

COMBINATIONAL MOS LOGIC CIRCUITS

Generic combinational logic circuit (gate)

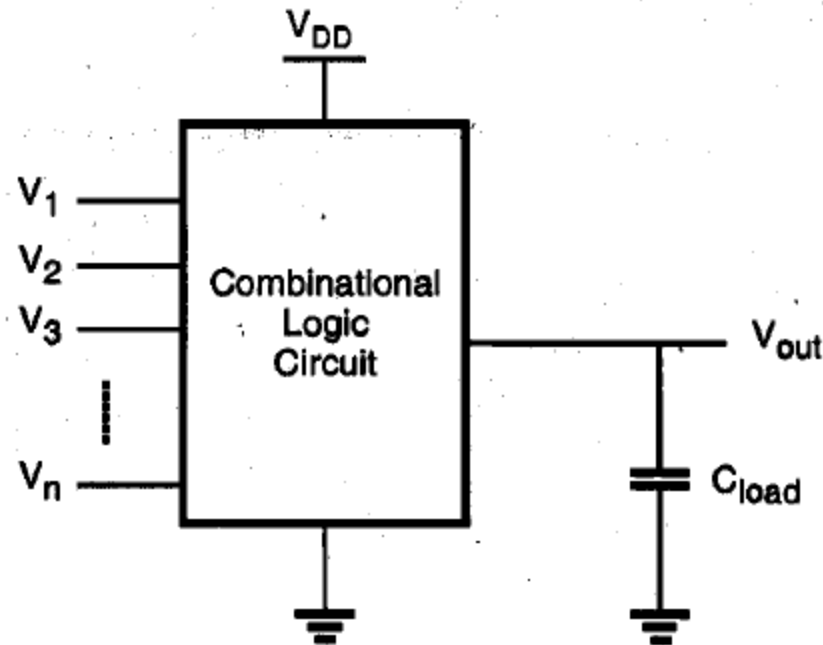


Fig. 1

MOS Logic Circuits with Depletion nMOS Loads

(Two-Input NOR Gate)

- The circuit diagram, the logic symbol, and the corresponding truth table of the gate are given in Fig. 2.
- If the input voltage V_A or the input voltage V_B is equal to the logic-high level, the corresponding driver transistor turns on and provides a conducting path between the output node and the ground.
- Hence, the output voltage becomes low.

MOS Logic Circuits with Depletion nMOS Loads

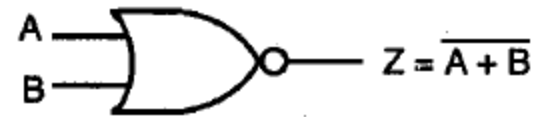
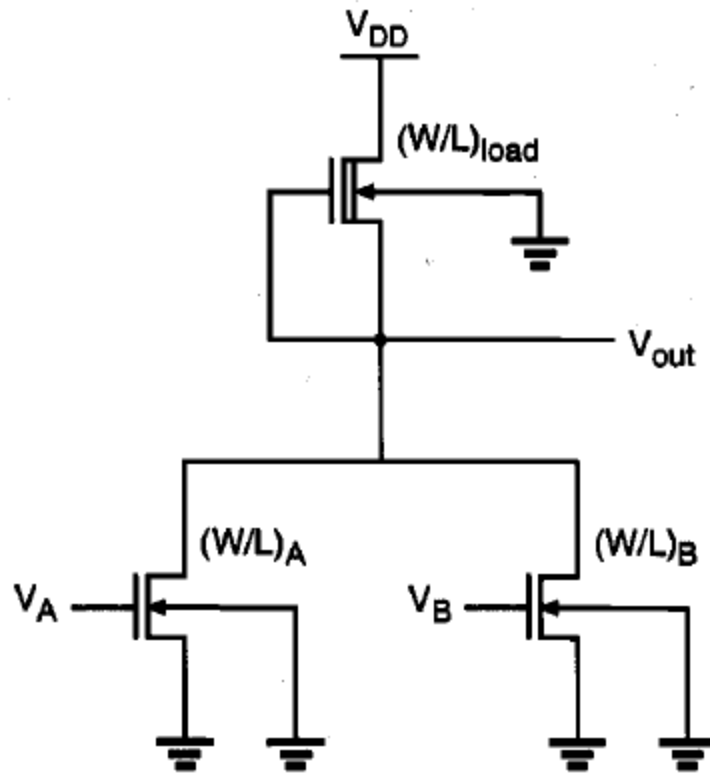
(Two-Input NOR Gate)

- In this case, the circuit operates like a depletion-load inverter with respect to its static behavior.
- A similar result is achieved when both V_A and V_B are high, in which case two parallel conducting paths are created between the output node and the ground.
- If, on the other hand, both V_A and V_B are low, both driver transistors remain cutoff.

MOS Logic Circuits with Depletion nMOS Loads (Two-Input NOR Gate)

- The output node voltage is pulled to a logic-high level by the depletion-type nMOS load transistor.

(Two-Input NOR Gate)



V_A	V_B	V_{out}
low	low	high
low	high	low
high	low	low
high	high	lower

Figure 2. A two-input depletion-load NOR gate, its logic symbol, and the corresponding truth table. Note that the substrates of all transistors are connected to ground.

(Two-Input NOR Gate)

Calculation of V_{OH}

- When both input voltages V_A and V_B are lower than the corresponding driver threshold voltage, the driver transistors are turned off and conduct no drain current.
- Consequently, the load device, which operates in the linear region, also has zero drain current.

(Two-Input NOR Gate)

Calculation of V_{OH}

- Thus the linear region current equation becomes;

$$I_{D,load} = \frac{k_{n,load}}{2} \cdot \left[2|V_{T,load}(V_{OH})| \cdot (V_{DD} - V_{OH}) - (V_{DD} - V_{OH})^2 \right] = 0 \quad \dots(1)$$

- The solution of this equation gives $V_{OH} = V_{DD}$.

(Two-Input NOR Gate)

Calculation of V_{OL}

- To calculate the output low voltage V_{OL} we must consider three different cases, i.e., three different input voltage combinations, which produce a conducting path from the output node to the ground. These cases are

(i) $V_A = V_{OH} \quad V_B = V_{OL}$

(ii) $V_A = V_{OL} \quad V_B = V_{OH}$

(iii) $V_A = V_{OH} \quad V_B = V_{OH}$

(Two-Input NOR Gate)

Calculation of V_{OL}

- For the first two cases, (i) and (ii), the NOR circuit reduces to a simple nMOS depletion load inverter.
- Assuming that the threshold voltages of the two enhancement-type driver transistors are identical ($V_{T0,A} = V_{T0,B} = V_{T0}$), the driver-to-load ratio of the corresponding inverter can be found as follows.

(Two-Input NOR Gate)

Calculation of V_{OL}

- In case (i), where the driver transistor A is on, the ratio is

$$k_R = \frac{k_{driver,A}}{k_{load}} = \frac{k'_{n,driver} \left(\frac{W}{L} \right)_A}{k'_{n,load} \left(\frac{W}{L} \right)_{load}} \dots (2)$$

- In case (ii), where the driver transistor B is on, the ratio is

$$k_R = \frac{k_{driver,B}}{k_{load}} = \frac{k'_{n,driver} \left(\frac{W}{L} \right)_B}{k'_{n,load} \left(\frac{W}{L} \right)_{load}} \dots (3)$$

(Two-Input NOR Gate)

Calculation of V_{OL}

- The output low voltage level V_{OL} in both cases is found as follows;

$$V_{OL} = V_{OH} - V_{T0} - \sqrt{(V_{OH} - V_{T0})^2 - \left(\frac{k_{load}}{k_{driver}}\right) \cdot |V_{T,load}(V_{OL})|^2} \quad \dots(4)$$

- If the (W/L) ratios of both drivers are identical, i.e., $(W/L)_A = (W/L)_B$ the output low voltage (V_{OL}) values calculated for case (i) and case (ii) will be identical.

(Two-Input NOR Gate)

Calculation of V_{OL}

- In case (iii), where both driver transistors are turned on, the saturated load current is the sum of the two linear-mode driver currents.

$$I_{D,load} = I_{D,driverA} + I_{D,driverB} \quad \dots(5)$$

$$\begin{aligned} \frac{k_{load}}{2} |V_{T,load}(V_{OL})|^2 = & \frac{k_{driver,A}}{2} [2(V_A - V_{T0})V_{OL} - V_{OL}^2] \\ & + \frac{k_{driver,B}}{2} [2(V_B - V_{T0})V_{OL} - V_{OL}^2] \quad \dots(6) \end{aligned}$$

(Two-Input NOR Gate)

Calculation of V_{OL}

- Since the gate voltages of both driver transistors are equal ($V_A = V_B = V_{OH}$), we can devise an equivalent driver-to-load ratio for the NOR structure:

$$k_R = \frac{k_{driver,A} + k_{driver,B}}{k_{load}} = \frac{k'_{n,driver} \left[\left(\frac{W}{L} \right)_A + \left(\frac{W}{L} \right)_B \right]}{k'_{n,load} \left(\frac{W}{L} \right)_{load}} \dots (7)$$

(Two-Input NOR Gate)

Calculation of V_{OL}

- Thus, the NOR gate with both of its inputs tied to a logic-high voltage is replaced with an nMOS depletion-load inverter circuit with the driver-to-load ratio given by (Eqn 7).
- The output voltage level in this case is

$$V_{OL} = V_{OH} - V_{T0} - \sqrt{(V_{OH} - V_{T0})^2 - \left(\frac{k_{load}}{k_{driver,A} + k_{driver,B}} \right) \cdot |V_{T,load}(V_{OL})|^2} \quad \dots(8)$$

(Two-Input NOR Gate)

Calculation of V_{OL}

- The V_{OL} given by (Eqn. 8) is lower than the V_{OL} values calculated for case (i) and for case (ii), when only one input is logic-high.
- We conclude that the worst-case condition from the static operation viewpoint, i.e., the highest possible V_{OL} value, is observed in case (i) or in (ii).

