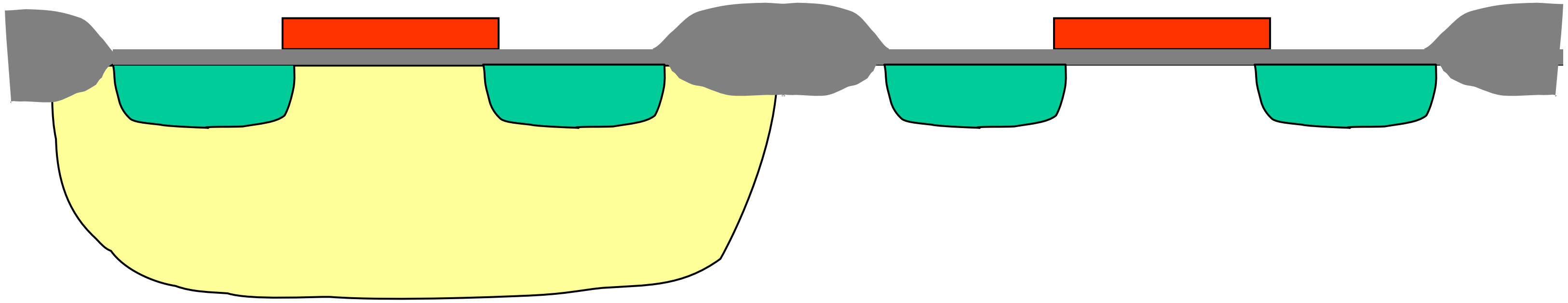
MOS transistors (in subthreshold)



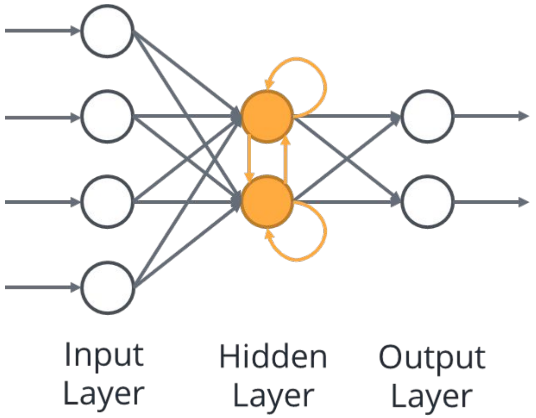
- Poll for lecture 1+Lab exercise sessions update
- Revisiting the cost of digital computation with real example of recurrent neural network hardware accelerator
- Quiz for lecture 1
- Standards for lab exercise reports
- Preview of this week's lab exercise: Measuring subthreshold FET IV characteristics
- History of MOSFET
- Review of Semiconductors
- What is a MOSFET? CMOS?
- Relation between physics of transistors and voltage-sensitive nerve membrane channels
- MOS capacitor structure
- Surface: *accumulation, depletion, inversion*
- Capacitive dividers: The *back-gate/body effect* parameter κ
- MOS transistor in *subthreshold / weak inversion* operation

Feedback poll and lab sessions update

1. Was the pace satisfactory? (Single Choice)
   - Answer 1: Too slow
   - Answer 2: About right
   - Answer 3: Too fast
2. Were you already familiar with the material? (Si
   - Answer 1: Yes, with most of it
   - Answer 2: Knew about some of it
   - Answer 3: Didn't know about most of it
3. Did you sign up for exercises? You must complete the exercises to take the exam. (Single Choice)
   - Answer 1: Yes
   - Answer 2: No
4. If you did not sign up for exercises, why not? (Single Choice)
   - Answer 1: Just auditing (you must be registered)
   - Answer 2: No slot suits me

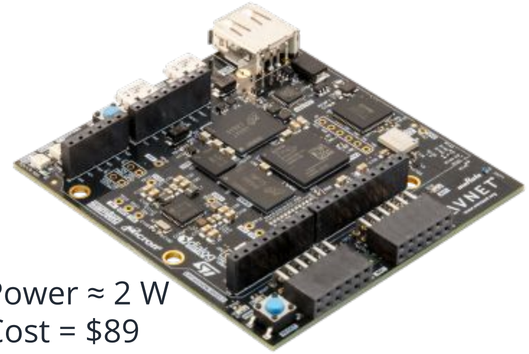
	Sep 24 THU 8:00 AM 10:00 AM	Sep 24 THU 10:00 AM 12:00 PM	Sep 24 THU 1:00 PM 3:00 PM	Sep 24 THU 3:00 PM 5:00 PM	Sep 25 FRI 8:00 AM 10:00 AM	Sep 25 FRI 10:00 AM 12:00 PM	Sep 28 MON 8:00 AM 10:00 AM	Sep 28 MON 10:00 AM 12:00 PM
32 participants	✓0/8	✓2/8	✓4/8	✓5/8	✓0/8	✓5/8	✓8/8	✓8/8

Energy consumption of EdgeDRNN digital recurrent neural network

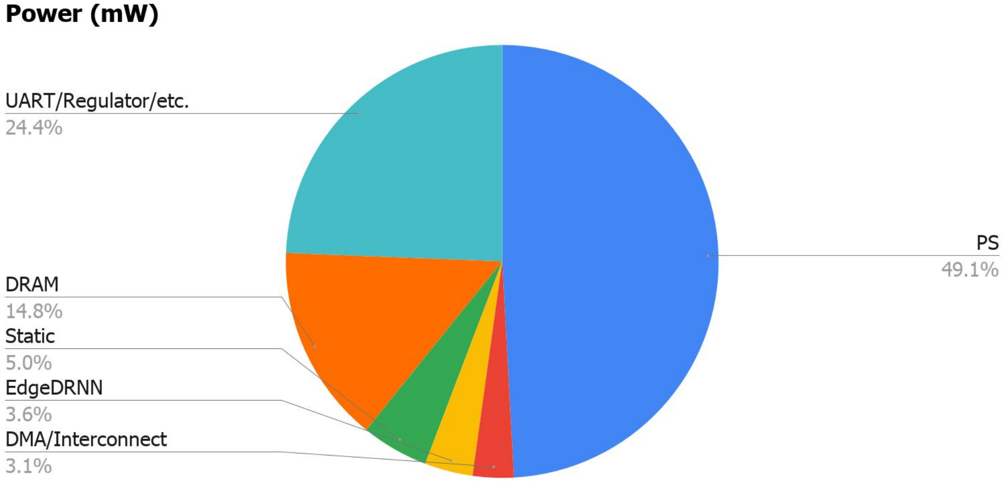


2L-768 unit GRU RNN

Throughput:
8 PEs * 2 Op/PE/clk * 125 MHz
=2 GOp/s
2 GOp/s / 87mW ➡️ **E_{Op}=45pJ/Op**



Power ≈ 2 W
Cost = $89



Arithmetic and local memory: 87mW

28nm technology

t_{ox}=1.4nm

C_{ox}=4*8.8e-12F/m/t_{ox}=25e-3F/m^2

A_{chan}=28nm x 35nm=e-15m^2

C_{chan}=C_{ox}*A_{chan}=2.5e-17F

Energy to activate a transistor

$E_{FET}=C_{chan}(\Delta V)^2/2$

V_{dd}=1V, ΔV=Vdd/2

E_{FET}=3e-18J

Equiv # transistor activations for each 8*16 bit SRAM memory IO + math operations:

E_{Op}/E_{FET}=45e-12 J/Op / 3e-18 J/FET

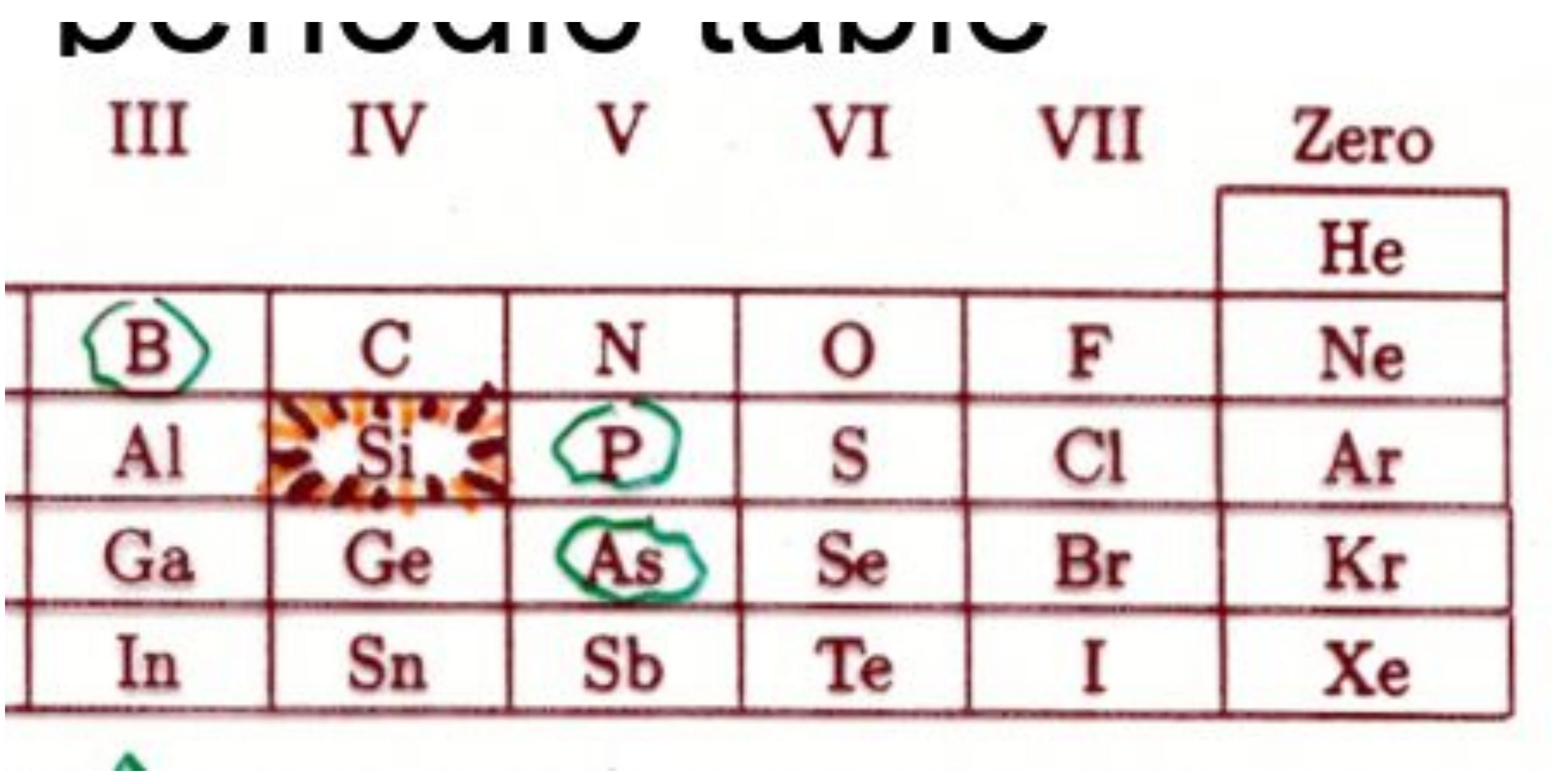
= **1.5e7 FET/Op**

Any questions about device physics?

Let’s do a quiz

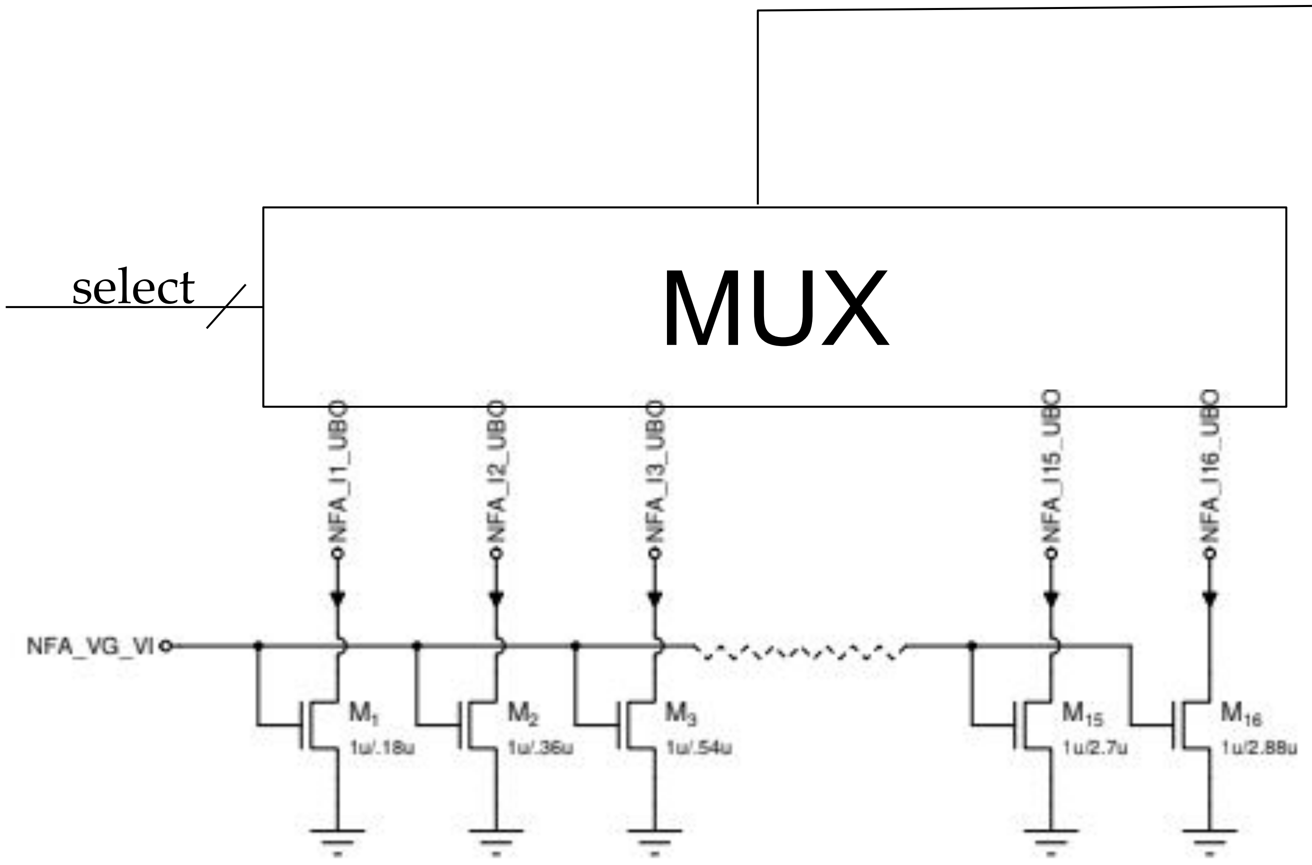
Quiz on device physics

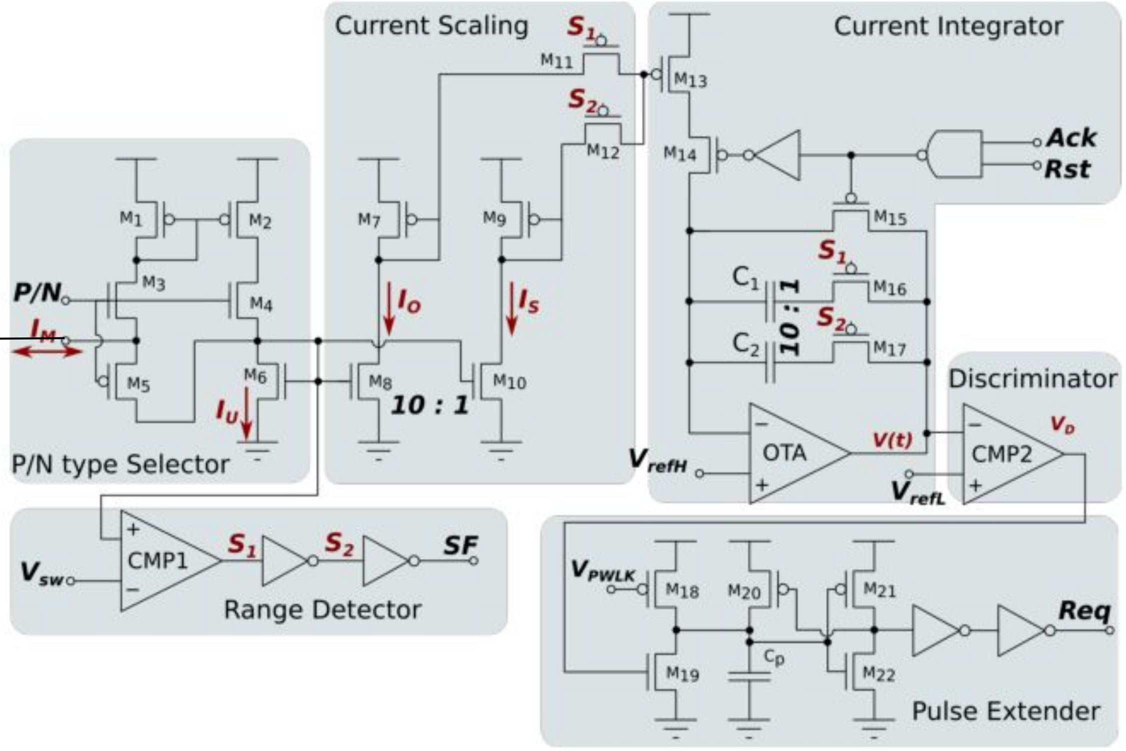
<u>Without looking at periodic table,</u> is Phosphorus a donor or acceptor?



III	IV	V	VI	VII	Zero
					He
B	C	N	O	F	Ne
Al	Si	P	S	Cl	Ar
Ga	Ge	As	Se	Br	Kr
In	Sn	Sb	Te	I	Xe

Preview of lab exercise





W/L=1/(0.18, .36, .054….. 2.7, 2.88) um/um

$$\delta T = \frac{\beta C(V_{refH} - V_{refL})}{\alpha I_{mon}}$$

$\beta = 10$

$\alpha = 1$ or 10

$V_{refH,L}$=?,?

Qiao, N., and G. Indiveri. 2016. "An Auto-Scaling Wide Dynamic Range Current to Frequency Converter for Real-Time Monitoring of Signals in Neuromorphic Systems." In *2016 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 160–63. https://doi.org/10.1109/BioCAS.2016.7833756 .

Avoid these common mistakes in your lab reports

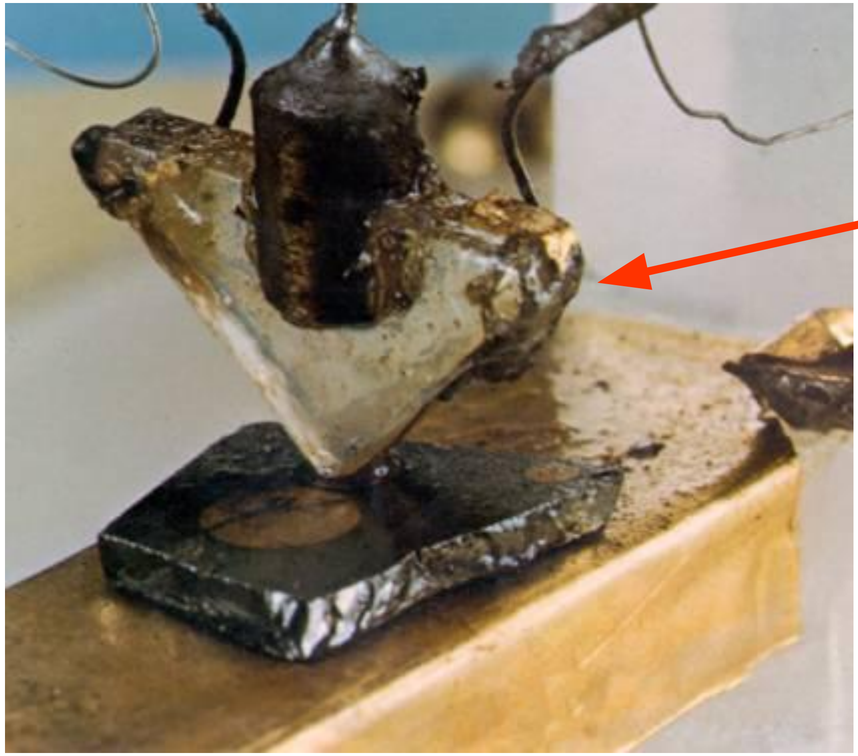
- **Not discussing your data sufficiently.** Think about a publication. The readers want to understand your reasoning with you. They want to be able to reproduce your results.
- **Using axes/plot fonts that are too small.** Use set(gca,'fontsize',18)
- **Not using cross-hair axes when 0,0 is relevant as the origin.**
- **Forgetting to mention what your plot shows.**
- **Forgetting units on your axes.**
- **Not labeling your figures with a caption,** e.g., "Fig. 1: Transistor drain current vs. gate voltage, Experiment 1."
- **Insufficiently labeling your data.** It's good to annotate your plots to indicate the slope of the curve, or the x / y intercepts.
- **Using identical markers for all plots.** Your curves must be distinguishable.
- **Forgetting units on measurements,** e.g. *"our conductance is 1.000653e-10"*. What are the units? is the reader supposed to guess?
- **Giving your measurements too many digits of precision**; see previous error. Do your instruments really give you 7 digits of precision?
- **Tip: Use notebooks cell editing mode,** where sections are separated to generate your measurements and plots. That way, you can quickly change and repeat your measurements and plotting.

Blackboard summary of subthreshold FET operation

History of the Transistor

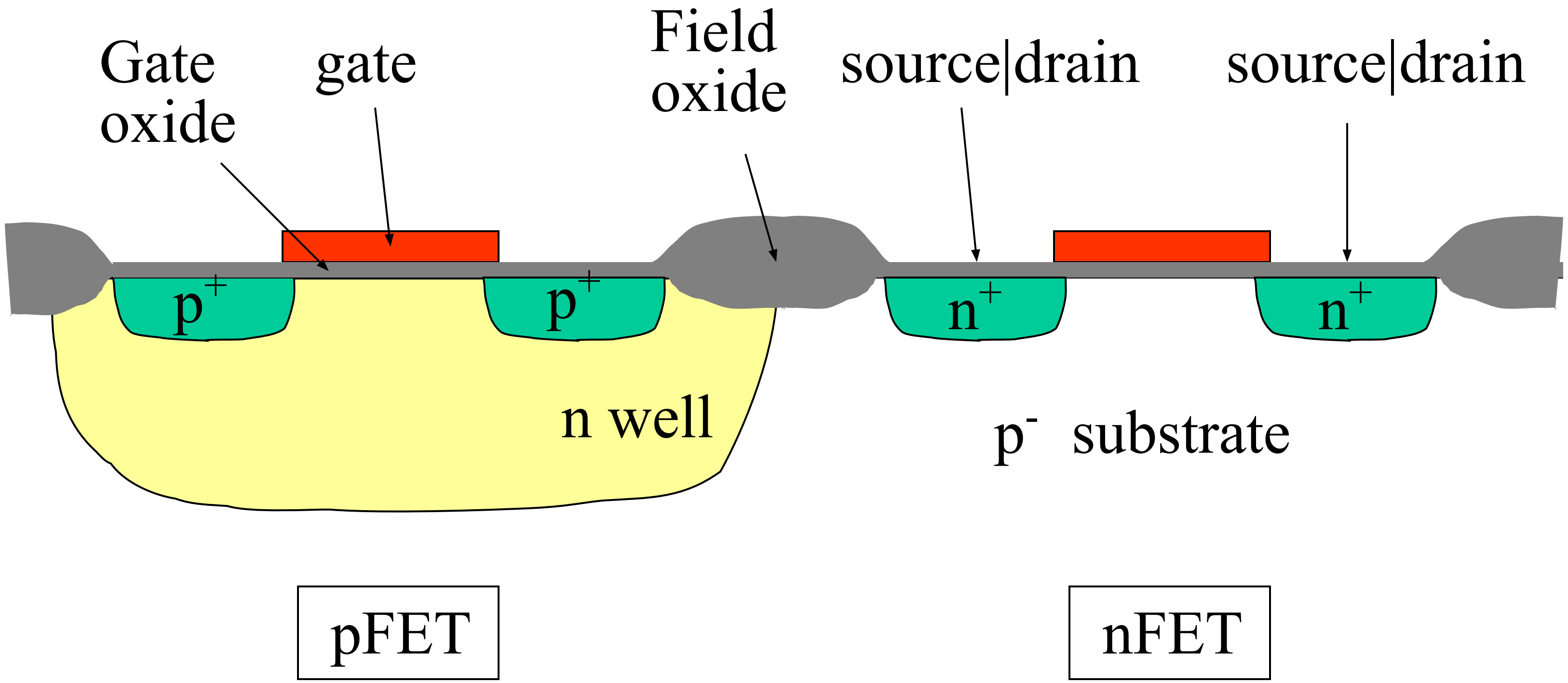
The term “transistor” is a generic name for a solid-state device with 3 or more terminals.

The field-effect transistor structure was first described in a patent by J. Lilienfeld in the 1930s! It took more than **40 years** before MOS transistors were in mass production.

