Chapter 11

Transistors

The MOSFET

The Metal Oxide Semi-conductor FET (MOSFET)

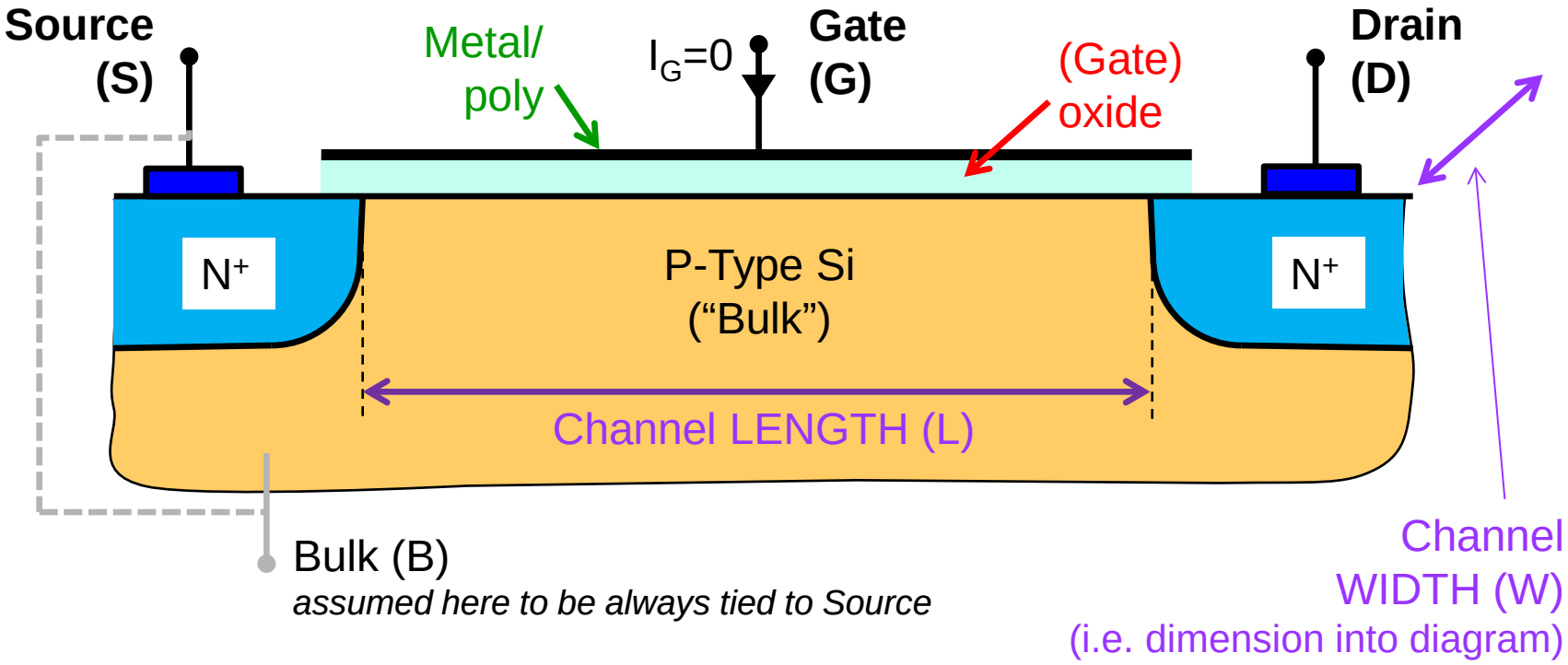
The MOSFET

- "MOSFET" stands for Metal Oxide Semiconductor Field Effect Transistor, where the "semiconductor" is understood to be Silicon and the "oxide" is usually silicon dioxide (SiO_2), which is an excellent insulator. The "metal" may be a material such as Al or Cu or is more likely to be highly doped polycrystalline Silicon ('poly');
- The fact that Silicon has a stable, easily grown oxide is immensely important in its dominance as a semiconductor material (e.g. GaAs has no stable native oxide);
- The MOSFET is by far the most important type of transistor now in production. It is used in nearly all digital ICs and also in more-and-more analogue functions

A Short History of the MOSFET

- The MOSFET device was proposed by Kahng and Atilla working in Bell Labs in the USA around 1960
- Early devices were plagued by drift or time-dependent changes in characteristics. Eventually it was realised that strict cleanliness and material purity are needed for reliable operation
- MOSFETs come in two basic varieties, P-channel or P-MOS (holes dominate) and N-channel or N-MOS (electrons dominate)
- Early ICs used P-MOS due to easier of fabrication, but these were soon replaced by N-MOS ICs with better performance
- The idea of using both N-MOS and P-MOS together to form Complementary MOS or CMOS is credited to Frank Wanlass of Fairchild in 1963, although the technology took a long time to be established. CMOS is now by far the most dominant IC family.

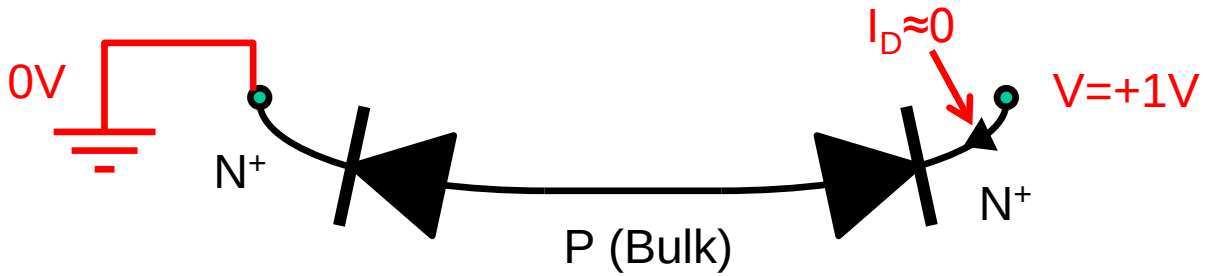
Simplified Structure of N-Channel MOSFET



- A critical dimension for device scaling is the channel Length 'L'. This has been driven down aggressively over time through Moore's Law;
- Note that there can be no DC current through the Gate in a MOSFET.

The Threshold Voltage

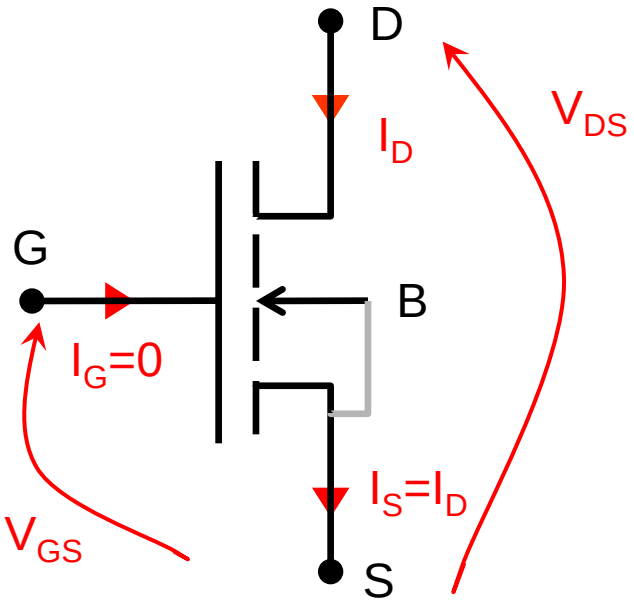
- Consider the Source at ground (0V), with no voltage on the Gate and say +1V on the Drain. We can model the path from S-to-D as follows:



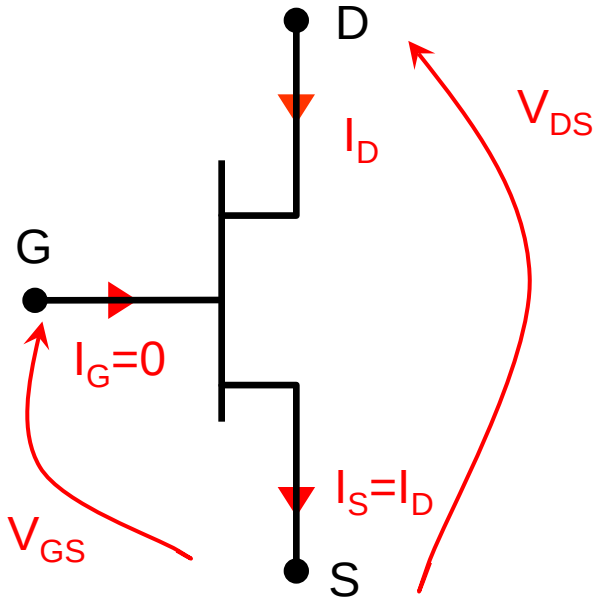
- The **Drain current (I_D)** is essentially zero no matter what the polarity of the Drain voltage, since one diode is always reverse biased
- For an N-channel MOSFET, it is found that if an increasing (positive) voltage is placed on the Gate, then at a particular critical value called the **Threshold Voltage** and denoted by **V_T**, a thin, dense layer of electrons (called an ***inversion layer***) forms near the surface creating a conducting channel between S and D, allowing current to flow.
- The MOSFET is thus normally an **Enhancement Mode FET**

