SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Term 7- Sept 2024

Nanoelectronics and Technology (01.119/99.503)

Week 2 Day 2 (26-Sept 2024)

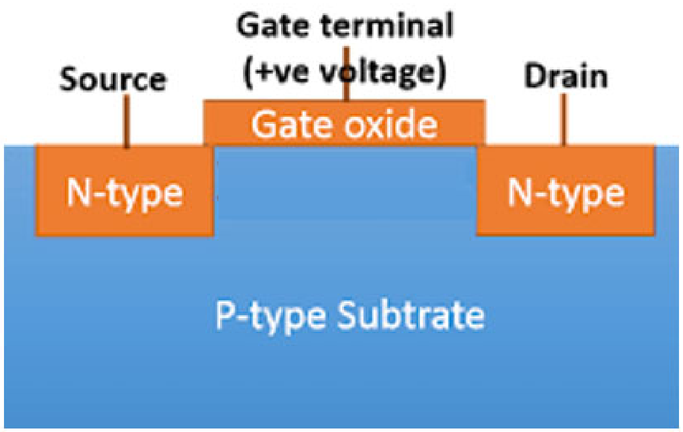
Outline

- CMOS scaling
- Impact of Scaling on different CMOS parameters
- Scaling Challenges

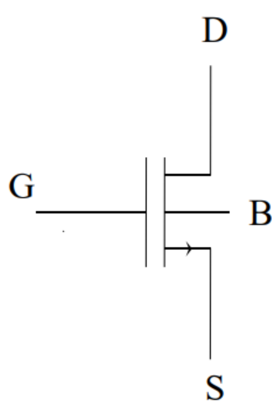
SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

Review of MOSFETs

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



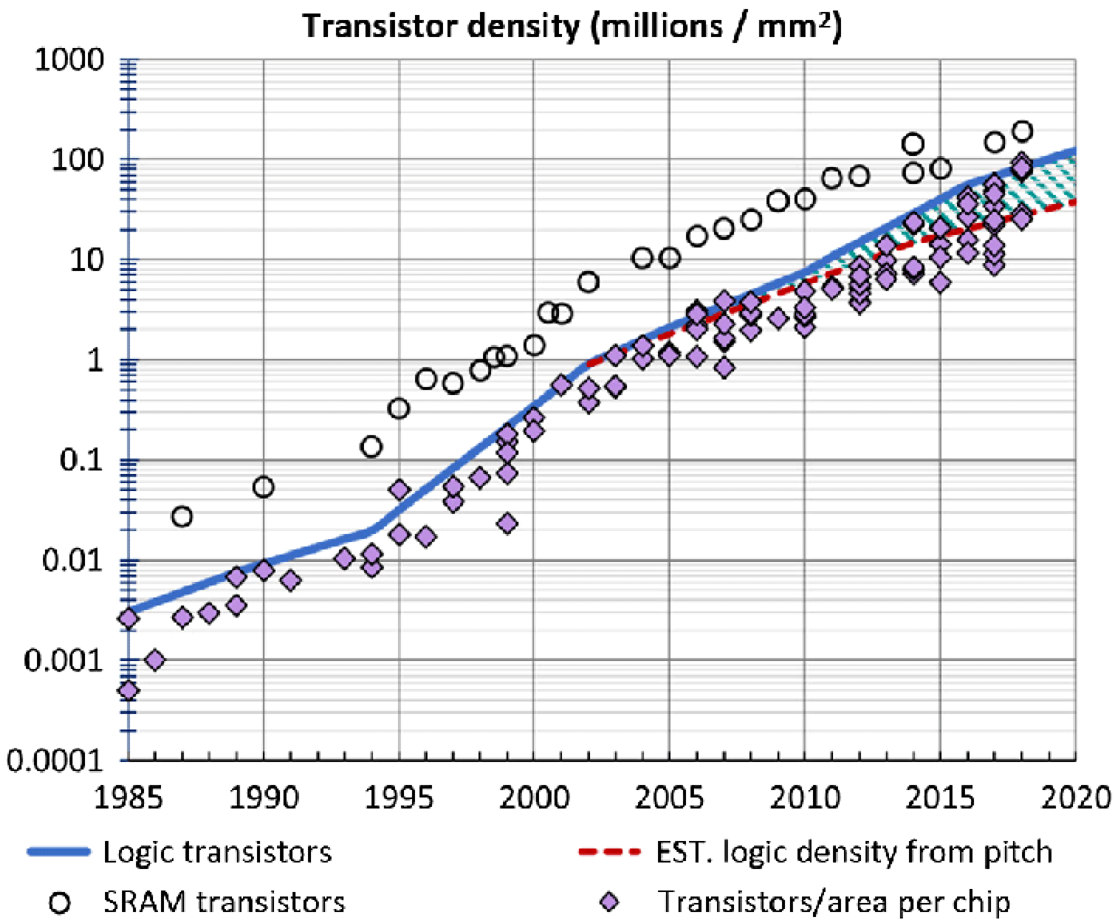
n-channel MOS
Transistor



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

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

CMOS scaling: Moore's Law

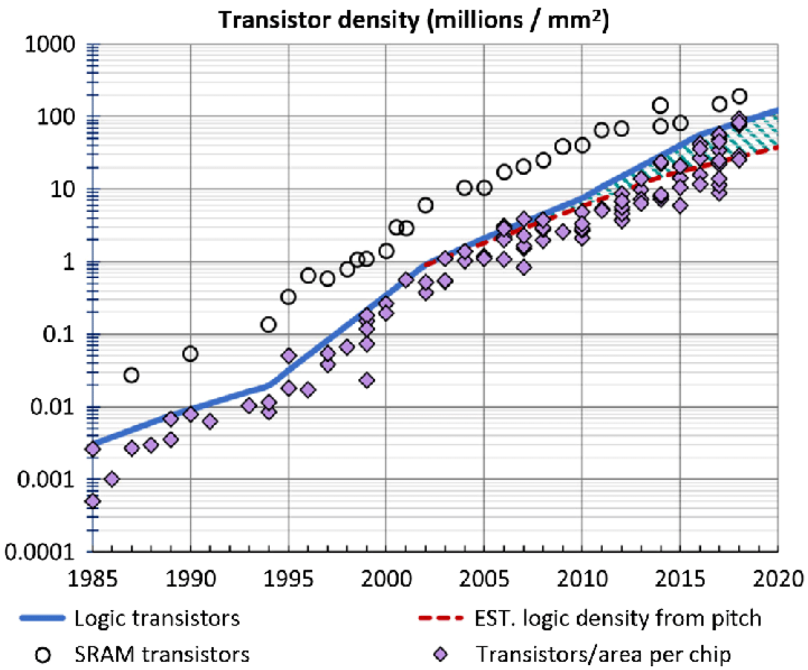


Moore's law says that the number of transistors doubles approximately every two years.

- CMOS scaling
  - Speed
  - High density
  - Less power
  - Reduced cost/transistor

Ref: M. L. Rieger, " Retrospective on VLSI value scaling and lithography" Journal of Micro/Nanolithography-2019

CMOS scaling: Dennard's scaling

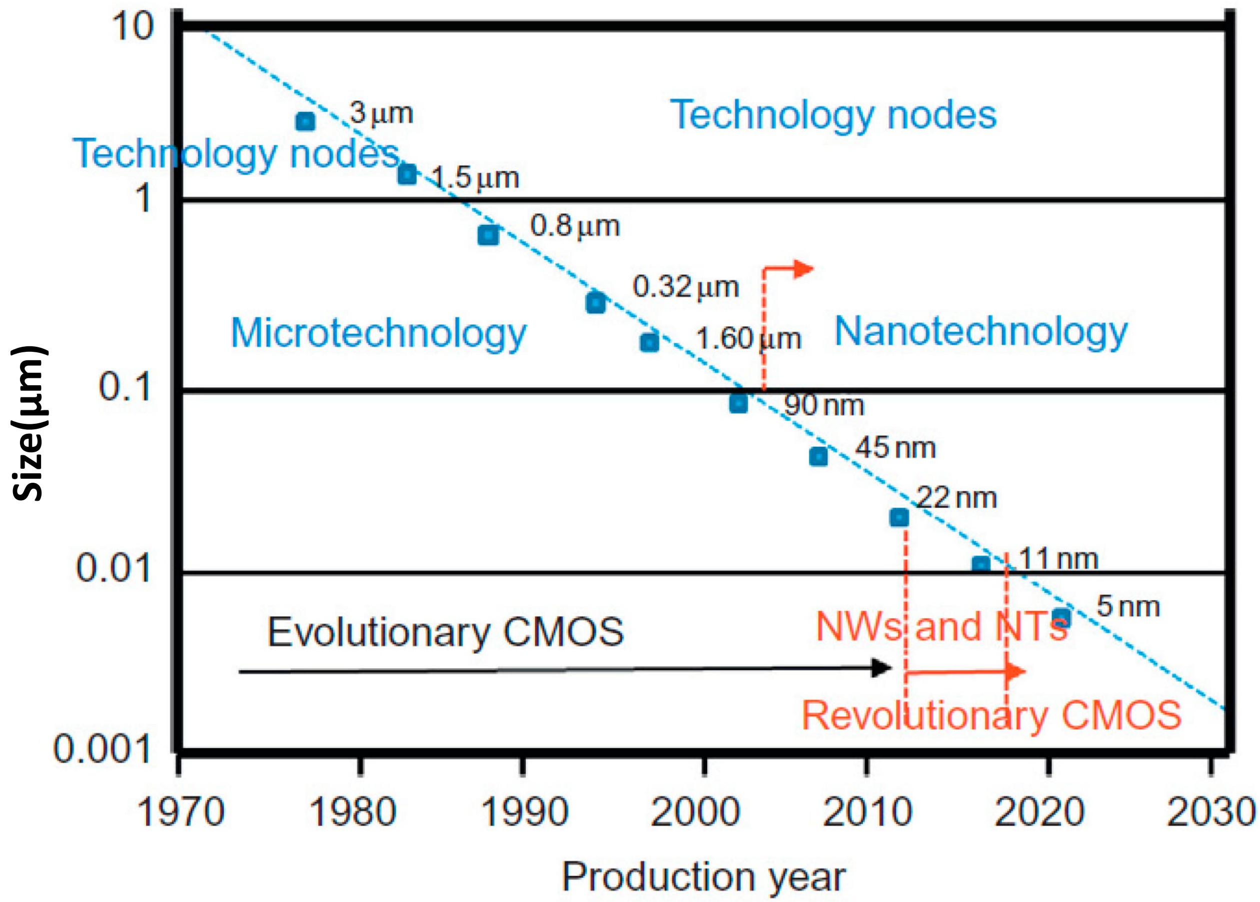


Dennard scaling, also known as MOSFET scaling, is a scaling law which states roughly that, as transistors get smaller, their power density stays constant, so that the power use stays in proportion with area; both voltage and current scale (downward) with length.

With feature sizes below 65nm, these rules could no longer be sustained, because of the exponential growth of the leakage current."

