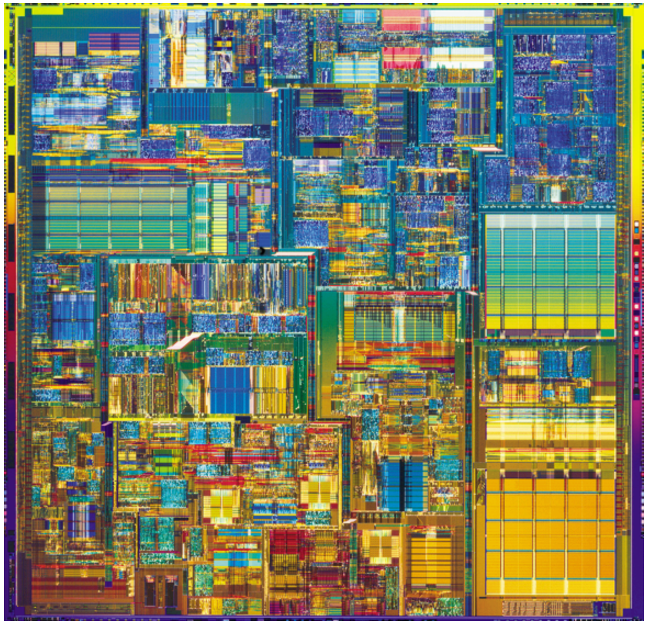
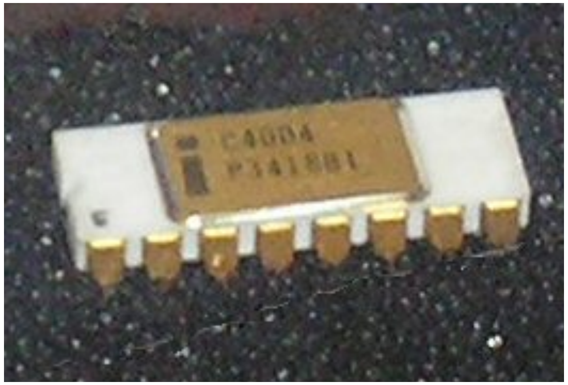
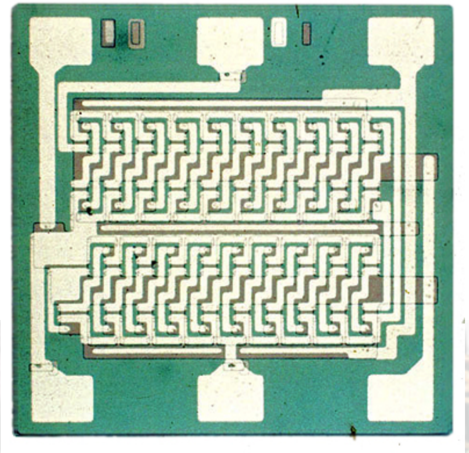
6.1– Transistor Operation
6.4.1 & 6.4.2 MOSFET Basics



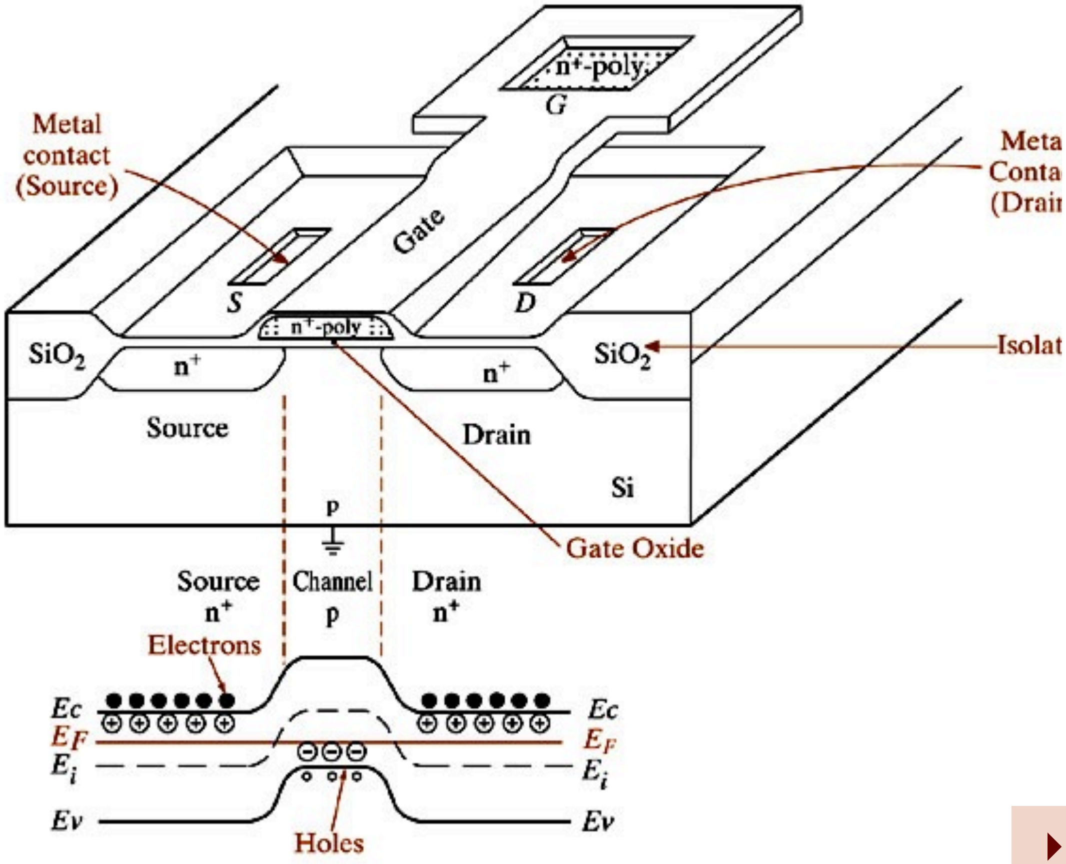
Images for this lecture: 1st MOS IC (RCA, 1964), 1st commercial microprocessor (Intel 4004, 4 bit, 2,300 transistors, 92 kHz, $60,1971), and a modern chip…

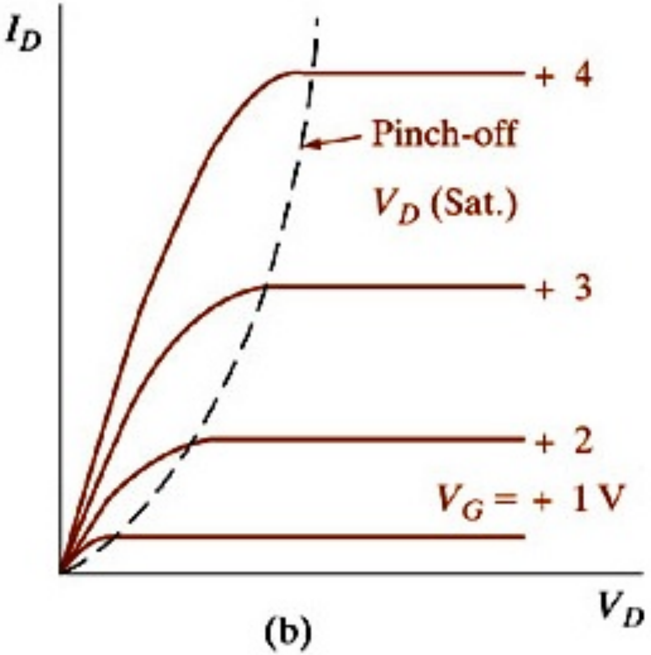
No other human artifact has been fabricated in larger numbers than MOSFETs!





And now… the MOSFET!



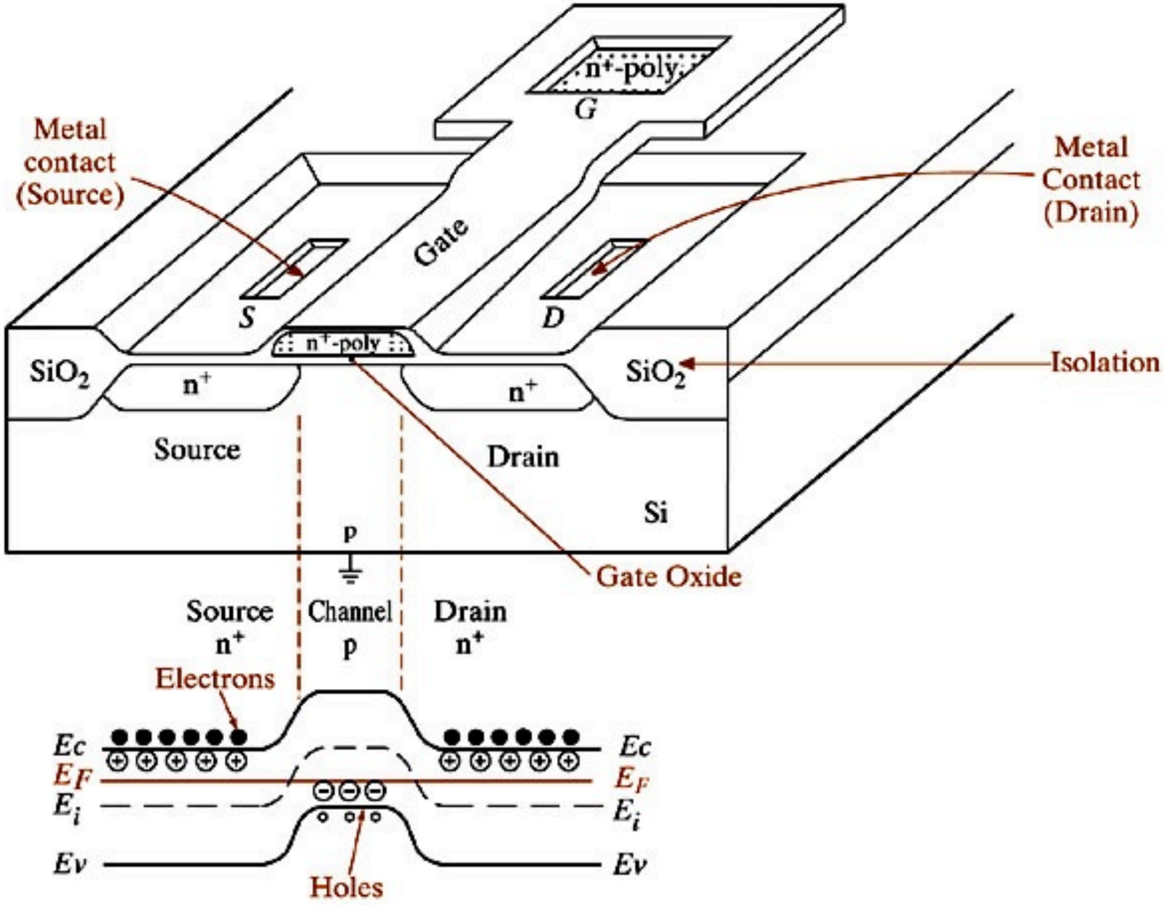


FABRICATION

- Substrate light doped p-Si
- n+ source/drain (diffused)
- thermal oxide (SiO_2)
- n+ poly-Si gate electrode (*thermally stable and best adhesion to oxides*)
- metal contacts (apertures)
- thick isolation oxide

- You already know enough to figure this out! So, tell me right now, why can't we get current flow from source to drain? ☆

- If we applied voltage to gate turn on, + or - ? Why?



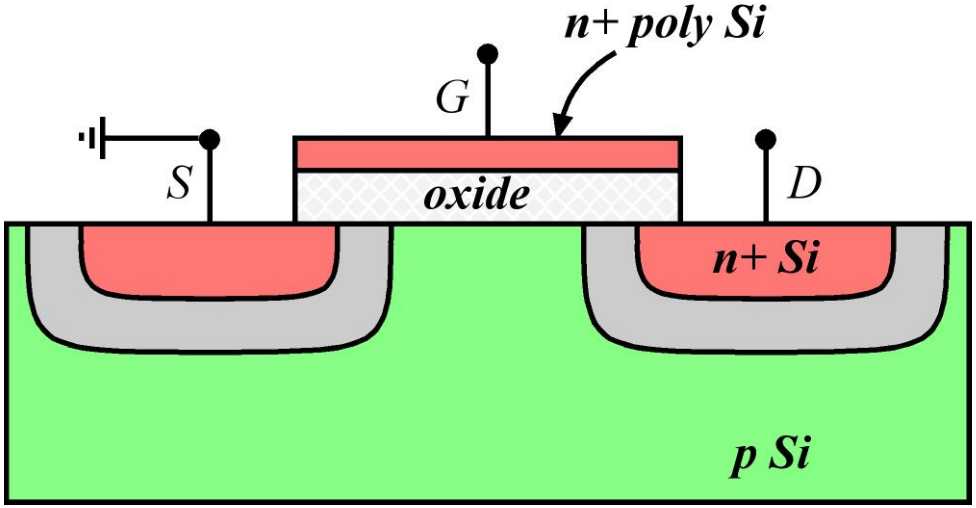
OPERATION

- Fermi levels flat in equilibrium
- Built-in barrier forms and prevents electron conduction in channel… (back to back PN’s!)
- Apply positive voltage, push bands down, provide a conductive channel for electrons to travel in…
- MOSFET TYPES:

(1) Enhancement mode n-channel device (at left) …normally OFF

(2) Depletion mode …normally ON

4 Basic Operation



- Unbiased device *(floating contacts)*

- Depletion regions formed between n+ and p, why n+? Two reasons…

- Now, even if we applied bias between drain and source these back-to-back diodes would prevent current flow

- So what is our ONLY option to get current flow from source to drain? Need to change the channel… We need to create lots of electrons in the channel (n-type!). ☆