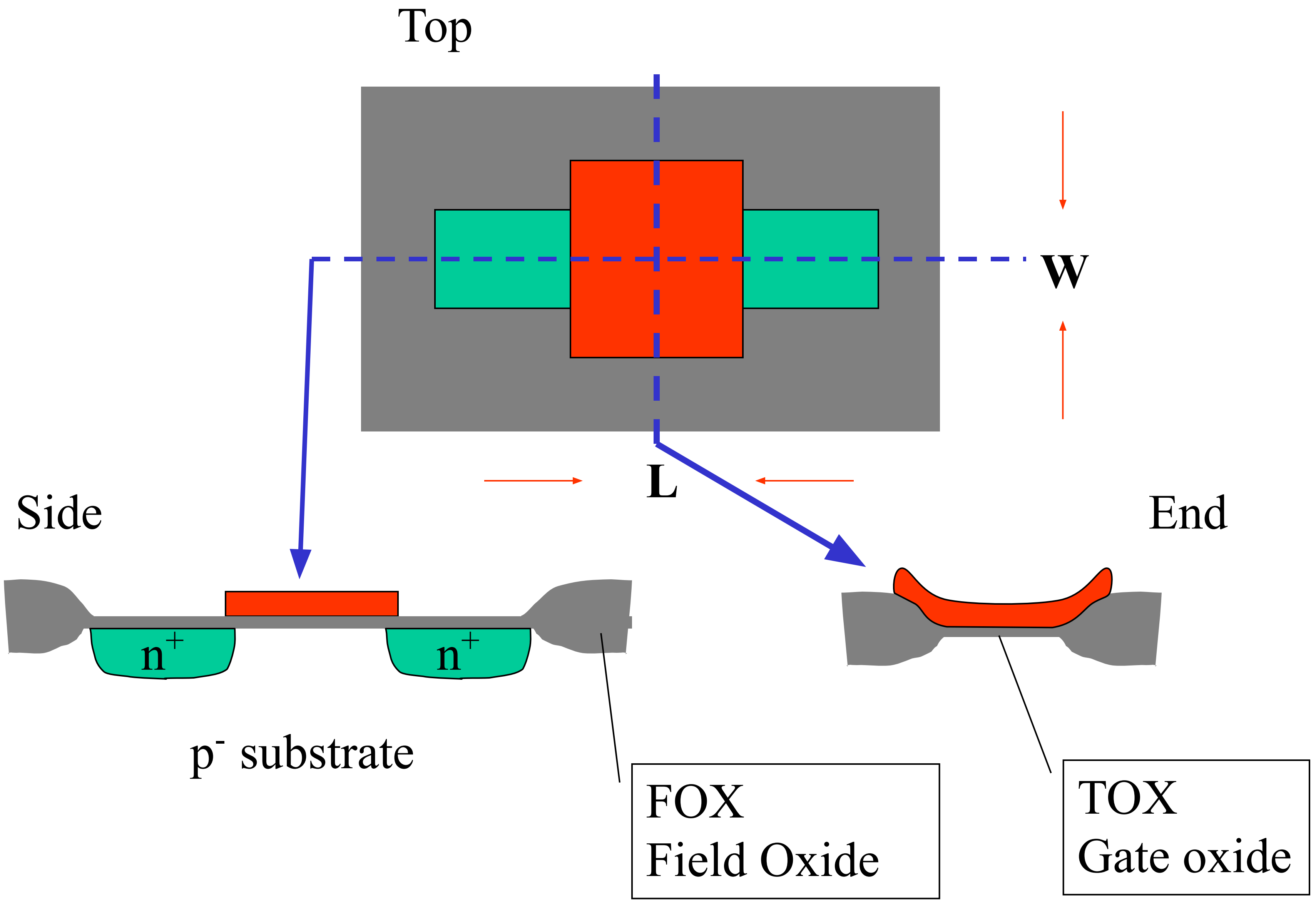


The first transistor (point-contact bipolar) fabricated at Bell Labs in 1947 (Bardeen, Brattain, Shockley). MOS transistors were not commercialized until mid 1970’s.

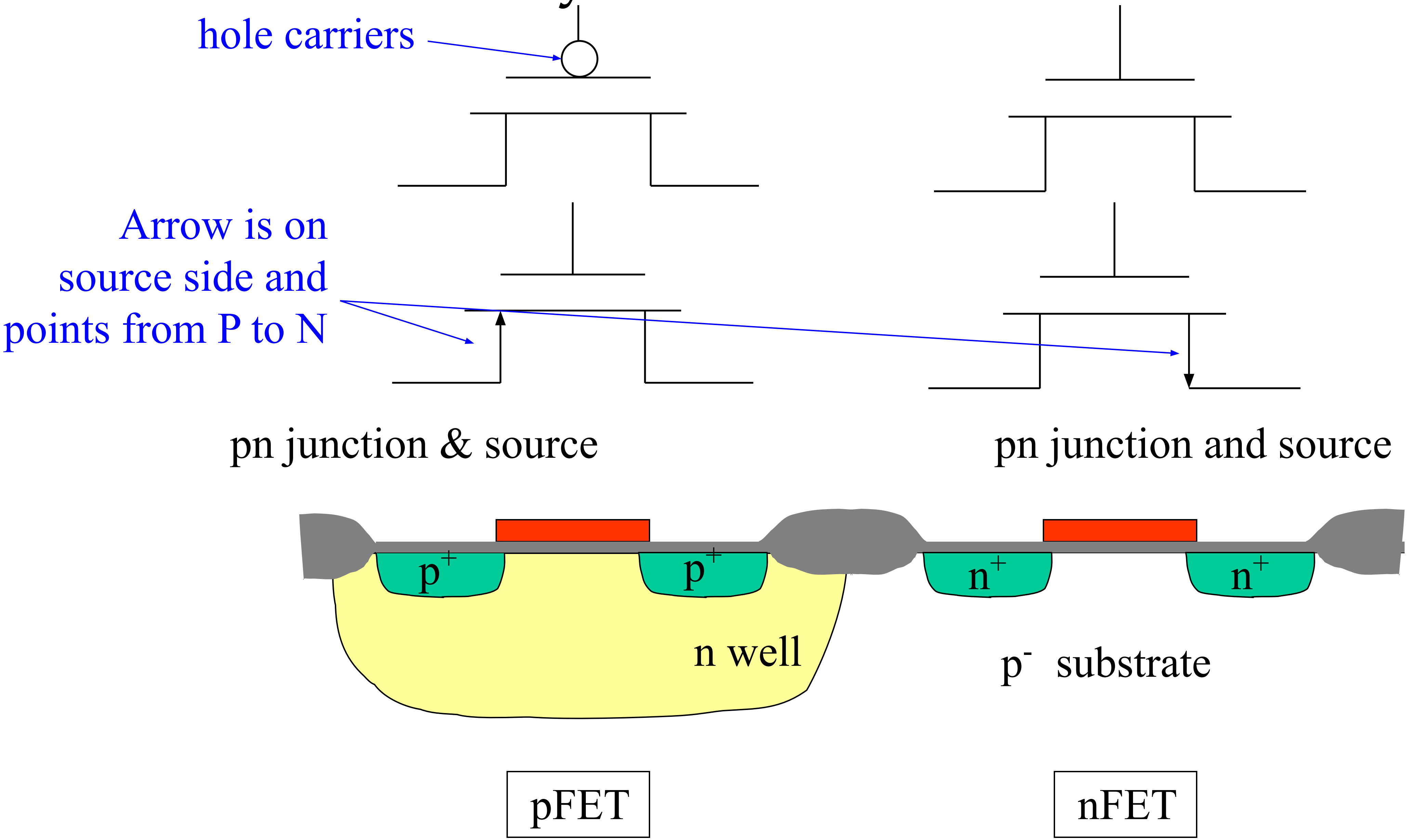
Cross-section of a complementary pair of Field-Effect Transistors (FETs)



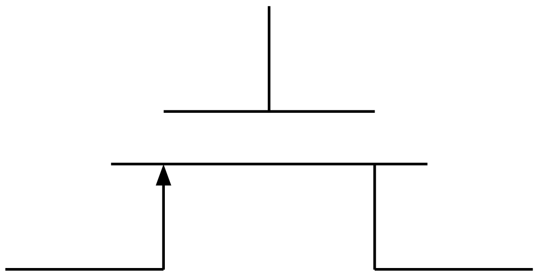
Top and Side Views of Field-Effect Transistor (FET)

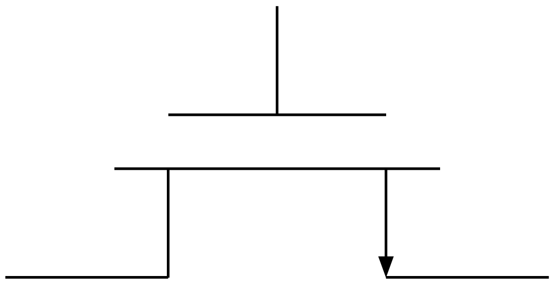


Symbols for transistors



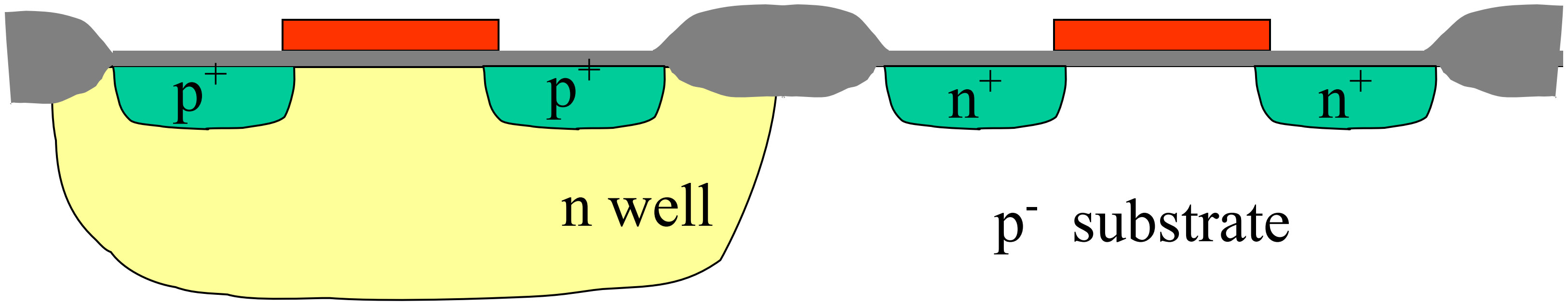
Alternative symbols for transistors





pn junction & source

pn junction and source



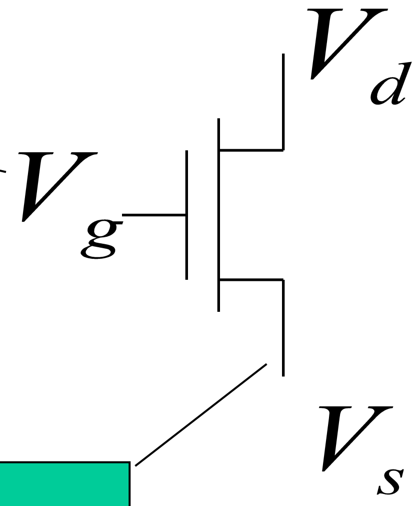
pFET

nFET

nFET terminology

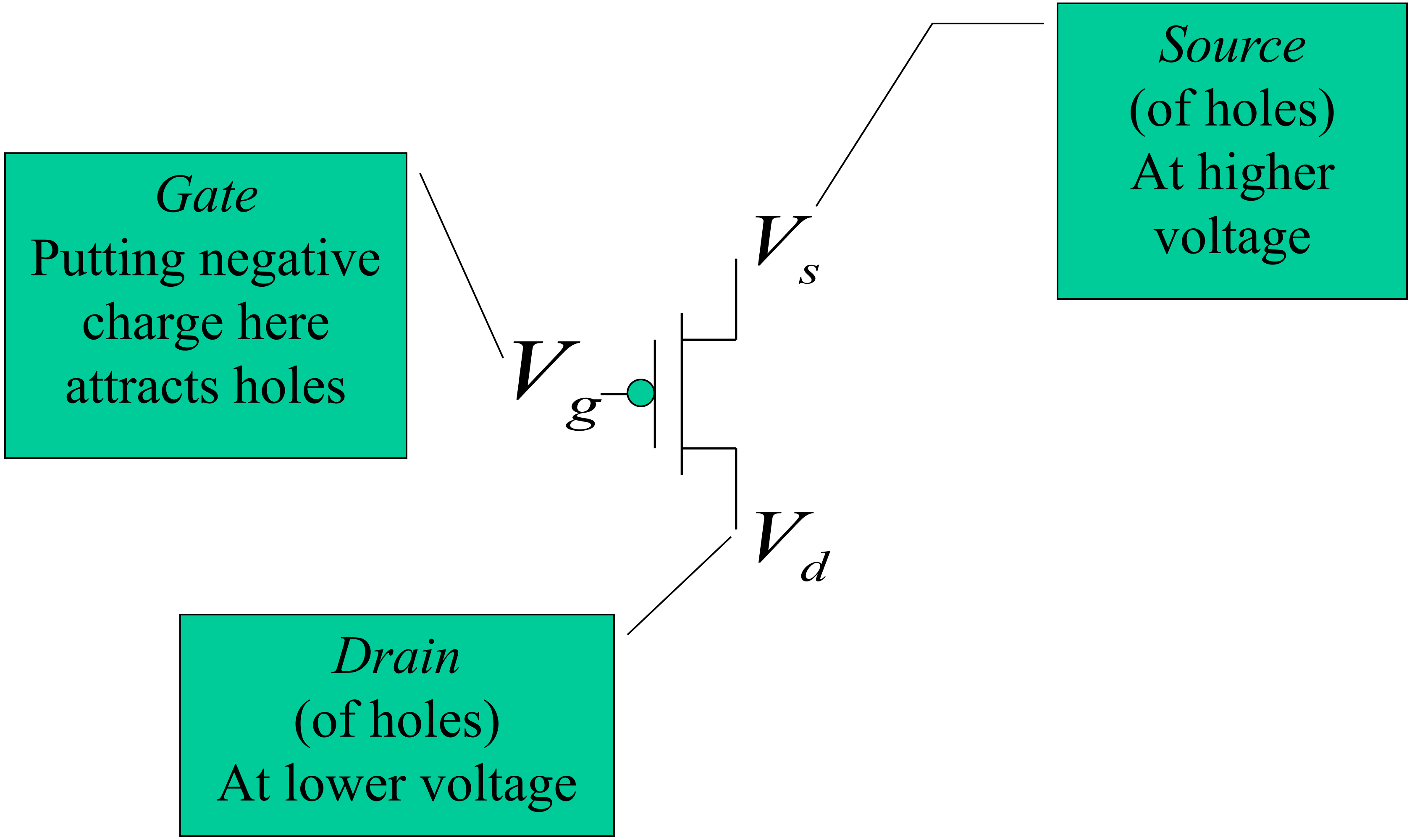
Drain
(of electrons)
At higher voltage

Gate
Putting positive charge here attracts electrons

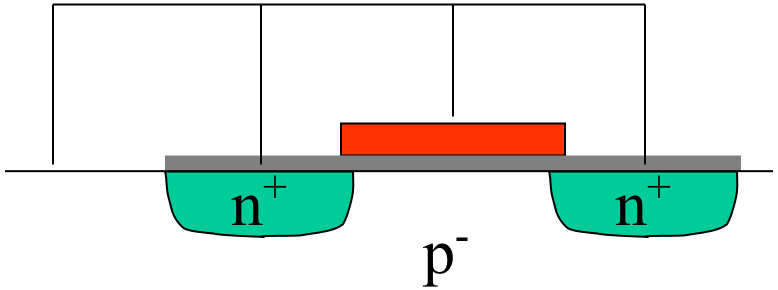


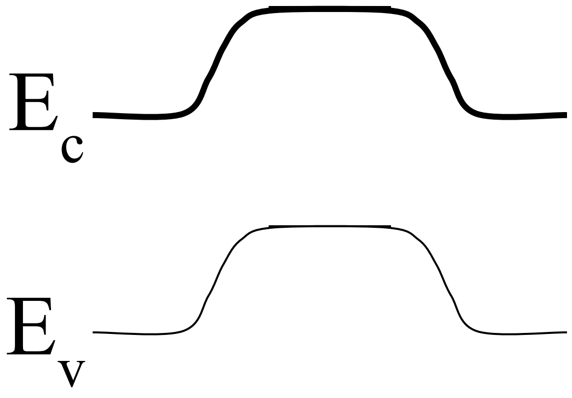
Source
(of electrons)
At lower voltage

pFET terminology



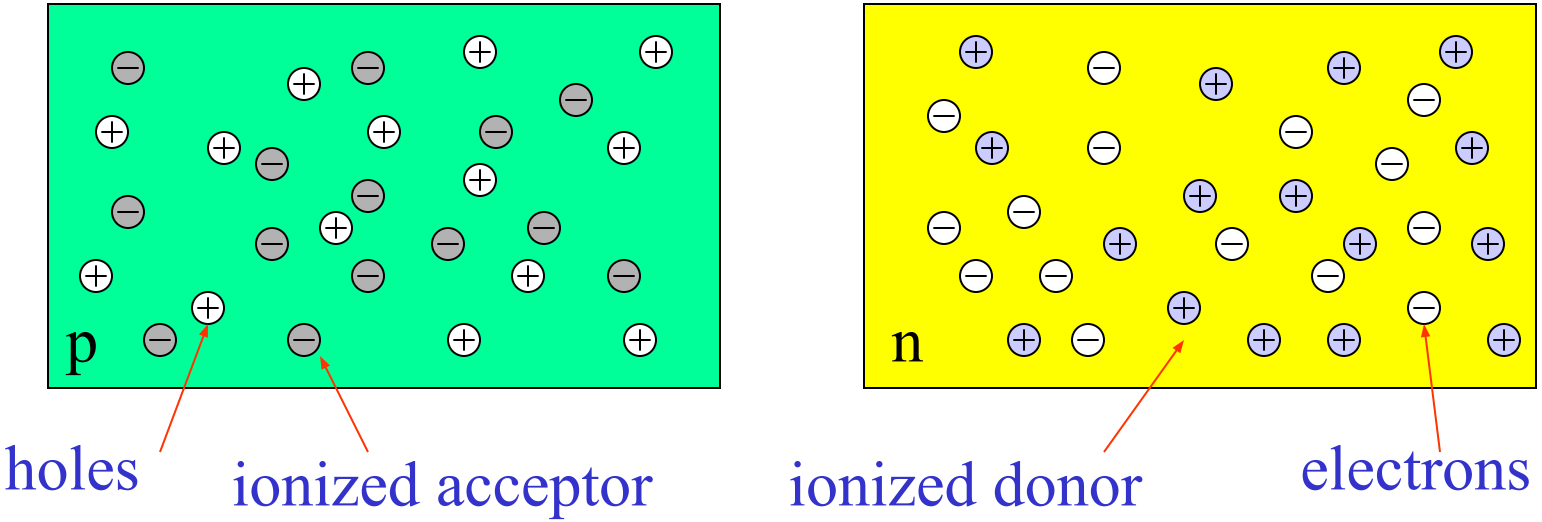
The built-in potential barrier in a FET channel arise from *pn* junction potentials





Review on Semiconductors

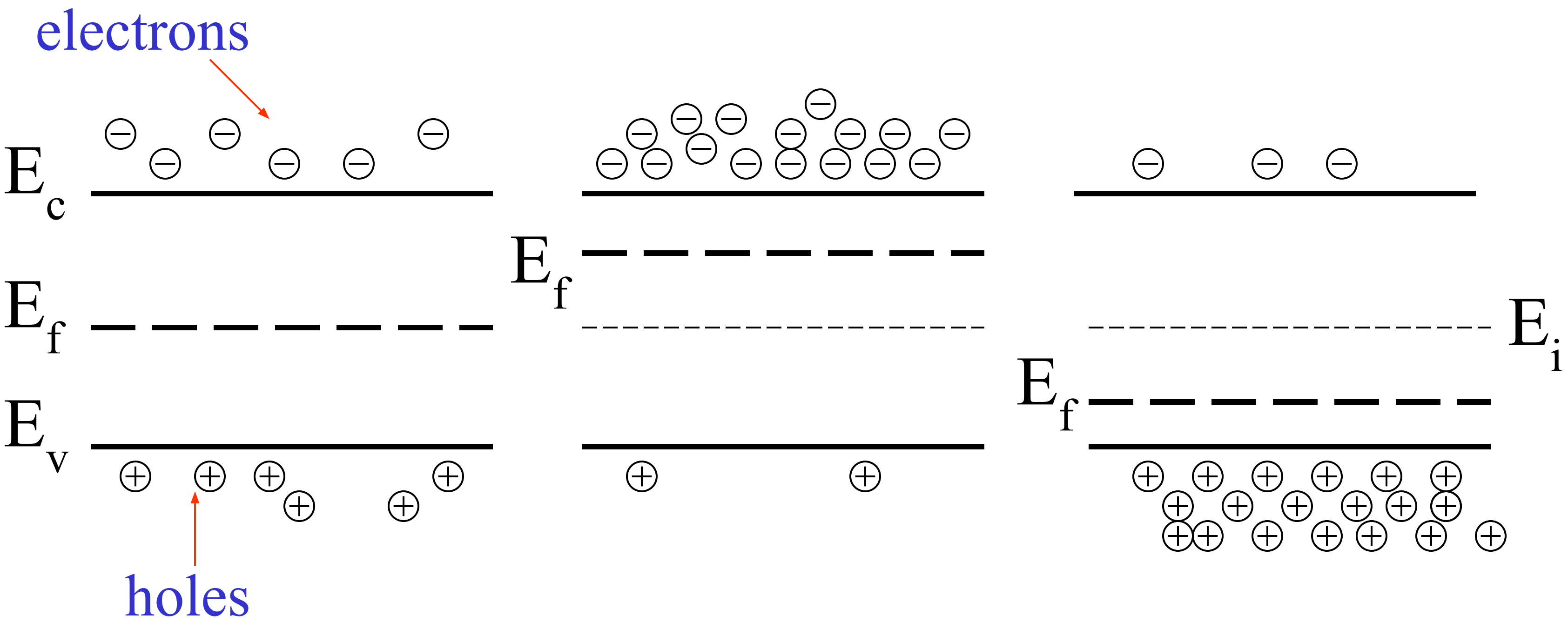
Intrinsic silicon is undoped
Extrinsic silicon is doped



Majority carriers are holes
Minority carriers are electrons

Majority carriers are electrons
Minority carriers are holes

Review on Energy Band Diagrams



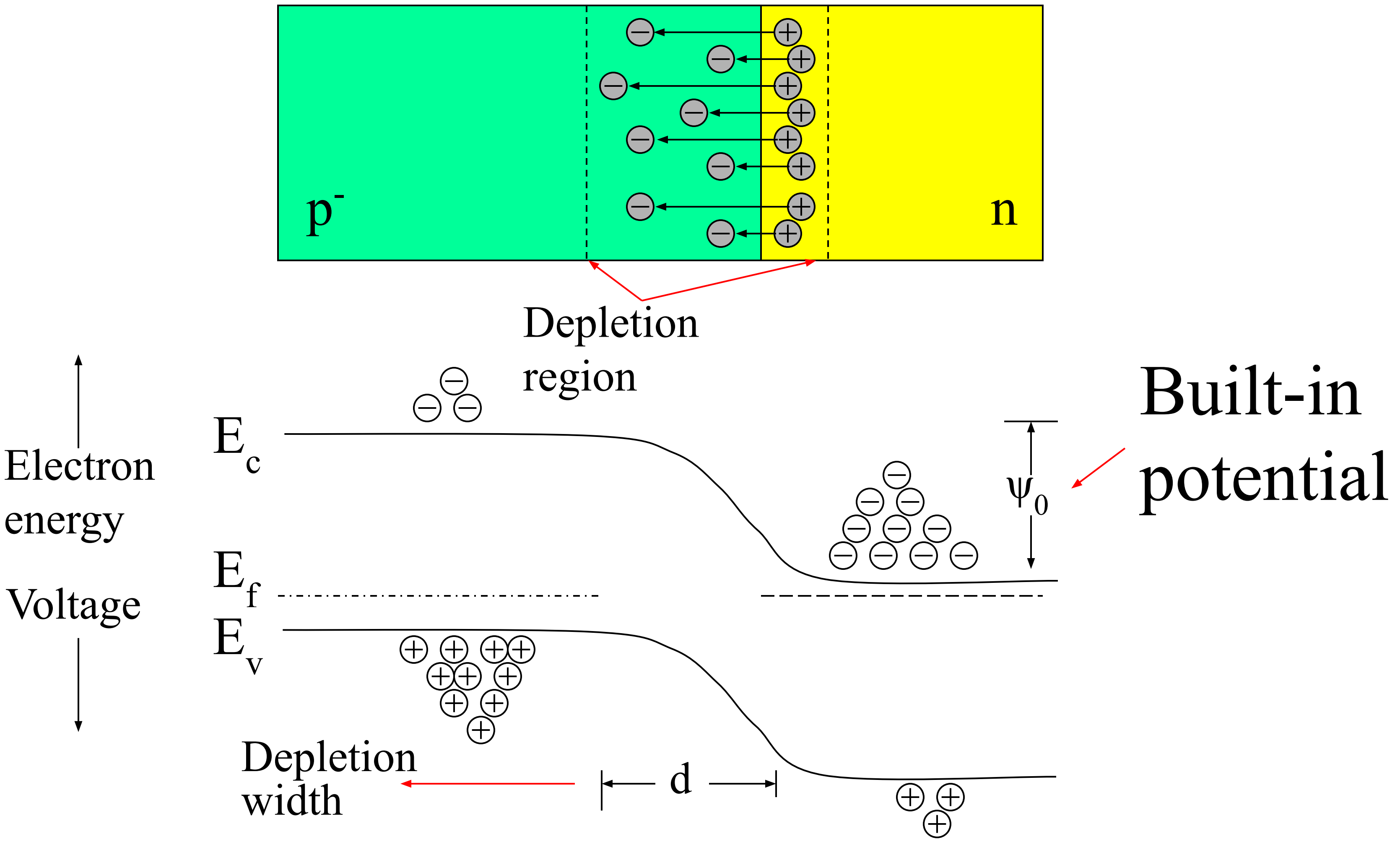
Intrinsic silicon | n-type silicon | p-type silicon

$$np = n_i^2$$

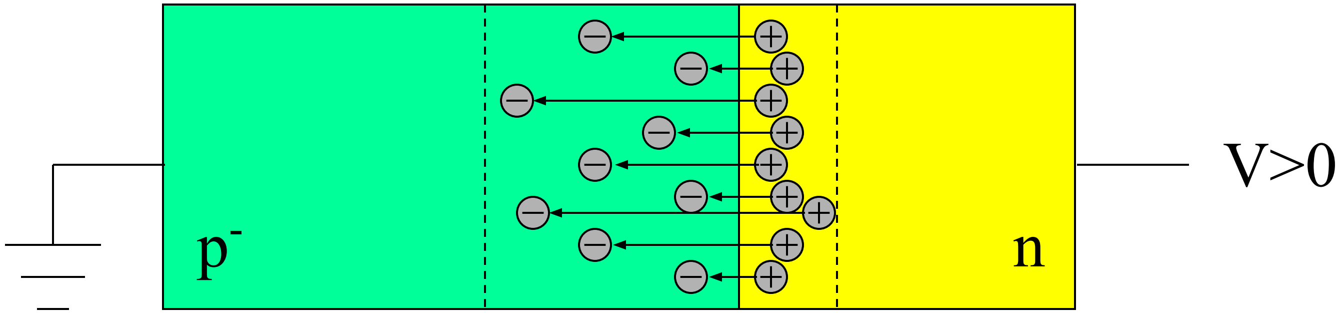
$$n = n_i \exp\left(\frac{E_f - E_i}{kT}\right) = N_D$$