(Two-Input NOR Gate)

Calculation of V_{OL}

- If $k_{\text{driver}, A} = k_{\text{driver}, B} = k_R k_{\text{load}} \dots (9)$
- This design choice yields two identical driver transistors, which guarantee the required value of V_{OL} in the worst case. When both inputs are logic-high, the output voltage is even lower than the required maximum V_{OL} thus the design constraint is satisfied.

Generalized NOR Structure with Multiple Inputs

- A generalized n-input NOR gates, which consist of n parallel driver transistors, is shown in Fig. 3.
- The combined current I_D in this circuit is supplied by the driver transistors which are turned on, i.e., transistors which have gate voltages higher than the threshold voltage V_{T0} .

Generalized NOR Structure with Multiple Inputs

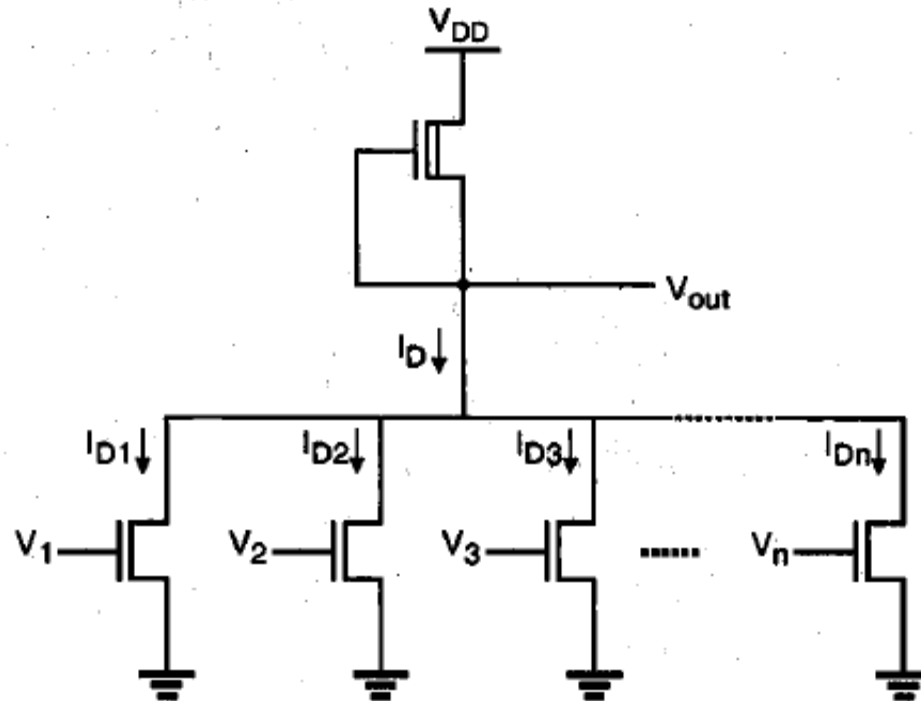


Figure 3. Generalized n -input NOR gate

Generalized NOR Structure with Multiple Inputs

- The combined pull-down current can then be expressed as follows;

$$I_D = \sum_{k(on)} I_{D,k} = \begin{cases} \sum_{k(on)} \frac{\mu_n C_{ox}}{2} \left(\frac{W}{L} \right)_k \left[2(V_{GS,k} - V_{T0})V_{out} - V_{out}^2 \right] & \text{linear} \\ \sum_{k(on)} \frac{\mu_n C_{ox}}{2} \left(\frac{W}{L} \right)_k (V_{GS,k} - V_{T0})^2 & \text{saturation} \end{cases}$$

....(10)

Generalized NOR Structure with Multiple Inputs

- Assuming that the input voltages of all driver transistors are identical the pull-down current expression can be rewritten as;

$$I_D = \begin{cases} \frac{\mu_n C_{ox}}{2} \left(\sum_{k(on)} \left(\frac{W}{L} \right)_k \right) [2(V_{GS} - V_{T0})V_{out} - V_{out}^2] & \text{linear} \\ \frac{\mu_n C_{ox}}{2} \left(\sum_{k(on)} \left(\frac{W}{L} \right)_k \right) (V_{GS} - V_{T0})^2 & \text{saturation} \end{cases} \dots(11)$$

Generalized NOR Structure with Multiple Inputs

- Thus, the multiple-input NOR gate can also be reduced to an equivalent inverter, shown in Fig. 4, for static analysis.
- The (W/L) ratio of the driver transistor here is

$$\left(\frac{W}{L}\right)_{\text{equivalent}} = \sum_{k(\text{on})} \left(\frac{W}{L}\right)_k \quad \dots(12)$$

Generalized NOR Structure with Multiple Inputs

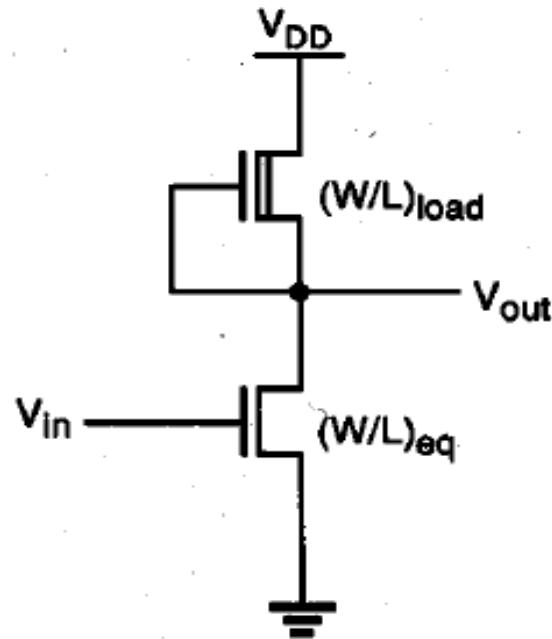


Figure 4. =====

Generalized NOR Structure with Multiple Inputs

- The source terminals of all enhancement-type nMOS driver transistors in the NOR gate are connected to ground.
- Thus, the drivers do not experience any substrate-bias effect.
- The depletion-type nMOS load transistor, is subject to substrate-bias effect, since its source is connected to the output node, and its source-to-substrate voltage is $V_{SB} = V_{out}$.

Transient Analysis of NOR Gate

- Figure 5 shows the two-input NOR (NOR2) gate with all of its relevant parasitic device capacitances.
- As in the inverter case, we can combine the capacitances seen in Fig. 5 into one lumped capacitance, connected between the output node and the ground.

Transient Analysis of NOR Gate

- The value of this combined load capacitance, C_{load} , can be found as;

$$C_{load} = C_{gd,A} + C_{gd,B} + C_{gd,load} + C_{db,A} + C_{db,B} + C_{sb,load} + C_{wire} \quad \dots(13)$$

Transient Analysis of NOR Gate

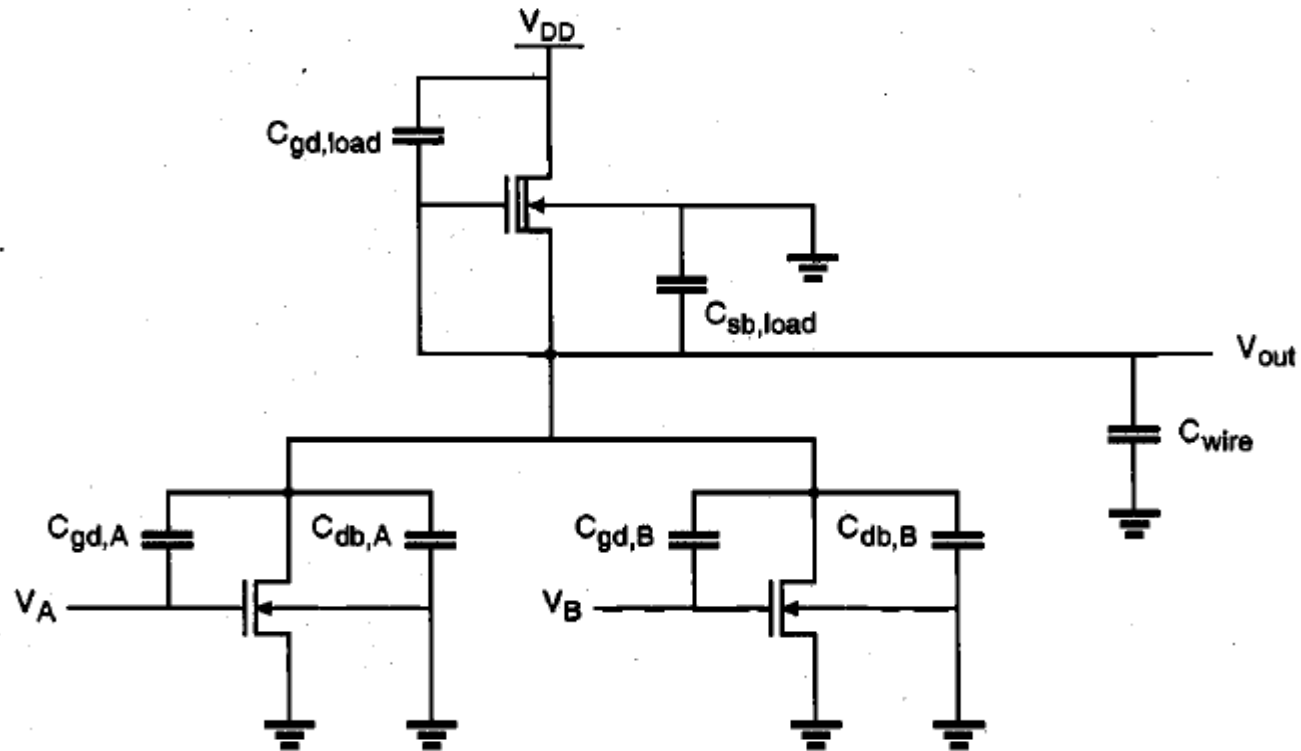


Figure 5. Parasitic device capacitances in the NOR2 gate and the lumped equivalent load capacitance. The gate-to-source capacitances of the driver transistors are included in the load of the previous stages driving the inputs A and B.