N-MOSFET Circuit Schematic Symbol

Full Symbol

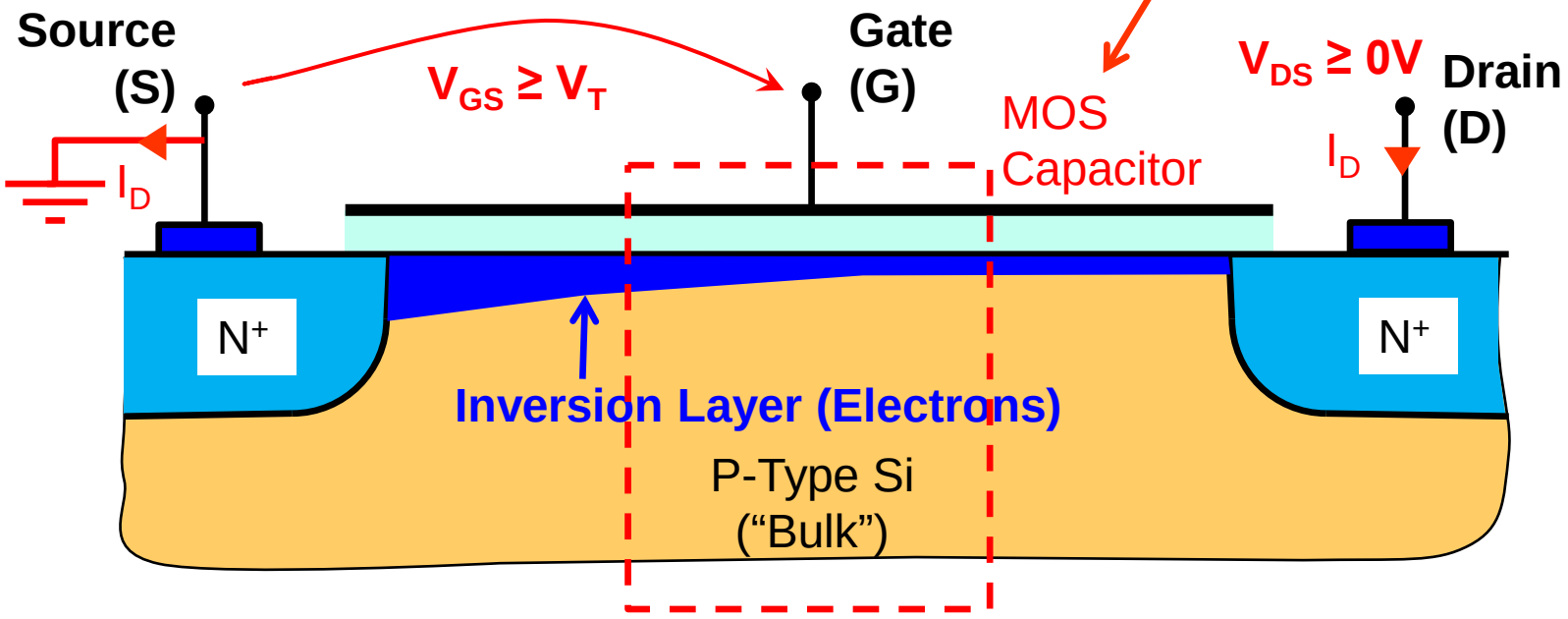


Simplified Symbol



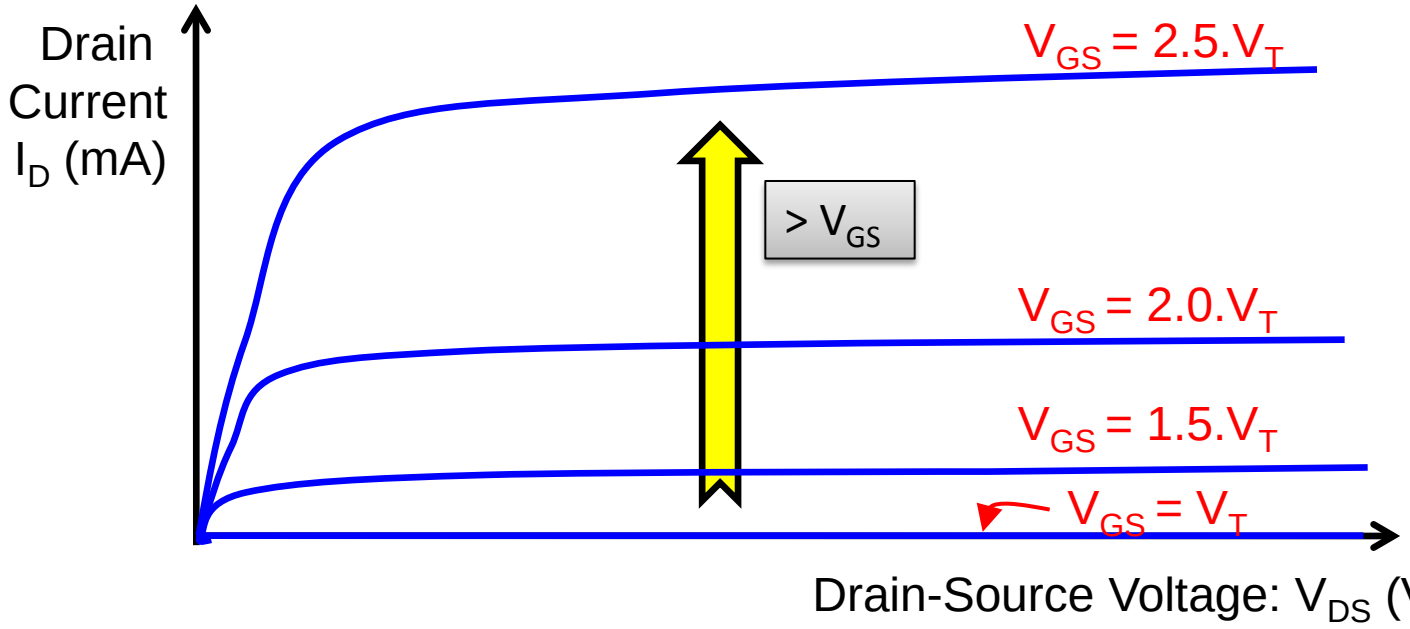
Inversion Layer in an N-Channel MOSFET

- A natural question is how does the Inversion Layer come to be formed – this will be answered shortly through a detailed study of a simpler structure, the MOS capacitor
- For now, we just consider the observed terminal behaviour of the device:



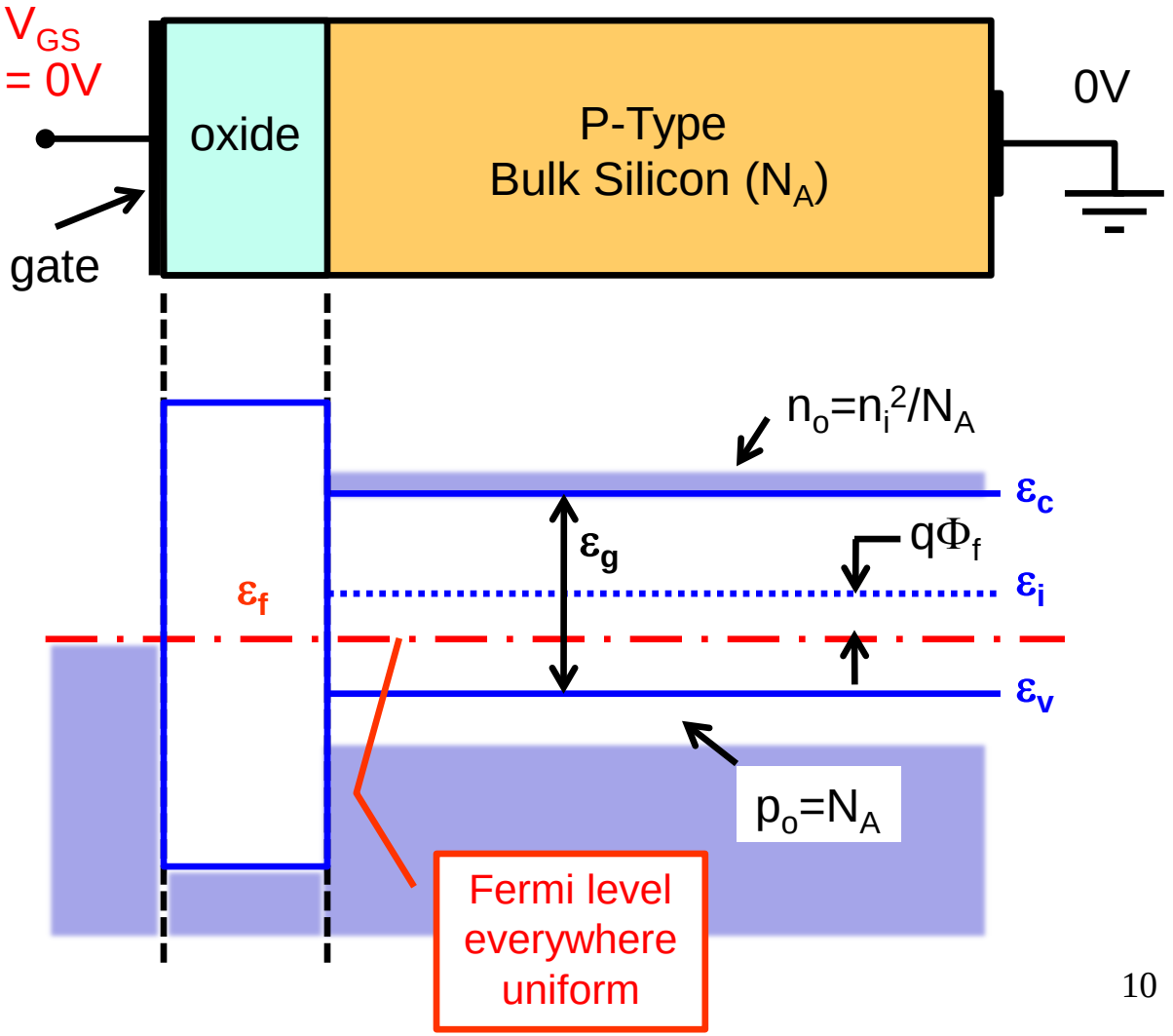
N-MOSFET DC Drain Characteristics

- When $V_{GS} < V_T$ negligible Drain current flows. The more V_{GS} exceeds V_T (which is typically ≈1V), the stronger the inversion layer (IL) becomes, and the more drain current flows
- However, for a fixed $V_{GS} > V_T$, making V_{DS} larger and more positive causes the IL to weaken at the Drain end, which is found to lead to Drain current saturation, similar to that found in JFETs:



