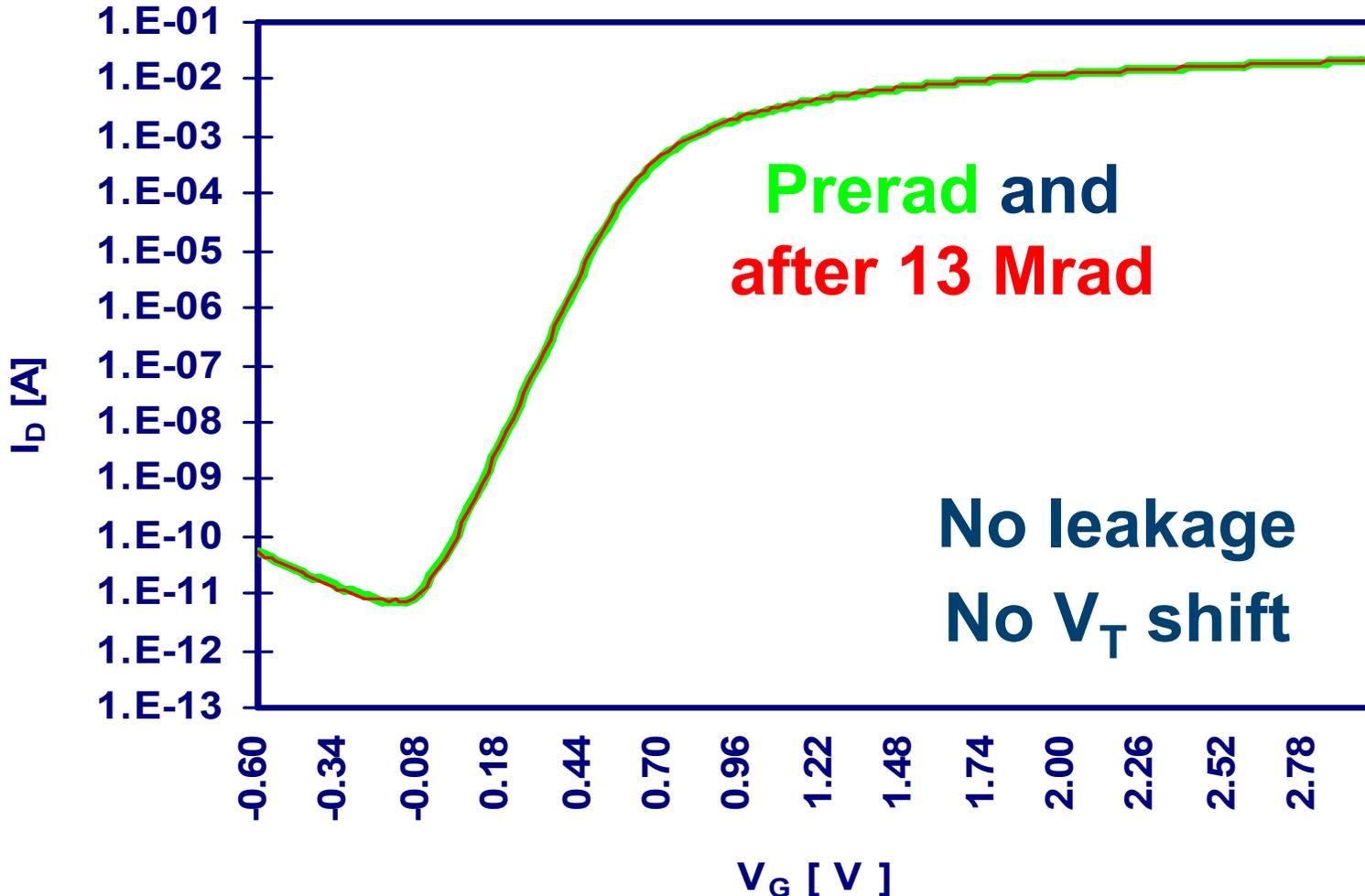
Effectiveness of ELTs

NMOS - 0.7 μm technology - $t_{\text{ox}} = 17 \text{ nm}$

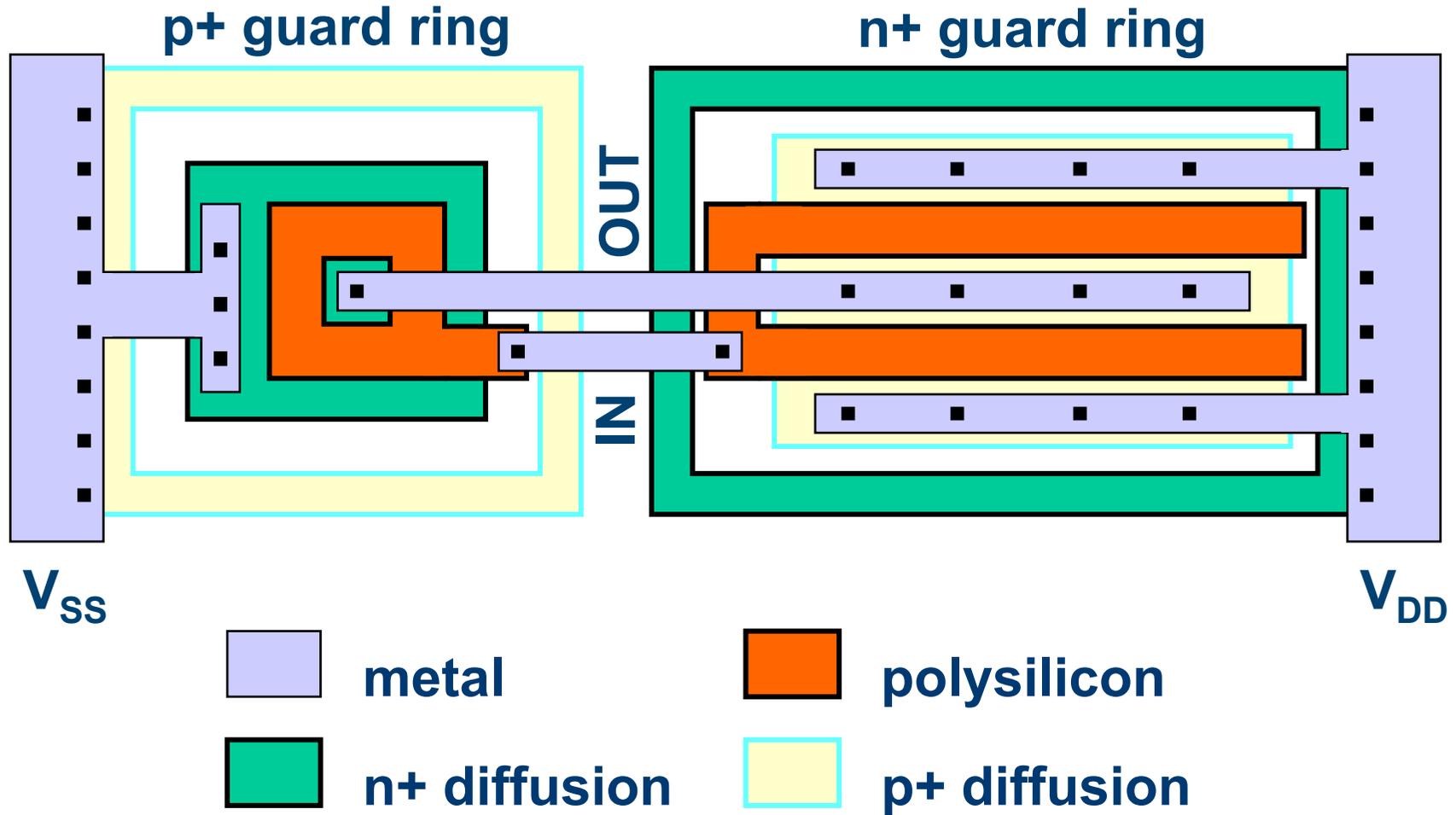


ELT & deep submicron

NMOS - 0.25 μm technology - $t_{\text{ox}} = 5 \text{ nm}$



A radiation-hard inverter



Total Ionizing Dose tolerance

$$\Delta V_{th} \propto t_{ox}^n + \text{ELT's and guard rings} = \text{TID Radiation Tolerance}$$

Deep sub- μm means also:

- speed
- low power
- VLSI
- low cost
- high yield

Single Event Effects

SEL

The systematic use of guard rings is ALSO an effective tool against SEL

SEGR

Never observed in our circuits

SEU

The higher gate capacitance of ELTs decreases the sensitivity, but other techniques might be necessary (at the circuit level)!

Reliability

Reliability: probability that a circuit will perform a required function under stated conditions for a stated period of time.

Reliability has become very important due to the aggressive scaling of device dimensions (without a corresponding scaling of the voltages).