Ref: M. L. Rieger, " Retrospective on VLSI value scaling and lithography" Journal of Micro/Nanolithography-2019

CMOS scaling: Moore's Law



Ref: H. H. Radamson, et al., " Miniaturization of CMOS" Micromachines 10(5) 293-2019

CMOS scaling: ITRS Roadmap

ITRS Projections

Year of Production	2007	2010	2013
Technology Node (nm)	65	45	32
Transistor Gate Length in Microprocessors circuits (nm)	25	18	13
Wafer diameter (inch)	12	12	18
Number of masks required for fabrication of Microprocessor	33	35	37
Number of Transistors in Microprocessor (billion)	1.1	2.2	4.4
Number of interconnect wiring levels in the Microprocessor	15	16	17

http://www.itrs2.net/

CMOS scaling: ITRS Roadmap

ITRS Projections

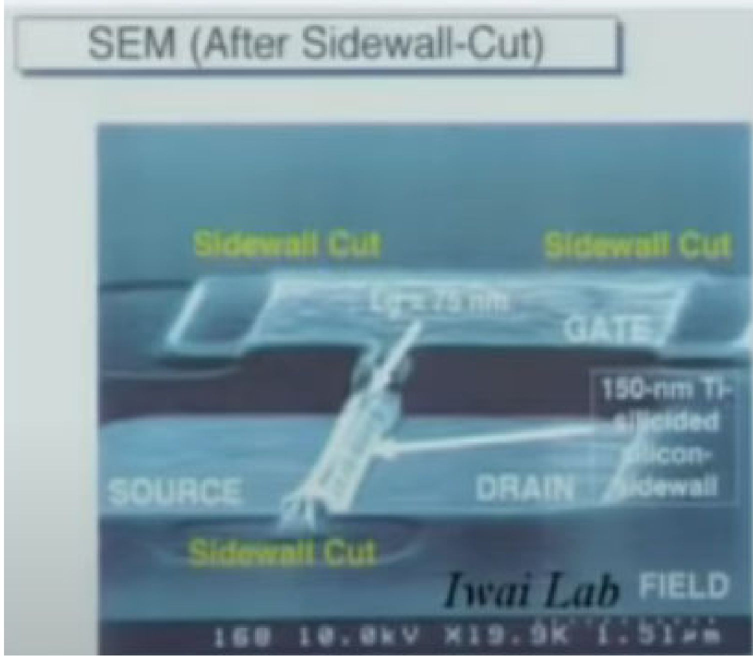
Year of Production	2007	2010	2013	2016	2019	2022
Technology Node (nm)	65	45	32	22	16	11
Transistor Gate Length in Microprocessors circuits (nm)	25	18	13	9	6.3	4.5
Wafer diameter (inch)	12	12	18	18	18	18

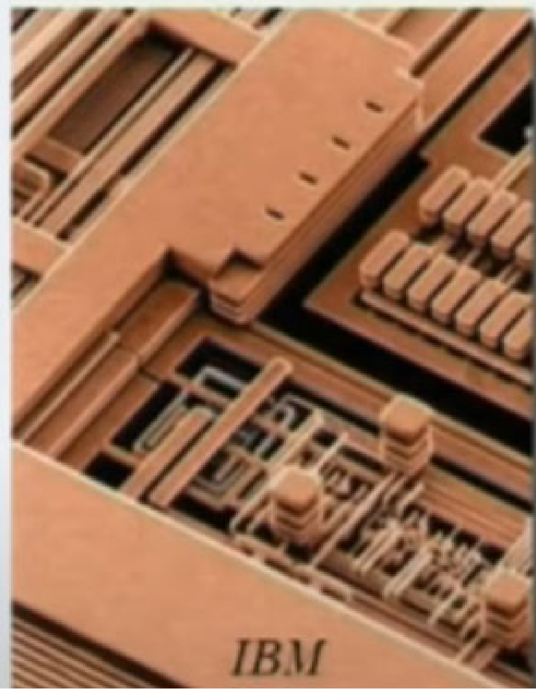
	2021	2022	2025	2028	2031	2034
	G51M30	G48M24	G45M20	G42M16	G40M16/T2	G38M16/T4
Logic industry "Node Range" Labeling (nm)	"5"	"3"	"2.1"	"1.5"	"1.0 eq"	"0.7 eq"
IDM-Foundry node labeling	i7-f5	i5-f3	i3-f2.1	i2.1-f1.5	i1.5e-f1.0e	i1.0e-f0.7e
Logic device structure options	FinFET	finFET LGAA	LGAA	LGAA	LGAA-3D	LGAA-3D
Platform device for logic	finFET	finFET	LGAA	LGAA	LGAA-3D	LGAA-3D
Frequency scaling - node-to-node	-	0.02	0.16	0.09	-0.08	-0.01
CPU frequency at constant power density (GHz)	3.13	2.83	3.53	2.50	1.48	0.86
Power at iso frequency - node-to-node	-	-0.16	-0.27	-0.05	-0.06	-0.08
Power density - relative	1.00	1.12	1.04	1.59	2.51	4.27
LOGIC TECHNOLOGY ANCHORS						
Patterning technology inflection for Mx interconnect	193i, EUV DP	193i, EUV DP	193i, EUV DP	193i, High-NA EUV	193i, High-NA EUV	193i, High-NA EUV
Beyond-CMOS as complimentary to platform CMOS	-	-	-	2D Device, FeFET	2D Device, FeFET	2D Device, FeFET
Channel material technology inflection	SiGe25%	SiGe50%	SiGe50%	Ge, 2D Mat	Ge, 2D Mat	Ge, 2D Mat
Process technology inflection	Conformal Doping, Contact	Channel, RMG	Lateral/AtomicEtch	Non-Cu Mx	3DVLSI	3DVLSI
Stacking generation inflection	2D	3D-stacking: W2W, D2W Mem-on-Logic	3D-stacking: W2W, D2W Mem-on-Logic	3D-stacking, Fine-pitch stacking, P-over-N, Mem-on-Logic	3D-stacking, 3DVLSI: Mem-on-Logic with Interconnect	3D-stacking, 3DVLSI: Logic-on-Logic

http://www.itrs2.net/

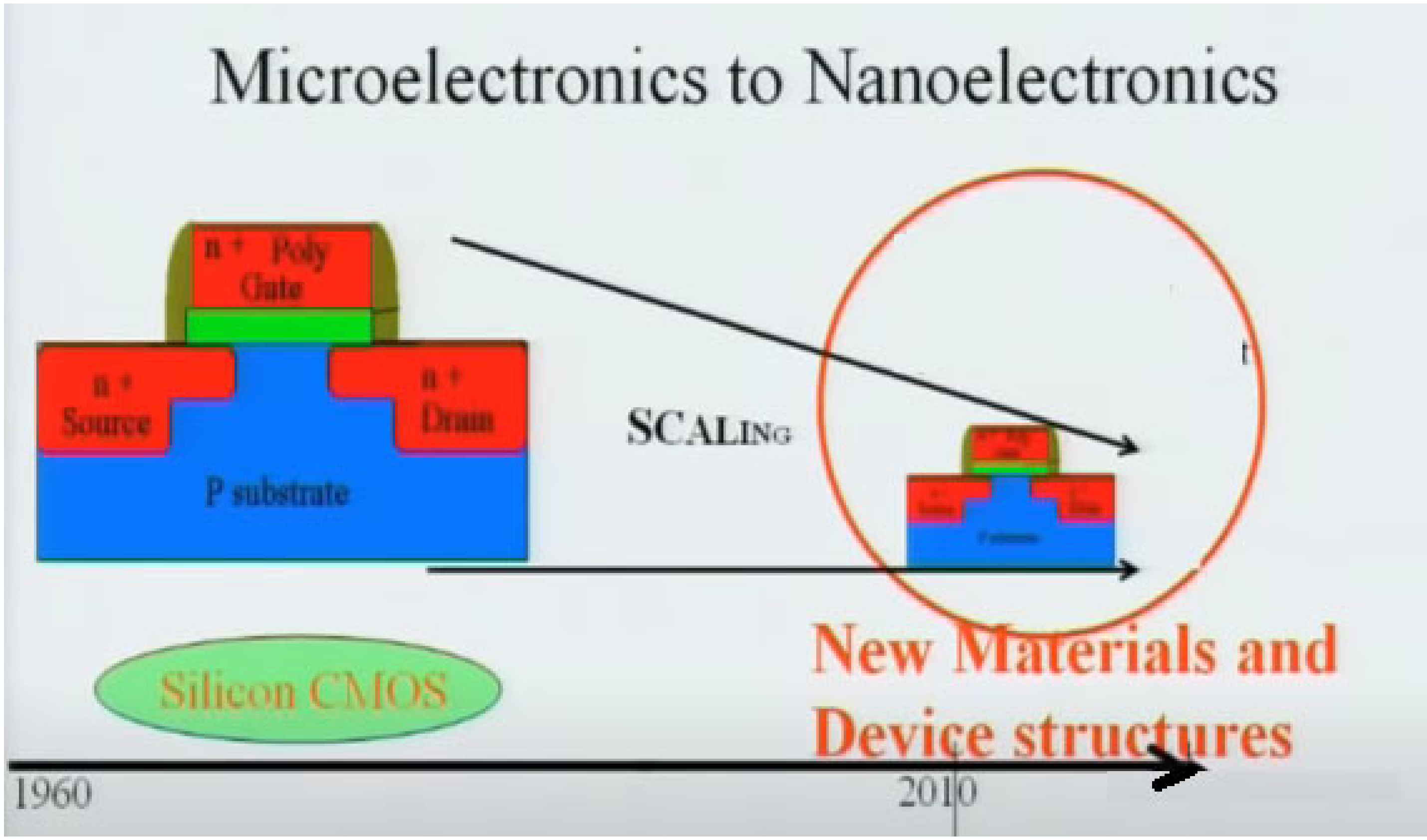
CMOS chip

A large ensemble of **transistors** connected *"appropriately"* through multilevel **metal interconnection** lines…





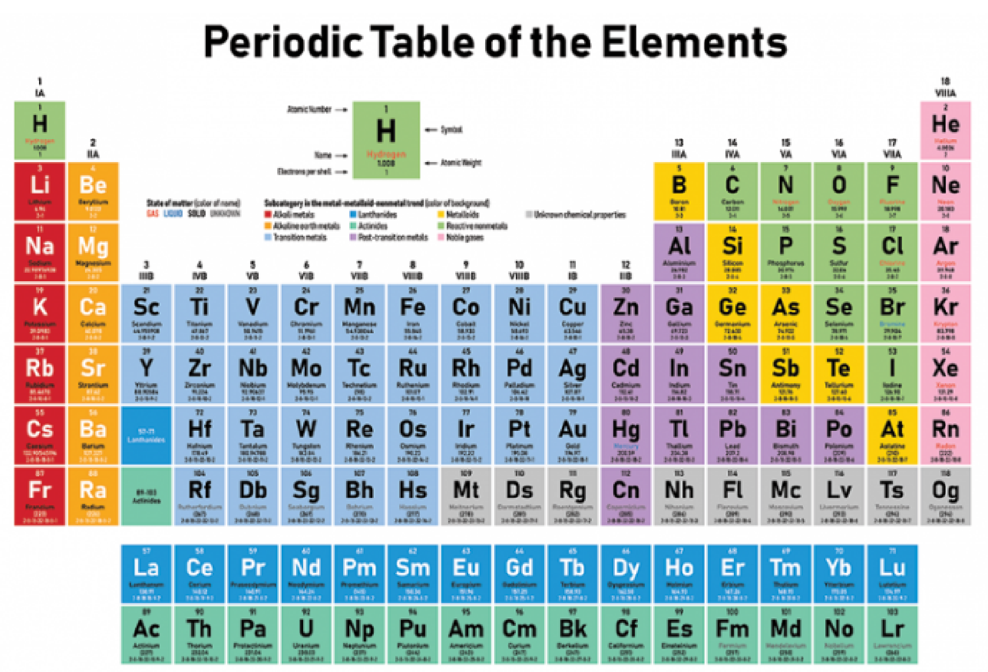
CMOS scaling



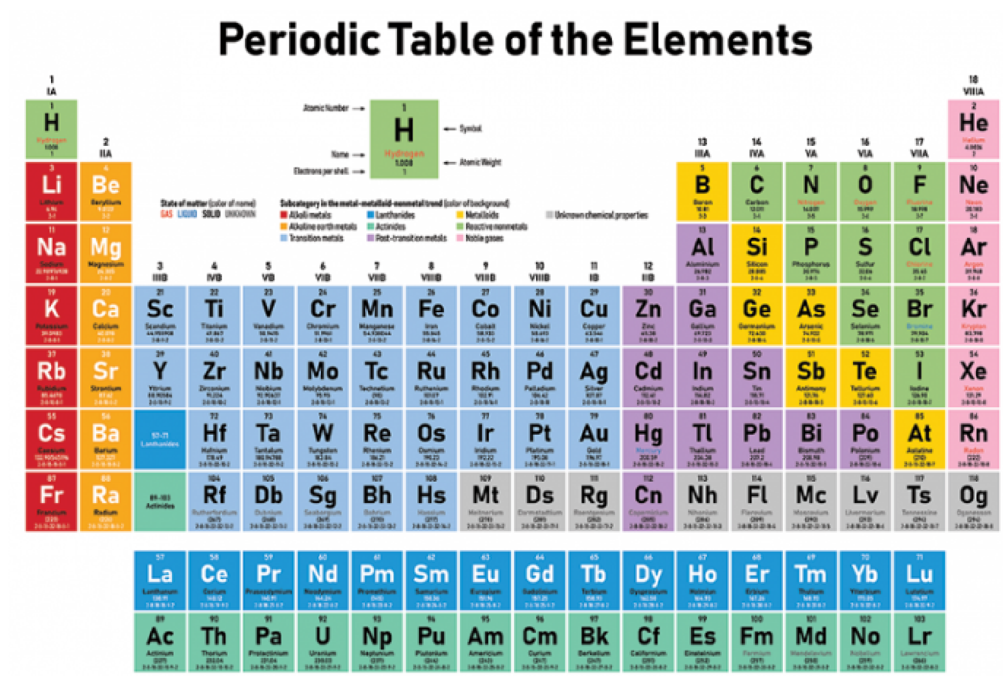
Ref: N. Bhat, CeNSE, IISc Bangalore- Nanoelectronic Device Technology