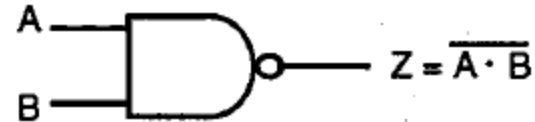
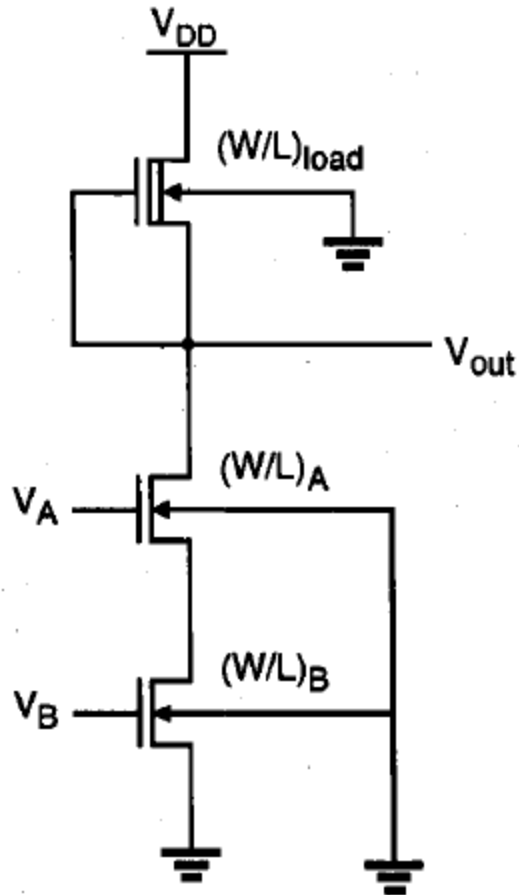
Two-Input NAND Gate

- Now, let us examine the two-input NAND (NAND2) gate.
- The circuit diagram, the logic symbol, and the corresponding truth table of the gate are given in Fig. 6.
- There is a conducting path between the output node and the ground only if the input voltage V_A and the input voltage V_B are equal to logic-high, i.e., only if both of the series-connected drivers are turned on.

Two-Input NAND Gate

- For all of the three input combinations which produce a logic-high output voltage, the corresponding V_{OH} value can easily be found as $V_{OH} = V_{DD}$.
- The calculation of the logic-low voltage V_{OL} , on the other hand, requires a closer investigation.

Two-Input NAND Gate



V_A	V_B	V_{out}
low	low	high
low	high	high
high	low	high
high	high	low

Figure 6. A two-input depletion-load NAND gate, its logic symbol, and the corresponding truth table. Notice the substrate-bias effect for all nMOS transistors except one.

Two-Input NAND Gate

- Consider the NAND2 gate with both of its inputs equal to V_{OH} as shown in Fig. 7. It can easily be seen that the drain currents of all transistors in the circuit are equal to each other.

$$I_{D,load} = I_{D,driverA} = I_{D,driverB} \quad \dots(14)$$

$$\begin{aligned} \frac{k_{load}}{2} |V_{T,load}(V_{OL})|^2 &= \frac{k_{driver,A}}{2} [2(V_{GS,A} - V_{T,A})V_{DS,A} - V_{DS,A}^2] \\ &= \frac{k_{driver,B}}{2} [2(V_{GS,B} - V_{T,B})V_{DS,B} - V_{DS,B}^2] \quad \dots(15) \end{aligned}$$

Two-Input NAND Gate

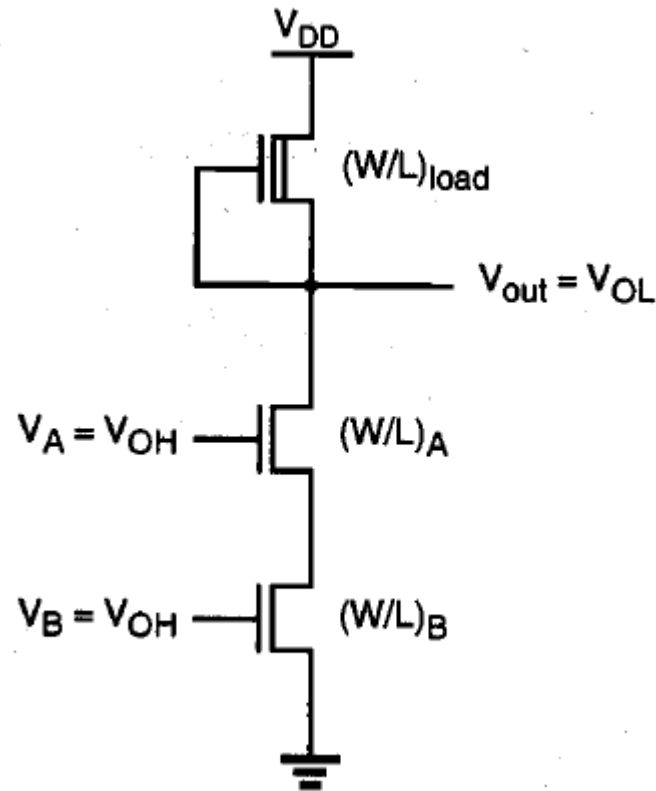


Figure 7. The NAND2 gate with both of its inputs at logic-high level.

Two-Input NAND Gate

- The gate-to-source voltages of both driver transistors can be assumed to be approximately equal to V_{OH} . Also, we may neglect, for simplicity, the substrate-bias effect for driver transistor A, and assume $V_{T,A} = V_{T,B} = V_{T0}$, since the source-to-substrate voltage of driver A is relatively low.

Two-Input NAND Gate

- The drain-to-source voltages of both driver transistors can then be represented as;

$$V_{DS,A} = V_{OH} - V_{T0} - \sqrt{(V_{OH} - V_{T0})^2 - \left(\frac{k_{load}}{k_{driver,A}} \right) \cdot |V_{T,load}(V_{OL})|^2} \quad \dots(16)$$

$$V_{DS,B} = V_{OH} - V_{T0} - \sqrt{(V_{OH} - V_{T0})^2 - \left(\frac{k_{load}}{k_{driver,B}} \right) \cdot |V_{T,load}(V_{OL})|^2} \quad \dots(17)$$

Two-Input NAND Gate

- Let the two driver transistors be identical, i.e., $k_{\text{driver,A}} = k_{\text{driver,B}} = k_{\text{driver}}$.
- Noting that the output voltage V_{OL} is equal to the sum of the drain-to-source voltages of both drivers, we obtain

$$V_{OL} \approx 2 \left(V_{OH} - V_{T0} - \sqrt{(V_{OH} - V_{T0})^2 - \left(\frac{k_{load}}{k_{driver}} \right) \cdot |V_{T,load}(V_{OL})|^2} \right) \dots (18)$$

Two-Input NAND Gate

- Assume that $V_{T,A} = V_{T,B} = V_{T0}$
- When both driver transistors are in the linear region, the drain currents can be written as

$$I_{D,A} = \frac{k_{driver}}{2} \left[2(V_{GS,A} - V_{T0})V_{DS,A} - V_{DS,A}^2 \right] \quad \dots(19)$$

$$I_{D,B} = \frac{k_{driver}}{2} \left[2(V_{GS,B} - V_{T0})V_{DS,B} - V_{DS,B}^2 \right] \quad \dots(20)$$

Two-Input NAND Gate

- Since $I_{D,A} = I_{D,B}$ this current can also be expressed as

$$I_D = I_{D,A} = I_{D,B} = \frac{I_{D,A} + I_{D,B}}{2} \quad \dots(21)$$

Two-Input NAND Gate

- Using $V_{GS,A} = V_{GS,B} - V_{DS,B}$ (Eqn. 21) yields

$$I_D = \frac{k_{driver}}{4} \left[2(V_{GS,B} - V_{TO})(V_{DS,A} + V_{DS,B}) - (V_{DS,A} + V_{DS,B})^2 \right] \dots(22)$$

Two-Input NAND Gate

- Now let $V_{GS} = V_{GS,B}$ and $V_{DS} = V_{DS,A} + V_{DS,B}$. The drain-current expression can then be written as follows.

$$I_D = \frac{k_{driver}}{4} \left[2(V_{GS} - V_{T0})V_{DS} - V_{DS}^2 \right] \quad \dots(23)$$

- Thus, two nMOS transistors connected in series and with the same gate voltage behave like one nMOS transistor with $k_{eq} = 0.5 k_{driver}$.