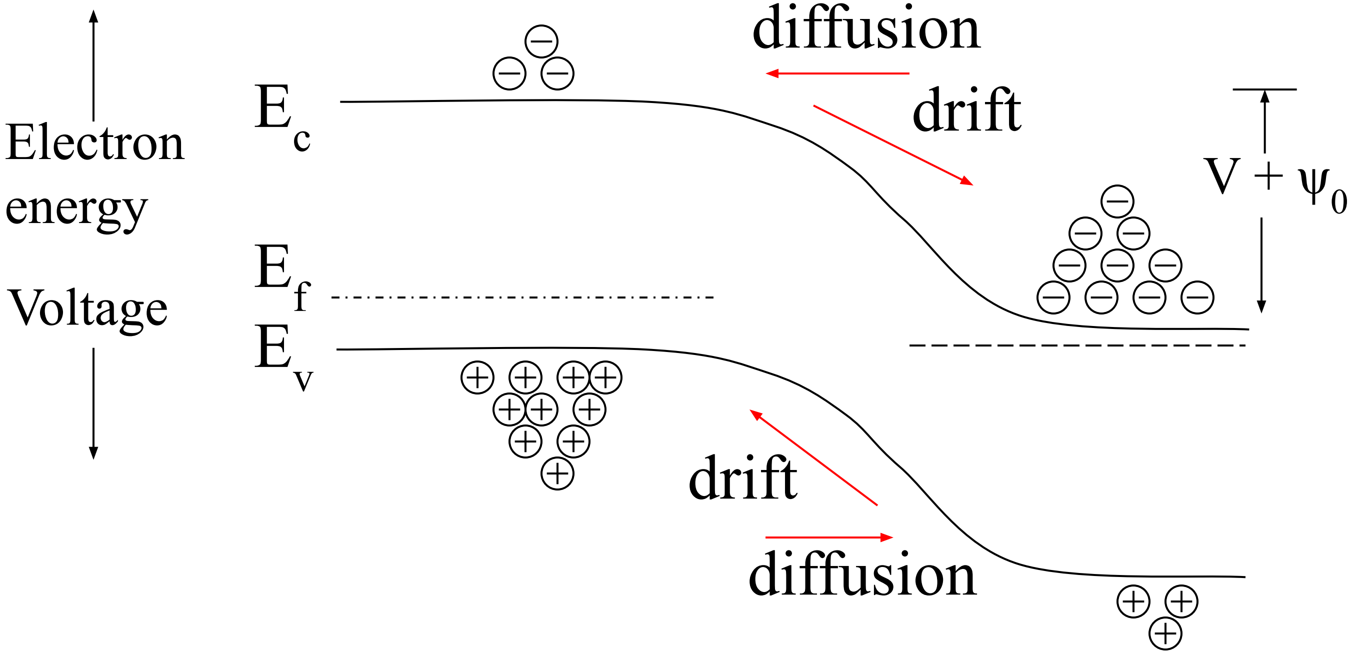
$$p = n_i \exp\left(\frac{E_i - E_f}{kT}\right) = N_A$$

Equilibrium in a p-n Junction



Reverse-biased p-n Junction





Electrostatics in 1-D

Relationship between E-field and charge density (Gauss' Law)

$$\frac{\partial \xi}{\partial x} = \frac{\rho}{\varepsilon}$$

Charge density

Permittivity

Relationship between electrical potential voltage and electric field

$$\frac{\partial V}{\partial x} = -\xi$$

Electric field

Electric field boundary condition at a dielectric interface

$$\varepsilon_1 \xi_1 = \varepsilon_2 \xi_2$$

Mechanisms of Carrier Transport

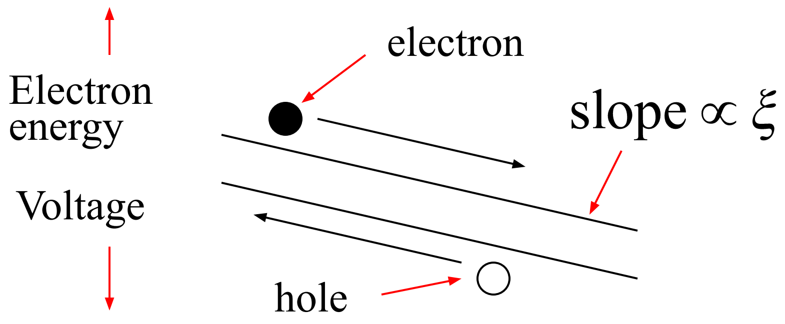
<u>Drift</u>: Movement of charge carriers due to an external field

Current $\rightarrow$ $I = qN\mu\xi$

Carrier density

Carrier mobility

Electric field

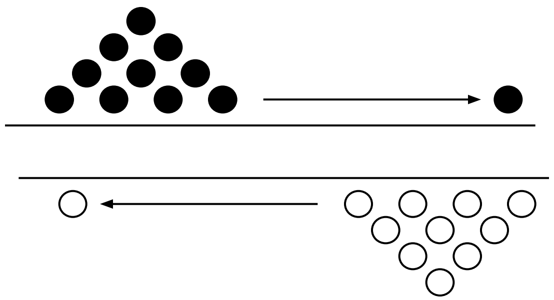


<u>Diffusion</u>: Movement of carriers due to a concentration gradient

Current $\rightarrow$ $I = -qD\dfrac{\partial N}{\partial x}$

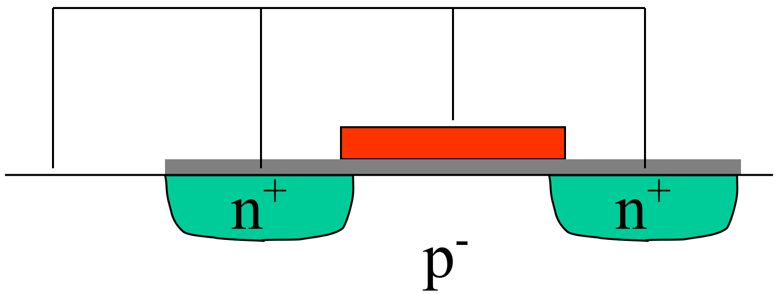
Diffusion constant

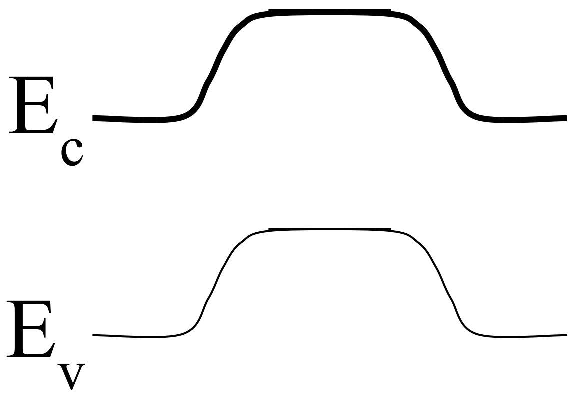
Concentration gradient

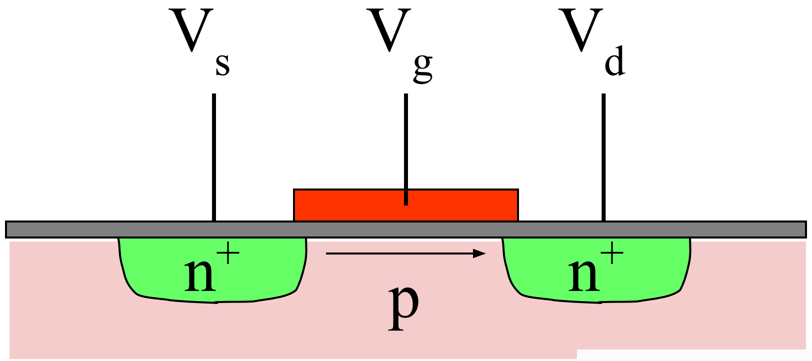


Einstein relation: $\boxed{D = \dfrac{kT}{q}\mu}$

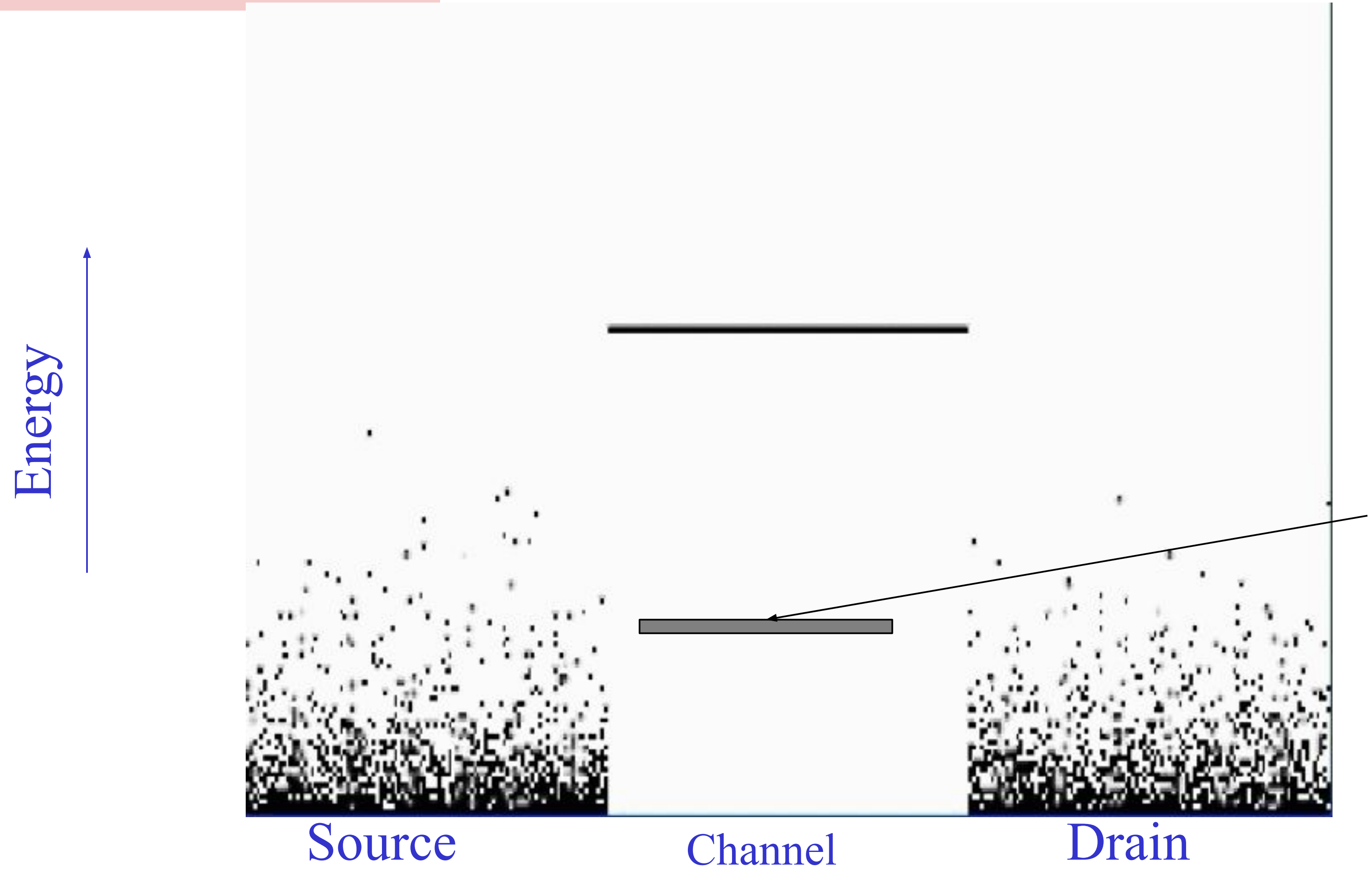
The built-in potentials in the *pn* junctions create an *energy barrier*. In *subthreshold,* controlling the barrier height controls the diffusion current.







Small Vgs, Vds=0



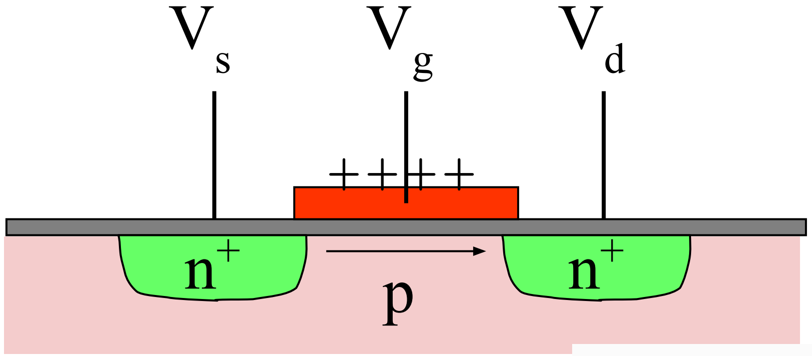
Is the scale realistic at room temperature?

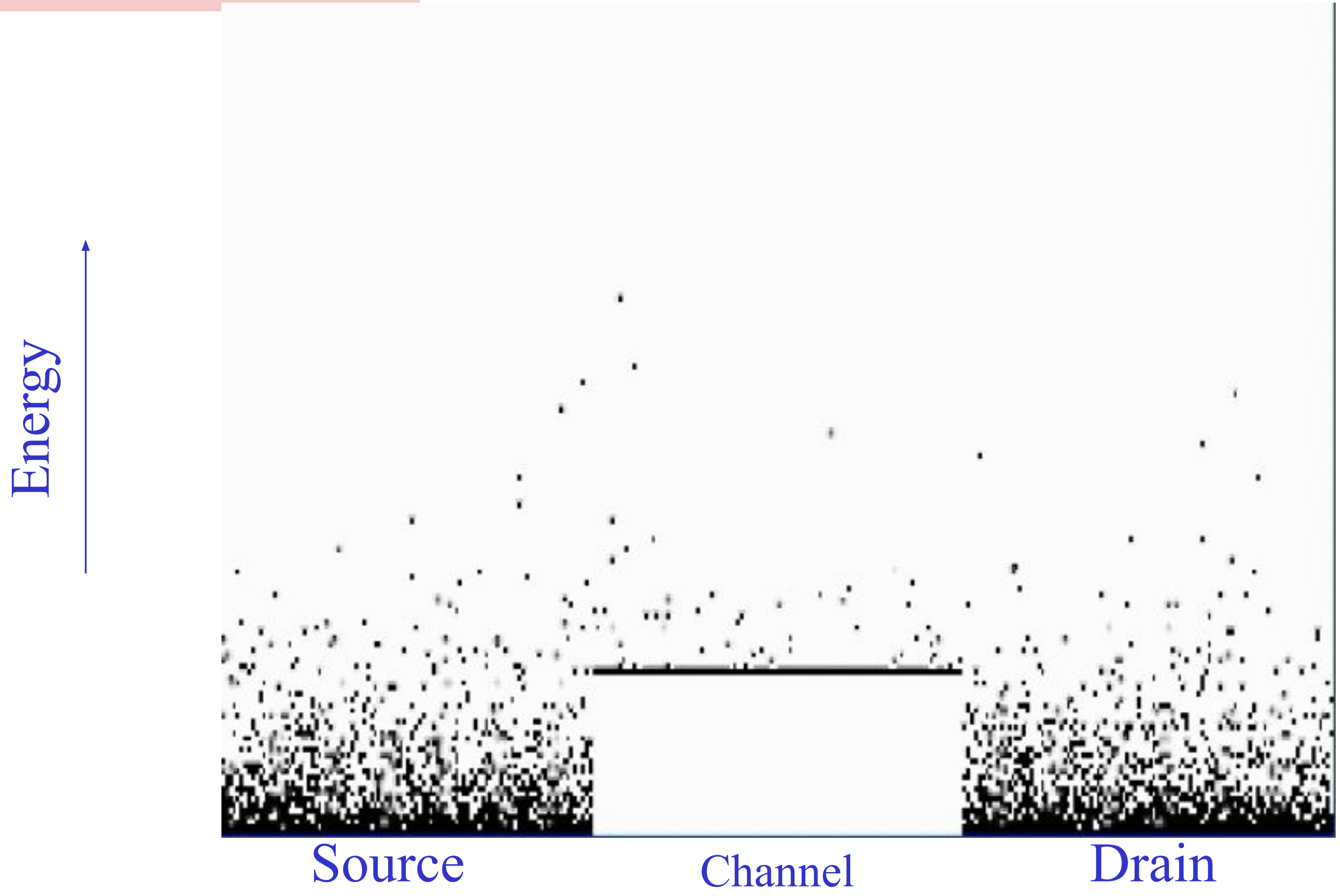
No!

Barrier height is about 700mV

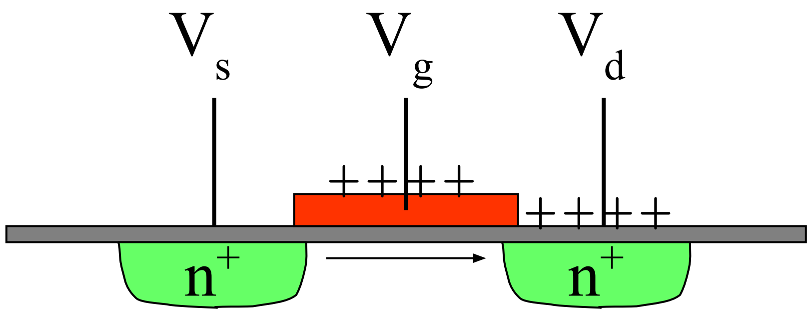
kT/c (the e-folding voltage for concentration) is only 25mV, which is 30X smaller.

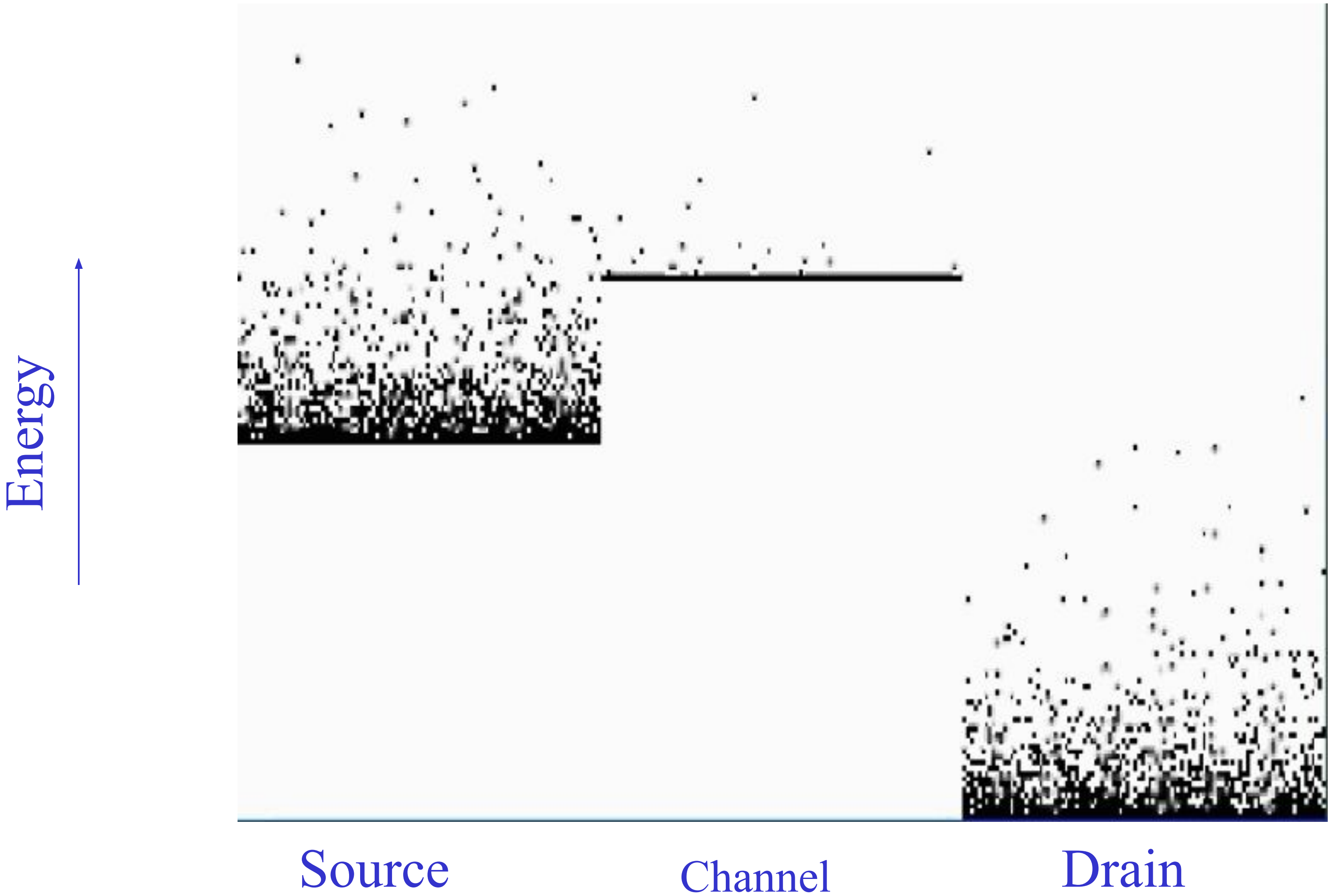
Larger Vgs, Vds=0





Larger Vgs, Large Vds





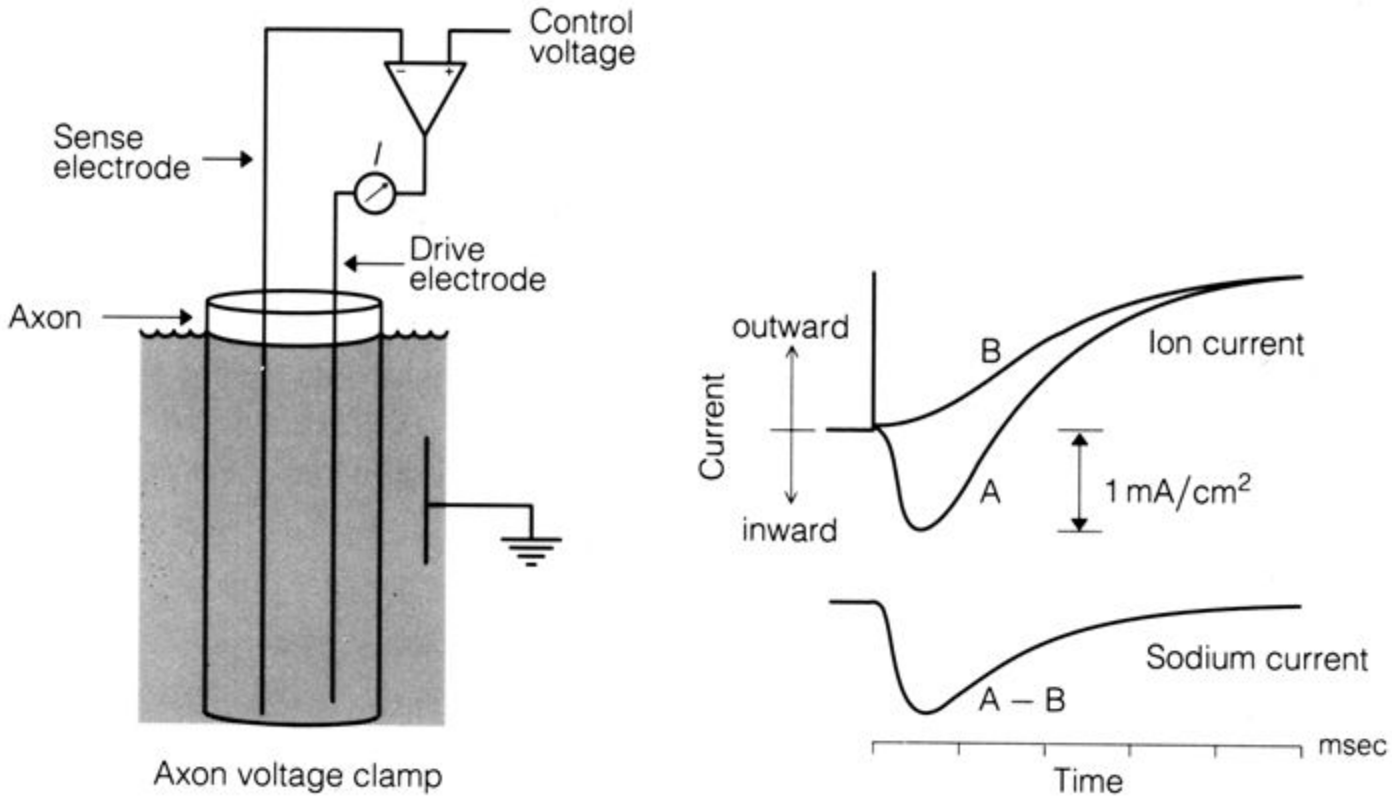
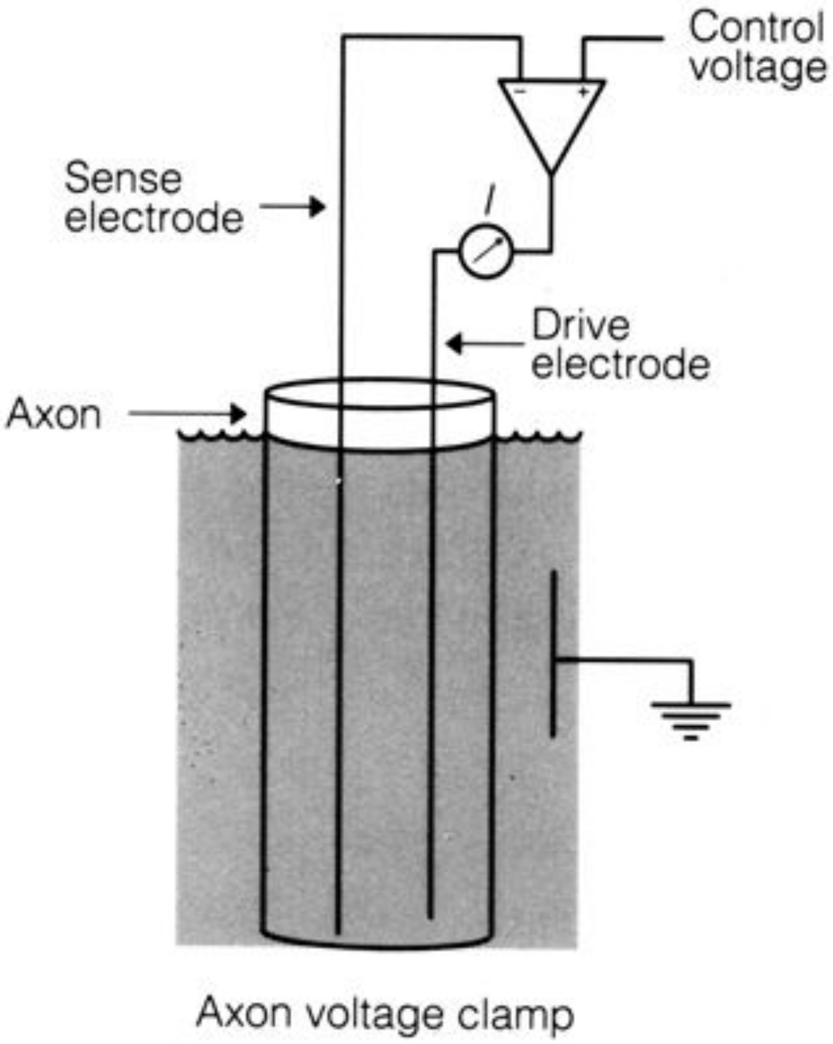
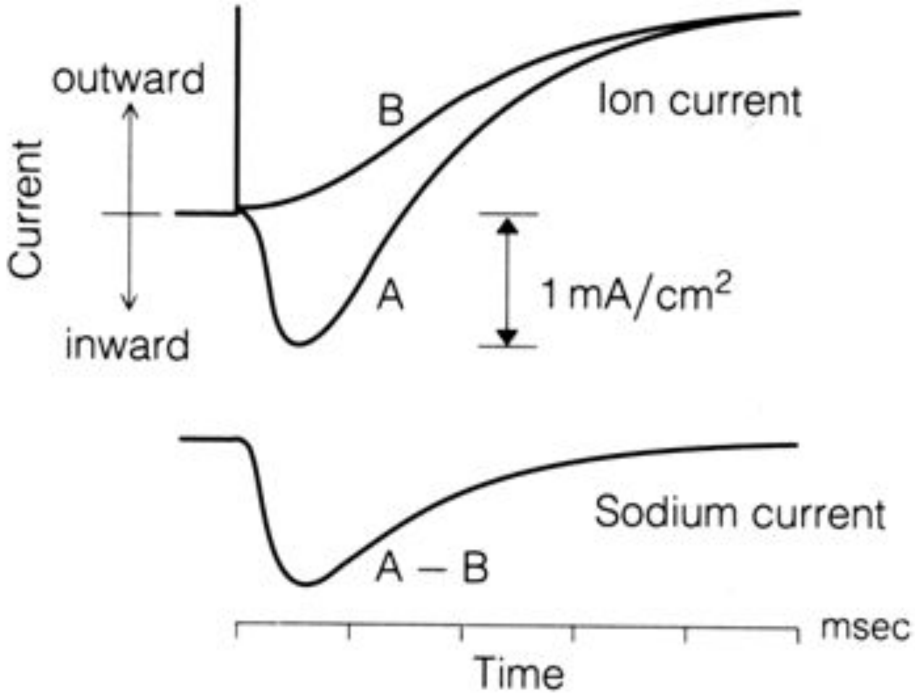


FIGURE 4.5 Schematic of the arrangement used by Hodgkin, Huxley, and Katz to measure the current through the membrane of a squid axon under conditions where the membrane potential was controlled precisely. The sense electrode assumes the potential of the cytoplasm. The amplifier generates a current *I* proportional to the difference between the actual potential and the desired potential. This current is in the direction to move the actual potential toward the desired value. The current is sensed by an oscilloscope, shown as a meter on the diagram; the extracellular fluid is ground for the entire arrangement. (*Source:* [Hodgkin et al., 1952a].)