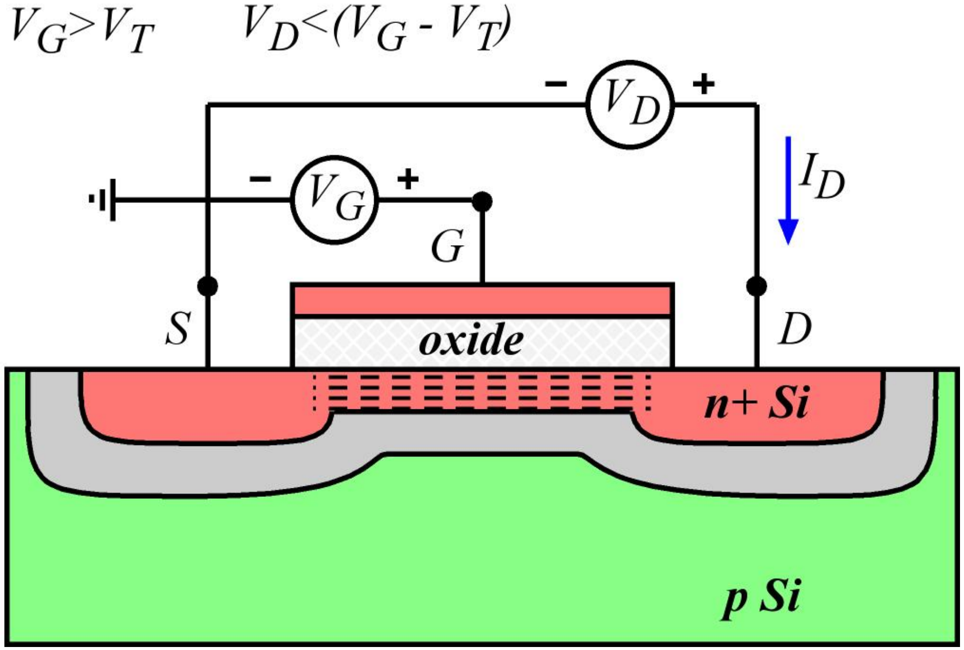
- What if we reversed all the doping types? ☆

- If you can answer these questions then you are well on your way to understanding basic MOSFET principles!

- Positive gate voltage (charge) greater than threshold voltage

- This requires negative voltage (charge) below gate oxide (capacitor charge up!)

- Small drain voltage to allow drift current (no substantial effect on PN junctions)



- Electron accumulation (*inversion*) forms a channel through which current can flow, source/drain current is allowed, oxide prevents any gate current…

☆

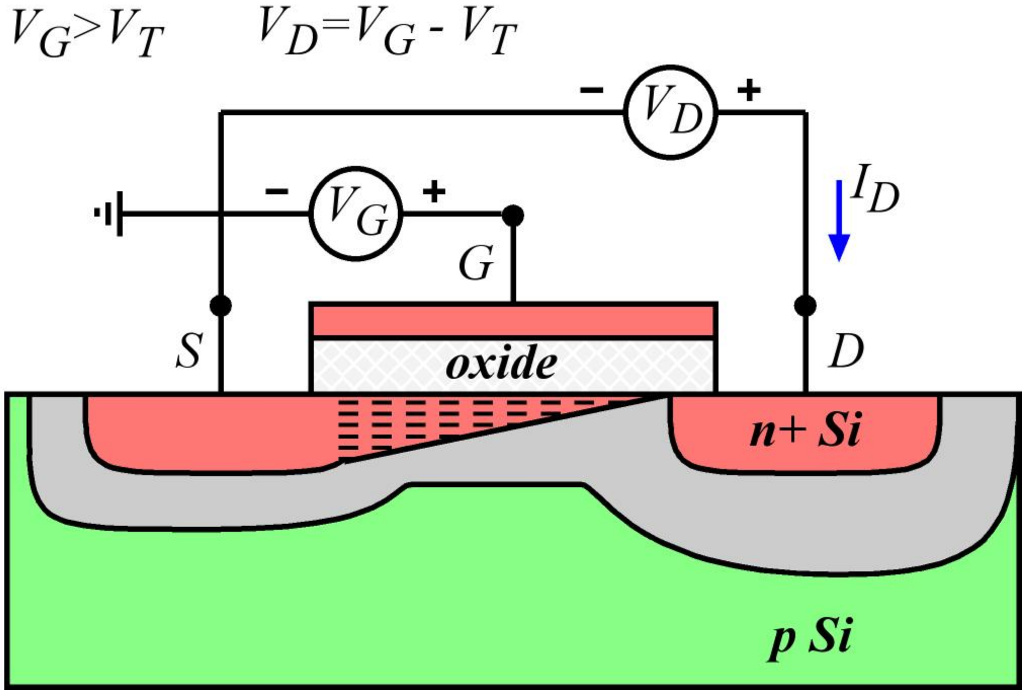
- Electron accumulation mimics n-type material (hence why called NMOS), so a depletion region is formed outside channel

- This depletion region isolates the device from the substrate (which is good for multi-device integration, VLSI)

- Now we have increased our drain voltage significantly

- Like the JFET and MESFET at this point the current flow saturates, what happened?

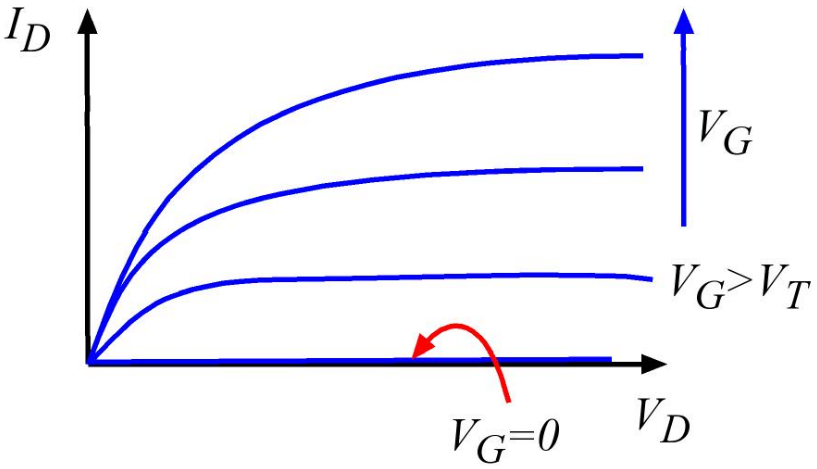
- We can counteract this pinchoff by increasing gate bias



- Good animated example: http://www-g.eng.cam.ac.uk/mmg/teaching/linearcircuits/mosfet.html

Basic Operation

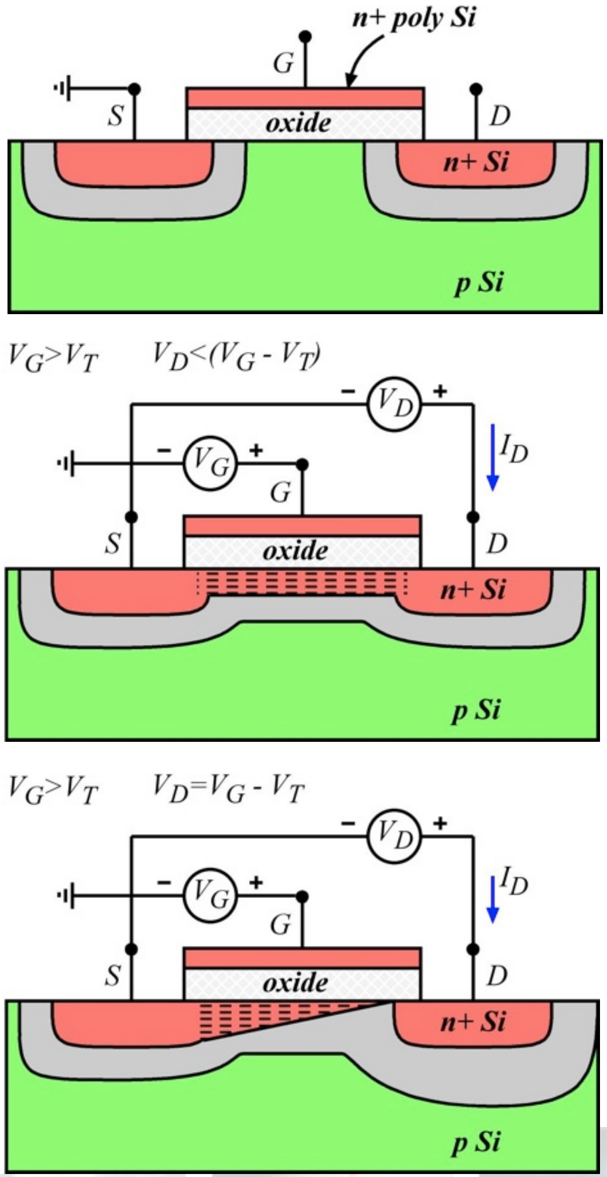
- Bringing it all together (review)



- We won’t spend much time on $I(V_d)$ or even $I(V_g)$! all we care about normally is Vt. *why? Think about the applications…*

- To do this, we need to answer how inversion creates electrons in a p-type material (and we have no current injection, just a capacitor). What has to shift?

- Lets derive Vt, hang in there…



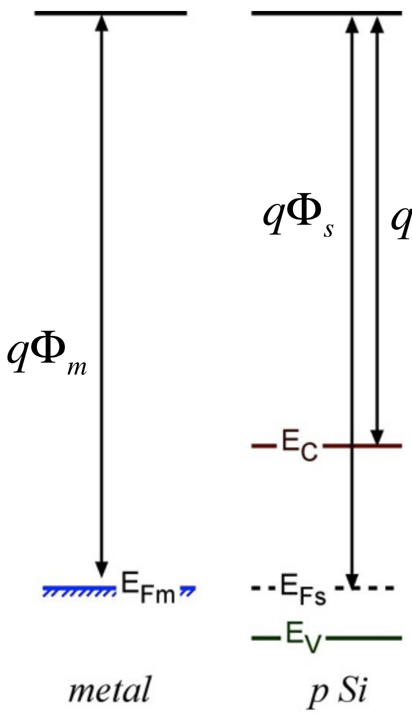
Ideal MOS Capacitor

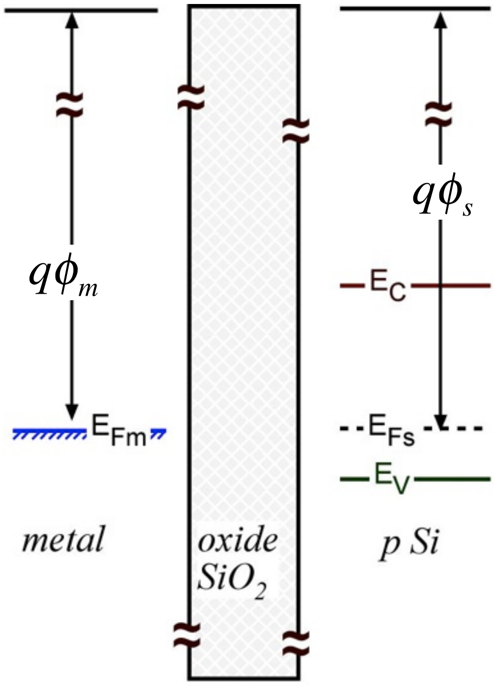
- Assume $q\Phi_m$ $q\Phi_s$ are relatively equal.

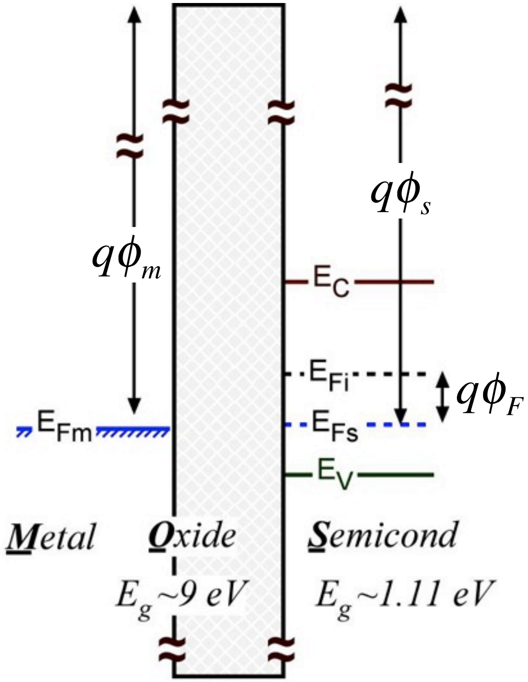
- For convenience use modified work functions (lower case) measured with respect to E_c of oxide.