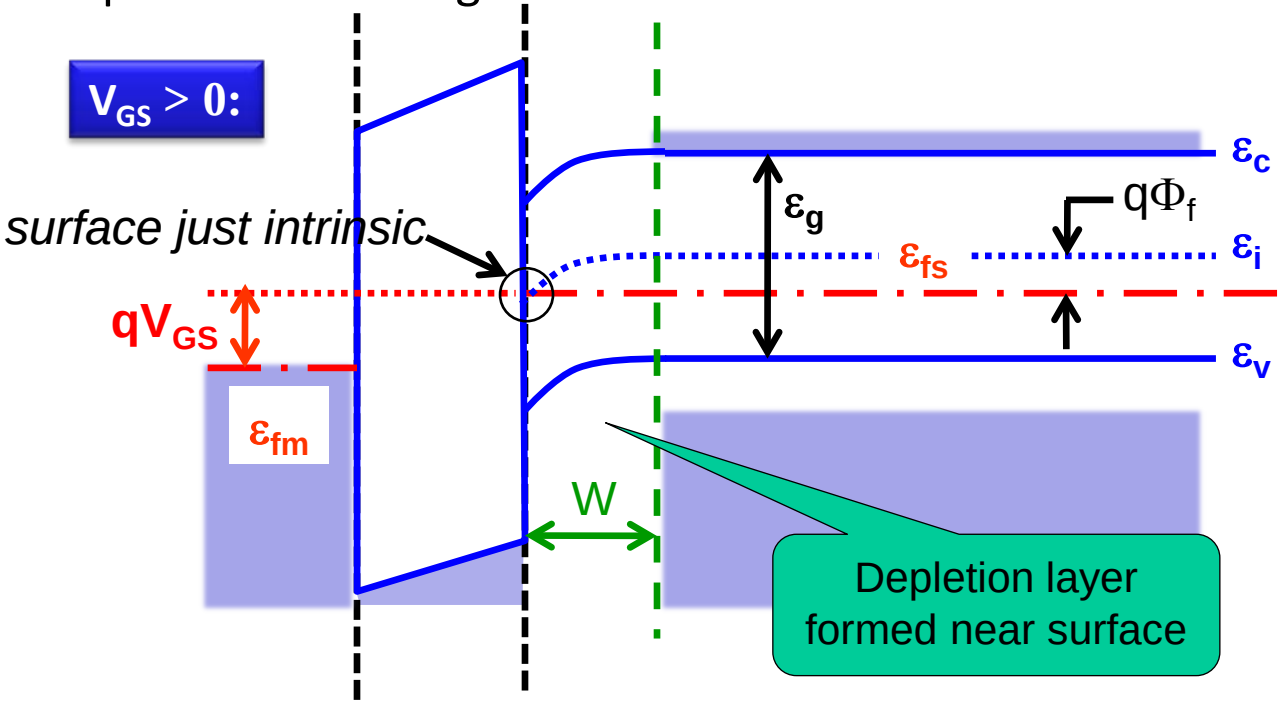
The **MOS Capacitor** – Thermal Equilibium

We can understand how the ***IL*** comes to be formed by considering the **MOS capacitor** structure (rotated thro' 90° for conv-enience) ["Flat-Band" assumption]



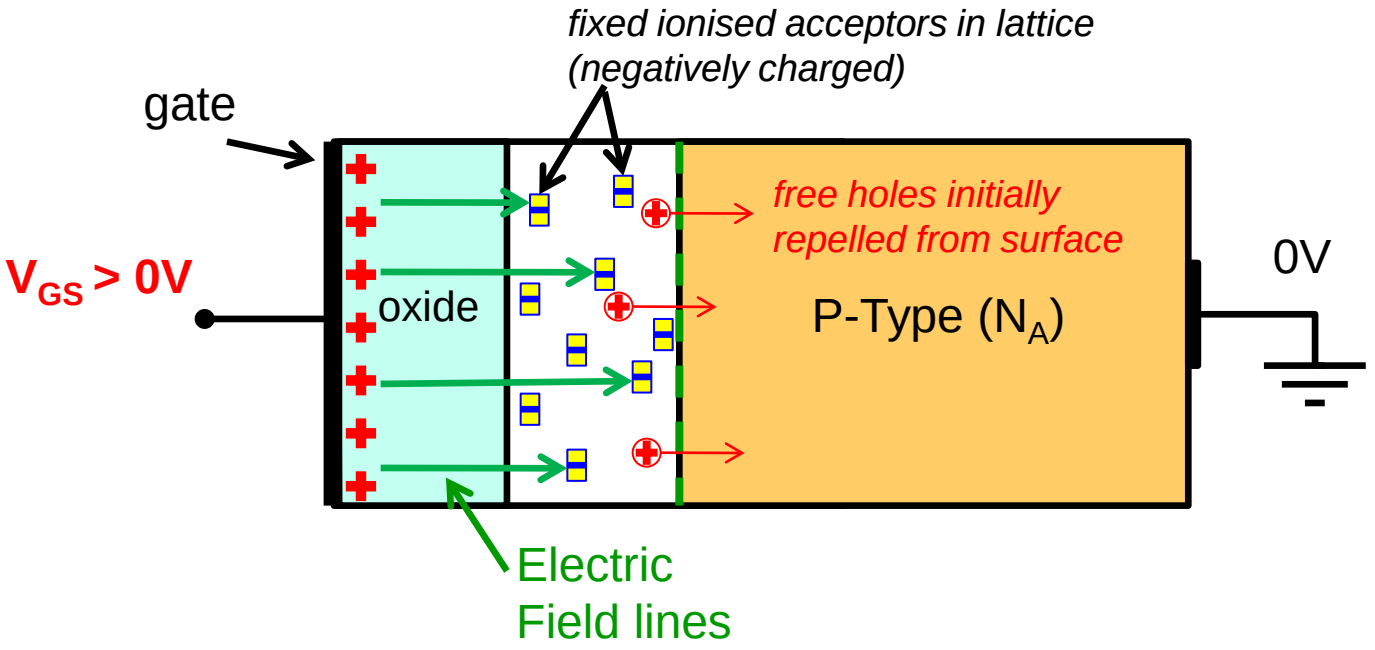
Quasi-Equilibrium in MOS-C

- If we apply $V_{GS} > 0$, no current can flow (due to the oxide) but the free holes near the surface are repelled away to form a depletion layer;
- The fact that the oxide is a barrier to flow means that the Fermi levels are separately constant in the metal and semiconductor, respectively (ε_{fm} and ε_{fs}). This is referred to as ***quasi-equilibrium***.
- Note that the energy bands must then **bend** to accommodate this hole depletion as the region near the surface becomes less P-type .



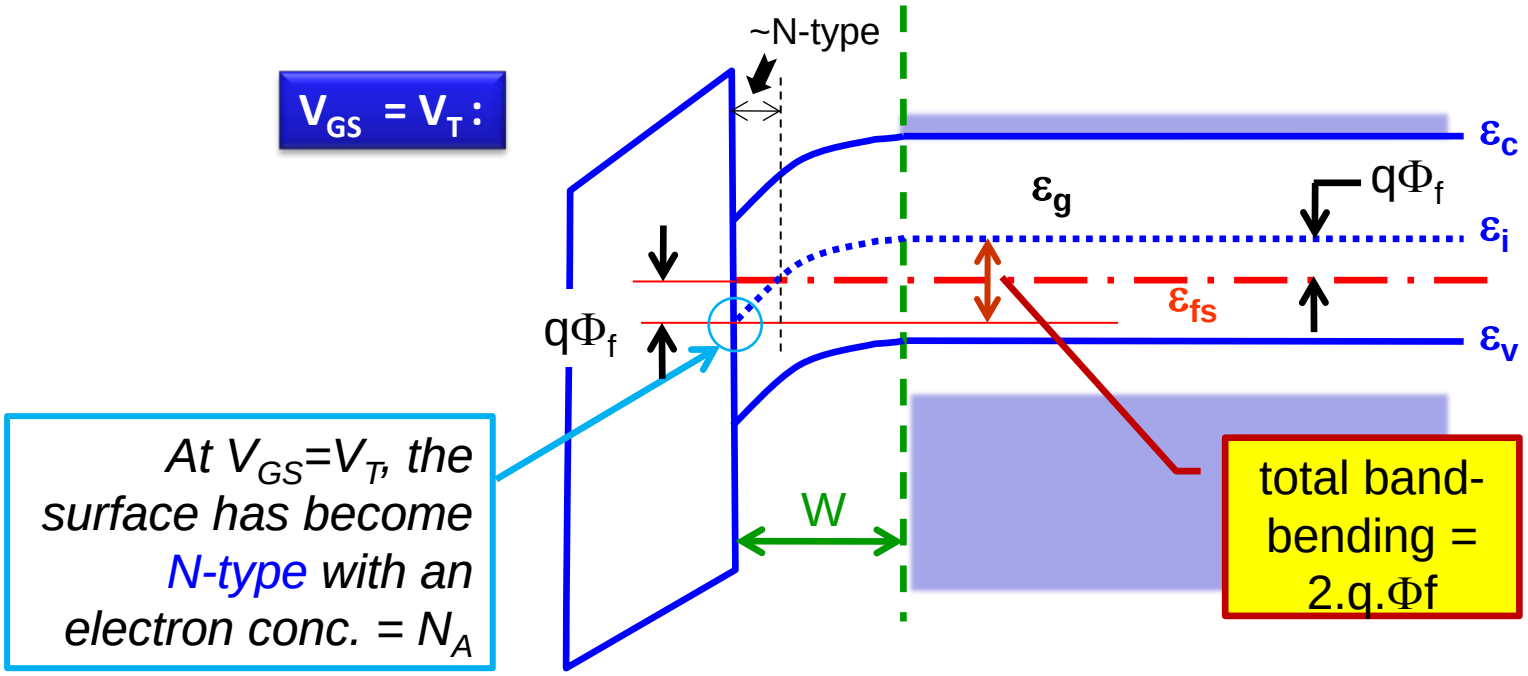
Charge Balance Picture

- A positive voltage on the Gate can be viewed as introducing a sheet of positive charge on the metal Gate
- Free holes in the P-type semiconductor are repelled from a region close to the surface, exposing ionised acceptors
- Electric field lines originating from the positive charges terminate on the negatively charged acceptor atoms