Generalized NAND Structure with Multiple Inputs

- A generalized n-input NAND gates, which consist of n series-connected driver transistors, as shown in Fig. 8.
- Neglecting the substrate bias effect, and assuming that the threshold voltages of all transistors are equal to V_{T0} the driver current I_D in the linear region can be derived as;

$$I_D = \frac{\mu_n C_{ox}}{2} \left(\frac{1}{\sum_{k(on)} \left(\frac{W}{L} \right)_k} \right) \cdot \begin{cases} [2(V_{in} - V_{T0})V_{out} - V_{out}^2] & \text{linear} \\ (V_{in} - V_{T0})^2 & \text{saturation} \end{cases} \dots (24)$$

Generalized NAND Structure with

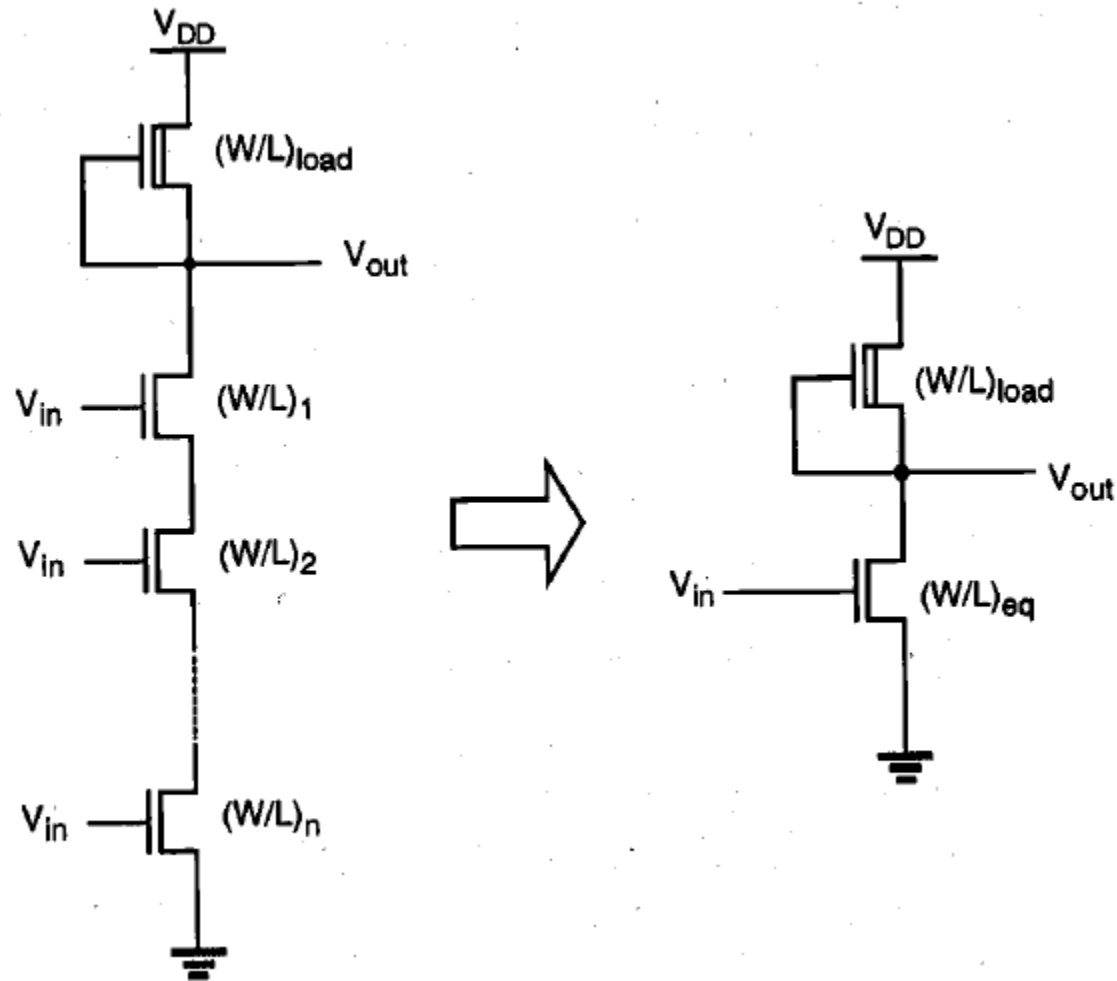


Figure 8. The generalized NAND structure and its inverter equivalent.

Generalized NAND Structure with Multiple Inputs

- Hence, the (W/L) ratio of the equivalent driver transistor is

$$\left(\frac{W}{L}\right)_{\text{equivalent}} = \frac{1}{\sum_{k(\text{on})} \frac{1}{\left(\frac{W}{L}\right)_k}} \quad \dots(25)$$

Generalized NAND Structure with Multiple Inputs

- If the series-connected transistors are identical, i.e., $(W/L)_1 = (W/L)_2 = \dots = (W/L)$, the width-to-length ratio of the equivalent transistor becomes

$$\left(\frac{W}{L}\right)_{\text{equivalent}} = \frac{1}{n} \left(\frac{W}{L}\right) \quad \dots(26)$$

Transient Analysis of NAND Gate

- Figure 9 shows a NAND2 gate with all parasitic device capacitances.
- As in the inverter case, we can combine the capacitances seen in Fig. 9 into one lumped capacitance, connected between the output node and the ground.

Transient Analysis of NAND Gate

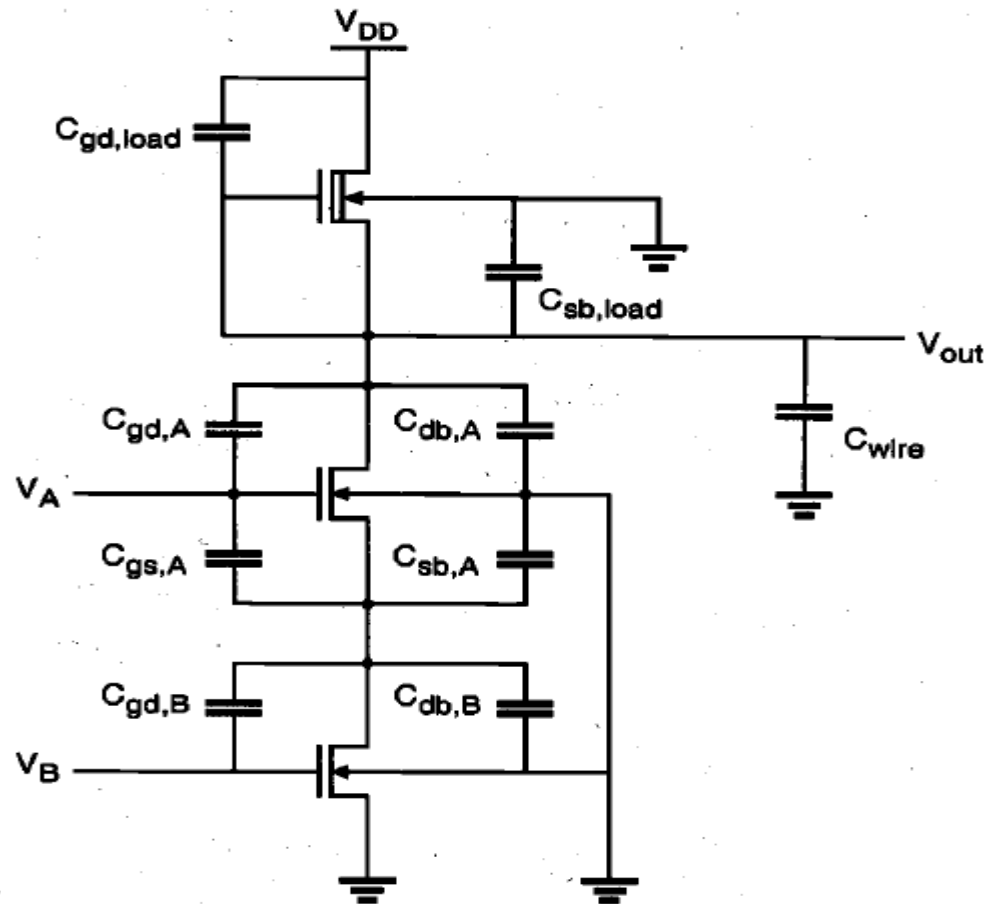


Figure 9. Parasitic device capacitances in the NAND2 gate.

Transient Analysis of NAND Gate

- The value of the lumped capacitance C_{load} however, depends on the input voltage conditions.
- Assume, for example, that the input V_A is equal to V_{OH} and the other input V_B is switching from V_{OH} to V_{OL} .
- In this case, both the output voltage V_{out} , and the internal node voltage V_x will rise, resulting in

$$C_{load} = C_{gd,load} + C_{gd,A} + C_{gd,B} + C_{gs,A} + C_{db,A} + C_{db,B} + C_{sb,A} + C_{sb,load} + C_{wire} \quad \dots(27)$$

Transient Analysis of NAND Gate

- Now let us consider another case where V_B is equal to V_{OH} and V_A switches from V_{OH} to V_{OL} .
- In this case, the output voltage V_{out} , will rise, but the internal node voltage V_x will remain low because the bottom driver transistor is on.
- Thus, the lumped output capacitance

$$C_{load} = C_{gd,load} + C_{gd,A} + C_{db,A} + C_{sb,load} + C_{wire} \quad \dots(28)$$

CMOS Logic Circuits

CMOS NOR2 (Two-Input NOR) Gate

- Figure 10 shows the circuit diagram of a two-input CMOS NOR gate.
- The circuit consists of a parallel-connected n-net and a series-connected complementary p-net.
- The input voltages V_A and V_B are applied to the gates of one nMOS and one pMOS transistor.