- Also note $q\phi_F$

vacuum level





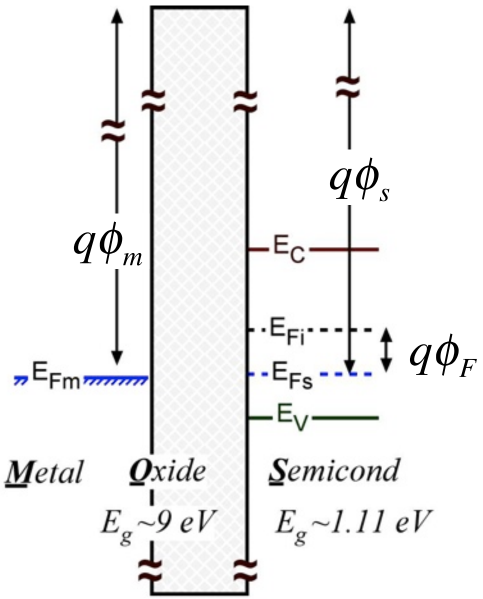


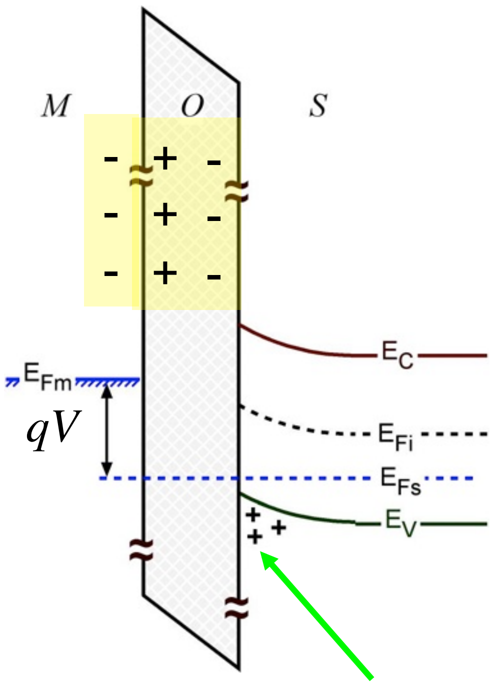
Oxide has conduction band, why insulating?

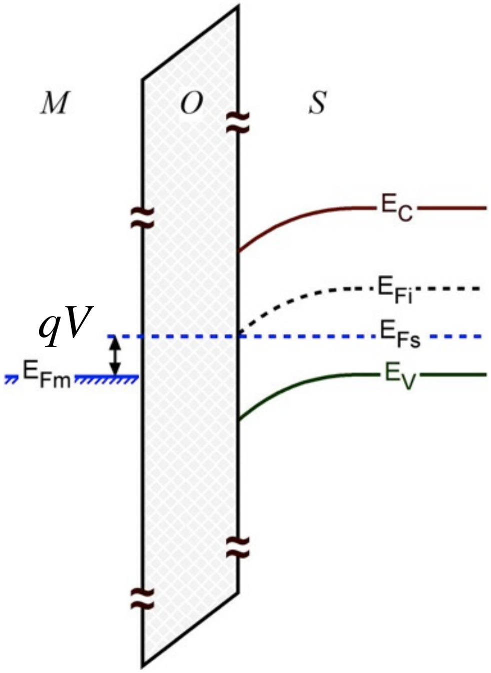
Ideal MOS Capacitor

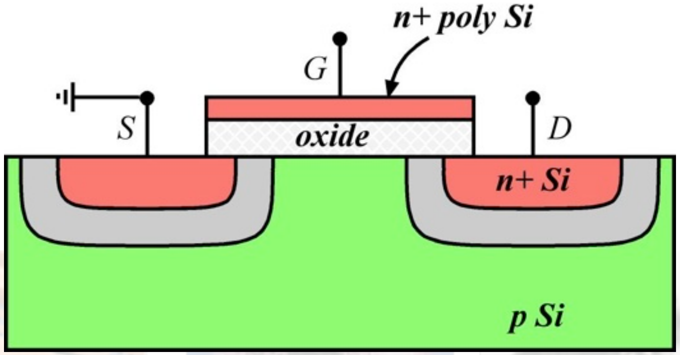
- For metal/oxide/**p-type** Si… voltage applied to metal… V drop across oxide (slope)

- Equilibrium (V=0)
- Accumulation (V<0)
- Depletion (V>0)







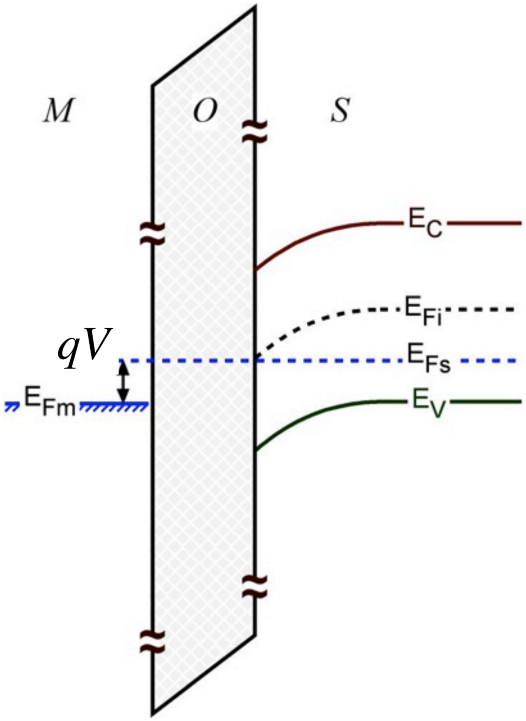


- Here is a channel of holes! why won’t this help us?

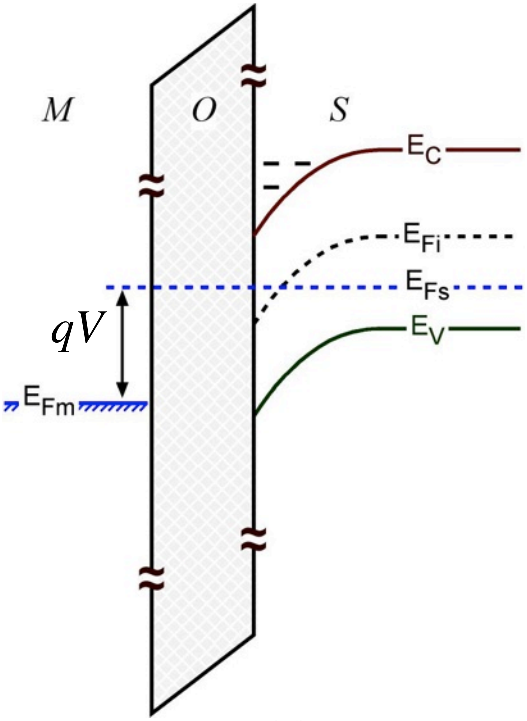
- *Back to back PN junctions!* ☆

- Wait, I said (+) gate bias would get us e’s? Why don’t we have any? How many e’s can drift? Lots? Few? ***Lets derive V_{th}!***

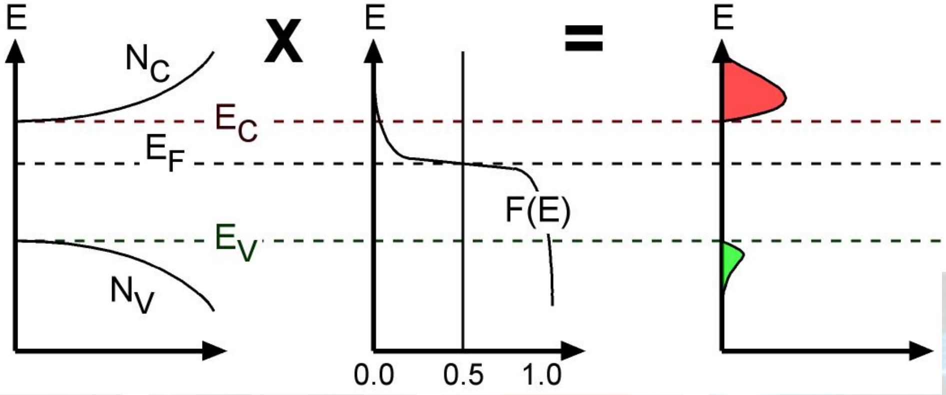
- Depletion (V>0)



- We need **<u>Inversion</u>** (V>>0)!



- Under inversion the Fermi-level ***near oxide*** is closer to the conduction band than the valance band! Makes sense!



Ideal MOS Capacitor

- However, to form a true n-type conducting channel (n+) we need to have **Strong** Inversion

the surface should be just as n-type as it was orginally p-type… (will explain why in a moment)

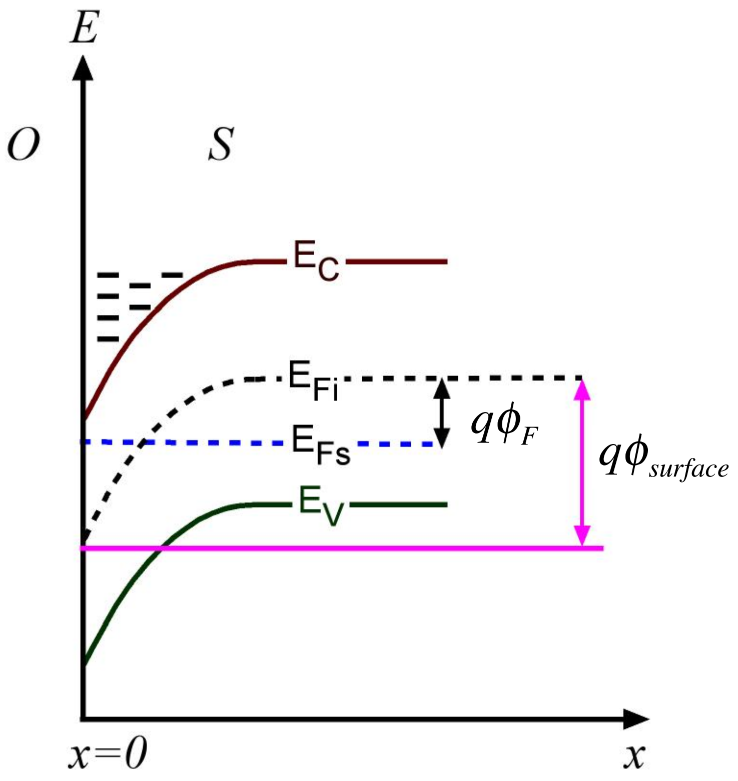
… E_{Fi} should be just as far below E_{Fs} at the surface as it is above E_{Fs} in the bulk

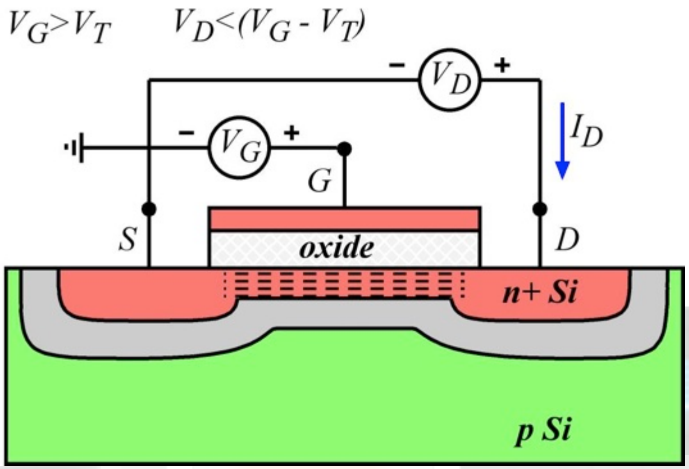
Another way to put this is that to get strong inversion we need the surface potential ($\phi_{surface}$) to be twice the Fermi offset (ϕ_F).

$$\phi_{surface}(inv.) > 2\phi_F = 2\frac{kT}{q}\ln\frac{N_a}{n_i}$$

Again, we will see why we need $2\phi_F$ in a moment…

Hmm…. Surface potential (sounds like it will be part of our threshold voltage). This will also make sense later! Stay tuned!





- The channel conductivity is based on the electron concentration, lets calculate…
  - In the bulk:

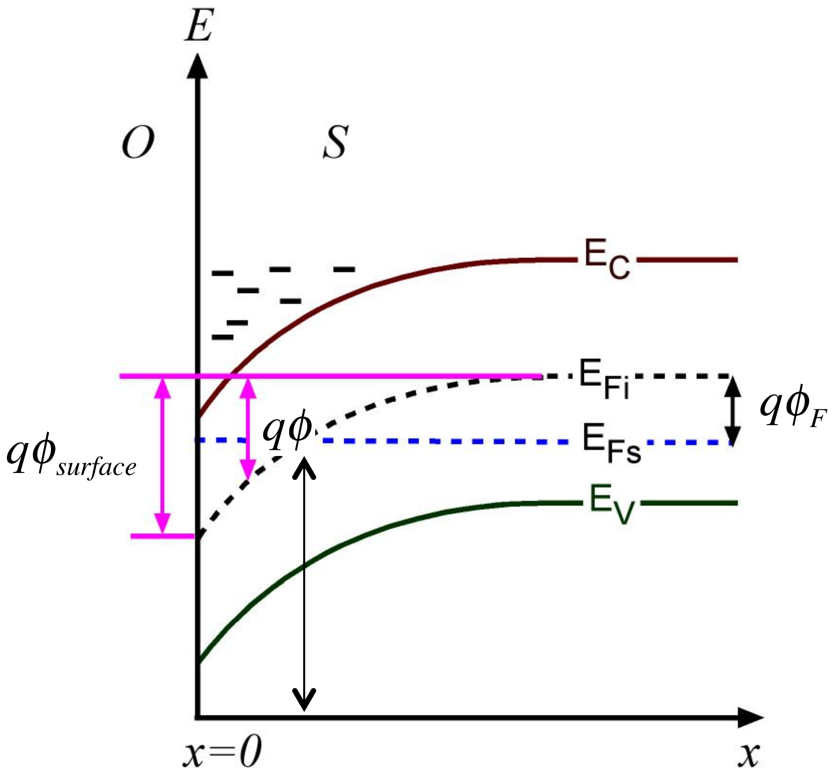
$$n_0 = n_i e^{(E_f - E_i)/kT}$$
$$= n_i e^{-q\phi_F/kT}$$