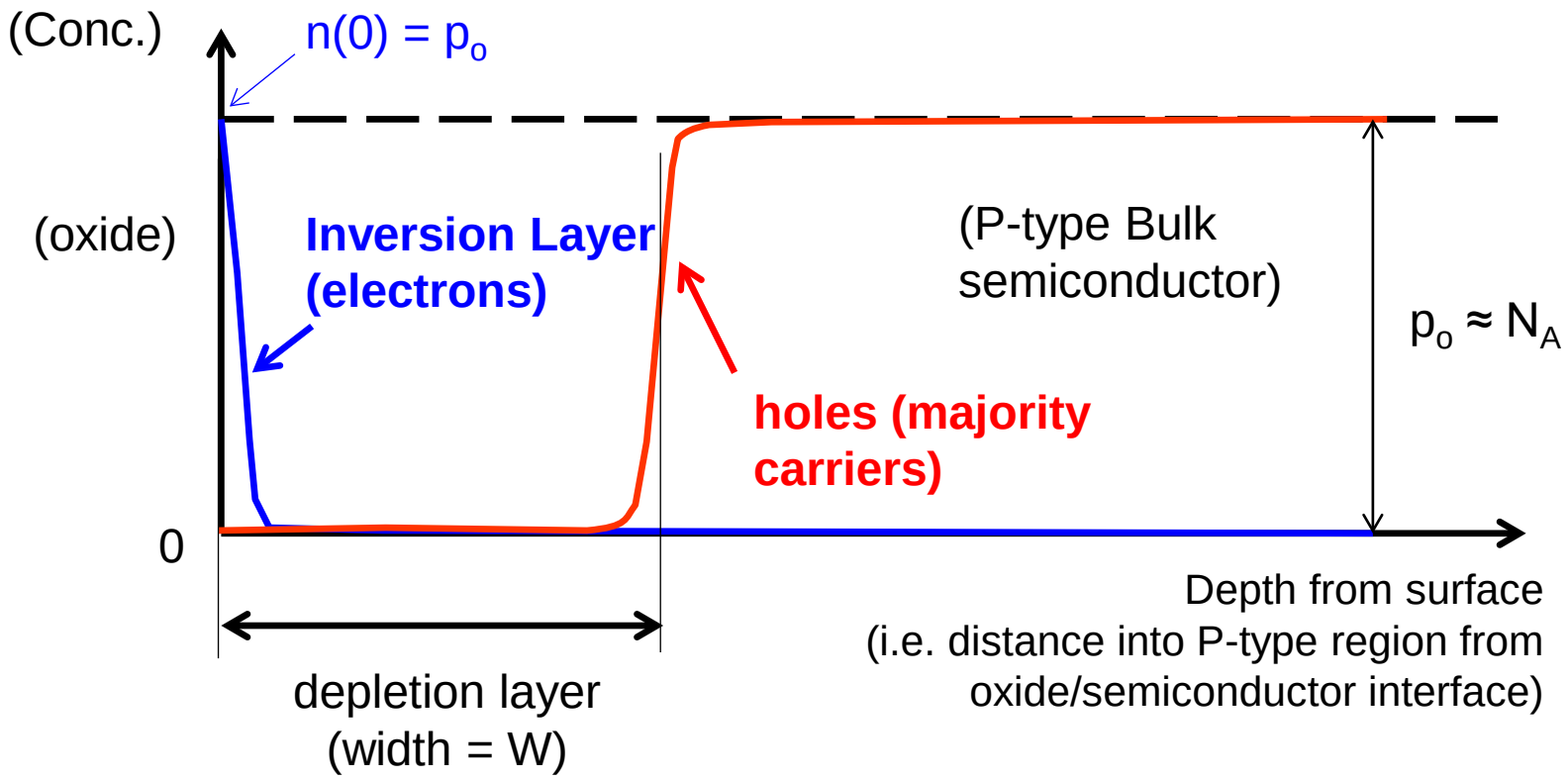


Threshold Voltage in MOS-Capacitor

- As $V_{GS} > 0$ is increased, the bending of the energy bands in the semiconductor continues until the intrinsic level at the surface falls *below* ε_{fs} - the surface then becomes 'inverted' from P-type to N-type.
- Further increase in V_{GS} eventually causes a critical condition when the surface has an electron concentration exactly equal to the bulk hole concentration. This value of V_{GS} is the **Threshold Voltage V_T**



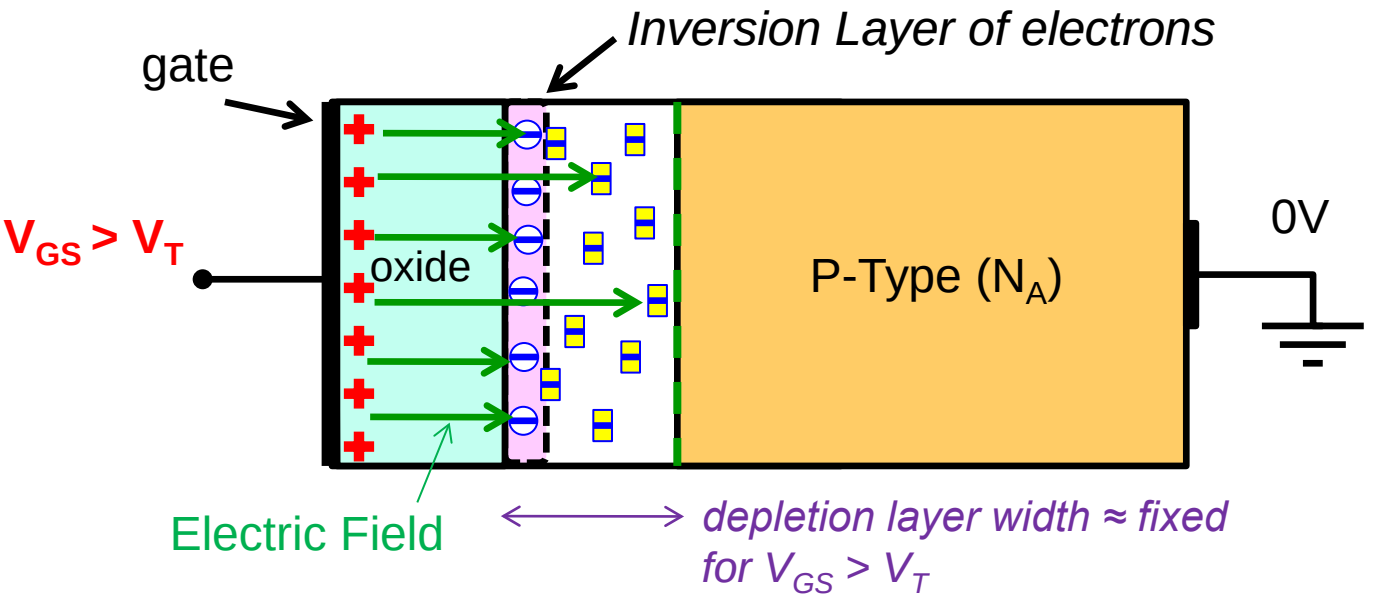
Internal Carrier Concentrations in the Semiconductor at the Threshold Voltage $V_{GS} = V_T$



http://jas.eng.buffalo.edu/education/mos/mosCap/biasBand10.html

Inversion Layer Formation in MOS-C

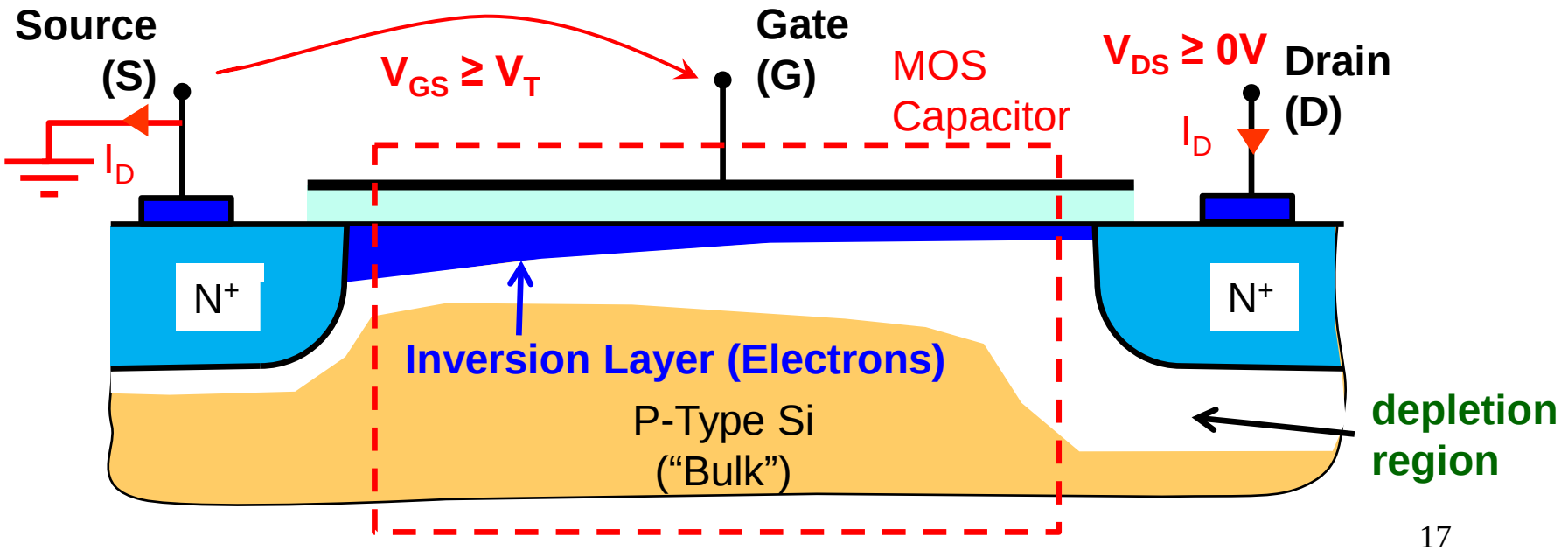
- Any increase in V_{GS} beyond V_T in the MOS-capacitor structure quickly produces a dense layer of electrons just beneath the oxide/ semiconductor interface – an ***inversion layer***
- Electric field lines from the Gate increasingly terminate on inversion layer charge: i.e. the depletion layer width tends not to increase further (cf. oxide electric field and a parallel plate capacitor)



