

## 2 MOS Devices and Basic Circuits

The purpose of this chapter is to describe the operation of the different types of MOS devices available and then to develop the characteristic equations which describe their behaviour. This enables the simple circuits which form the basis of MOS digital circuit design to be presented.

### 2.1 The MOS Structure

An insight into the behaviour of an MOS device can be gained by considering its structure, and figure 2.1 shows the structure of an n-channel or NMOS transistor.

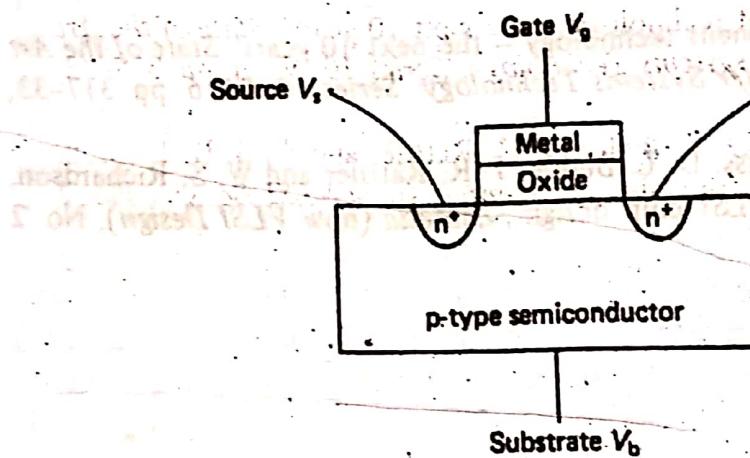


Figure 2.1 NMOS transistor structure

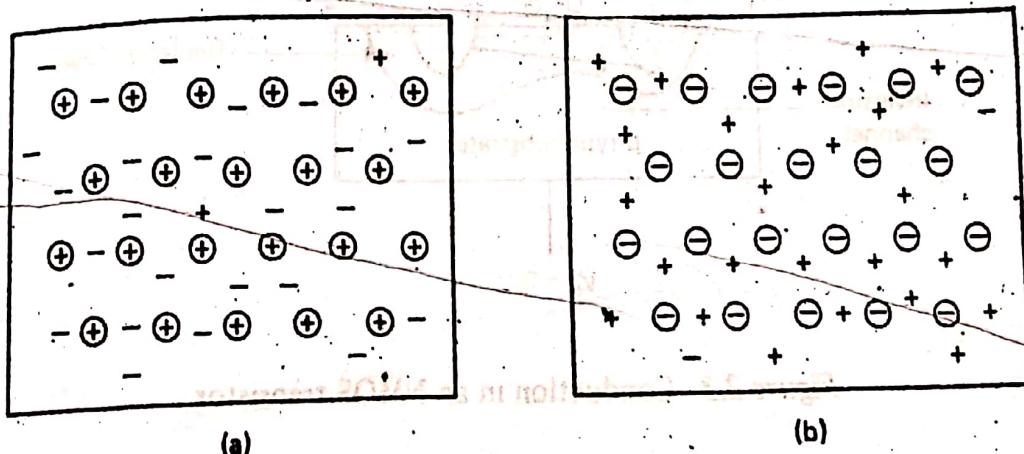
The basis of the transistor is a metal-oxide-semiconductor structure, hence the name MOS. The device input is called the 'gate'. Originally this was a metal plate, although nowadays it is usually made of polycrystalline silicon (commonly called poly or polysilicon). The oxide is very pure silicon dioxide and it separates the gate from the semiconductor material; it acts as an insulator.

The semiconductor is pure silicon which has been doped with relatively small amounts of an impurity such that the impurity atoms can easily replace silicon atoms in the regular, fixed crystal structure. Pure silicon is not a good conductor at room temperature as there are very few electrons of negative charge which

acquire a high enough energy to break away from the silicon atoms. The electrons which escape are free to move about the material and are referred to as 'free charge carriers'. They leave a vacancy or hole in the parent atom which now has a net positive charge. Other electrons of lower energy move to fill these vacancies and this movement of free carriers is equivalent to the movement of holes of positive charge.

An n-type semiconductor is obtained by doping pure silicon with an impurity possessing one more electron (in the outermost orbit) than silicon. This electron is only loosely bound to the impurity atom and can easily acquire enough energy to break away, leaving the fixed impurity atom positively charged. The freed electrons form the majority of free charge in the material, significantly reducing its resistivity below that of pure silicon. A few free holes still exist and these are referred to as the 'minority charge carrier'. This is depicted in figure 2.2a.

A p-type semiconductor arises when the impurity atoms have one fewer electron in the outer orbit than does silicon. There is thus a vacancy for an extra electron and electrons move to fill these vacancies, causing the fixed impurity atoms to become negatively charged. This electron extraction is equivalent to the injection of positive holes. Thus the majority of free charge carriers are holes and the minority charge carriers are electrons (see figure 2.2b).



+ Free hole

- Free electron

⊕ Impurity atom

⊖ Impurity atom

**Figure 2.2** Charge within doped semiconductor: (a) n-type, (b) p-type

The semiconductor material of figure 2.1 consists of a lightly doped p-type substrate and heavily doped n-type regions, denoted  $n^+$  and called the source and the drain, which are located at each end of the gate. Conventionally, the drain is the device output terminal and the source is the terminal that is common

to both the input and output circuits. It is therefore usual to specify the device's input voltage as  $V_{gs}$ , meaning  $V_g - V_s$ , and its output voltage as  $V_{ds}$ , meaning  $V_d - V_s$ .

A terminal is connected to the bulk substrate and in NMOS this is always connected to the most negative voltage available. This is so that the diodes formed by the substrate-source and substrate-drain pn junctions are always reverse-biased and hence never conduct.

## 2.2 Conduction

To make the transistor conduct, appropriate voltages have to be applied to the terminals. Figure 2.3 shows the effect of applying a positive bias to the gate and drain with the source and bulk substrate tied to 0 V.

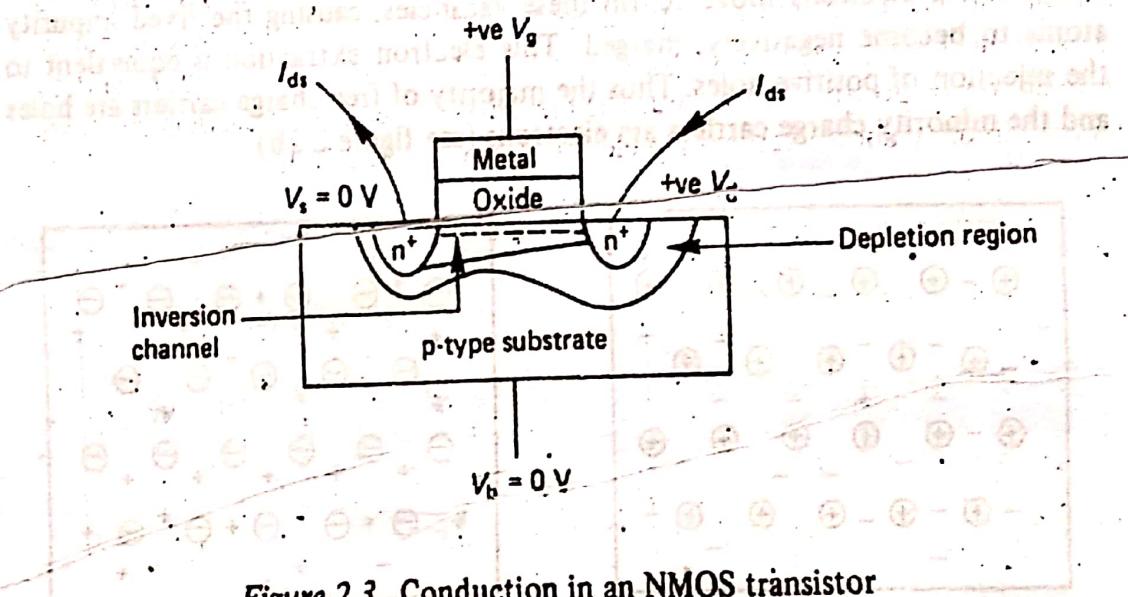


Figure 2.3 Conduction in an NMOS transistor

Although the conductivity of the semiconductor is less than that of the metal, it can nevertheless be considered to be a conducting material. Thus in figure 2.3, the oxide acts as an insulator between two conductors, so the structure resembles that of a capacitor.

The application of a positive gate bias with respect to the source causes a positive charge to accumulate on the metal and an equal negative charge, supplied by the source and the drain, to be induced in the semiconductor surface just beneath the oxide. This charge is in addition to the existing charge. If the induced charge is small then it only causes the surface layer to become less p-type than the bulk. If, however, the charge induced is large enough then the surface layer inverts from p-type to n-type, as shown in figure 2.3. There is now a continuous electron channel from the drain to the source, and current flows if there is a bias between them. In figure 2.3, current flows from the drain to the source as the drain potential is higher than that of the source.

A region depleted of free charge carriers separates all p-type regions from n-type and prevents conduction of reverse-biased pn junctions. Hence the device conduction path is isolated from the bulk substrate by a depletion region and this effect is also used to provide isolation between devices. Note that since the depletion width is dependent upon the junction reverse voltage, the depletion region is wider around the drain than around the source.

The device conduction path is also isolated from the gate by the oxide. Thus it follows that all the current flowing into the drain flows out of the source; this current is referred to as  $I_{ds}$ .

### 2.3 Threshold Voltage

The input potential  $V_{gs}$  at which the surface just becomes inverted is called the threshold voltage  $V_t$ . Below the threshold voltage an NMOS transistor is off and no current flows, while above the threshold the inversion channel is established and the device conducts.

A detailed analysis of the magnitude of the threshold voltage is complex and beyond the scope of this book. However, the factors determining  $V_t$  can be appreciated from a consideration of the excess charge within an NMOS device when the inversion layer is established. This is depicted in figure 2.4.

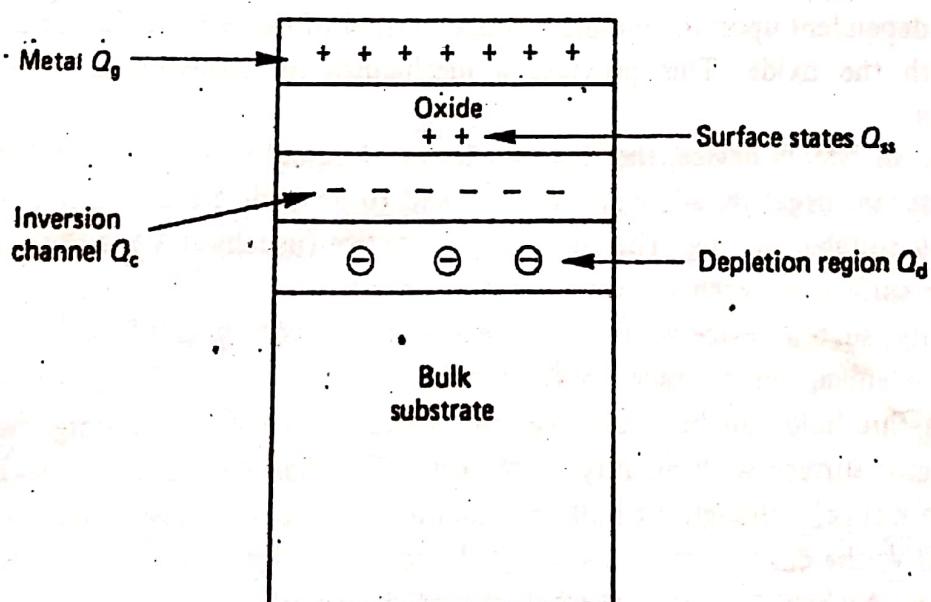


Figure 2.4 Excess charge within conducting NMOS transistor

The surface state charge  $Q_{ss}$  represents positive charge trapped at the oxide-semiconductor interface as a result of imperfections at this surface. This, plus the gate charge  $Q_g$ , must equal the induced charge in the inversion channel  $Q_c$ , plus the impurity atom charge  $Q_d$  in the depletion region. Thus

$$Q_{ss} + Q_g = Q_c + Q_d$$

When the surface is just at the point of inversion,  $Q_c = 0$  and

$$V_t = \frac{Q_g}{C_g} = \frac{Q_d - Q_{ss}}{C_g}$$

where  $C_g$  is the capacitance across the insulator.