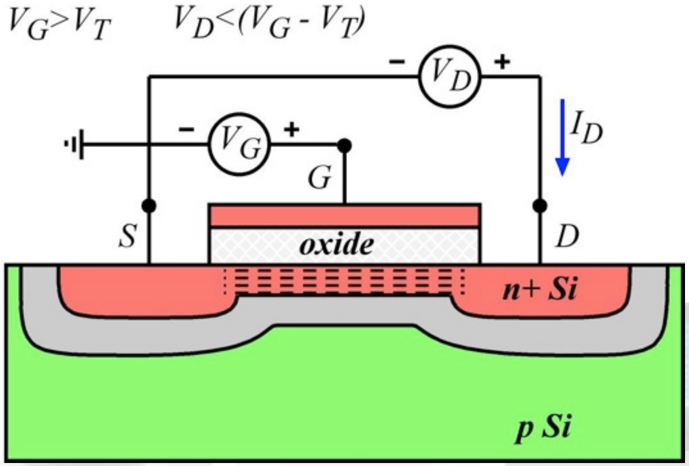
  - As a function of x: qϕ

$$n = n_i e^{-q(\phi_f - \phi)/kT}$$
$$= \underbrace{n_i e^{-q\phi_f/kT}} e^{q\phi/kT}$$
$$= n_0 e^{q\phi/kT}$$

  - Same approach for holes…

$$p = p_0 e^{-q\phi/kT}$$





- Near surface, what does this tell us for n, p? Device at right I(Vg)? ☆

Ideal MOS Capacitor

- Use Poisson's equation (ϕ for V), what does this mean?

$$\frac{\partial^2\phi}{\partial x^2} = -\frac{\rho(x)}{\varepsilon_s}$$

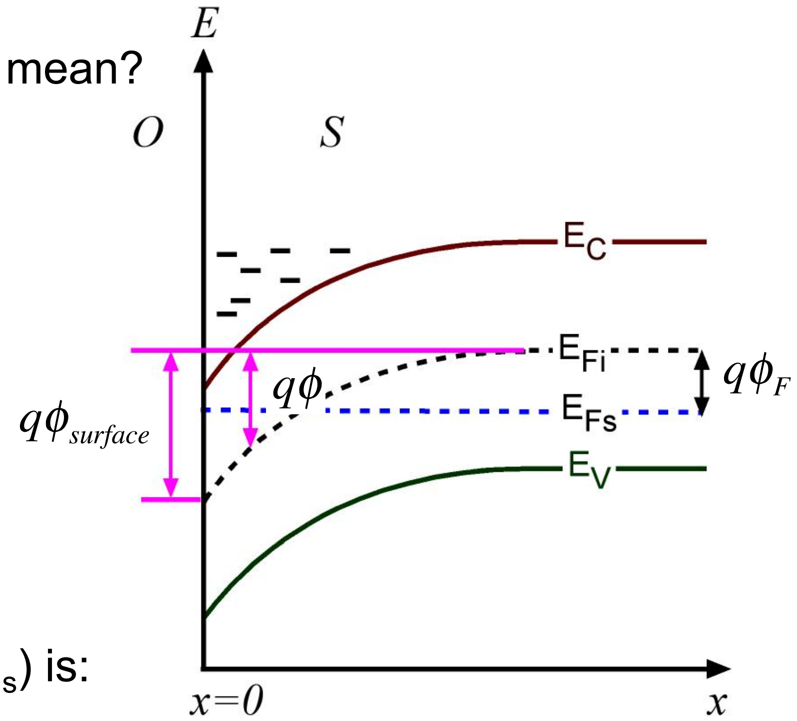
the typical charge density:

$$\rho(x) = q(N_d^+ - N_a^- + p - n)$$

and the relation:

$$\mathrm{E} = -\frac{\partial \phi}{\partial x}$$

it can be shown that *at the surface and perpendicular to the surface* (x=0) that the E-field (E_s) is:



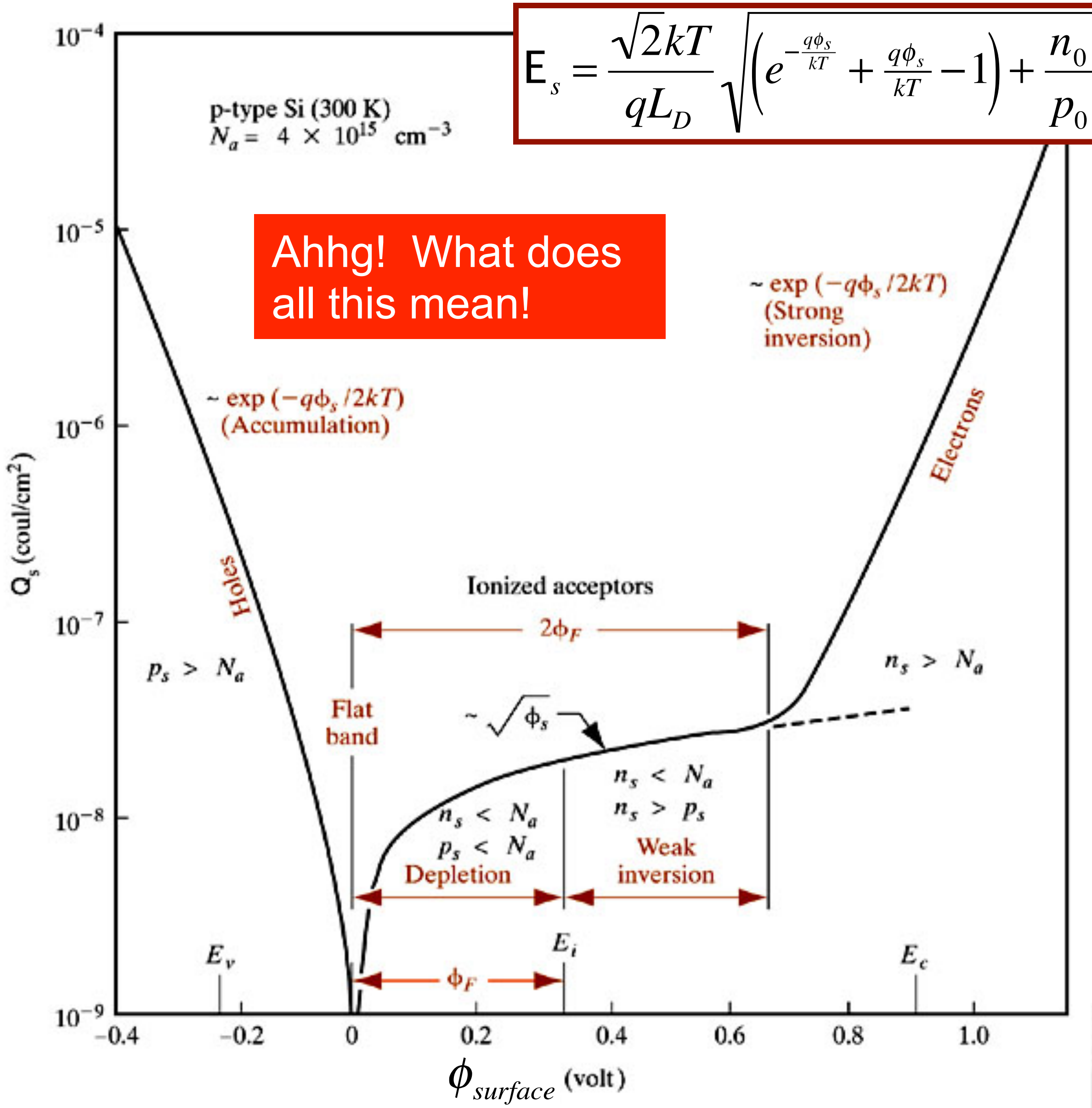
$$\mathrm{E}_s = \frac{\sqrt{2}kT}{qL_D}\sqrt{\left(e^{-\frac{q\phi_s}{kT}} + \frac{q\phi_s}{kT} - 1\right) + \frac{n_0}{p_0}\left(e^{\frac{q\phi_s}{kT}} - \frac{q\phi_s}{kT} - 1\right)} \qquad L_D = \sqrt{\frac{\varepsilon_s kT}{q^2 p_0}}$$

- L_D is the Debye length. It comes up a lot in electrostatics.

We cannot bring all the electrons to x=0, diffusion forces want to push them away…. (look inside the equation, for really heavy doping L_D is small, why?).

Ideal MOS Capacitor

$$E_s = \frac{\sqrt{2}kT}{qL_D}\sqrt{\left(e^{-\frac{q\phi_s}{kT}}+\frac{q\phi_s}{kT}-1\right)+\frac{n_0}{p_0}\left(e^{\frac{q\phi_s}{kT}}-\frac{q\phi_s}{kT}-1\right)}$$

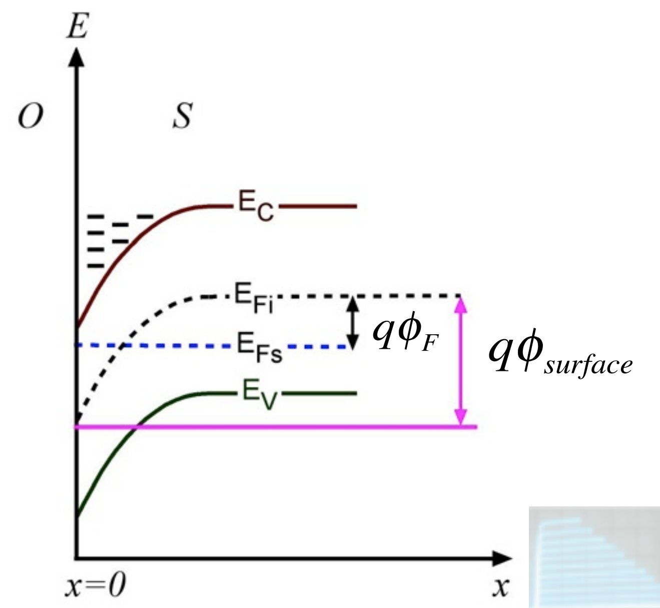


Ahhg! What does all this mean!

- We can apply Gauss' Law at the surface

$$Q_{surf} = -\varepsilon_s E_{surf} \;(coul/cm^2)$$

- Note, this plot is only at the surface (x=0)



Ideal MOSFET V_{TH}

- KEY!!!!

To understand this plot…

- and capacitance vs. voltage,
- and charge distribution vs. voltage,
- and threshold voltage,

we must understand that there are a series of events that take place in biasing the MOSFET:

Accumulation <-> Flatband <-> Depletion <-> Inversion

You cannot move to one state, without having passed through the other. This will have a large implication on capacitance (switching speed) and theshold voltage!

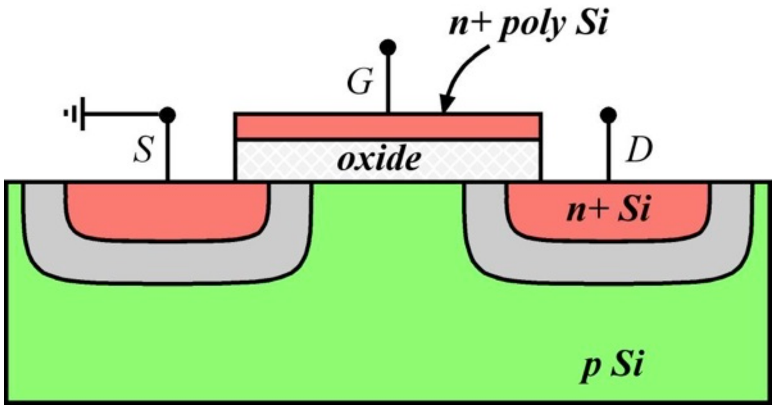
- Why can't I get current flow from source to drain without gate voltage? *No hint needed, answer should be obvious now!*

- If I apply negative voltage to the gate what will happen? Why no source-drain current? *Hint: negative voltage is negative charge on the gate electrode, which on the other side of the capacitor is therefore positive charge (holes).*

- If I apply positive voltage to the gate, what happens BEFORE the MOSFET is turned on? *Hint: before you get electrons (n-type channel), you first get something that still does not allow source-to-drain current, what is it?*

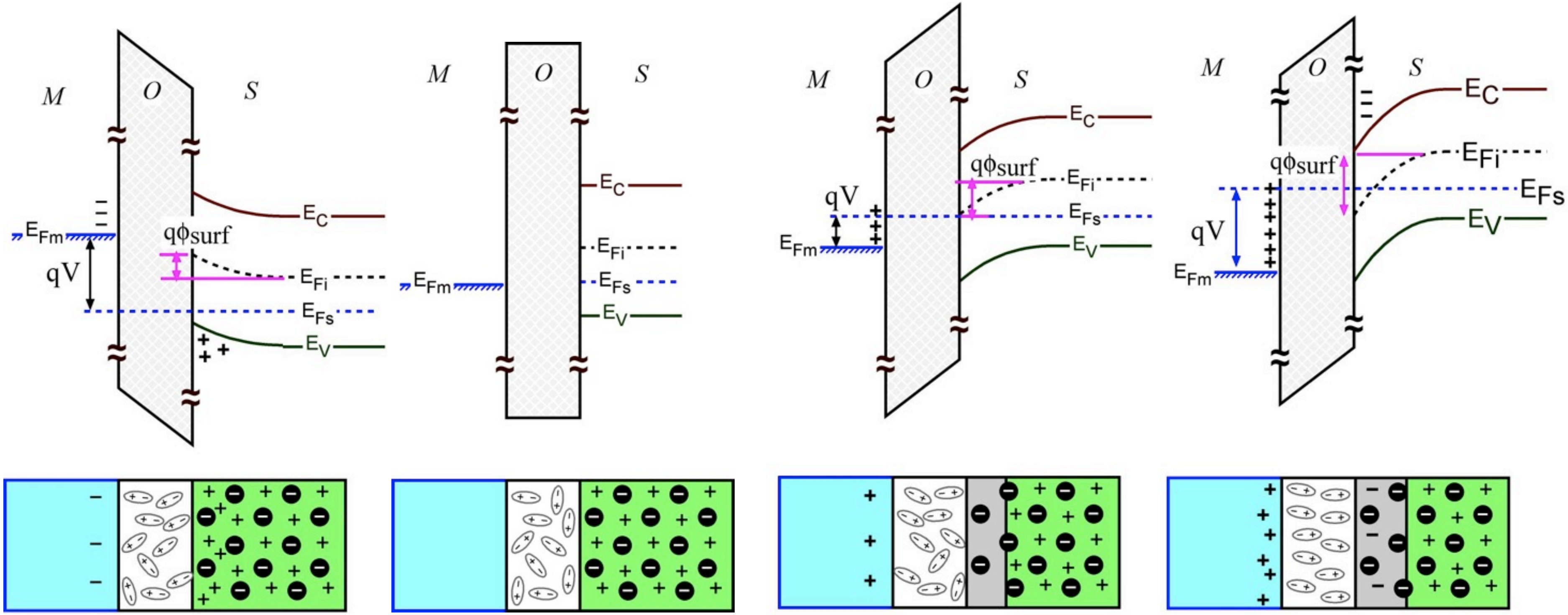
- If the MOSFET is on and I keep increasing the source drain voltage, what will happen and why? What type of current flow is this? *Hint, answer with the same terms we used for previous transistors!*