CMOS Logic Circuits

CMOS NOR2 (Two-Input NOR) Gate

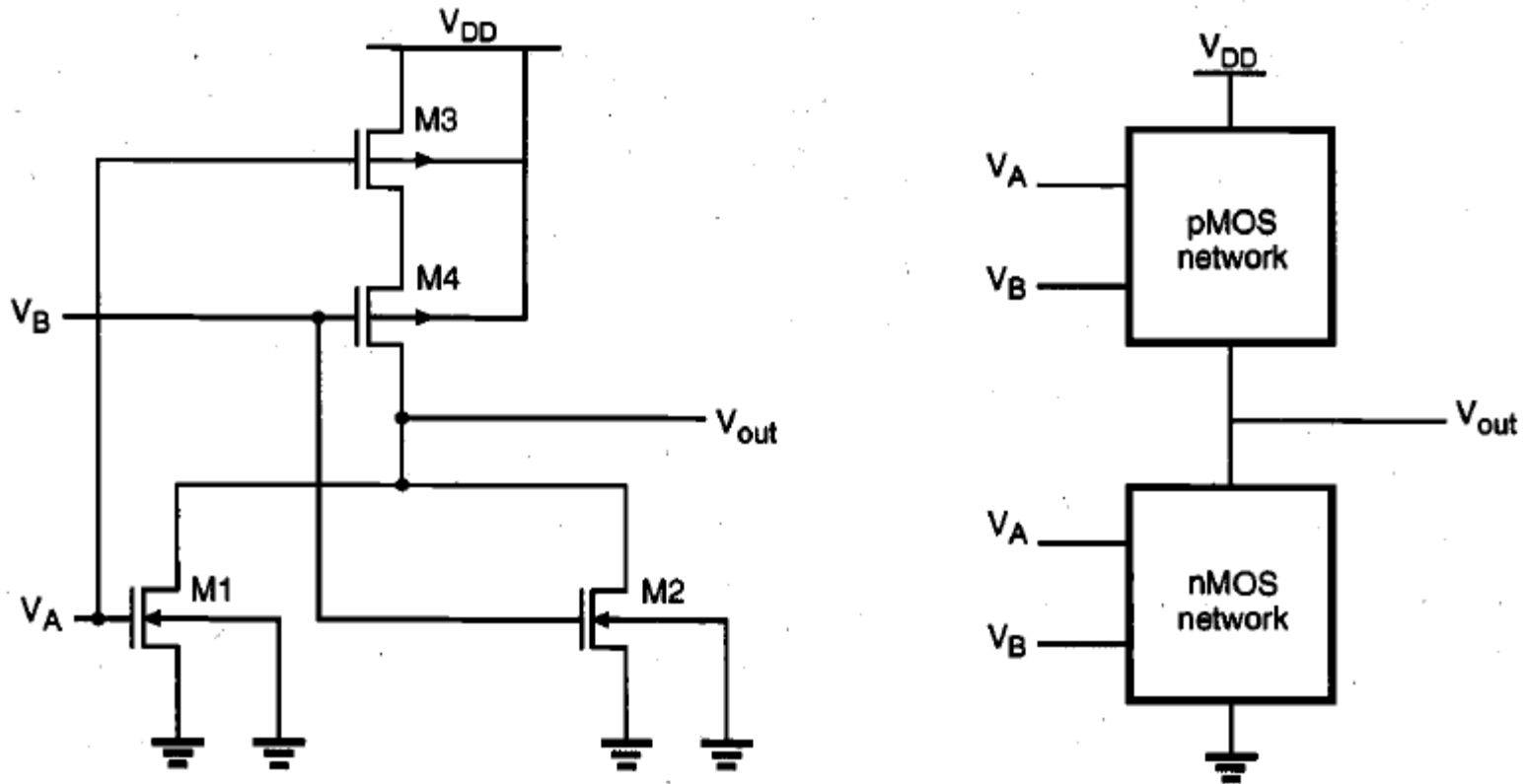


Figure 10. A CMOS NOR2 gate and its complementary operation: Either the nMOS network is on and the pMOS network is off, or the pMOS network is on and the nMOS network is off.

CMOS Logic Circuits

CMOS NOR2 (Two-Input NOR) Gate

- The output voltage of the CMOS NOR2 gate will attain a logic-low voltage of $V_{OL} = 0$ and a logic-high voltage of $V_{OH} = V_{DD}$.
- We start our analysis of the switching threshold by assuming that both input voltages switch simultaneously, i.e., $V_A = V_B$.
- It is assumed that the device sizes in each block are identical, $(W/L)_{n,A} = (W/L)_{n,B}$ and $(W/L)_{p,A} = (W/L)_{p,B}$ and the substrate-bias effect for the pMOS transistors is neglected for simplicity.

CMOS Logic Circuits

CMOS NOR2 (Two-Input NOR) Gate

- By definition, the output voltage is equal to the input voltage at the switching threshold.

$$V_A = V_B = V_{out} = V_{th} \quad \dots(29)$$

- The two parallel nMOS transistors are saturated at this point, because $V_{GS} = V_{DS}$. The combined drain current of the two nMOS transistors is

$$I_D = k_n (V_{th} - V_{T,n})^2 \quad \dots(30)$$

CMOS Logic Circuits

CMOS NOR2 (Two-Input NOR) Gate

- Thus, we obtain the first equation for the switching threshold V_{th}

$$V_{th} = V_{T,n} + \sqrt{\frac{I_D}{k_n}} \quad \dots(31)$$

CMOS Logic Circuits

CMOS NOR2 (Two-Input NOR) Gate

- Examination of the p-net in Fig. 10 shows that the pMOS transistor M_3 operates in the linear region, while the other pMOS transistor, M_4 is in saturation for $V_{in} = V_{out}$,
- Thus

$$I_{D3} = \frac{k_p}{2} \left[2(V_{DD} - V_{th} - |V_{T,p}|)V_{SD3} - V_{SD3}^2 \right] \quad \dots(32)$$

$$I_{D4} = \frac{k_p}{2} (V_{DD} - V_{th} - |V_{T,p}| - V_{SD3})^2 \quad \dots(33)$$

CMOS Logic Circuits

CMOS NOR2 (Two-Input NOR) Gate

- The drain currents of both pMOS transistors are identical, i.e., $I_{D3}=I_{D4}=I_D$.
- Thus

$$V_{DD} - V_{th} - |V_{T,p}| = 2 \sqrt{\frac{I_D}{k_p}} \quad \dots(34)$$

CMOS Logic Circuits

CMOS NOR2 (Two-Input NOR) Gate

- This yields the second equation of the switching threshold voltage V_{th} .
- Combining (Eqn. 31) and (Eqn. 34), we obtain

$$V_{th}(\text{NOR2}) = \frac{V_{T,n} + \frac{1}{2} \sqrt{\frac{k_p}{k_n}} (V_{DD} - |V_{T,p}|)}{1 + \frac{1}{2} \sqrt{\frac{k_p}{k_n}}} \dots (35)$$

CMOS Logic Circuits

CMOS NOR2 (Two-Input NOR) Gate

- The switching threshold voltage of the NOR2 gate can also be obtained by using the equivalent-inverter approach.
- When both inputs are identical, the parallel-connected nMOS transistors can be represented by a single nMOS transistor with $2k_n$.
- Similarly, the series-connected pMOS transistors are represented by a single pMOS transistor with $k_p/2$.
- The resulting equivalent CMOS inverter is shown in Fig. 11.

CMOS Logic Circuits

CMOS NOR2 (Two-Input NOR) Gate

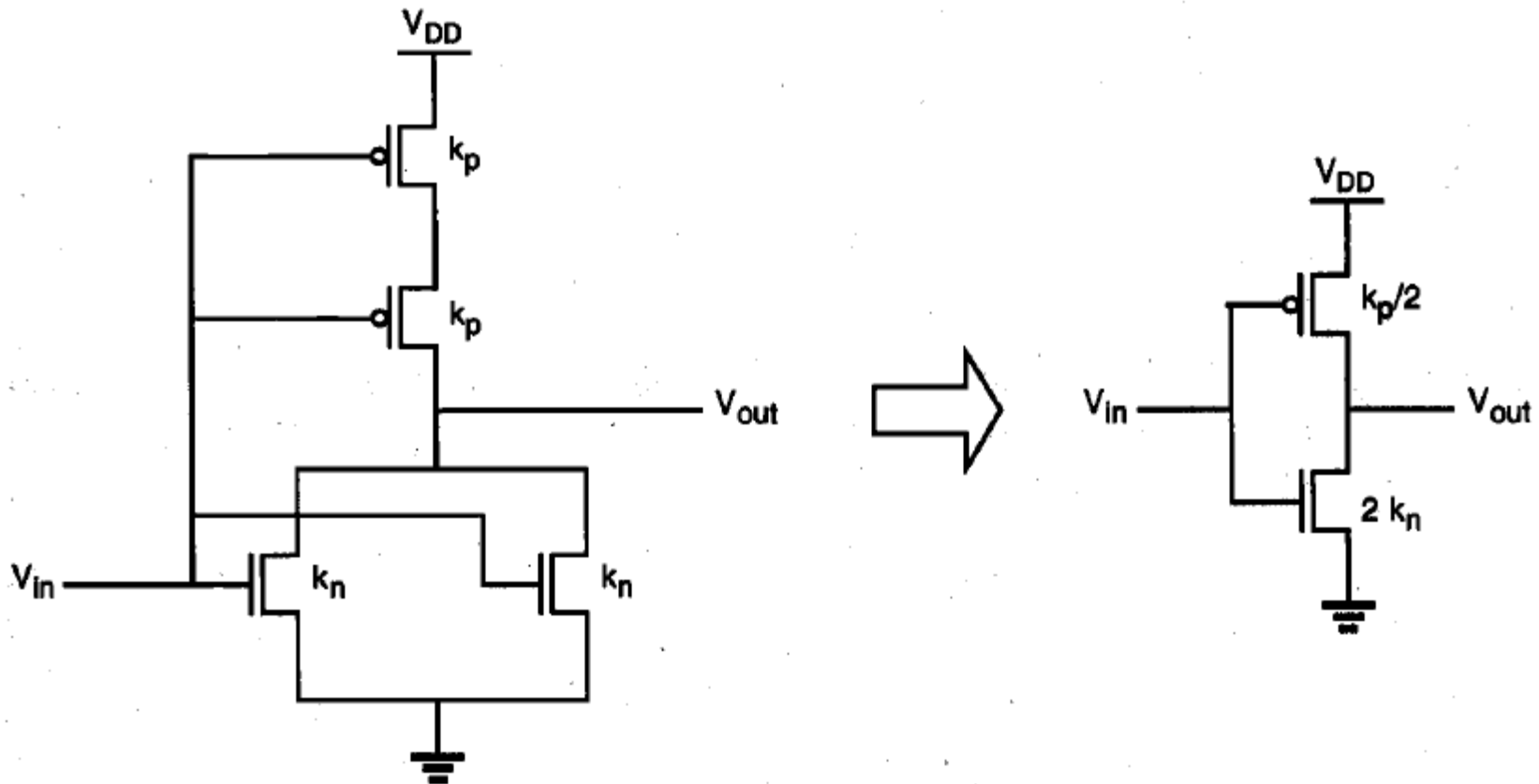


Figure 11. A CMOS NOR2 gate and its inverter equivalent

CMOS Logic Circuits

CMOS NOR2 (Two-Input NOR) Gate

- Using the inverter switching threshold expression for the equivalent inverter circuit, we obtain

$$V_{th}(\text{NOR2}) = \frac{V_{T,n} + \sqrt{\frac{k_p}{4k_n}} (V_{DD} - |V_{T,p}|)}{1 + \sqrt{\frac{k_p}{4k_n}}} \dots(36)$$