In practice,  $V_t$  has to overcome some in-built potential differences before the transistor is brought to the edge of conduction. As a result, another two voltage terms have to be included in the above expression for  $V_t$ .  $V_{dif}$  represents the voltage arising as a result of the difference between the gate and the semiconductor material. Since silicon gates are used nowadays,  $V_{dif}$  is small. The other voltage term,  $V_r$ , is the voltage across the depletion region just at the point of inversion; it is dependent upon the impurity concentration and is usually less than 1 volt. Thus

$$V_t = \frac{Q_d}{C_g} + V_r - \frac{Q_{ss}}{C_g} + V_{dif} \quad (2.1)$$

The last three terms can be regarded as constant potentials, while  $Q_d$  is dependent upon transistor parameters and the applied voltages. In particular,  $Q_d$  is dependent upon the impurity concentration of the semiconductor material beneath the oxide. This provides a mechanism for adjusting the threshold voltage.

For an NMOS device, the first two terms of equation (2.1) are positive and the last two negative, allowing the threshold to be made either positive or negative by suitable doping. The threshold is positive (usually 1 V) if the semiconductor surface between the source and the drain is heavily doped with a p-type impurity; such a device is off when the gate-source voltage is 0 V and is referred to as an 'enhancement mode NMOS transistor'.

The threshold can be made negative (typically -4 V) by doping the semiconductor surface with an n-type impurity. This has the effect of making the surface n-type, although the bulk semiconductor remains p-type. Thus even with  $V_{gs} = 0$  V, the channel region is inverted and the device is on; an NMOS device which is on when  $V_{gs}$  is 0 V is called an 'NMOS depletion mode transistor'. Here, it is necessary to apply a negative gate-source voltage in order to repel electrons from the surface and turn the device off.

As well as NMOS devices, there are p-channel or PMOS transistors where the substrate is n-type and the drain and the source are heavily doped p-type regions. The substrate is connected to the most positive voltage available so that the drain-substrate and source-substrate pn junctions are always reverse-biased. A negative gate-source voltage causes holes to be attracted to and electrons to be

repelled from the semiconductor surface just beneath the oxide. This surface inverts to p-type if a sufficiently negative  $V_{gs}$  is applied.

Consideration of the excess charge in a PMOS device results in an expression similar to equation (2.1), except that all four terms are negative. Again, the threshold voltage can be adjusted by altering the impurity doping level in the surface beneath the oxide. However, logic circuits require only PMOS enhancement devices. These are fabricated with a negative threshold (normally -1 V) and are off when  $V_{gs} = 0$  V.

Thresholds are quoted for transistors assuming a source-substrate voltage of 0 V. Changing the substrate voltage causes the threshold to change, and this effect is known as the 'body effect'. In NMOS devices, taking the substrate voltage negative of the source causes the depletion region surrounding the conduction path to widen and thus increases  $Q_d$ . As a result,  $V_t$  has to be increased to bring the transistor to the edge of conduction. If  $V_{to}$  is defined as the threshold when the source-substrate voltage  $V_{sb}$  is 0 V, then the modified threshold  $V_t$  to take into account the body effect is

$$V_t = V_{to} + \gamma(V_{sb})^{1/2}$$

$\gamma$  is a constant dependent upon transistor parameters and tends to lie between 0.3 and 0.7 for MOS transistors. It is usual to use a value of 0.5 in calculations.

Similar reasoning for a PMOS transistor shows that increasing the substrate voltage above the source potential causes the threshold to become more negative.

#### 2.4. $I_{ds}$ versus $V_{ds}$ Characteristic for NMOS Devices

Figure 2.5 shows the symbols used in this book to denote MOS transistors. It is assumed that the substrate of NMOS devices is tied to the most negative voltage available and a PMOS substrate to the most positive available voltage. Thus this connection is usually omitted from circuit diagrams.

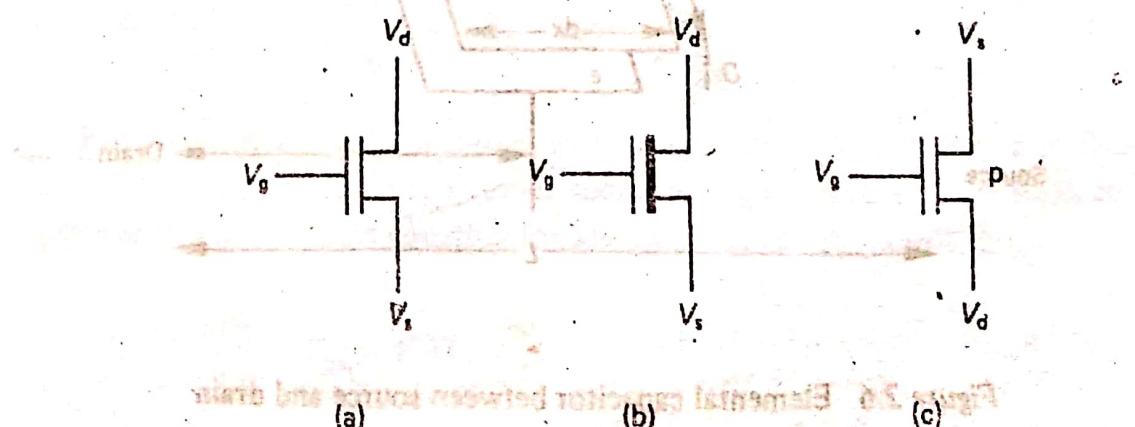


Figure 2.5 Symbols for MOS transistors: (a) enhancement mode NMOS,  
(b) depletion mode NMOS, (c) enhancement mode PMOS.

Consider an NMOS device. When  $V_{gs} < V_t$  the transistor is off, regardless of the drain voltage; the device does not conduct and no current flows. The device conducts when  $V_{gs} > V_t$ , and if a constant  $V_{gs}$  is applied then the resulting  $I_{ds}$  versus  $V_{ds}$  curve can be split into two regions.

### (a) Resistive Region

Here,  $V_{ds} < V_{gs} - V_t$ . The voltage across the insulator at the source is  $V_{gs}$  and at the drain is  $V_{gd}$  (meaning  $V_g - V_d$ ). Although the voltage across the insulator is not constant, a voltage in excess of  $V_t$  exists at all points across the oxide, causing the formation of a continuous inversion channel between the drain and the source. It will be assumed that the increase in voltage along the channel from the drain to the source is linear with distance.

The device structure therefore resembles an infinite number of capacitances between the drain and source, each one having a different voltage across it and therefore a different charge from its neighbours. The total charge induced in the channel is the sum of the charge induced on each of these capacitances.

Consider one of these capacitances of length  $dx$ -situated at a distance  $x$  metres from the drain, as shown in figure 2.6.

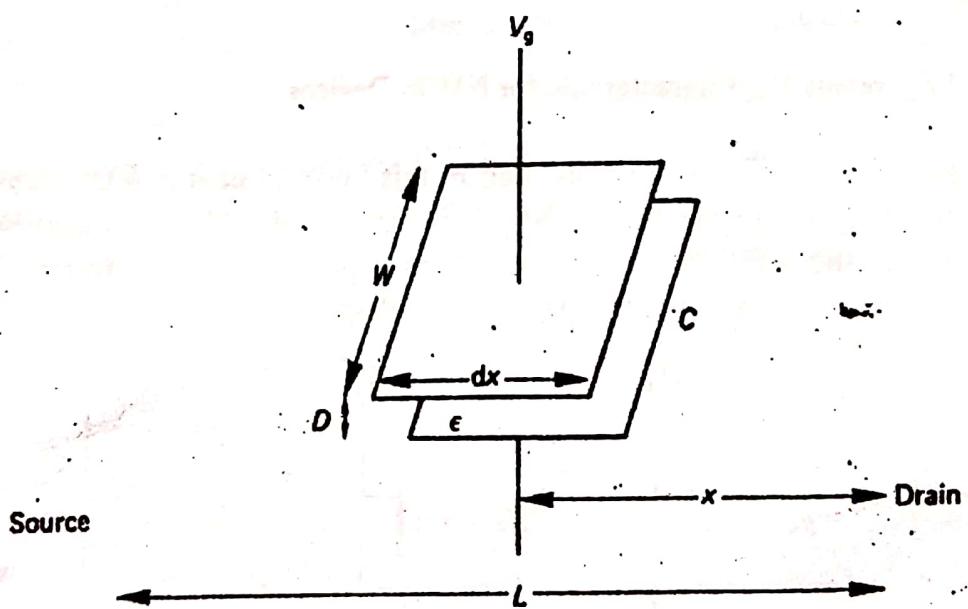


Figure 2.6 Elemental capacitor between source and drain

The channel width and length are  $W$  and  $L$  metres respectively. Thus the capacitance,  $C$  farads, of the structure shown in figure 2.6 is

$$C = \frac{W\epsilon_{ox}x}{D} \quad \left( \frac{A}{D} \right)$$

where  $\epsilon$  is the permittivity of the insulator in farads/metre and  $D$  is the thickness of the oxide in metres. The voltage  $v$  in excess of  $V_t$  across this capacitor is

$$v = (V_{gd} + \frac{x}{L} V_{ds} - V_t) = (V_{gs} - V_{ds} + \frac{x}{L} V_{ds} - V_t)$$

Thus the charge,  $q$  coulombs, induced on this capacitor is

$$q = C_v = \frac{W\epsilon_{ox}x}{D} (V_{gs} - V_{ds} + \frac{x}{L} V_{ds} - V_t)$$

The total charge  $Q$  induced in the channel is

$$Q = \int_0^L \frac{\epsilon W}{D} (V_{gs} - V_{ds} + \frac{x}{L} V_{ds} - V_t) dx$$

$$= \frac{\epsilon WL}{D} \left[ (V_{gs} - V_t) - \frac{V_{ds}}{2} \right]$$

Now  $Q = tI_{ds}$ , where  $t$  is the time in seconds for an electron to move across the channel and

$$t = \frac{\text{channel length } L}{\text{electron velocity}}$$

$\mu_n$  is the electron velocity per unit electric field (measured in metre<sup>2</sup>/volt-second) and is called the 'electron mobility'. Thus the electron velocity in the channel is  $\mu_n V_{ds}/L$ . Hence the current  $I_{ds}$  in amps is

$$I_{ds} = \frac{Q}{t} = \frac{\epsilon W \mu_n}{LD} \left[ (V_{gs} - V_t) V_{ds} - \frac{V_{ds}^2}{2} \right] \quad (2.2)$$

For a constant  $V_{gs}$ ,  $I_{ds}$  increases with an increase in  $V_{ds}$ , as shown in figure 2.7. It should be noted that at drain-source voltages which are very small compared with  $V_{gs} - V_t$ , the equation for the channel current reduces to

$$I_{ds} = \frac{\epsilon W \mu_n}{LD} (V_{gs} - V_t) V_{ds}$$

giving a linear relationship between  $I_{ds}$  and  $V_{ds}$  for a constant  $V_{gs}$ .

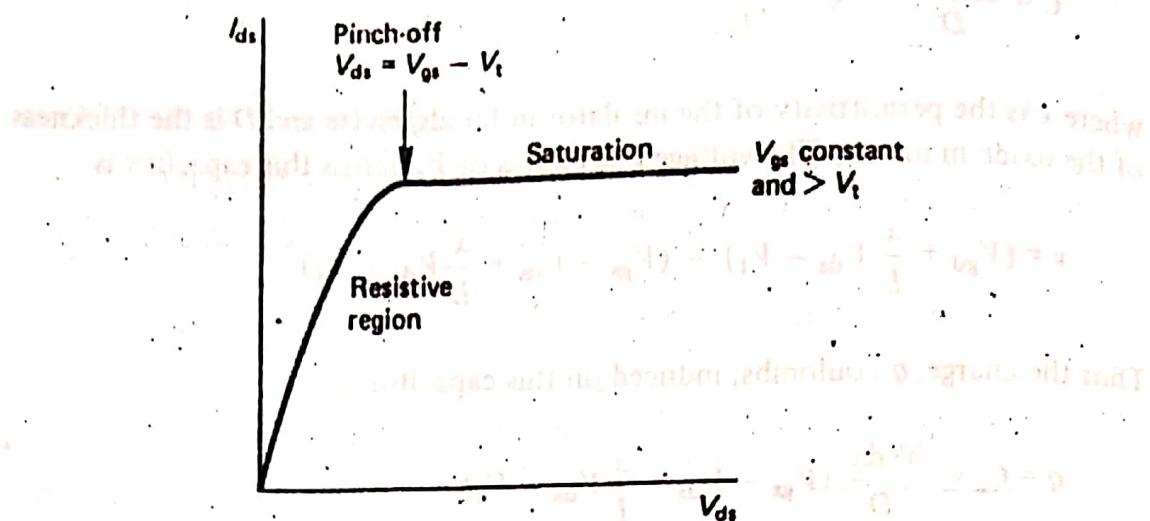


Figure 2.7  $I_{ds}$  versus  $V_{ds}$  for an NMOS transistor

### (b) Saturation

In this region  $V_{ds} \geq V_{gs} - V_t$ . As the drain voltage rises, the voltage across the insulator at the drain drops, and at  $V_{ds} = V_{gs} - V_t$  it is  $V_t$ . This is the voltage necessary to just support inversion, and this point on the  $I_{ds}$  versus  $V_{ds}$  characteristic is called 'pinch-off'. At this point, the inversion channel ends just at the drain. As  $V_{ds}$  increases beyond pinch-off, the point at which inversion ceases moves away from the drain as shown in figure 2.8.