- The MOSFET at right is NMOS, why called NMOS? *No hint needed!*

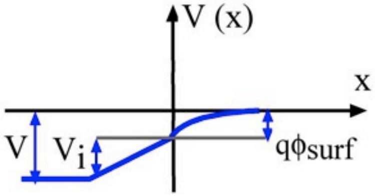


Ideal MOSFET V_{TH}

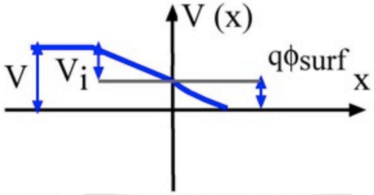
- Accumulation
- Flat Band
- Depletion & Weak Inversion
- Inversion

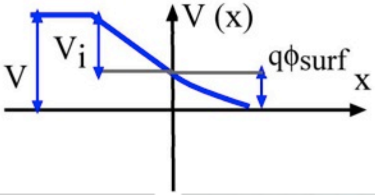


charge, but no channel!

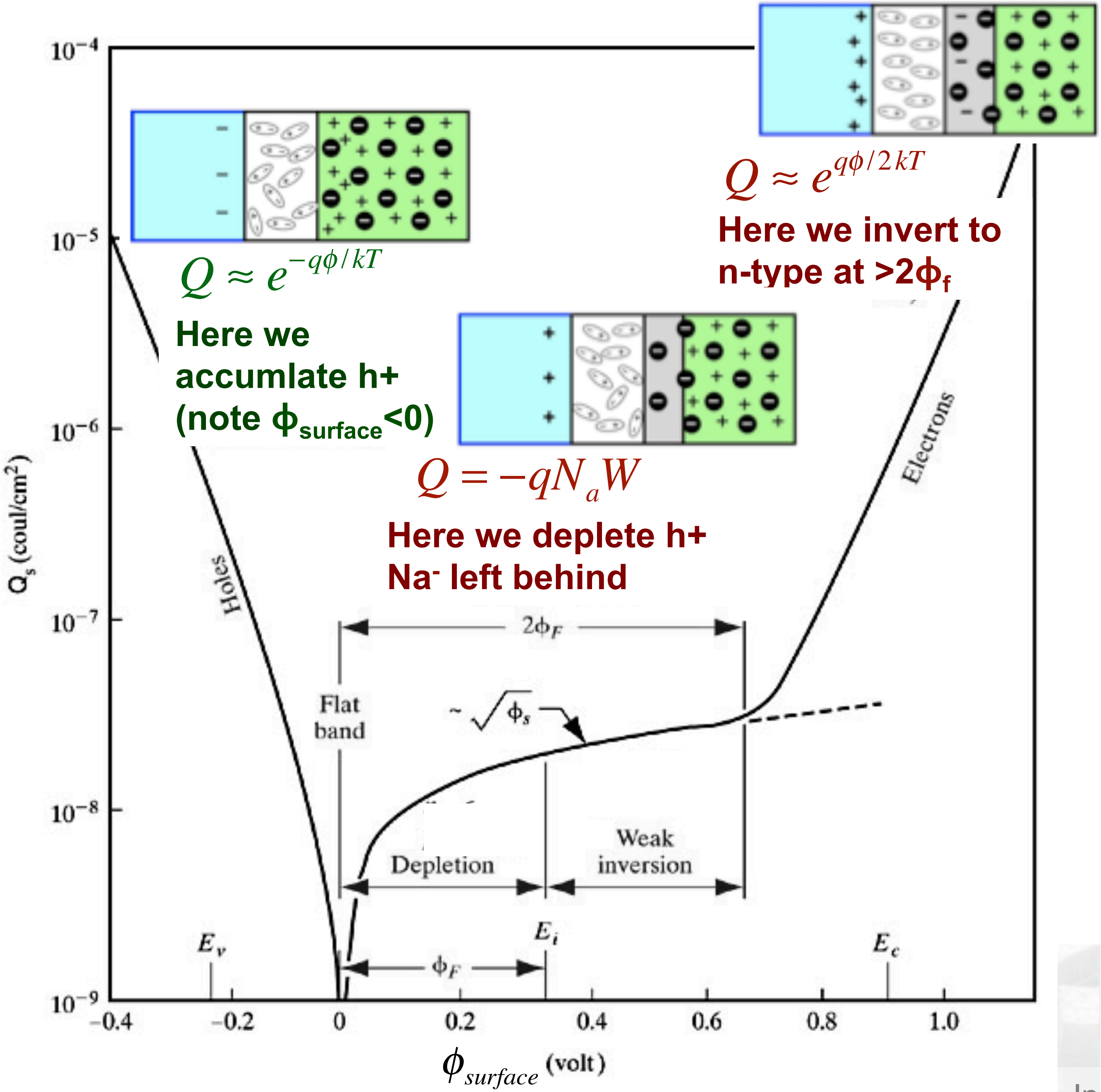


Note how draw dipoles inside dielectric…





Straight band bend = constant V drop (constant E), curved band = non-constant V drop…
Curved for semicon. because contributing charge decreases (and screens) as get toward edge.



$Q \approx e^{-q\phi/kT}$

Here we accumlate h+ (note $\phi_{surface}$<0)

$Q = -qN_aW$

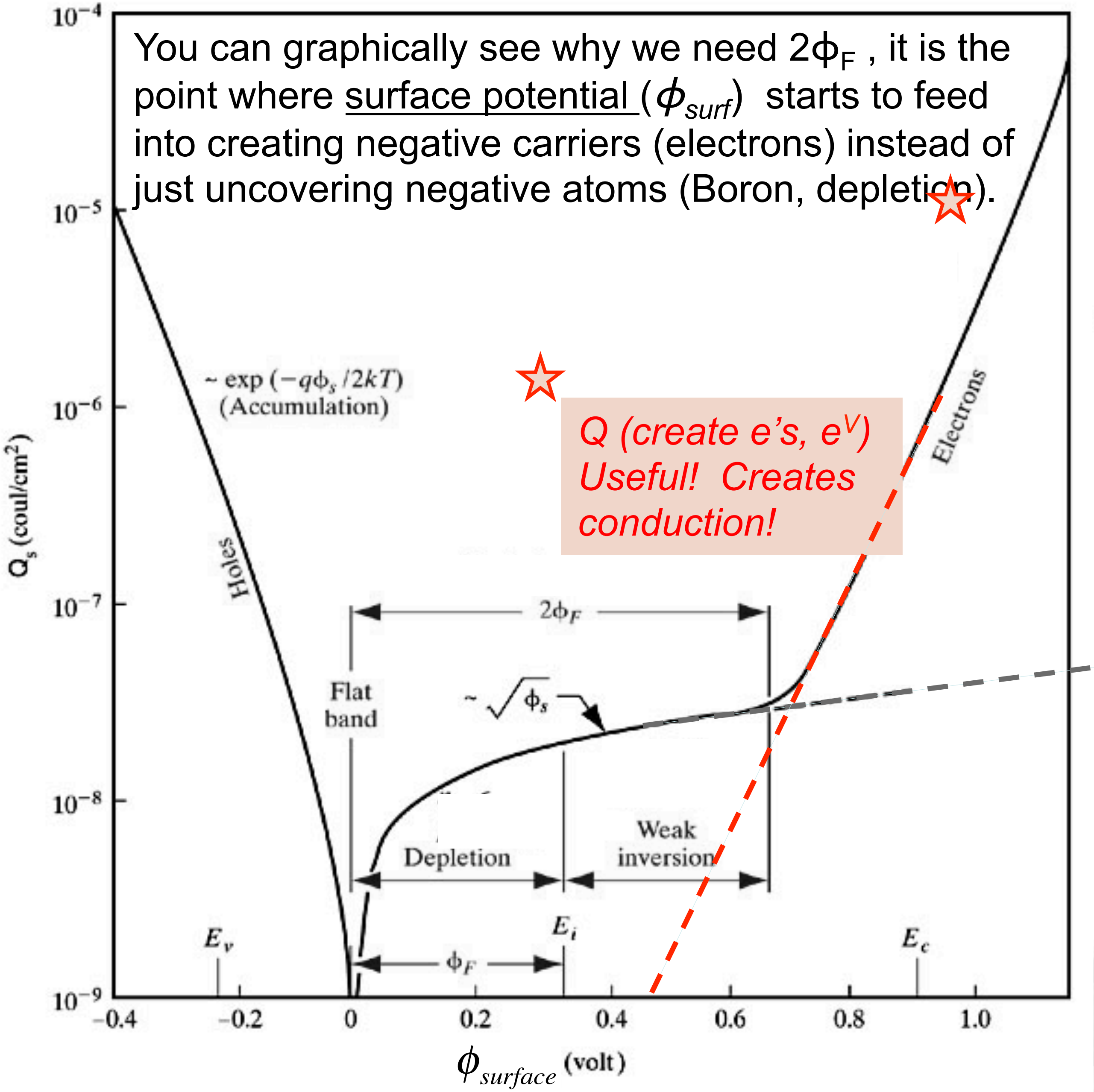
Here we deplete h+ Na^- left behind

$Q \approx e^{q\phi/2kT}$

Here we invert to n-type at >$2\phi_f$

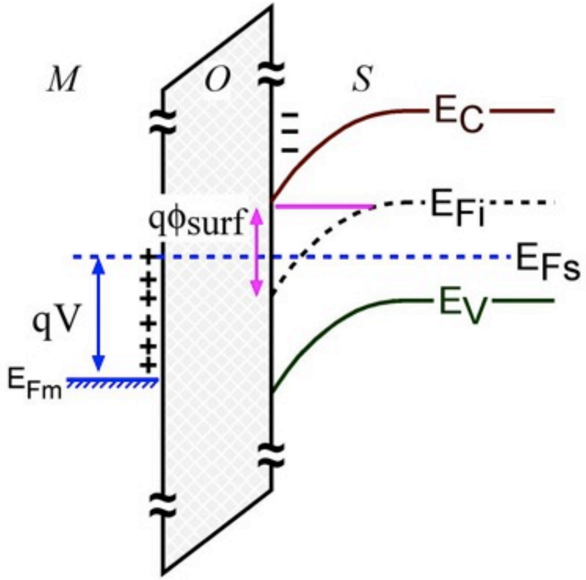
But, why do we need to still pay for $2\phi_F$ of voltage before we can get the n-type channel to form?

Why We Have to Pay 2ϕ$_f$



You can graphically see why we need 2ϕ$_F$, it is the point where surface potential (ϕ_{surf}) starts to feed into creating negative carriers (electrons) instead of just uncovering negative atoms (Boron, depletion).

☆ *Q (create e's, e^V) Useful! Creates conduction!*



☆ *Q (deplete + h's and leave behind - Borons, V$^{1/2}$) this created Q is not useful for us…*

Ideal MOSFET V_{TH}

- Similar to what we did with the PN junction, lets plot Q, E, V for the MOS capacitor… <u>*WHY IS Q IMPORTANT?*</u>

  - speed! power!

- Charge on metal side:

  - high-density thin layer (Q_m, positive)

- Charge in dielectric is zero (dipoles)

- Charge in p-type semiconductor is:

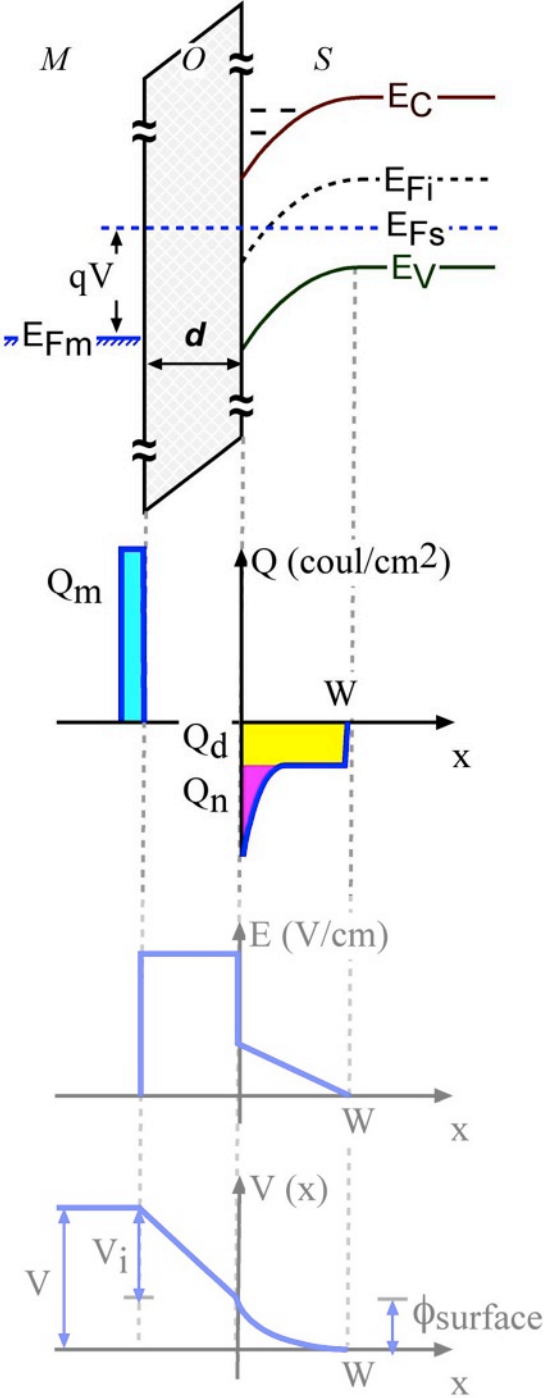
  - depletion (Q_d, <u>uncompensated ionized acceptors, N_a^-</u>, negative)

  - inversion (Q_n, electrons, negative)

  - Why Q zero for x>W? ☆

$$Q_m = -Q_d - Q_n$$

$$Q_m = -qN_aW - Q_n \quad (coul/cm^2)$$



Ideal MOSFET V_{TH}

- Note, channel is exaggerated in figure at right, typically it is only ~10 nm.