The waveforms shown are a simplification of records taken, using this apparatus, for a step increase in membrane potential. The initial transient is the current required to charge the membrane capacitance. Curve A is the total current as a function of time. Curve B is the potassium current alone. The difference, A – B, is thus attributed to the sodium current, which rises to a maximum and then decays.

Measuring voltage-dependent nerve membrane currents





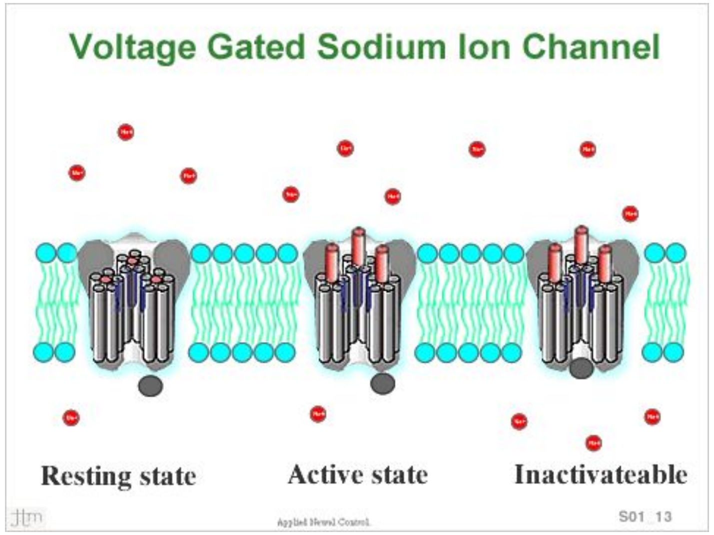
Hodgkin & Huxley 1952

Neuron channels and Transistors

Both depend on exponential Boltzmann distributions of Concentration vs. Energy.

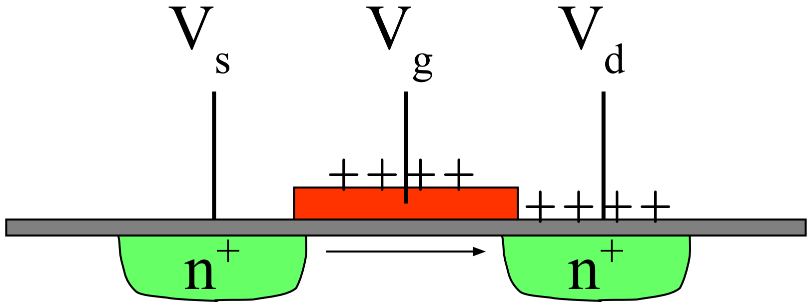
Neurons

- **Membrane ionic conductance** is exponentially dependent on the **voltage across the neuron membrane**.
- The **population of open channels** depends exponentially on potential across barrier.

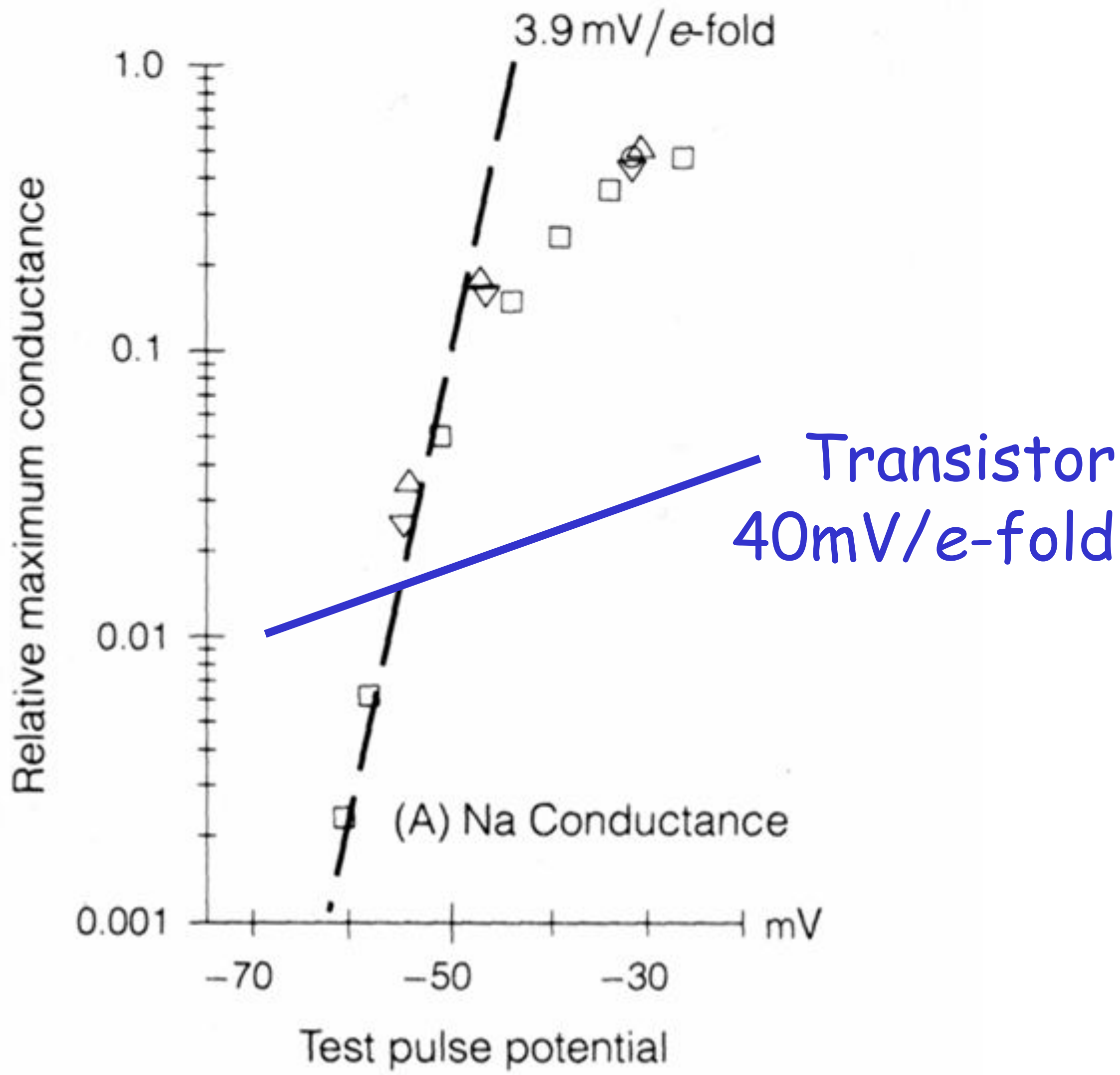


Transistors

- **Current flow** in transistors is exponentially dependent on **barrier height**.
- The **population of carriers** depends exponentially on the barrier height.



Comparing transistor and membrane channel currents



How can biology achieve such a high exponential transconductance?

Hint: Biological voltage sensitive channels carry multiple charges that sense the voltage.

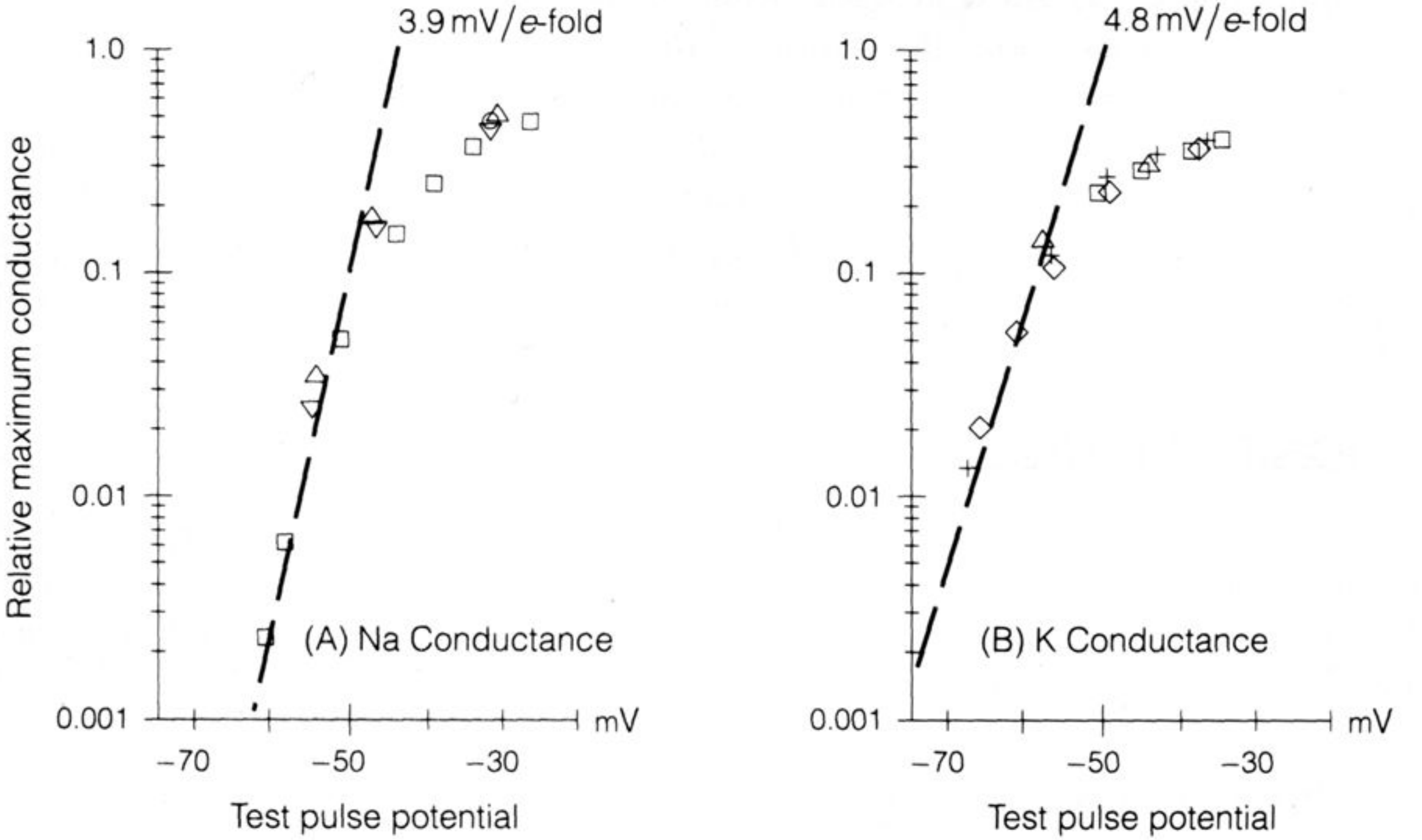
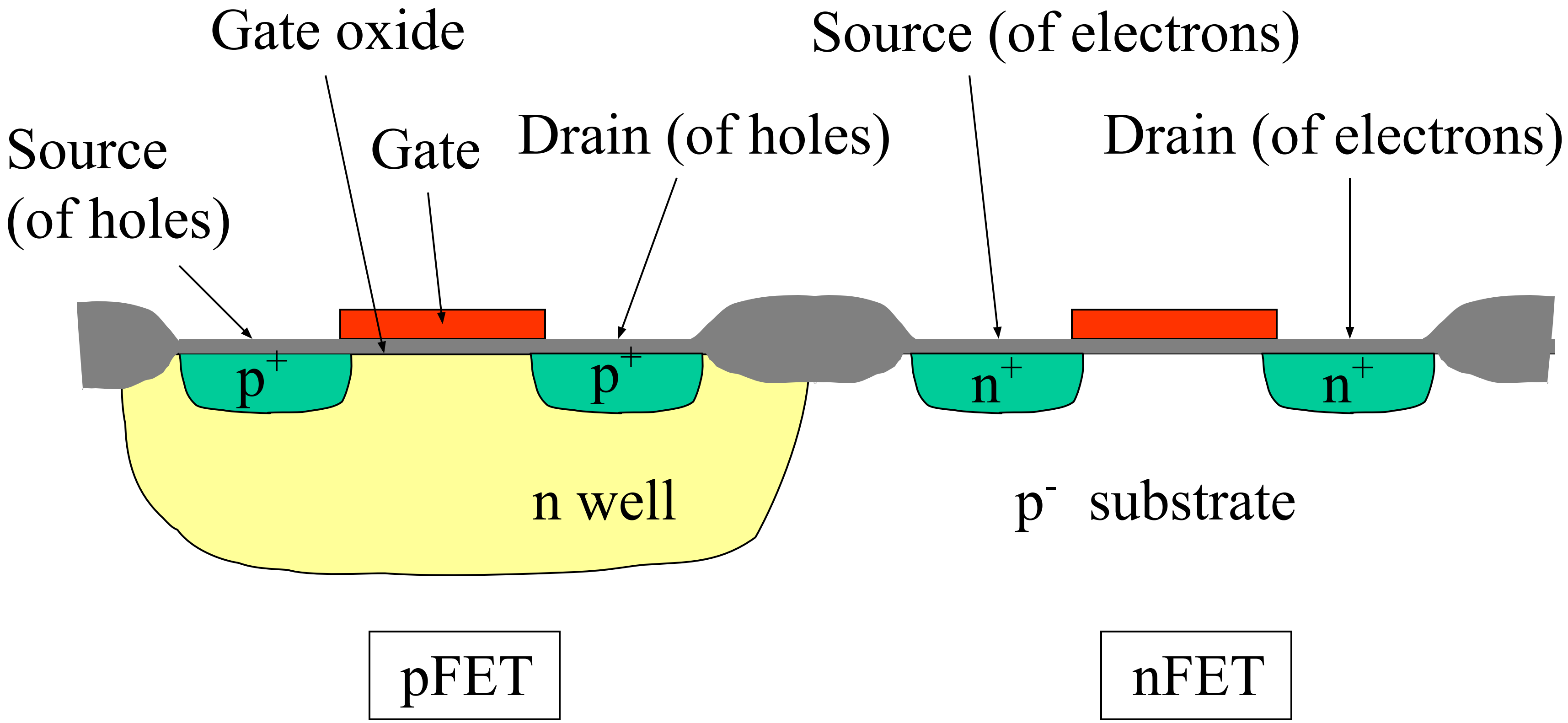
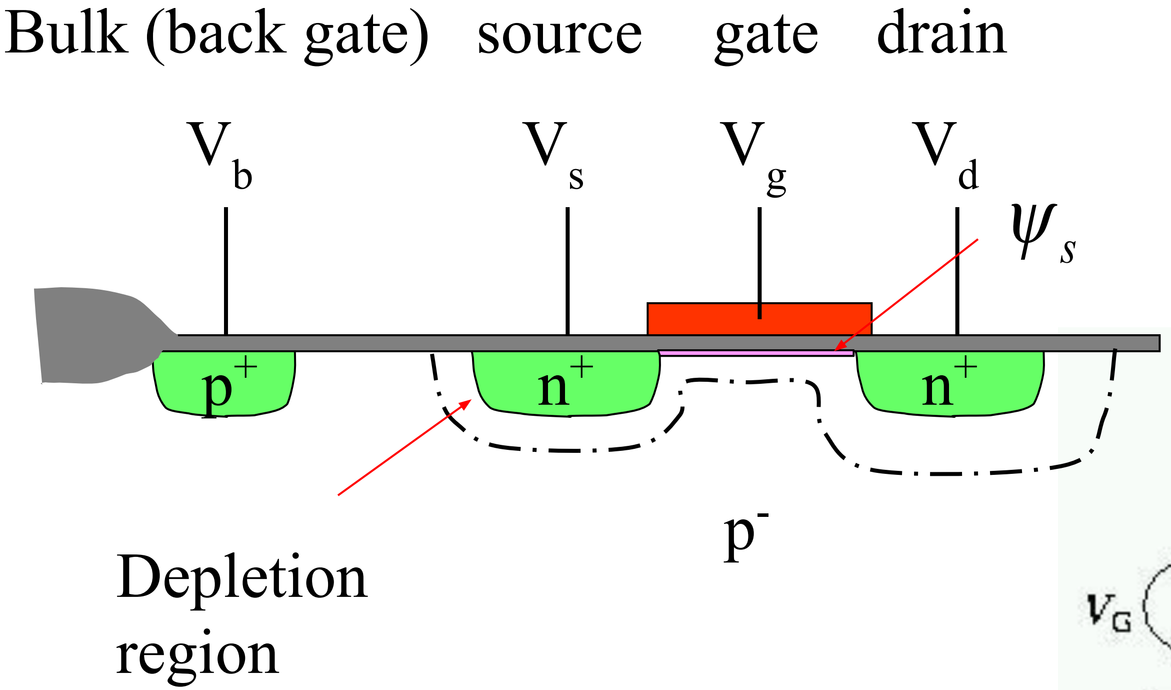


FIGURE 4.6 Exponential current–voltage characteristic of voltage-dependent channels. At high voltages, the fraction of channels that are open approaches unity, causing a saturation of the curves. (*Source:* [Hodgkin et al., 1952b, p. 464].)

Cross-section of Field-Effect Transistor (FET)

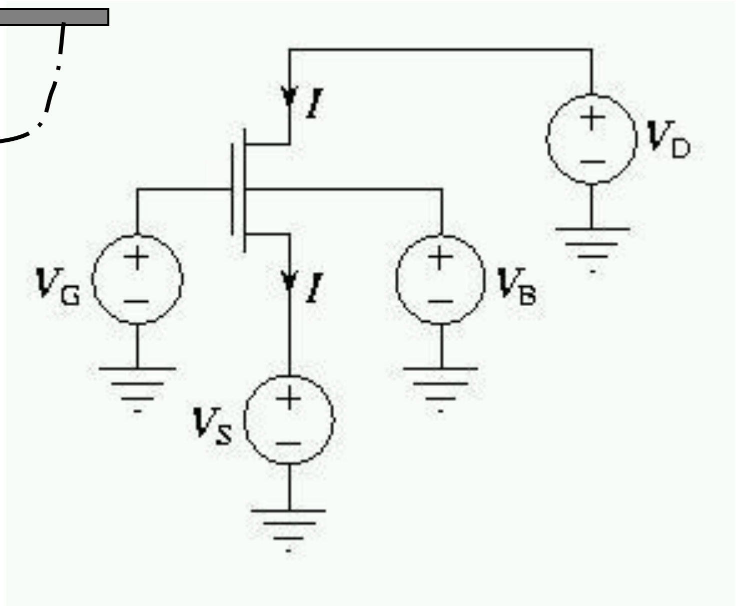


n-type MOSFET



$$V_s \geq 0V$$

$$V_g, V_d \geq V_s$$



All voltages are referenced to $V_b = 0V$

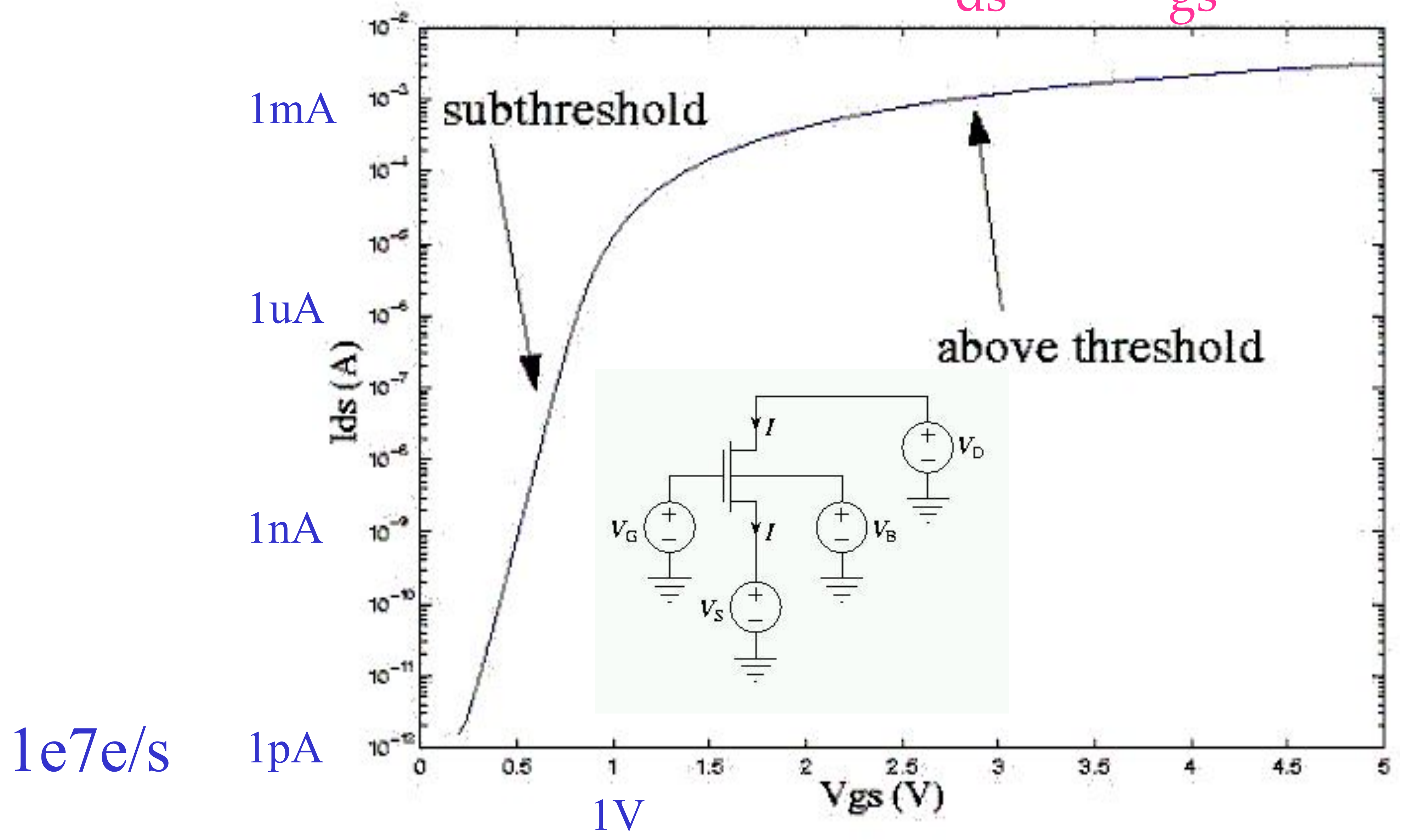
Regimes of operation for FET (dependent on V_{gs})

- <u>Cutoff</u> - Surface is accumulated
- <u>Subthreshold (Weak Inversion) Regime</u>