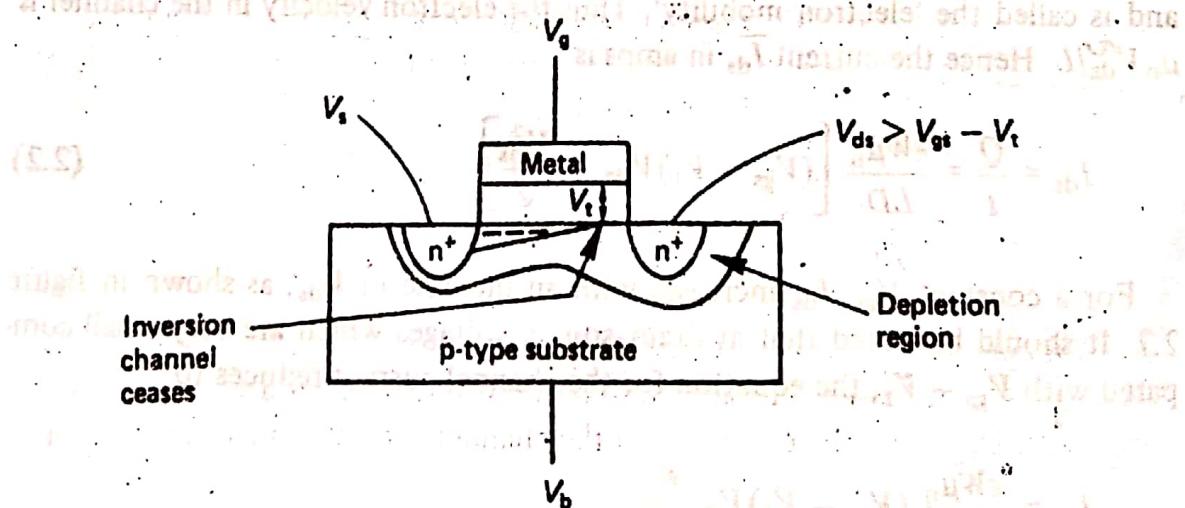


Figure 2.8 Inversion channel beyond pinch-off

The voltage difference along the inversion channel from the source to where it ceases is  $V_{gs} - V_t$ , and the excess potential  $V_{ds} - V_{gs} + V_t$  is dropped between the end of the inversion channel and the drain. This creates a high electric field across this very short distance and the electrons from the inversion channel are quickly swept across this area to the drain.

At and above pinch-off the voltage between the source and the end of the inversion channel is constant at  $V_{gs} - V_t$ . The channel length can also be considered to be constant (equal to  $L$ ). Thus the current flow is constant and the device is said to be 'saturated'. Replacing  $V_{ds}$  by  $V_{gs} - V_t$  in equation (2.2),  $I_{ds}$  becomes

$$I_{ds} = \frac{eW\mu_n}{2LD} (V_{gs} - V_t)^2 \quad (2.3)$$

The  $I_{ds}$  versus  $V_{ds}$  characteristic for a constant  $V_{gs}$  is shown in figure 2.7. In practice, there is a slight increase in current with increasing  $V_{ds}$  above pinch-off, as a result of the reduction in the channel length.

## 2.5 Characteristic Equation for PMOS Devices

A PMOS enhancement device is off when  $V_{gs}$  is 0 V. Here it is necessary to take the gate voltage negative of the source in order to exceed the negative threshold and turn the transistor on. It is thus normal to connect the source terminal to a positive potential and operate the gate and the drain at voltages equal to or negative of this potential. Consequently current,  $I_{sd}$ , flows from the source to the drain in a PMOS device.

A similar analysis to that performed for NMOS devices allows the equations for  $I_{sd}$  versus  $V_{sd}$  to be obtained for the resistive and saturated regions. If the threshold voltage  $V_t$  for PMOS devices (only) is redefined to be the positive source-gate voltage at which the transistor just turns on, then equations similar to those for NMOS are obtained. In the resistive region where  $V_{sd} < V_{sg} - V_t$

$$I_{sd} = \frac{e\mu_p W}{DL} \left[ (V_{sg} - V_t)V_{sd} - \frac{V_{sd}^2}{2} \right] \quad (2.4)$$

Pinch-off occurs when  $V_{sd} = V_{sg} - V_t$  and for  $V_{sd} > V_{sg} - V_t$ , the device is saturated. At and above pinch-off

$$I_{sd} = \frac{e\mu_p W}{2DL} (V_{sg} - V_t)^2 \quad (2.5)$$

$\mu_p$  is the hole velocity per unit electric field and is known as the 'hole mobility'. It is two to three times less than  $\mu_n$ . Since the current flow is proportional to the carrier mobility, an NMOS transistor will conduct more current than a PMOS device of similar size. This can be seen in figure 2.9 which shows characteristic curves for PMOS and NMOS transistors if typical values are assumed for the parameters.

The speed of a circuit is dependent upon the rate at which circuit capacitances can be charged and discharged. This in turn is dependent upon the current available from devices within the design. The current capability of PMOS devices can be made equal to that for NMOS devices by increasing their size to compensate for the difference between hole and electron mobility. However, these larger devices require a greater silicon area and have an increased circuit capacitance. For this reason, NMOS devices are used in preference to PMOS in circuit design.

## 2.6 Principles of Inverters

Now that the characteristics of MOS devices have been established, their use in some basic circuits can be discussed. The inverter will be considered first as, although it is the simplest logic function that can be implemented, it often forms the basis of more complex circuits.

Figure 2.10 illustrates the principles involved in designing an inverter. It consists of a digital switch S which is closed if a high input voltage is applied and is open for a low input voltage. The switch output is connected via a load to the power rail. Thus a high input causes the output to be 0 V (low output) and current flows through the load and the switch. A low input leaves the switch open and no current flows in the circuit. As a result, no voltage is dropped across the load and the output is  $V_p$  (high output).

In MOS inverter design, an n-channel enhancement transistor is used as the digital switch. This is convenient as it has a positive threshold and a low input voltage below the device threshold causes the transistor to be off, while a high input voltage above the threshold causes it to be on.

The load can be implemented in a number of ways and in particular by the use of an MOS device. This results in a range of inverter circuits, each of which illustrates different design principles.

Many designers learn about bipolar junction transistors before being introduced to MOS devices and in that technology the most common implementation of an inverter uses a resistor as a load. Thus a convenient starting point for a description of MOS inverters is to consider a circuit which uses a resistor as the load.

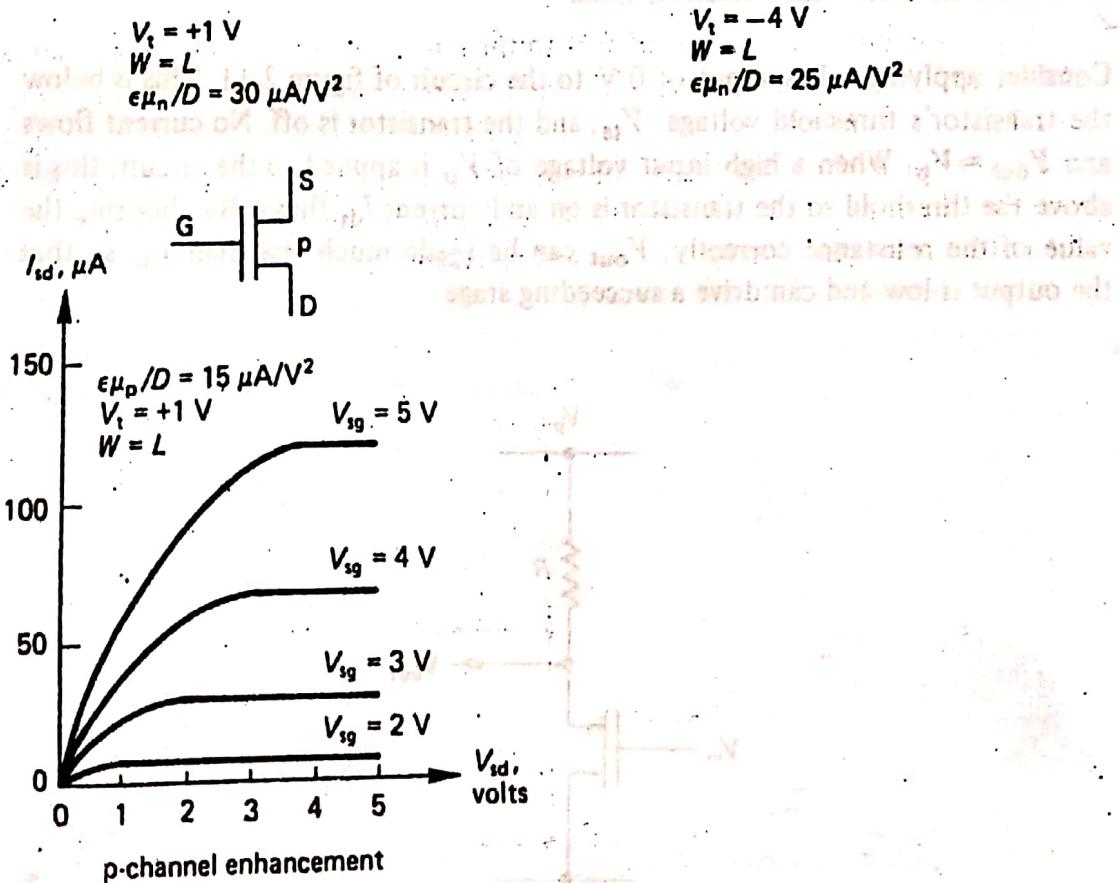
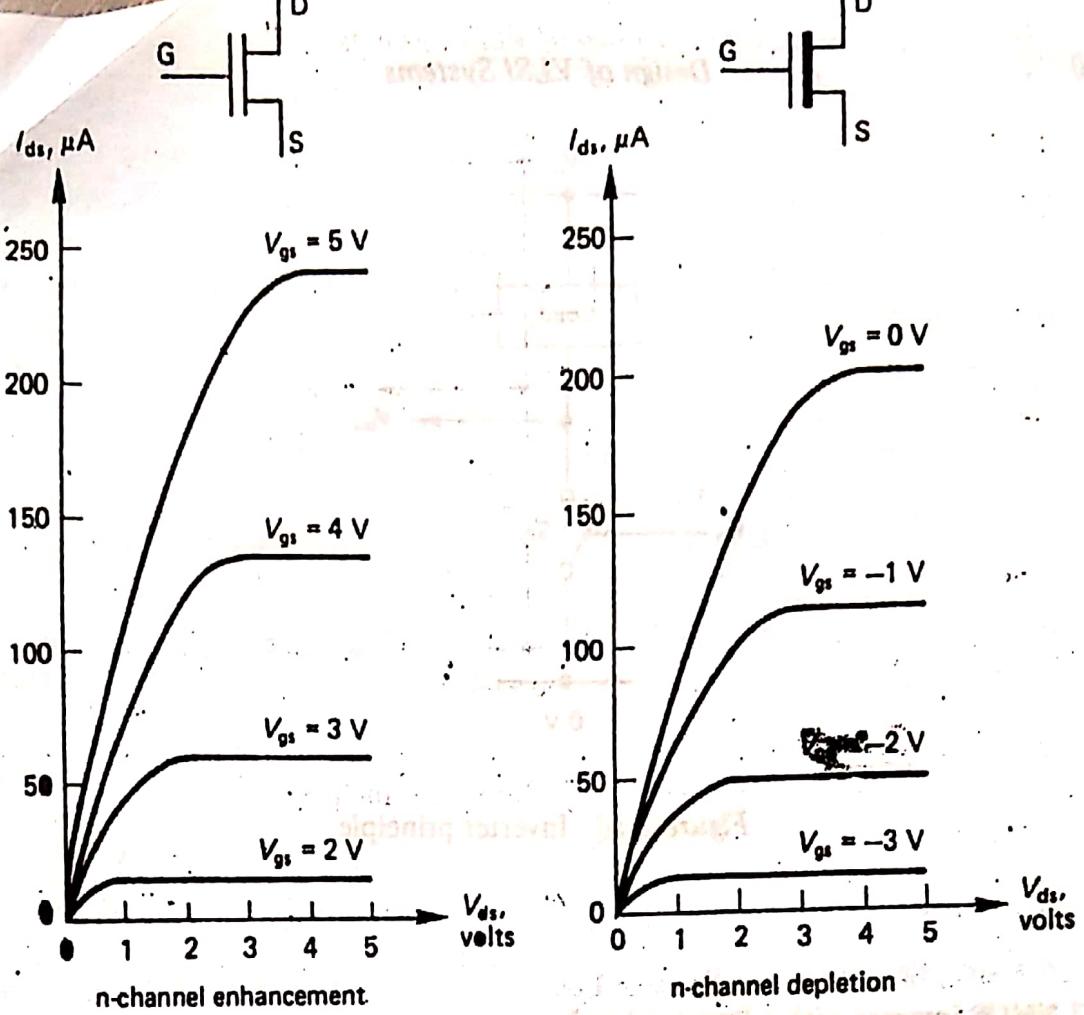