- Our applied voltage is split up as voltage across the oxide insulator and the bands (both are sloped, right?!):

$$V = V_i + \phi_{surface} \qquad C = \frac{\varepsilon A}{d} \qquad V_i = Q_i / C_i$$

Vi is wasted… *V_i will be part of the price we have to pay for V_T*

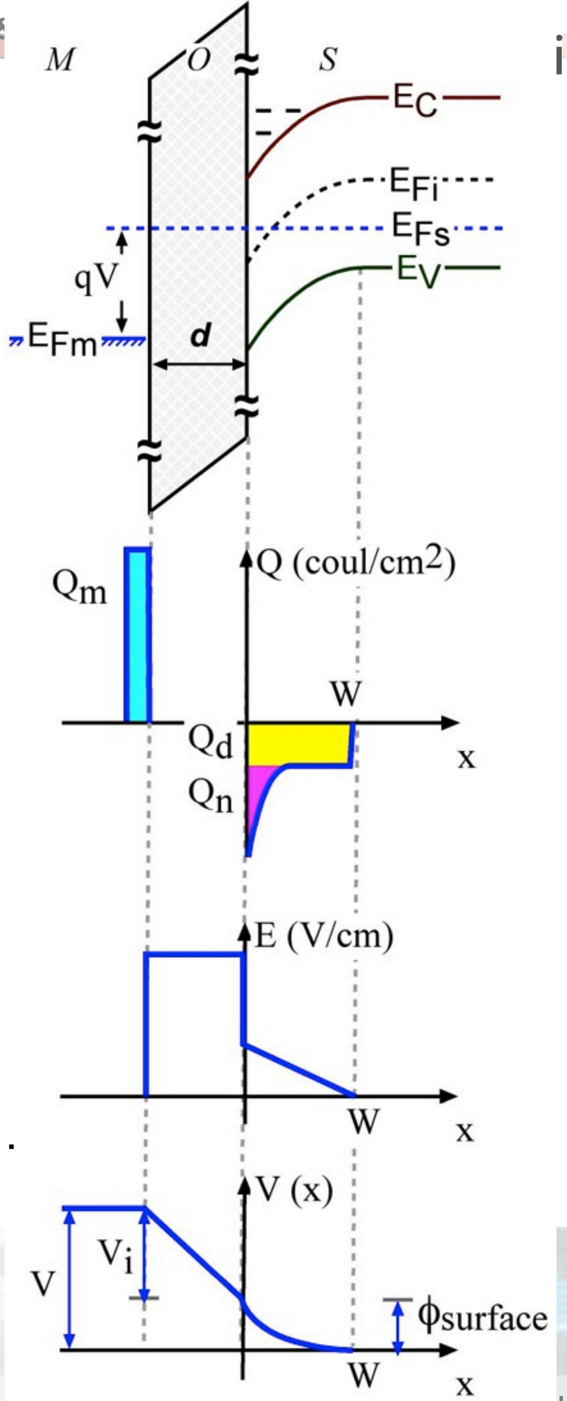
- Next, figure out how much depletion we need (we also have to pay with voltage for that too, right?)…. we can treat like a n+p junction and assume all deplete into p-side:

$$W = \left[\frac{2\varepsilon(V_0 - V_{app})}{q}\left(\frac{N_a + N_d}{N_a N_d}\right)\right]^{1/2}$$

$$W = \sqrt{\frac{2\varepsilon_s \phi_{surface}}{qN_a}}$$

Note similarity to Debye Length…

$$L_D = \sqrt{\frac{\varepsilon_s kT}{q^2 p_0}} \qquad \frac{kT}{q} = thermal\ volt.$$



Ideal MOSFET V_{TH}

- Like a PN junction, W increases as we apply more V and further deplete the p-type material…

- However, eventually inversion sets in exponentially and takes over the charge increase as voltage is added…

- Therefore depletion region (W) stops growing at a maximum value of:

$$W = \sqrt{\frac{2\varepsilon_s \phi_{surface}}{qN_a}} \leftarrow \phi_{surface}(inv.) = 2\phi_F = 2\frac{kT}{q}\ln\frac{N_a}{n_i}$$

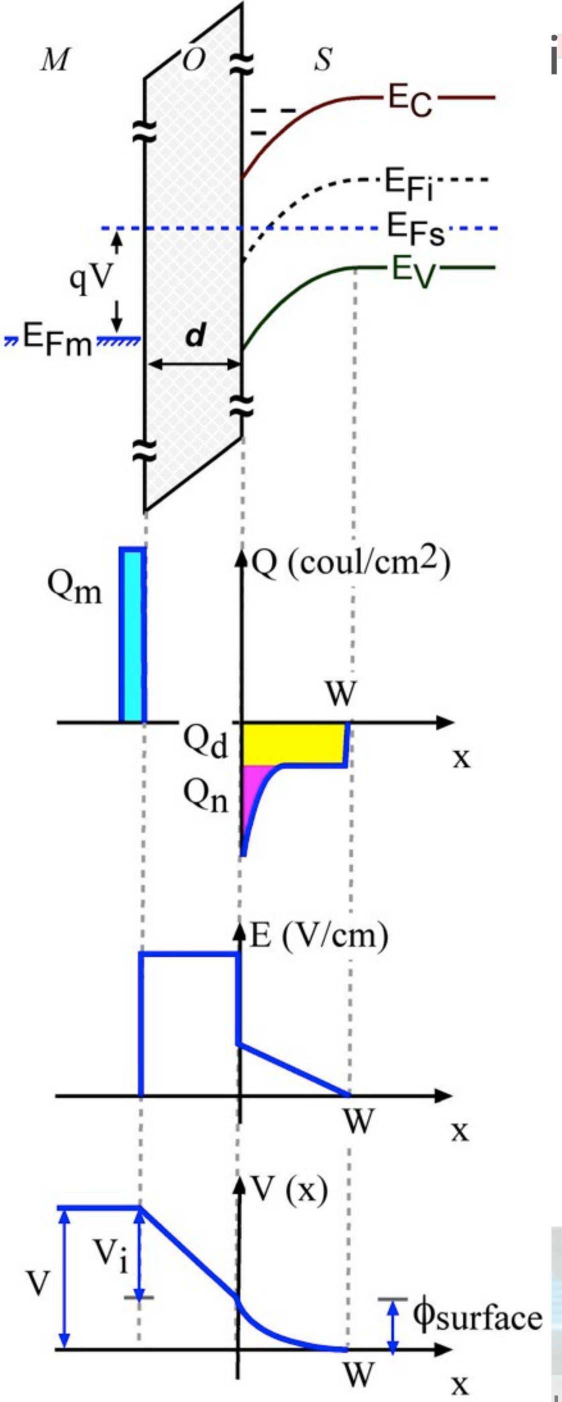
$$\Downarrow$$

$$W_m = \sqrt{\frac{2\varepsilon_s \phi_{surface}(inv.)}{qN_a}}$$

$$= 2\sqrt{\frac{\varepsilon_s kT \ln(N_a/n_i)}{q^2 N_a}}$$

☆

- Key point! W maximizes! Inversion takes over all the new Q at some point!



Ideal MOSFET V_{TH}

- We had already shown depletion charge:

$$Q_d = -qN_aW \quad (coul/cm^2)$$

- We can substitute W_m into Q_d

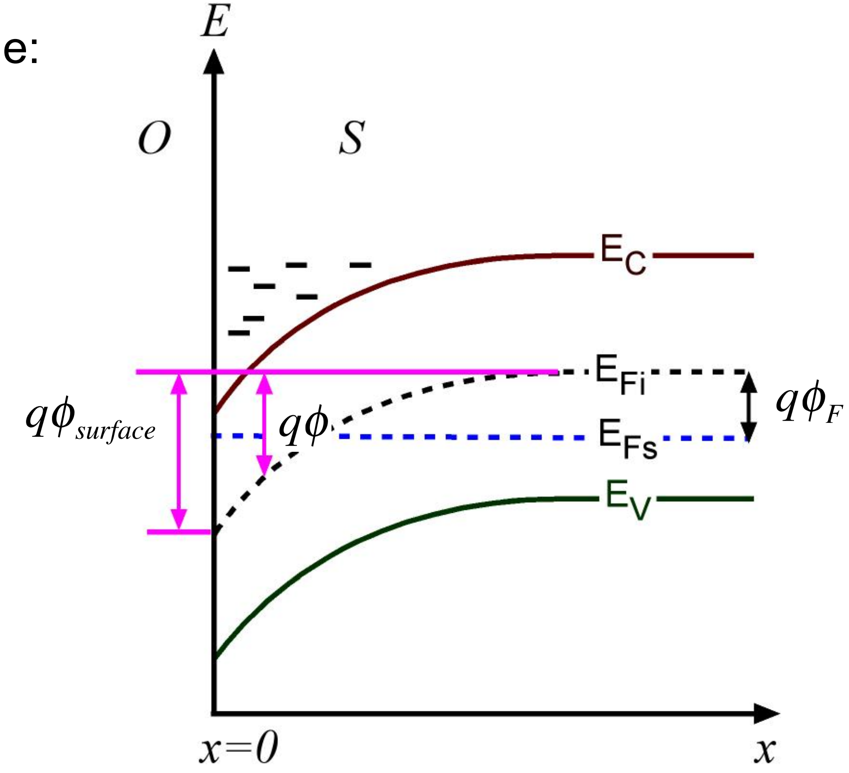
$$W_m = 2\sqrt{\frac{\varepsilon_s kT\ln(N_a/n_i)}{q^2N_a}}$$

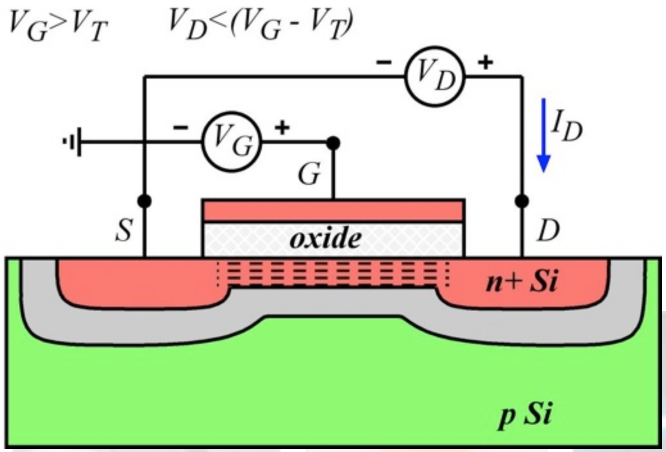
$$\frac{kT}{q}\ln\frac{N_a}{n_i} = \phi_F$$

- And obtain the maximum depletion charge as:

$$Q_{d,\max} = -2\sqrt{\varepsilon_s qN_a\phi_F} \quad (coul/cm^2)$$

☆ *HOW CAN WE TURN THIS INTO A VOLTAGE?*

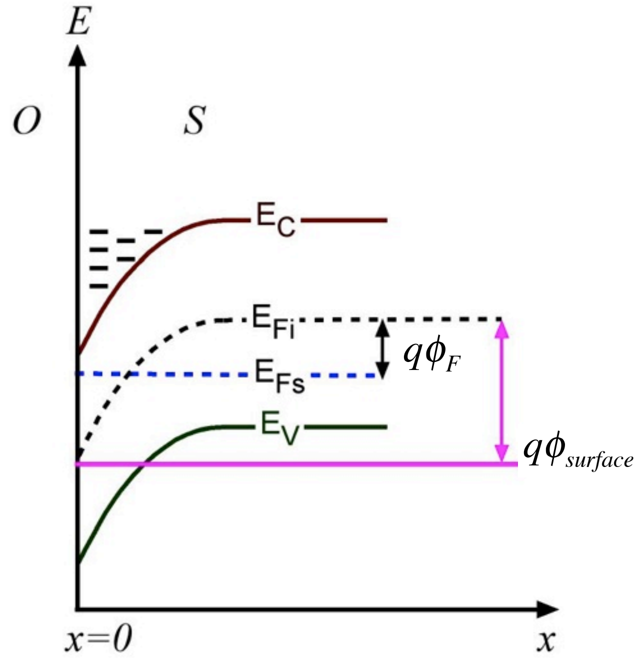




Ideal MOSFET V_{TH}

- Okay, lets calculate V_{TH}! 1st, recall for conducting channel we need to have **strong** Inversion, *the surface should be just as n-type as the substrate is p-type…*