CMOS Logic Circuits

CMOS NOR2 (Two-Input NOR) Gate

- From (Eqn. 35), we can easily derive simple design guidelines for the NOR2 gate.
- For example, in order to achieve a switching threshold voltage of $V_{DD}/2$ for simultaneous switching, we have to set $V_{T,n} = |V_{T,p}|$
- and $k_p = 4 k_n$.

CMOS Logic Circuits

CMOS NOR2 (Two-Input NOR) Gate

- Figure 12 shows the CMOS NOR2 gate with the parasitic device capacitances, the inverter equivalent, and the corresponding lumped output load.

CMOS Logic Circuits

CMOS NOR2 (Two-Input NOR) Gate

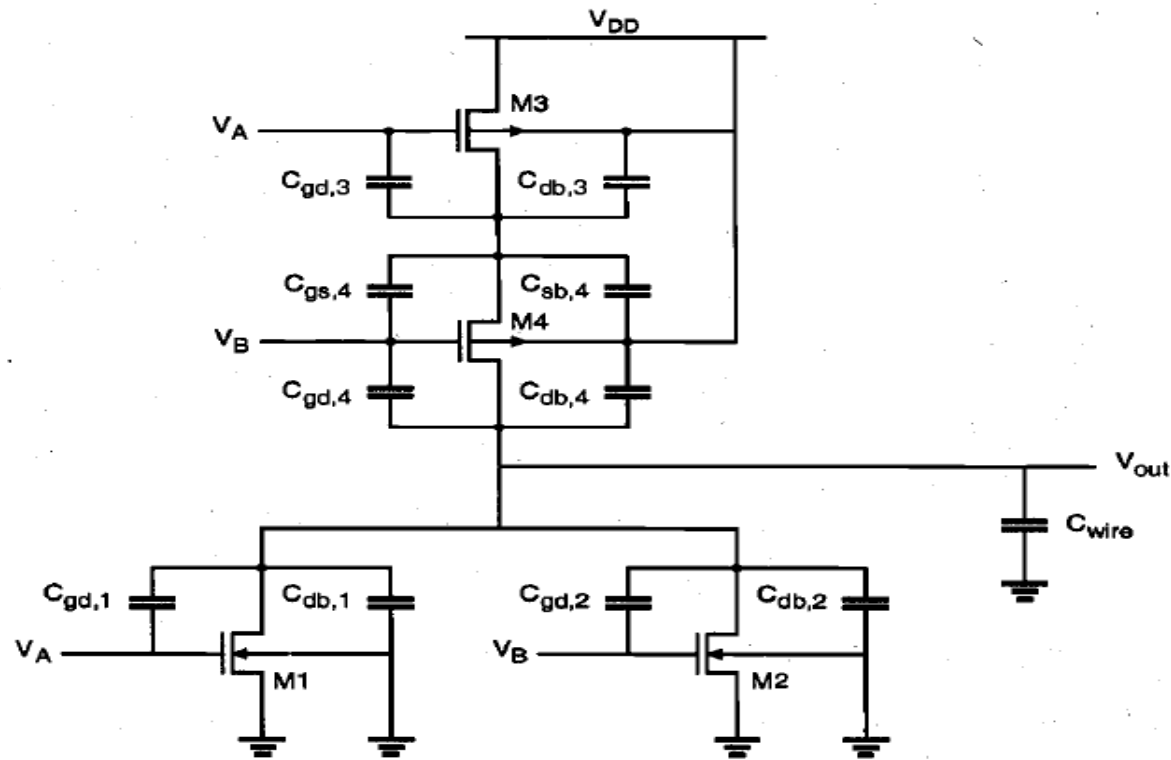


Figure 7.12. Parasitic device capacitances of the CMOS equivalent with the lumped output load capacitance. NOR2 circuit and the simplified

CMOS NAND2 (Two-Input NAND) Gate

- Figure 13 shows a two-input CMOS NAND (NAND2) gate.
- Using an analysis similar to the one developed for the NOR2 gate, we can easily calculate the switching threshold for the CMOS NAND2 gate.

CMOS NAND2 (Two-Input NAND) Gate

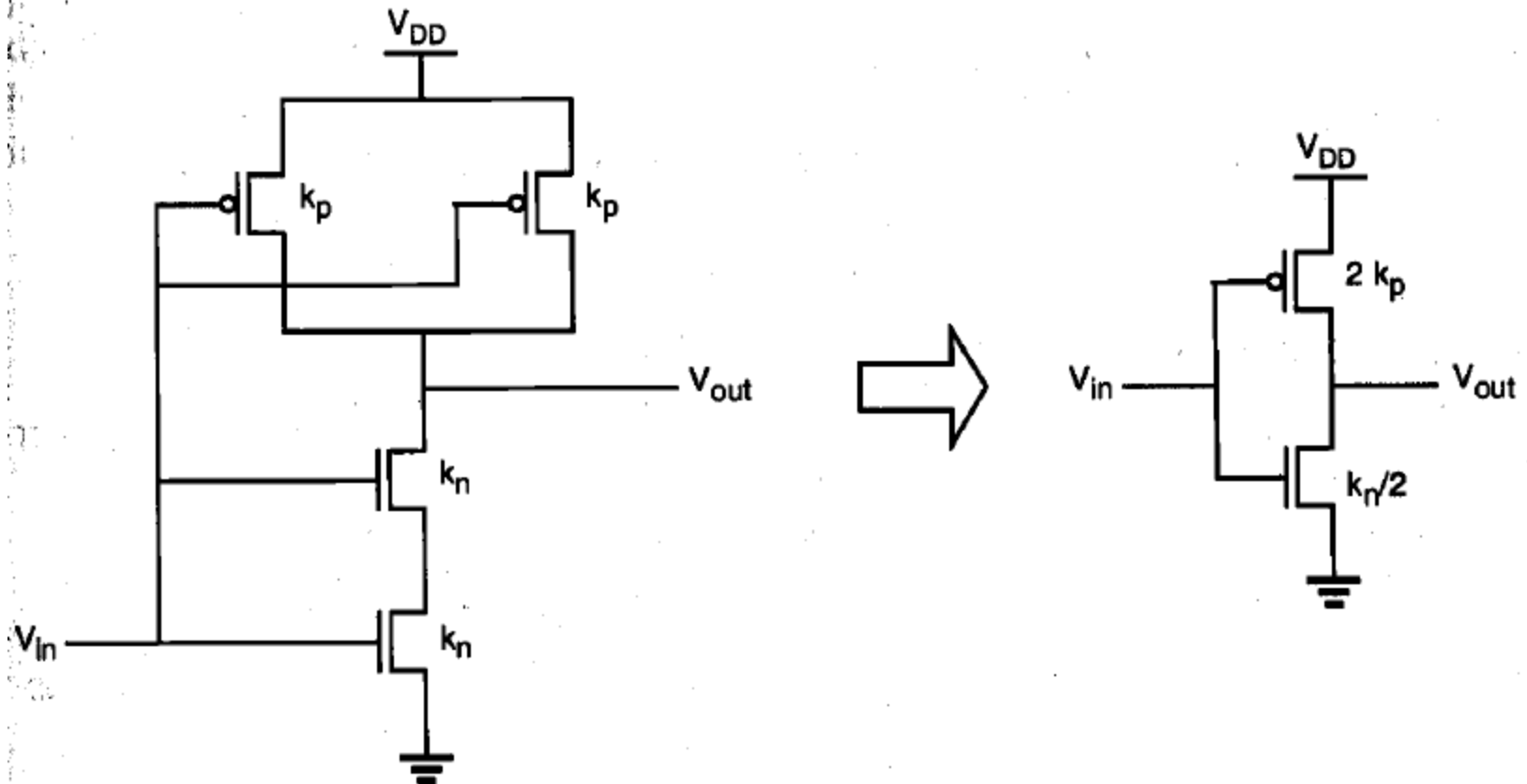


Figure 13. A CMOS NAND2 gate and its inverter equivalent.

CMOS NAND2 (Two-Input NAND) Gate

- The switching threshold for this gate is found as

$$V_{th}(\text{NAND2}) = \frac{V_{T,n} + 2\sqrt{\frac{k_p}{k_n}}(V_{DD} - |V_{T,p}|)}{1 + 2\sqrt{\frac{k_p}{k_n}}} \dots(37)$$

CMOS NAND2 (Two-Input NAND) Gate

- A switching threshold voltage of $V_{DD}/2$ (for simultaneous switching) is achieved by setting $V_{T,n} = |V_{T,p}|$
- and $k_n = 4 k_p$ in the NAND2.