Figure 2.9 MOS transistors – typical characteristics

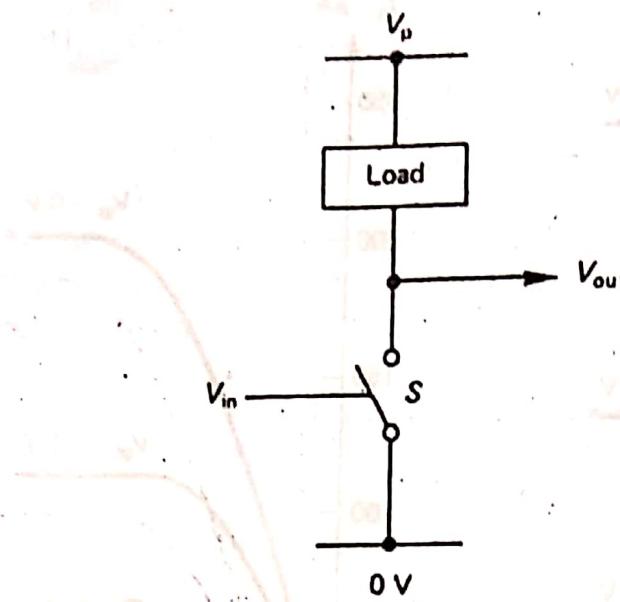


Figure 2.10 Inverter principle

### ✓ 2.7 NMOS Inverter with a Resistor Load

Consider applying a low input of 0 V to the circuit of figure 2.11. This is below the transistor's threshold voltage,  $V_{te}$ , and the transistor is off. No current flows and  $V_{out} = V_p$ . When a high input voltage of  $V_p$  is applied to the circuit, this is above the threshold so the transistor is on and current  $I_{ds}$  flows. By choosing the value of the resistance correctly,  $V_{out}$  can be made much less than  $V_{te}$  so that the output is low and can drive a succeeding stage.

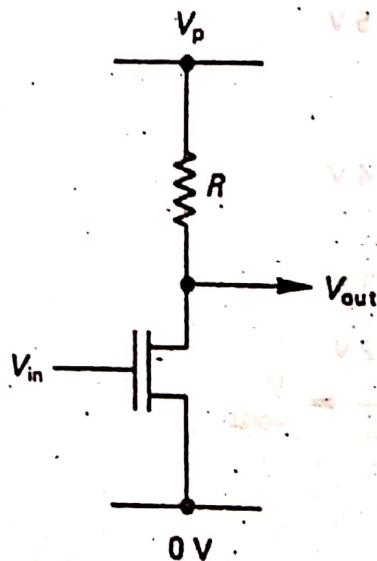


Figure 2.11 NMOS inverter with a resistor load

In order to see what value of  $R$  is required, it is necessary to take typical values for the parameters.  $V_p$  will be taken as 5 V since most MOS and bipolar logic families operate at this voltage.  $\epsilon\mu_n/D$  will be taken as  $30 \mu\text{A}/\text{V}^2$  and  $V_{te}$  as 1 V. An equal gate width and length are assumed, so the width-to-length ratio of the transistor, known as the 'aspect ratio', is 1/1.

$R$  is calculated to give a low  $V_{out}$  of  $0.3 V_{te}$ , since this gives a reasonable noise margin between  $V_{out}$  and the gate threshold. Thus  $V_{ds}$  of the transistor is 0.3 V and, since  $V_{gs}$  is 5 V,  $V_{ds} < V_{gs} - V_{te}$  and the device is operating in the resistive region. Equation (2.2) can now be used to find that

$$I_{ds} = 30 \left( \frac{1}{1} \right) \cdot \left[ (5 - 1) 0.3 - \frac{0.3^2}{2} \right] = 34.7 \mu\text{A}$$

This current also flows through the resistor. Hence

$$R = \frac{V_p - V_{out}}{I_{ds}} = \frac{5 - 0.3 \text{ V}}{34.7 \mu\text{A}} = 135.4 \text{ k}\Omega$$

The silicon area required to implement this resistor is far larger than that for the transistor (typically by a factor of 300). Thus it is not practical to use a resistor as a load and so an MOS device is used instead.

### 2.8 NMOS Inverter with an NMOS Enhancement Transistor Load

This section considers the use of the same transistor type for both the load and the digital switch, see figure 2.12. T1 acts as the switch and is also referred to as the 'driver'. T2 is the load and its gate is connected to  $V_p$  in order to maximise its  $V_{gs}$ . Both transistors have the same threshold  $V_{te}$ .

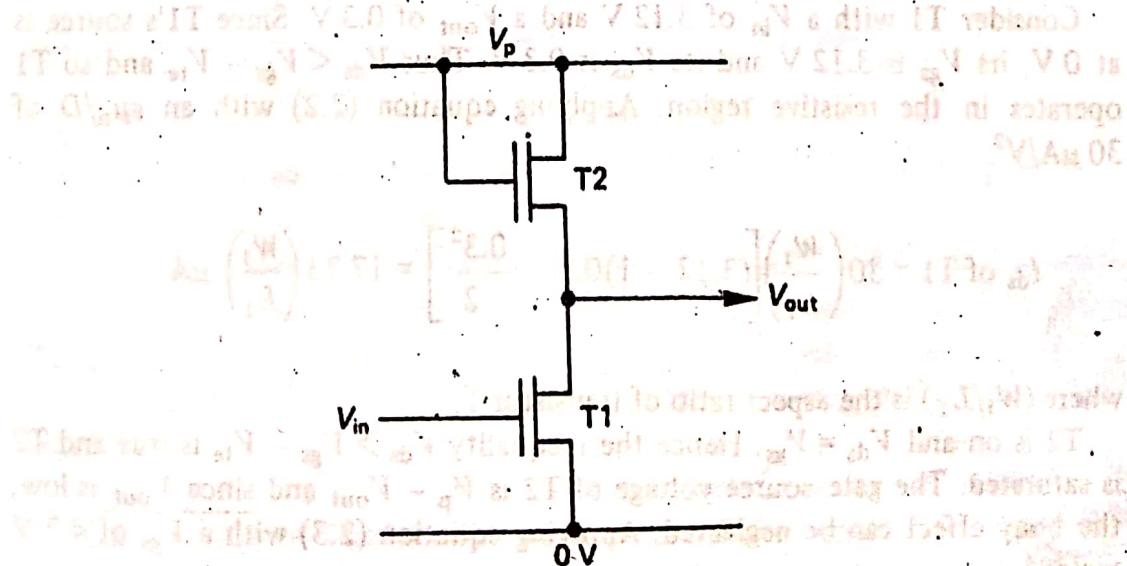


Figure 2.12 NMOS inverter with NMOS enhancement load

When  $V_{in}$  is low, it is less than  $V_{te}$  so transistor T1 is off. However, some very small leakage current flows through T1 and this is supplied by T2. Thus T2 is just brought to the edge of conduction and  $V_{gs} - V_{te}$  for T2 is approximately 0 V, so

$$V_{out} = V_p - V_{te}$$

Again, taking  $V_p$  as 5 V and  $V_{te}$  as 1 V, the high level output voltage is 4 V if the body effect of T2 is neglected. However, T2's source-substrate voltage is significant at this magnitude of output voltage and the effect on the threshold voltage cannot be ignored. Using the expression for modifying the threshold voltage (given in section 2.3) with a substrate voltage of 0 V

$$V_{out} = V_p - V_{te} = V_p - [V_{te0} + \gamma(V_{sb})^{1/2}] = 5 - [1 + 0.5(V_{out})^{1/2}]$$

where the body effect constant  $\gamma$  has been taken as 0.5. Rearranging yields

$$(V_{out})^{1/2} = 8 - 2V_{out}$$

By squaring each side and solving the resulting quadratic equation, the high level  $V_{out}$  is found to be only 3.12 V.

The output from a gate is normally used to drive the input of other gates. Hence the high level input voltage for a gate is 3.12 V. When  $V_{in}$  is high at this level, it exceeds T1's threshold so T1 is on and conducts current. This current is supplied by T2 which is also on. The circuit output is required to be low for a high input and this can be obtained by selecting suitable widths and lengths for the gates of the driver and the load transistors. Again, it is appropriate to choose the gate sizes so that the low level output is 0.3  $V_{te}$ .

Consider T1 with a  $V_{in}$  of 3.12 V and a  $V_{out}$  of 0.3 V. Since T1's source is at 0 V, its  $V_{gs}$  is 3.12 V and its  $V_{ds}$  is 0.3 V. Thus  $V_{ds} < V_{gs} - V_{te}$  and so T1 operates in the resistive region. Applying equation (2.2) with an  $\epsilon\mu_n/D$  of  $30 \mu\text{A}/\text{V}^2$

$$I_{ds} \text{ of T1} = 30 \left( \frac{W_1}{L_1} \right) \left[ (3.12 - 1)0.3 - \frac{0.3^2}{2} \right] = 17.73 \left( \frac{W_1}{L_1} \right) \mu\text{A}$$

where  $(W_1/L_1)$  is the aspect ratio of transistor T1.

T2 is on and  $V_{ds} = V_{gs}$ . Hence the inequality  $V_{ds} > V_{gs} - V_{te}$  is true and T2 is saturated. The gate-source voltage of T2 is  $V_p - V_{out}$  and since  $V_{out}$  is low, the body effect can be neglected. Applying equation (2.3) with a  $V_{gs}$  of 4.7 V yields

$$I_{ds} \text{ of T2} = \frac{30}{2} \left( \frac{W_2}{L_2} \right) (4.7 - 1)^2 = 205.4 \left( \frac{W_2}{L_2} \right) \mu\text{A}$$

where  $(W_2/L_2)$  is the aspect ratio of T2.  $I_{ds}$  of T1 and T2 are equated to obtain an inverter ratio  $k$  for the circuit,

$$k = \frac{(W_1/L_1)}{(W_2/L_2)} = 11.6$$

It is usual to split the inverter ratio obtained between transistors T1 and T2 in order to prevent T1 from occupying a much larger area than T2. Thus a ratio of 11.6 would be taken as 12 for convenience and could be split as  $(W_1/L_1) = 3/1$  and  $(W_2/L_2) = 1/4$ .