CMOS scaling:

Micro-MOSFET



Nano-MOSFET



CMOS scaling

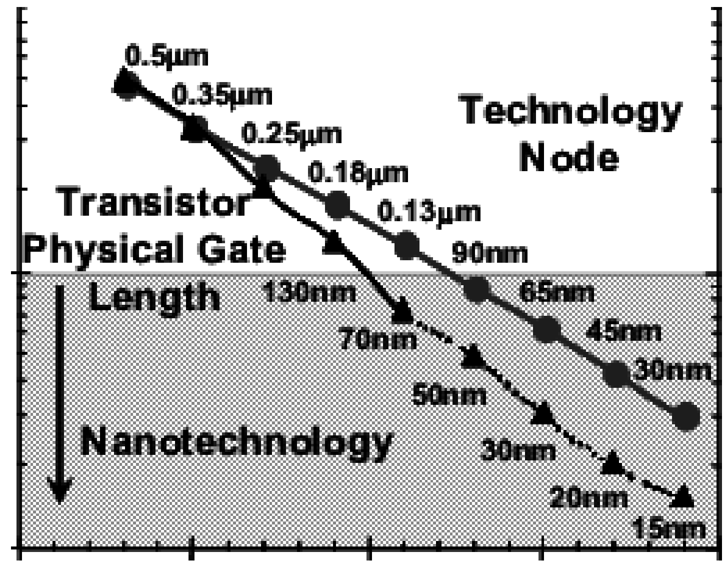
Technology scaling

Parameter	Constant Field	
Supply voltage (V_{dd})	$1/\alpha$	**Scaling Variables**
Length (L)	$1/\alpha$	
Width (W)	$1/\alpha$	
Gate-oxide thickness (t_{ox})	$1/\alpha$	
Junction depth (X_j)	$1/\alpha$	
Substrate doping (N_A)	α	
Electric field across gate oxide (E)	1	**Device Repercussion**
Depletion layer thickness	$1/\alpha$	
Gate area (Die area)	$1/\alpha^2$	
Gate capacitance (load) (C)	$1/\alpha$	
Drain-current (I_{dss})	$1/\alpha$	
Transconductance (g_m)	1	
Gate delay	$1/\alpha$	**Circuit Repercussion**
Current density	α	
DC & Dynamic power dissipation	$1/\alpha^2$	
Power density	1	
Power-Delay product	$1/\alpha^3$	

CMOS scaling

Typical Scaling Scenario

- 1974 : 5μm Technology, Vdd = 10V
- 1984 : 1μm Technology, Vdd = 5V
- 1994 : 0.35μm Technology, Vdd = 3.5V
- 2004 : 90nm Technology, Vdd = 1V



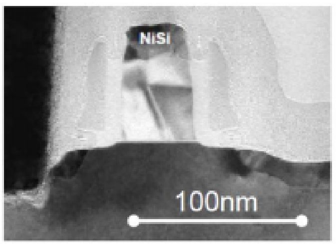
CMOS scaling and FINFET technology

SiO_2/SiON dielectric based

High-k/Metal gate based

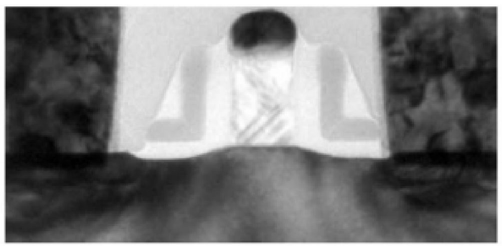
FINFET based

90 nm node



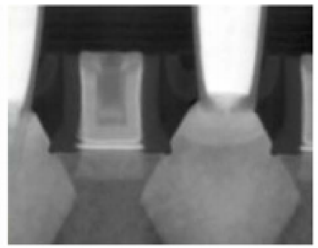
T. Ghani *et al.*, *IEDM* 2003

65 nm node



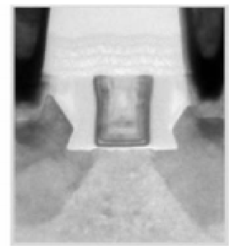
(after S. Tyagi *et al.*, *IEDM* 2005)

45 nm node



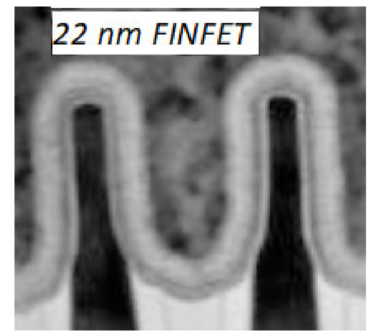
K. Mistry *et al.*, *IEDM* 2007

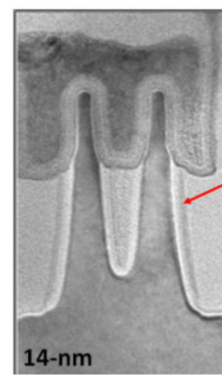
32 nm node

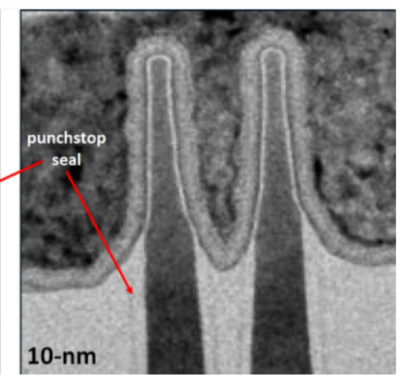


P. Packan *et al.*, *IEDM* 2009

22nm/14nm/10nm/7nm/5nm







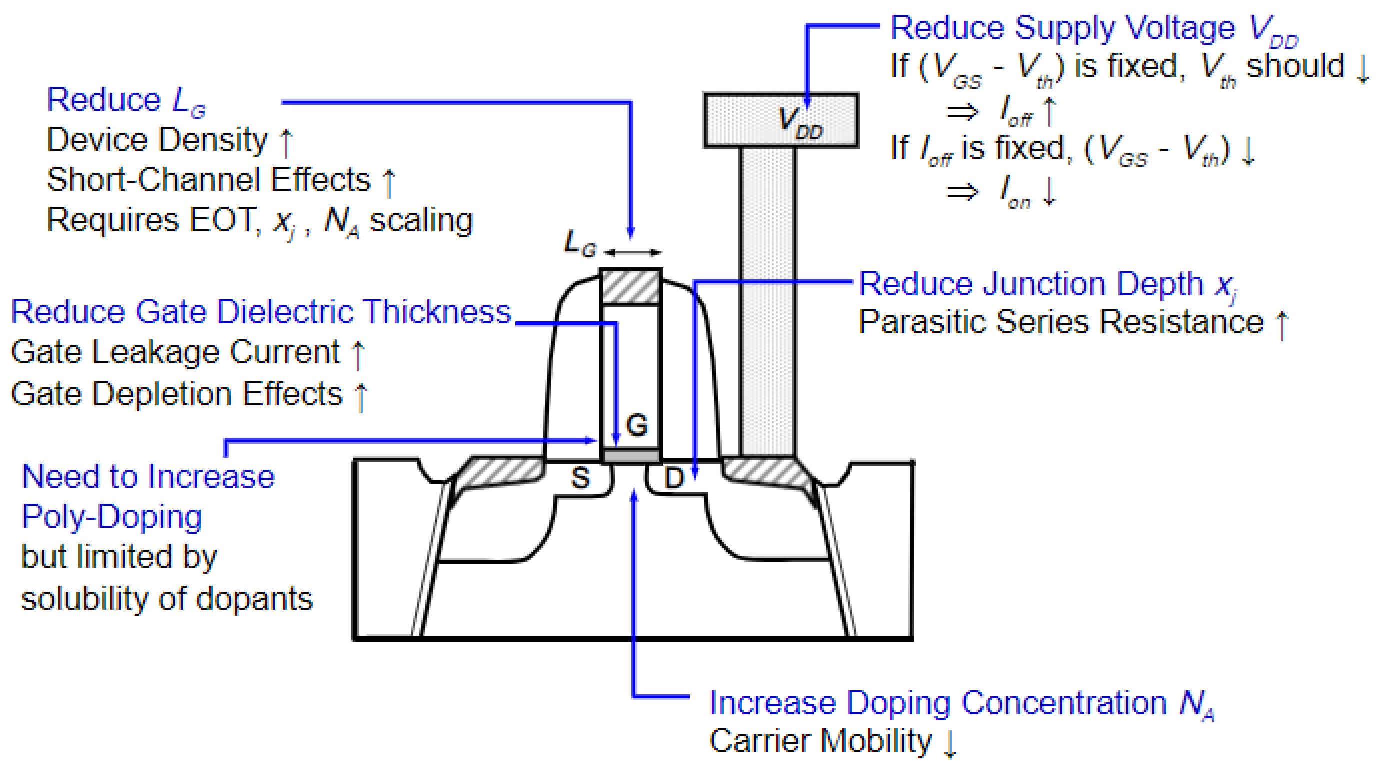
Gate length has not scaled proportionately with device pitch (0.7x per generation) in recent generations. – Transistor performance has been boosted by other means.

SOURCE: Intel

CMOS scaling:

- Gate length scaling
- Supply voltage scaling
- Gate oxide scaling
- Scaling of Doping
- Shallow S/D

Overview of Scaling Limitations



<u>CMOS scaling Issues:</u>

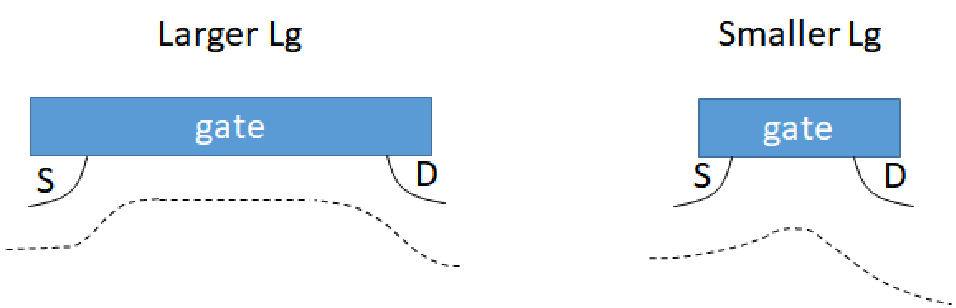
- Short channel effects
- Threshold voltage variation
- DIBL (Drain Induced Barrier Lowering)
- Gate leakage current
- GIDL (Gate Induced Drain Leakage)
- Shallow S/D – Parasitic resistance
- Mobility issues/Velocity saturation/Hot carrier effect

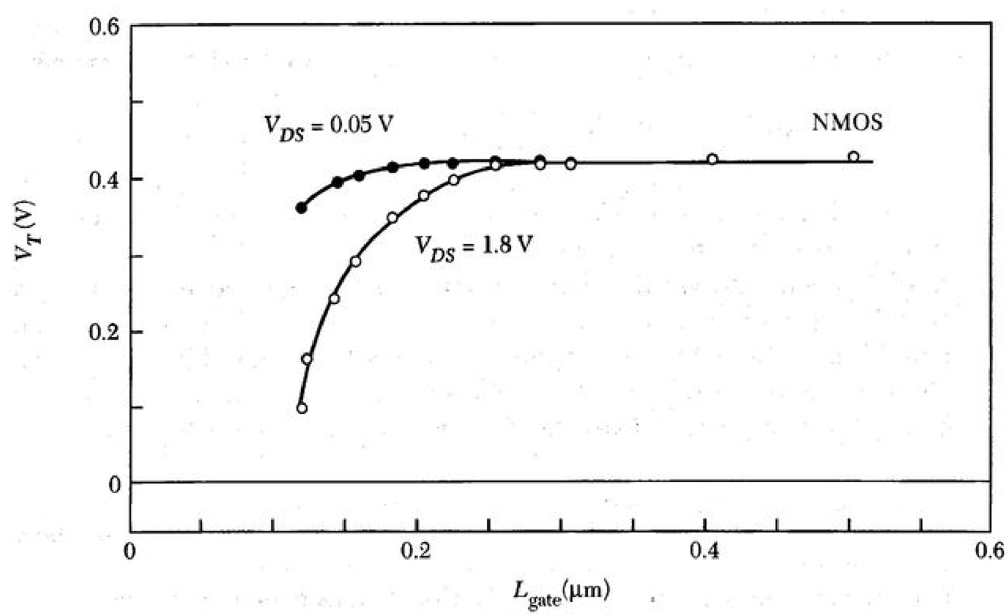
<u>CMOS scaling:</u>

- Gate length scaling

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

Threshold Voltage: (V_{th}) Variation





CMOS scaling:

- Gate length scaling

Channel Length Modulation

- It occurs when transistor is in *Saturation region*.