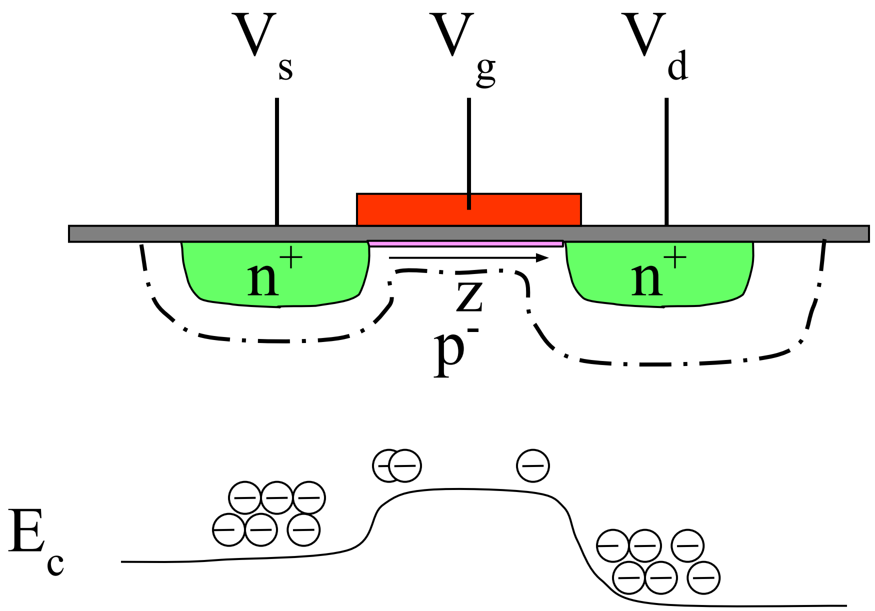
  Current flows through diffusion
- <u>Above threshold (Strong Inversion) Regime</u>

  Current flows through drift

nFet curve: I_{ds} *vs.* V_{gs}



Subthreshold nFET: Current is diffusion current



$$I = -qD\frac{dN}{dz}$$

Current (A)

Electron charge (1.6e-19C)

Diffusion Constant m^2/s

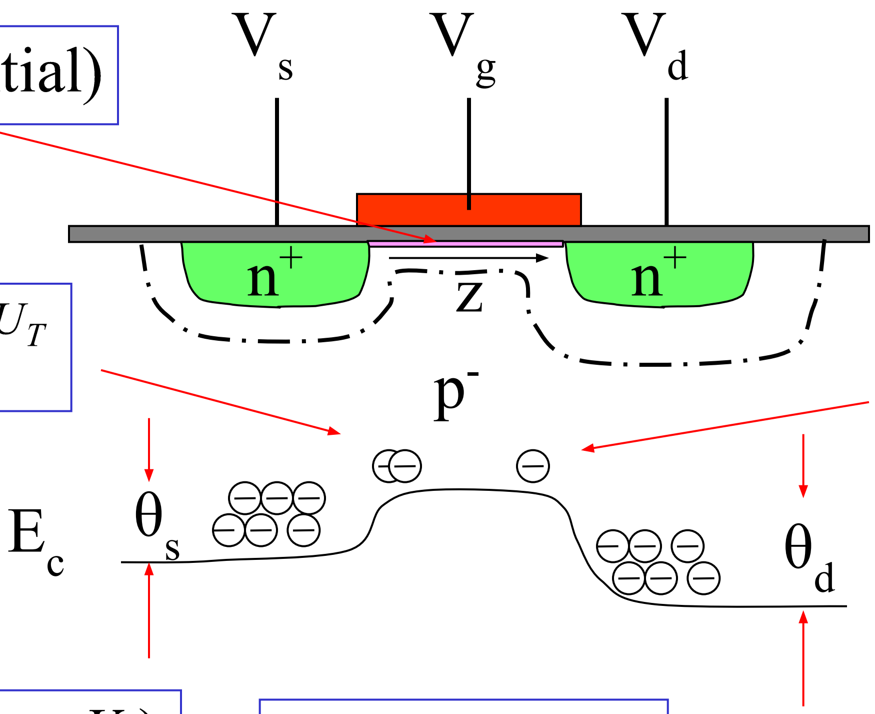
Concentration Gradient $\#/m^3/m$

N=carrier density $\#/m^3$

D=diffusion constant

Subthreshold nFET: Current is diffusion current

ψ_s (surface potential)



$$N_s = N_o e^{-\theta_s / U_T}$$

$$N_d = N_o e^{-\theta_d / U_T}$$

$$\theta_s = \theta_0 - q(\psi_s - V_s)$$
$$\theta_d = \theta_0 - q(\psi_s - V_d)$$

$$\frac{dN}{dz} = \frac{N_s - N_d}{L}$$

$$I = -qWD_n \frac{dN}{dz} = I_0 e^{\psi_s / U_T} \left(e^{-V_d / U_T} - e^{-V_s / U_T}\right)$$

Fwd Rev

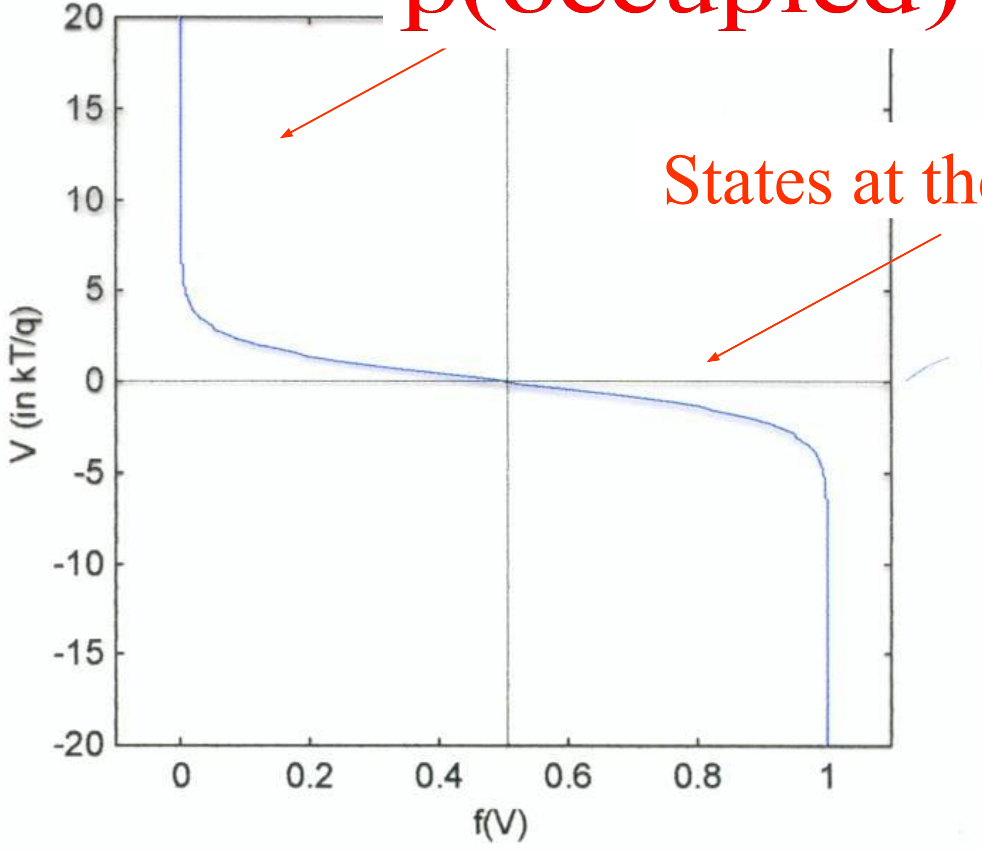
N=carrier density per unit volume
W=channel width
L=channel length
D=diffusion constant
θ_o=built-in potential barrier
N_s=built in concentration
I_0=subthreshold leakage current

Reminder: The Fermi-Dirac distribution

States above Fermi level (e.g. in conduction band) are **occupied** with Boltzmann distribution

$$\mathbf{p(occupied)} \approx e^{-(E-E_f)/kT}$$

Energy relative to Fermi level in kT units



States at the Fermi level are ½ occupied

E_f

$$\boxed{N_d = N_o e^{-\theta_d / U_T}}$$

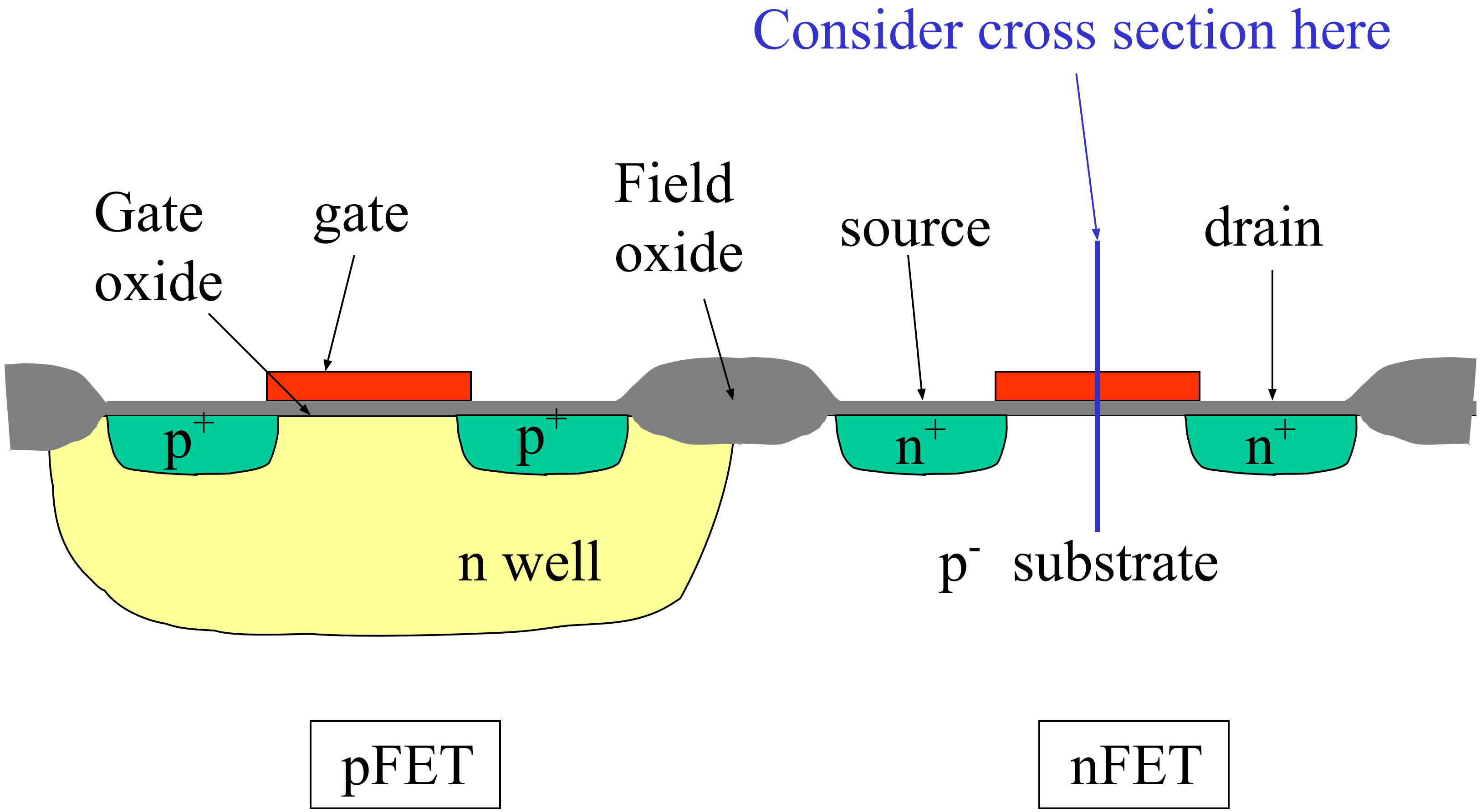
Probability of occupation of a state

We have equation for subthreshold current, but we don't directly control the surface potential

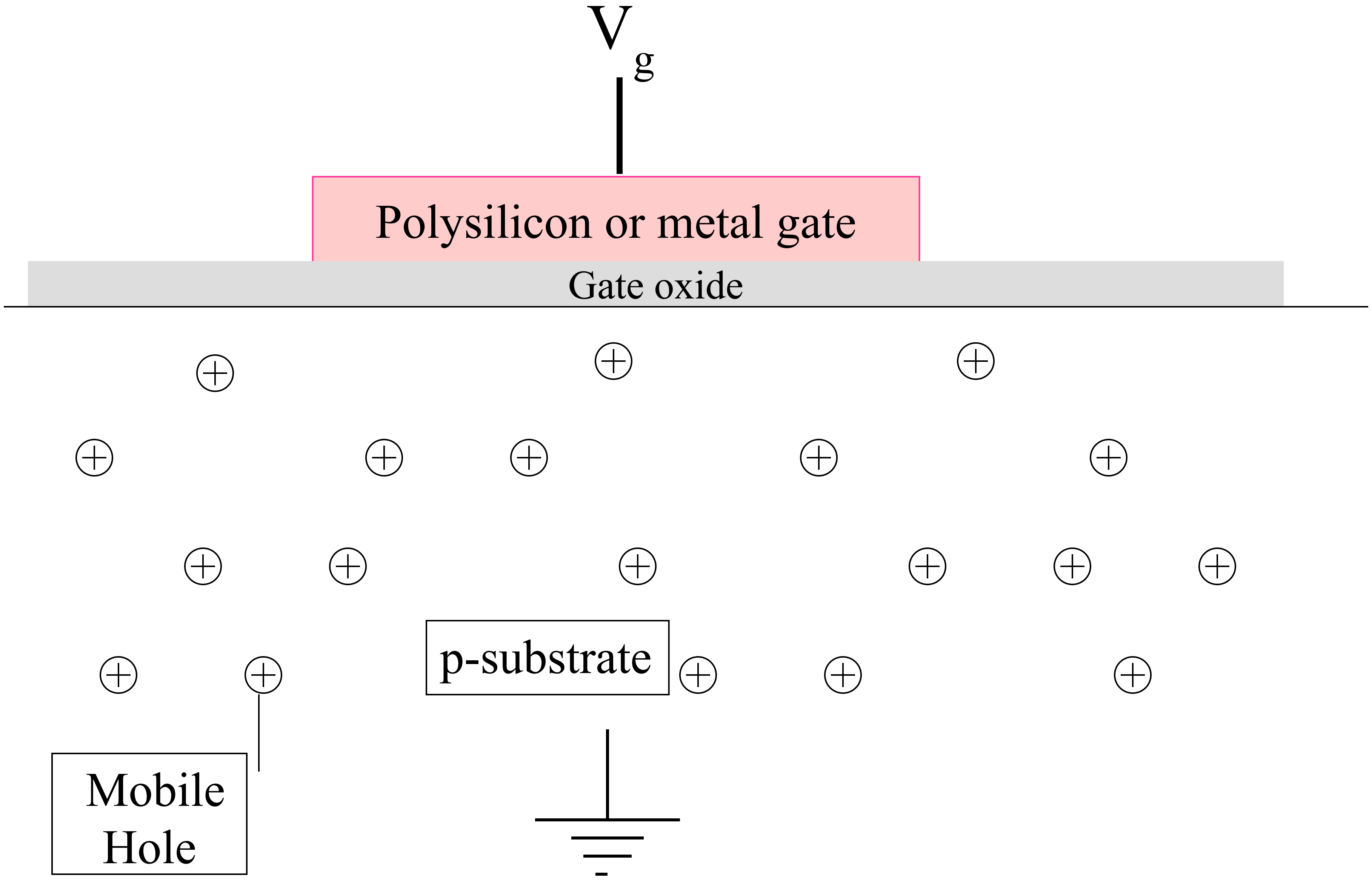
$$\boxed{I = -qWD_n \frac{dN}{dz} = I_0 e^{\psi_s / U_T} \left(e^{-V_d / U_T} - e^{-V_s / U_T}\right)}$$

How is the <u>surface potential</u> related to the <u>gate voltage</u>?

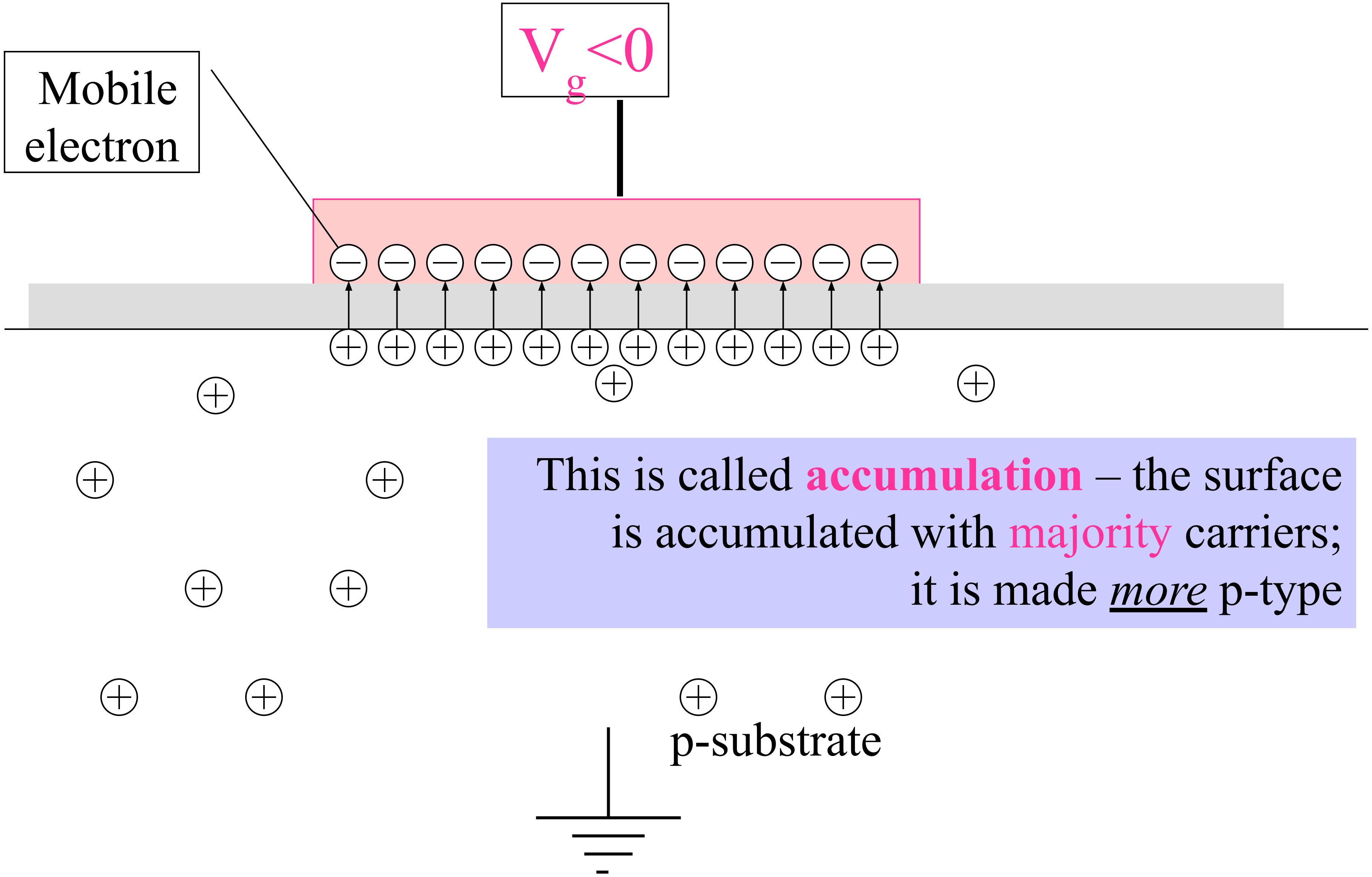
We need to understand effect of gate on surface potential



MOS capacitor structure

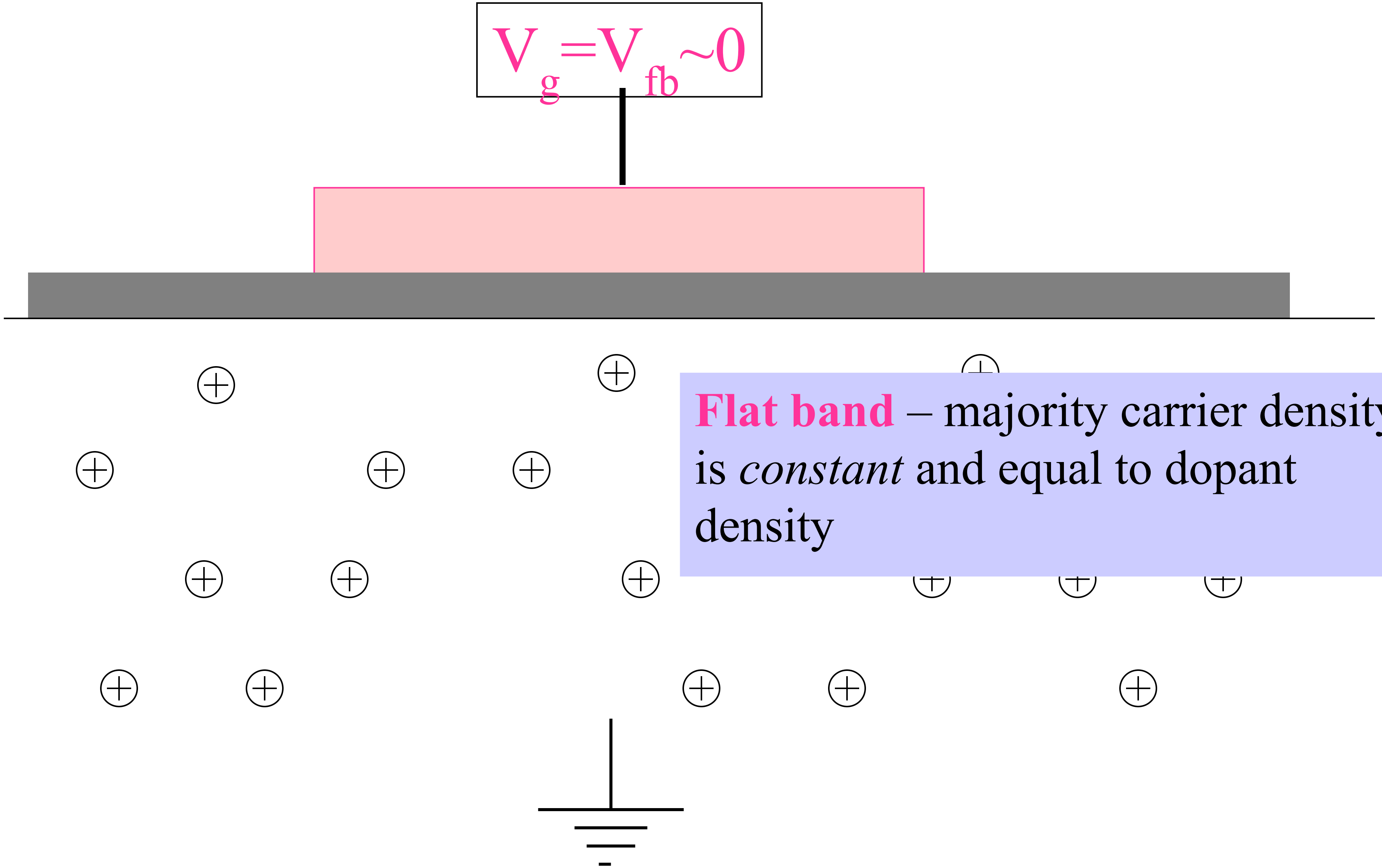


MOS capacitor structure: accumulation



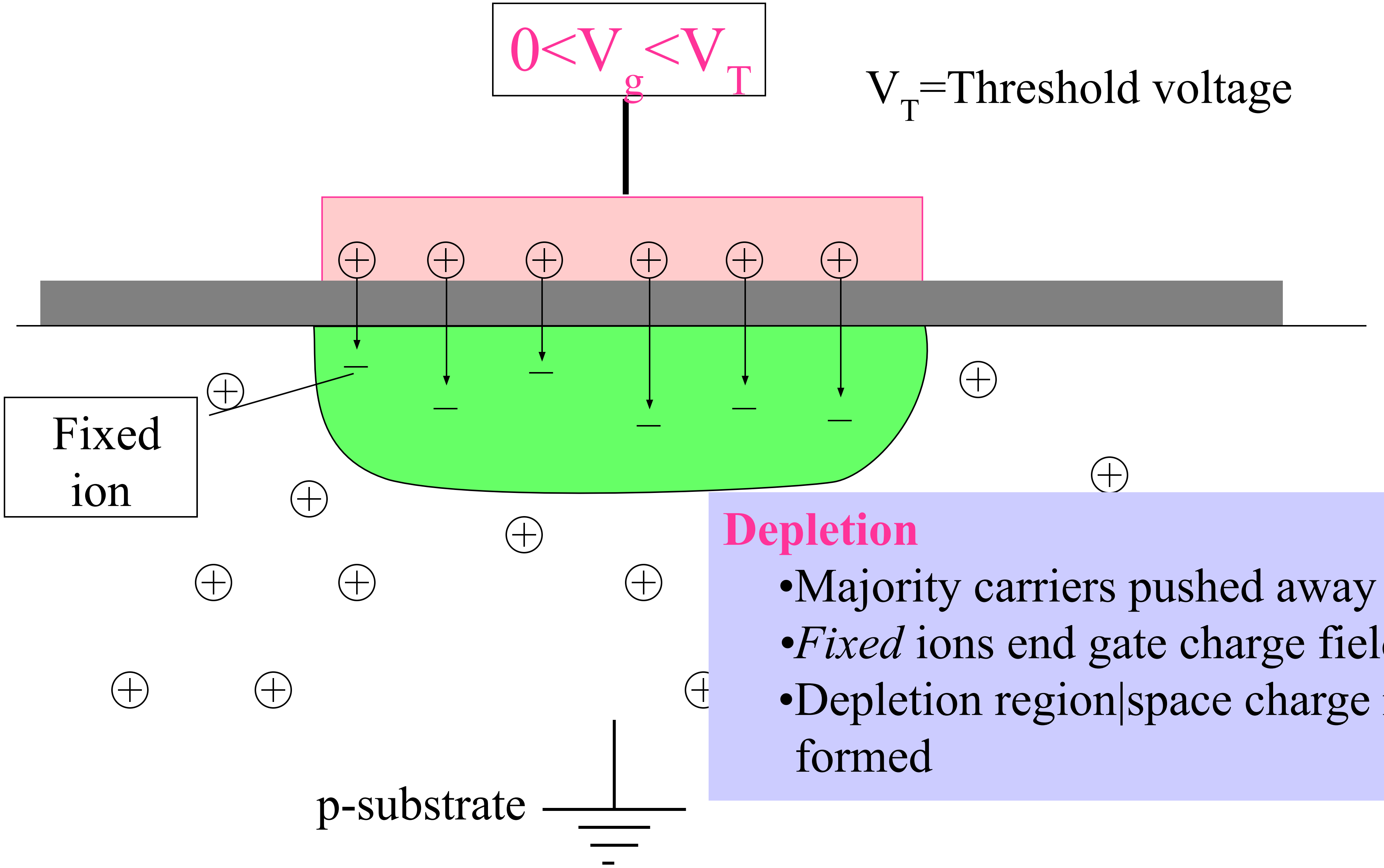
This is called **accumulation** – the surface is accumulated with majority carriers; it is made *more* p-type