This implied:

- **Higher electric fields**
- **Higher current densities**
- **Higher power dissipation**
- **Higher chip temperature**
- **Higher technology complexity (e.g. number of metal levels)**

<http://www.imec.be/mtc/>

G. Groeseneken et al., “Reliability, Yield and Failure Analysis”, IMEC Microelectronics Training Center Course, Leuven, Belgium, 24-26/11/03

Failure mechanisms at the chip level

- **Devices:**

- **Oxide breakdown** → due to high electric field in the oxide
- **Hot Carriers** → create interface traps and charge trapping @ the drain
- **Electro Static Discharge (ESD)** → discharge of static charge coming from human body
- **Latch-up**

- **Interconnects:**

- **Electromigration (wires and contacts)** → mass transport of metal ions by momentum exchange with conduction electrons

- **Packages:**

- **Thermomechanical fatigue**
- **Die cracking**
- **Bond aging**

Outline

- Operation and characteristics of MOS and Bipolar transistors
- Sub-micron CMOS and BiCMOS technologies
- Feature size scaling
- Radiation effects and reliability
- **Mixed-signal circuits**
 - Analog design in digital CMOS processes
 - Digital noise in mixed-signal circuits



Analog design in digital processes

The integrated circuit market is driven by digital circuits, such as memories and microprocessors. This led to an increasing interest in integrating analog circuits together with digital functions in processes optimized for digital circuits, making what it is called a System on a Chip (SoC). This approach has several **advantages** and **disadvantages**.

ADVANTAGES:

- Lower wafer cost
- Higher yield
- Higher speed
- Lower power consumption (not always)
- Complex digital functions on chip (DSP)

DISADVANTAGES:

- Low power supplies
- Lack of “analog” components
- Inadequate modeling
 - Output conductance
 - Different inversion regions
- “Digital” noise

- E. A. Vittoz, "The Design of High-Performance Analog Circuits on Digital CMOS Chips", *IEEE JSSC*, vol. 20, no. 3, June 1985, pp. 657-665.
- W. Sansen, "Challenges in Analog IC Design in Submicron CMOS Technologies", *Proceedings of the 1996 IEEE-CAS Region 8 Workshop on Analog and Mixed IC Design*, Pavia, Italy, 13-14 September 1996, pp. 72-78.
- C. Azaredo Leme and J. E. Franca, "Analog-Digital Design in Submicrometric Digital CMOS Technologies", *Proceedings of the 1997 IEEE International Symposium on Circuits and Systems*, Hong Kong, 9-12 June 1997, vol. 1, pp. 453-456.

Analog components availability

Analog design needs high-quality passive components. These are not always present in processes optimized for digital design, or at least not in the first stages of the process development. These analog “options” are:

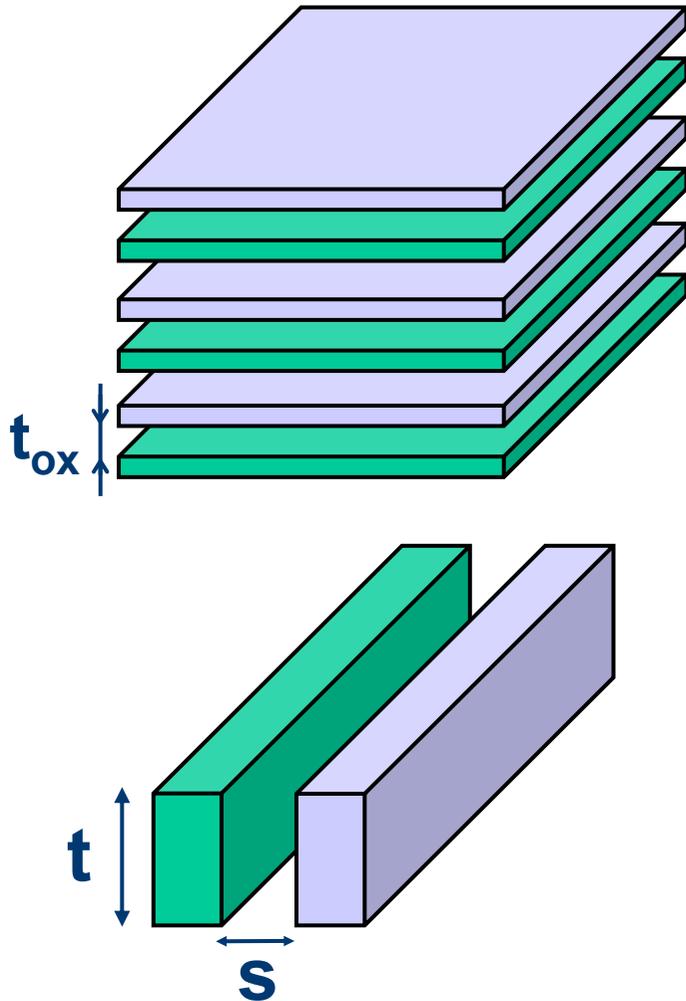
- **High-resistivity polysilicon for resistors**
- **Diffusion resistors**
- **Trimming options**
- **Linear and dense capacitors**
 - **Metal to metal (at least one special metal layer required)**
 - **Metal to poly**
 - **Poly to poly**