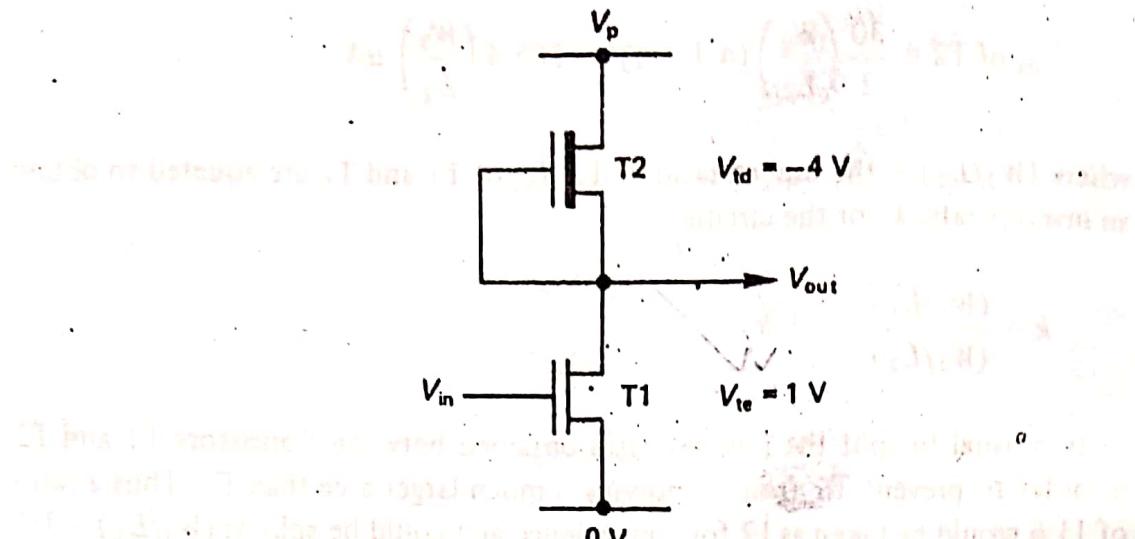
The main drawback of an NMOS enhancement load, apart from the loss of voltage for high outputs, is the speed of the rising edge of the gate. This can be appreciated by considering that there is some capacitance  $C_{out}$  on  $V_{out}$ , and  $V_{out}$  is at 0.3 V corresponding to a  $V_{in}$  of 3.12 V. If  $V_{in}$  now switches to 0.3 V, T1 turns off and  $V_{out}$  starts to rise. Initially,  $V_{gs}$  of T2 is 4.7 V but as  $V_{out}$  rises,  $V_{gs}$  of T2 decreases and T2's threshold rises (because of the body effect) so that less current flows in T2. This progressively slows down the rate of charging up  $C_{out}$ , and as  $V_{out}$  approaches 3.12 V, T2 tends to turn off and very little current flows to finish the charging of  $C_{out}$ . For this reason, n-channel depletion devices are preferred as a load.

## 2.9 NMOS Inverter with an NMOS Depletion Transistor Load

Referring to figure 2.13, the NMOS depletion transistor T2 forms the circuit load. The depletion threshold  $V_{td}$  of T2 is negative and since the gate-source voltage of T2 is zero, T2 is always on.

When  $V_{in}$  is low, it is less than the enhancement threshold  $V_{te}$  of T1, and T1 is off. T2 is on and supplies the very small leakage current of T1. Since  $I_{ds}$  of T2 is negligible, the drain-source voltage of T2 is very small, leading to a  $V_{out}$  of  $V_p$ . Thus there is no high level output voltage loss as with the NMOS enhancement load.

The high level input voltage to a gate is therefore  $V_p$ . This exceeds T1's threshold voltage and hence T1 and T2 are both on. Again it is necessary to choose the gate dimensions of the transistors in order to obtain the desired low level  $V_{out}$ . Taking  $V_{out}$  as  $0.3V_{te}$  and  $V_p$  as 5 V again, T1's  $V_{gs}$  is 5 V and its  $V_{ds}$  is 0.3 V. Thus  $V_{ds} < V_{gs} - V_{te}$  and T1 is operating in the resistive region. Equation (2.2) is applied to express  $I_{ds}$  of T1 in terms of its aspect ratio.



**Figure 2.13** NMOS inverter with NMOS depletion load

$I_{ds}$  of T1 =  $30\left(\frac{W_1}{L_1}\right)\left[(5-1)0.3 - \frac{0.3^2}{2}\right] = 34.65\left(\frac{W_1}{L_1}\right)\mu A$

T2 is on with a  $V_{gs}$  of 0 V and a  $V_{ds}$  of  $V_p - V_{out} = 4.7$  V. Thus  $V_{ds} > V_{gs} - V_{td}$  and T2 is saturated. Equation (2.3) is applied to find T2's  $I_{ds}$ :

$$I_{ds} \text{ of T2} = \frac{25}{2}\left(\frac{W_2}{L_2}\right)[0 - (-4)]^2 = 200\left(\frac{W_2}{L_2}\right)\mu A$$

Note that  $\epsilon\mu_n/D$  for a depletion transistor is typically  $25 \mu A/V^2$  compared with  $30 \mu A/V^2$  for an NMOS enhancement device (owing to the fact that a depletion transistor requires additional impurity doping to define a threshold of  $-4$  V). Equating the current in T1 and T2 yields

If  $k = \frac{(W_1/L_1)}{(W_2/L_2)} = 5.8$ , then equating unit areas of T1 and T2 gives  $(W_1/L_1) = 5.8(W_2/L_2)$ . Since  $k = 5.8$ , then  $(W_1/L_1) = 5.8$ . For convenience, this inverter ratio would be taken as 6 and split as  $(W_1/L_1) = 3/1$  and  $(W_2/L_2) = 1/2$ .

It is possible to reduce the inverter ratio and thus the silicon area occupied by T1 and T2 at the expense of a reduced noise margin. The lower limit for  $k$  is generally accepted to be 4 and this allows 0.5 V between the low level  $V_{out}$  and the enhancement threshold. However, the author recommends that a value of  $k = 6$  or trade is adopted for practical designs.

## 2.10 Edge Times for NMOS Inverter with a Depletion Load

The circuit of figure 2.13 represents a practical inverter that can be implemented in NMOS technology. It is thus worth examining the edge time response of the gate, as this will indicate the maximum speed of operation and the factors upon which the speed is dependent.

In the circuit of figure 2.13 the maximum current capability of T1 and T2 is different. This can be seen by considering the saturation currents of the devices, and it has already been shown that this current for T2 is  $200(W_2/L_2) \mu\text{A}$ . The maximum current through T1 occurs when the gate-source voltage is a maximum (5 V) and the device is saturated. Using equation (2.3)

$$\text{maximum } I_{ds} \text{ of T1} = \frac{30}{2} \left( \frac{W_1}{L_1} \right) (5 - 1)^2 = 240 \left( \frac{W_1}{L_1} \right) \mu\text{A}$$

thus

$$\frac{\text{maximum } I_{ds} \text{ of T1}}{\text{maximum } I_{ds} \text{ of T2}} = \frac{6(W_1/L_1)}{5(W_2/L_2)} = \frac{6k}{5}$$

Let the total capacitance on the output be  $C_{out}$ , as shown in figure 2.14.  $C_{out}$  arises because of the capacitance of the gate of T2, the source of T2 and the drain of T1. In addition,  $C_{out}$  includes the capacitance of the connections between  $V_{out}$  and these transistor terminals. Thus  $C_{out}$  is proportional to the silicon area necessary to implement these features. If the minimum length or width allowable is  $6 \mu\text{m}$ , then a typical value of  $C_{out}$  for this circuit is  $0.1 \text{ pF}$ .

Consider that the input is high and that the output has settled to its low level value of 0.3 V. If the input now changes instantaneously from high to low, then T1 turns off. T2 remains on and supplies current to charge the load capacitance (see figure 2.14). This causes the output to gradually rise to 5 V. The current  $I_c$  in the capacitor is related to the change in voltage  $dV_{out}$  across it with respect to time  $dt$  by

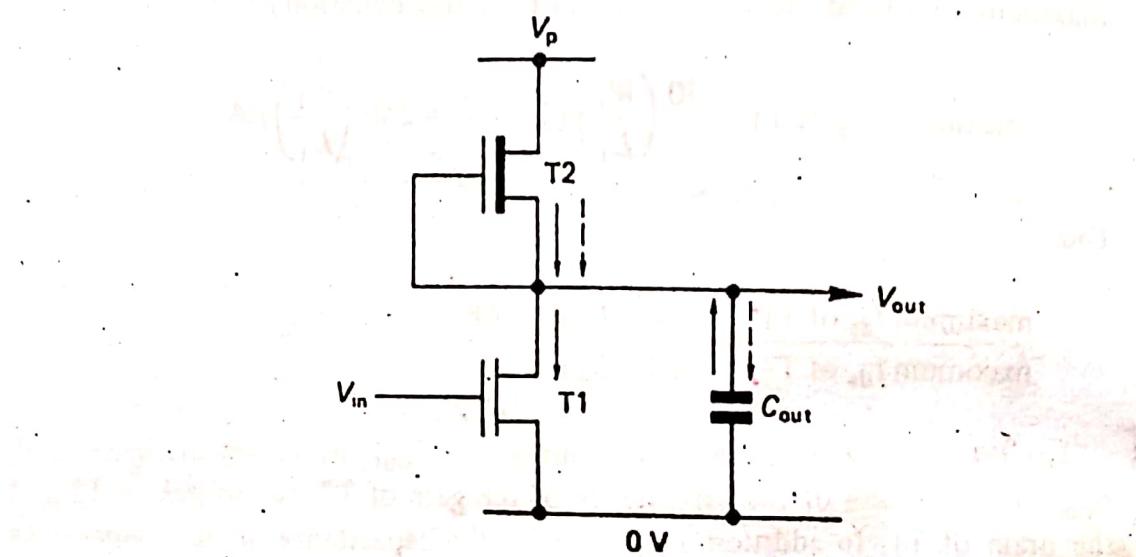
$$I_c = C_{out} \frac{dV_{out}}{dt}$$

The rise time is dependent upon the rate at which the capacitance can be charged during this period. This in turn depends upon the current available from T2.

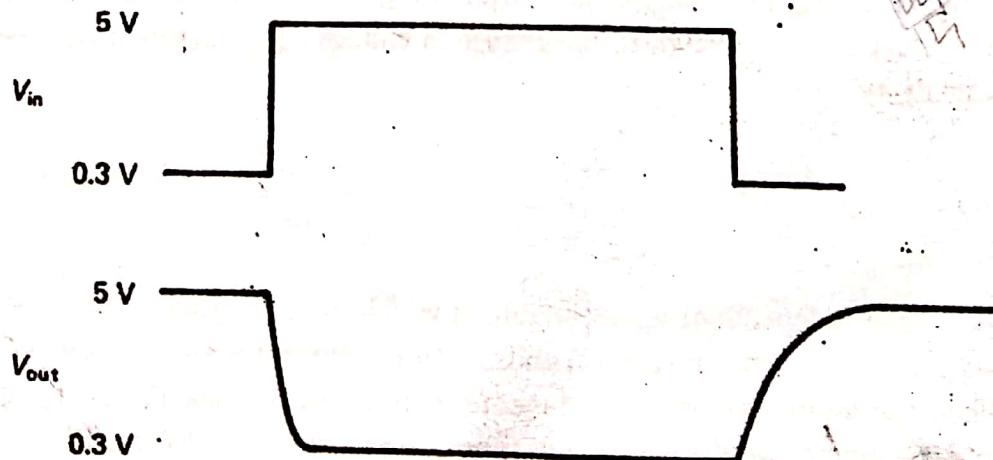
When the input changes instantaneously from low to high, T1 turns on. Both T1 and T2 are on and during the fall time, T1 accepts current from T2 and from  $C_{out}$  as it discharges (see figure 2.14). Since the current in T2 is small compared with that in T1, the current T1 can accept determines the rate at which the capacitor discharges and hence the fall time.

It can therefore be expected that the falling and rising edge times will differ by a factor equal to the ratio of the current capability of T1 and T2. On this basis, the rising edge time will be  $6k/5$  times that of the falling edge time. This is illustrated in figure 2.15.

Consider in detail the rising edge at  $V_{out}$ . The current supplied by T2 causes the output to rise from 0.3 V to 5 V, as shown in figure 2.16a. It is usual to calculate edge times from the 10 per cent to 90 per cent points on the output waveform, as many edges are exponential in shape and take a disproportionately long time to settle to their final value. Applying this to the depletion load inverter, the rise time is the time for the output to rise from 0.8 V to 4.5 V.