MOS capacitor structure: flat band



Flat band – majority carrier density is *constant* and equal to dopant density

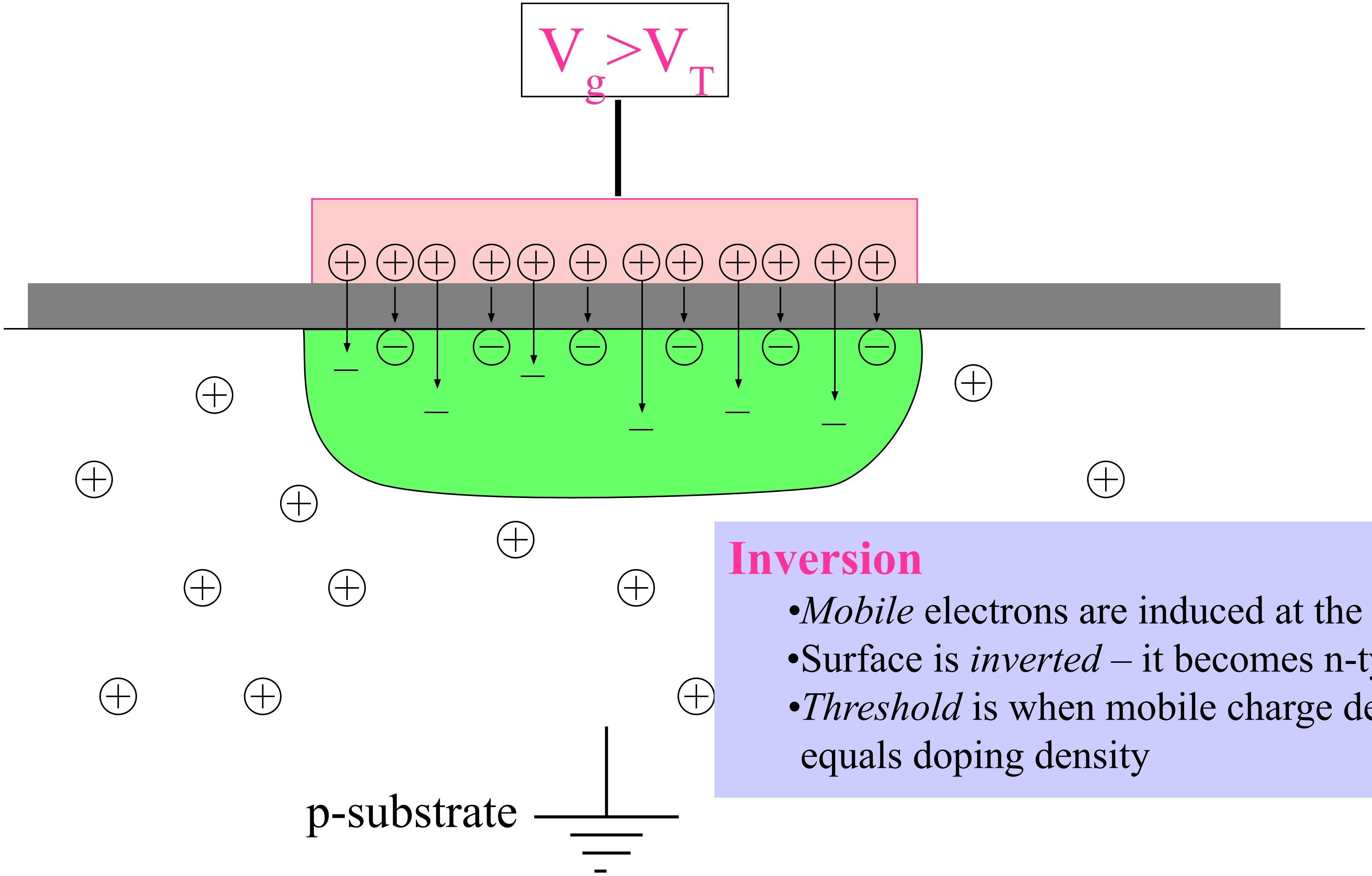
MOS capacitor structure: depletion



Depletion

- Majority carriers pushed away
- *Fixed* ions end gate charge field lines
- Depletion region|space charge region formed

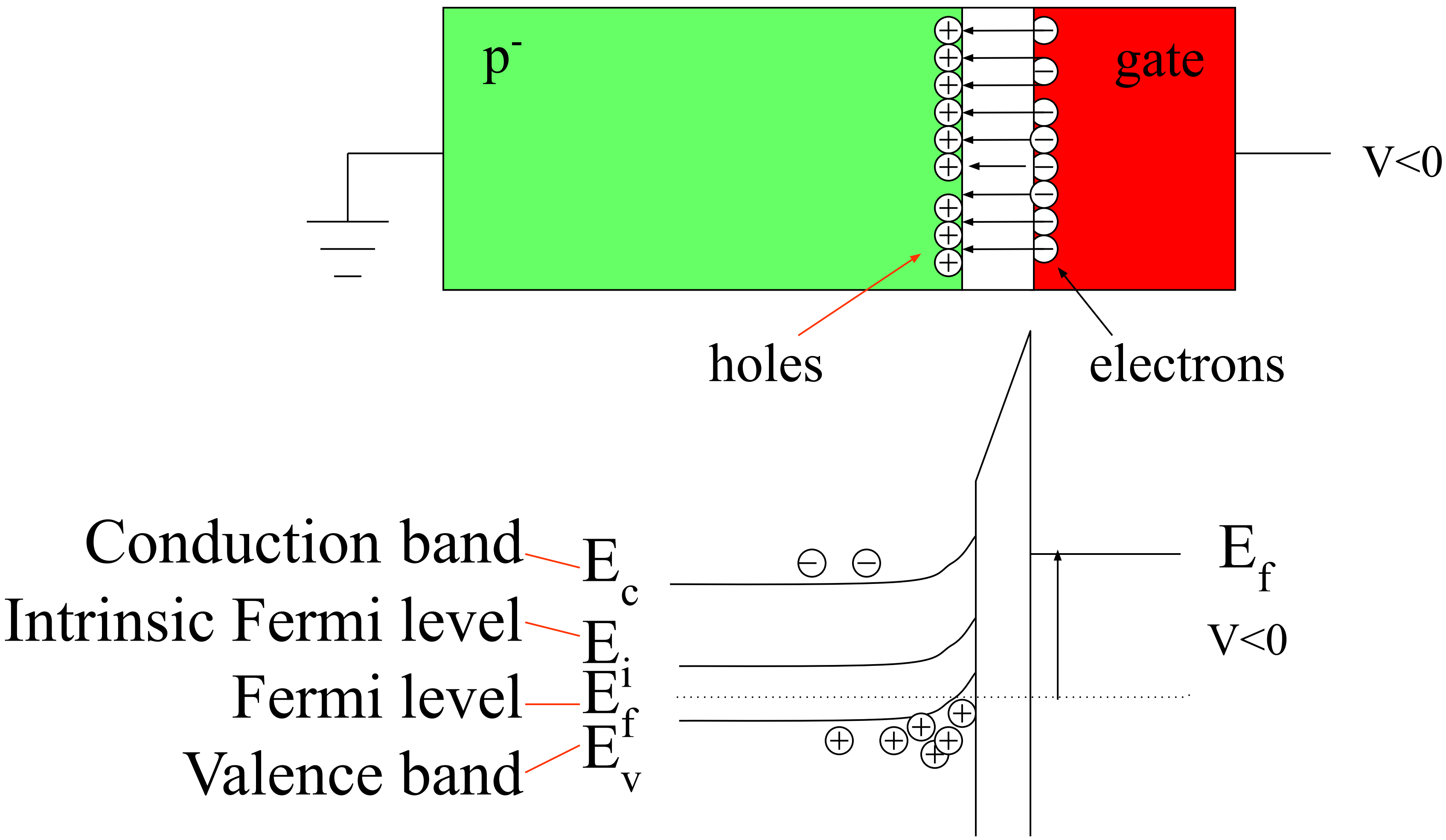
MOS capacitor structure: inversion



Inversion

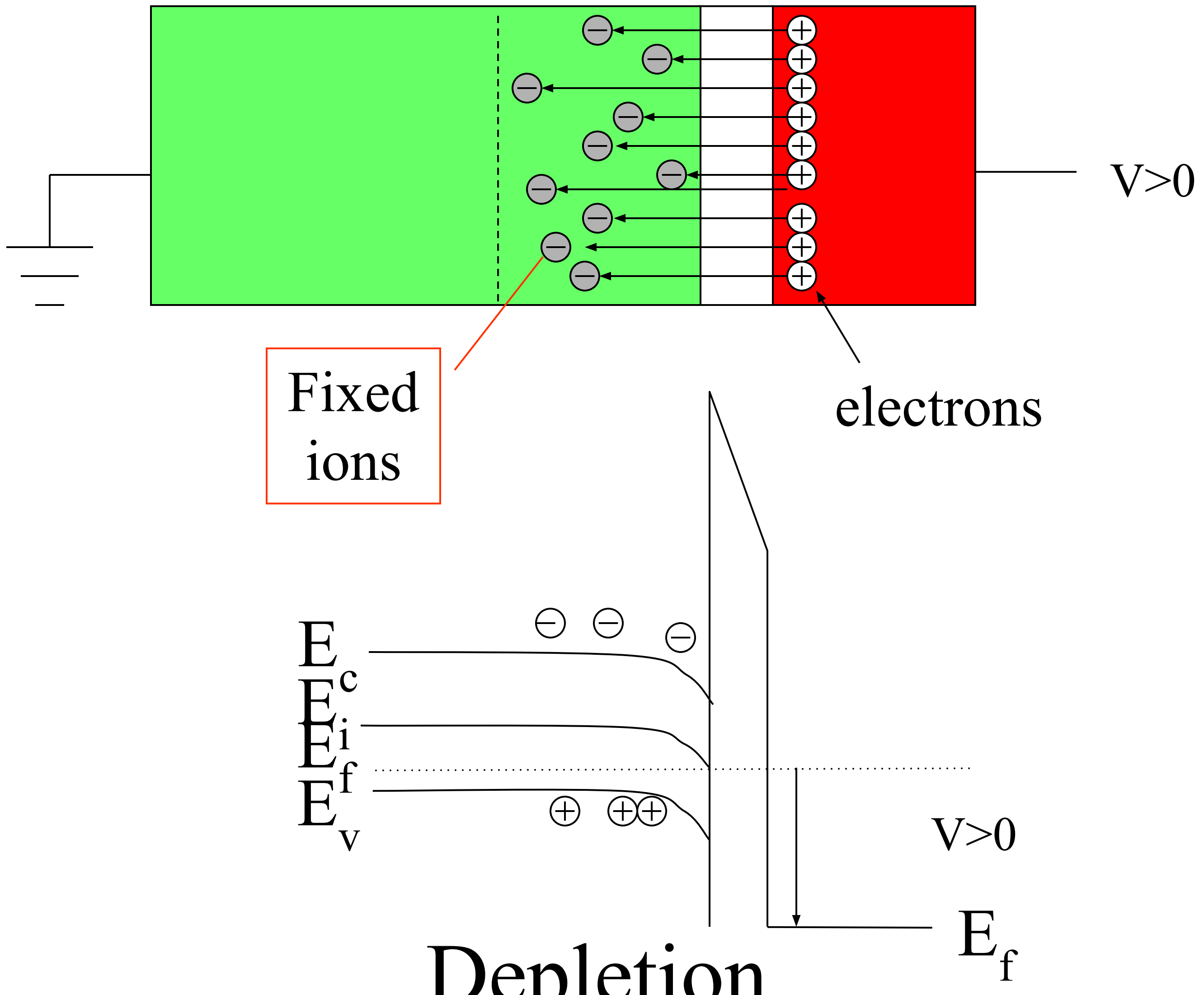
- *Mobile* electrons are induced at the surface
- Surface is *inverted* – it becomes n-type
- *Threshold* is when mobile charge density equals doping density

Electrostatics of the MOS Structure



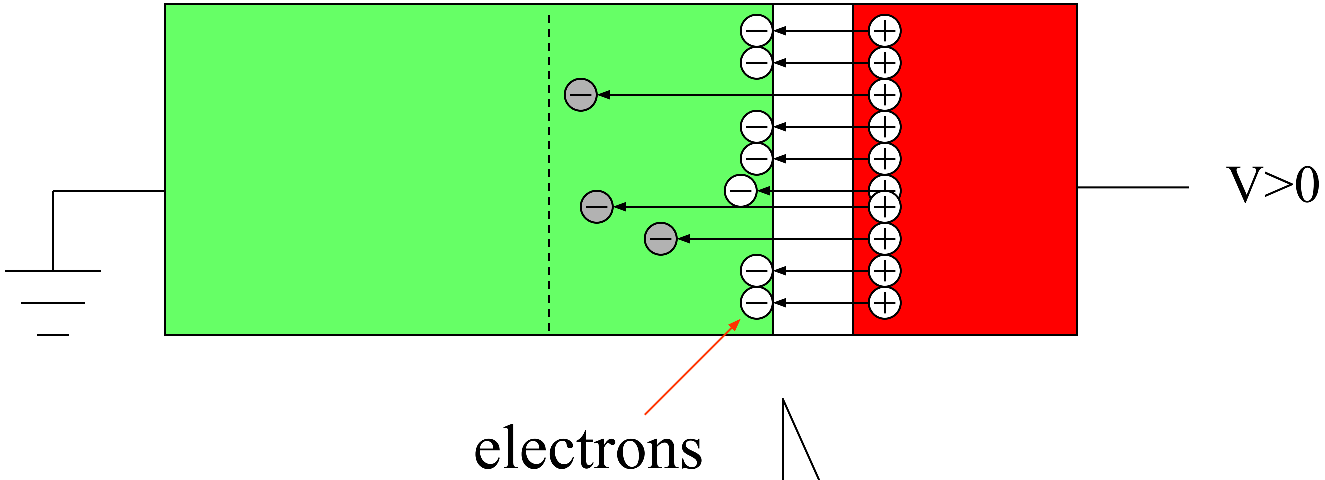
Accumulation

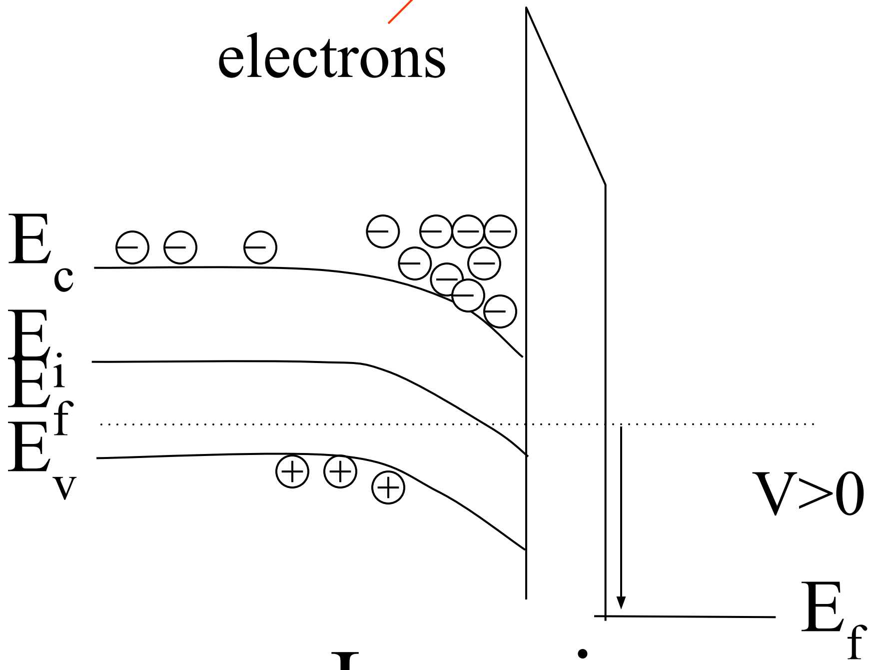
Electrostatics of the MOS Structure (II)



<u>Depletion</u>

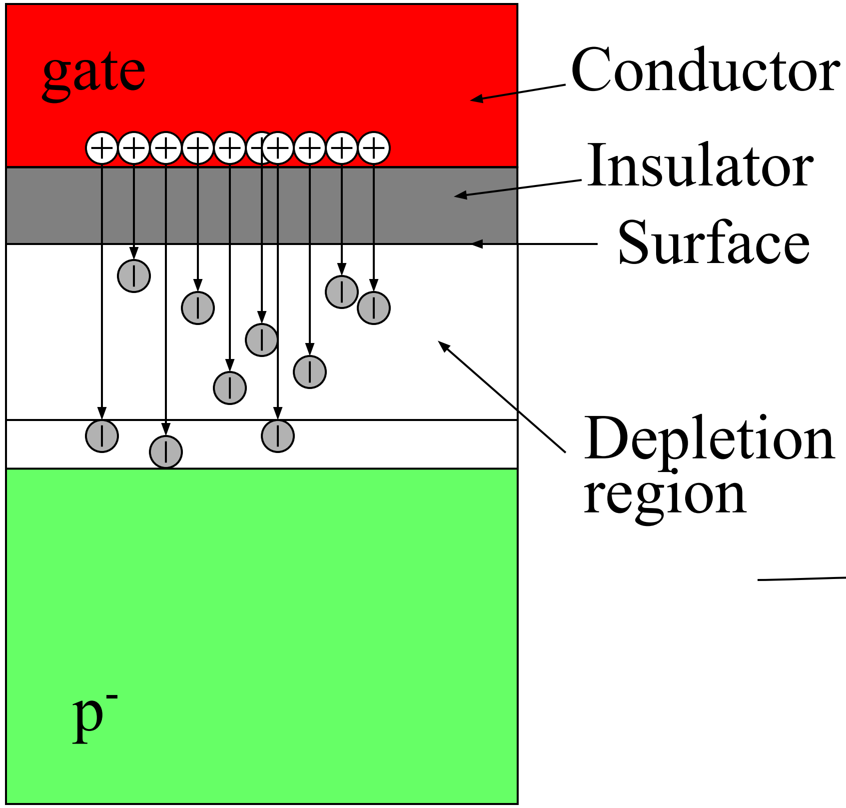
Electrostatics of MOS structure (III)



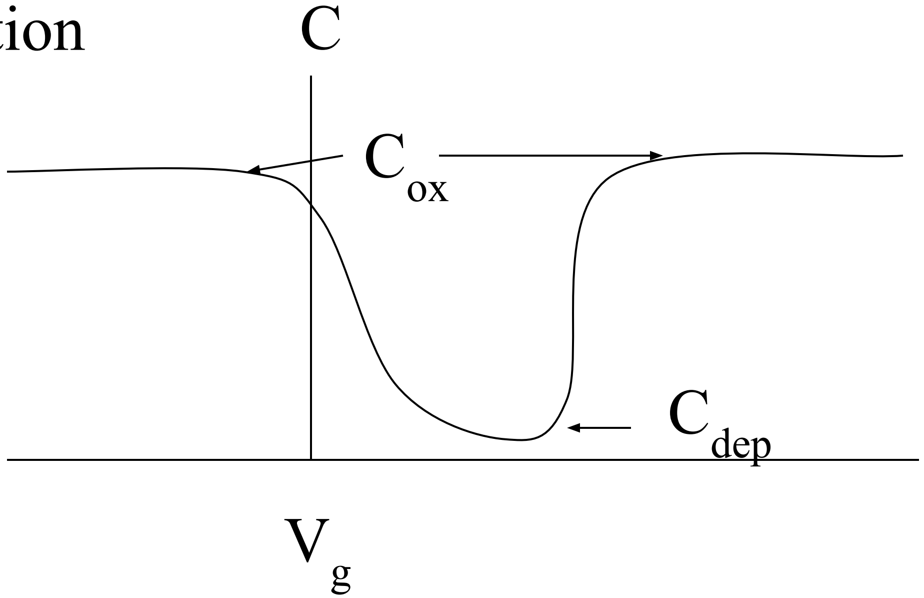


<u>**Inversion**</u>

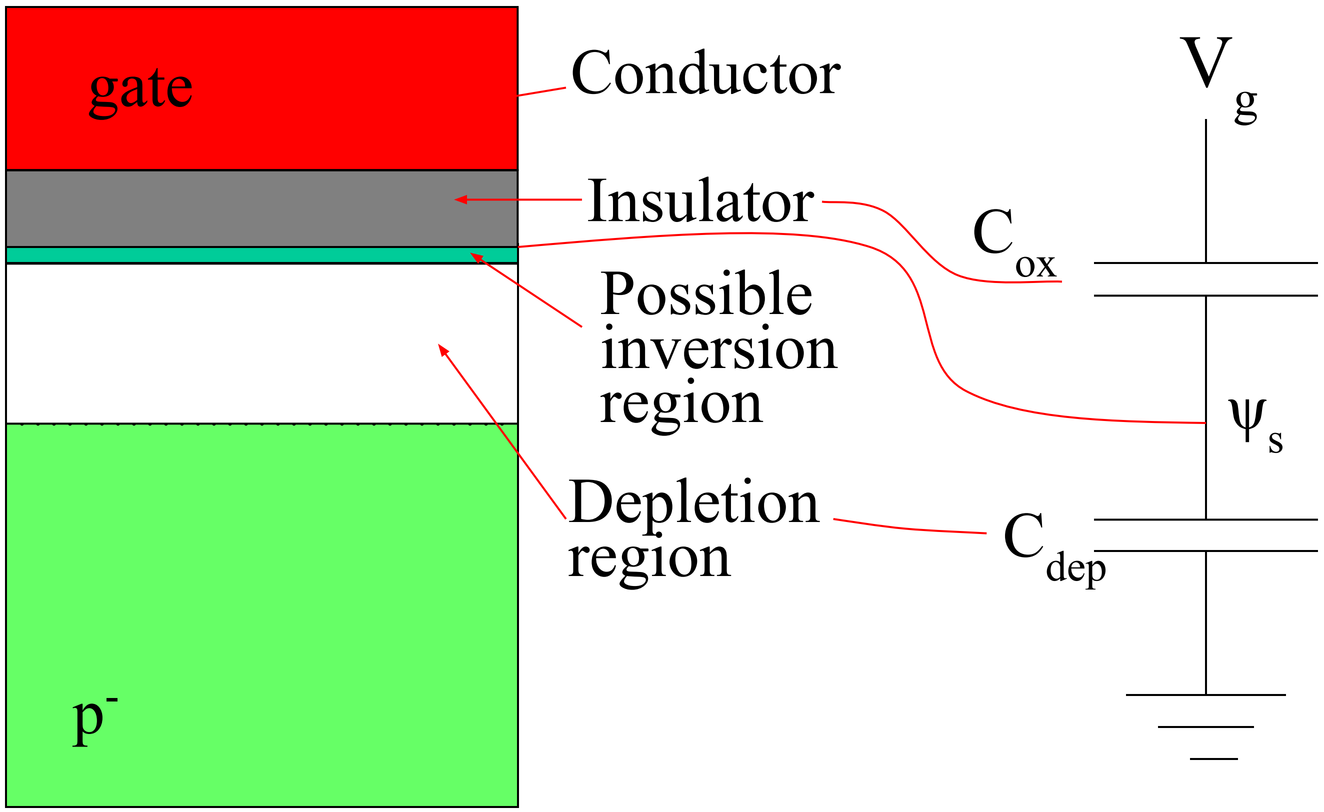
What is a *depletion capacitor*?



$$C = \frac{dQ}{dV}$$



Influence of gate on surface potential



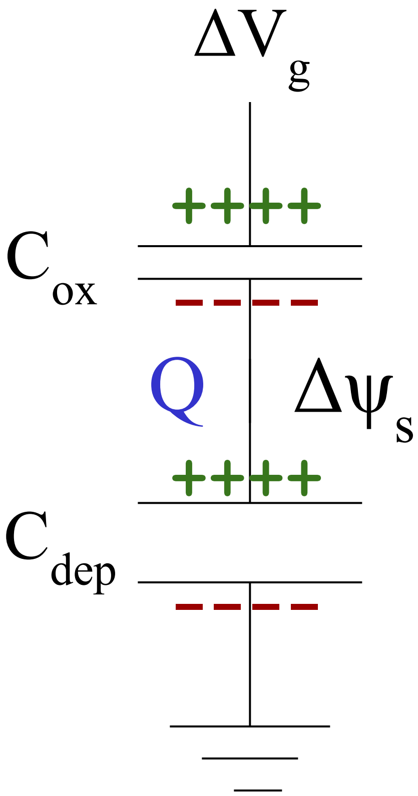
$$\psi_s = \textit{Surface potential}$$

$$\kappa(kappa) = \frac{\partial \psi_s}{\partial V_g} = \frac{C_{ox}}{C_{ox} + C_{dep}}$$

Gate-depletion capacitive divider

How does changing V_g change ψ_s?