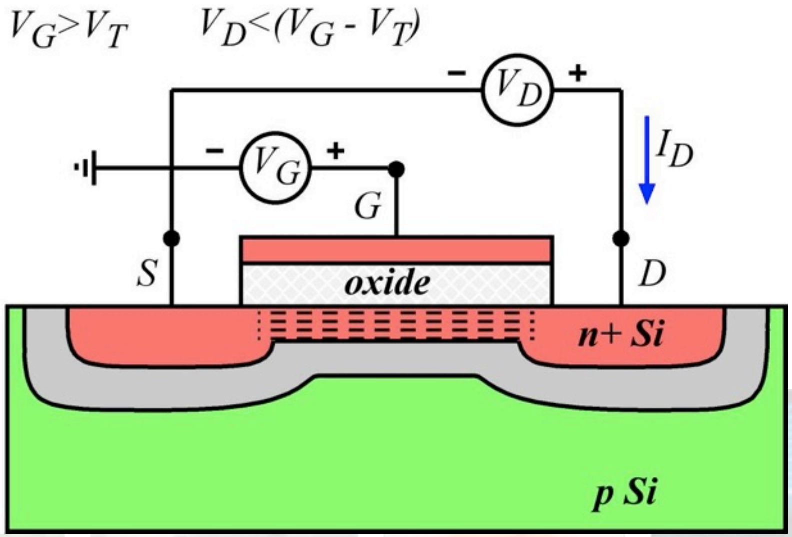
$$\phi_{surface}(inv.) > 2\phi_F = 2\frac{kT}{q}\ln\frac{N_a}{n_i}$$



- To get the threshold voltage we therefore need $2\phi_F$

- However, before we could achieve $2\phi_F$ we needed to max out the depletion region, and that creates charge, and using Q=CV means it requires additional voltage!

$$Q_{d,\max} = -2\sqrt{\varepsilon_s q N_a \phi_F} \quad (coul/cm^2)$$



- Therefore the 'ideal' case for MOSFET threshold voltage is:

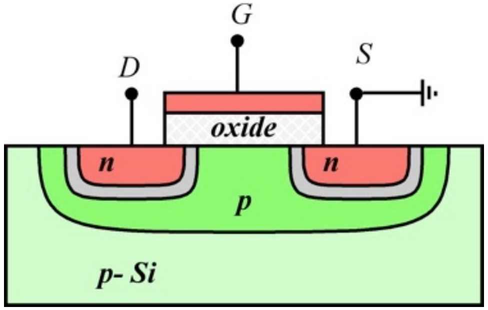
☆ $$V_T = -\frac{Q_{d,\max}}{C_i} + 2\phi_f$$

$$C_i = \varepsilon / t$$

Look Familiar?

At Intel: Andrew Grove (left), Bruce Deal (center) and Ed Snow (left)

Reducing V_{TH}

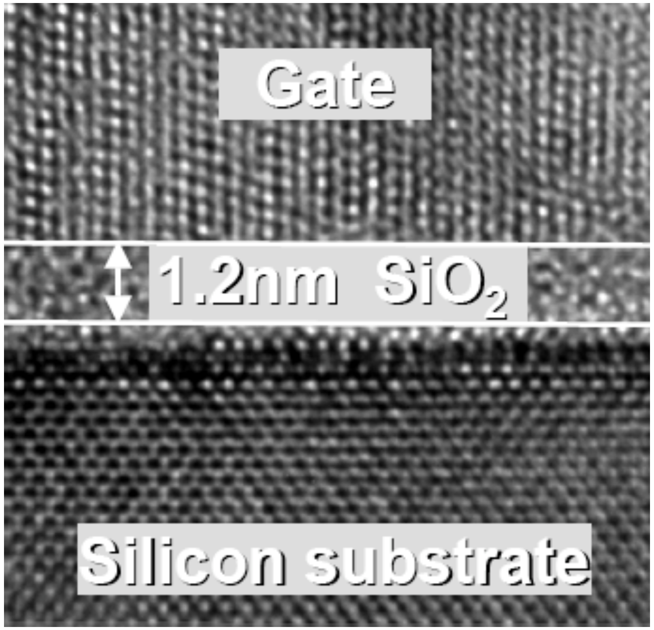


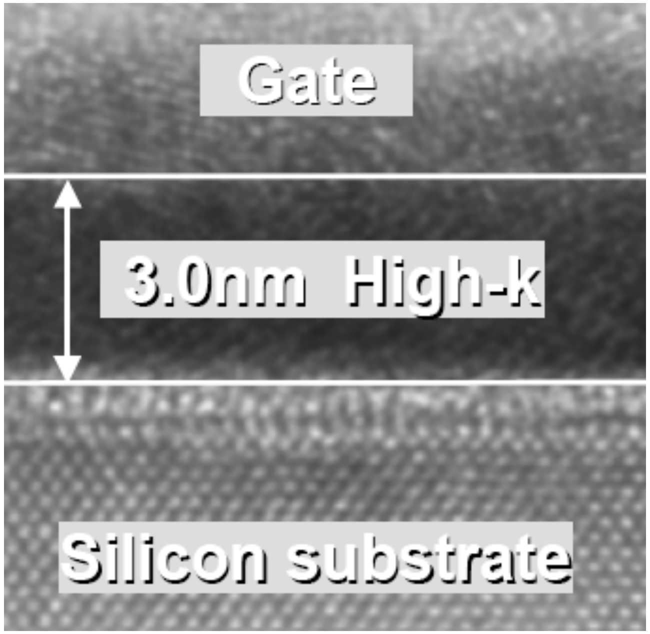
$$V_T = -\frac{Q_{d,\max}}{C_i} + 2\phi_f \qquad C_i = \varepsilon / t$$

☆

There is a limit to decreasing t (oxide thickness)… Why?

- Down to a few layers of SiO_2 molecules… chance for defects increases, and tunnel current increases… now using high k (ε) HfO_2 etc..

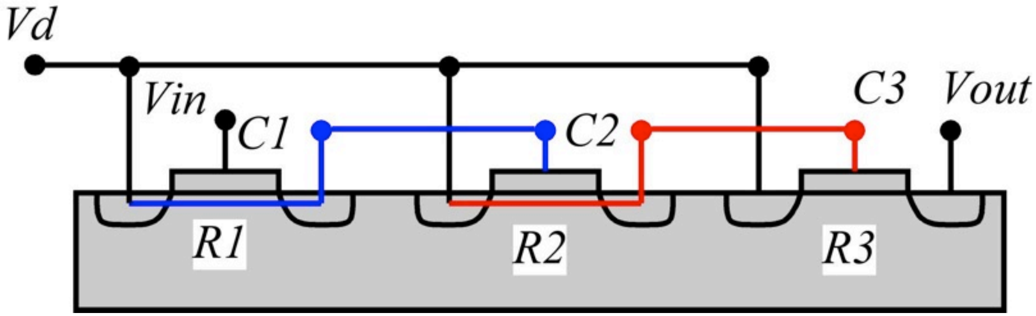




$\varepsilon_r \sim 4$ $\varepsilon_r \sim 18$

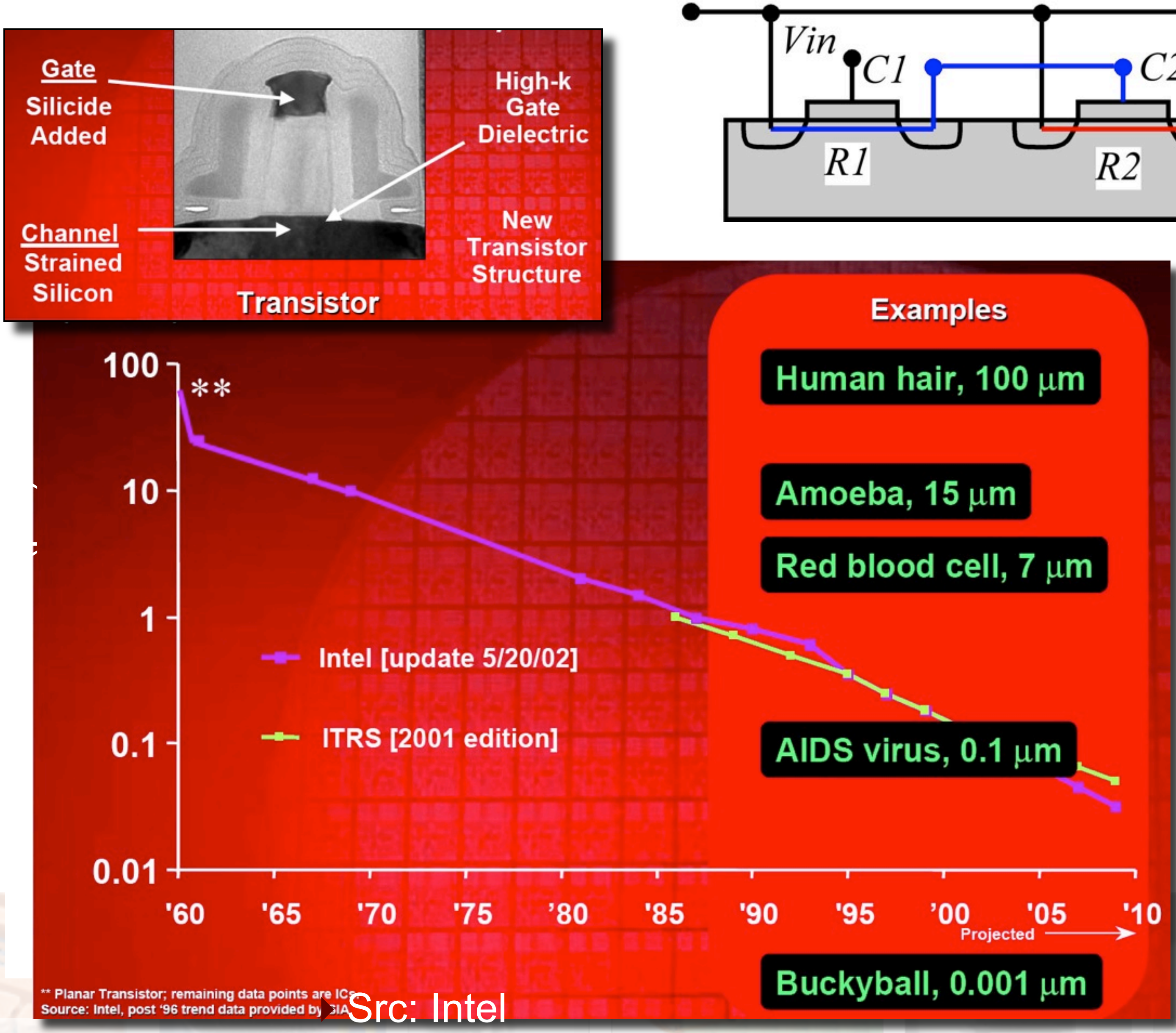
What is the capacitance difference? So why sometimes still use SiO_2?

Smaller, what happens to R&C? ☆



$$\tau = R1 \times C2$$

$$\tau = R2 \times C3$$



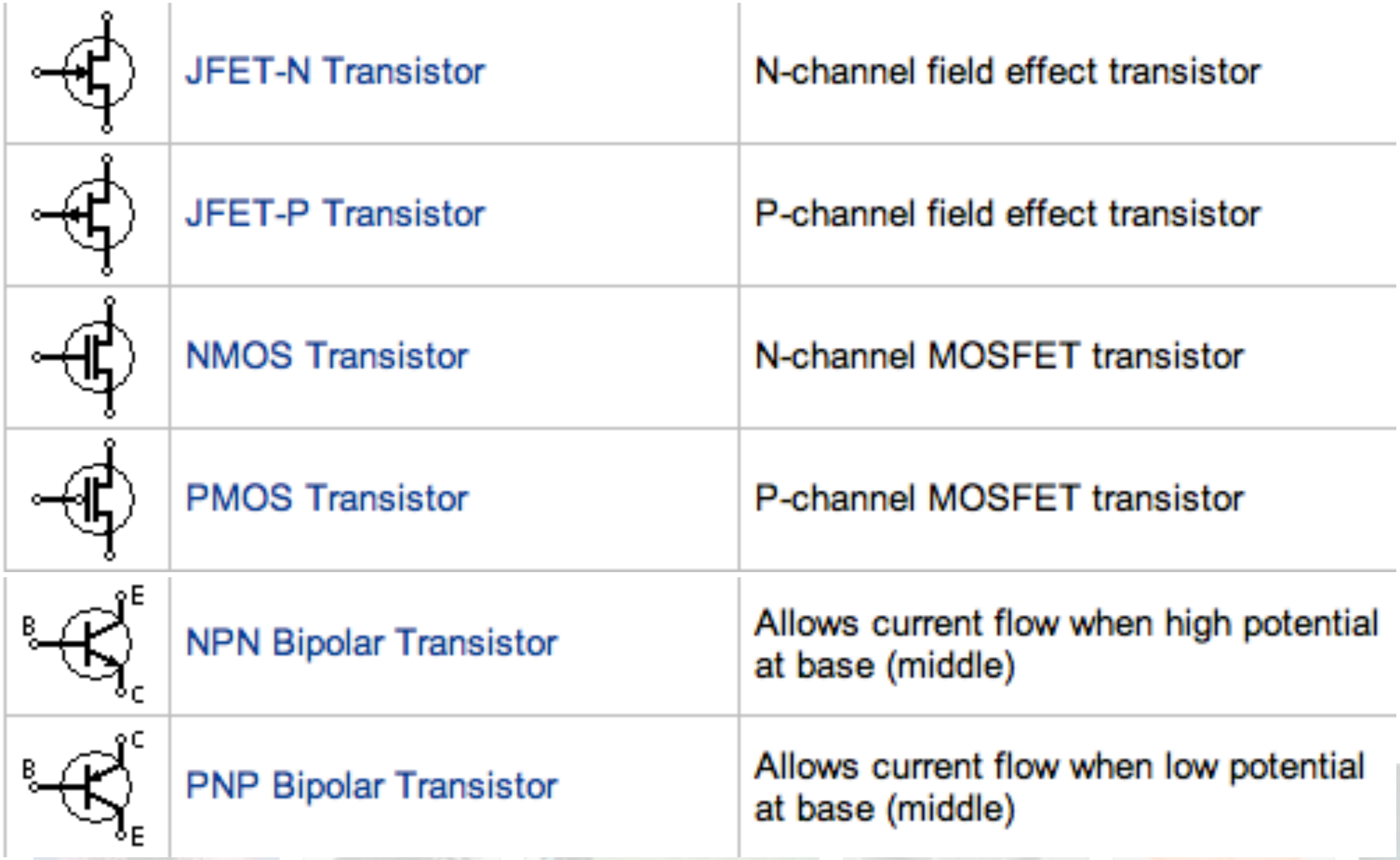
** Planar Transistor; remaining data points are ICs
Source: Intel, post '96 trend data provided by SIA

Src: Intel

So many transistor types!