Integrated capacitors

- **High-linearity capacitors can be obtained with metal-to-metal structures. This requires adding a special metal layer to the technology, in order to reduce the dielectric thickness between the metal plates. The density reached is generally around a few fF/ μm^2 .**
- **Dense capacitors can be obtained exploiting the high capacitance density of the thin gate oxide. This allows having dense and precise capacitors, good matching but very poor linearity. This solution can be adopted in any process.**
- **A third possible solution is suggested by the availability of many interconnection layers. Exploiting the parasitic capacitance between metal wires in a clever way, one can obtain linear capacitors with good matching and linearity and densities up to 1.5 fF/ μm^2 . These capacitors can be integrated in any process!!**

- A. T. Behr et al., "Harmonic Distortion Caused by Capacitors Implemented with MOSFET Gates", *IEEE JSSC*, vol. 27, no. 10, Oct. 1992, pp. 1470-1475.
- D. B. Slater, Jr. and J. J. Paulos, "Low-Voltage Coefficient Capacitors for VLSI Processes", *IEEE JSSC*, vol. 24, no. 1, February 1989, pp. 165-173.
- J. L. McCreary, "Matching Properties, and Voltage and Temperature Dependence of MOS Capacitors", *IEEE JSSC*, vol. 16, no. 6, Dec. 1981, pp. 608.
- S. Pavan, Y. Tsvividis and K. Nagaraj, "Modeling of accumulation MOS capacitors for analog design in digital VLSI processes", *Proceedings of the 1999 IEEE International Symposium on Circuits and Systems*, Orlando, Florida, USA, 30 May – 2 June 1999, vol. 6, pp. 202-205.

Multi-metal-layer capacitors



This solution is a possibility, but it does not exploit the fact that in deep submicron processes the highest parasitic capacitance can be obtained “horizontally” rather than vertically, i.e. $t_{ox} > s$