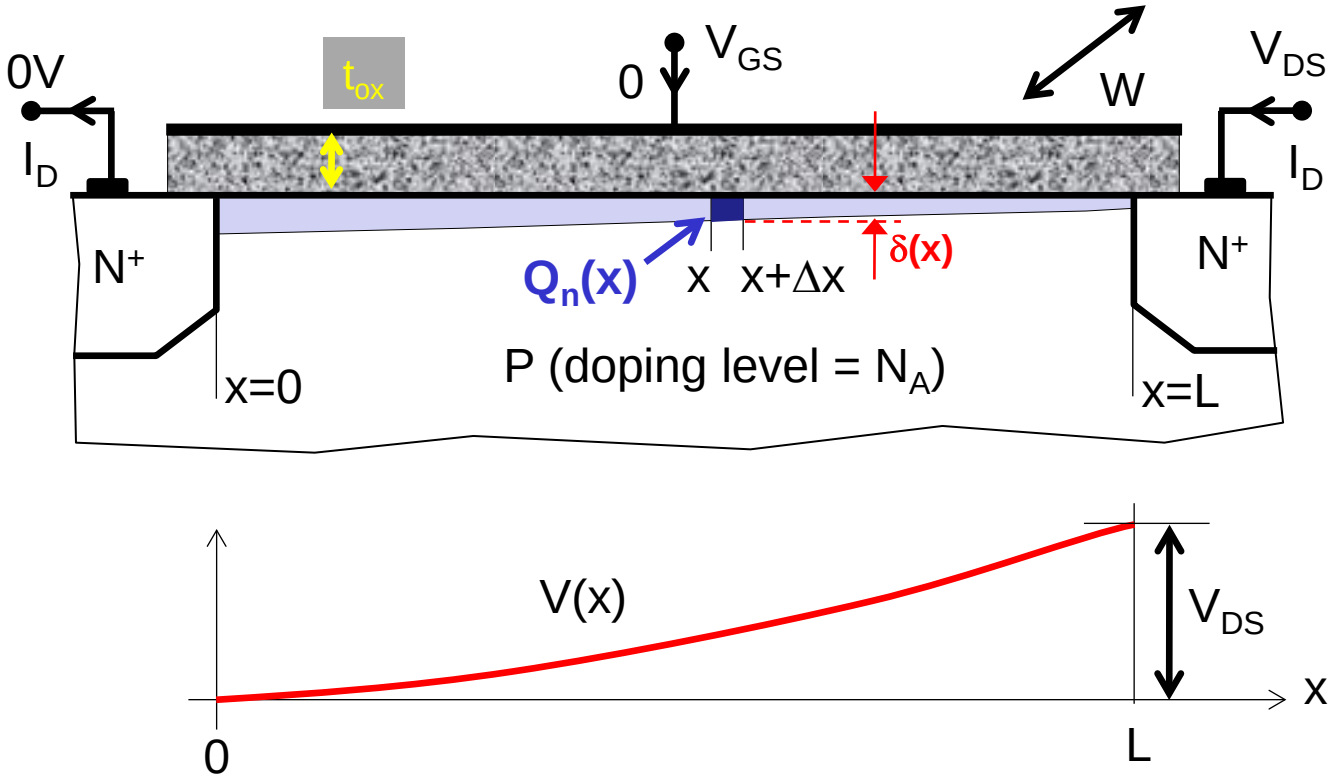
EXAMPLE 11.1

N-Channel Enhancement-Type MOSFET

- In the MOS-C, the inversion layer can be slow to form, as it relies on thermal EHP generation to produce the electrons. We can view the MOSFET as just a MOS-C with a Source and Drain on either side.
- The availability of the N^+ regions means that the IL forms very quickly
- Note that the MOSFET has **intrinsic isolation** due to the depletion region that surrounds the active zone – a huge advantage in ICs



Simplified N-Channel MOSFET



- V(x) is the channel potential, which varies along the Inversion Layer, with a value 0 at x=0 and V_{DS} at x=L

Surface Charge Density $Q_n(x)$

- We define $Q_n(x)$ as the surface charge density of the Inversion Layer (units: Coul/m^2). If the depth of the IL at a particular 'x' is $\delta(x)$, then the electron concentration at that 'x' (=n(x)) is related to $Q_n(x)$ as: $Q_n(x) = q.n(x).\delta(x)$;
- As a first approximation to relating $Q_n(x)$ to V_{GS}, we ignore the effect of depletion layer charge, and simply consider that once the potential difference across the oxide at any 'x' exceeds V_T, then the oxide behaves as an ideal parallel plate capacitor, with the IL as one "plate";
- Hence we can write: $Q_n(x) = C_{OX}.[(V_{GS} - V_T) - V(x)]$

  where $C_{OX} = (\varepsilon_{ox}/t_{ox})$ is the *oxide capacitance per unit area*
  (Note: $\varepsilon_{ox} = \varepsilon_{rox}.\varepsilon_o$, where $\varepsilon_{rox} = 3.82$ for SiO_2)

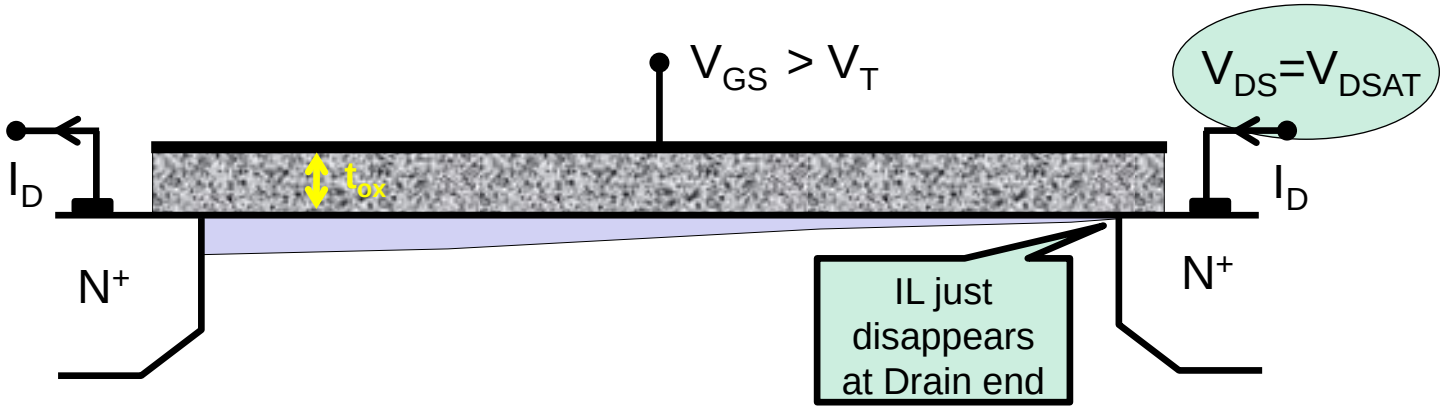
Drain Pinch-Off

- Suppose $V_{GS} \geq V_T$ and $V_{DS} \geq 0$. Notice that as V(x) rises along the channel it tends to oppose the tendency for V_{GS} to create the Inversion Layer, hence the IL weakens towards the Drain end of the channel;
- This causes I_D to rise sub-linearly with increasing V_{DS} at a given V_{GS} ("triode" region operation)
- A critical point is reached when V(L) is such that: $Q_n(L) = C_{OX}.[V_{GS} - V_T - V(L)] = 0$, i.e. when the IL just disappears at the Drain end. This is called Drain pinch-off and occurs at the following value of $V(L) = V_{DSAT}$:

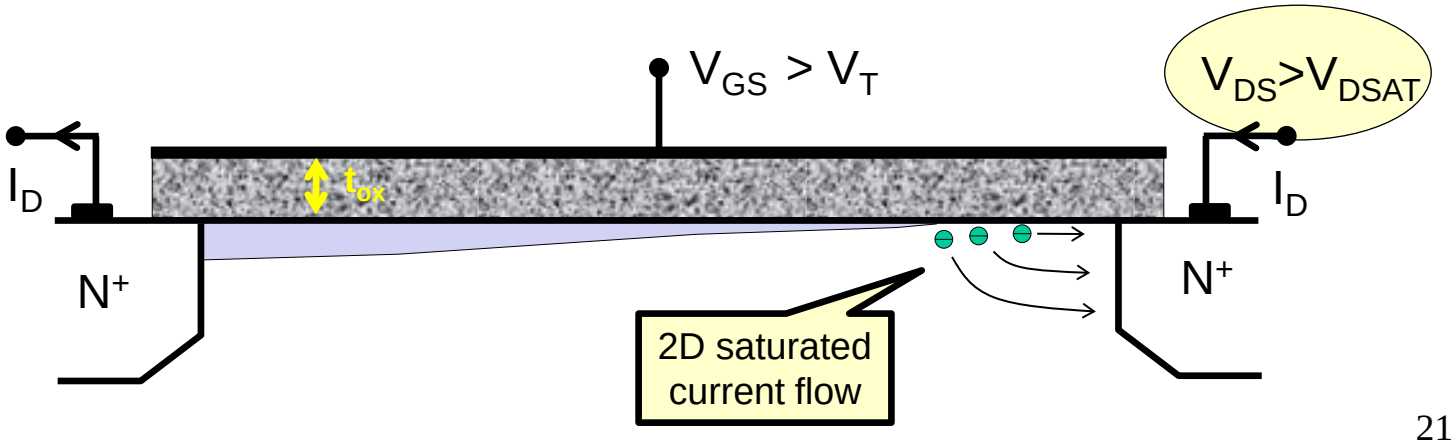
$$V_{DSAT} = (V_{GS} - V_T)$$

- As V_{DS} is increased beyond V_{DSAT}, the Drain current I_D tends to saturate to an approximately constant value, increasing slowly with V_{DS} ("saturation" region operation)