Complex Logic Circuits

- Consider the following Boolean function as an example

$$Z = \overline{A(D + E) + BC} \quad \dots(38)$$

Complex Logic Circuits

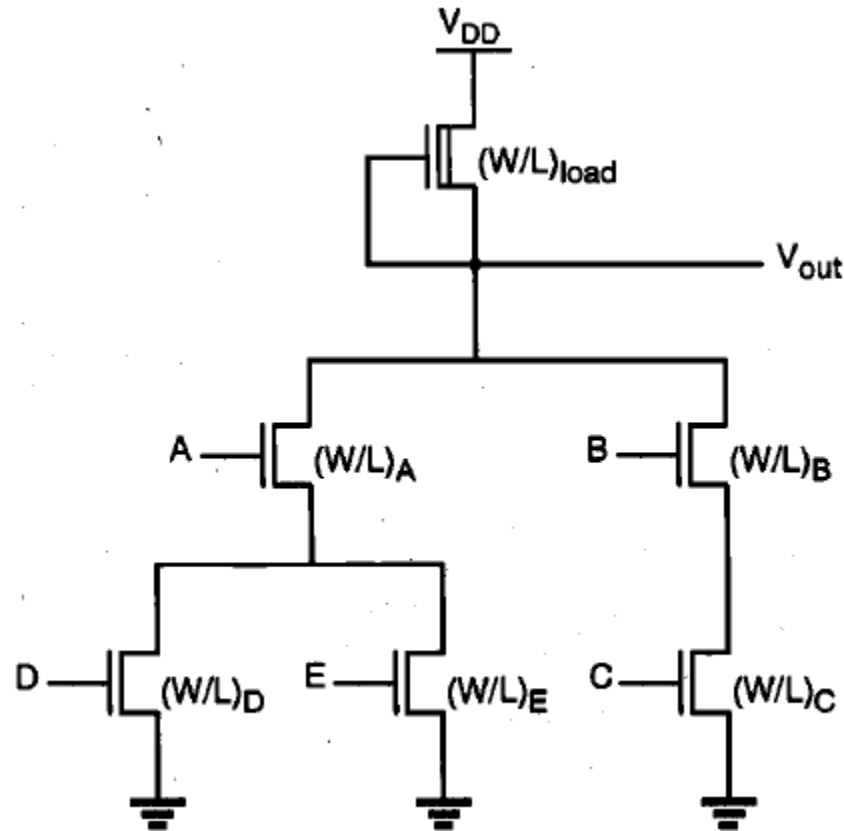
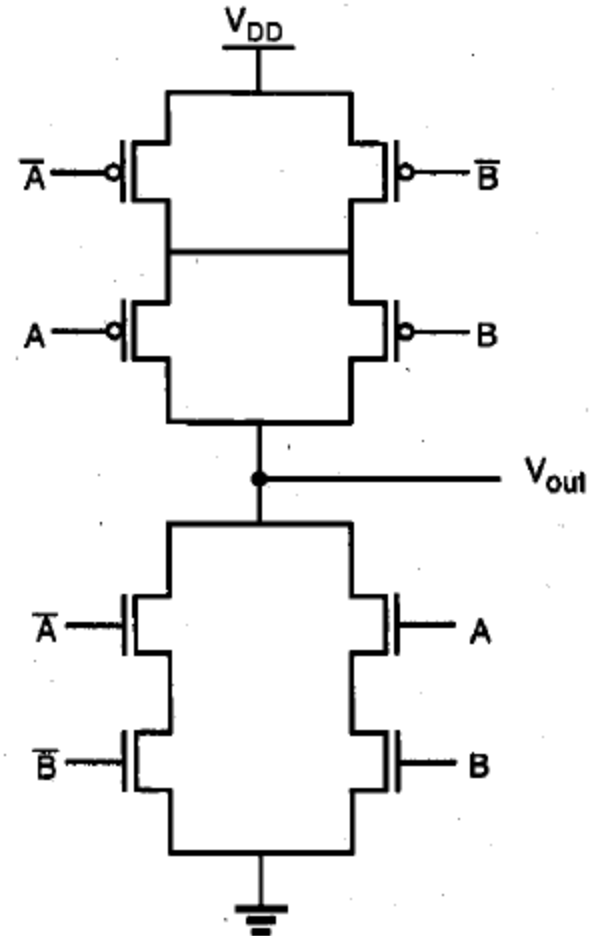
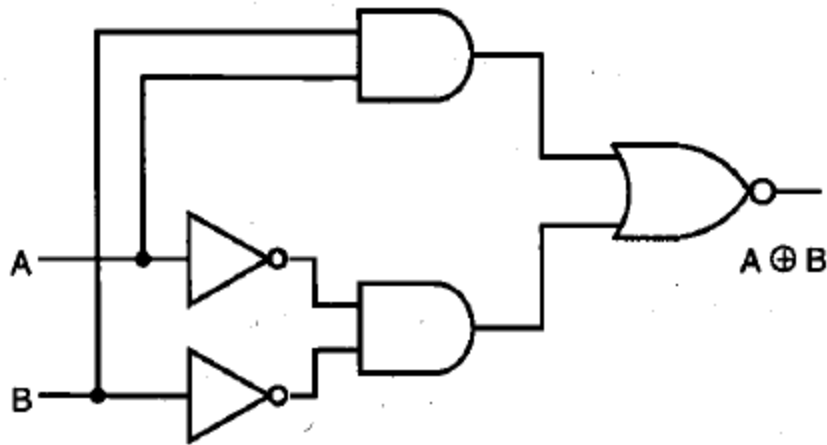
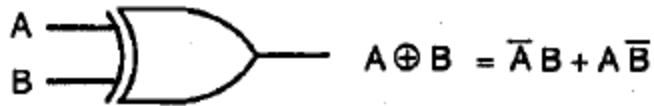


Figure 14. nMOS complex logic gate realizing the Boolean function given in

Complex Logic Circuits

$$\left(\frac{W}{L}\right)_{\text{equivalent}} = \frac{1}{\frac{1}{\left(\frac{W}{L}\right)_B} + \frac{1}{\left(\frac{W}{L}\right)_C}} + \frac{1}{\frac{1}{\left(\frac{W}{L}\right)_A} + \frac{1}{\left(\frac{W}{L}\right)_D + \left(\frac{W}{L}\right)_E}} \quad \dots(39)$$

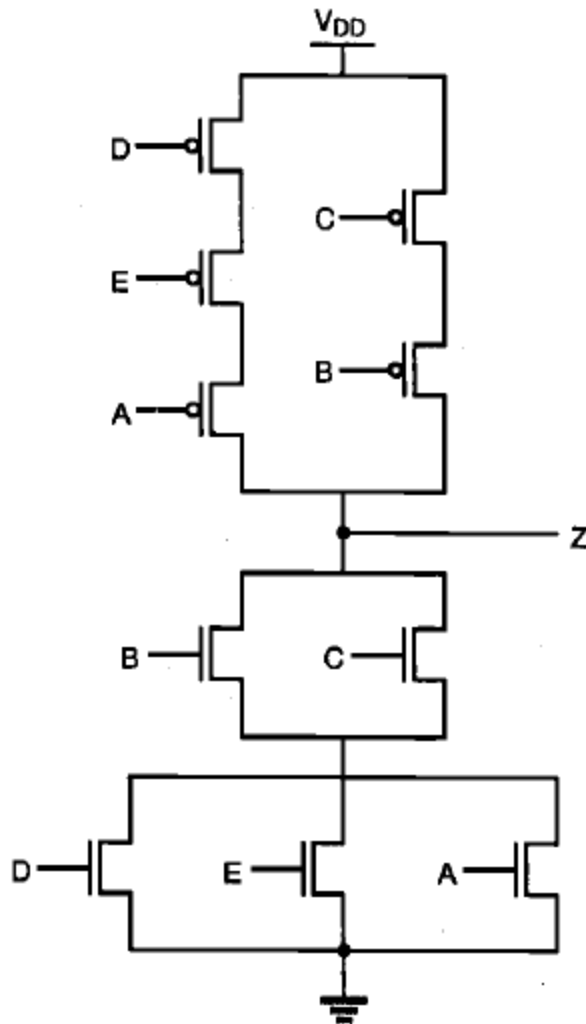
Full-CMOS implementation of the XOR function



Problem 1

- The simplified layout of a CMOS complex logic circuit is given below. Draw the corresponding circuit diagram, and find an equivalent CMOS inverter circuit for simultaneous switching of all inputs, assuming that $(W/L)_p = 15$ for all pMOS transistors and $(W/L)_n = 10$ for all nMOS transistors.

Solution



$$Z = \overline{(D + E + A)(B + C)}$$

Solution

$$\begin{aligned}\left(\frac{W}{L}\right)_{n,eq} &= \frac{1}{\frac{1}{\left(\frac{W}{L}\right)_D + \left(\frac{W}{L}\right)_E + \left(\frac{W}{L}\right)_A} + \frac{1}{\left(\frac{W}{L}\right)_B + \left(\frac{W}{L}\right)_C}} \\ &= \frac{1}{\frac{1}{30} + \frac{1}{20}} = 12\end{aligned}$$

$$\begin{aligned}\left(\frac{W}{L}\right)_{p,eq} &= \frac{1}{\frac{1}{\left(\frac{W}{L}\right)_D} + \frac{1}{\left(\frac{W}{L}\right)_E} + \frac{1}{\left(\frac{W}{L}\right)_A}} + \frac{1}{\frac{1}{\left(\frac{W}{L}\right)_B} + \frac{1}{\left(\frac{W}{L}\right)_C}} \\ &= \frac{1}{\frac{1}{15} + \frac{1}{15} + \frac{1}{15}} + \frac{1}{\frac{1}{15} + \frac{1}{15}} = 12.5\end{aligned}$$

AOI and OAI Gates

- Two important circuit categories as subsets of the general complex CMOS gate topology is represented.
- These are the AND-OR-INVERT (AOI) gates and the OR-AND-INVERT (OAI) gates.
- The AOI gate, as its name suggests, enables the sum-of-products realization of a Boolean function in one logic stage (Fig. 15).
- The pull-down net of the AOI gate consists of parallel branches of series-connected nMOS driver transistors. The corresponding p-type pull-up network can simply be found using the dual-graph concept.