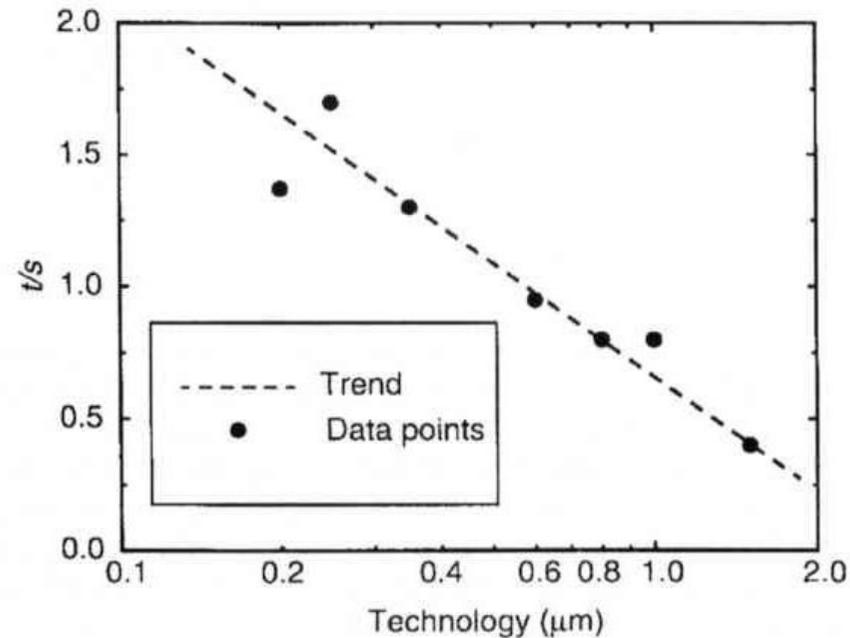
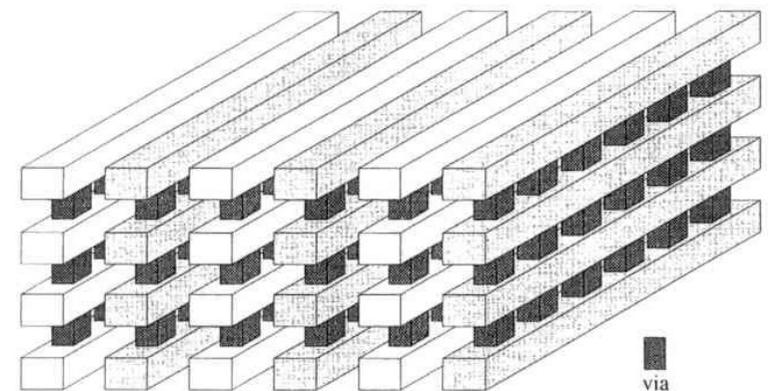
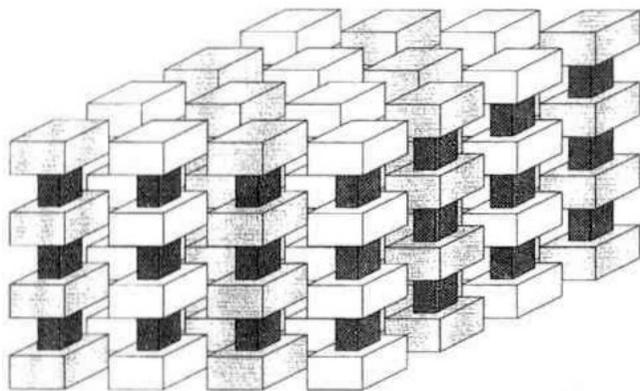
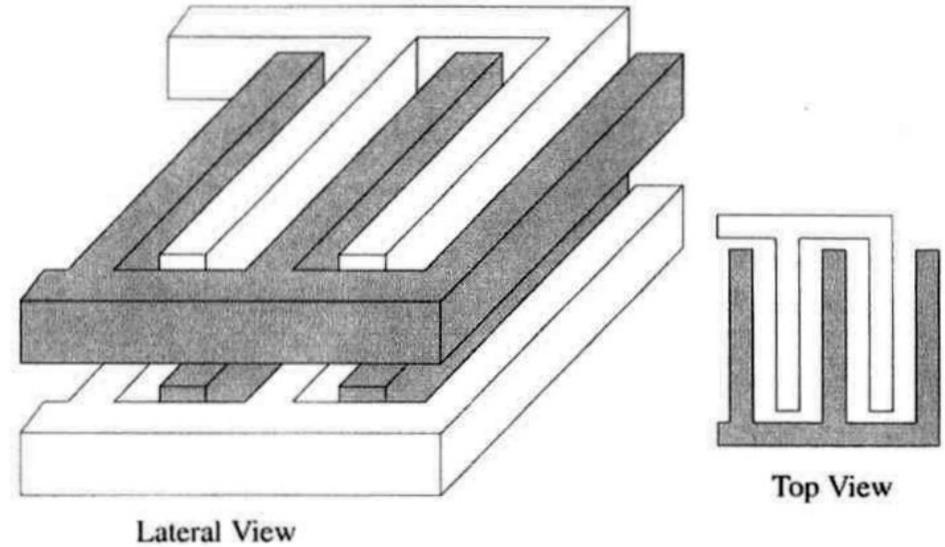
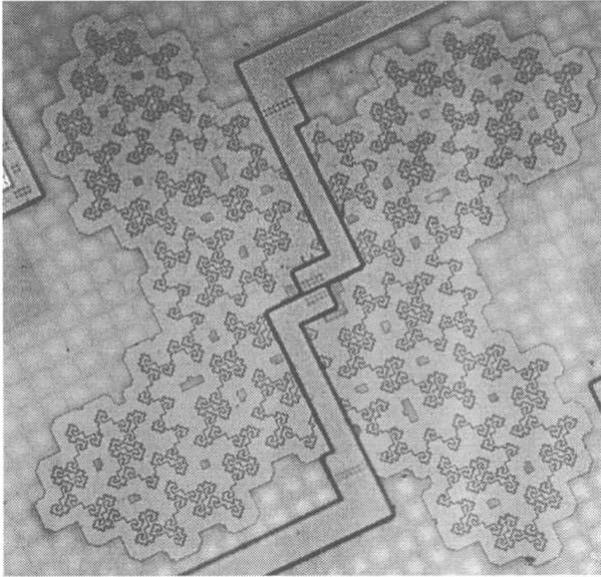


Fig. 3. Ratio of metal thickness to horizontal metal spacing versus technology (channel length).