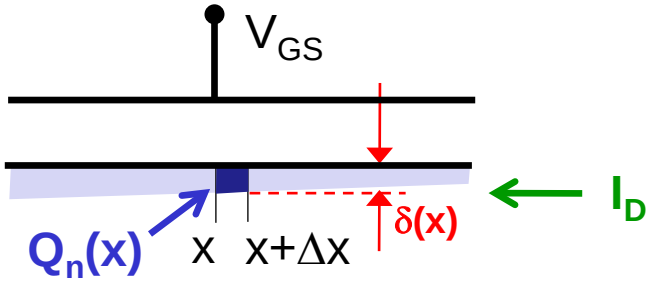
Onset of Saturation in MOSFET



MOSFET Operation in Saturation Region



MOSFET Analysis in **Triode** Region



- We assume $V_{GS} > V_T$ and that $V_{DS} < V_{DSAT}$, i.e. the Inversion Layer exists all along the channel length [0, L]
- Consider the elemental section of the IL between x and x+Δx (shown above) The Drain current I_D is a constant, independent of "x", so that the voltage dropped across the section of length Δx is given by: $\Delta V(x) = I_D.\Delta R(x)$ where ΔR(x) is the resistance of the section:

$$\Delta R(x) = \left|\frac{\rho \cdot l}{A}\right| = \frac{\Delta x}{q \cdot n(x) \cdot \mu_n^* \cdot [W \cdot \delta(x)]} = \frac{\Delta V(x)}{I_D}$$

l ; $\sigma = 1/\rho$; A

MOSFET Analysis in Triode Region

- Notice that we have used μ_n^* to denote the mobility here rather than just μ_n. This is because the electrons in the Inversion Layer are travelling close to the interface with the oxide where more trapping and scattering sources exist, so that the *surface mobility* μ_n^* may be significantly lower than the normal *bulk mobility* μ_n.
- We have already seen that $Q_n(x) = q.n(x).\delta(x)$, so we can use this substitution and also re-arrange, taking the limit as $\Delta x \rightarrow 0$ to get:

$$I_D = W \cdot \mu_n^* \cdot Q_n(x) \cdot \frac{dV(x)}{dx}$$

$$= W \cdot \mu_n^* \cdot C_{OX} \cdot [V_{GS} - V_T - V(x)] \cdot \frac{dV(x)}{dx}$$

- Now take (1/L) times the integral from "0" to "L" w.r.t. "x", on both sides of this equation, noting that I_D is constant:

MOSFET Analysis in Triode Region

$$I_D = \frac{1}{L}\int_0^L I_D \cdot dx = \frac{W \cdot \mu_n^* \cdot C_{OX}}{L} \cdot \int_0^L \left[V_{GS} - V_T - V(x)\right] \cdot \frac{dV(x)}{dx} \cdot dx$$

- Now change variable from "x" to "V":

$$I_D = \frac{W \cdot \mu_n^* \cdot C_{OX}}{L} \cdot \int_0^{V_{DS}} \left[V_{GS} - V_T - V\right] \cdot dV$$

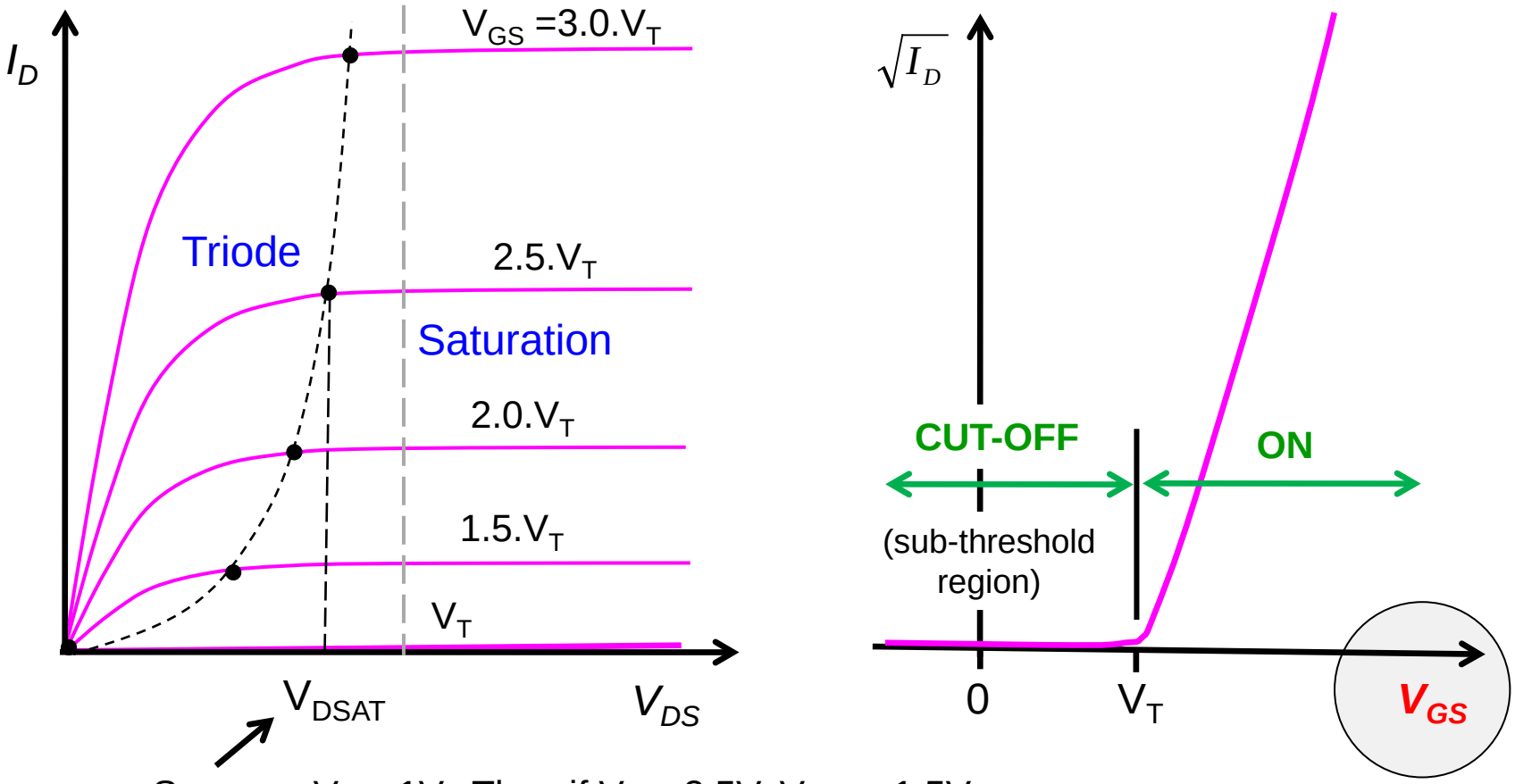
$$\Rightarrow \quad \boxed{I_D = \frac{W \cdot \mu_n^* \cdot C_{OX}}{L} \cdot \left[\left(V_{GS} - V_T\right) \cdot V_{DS} - \frac{V_{DS}^2}{2}\right]}$$

(prove as exercise)

- This is an approximate result (since depletion charge has been ignored) but it is useful and widely-used in simple circuit-level design;
- The result is only valid in the triode region. In saturation, we can just assume the Drain current is constant at the value it had at V_{DSAT} (which $=(V_{GS}-V_T)$):

$$I_D(V_{DSAT}) = \frac{W \cdot \mu_n^* \cdot C_{OX}}{L} \cdot \left[\left(V_{GS} - V_T\right) \cdot V_{DSAT} - \frac{V_{DSAT}^2}{2}\right] = \frac{W \cdot \mu_n^* \cdot C_{OX}}{L} \cdot \left[\frac{V_{DSAT}^2}{2}\right]$$

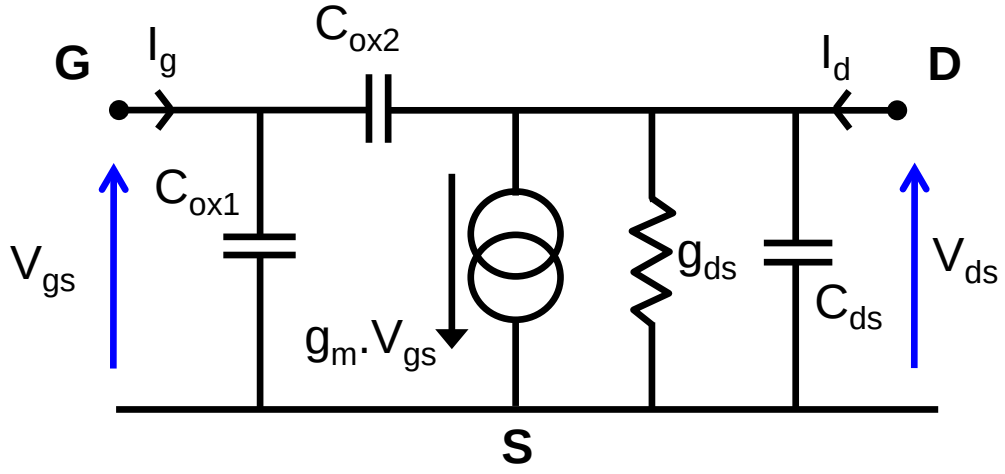
N-Channel MOSFET DC Characteristics



e.g. Suppose V_T = 1V. Then if V_{GS}=2.5V, V_{DSAT}=1.5V

http://jas.eng.buffalo.edu/education/mos/mosfet/mos_2.html

Small-Signal MOSFET Model



- The capacitances are now associated with the Gate oxide, and a drain-source capacitance (C_{ds}) also needs to be included;
- We can calculate g_m and g_{ds} from the DC current expression:
  - Triode Region:

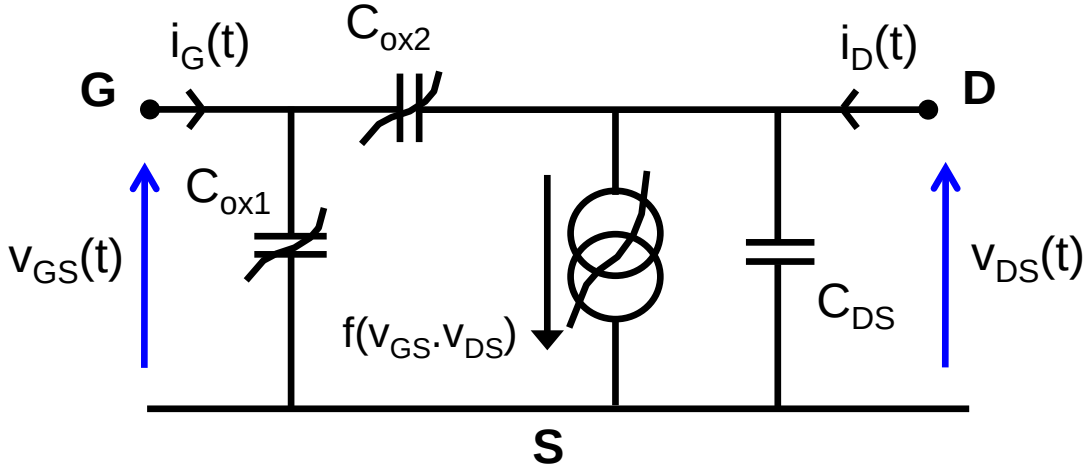
$$g_m = \left.\frac{\partial I_D}{\partial V_{GS}}\right|_{DC} = \frac{W \cdot \mu_n^* \cdot C_{OX}}{L} \cdot V_{DS} \qquad g_{ds} = \left.\frac{\partial I_D}{\partial V_{DS}}\right|_{DC} = \frac{W \cdot \mu_n^* \cdot C_{OX}}{L} \cdot \left[\left(V_{GS} - V_T\right) - V_{DS}\right]$$