1. $CV=Q$
2. Charge Q on ψ_s is constant
3. Change V, hold Q constant

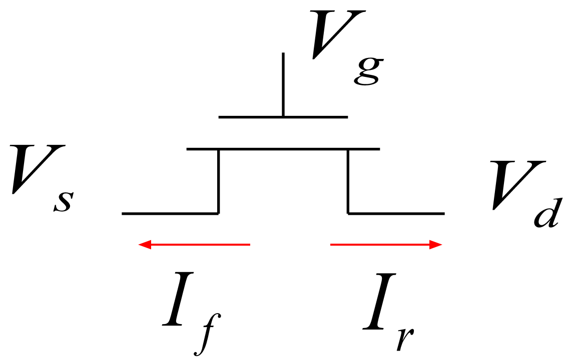


$$C_{ox}(\Delta V_g - \Delta\psi_s) = C_{dep}\Delta\Psi_s$$

$$C_{ox}\Delta V_g = (C_{ox} + C_{dep})\Delta\Psi_s$$

$$\frac{\Delta\Psi_s}{\Delta V_g} = \frac{C_{ox}}{C_{ox} + C_{dep}} \equiv \kappa$$

Equations for Subthreshold nFET

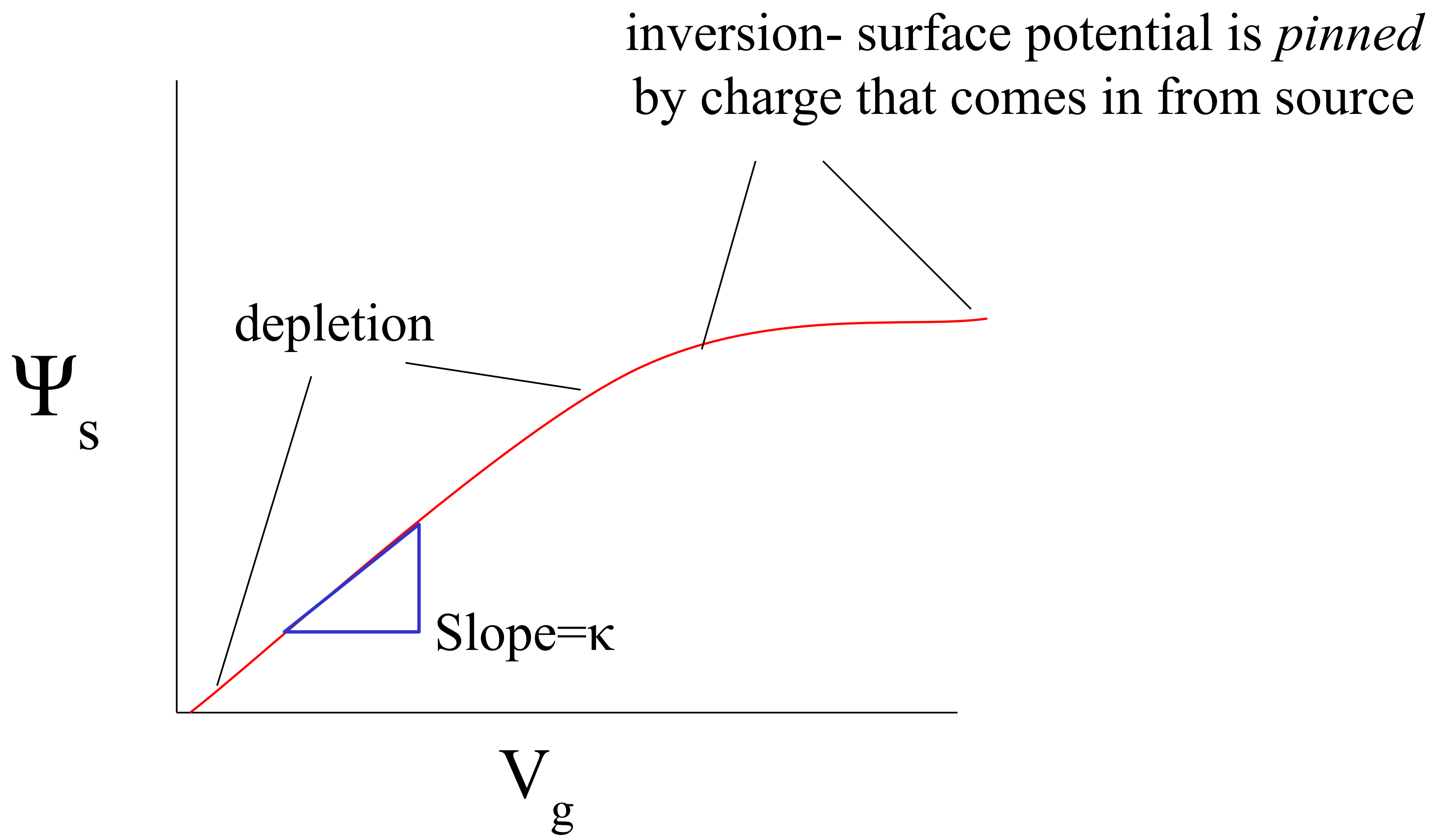


I_f = forward current
I_r = reverse current
I_0 = off current

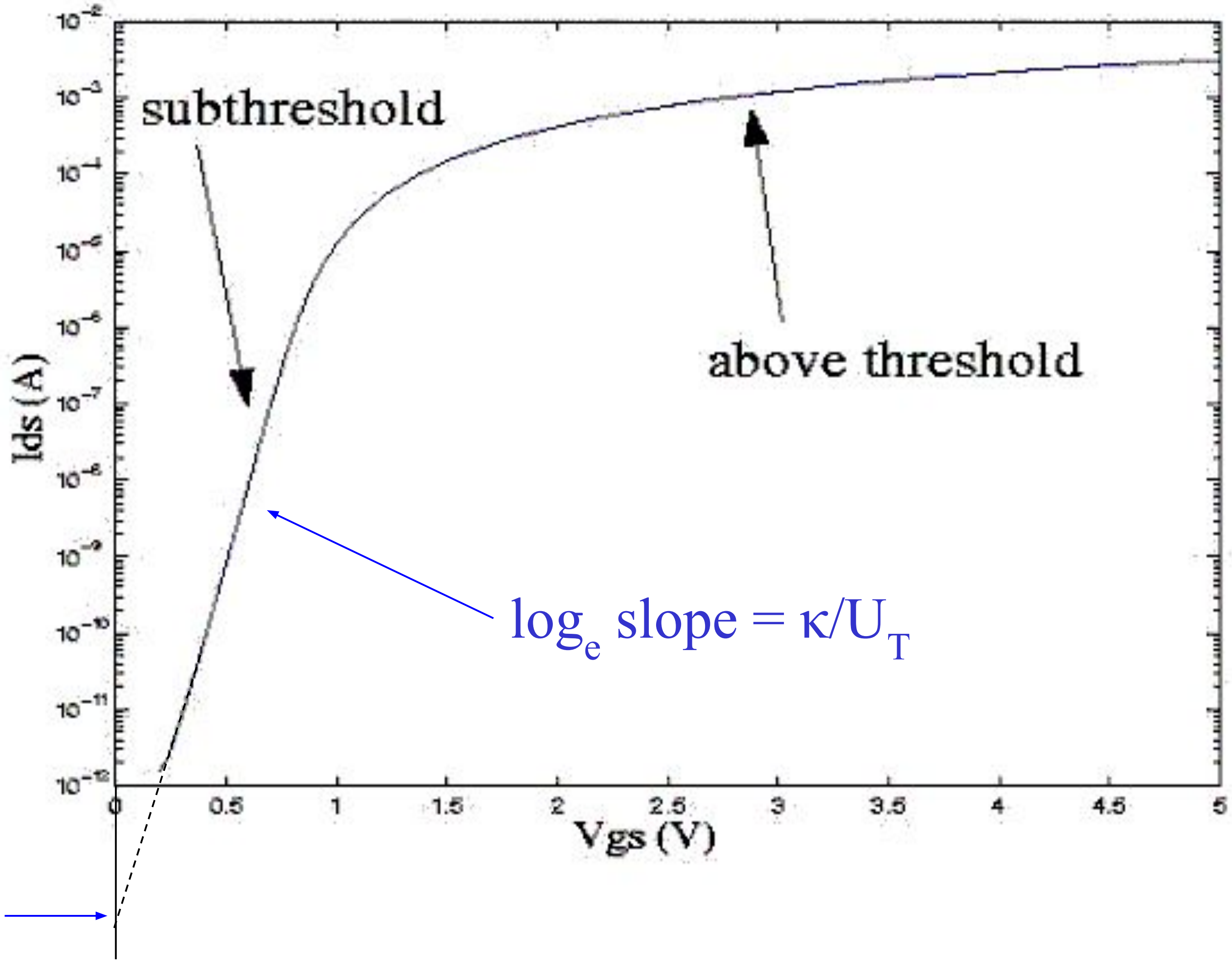
$$\boxed{\begin{aligned} I &= I_0 e^{\kappa V_g/U_T}\left(e^{-V_s/U_T} - e^{-V_d/U_T}\right) \\ &= I_f - I_r \end{aligned}}$$

$$I_f = I_0 e^{\kappa V_g/U_T} e^{-V_s/U_T} \qquad I_r = I_0 e^{\kappa V_g/U_T} e^{-V_d/U_T}$$

Surface potential as function of V_g

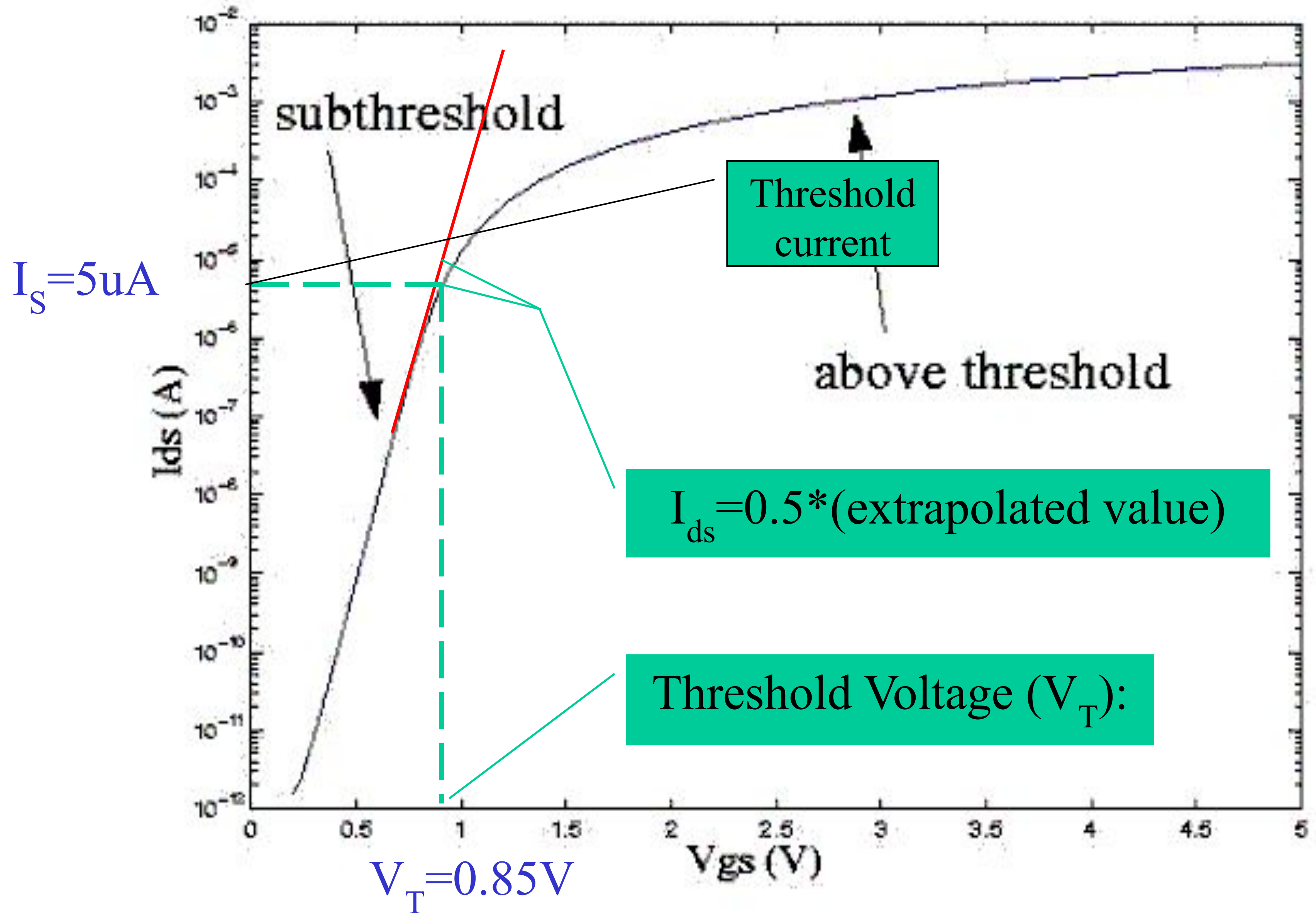


nFET curve: I vs V_{gs}



Intercept I_0

nFet Threshold



Regimes of Subthreshold Operation
(dependence on V_{ds})

<u>Triode/Linear Region</u>

$$\boxed{I = I_0 e^{(\kappa V_g - V_s)/U_T}\left(1 - e^{-(V_d - V_s)/U_T}\right)}$$

<u>Saturation Region (Vds>few UT)</u>

$$\boxed{I = I_f = I_0 e^{(\kappa V_g - V_s)/U_T}}$$

nFET subthreshold Operation

V in units of U_T

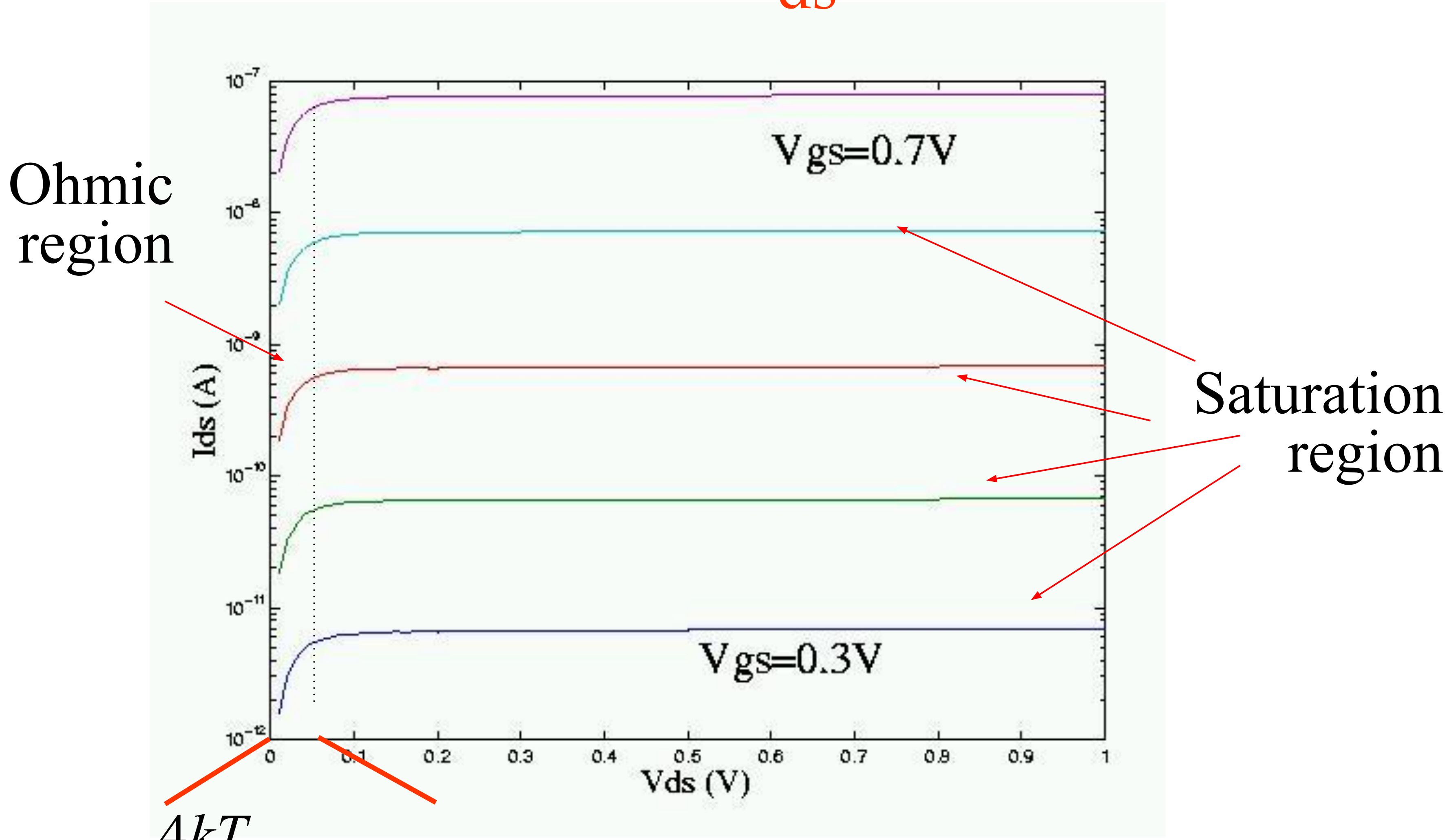
Triode/Linear/Ohmic Region

$$\boxed{I = I_0 e^{\kappa V_g - V_s}(1 - e^{-V_{ds}})}$$

Saturation Region, V_{ds} > a few U_T

$$\boxed{I = I_f = I_0 e^{\kappa V_g - V_s}}$$

nFET drain curve: I vs V_{ds} for long transistors



$$\frac{4kT}{q} \approx 100mV$$

What about the pre-exponential I_0?

$$\boxed{I = I_f = I_0 e^{(\kappa V_g - V_s)/U_T}}$$

I_0 comes from the built-in barrier and the doping concentrations. It takes the form

$$I_0 = N_s U_T{}^2 \beta(T) \exp\left(\frac{-\kappa V_T}{U_T}\right)$$

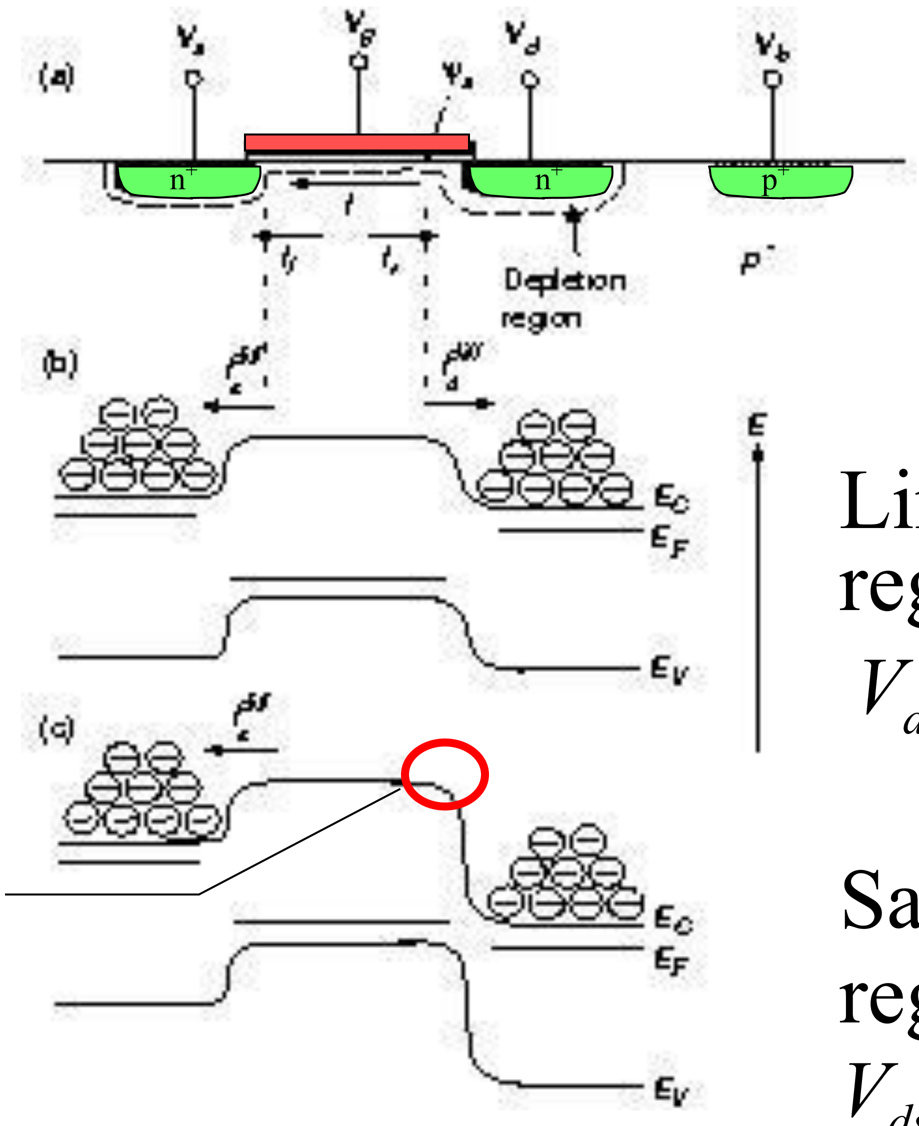
Dimensionless source concentration

$U_T\beta$: diffusivity

U_T : factor for density of states

Concentration at source reduced by barrier

Band Diagram for subthreshold nFET



Linear regime

$V_{ds} < 100mV$

Saturation regime

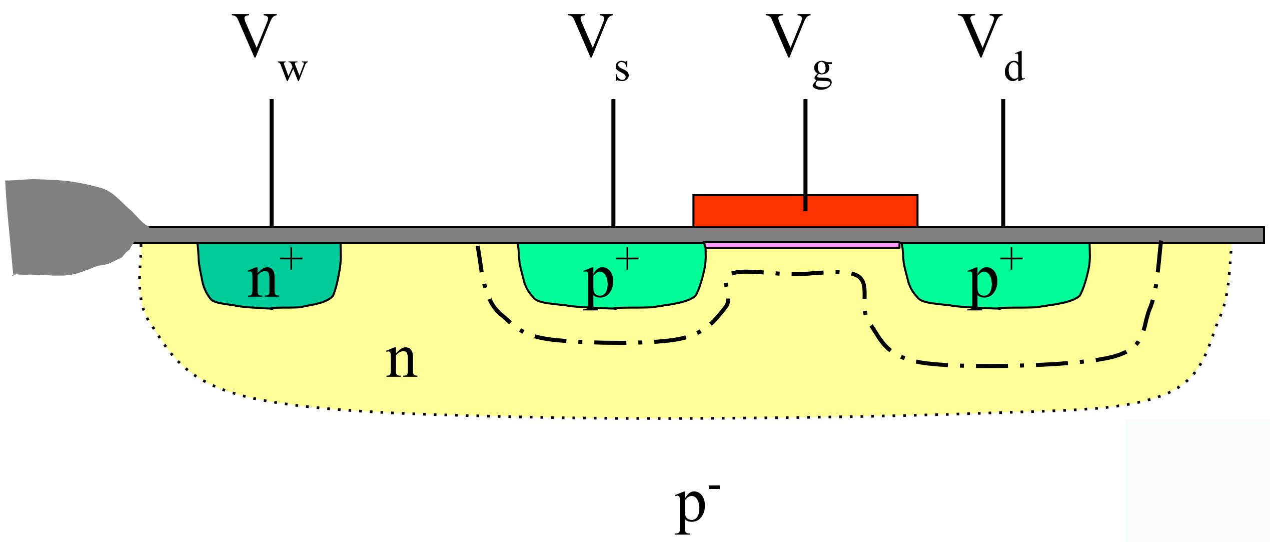
$V_{ds} \geq 100mV$

Drain density is zero

p-type MOSFET

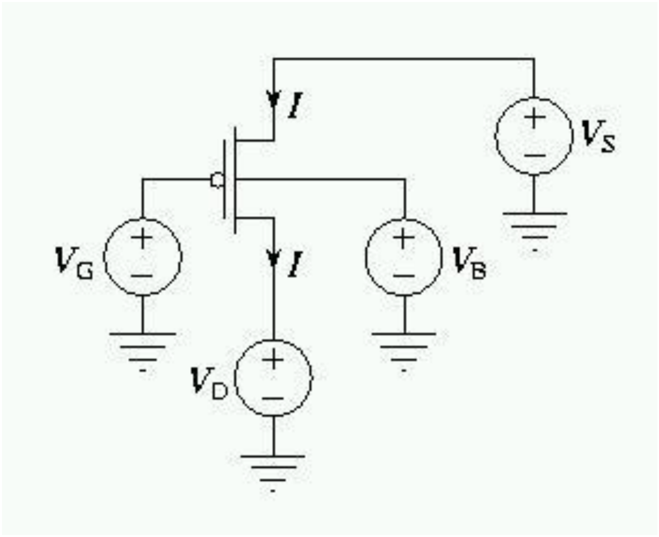
N-well (back gate),
usually Vdd

source gate drain



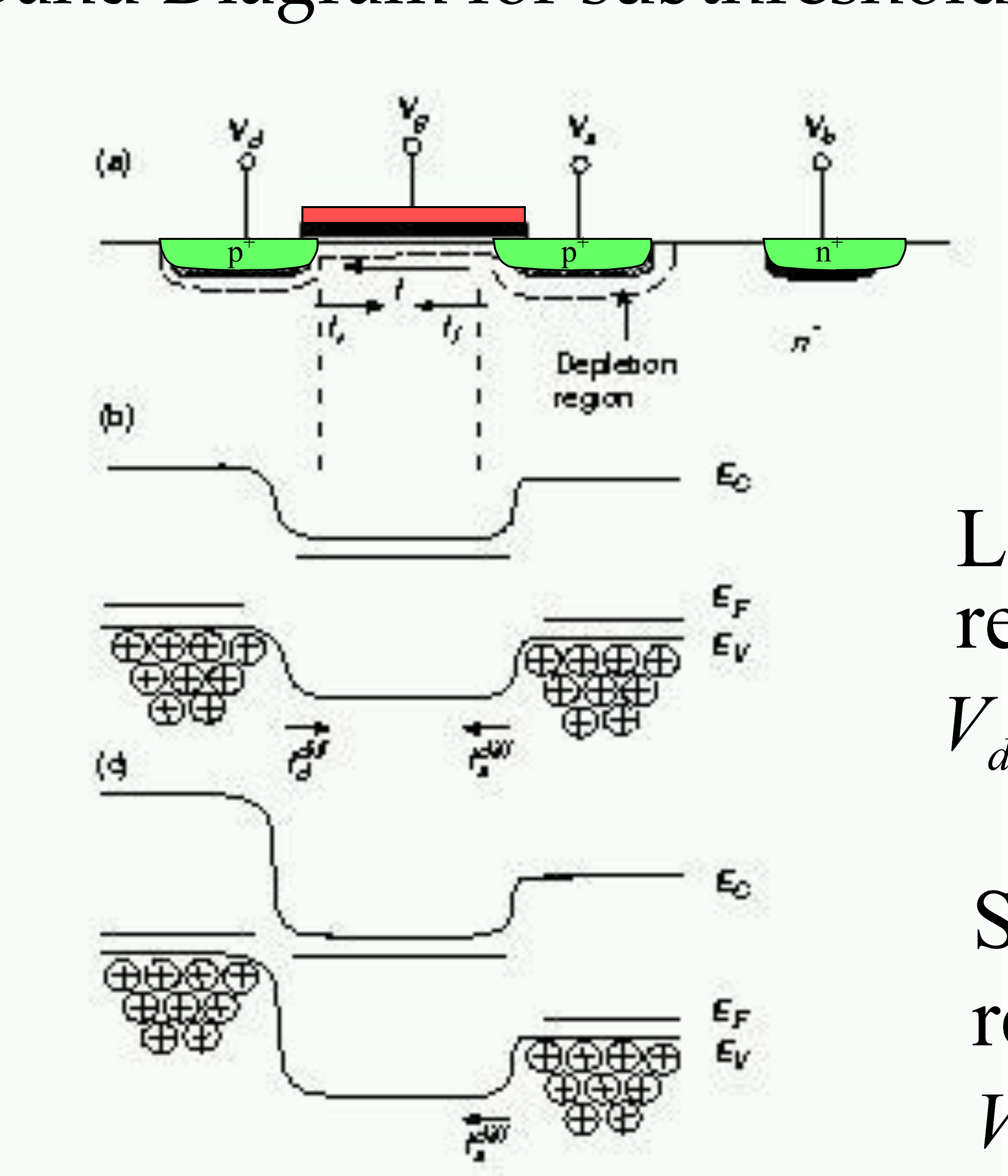
$$V_g, V_d, V_s \le V_w$$

$$V_s \ge V_d$$



All voltages are referenced to $V_w = V_{dd}$

Band Diagram for subthreshold pFET



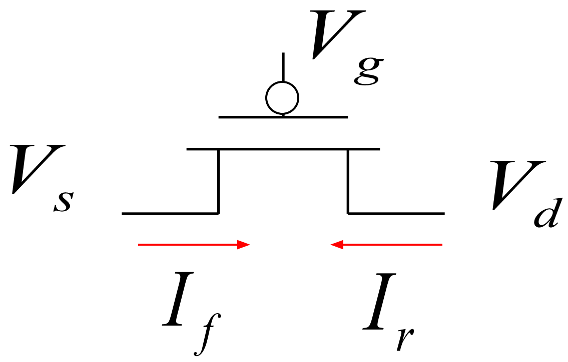
Linear regime

$V_{ds} < 100mV$

Saturation regime

$V_{ds} \geq 100mV$

Equations for Subthreshold pFET



$$\boxed{\begin{aligned} I &= I_0 e^{-\kappa V_g / U_T} \left(e^{V_s / U_T} - e^{V_d / U_T} \right) \\ &= I_f - I_r \end{aligned}}$$