Multi-metal-layer capacitors

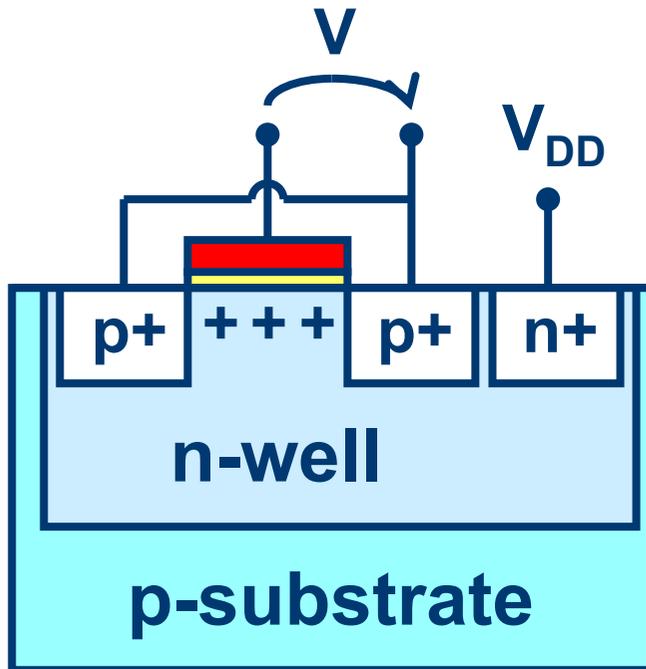


- Hiran Samavati et al., "Fractal Capacitors", *IEEE Journal of Solid-State Circuits*, vol. 33, no. 12, December 1998, pp. 2035-2041.
- R. Aparicio and A. Hajimiri, "Capacity Limits and Matching Properties of Integrated Capacitors", *IEEE JSSC*, vol. 37, no. 3, March 2002, pp. 384-393.