  - Saturation:

$$g_m = \left.\frac{\partial I_D}{\partial V_{GS}}\right|_{DC} = \frac{W \cdot \mu_n^* \cdot C_{OX}}{L} \cdot V_{DSAT} \qquad g_{ds} = \left.\frac{\partial I_D}{\partial V_{DS}}\right|_{DC} = 0$$

EXAMPLE 11.2

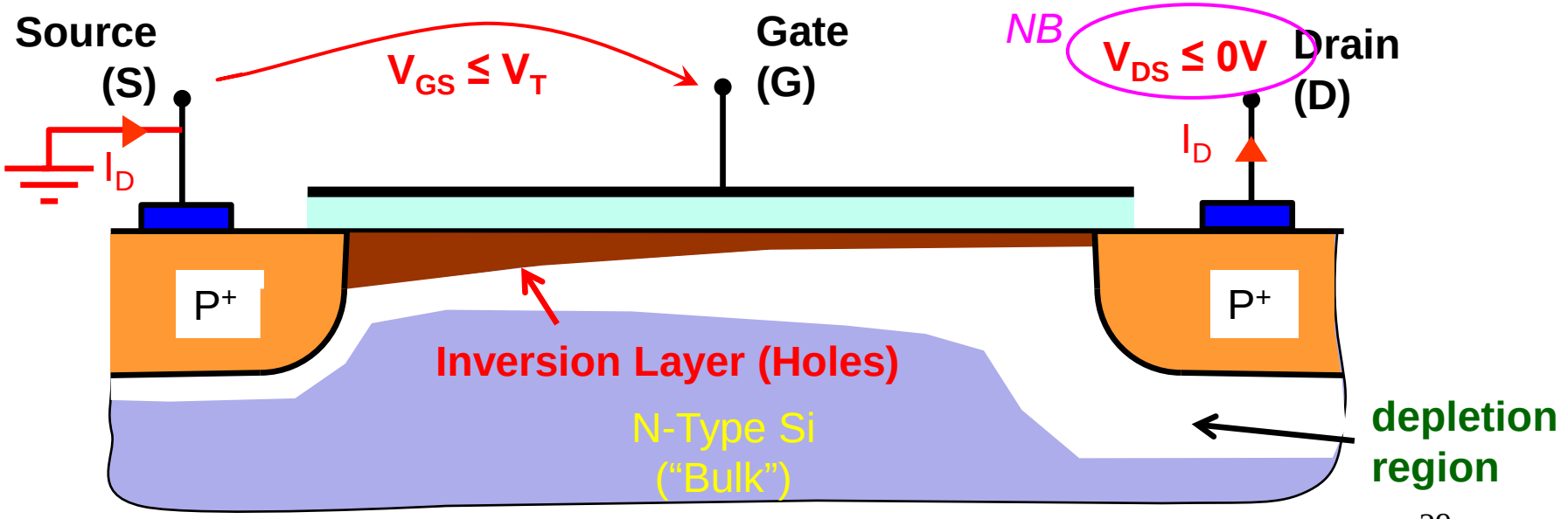
Large-Signal MOSFET Model



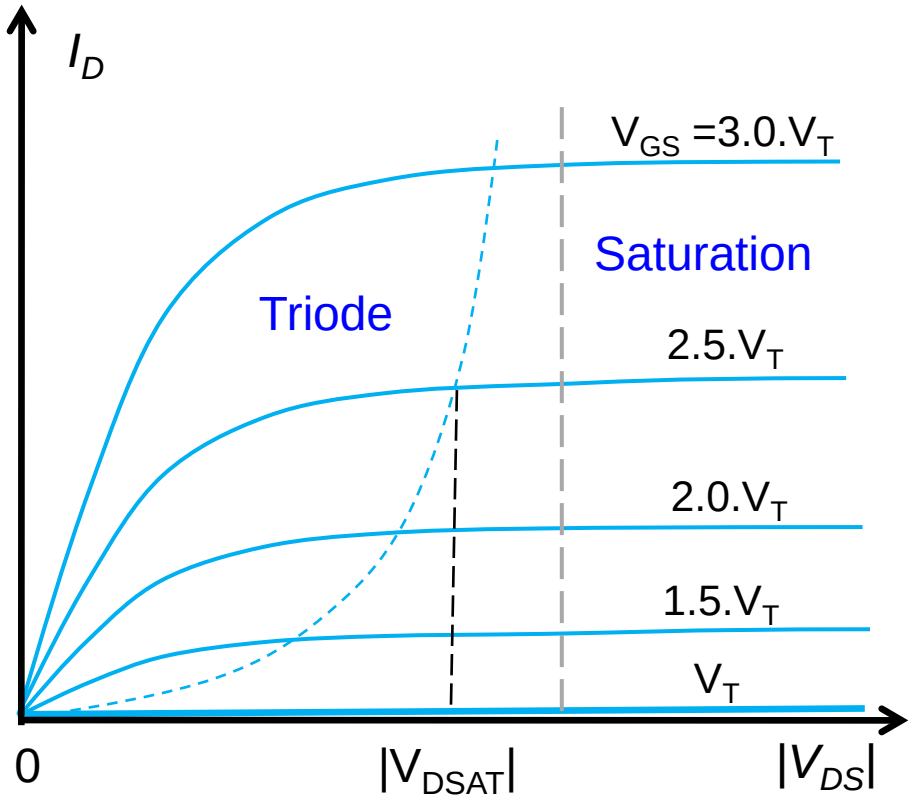
- This is a very simple version of a non-linear MOSFET model, but it emphasises the fact that there is no DC current path at the Gate terminal;
- Some extremely complex and general non-linear MOSFET models exist. For example, BSIM4 (Berkeley SPICE Integrated Model 4) is a widely-used MOSFET model with over 300 different model parameters!

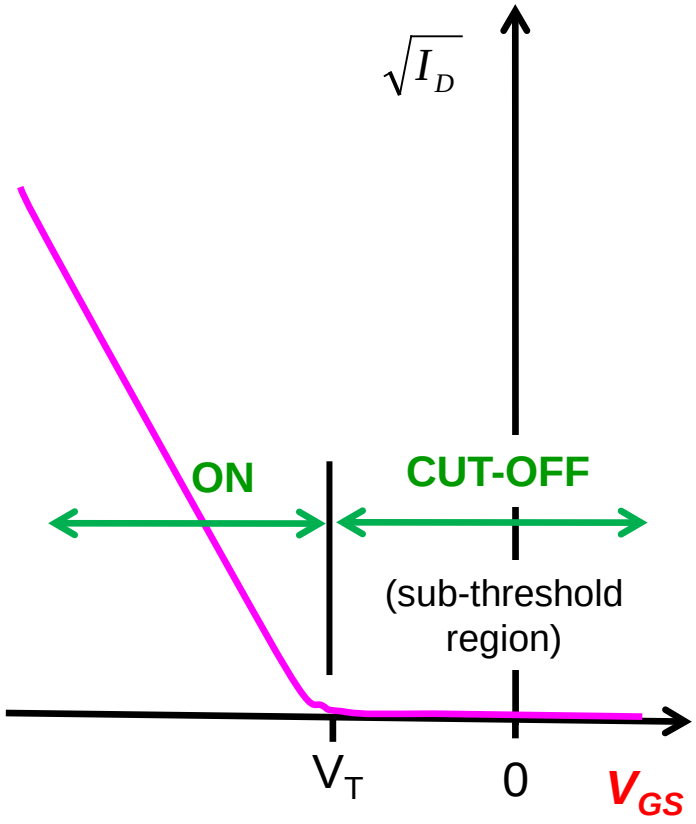
P-Channel Enhancement-Type MOSFET

- By reversing the N and P regions, we can create a P-channel enhancement-type MOSFET, in which the inversion layer consists of holes;
- A P-channel MOSFET requires a negative threshold voltage on the Gate compared to the Source. If a negative voltage is then applied on the Drain, a Drain current will flow from Source to Drain;
- Performance is inferior to an equivalent Si N-MOS device because of the approximately 3 times lower mobility of holes compared to electrons