- How to tell different transistors apart… (but you will find not everyone follows this!).

- So for MOSFETs, why the two parallel lines at the input? Why the ‘dot’ at the input of PMOS?
- For the JFETs, why no parallel lines and why the arrows at the input?



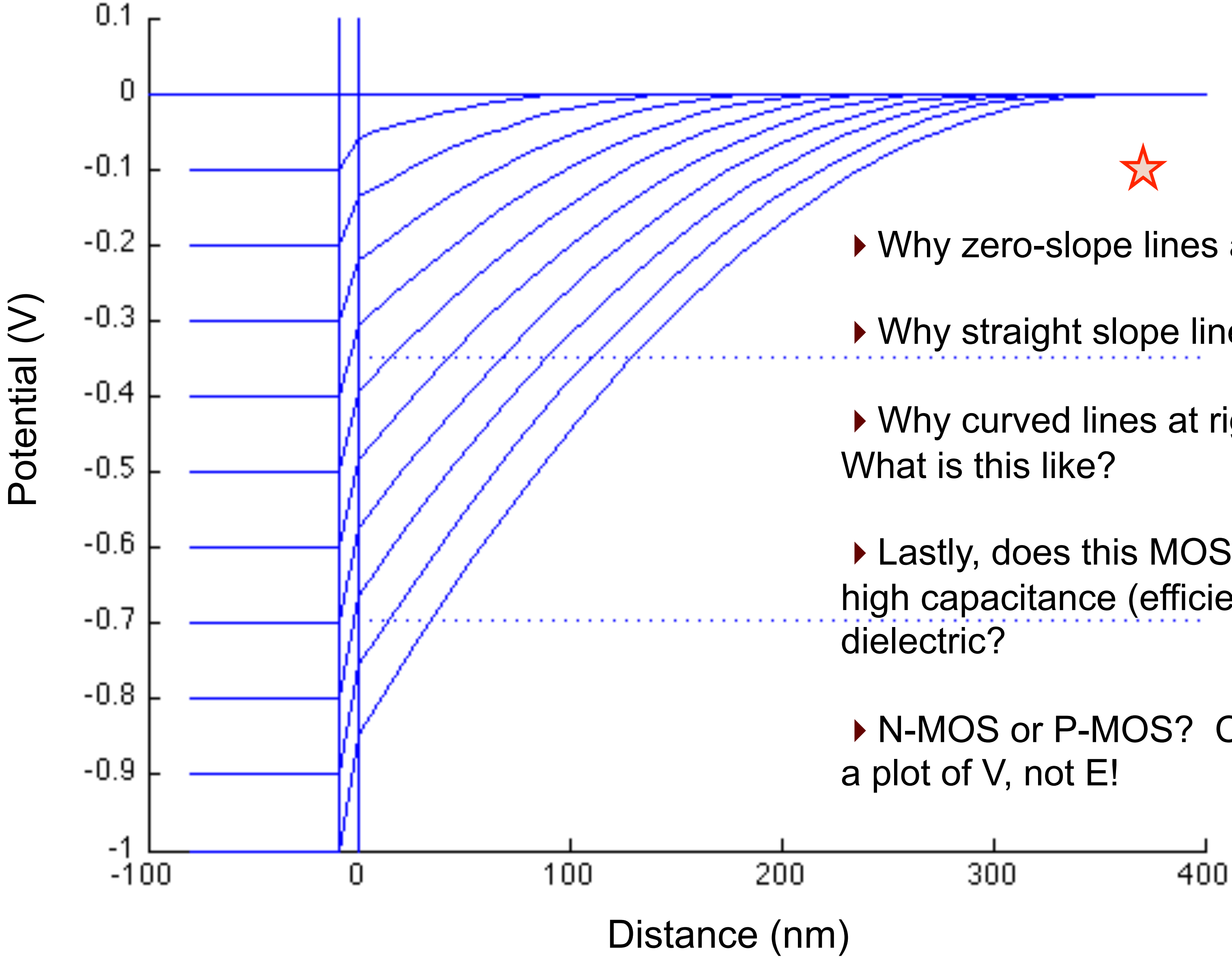
	JFET-N Transistor	N-channel field effect transistor
	JFET-P Transistor	P-channel field effect transistor
	NMOS Transistor	N-channel MOSFET transistor
	PMOS Transistor	P-channel MOSFET transistor
	NPN Bipolar Transistor	Allows current flow when high potential at base (middle)
	PNP Bipolar Transistor	Allows current flow when low potential at base (middle)

MATLAB Example (Pierret)

```
% Constants
eps_0 =   8.85e-14       ; % Units: F/cm
kToq  =   0.0259         ; % Units: V
q     =   1.6e-19        ; % Units: C
cm    =   1.0e4          ; % Units: micron

% Parameters for Silicon
eps_si =  11.8 * eps_0    ; % Units: F/cm
n_i    =   1.5e10         ; % Units: 1/cm^3
N_a    =   1.0e16         ; % Units: 1/cm^3  <--- INPUT DOPING DENSITY


% Parameters for Oxide
eps_ox =   3.9 * eps_0    ; % Units: F/cm
d_ox   =  10.0            ; % Units: nm     <--- INPUT OXIDE THICKNESS


Phi_F  =  kToq * log( N_a/n_i)   ;
C_ox   =  eps_ox / (d_ox * 1e-7) ;
W_max  =  2 * sqrt( (eps_si*Phi_F) / (q*N_a) ) ;
Q_d    =  q * N_a * W_max  ;
V_T    =  Q_d / C_ox + 2 * Phi_F  ;

V  = linspace(0,1,11);          % Define Range of Depletion Bias Voltages
A = (C_ox/eps_si)*(C_ox/2/q/N_a) ;  a =1/(2*A) ;
V_ox = -a + sqrt(a*a + V/A) ;
V_si = V - V_ox             ;

W = sqrt( 2*eps_si*V/q/N_a) ;
```

```
X_si = linspace(0,500,51) ;
X_ox = [-10 0] ;
X_m  = linspace(-50,-10,5) ;
XX   = [X_m X_si];

axis( [-100 400 -1 .1]); hold on
plot( [-100 400] , [ 0 0] ) ; hold on
plot( [0 400] , [-Phi_F -Phi_F], ':') ; hold on
plot( [0 400] , [-2*Phi_F -2*Phi_F], ':') ; hold on
plot( [0 0] , [-1 .1] ) ; plot( [-10 -10] , [-1 .1] ) ;

for iV =1:11

Vm = V(iV)     ;
Vi = V_ox(iV)  ;
Vs = V_si(iV)  ;
WW = W(iV)*cm*1000 + 0.1 ;

xV = linspace(0,WW,15) ;
vV = Vs * (xV/WW - 1) .* (xV/WW - 1) ;


plot( [-80 -10]  , [-Vm -Vm] ) ; hold on
plot( X_ox , [ -Vs-Vi , -Vs] ) ; hold on
plot( xV   , -vV) ; hold on

end
```

MATLAB Example (Pierret)

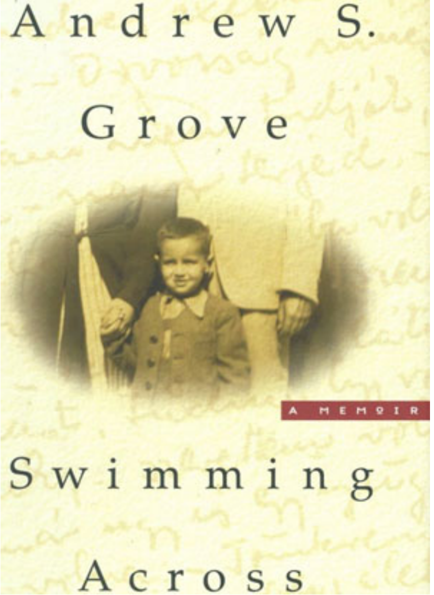


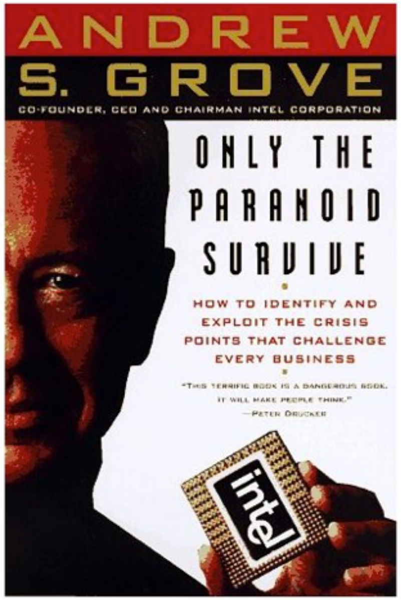
☆

- Why zero-slope lines at left side?
- Why straight slope lines in middle?
- Why curved lines at right side? What is this like?
- Lastly, does this MOSFET have a high capacitance (efficient) gate dielectric?
- N-MOS or P-MOS? Careful, this is a plot of V, not E!

Some suggested reading…

- Last Topic… Who is Andy Grove?



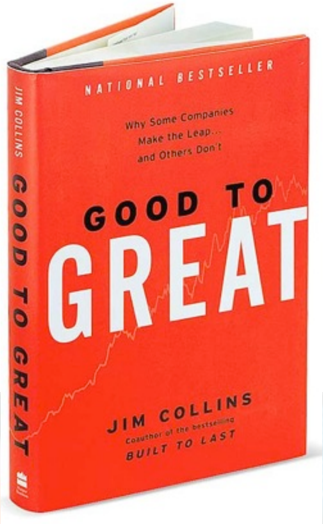

ANDREW S. GROVE
CO-FOUNDER, CEO AND CHAIRMAN INTEL CORPORATION
ONLY THE PARANOID SURVIVE
HOW TO IDENTIFY AND EXPLOIT THE CRISIS POINTS THAT CHALLENGE EVERY BUSINESS
"THIS TERRIFIC BOOK IS A DANGEROUS BOOK. IT WILL MAKE PEOPLE THINK." —PETER DRUCKER
intel

1997

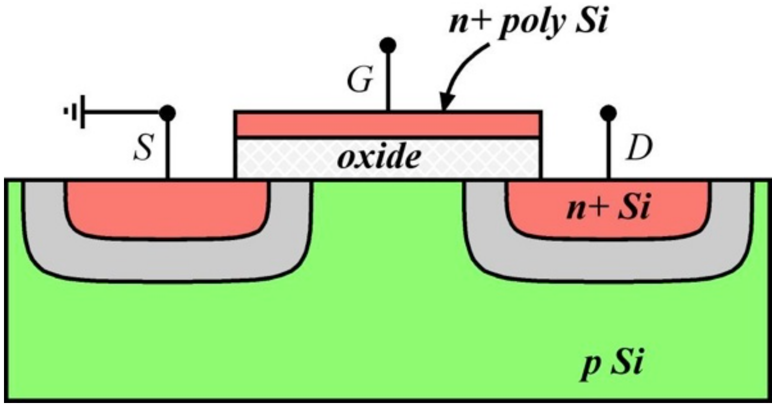
- First read 'Only the Paranoid Survive' to appreciate what he accomplished as an Engineer.

- Second, read 'Swimming Across' to realize how fortunate you are to be here as an EE.

- Third, a business-related book every engineer should read is 'Good to Great'.


NATIONAL BESTSELLER
Why Some Companies Make the Leap... and Others Don't
GOOD TO GREAT
JIM COLLINS
Coauthor of the bestselling BUILT TO LAST
JIM COLLINS GOOD TO GREAT

Review!



- The depletion I can create under the gate oxide maximizes, why? *Hint, something else takes over that dominates mathematically in terms of charge generation…*

- Once my surface potential is (ϕ_{surf}) is above threshold voltage, at what mathematical rate are carriers created in the channel? *Hint: think back to that Qs plot…, think how we calculate carrier concentration*.

Note – we will see next lecture, however, that surface potential and external gate voltage are not proportional…

- Why don't we see strong exponential increase in carriers until the band-bending (ϕ_s) reaches $2\phi_F$? *Hint: see the Qs plot*

- How does the MOSFET gate voltage change if I reverse all the doping types? *No hint needed!*

- Why can't I make the oxide thinner and thinner? *One word answer will do!*

- Why do <u>smaller</u> MOSFETs make faster chips? *But remember, more transistors per unit area means more heat generation… which is often the limiter today…*