**Figure 2.14** Current flow during edge times: ↑ rising edge current flow,  
↓ current flow during falling edge



**Figure 2.15** Output response to an input step

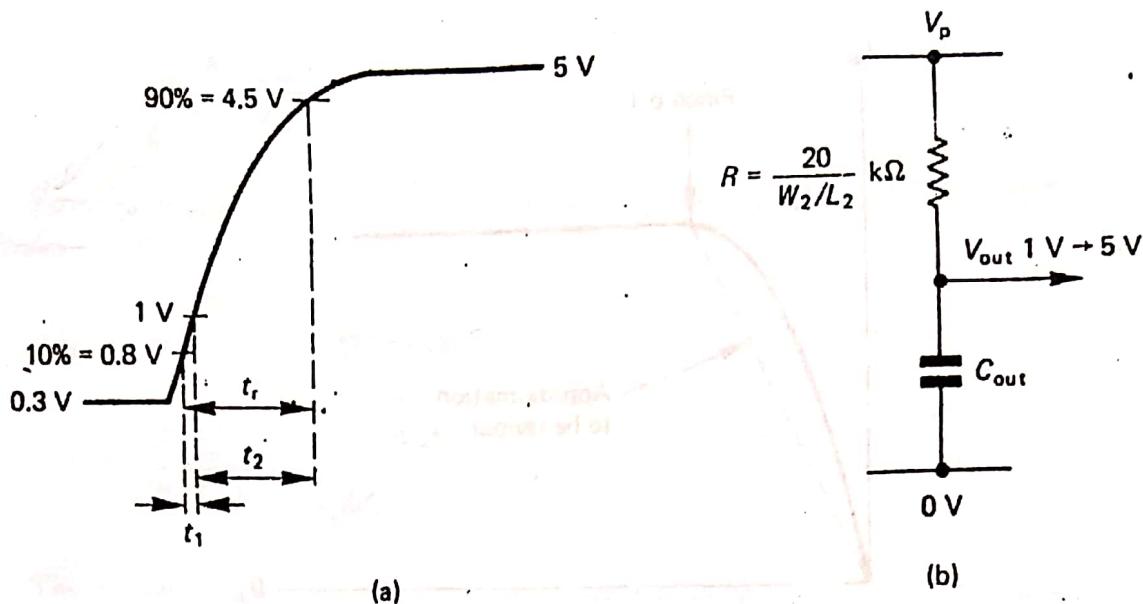


Figure 2.16 Rise time behaviour: (a) rising edge, (b) circuit approximation

During the output rise, T2 is saturated until the output reaches 1 V and thereafter is in the resistive mode. Thus the rise time can be split into two regions:  $t_1$  is the time for the output to rise from 0.8 V to 1 V where T2 is in the saturated mode, and  $t_2$  is the time for the output to rise from 1 V to 4.5 V where T2 operates in the resistive mode. It will be appreciated, from figure 2.16a, that since the initial part of the rise at  $V_{out}$  is fast in comparison with the latter parts and since the output voltage change during  $t_1$  is small in comparison with the voltage change during  $t_2$ , the rise time is effectively determined by  $t_2$  and  $t_1$  can be neglected.

The calculation of  $t_2$  is most easily approached by replacing T2 with a constant resistance  $R$  which is used to represent the transistor's behaviour in the resistive region of the characteristic. The resistance's characteristic is shown as a dotted line on figure 2.17. It has been obtained by drawing a straight line between the origin and the pinch-off point. Using this approximation to transistor behaviour, a value of  $R$  (applicable to any NMOS transistor) can be obtained.

$$R = \frac{\text{pinch-off voltage}}{\text{pinch-off current}} = \frac{(V_{gs} - V_t)}{\frac{\epsilon \mu_n}{2D} \left(\frac{W}{L}\right) (V_{gs} - V_t)^2}$$

$$= \frac{1}{\frac{\epsilon \mu_n}{2D} \left(\frac{W}{L}\right) (V_{gs} - V_t)}$$

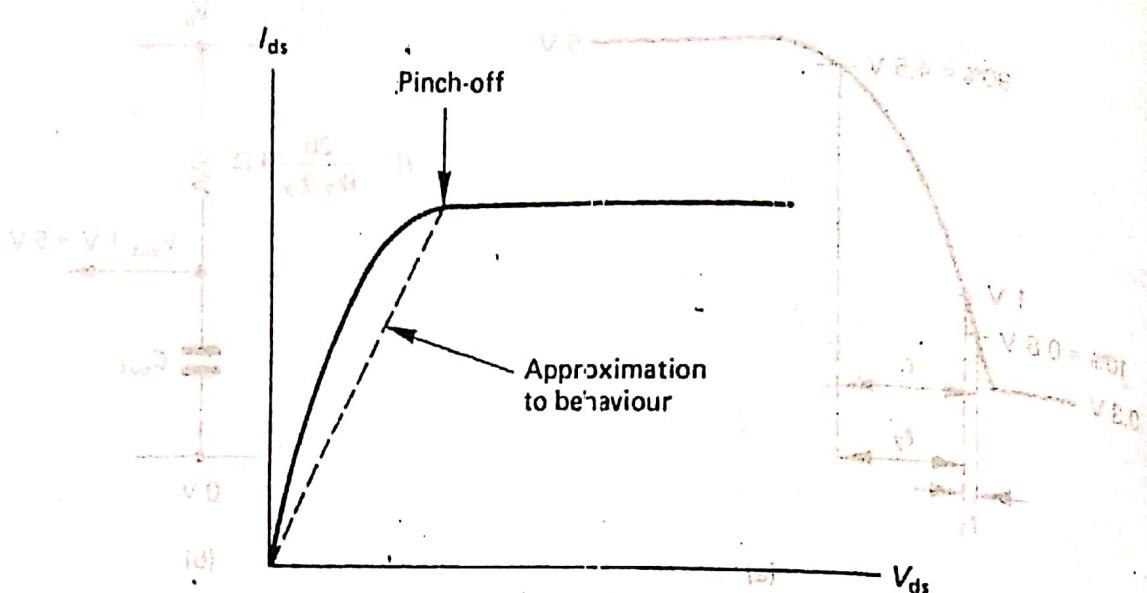


Figure 2.17 Characteristic approximation

Considering transistor T2 and substituting for  $V_{gs}$ ,  $V_{td}$  and  $\epsilon\mu_n/D$ , the resistance can be expressed in terms of T2's aspect ratio

$$R = \frac{1}{25(W_2/E_2)[0 - (-4)]} \text{ M}\Omega = \frac{20}{(W_2/L_2)} \text{ k}\Omega$$

Thus the equivalent circuit during time  $t_2$  simplifies to an RC network, as shown in figure 2.16b. The output voltage  $V_{out}$  across such a network, where the initial voltage across the capacitor is  $V_i$  and the final voltage is  $V_p$ , is given by

$$V_{out} = V_p - (V_p - V_i) \exp[-t/(RC_{out})]$$

The time for the output to reach 4.5 V with an initial output of 1 V is

$$4.5 = 5 - (5 - 1) \exp[-t_2/(RC_{out})]$$

giving

$$t_2 = 2.08RC_{out}$$

If  $C_{out}$  is in picofarads and  $R$  in kilohms, then  $t_2$  is given in nanoseconds. Substituting for T2's value of  $R$ , the rise time is

$$t_r = t_2 = \frac{42C_{out}}{(W_2/L_2)} \text{ ns}$$

From previous considerations concerning the current-drive capability of T1 relative to T2, the fall time  $t_f$  is

$$t_f = \frac{5t_r}{6k} = \frac{5(W_2/L_2)42C_{out}}{6(W_1/L_1)(W_2/L_2)}$$

$$t_f = \frac{35C_{out}}{(W_1/L_1)} \text{ ns}$$

The expression for the fall time agrees well with results obtained by simulating the inverter circuit. However, the expression for the rise time is not so accurate as the body effect is significant for T2. As the output rises, the threshold of T2 rises, reducing the current in T2. This increases the time necessary to charge the output capacitance. Simulation results for the inverter show that

$$t_r = \frac{60C_{out}}{(W_2/L_2)} \text{ ns}$$

is more appropriate for the rise time.

In a fabrication process where the minimum allowable gate width or length is  $6 \mu\text{m}$ , an inverter having T1 and T2 aspect ratios of 3/1 and 1/2 respectively will have a typical input capacitance of  $0.05 \text{ pF}$  and an output capacitance of  $0.1 \text{ pF}$ . Thus the circuit exhibits a rise time of  $12 \text{ ns}$  and a fall time of  $1.2 \text{ ns}$ .

The speed at which the circuit operates is dominated by the load capacitance. In practice, the output of a circuit drives other inputs and each input causes an increase in the load capacitance and hence the edge times. The absolute maximum frequency (in hertz) at which a circuit can operate is  $1/(t_r + t_f)$  so, taking the capacitance figures of the last example, an inverter driving a similar inverter has a total load capacitance of  $0.15 \text{ pF}$  and an upper frequency limit of  $50 \text{ MHz}$ .

## 2.11 Ratioed and Ratioless Design

It is clear that for NMOS inverters with an NMOS depletion or enhancement load, it is necessary to choose suitable gate geometry ratios for both the load and drive transistors in order to obtain the desired low level output. Such designs are therefore called 'ratioed' circuits.

It should be noted that the width-to-length ratio of a transistor is an important design characteristic of MOS circuits. This ratio determines the circuit speed, as

edge times are inversely proportional to aspect ratios. The power dissipation is also determined by the gate geometry, since the current flow through a device is proportional to its aspect ratio.

Ratioless designs are obtained by implementing the inverter load of figure 2.10 with a digital switch where the load switch and the driver switch are never simultaneously closed. Consider that the load and driver switches operate in antiphase, as shown in figure 2.18. A high input causes the driver switch to close and the load switch to remain open,  $V_{out}$  is connected to 0 V. A low input closes the load switch and opens the driver switch,  $V_{out}$  is connected to  $V_p$ .

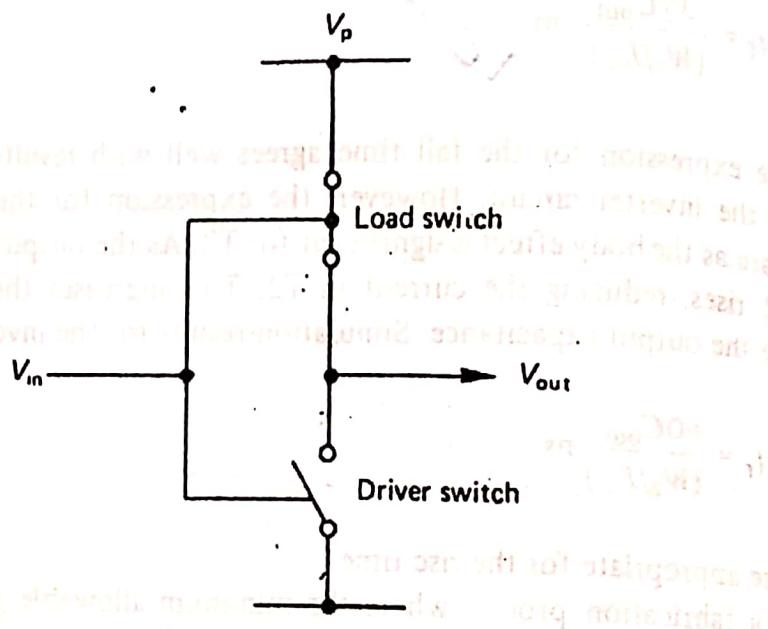


Figure 2.18 Ratioless inverter design

Since only one switch is normally closed at a time, no current flows between  $V_p$  and 0 V. Thus there is no static power dissipation. However, note that when the circuit changes state, both switches may become closed, causing a transient current to flow.

The switches are implemented with MOS transistors and, because the load and driver transistors are never on simultaneously, the correct output levels are obtained regardless of the transistor aspect ratios. For this reason, such circuits are described as 'ratioless'.

## 2.12 The CMOS Inverter

The complementary-channel MOS or CMOS inverter is shown in figure 2.19. It has an n-channel enhancement transistor as the driver switch and a p-channel enhancement device as the load switch. It is assumed that the substrate of T1 is

connected to 0 V and T2's substrate to  $V_p$ . The threshold of T1,  $V_{te}$ , is the minimum gate-source voltage at which T1 conducts, while T2's threshold,  $V_{tsp}$ , is the minimum source-gate voltage at which it turns on.

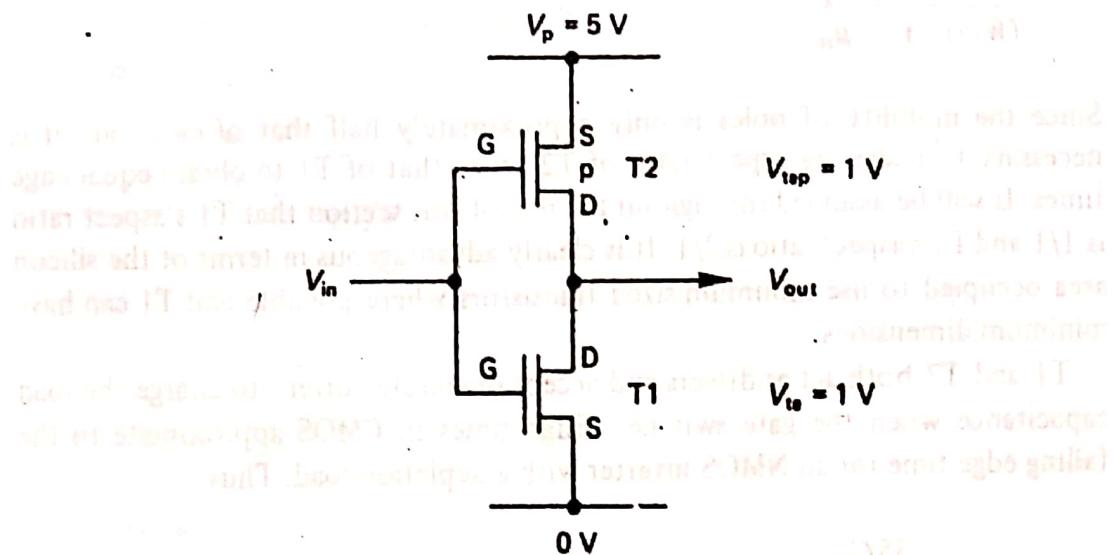


Figure 2.19 CMOS inverter

A low input voltage of 0 V causes T1 to be off since the input is less than T1's threshold. T2 is on as its source-gate voltage is 5 V, which exceeds its 1 V threshold. The current flowing through T2 is the leakage current of T1, which is very small. Thus  $V_{sd}$  of T2 is 0 V and  $V_{out}$  is 5 V; T2 is in the resistive mode.