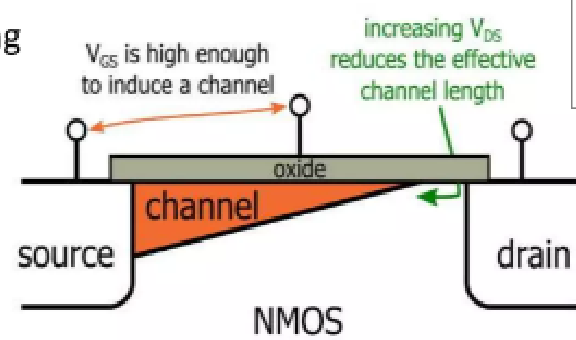
  i.e. **Saturation region**,

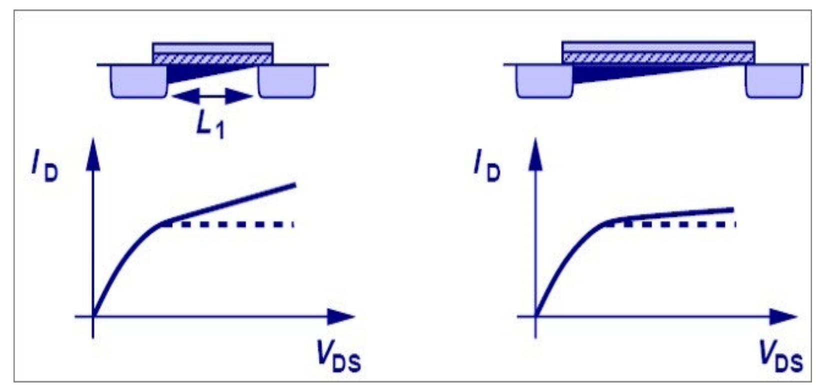
  V_{GS}>V_{th} and V_{DS}> V_{GS} - V_{th}

  I_D increases slightly with increasing V_{DS} .

- The pinch-off point moves toward the source as V_{DS} increases.

- The length of the channel becomes shorter with increasing V_{DS}.

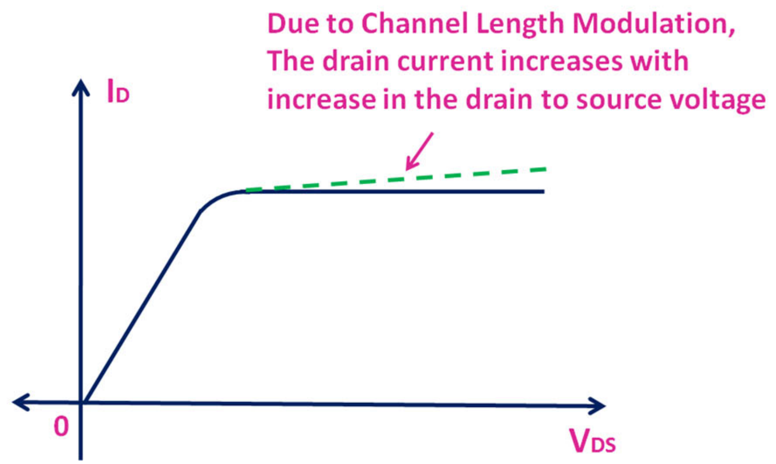


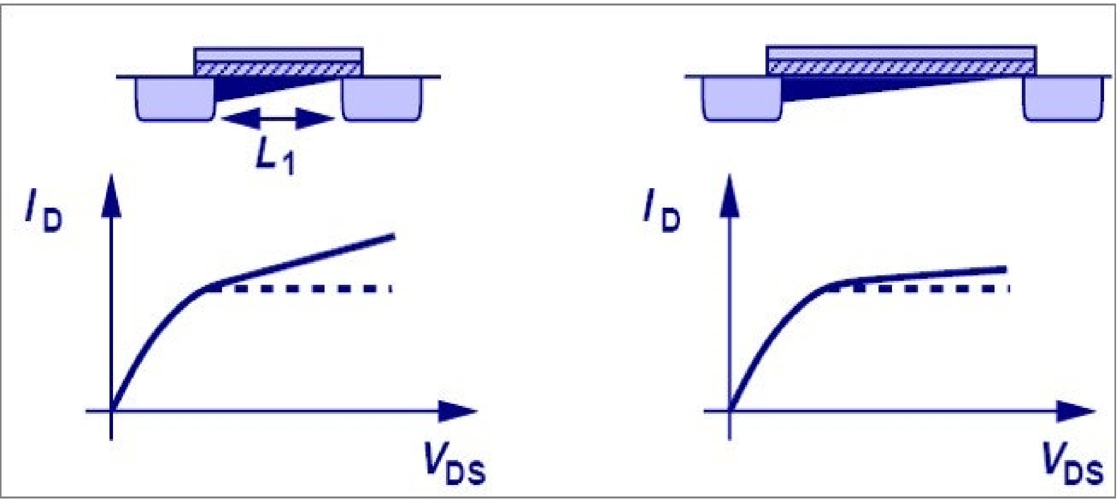


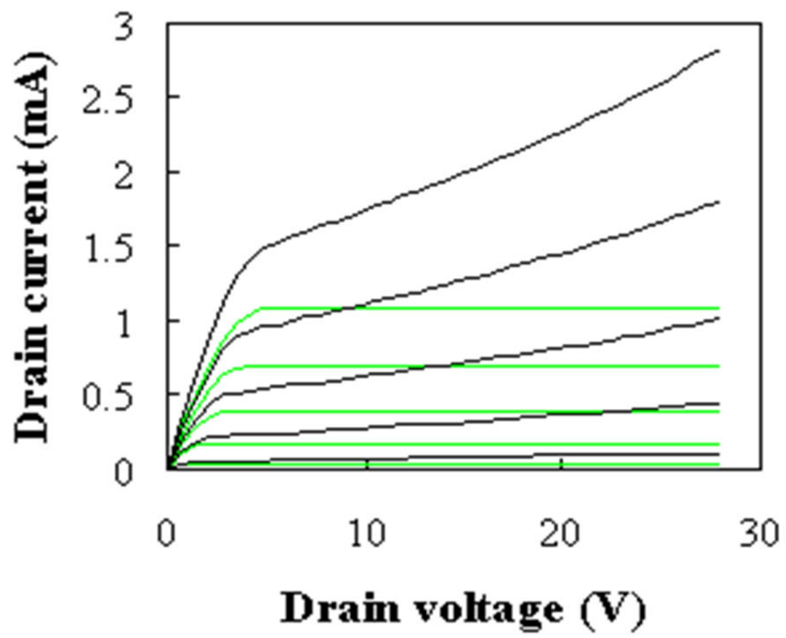
CMOS scaling:

- Gate length scaling

Channel length Modulation







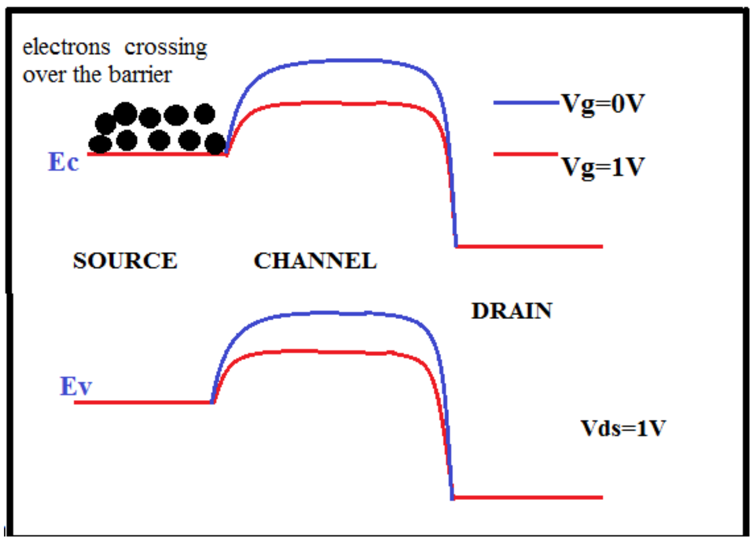
CMOS scaling:

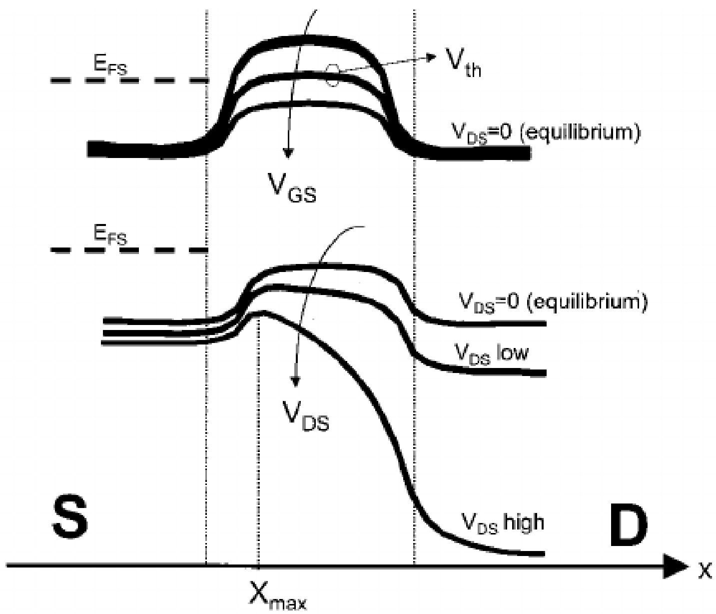
- Gate length scaling

Drain Induced Barrier Lowering (DIBL)

As the source and drain get closer, they become electrostatically coupled, so that the drain bias can affect the potential barrier to carrier diffusion at the source junction

→ V_T decreases (*i.e.* OFF state leakage current increases)