P-Channel MOSFET DC Characteristics



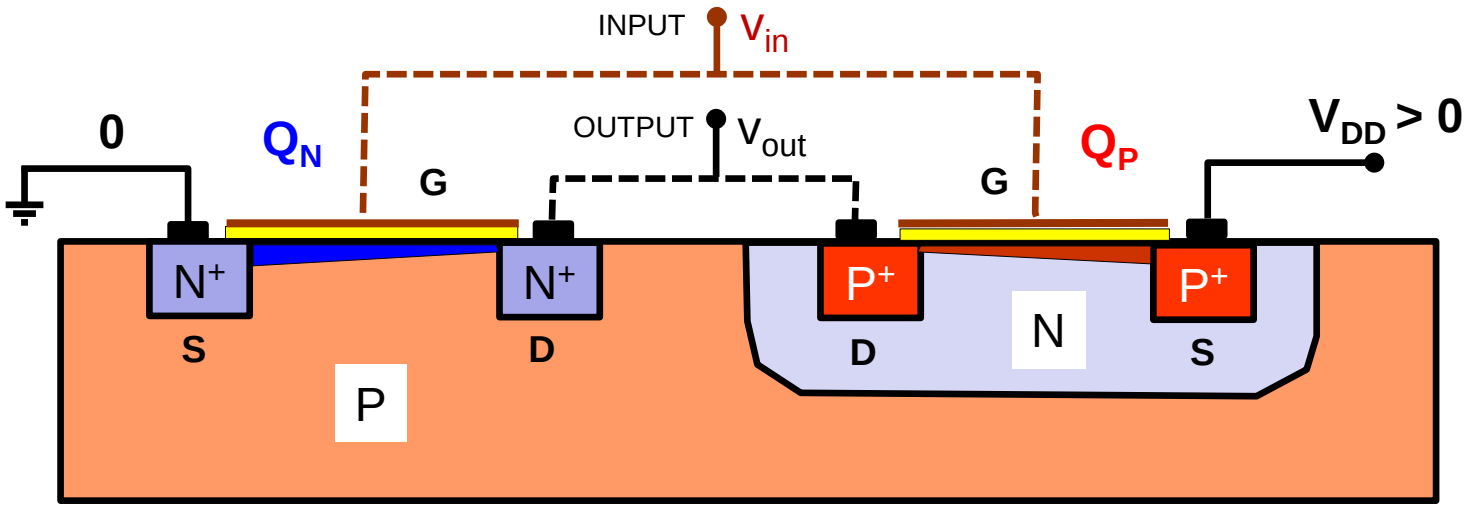


e.g. V_T = -1V. For V_{GS}= -2.5V, V_{DSAT}= -1.5V

Complementary MOS - CMOS

- By far the most successful ever Integrated Circuit (IC) technology has been found through combining N-channel MOSFETs with P-channel MOSFETs on the same Silicon substrate. This is called Complementary MOS or CMOS;
- A feature of CMOS is extremely low power consumption as one transistor in a series combination is usually turned OFF;
- This requires either an N-well technology (in a P-type epitaxial layer, to realise the P-MOS device) or a P-well technology (in an N-type epitaxial layer, to realise the N-MOS device), or even a double-well technology;
- The basic CMOS structure is a combination of a P-MOS and an N-MOS device to create a digital inverter.

(N-Well) CMOS Inverter

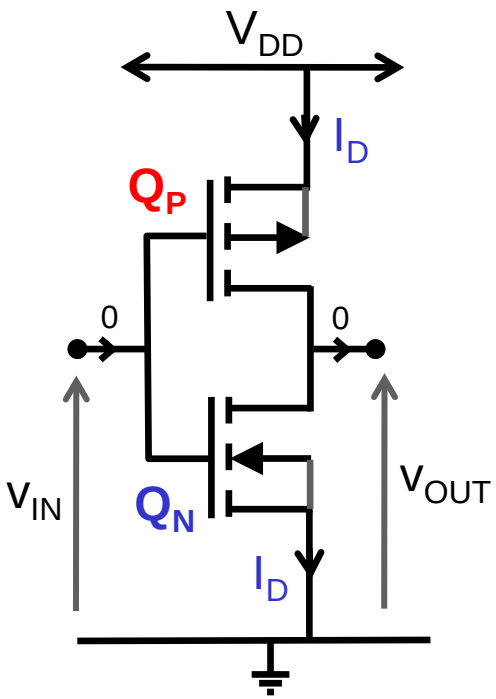


- An inverter is the most simple digital circuit that inverts a binary input, i.e. if the input is a binary "0" (LO) the output is binary "1" (HI) and conversely if the input is "1" the output is "0";
- It can be realised in CMOS by connecting a P-MOS and an N-MOS in series as shown with the 2 Gates tied together to form the input (v_{in}) and the two Drains are connected together to form the output (v_{out})

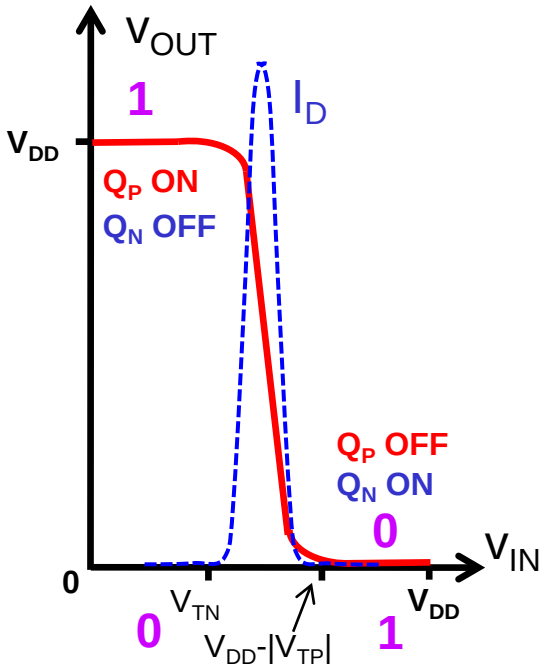
Please study this applet: click the "animate-to-next" button:
http://jas.eng.buffalo.edu/education/fab/invFab/index.html

CMOS Inverter

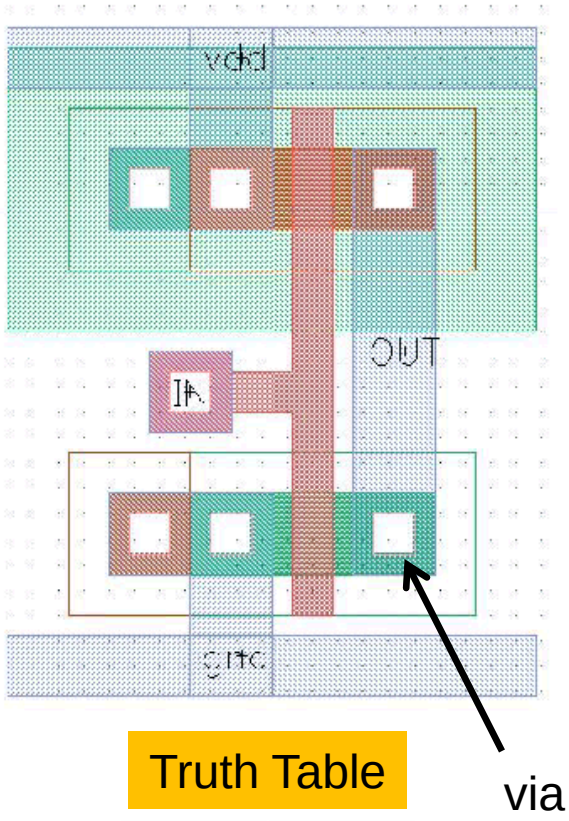
Circuit Schematic



DC Transfer Characteristic



IC Layout

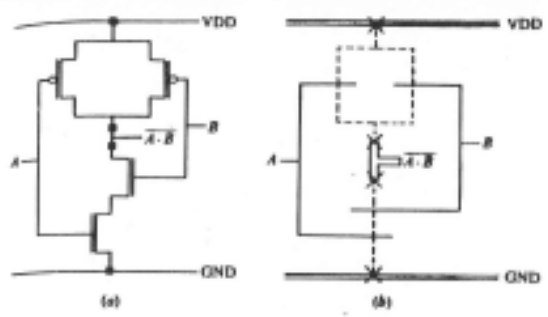


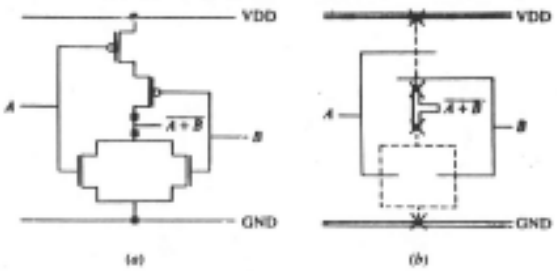
- When v_{IN} is LO (< V_{TN}), Q_N is OFF while Q_P is ON => the output v_{OUT} is HI (≈ V_{DD})
- When v_{IN} is HI (near V_{DD}), Q_N is ON while Q_P is OFF => the output v_{OUT} is LO (≈ 0)

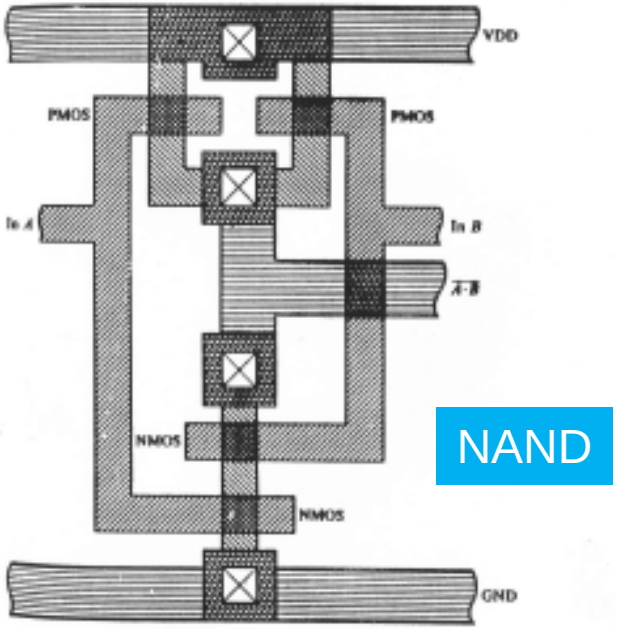
Truth Table

V_{IN}	V_{OUT}
0	1
1	0

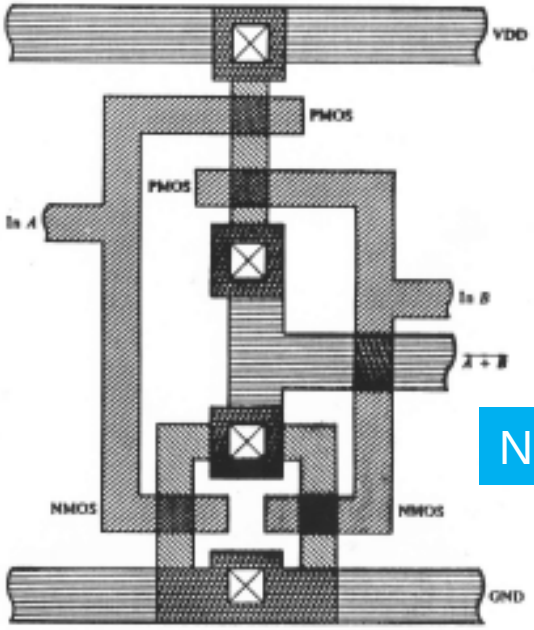
NAND Gate and NOR Gate in CMOS







NAND



NOR