For a high input voltage of 5 V, T1 is on since the input exceeds its threshold voltage. T2 is off since its source-gate voltage is 0 V, which is below its 1 V threshold. Again, the only current flowing is a very small leakage current (of T2), so  $V_{ds}$  of T1 and hence  $V_{out}$  equal 0 V; T1 is in the resistive mode.

If the current capability of T1 and T2 during switching is the same, then the rising and falling edge times are approximately equal. When  $V_{in}$  changes from a low to a high level instantaneously, T1 turns on and T2 turns off. The load capacitance on the output discharges from 5 V to 0 V via T1. T1 is saturated until the output falls to 4 V and thereafter is in the resistive region. Its  $V_{gs}$  is 5 V throughout the fall time.

Similarly, when  $V_{in}$  changes from high to low instantaneously, T1 turns off and T2 turns on. The load capacitance on the output charges from 0 V to 5 V via T2. During the rise, T2 is saturated until the output reaches 1 V and thereafter is in the resistive mode.  $V_{sg}$  of T2 is 5 V throughout the output rise.

Equations (2.3) and (2.5) are used to find the saturation current of the transistors and equating these currents yields

$$\frac{\epsilon \mu_n}{2D} \left( \frac{W_1}{L_1} \right) (5 - 1)^2 = \frac{\epsilon \mu_p}{2D} \left( \frac{W_2}{L_2} \right) (5 - 1)^2$$

giving

$$\frac{(W_1/L_1)}{(W_2/L_2)} = \frac{\mu_p}{\mu_n}$$

Since the mobility of holes is only approximately half that of electrons, it is necessary to make the aspect ratio of T2 twice that of T1 to obtain equal edge times. It will be assumed throughout the rest of this section that T1's aspect ratio is 1/1 and T2's aspect ratio is 2/1. It is clearly advantageous in terms of the silicon area occupied to use minimum-sized transistors where possible and T1 can have minimum dimensions.

T1 and T2 both act as drivers and accept or supply current to charge the load capacitance when the gate switches. Edge times in CMOS approximate to the falling edge time for an NMOS inverter with a depletion load. Thus

$$t_r = t_f = \frac{35 C_{out}}{(W_1/L_1)} \text{ ns}$$

where  $C_{out}$  is the output load capacitance in pF.

To understand the gate behaviour during switching, it is helpful to examine the gate transfer characteristic ( $V_{out}$  versus  $V_{in}$ ) and the variation of circuit current with input voltage. The easiest way of determining these features is graphically. This is done by first plotting T1's  $I_{ds}$  versus  $V_{ds}$  curves for different values of  $V_{gs}$  above the threshold. Effectively, this is a graph of the circuit current versus  $V_{out}$  for various values of  $V_{in}$ . The resulting curves are shown with broken lines in figure 2.20. It will be noted that the behaviour below pinch-off is taken to be linear. This assumption has a negligible effect upon the transfer and current characteristics to be obtained and greatly simplifies the preparation and plotting of the family of curves for the transistors. Table 2.1 illustrates the simple calculations from which the curves for T1 have been drawn in figure 2.20.

Next, the curves for T2 are superimposed on those for T1. Here,  $V_{sg}$  of T2 is  $5 - V_{in}$ ,  $V_{sd}$  of T2 is  $5 - V_{out}$  and the current  $I_{sd}$  is the circuit current  $I$ . Thus the T2 curves can be expressed in terms of  $V_{out}$  versus  $I$  for different values of  $V_{in}$ . These are shown in figure 2.20 as solid lines and table 2.1 shows in detail the calculations necessary to draw the curves. Again the behaviour below pinch-off is assumed to be linear.

By reading values off the superimposed characteristic plots for T1 and T2,  $V_{out}$  versus  $V_{in}$  and  $I$  versus  $V_{in}$  can be found. For example, when  $V_{in} = 2 \text{ V}$  figure 2.20 shows that the T1 curve for this value of input intersects the T2 curve for  $V_{in} = 2 \text{ V}$  at a  $V_{out}$  of 4.5 V and a circuit current of  $15 \mu\text{A}$ . It can also be seen that at this operating point, T1 is saturated and T2 is resistive.

Table 2.1 Current and voltage calculations for CMOS inverter

*Transistor T1*

$$(W_1/L_1) = 1/1, \epsilon\mu_n/D = 30 \mu\text{A}/\text{V}^2, V_{te} \text{ of T1} = 1 \text{ V}$$

$$\text{Saturation } I_{ds} = 15(V_{gs} - 1)^2 \mu\text{A}$$

Pinch-off occurs when  $V_{ds} = V_{gs} - 1$

$$V_{in} = V_{gs}, V_{out} = V_{ds}, I = I_{ds}$$

$V_{gs}$ (V)	Pinch-off $V_{ds}$ (V)	Saturation $I_{ds}$ ( $\mu\text{A}$ )
2	1	15
2.5	1.5	33.8
3	2	60
4	3	135
5	4	240

*Transistor T2*

$$(W_2/L_2) = 2/1, \epsilon\mu_p/D = 15 \mu\text{A}/\text{V}^2, V_{tep} \text{ of T2} = 1 \text{ V}$$

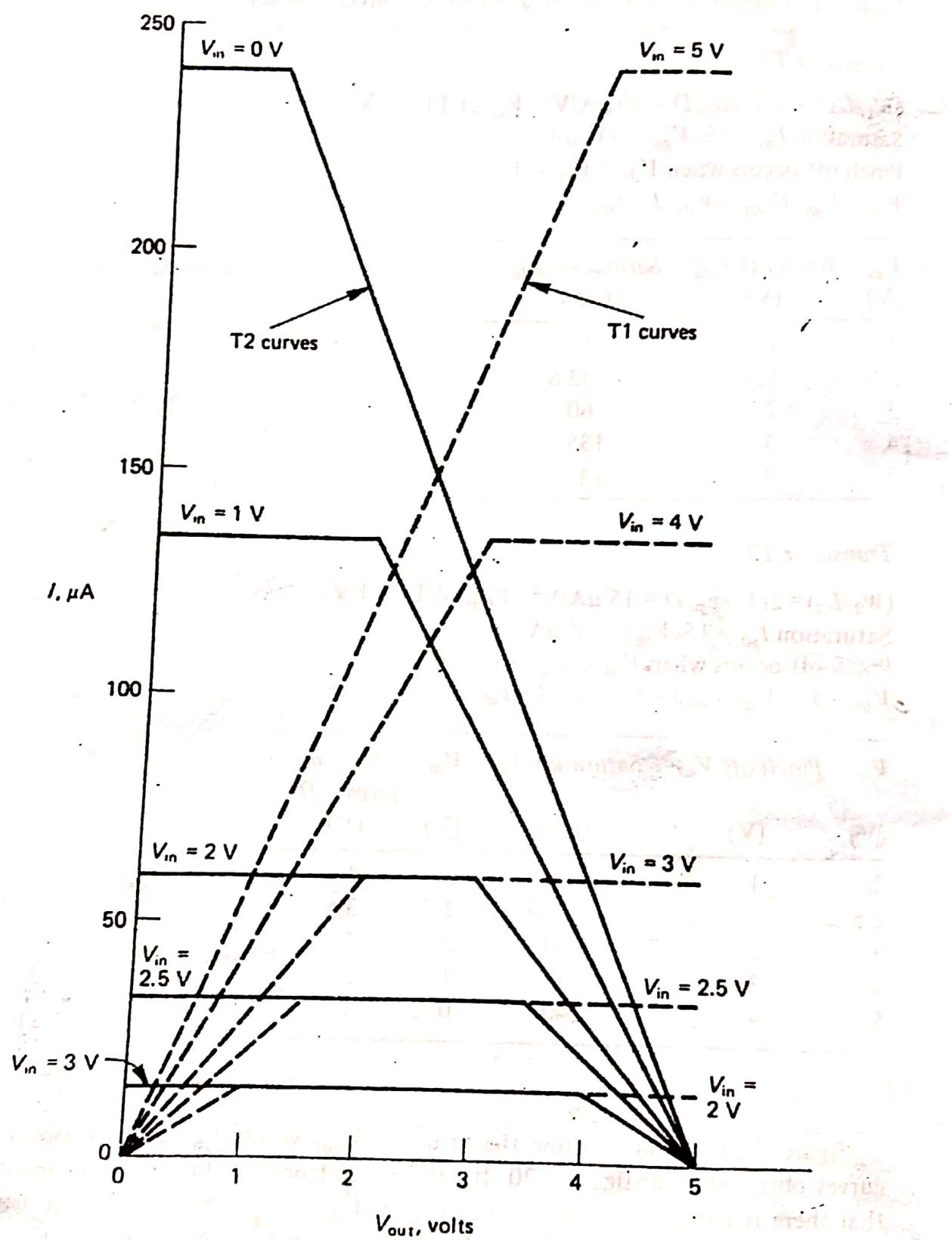
$$\text{Saturation } I_{sd} = 15(V_{sg} - 1)^2 \mu\text{A}$$

Pinch-off occurs when  $V_{sd} = V_{sg} - 1$

$$V_{in} = 5 - V_{sg}, V_{out} = 5 - V_{sd}, I = I_{sd}$$

$V_{sg}$ (V)	Pinch-off $V_{sd}$ (V)	Saturation $I_{sd}$ ( $\mu\text{A}$ )	$V_{in}$ (V)	$V_{out}$ at pinch-off (V)
2	1	15	3	4
2.5	1.5	33.8	2.5	3.5
3	2	60	2	3
4	3	135	1	2
5	4	240	0	1

Figure 2.21, a and b, show the resulting  $V_{out}$  versus  $V_{in}$  and  $I$  versus  $V_{in}$  curves obtained from figure 2.20. It can be seen from the transfer characteristic that there is a very sharp transition in  $V_{out}$  at  $V_{in} = V_p/2$ . During this transition, both T1 and T2 are saturated and the current  $I$  rises to  $33.8 \mu\text{A}$ . The frequency of switching therefore determines the overall power dissipation. The power dissipated is also dependent upon the value of  $V_p$ . A greater value for  $V_p$  increases the current capability, reducing the switching time at the expense of an increased power dissipation. Nowadays, it is common practice to operate CMOS logic from a 5 V supply.