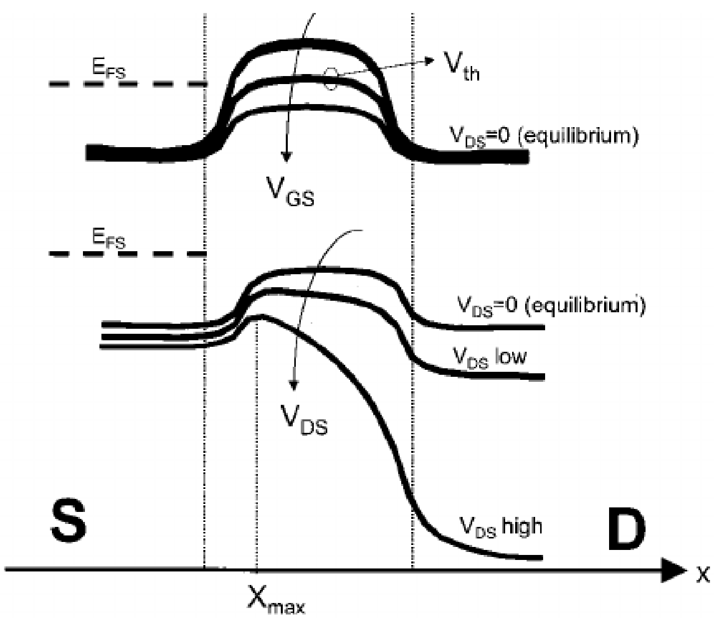


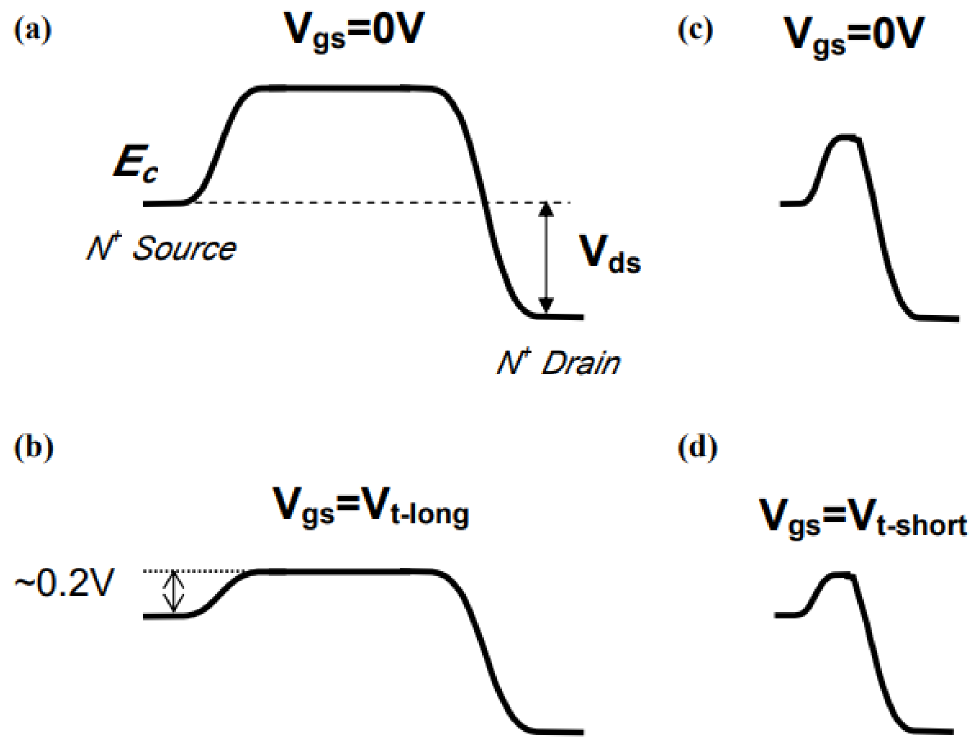


CMOS scaling:

- Gate length scaling

Drain Induced Barrier Lowering (DIBL)

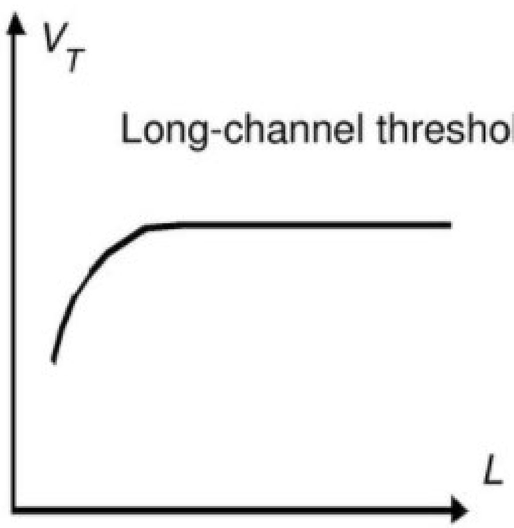




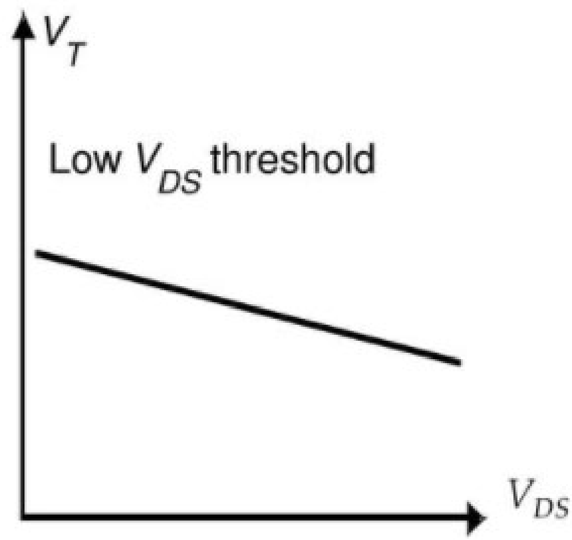
(a)-(d): Energy-band diagram from source to drain when V_{gs}=0V and $V_{gs}=V_t$

Gate length scaling

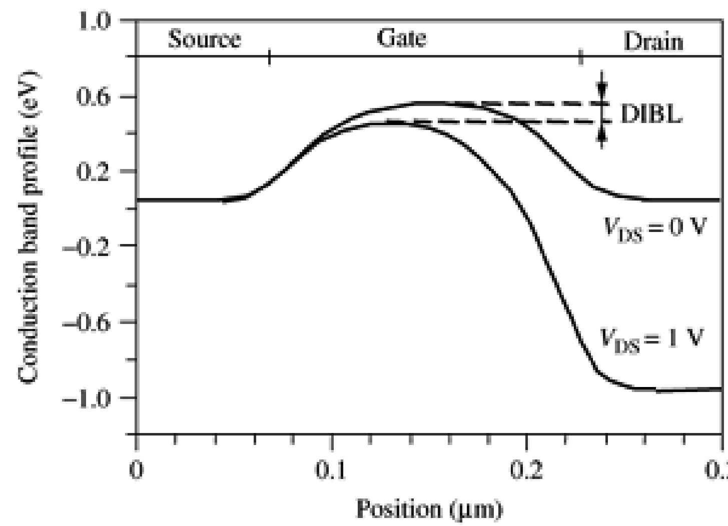
Threshold Voltage: (V_{th}) and DIBL



Threshold as a function of the length (for low V_{DS})



Drain-induced barrier lowering (for low L) …(DIBL)



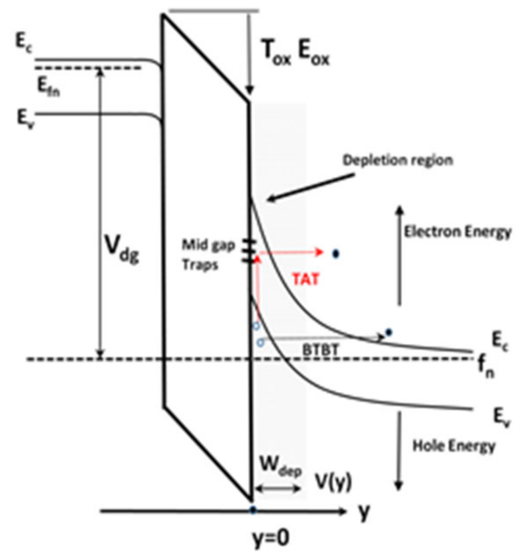
J. Ho, "Introduction and Short Channel Effects", 2014, Semiconductors

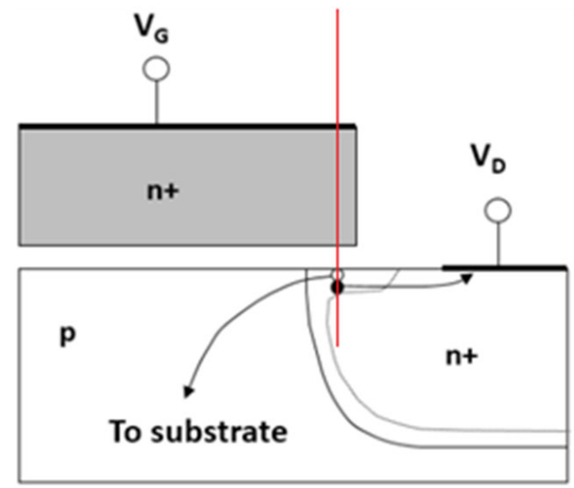
Gate length scaling

Gate Induced Drain Leakage (GIDL)

- GIDL occurs when the V_{DG} potential and band bending are high enough to generate electron/hole pairs by valence to conduction band tunneling

- GIDL arises in MOSFETs due primarily to band to band tunneling that occurs between the drain and the gate
- GIDL prevents the drain from reaching a current of zero when it is in its "off" state





CMOS scaling: Gate dielectric Scaling

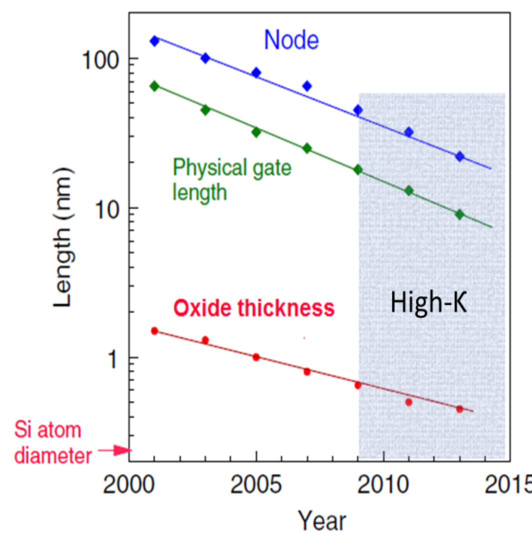
- SiO_2/SiON based dielectric till 65 nm CMOS Technology Node
- Best properties with silicon substrate
- Conventional SiON < 12Ǻ loses its intrinsic insulative property.

- SiON < 12Ǻ -Increased leakage current and poses reliability issues
- To reduce leakage**, high-κ dielectrics** with equivalent C_{OX} introduced.

SiO_2/SiON Dielectric based

1st Production	1997	1999	2001	2003	2005	2007	2009
Process Generation	0.25µm	0.18µm	0.13µm	90 nm	65 nm	45 nm	32 nm
Gate dielectric	SiO_2	SiO_2	SiO_2	SiO_2	SiO_2	High-k	High-k
Gate electrode	Poly-silicon	Poly-silicon	Poly-silicon	Poly-silicon	Poly-silicon	Metal	Metal

Source: Intel



$$\mathrm{EOT} = \left(\frac{\kappa_{\mathrm{ox}}}{\kappa_{\text{hi-}k}}\right) t_{\text{hi-}k}$$

Robertson, Rep. Prog. Phys. 69 (2006) 327-396

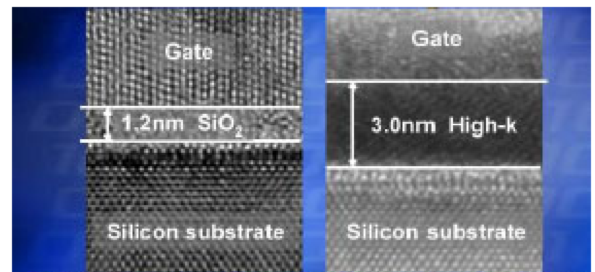
CMOS scaling: HK-MG technology

- HK-MG technology: 45 nm and 32 CMOS Technology Node
- *k* value - high enough and scalable
- Thermal stability
- Compatible with Si CMOS tech.
- Good interface with Si
- Lower defects in the bulk
- Not used after 32nm CMOS technology

HK-MG technology

1st Production	1997	1999	2001	2003	2005	2007	2009
Process Generation	0.25μm	0.18μm	0.13μm	90 nm	65 nm	45 nm	32 nm
Gate dielectric	SiO_2	SiO_2	SiO_2	SiO_2	SiO_2	High-k	High-k
Gate electrode	Poly-silicon	Poly-silicon	Poly-silicon	Poly-silicon	Poly-silicon	Metal	Metal

Source: Intel



Benefits compared to current process technologies

	High-k vs. SiO_2	Benefit
Capacitance	60% greater	*Much faster transistors*
Gate dielectric leakage	> 100x reduction	*Far cooler*