MOS capacitors

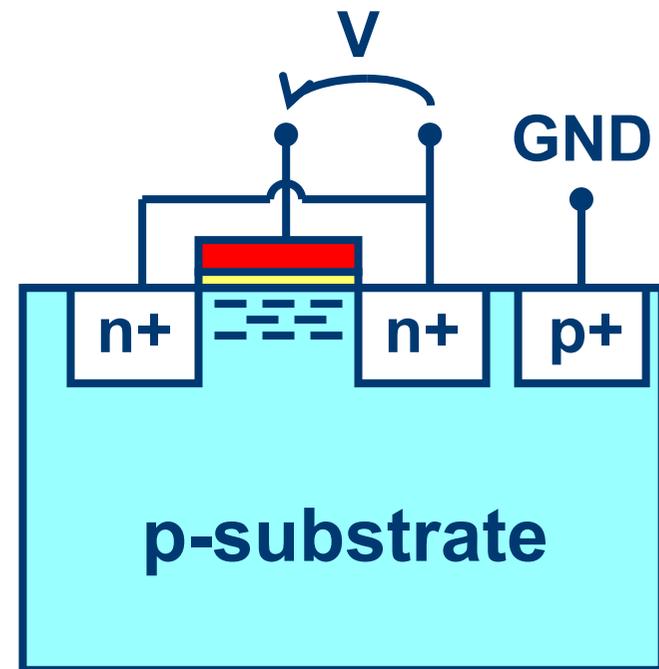
PMOS

Inversion Region



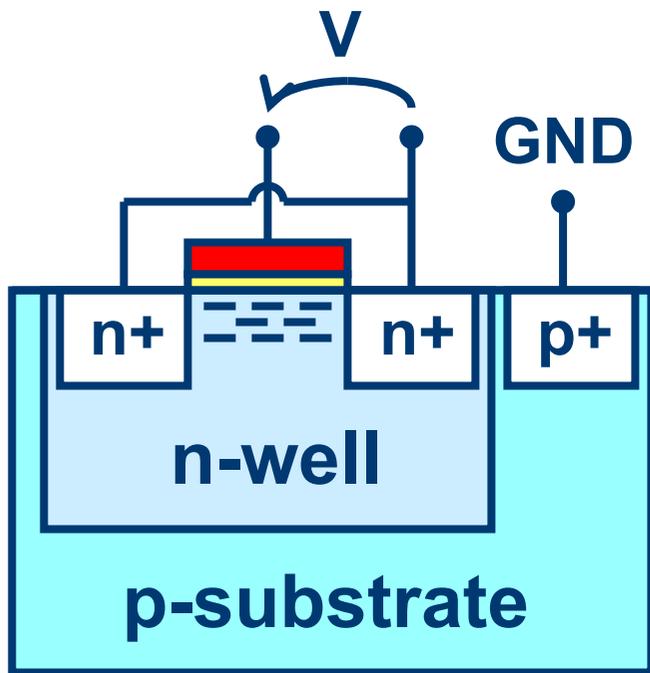
NMOS

Inversion Region

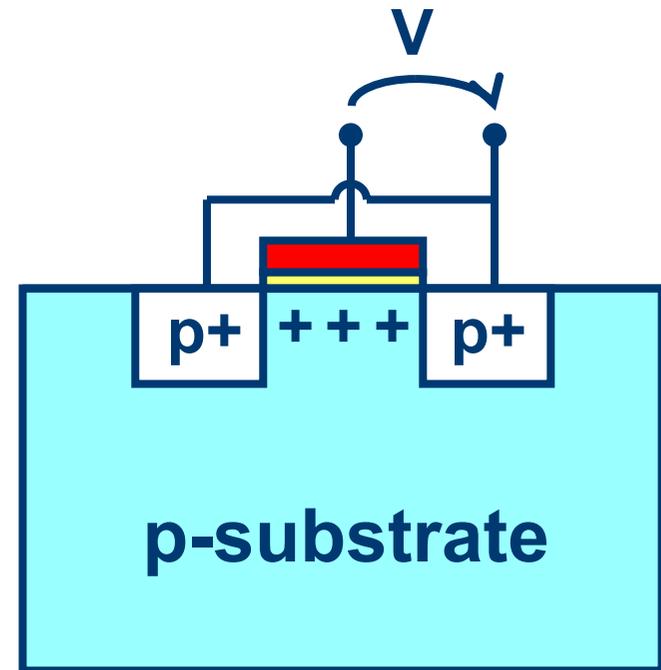


MOS capacitors

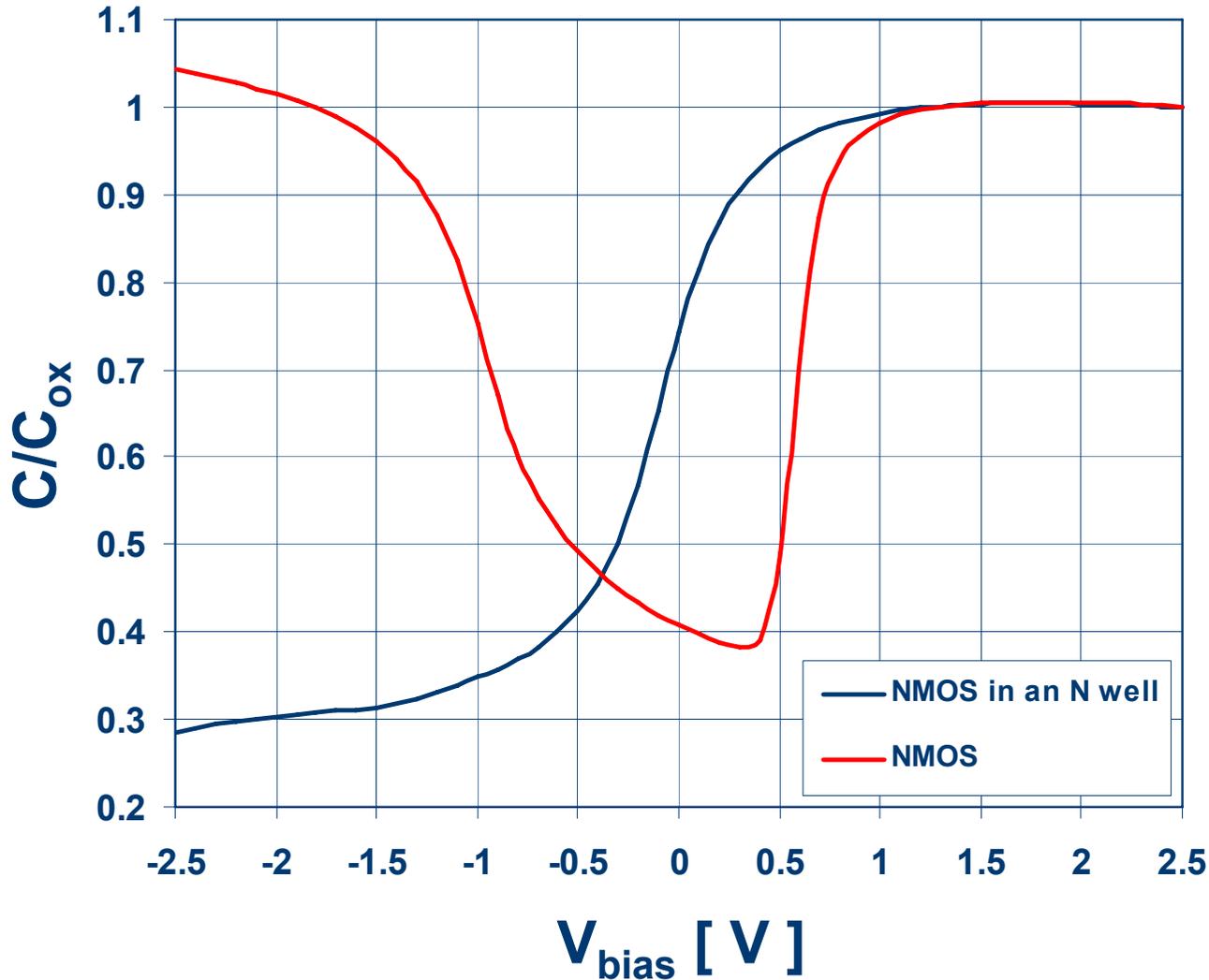
NMOS in an N well Accumulation Region



MOS structure Accumulation Region



C-V characteristics



The NMOS in an N well capacitor is in accumulation for $V > 0$ V.

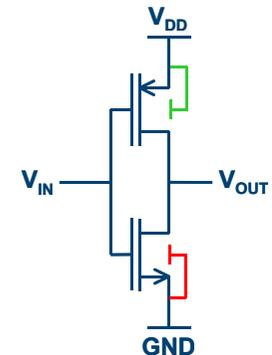
The NMOS capacitor is in inversion for $V > 0$.

Digital noise in mixed-signal ICs

Integrating analog blocks on the same chip with digital circuits can have some serious implications on the overall performance of the circuit, due to the influence of the “noisy” digital part on the “sensitive” analog part of the chip.

The switching noise originated from the digital circuits can be coupled in the analog part through:

- The power and ground lines
- The parasitic capacitances between interconnection lines
- The common substrate



The substrate noise problem is the most difficult to solve.

- A. Samavedam et al., "A Scalable Substrate Noise Coupling Model for Design of Mixed-Signal IC's", *IEEE JSSC*, vol. 35, no. 6, June 2000, pp. 895-904.
- N. K. Verghese and D. J. Allstot, "Computer-Aided Design Considerations for Mixed-Signal Coupling in RF Integrated Circuits", *IEEE JSSC*, vol. 33, no. 3, March 1998, pp. 314-323.
- M. Ingels and M. S. J. Steyaert, "Design Strategies and Decoupling Techniques for Reducing the Effects of Electrical Interference in Mixed-Mode IC's", *IEEE Journal of Solid-State Circuits*, vol. 32, no. 7, July 1997, pp. 1136-1141.

Noise reduction techniques

- Quiet the Talker. Examples (if at all possible !!!):
 - Avoid switching large transient supply current
 - Reduce chip I/O driver generated noise
 - Maximize number of chip power pads and use on-chip decoupling
- Isolate the Listener. Examples:
 - Use on-chip shielding
 - Separate chip power connections for noisy and sensitive circuits