Figure 2.20  $I$  versus  $V_{\text{out}}$  curves for T1 and T2

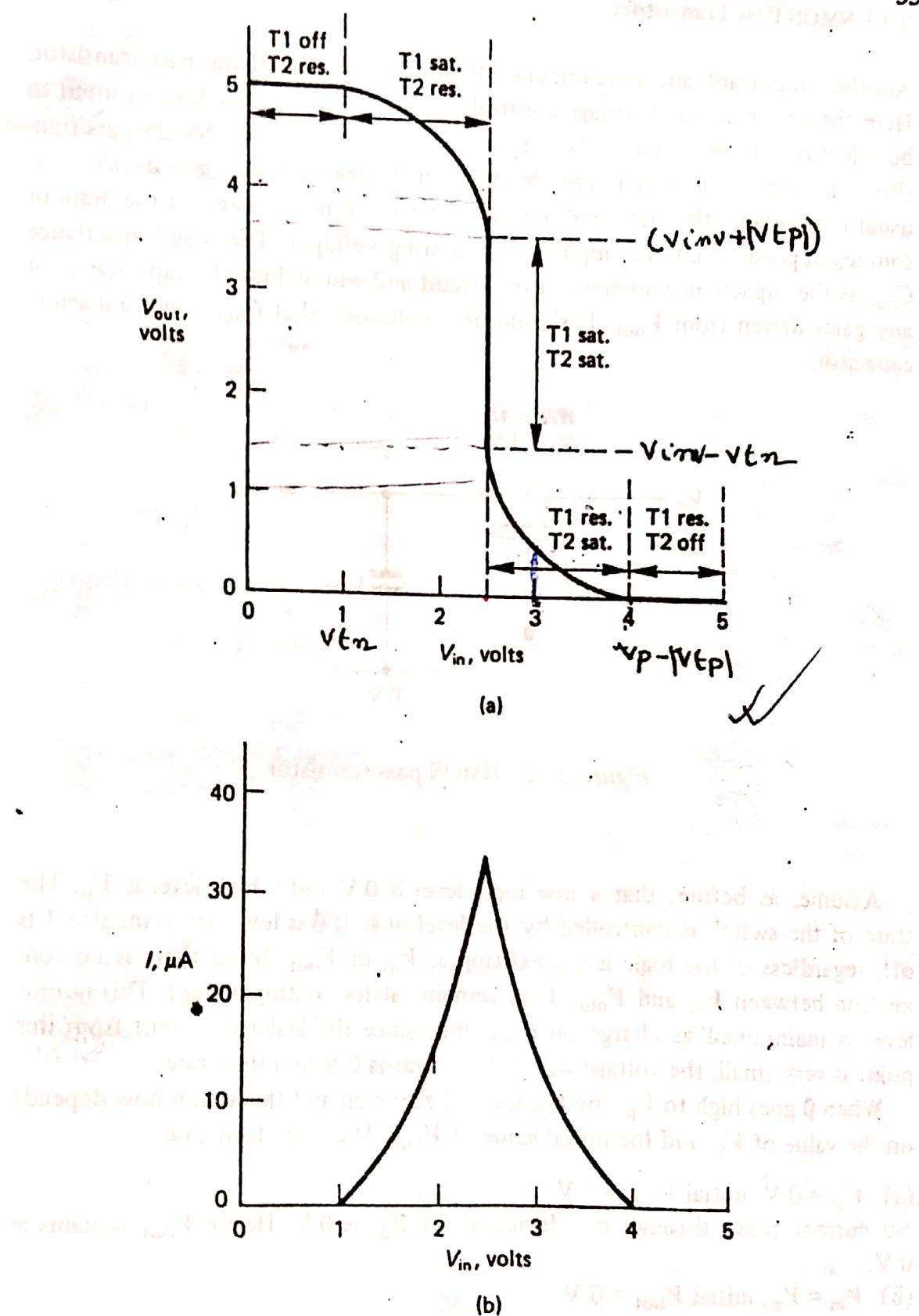


Figure 2.21 Characteristic curves for the CMOS inverter: (a) transfer characteristic. (b) current versus  $V_{in}$

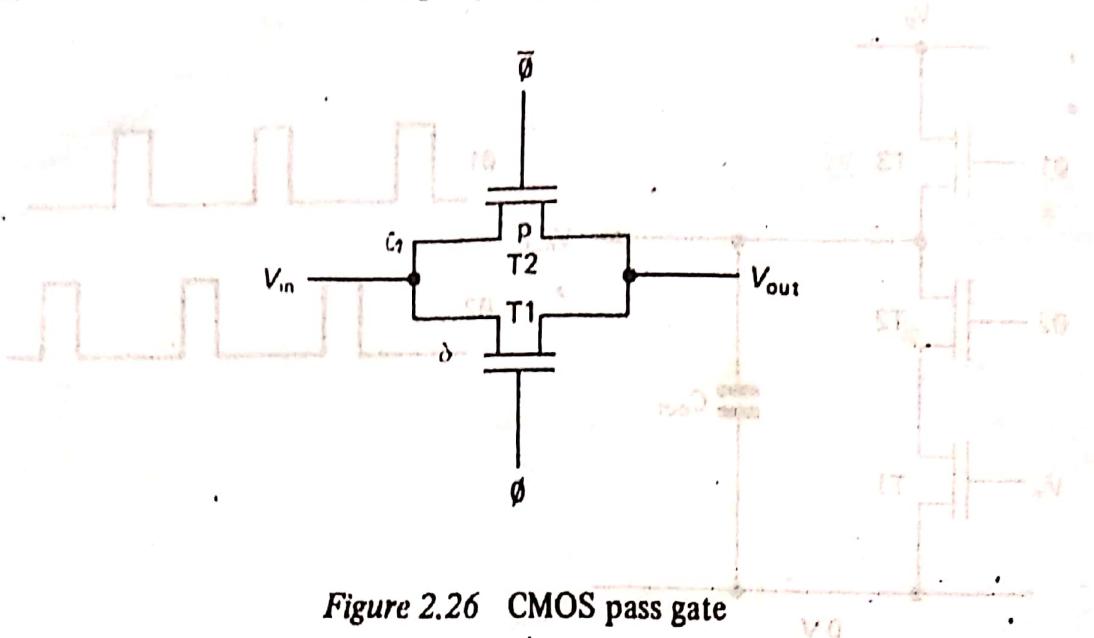


Figure 2.26 CMOS pass gate

(a)  $V_{in} = 0 \text{ V}$ , initial  $V_{out} = 0 \text{ V}$

The drain and source potentials of each device are equal, so no current flows.

$V_{out}$  remains at  $0 \text{ V}$ .

(b)  $V_{in} = V_p$ , initial  $V_{out} = V_p$

Again the drain and source potentials of T1 and T2 are equal, so no current flows and  $V_{out}$  remains at  $V_p$ .

(c)  $V_{in} = V_p$ , initial  $V_{out} = 0 \text{ V}$

Here  $V_{in}$  acts as the drain of T1 and the source of T2. Current flows from  $V_{in}$  to  $V_{out}$ , causing  $V_{out}$  to rise as the capacitance  $C_{out}$  on the output charges up.  $V_{gs}$  of T2 is  $V_p$  throughout this rise, while  $V_{gs}$  of T1 is initially  $V_p$  but decreases as  $V_{out}$  rises. At  $V_{out} = V_p - V_{te}$ , the NMOS transistor turns off and  $V_{out}$  continues to rise to  $V_p$  via T2.

(d)  $V_{in} = 0 \text{ V}$ , initial  $V_{out} = V_p$

$V_{in}$  acts as the source of T1 and the drain of T2. Current flows from  $V_{out}$  to  $V_{in}$ , causing  $V_{out}$  to fall as  $C_{out}$  discharges.  $V_{gs}$  of T1 is constant at  $V_p$  during the fall, while  $V_{gs}$  of T2 is initially  $V_p$  but decreases as  $V_{out}$  falls. At  $V_{out} = V_{te}$ , the PMOS device turns off and  $V_{out}$  continues to fall to  $0 \text{ V}$  via T1.

## 2.16 Buffer Circuits

Some gate outputs in a design, for example clock signals, need to be connected to a large number of gates and thus drive a large capacitive load. The effect of an output driving many inputs directly can be seen from the edge time results obtained in section 2.10. Here it was found that

~~edge time  $\propto \frac{C_{out}}{W/L}$~~

where  $W/L$  is the pull-down transistor aspect ratio.

Thus the edge time increases in proportion to the capacitance driven, and soon becomes unacceptable slow. This speed loss can be avoided by suitably increasing the width-to-length ratio of all transistors in the driving gate; this increases its current capability. Unfortunately, this increase in the dimensions of the driver gate causes its input capacitance to rise. This in turn increases the loading on its preceding gate, causing an unacceptable loss of speed here.

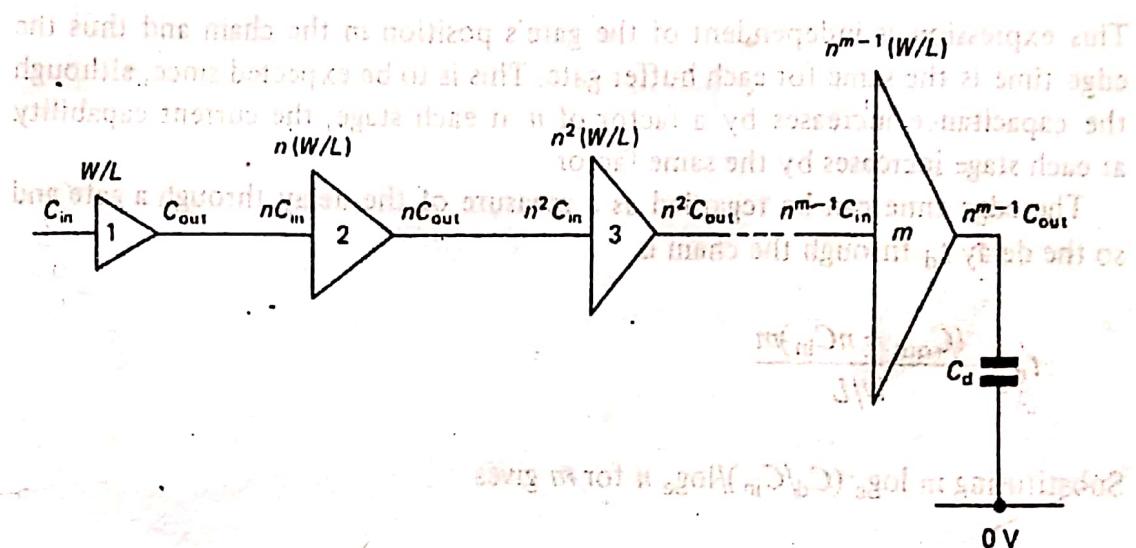


Figure 2.27 Buffer gate chain

For these reasons, a large capacitive load cannot be directly driven from a standard gate output. Instead, a buffer chain is used, as shown in figure 2.27. The aspect ratio of each gate in the chain is larger by a factor  $n$  than those of the preceding gate. This increase in dimensions causes the input and output capacitance of a gate to be  $n$  times that of the preceding gate. These are indicated on figure 2.27 relative to the input and output capacitance,  $C_{in}$  and  $C_{out}$ , of gate 1. If the number of stages in the chain,  $m$ , is chosen so that the capacitance  $C_d$  to be driven by gate  $m$  is  $n$  times this gate's input capacitance, then the load capacitance on each output increases by a factor of  $n$  at each stage and

$$C_d = n^m C_{in}$$

giving

$$m = \frac{\log_e(C_d/C_{in})}{\log_e n}$$

The effect of this cascade arrangement upon the edge times can be estimated from the aspect ratio and load capacitance of each gate in the chain. The load capacitance of gate  $i$  is  $n^{i-1} (C_{\text{out}} + nC_{\text{in}})$  and its aspect ratio is  $n^{i-1} (W/L)$ . So

$$\text{edge time of gate } i \propto \frac{n^{i-1} (C_{\text{out}} + nC_{\text{in}})}{n^{i-1} (W/L)}$$

$$\propto \frac{C_{\text{out}} + nC_{\text{in}}}{W/L}$$

This expression is independent of the gate's position in the chain and thus the edge time is the same for each buffer gate. This is to be expected since, although the capacitance increases by a factor of  $n$  at each stage, the current capability at each stage increases by the same factor.

The edge time can be regarded as a measure of the delay through a gate and so the delay  $t_d$  through the chain is

$$t_d \propto \frac{(C_{\text{out}} + nC_{\text{in}})m}{W/L}$$

Substituting in  $\log_e (C_d/C_{\text{in}})/\log_e n$  for  $m$  gives

$$t_d \propto \frac{(C_{\text{out}} + nC_{\text{in}}) \log_e (C_d/C_{\text{in}})}{(W/L) \log_e n}$$

$$\propto \frac{C_{\text{out}}/C_{\text{in}} + n}{\log_e n}$$

Differentiating this expression with respect to  $n$  yields

$$\frac{dt_d}{dn} \propto \frac{\log_e n - (C_{\text{out}}/C_{\text{in}} + n)/n}{(\log_e n)^2}$$

Taking  $C_{\text{out}}$  as 0.1 pF and  $C_{\text{in}}$  as 0.05 pF and then equating  $dt_d/dn$  to zero to obtain a value for  $n$  which minimises the delay, the optimum value for  $n$  is 4.3. If this is adopted, the number of stages is  $\log_e (C_d/C_{\text{in}})/1.46$ . Thus a 75 pF load representing 1500 driven gates requires a five-stage chain. In practice, the silicon area occupied by such a chain soon becomes significant and it is necessary to accept an increased delay in order to reduce the area and the number of stages required.

In NMOS, the rise and fall times are unequal. It is therefore usual to implement buffer gates with superbuffer circuits which have approximately equal rise