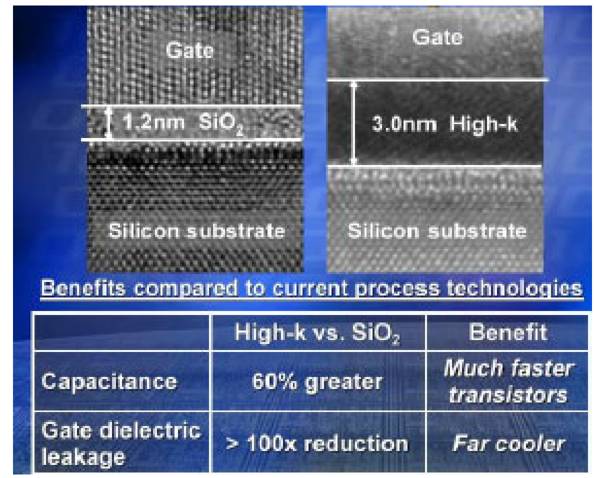
***Source*: Intel**

Same C_{OX}, ↑ t_{HK}

Robertson, Rep. Prog. Phys. 69 (2006) 327-396

CMOS scaling: HK-MG technology

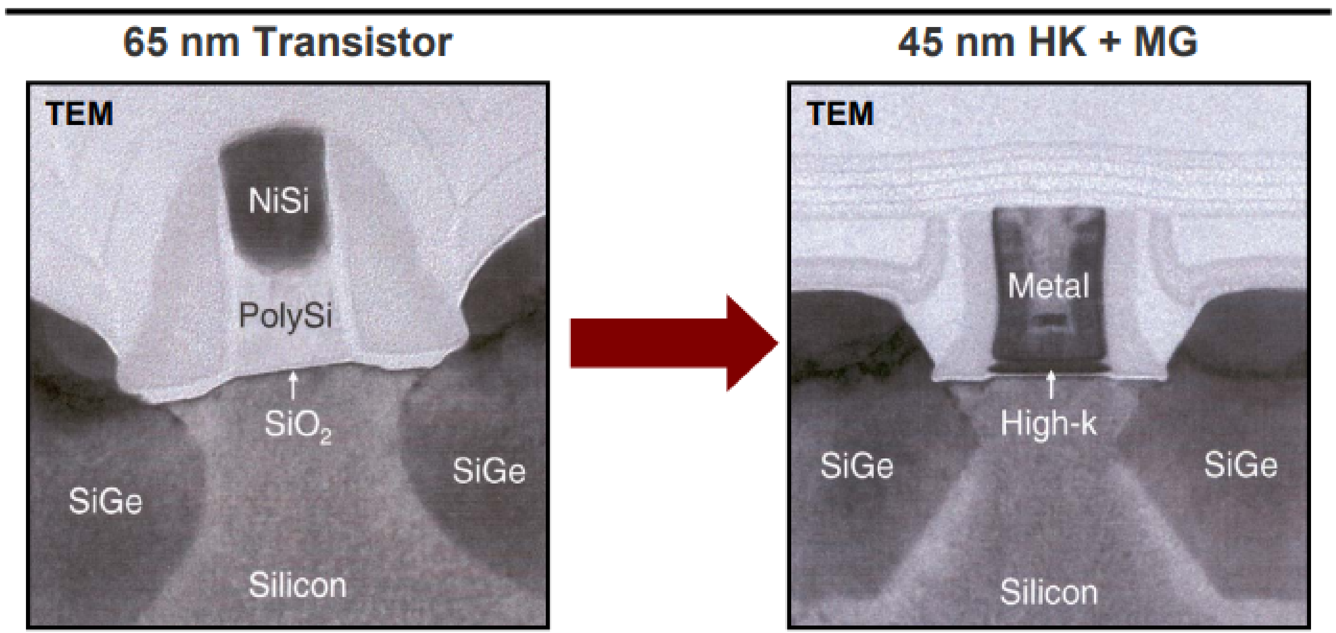
- Challenges:
  - High density of defects
  - Interaction with the polysilicon gate
  - Interaction with the substrate
  - Polycrystallization structure
  - Increased trap-assisted tunneling
  - Increased leakage current
  - Reliability and failure of high-k dielectric



Source: Intel

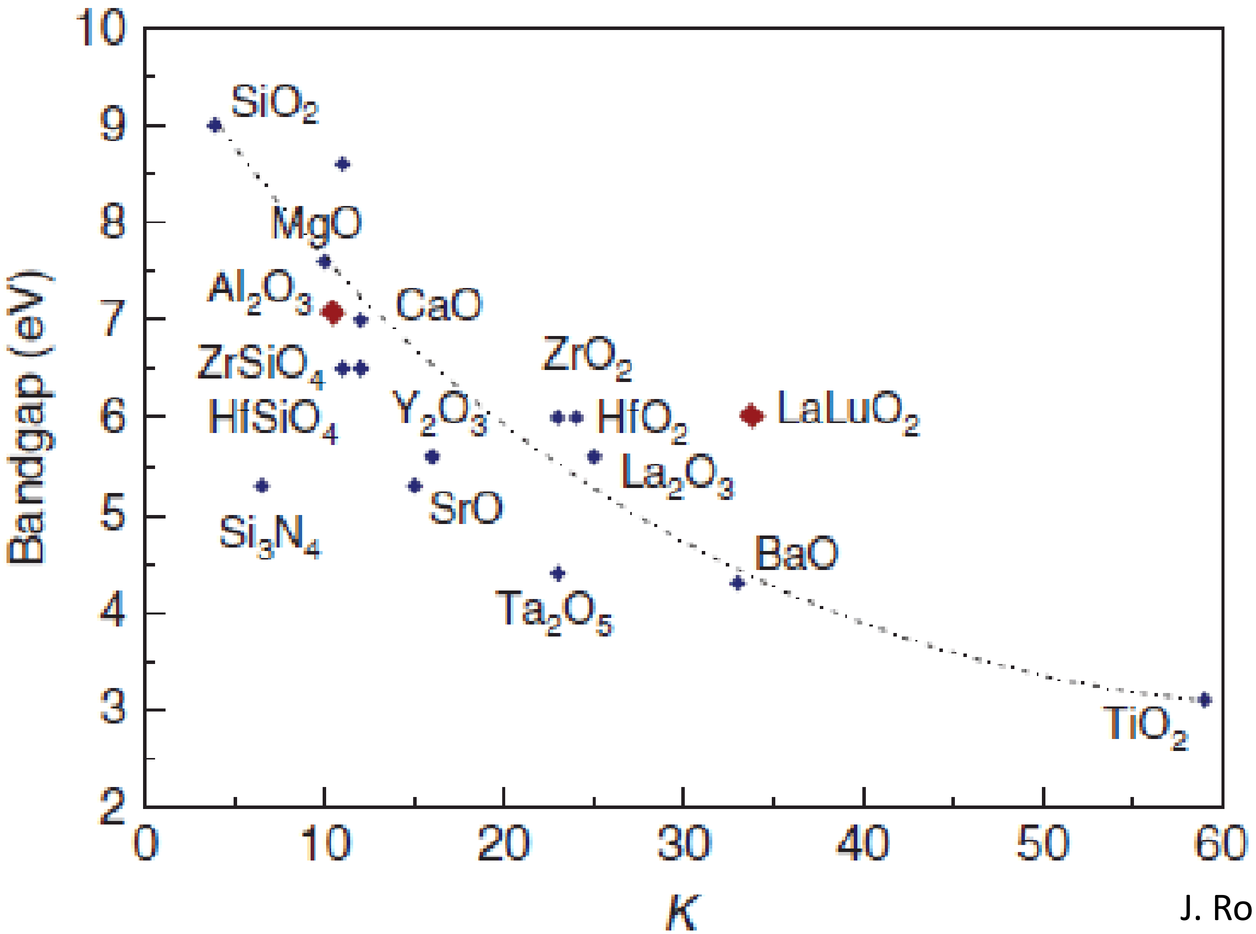
CMOS scaling: HK-MG technolgy

45nm High-k + Metal Gate Transistors



Source: Intel

CMOS scaling: HK-MG technolgy

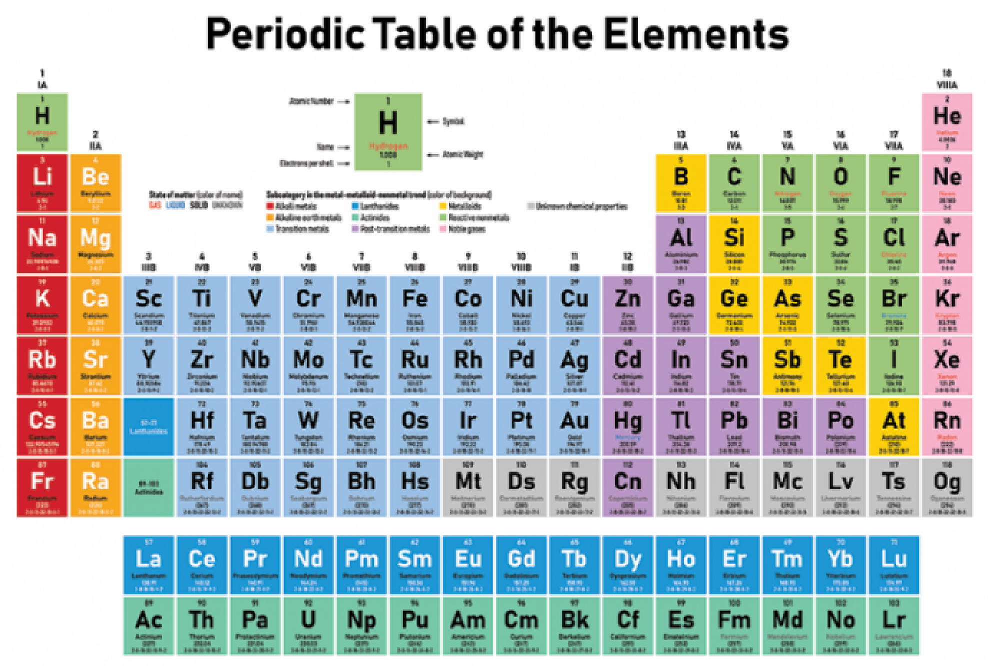


$$\mathrm{EOT} = \left(\frac{\kappa_{ox}}{\kappa_{hi\text{-}k}}\right) t_{hi\text{-}k}$$

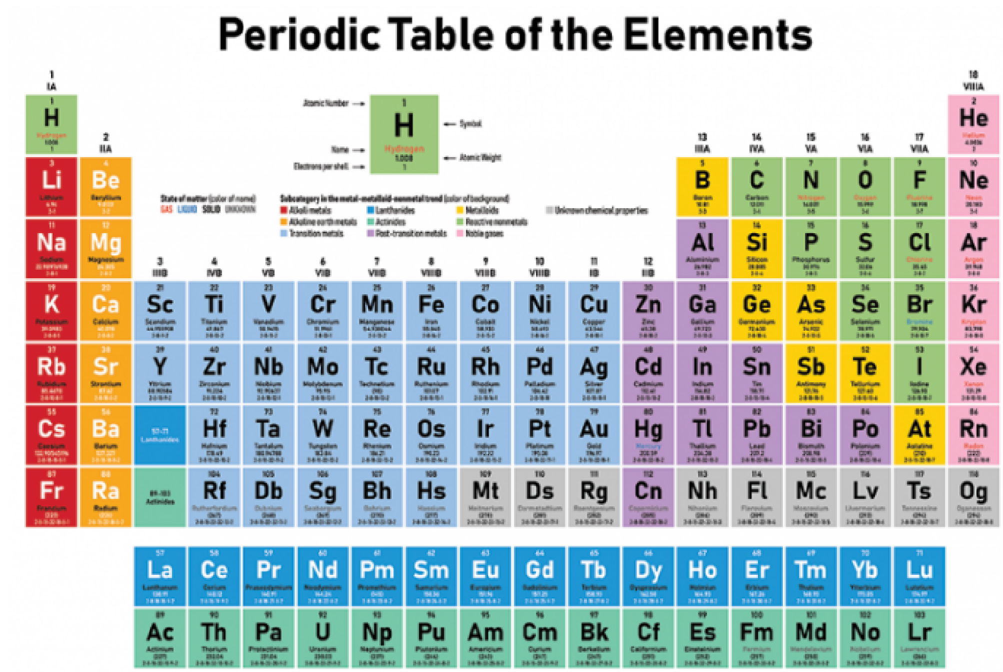
J. Robertson, J. Vac. Sci. Technol. B 18, 1785 (2000)