$I_f = \textit{forward current}$

$I_r = \textit{reverse current}$

$$I_f = I_0 e^{-\kappa V_g / U_T} e^{V_s / U_T} \qquad I_r = I_0 e^{-\kappa V_g / U_T} e^{V_d / U_T}$$

pFET subthreshold Operation
V in units of U_T

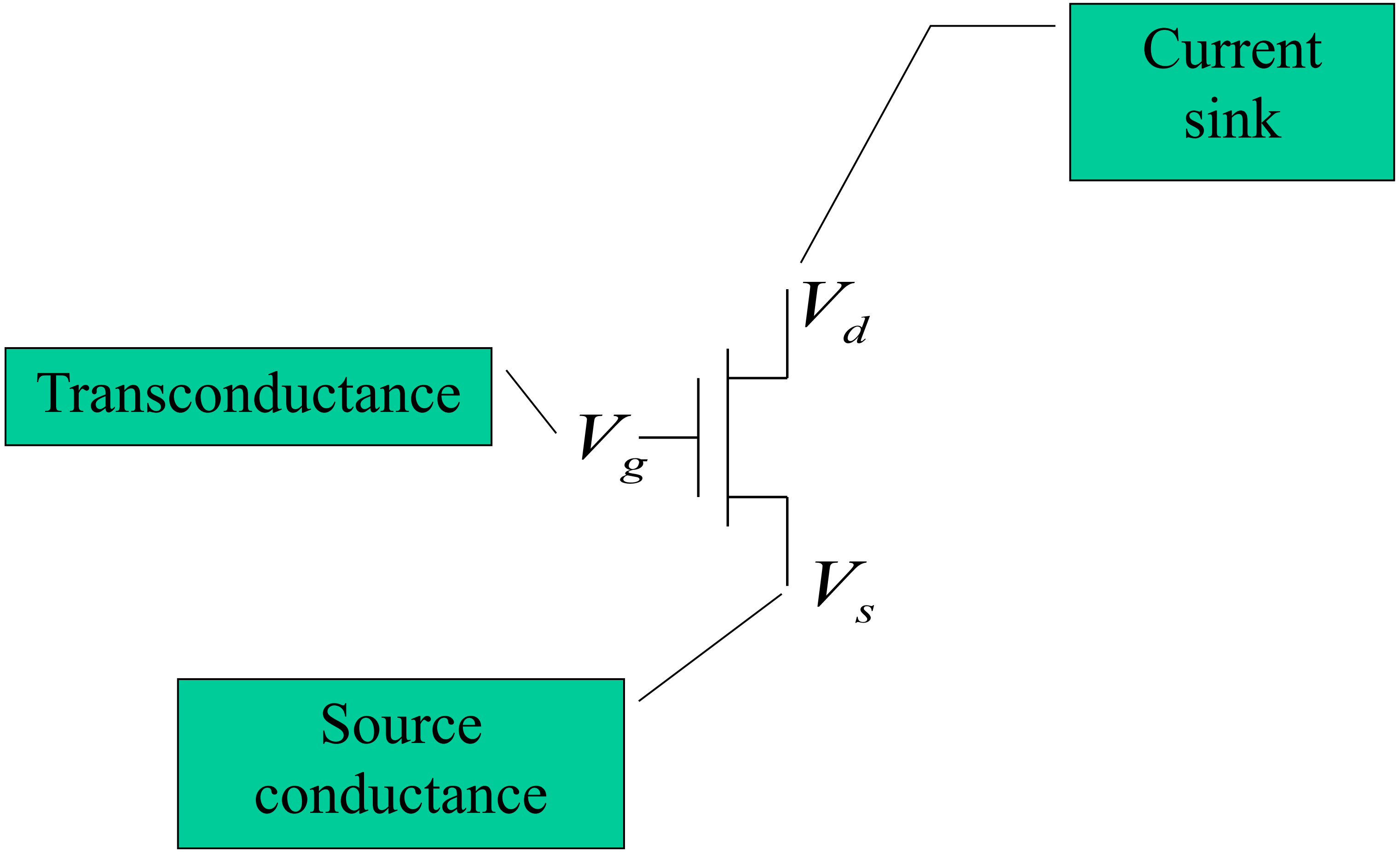
Triode/Linear Region

$$\boxed{I = I_0 e^{-\kappa V_g + V_s}\left(1 - e^{+V_{ds}}\right)}$$

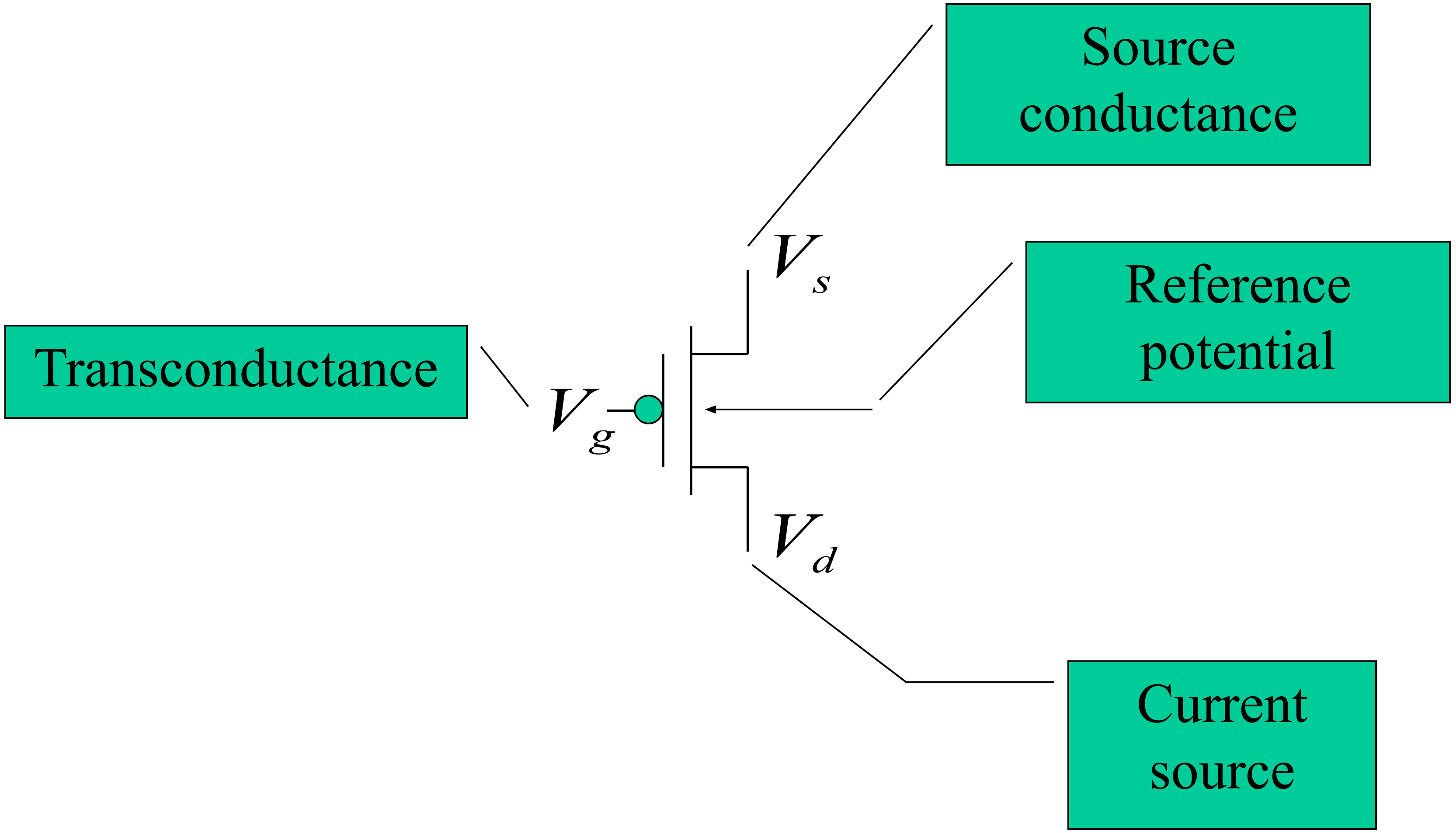
Saturation Region, V_{ds} > a few U_T

$$\boxed{I = I_f = I_0 e^{-\kappa V_g + V_s}}$$

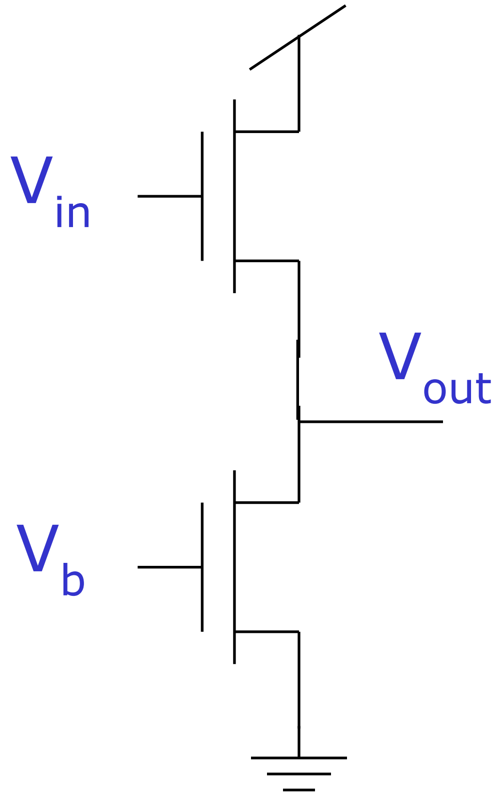
nFET functional behavior



pFET functional behavior



Circuit question



- What is V_{out} *vs.* V_{in}?
- Why is this circuit called a *source follower*?
- How can you use this circuit to measure kappa?

THE END

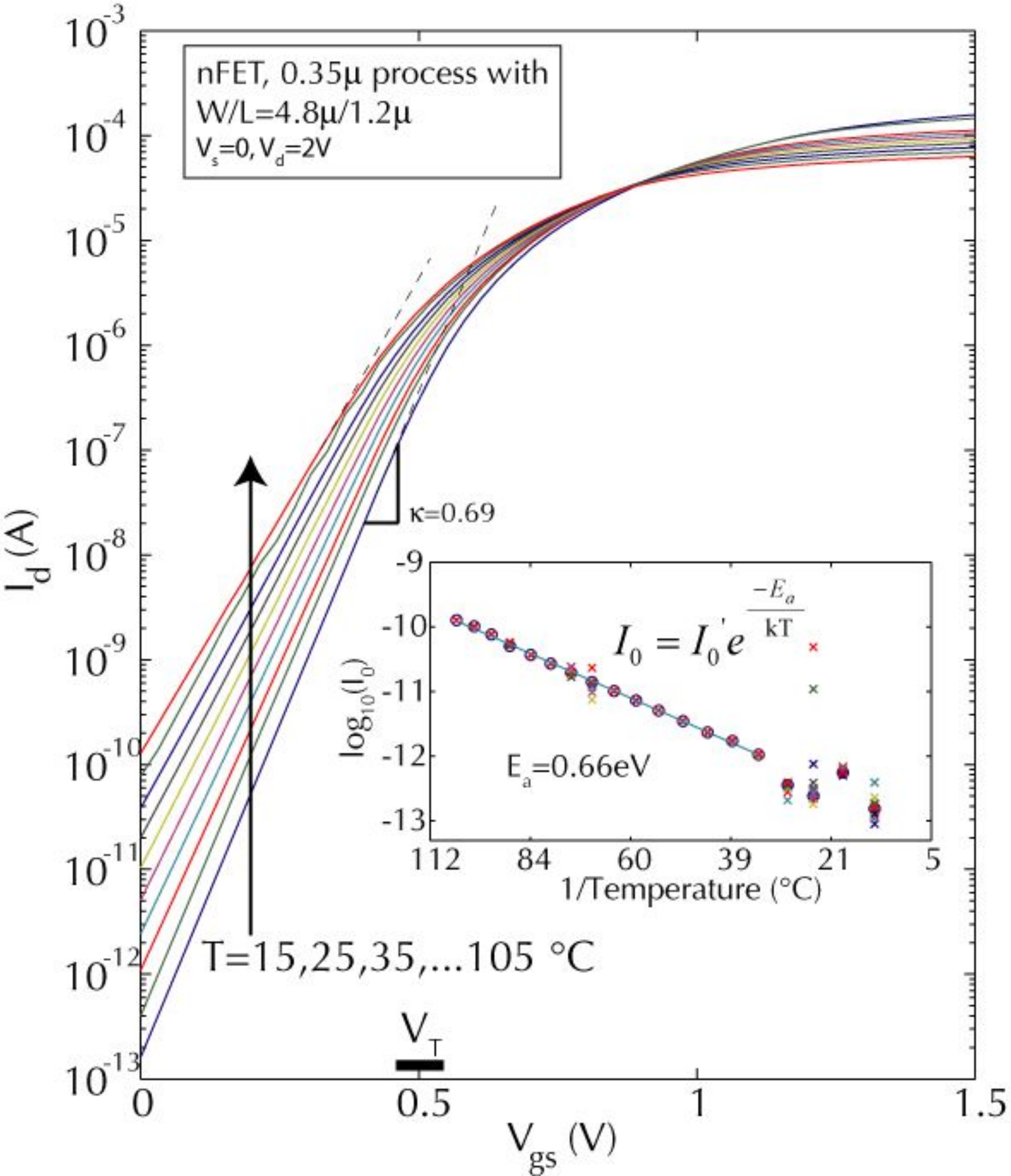
Next transistor lecture (Mellika):

What is the transistor threshold?
Above threshold operation.
Drain conductance-Early effect

Next lecture after this (Giacomo):

Static circuits in subthreshold
Current mirror
Differential pair
Bump and Anti-Bump circuits

Measured temperature dependence of I_d vs V_{gs}



Quiz on device physics

Do you know these constants?

q (elementary charge of electron)

  1.6e-19 C

U_T=kT/q (thermal voltage)

  25mV

ε(permittivity of vacuum)

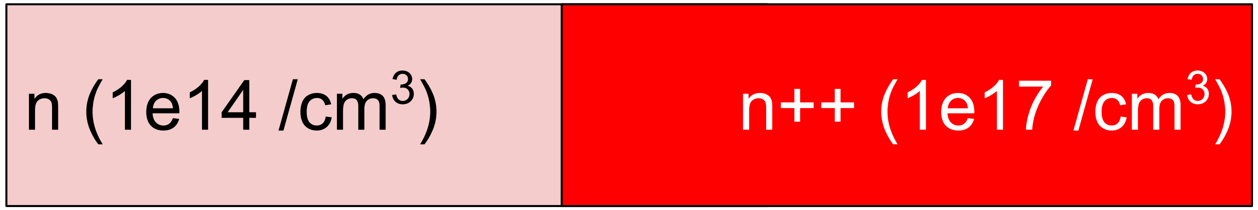
  8.85e-12 F/m

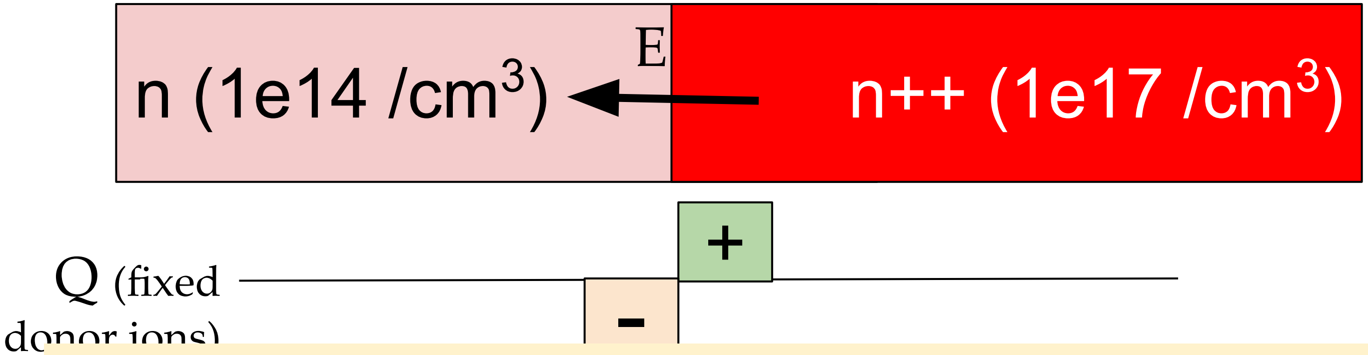
ε_{Si} and ε_{SO2} (relative permittivities of silicon and silicon dioxide)

  ε_{Si}~12 ε_{SO2}~4

Quiz on device physics

If we create an N to N++ doping profile, which way does the electric field point? Is there a space charge region or junction?





Gradient of electron density causes electrons to diffuse leftwards, resulting in electric field pulling them rightwards, so there will be a built in E field pointing to the left. There will be <u>space charge</u>, but <u>no depletion region</u>.

Doping profiles are often used to build in electric fields to push minority carriers in a desired direction, e.g towards a photodiode junction, or away from a switch transistor

Good lab examples

Subthre

1 Exper

1.1 Descr

The purpose o
function of the
a well transisto
process, therefo

while for well t

I

In the above e
instead refer th
and revert the

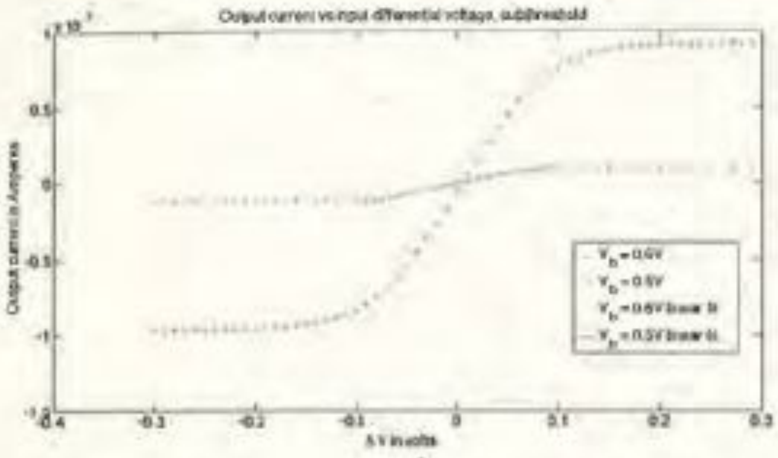
Neuromorphic Engineering I, Lab 05 (04b), Report

[redacted]

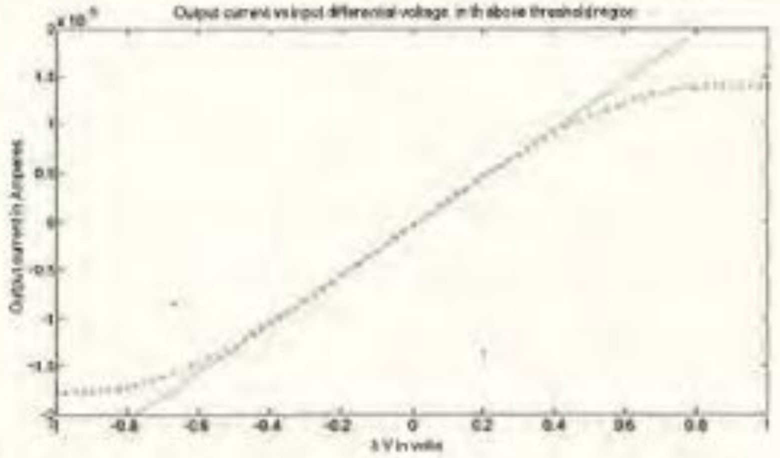
Experiment 03

In this experiment, the output current versus differential input voltage characteristics of the transconductance amplifiers for different bias currents are measured. Specifically, the output current versus input voltage for two different bias currents in the sub threshold region and one bias current in the above threshold regions are considered.

The plots for the simple transconductance amplifier are shown below in 1. We can see from the plot that



(a) Plots for bias currents below sub threshold



(b) Plot for a bias current above threshold

Figure 1: Output currents vs input voltage for different bias currents, in a simple transconductance amplifier, with linear fits to specify the regions of linear behavior. Note that the voltage scales in two subplots are different.

Bad exar

1 I

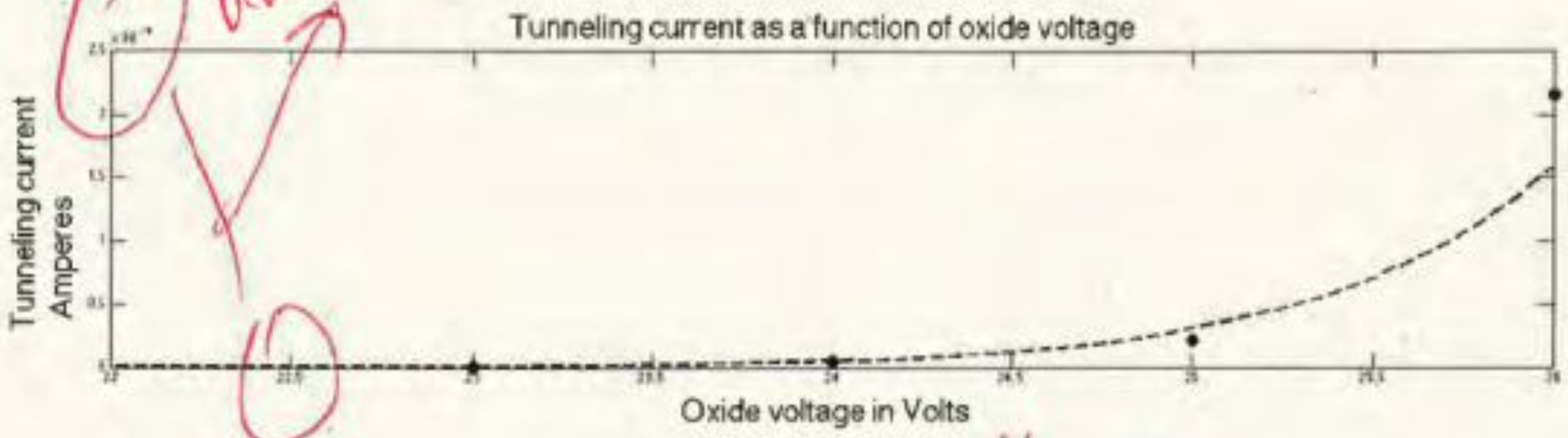
Figure 3 i
According
For nFET,
For pFE
β=5.6

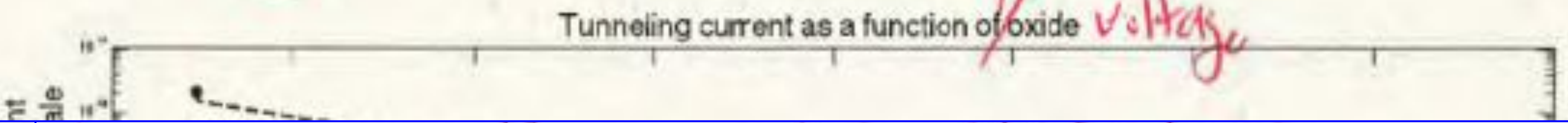
Experiment 3, Gate Oxide Tunneling

Not a sentence

In this experiment the dependence of tunneling current on the oxide voltage. The linear scale plot and the scale plot of the dependence is shown in 3. The curve is fitted to the theoretical function $I = I_0 e^{-\frac{V_0}{V_{ox}}}$. parameter I_0 is the maximum possible tunneling current when the oxide voltage moves to infinity, while parameter V_0 is the oxide voltage at which the tunneling current reaches $1/e = 0.3679$ times the maxim current I_0.

unreadable





Missing theory, setup, procedure

We can clearly see th
are at roughly the sa
on Vref. The smaller