 - Other techniques depend on the type of substrate. See next slide
- Close the Listener's ears. Examples:
 - Design for high CMRR and PSRR
 - Use minimum required bandwidth
 - Use differential circuit architectures
 - Pay a lot of attention to the layout

Different types of substrates

There are mainly two types of wafers:

1. **Lightly doped wafers: “high” resistivity, in the order of 10 Ω -cm.**
2. **Heavily doped wafers: usually made up by a “low” resistivity bulk (~ 10 m Ω /cm) with a “high” resistivity epitaxial layer on top.**

TSMC, UMC, IBM and STM (below 180 nm) offer type 1

Substrate noise: how to reduce it

To minimize the impact of disturbances coming from the substrate on the sensitive analog blocks, we have mainly three ways:

- **Separate the “noisy” blocks from the “quiet” blocks. This is effective especially in uniform lightly doped substrates. For heavily doped substrates, it is useless to use a separation greater than about 4 times the epitaxial layer thickness.**
- **In n-well processes, p⁺ guard rings can be used around the different blocks. Unfortunately, this is again effective mainly for lightly doped substrates. Guard rings (both analog and digital) should be biased with separate pins.**
- **The most effective way to reduce substrate noise is to ground the substrate itself in the most “solid” possible way (no inductance between the substrate and ground). This can be done using many ground pins to reduce the inductance, or, even better, having a good contact on the back of the chip (metallization) and gluing the chip with a conductive glue on a solid ground plane.**
- **Separate the ground contact from the substrate contact in the digital logic cells, to avoid to inject the digital switching current directly into the substrate.**

D. K. Su, M. J. Loinaz, S. Masui and B. A. Wooley, "Experimental Results and Modeling Techniques for Substrate Noise in Mixed-Signal Integrated Circuits", *IEEE Journal of Solid-State Circuits*, vol. 28, no. 4, April 1993, pp. 420-429.