CMOS scaling:

Micro-MOSFET

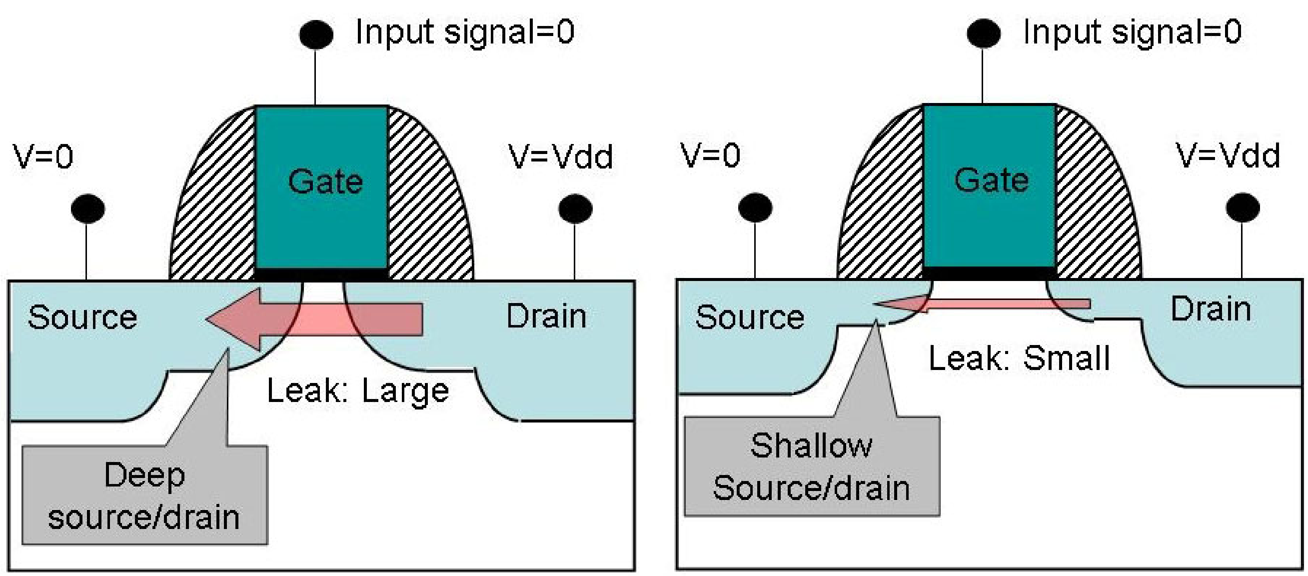


Nano-MOSFET



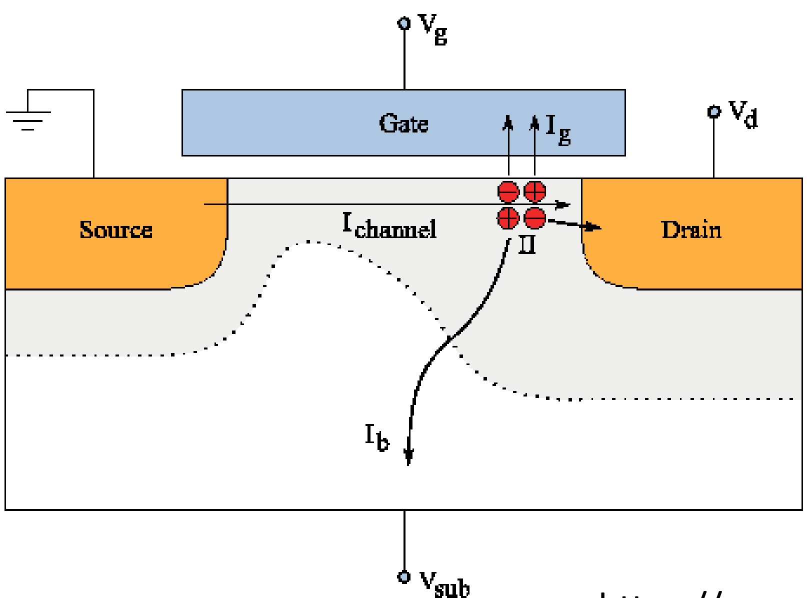
CMOS scaling: <u>Shallow Source/Drain:</u>

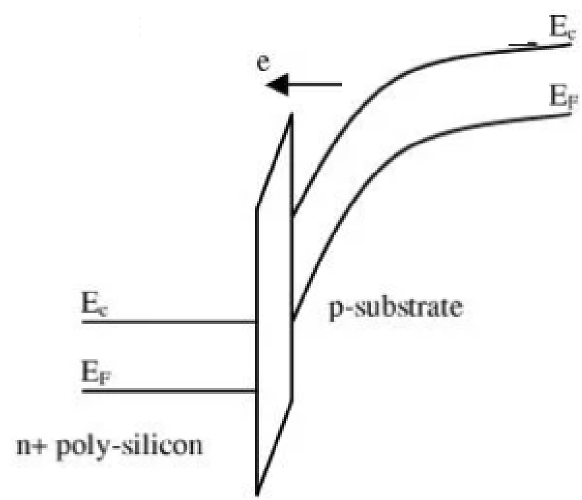
- To minimize the short channel effect and DIBL, we want shallow (small r_j) S/D regions – but the parasitic resistance of these regions increases when r_j is reduced.



Hot carrier injection

High electric field near the substrate-oxide interface energizes the electrons or holes and they cross the substrate-oxide interface to enter the oxide layer. This phenomenon is known as hot carrier injection.

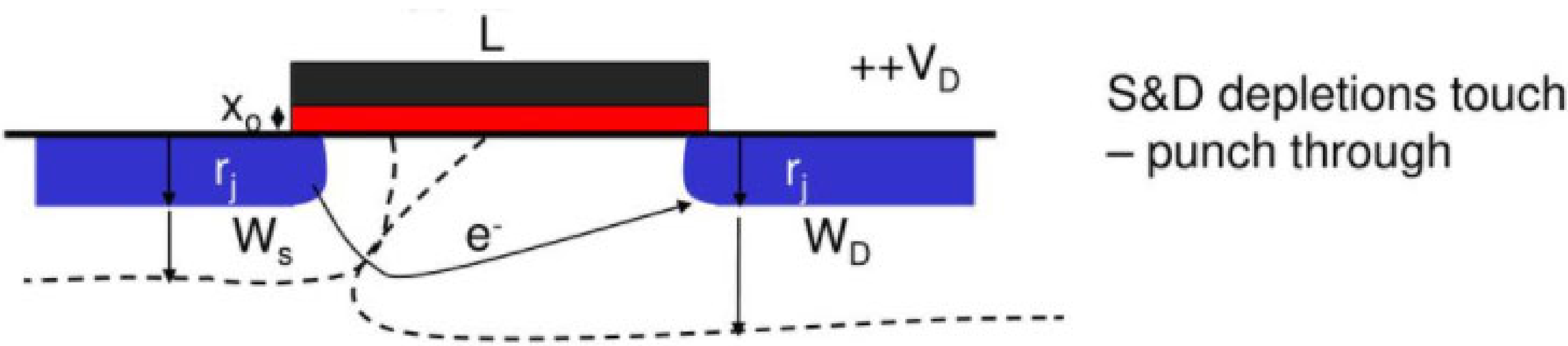




https://www.iue.tuwien.ac.at/phd/entner/node21.html

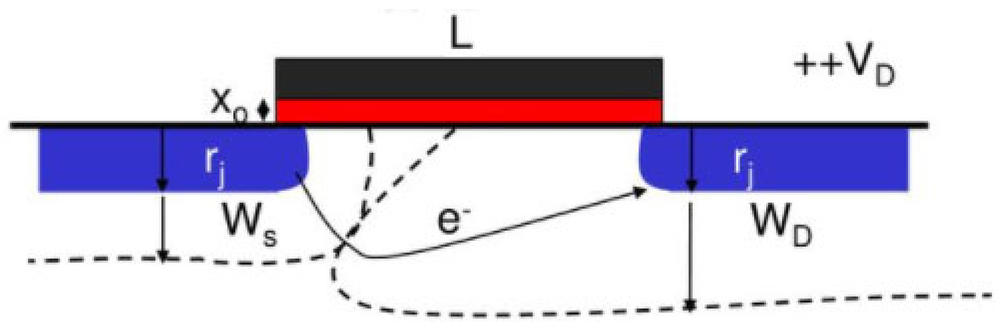
Punch through effect

- In short channel devices, due to the proximity of drain and source terminals, the depletion region of both the terminals come together and eventually merge. In such a condition, "punch-through" is said to have taken place.



J. Ho, “Introduction and Short Channel Effects”, 2014, Semiconductors

Punch through effect

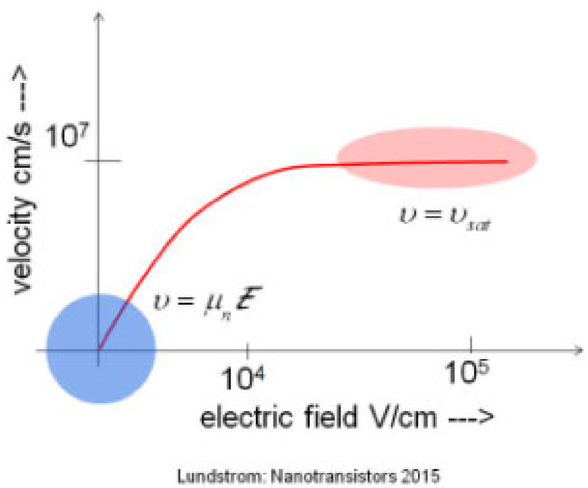


S&D depletions touch – punch through

- Drain current no longer controlled by gate voltage
- Drain current does not saturate
- High subthreshold current

J. Ho, “Introduction and Short Channel Effects”, 2014, Semiconductors

Velocity saturation



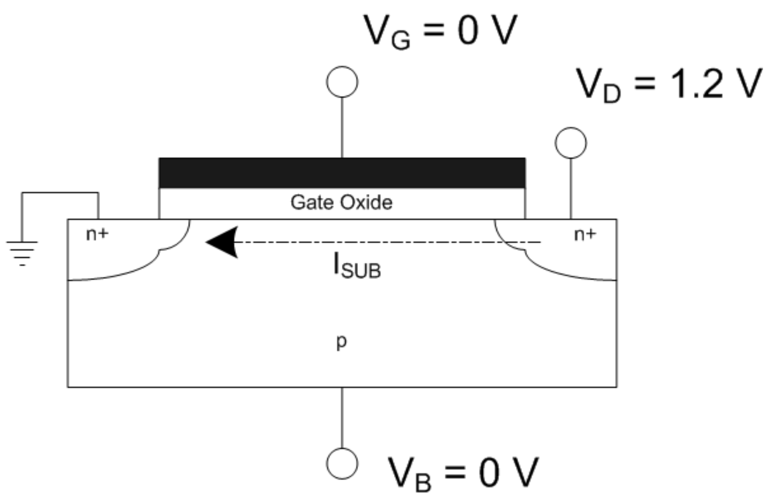
$V=\mu E$

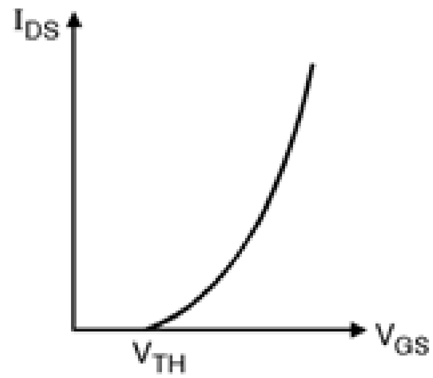
- Velocity of charge carriers is linearly proportional to the electric field and the proportionality constant is called as mobility of carrier.

- But when we increase the electric field beyond certain velocity called as the thermal velocity or saturated velocity the velocity of the charge carrier does not change with electric field

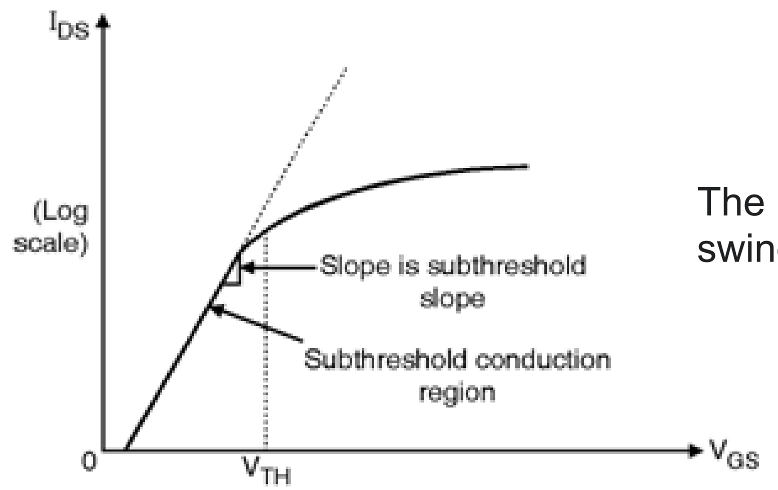
- The electric field at which the velocity of carrier saturates is called as the critical electric field. The loss of energy is because of the collisions of carriers called as scattering effect.

Subthreshold current





I_{DS} Vs V_{GS} characteristics in linear scale

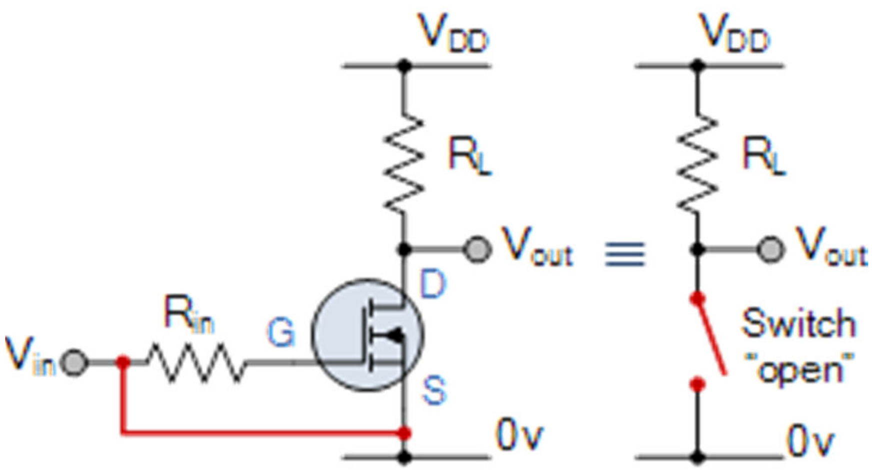


I_{DS} Vs V_{GS} characteristics in log scale

The minimum subthreshold swing= 60 mV/Dec

https://www.electronics-tutorial.net/Digital-CMOS-Design/Non-Ideal-Effects/Subthreshold-Conduction/

Subthreshold current

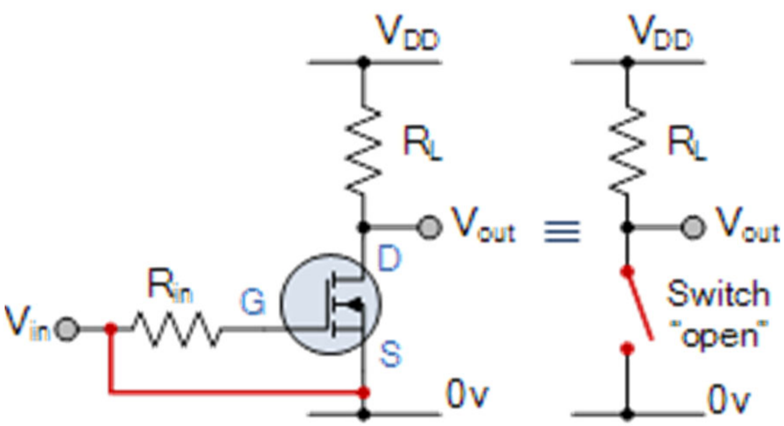


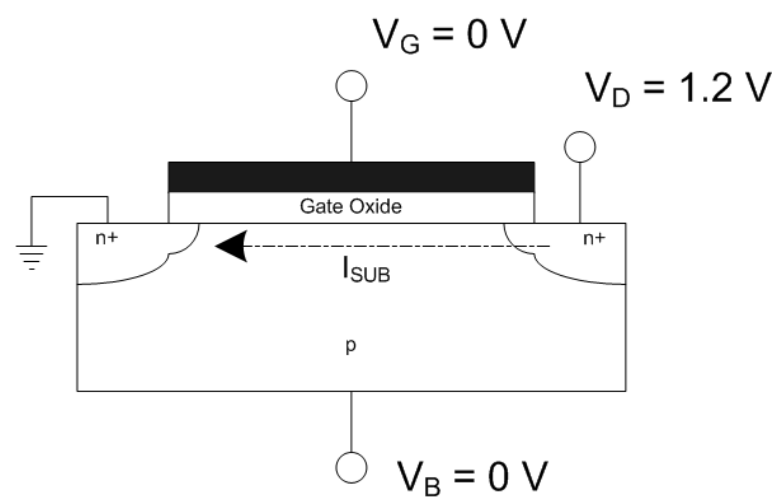
https://www.electronics-tutorials.ws/transistor/tran_7.html

Subthreshold current

Subthreshold conduction or subthreshold leakage t is **the current between the source and drain of a MOSFET** when

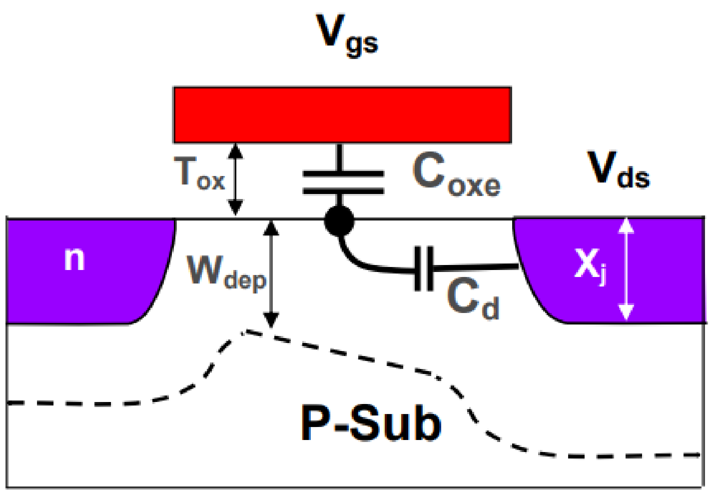
- the transistor is in subthreshold region, or
- weak-inversion region, that is, for gate-to-source voltages below the threshold voltage.

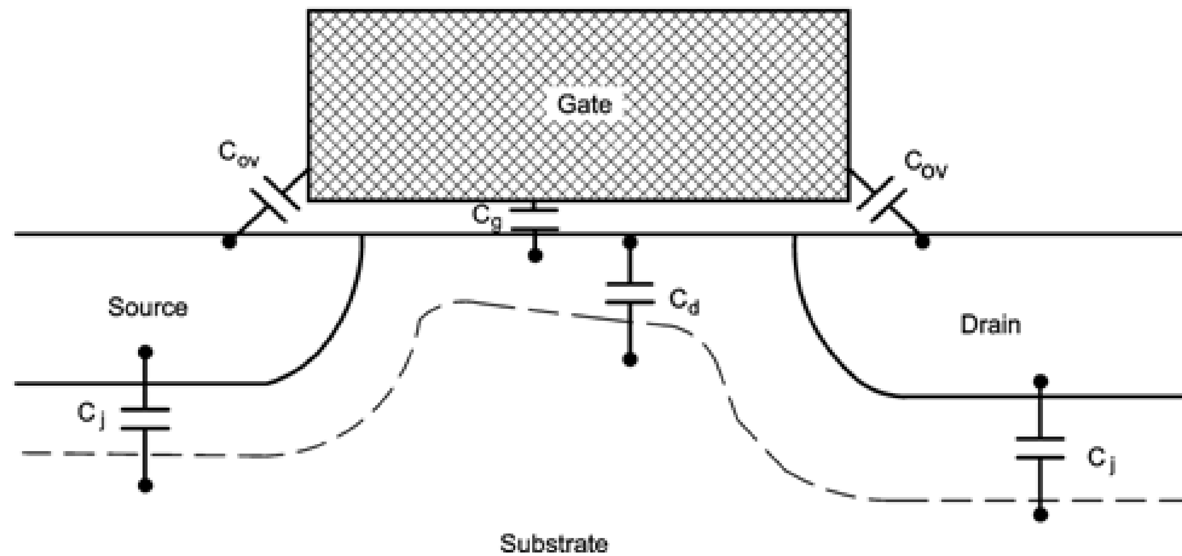




https://www.electronics-tutorials.ws/transistor/tran_7.html

Impact of scaling on parasitic C and R



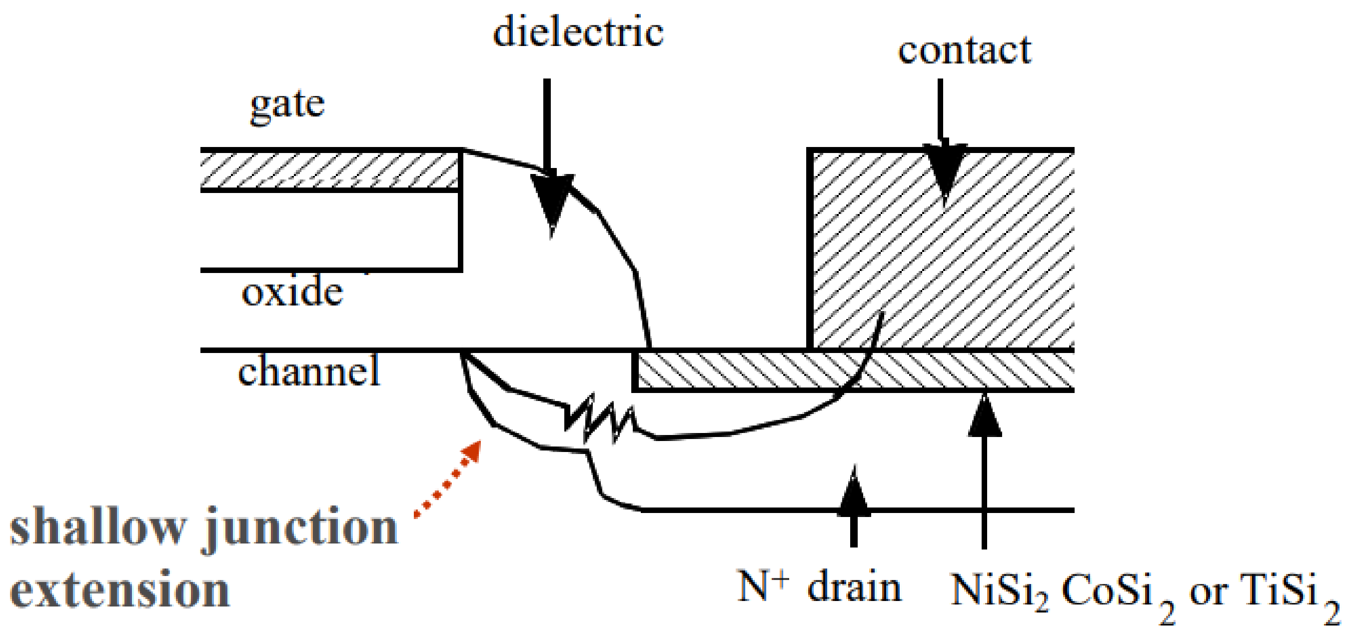


MOSFET parasitic capacitances

https://inst.eecs.berkeley.edu/~ee130/sp06/chp7full.pdf

https://www.electronics-tutorial.net/Analog-CMOS-Design/MOSFET-Parasitics/Parasitic-Capacitances-MOSFETS/

Impact of scaling on parasitic C and R



https://inst.eecs.berkeley.edu/~ee130/sp06/chp7full.pdf