AOI Gate

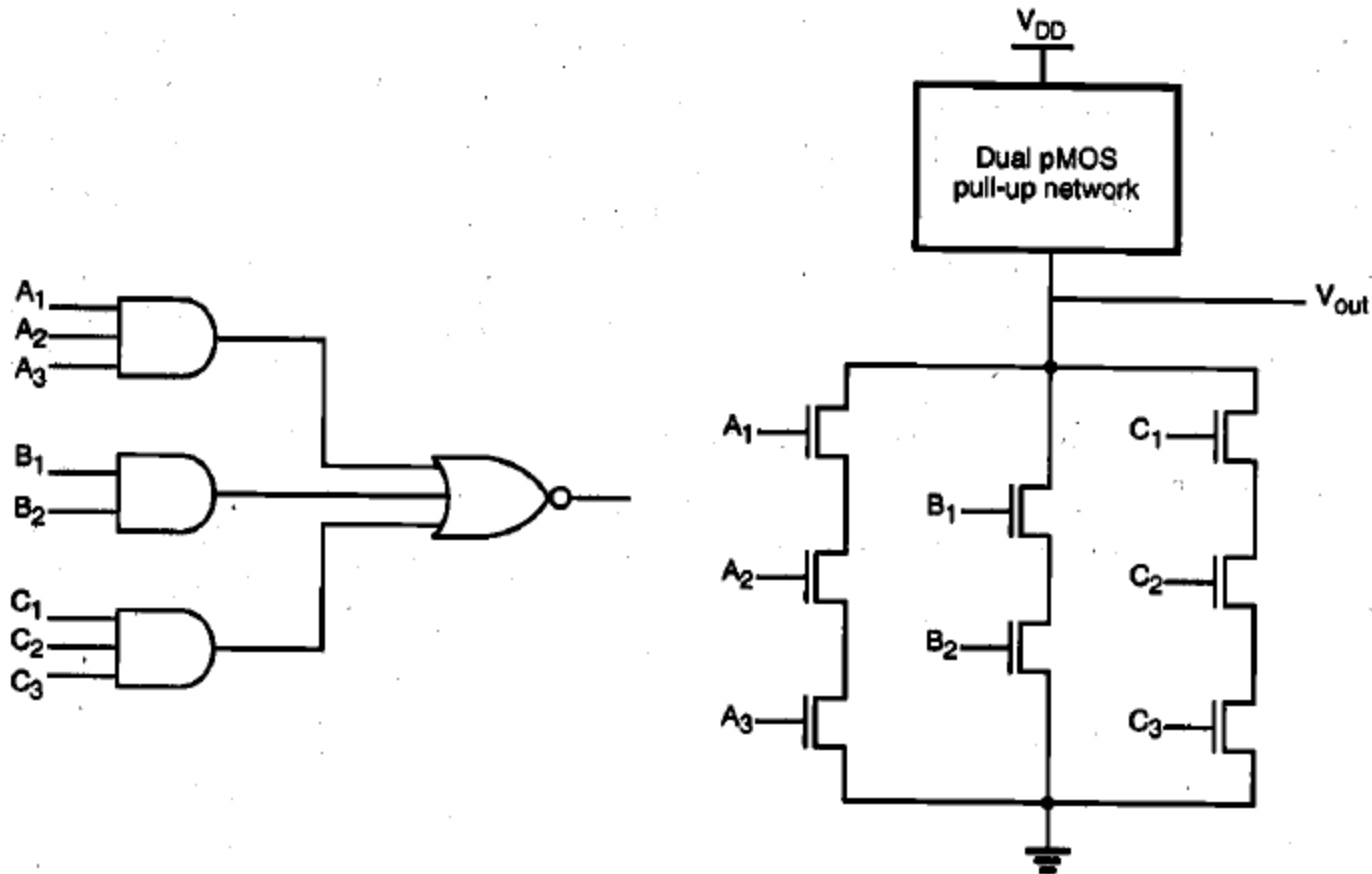


Figure 15. An AND-OR-INVERT (AOI) gate and the corresponding pull-down net.

OAI Gate

- The OAI gate, on the other hand, enables the product-of-sums realization of a Boolean function in one logic stage (Fig. 16).
- The pull-down net of the OAI gate consists of series branches of parallel-connected nMOS driver transistors, while the corresponding p-type pull-up network can be found using the dual-graph concept.

OAI Gate

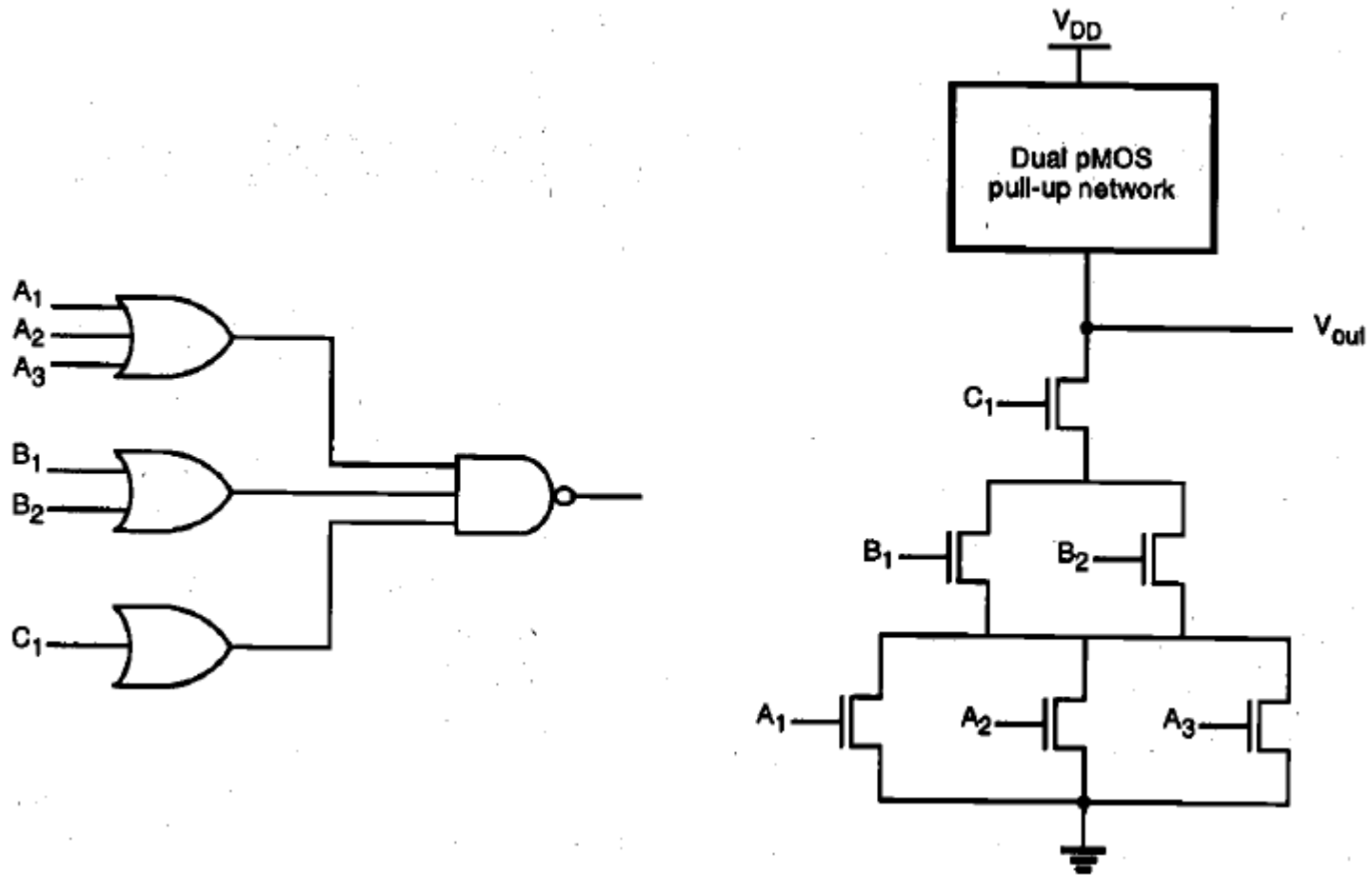


Figure 16. An OR-AND-INVERT (OAI) gate, and the corresponding pull-down net.

CMOS Transmission Gates (Pass Gates)

- Fig. 17 represents the CMOS transmission gate consists of one nMOS and one pMOS transistor, connected in parallel.
- The gate voltages applied to these two transistors are also set to be complementary signals.
- As such, the CMOS TG operates as a bidirectional switch between the nodes A and B which is controlled by signal C.

CMOS Transmission Gates (Pass Gates)

- If the control signal C is logic-high, i.e., equal to V_{DD} , then both transistors are turned on and provide a low-resistance current path between the nodes A and B .
- If, on the other hand, the control signal C is low, then both transistors will be off, and the path between the nodes A and B will be an open circuit. This condition is also called the high-impedance state.

CMOS Transmission Gates (Pass Gates)

- The substrate terminal of the nMOS transistor is connected to ground and the substrate terminal of the pMOS transistor is connected to V_{DD} .

CMOS Transmission Gates (Pass Gates)

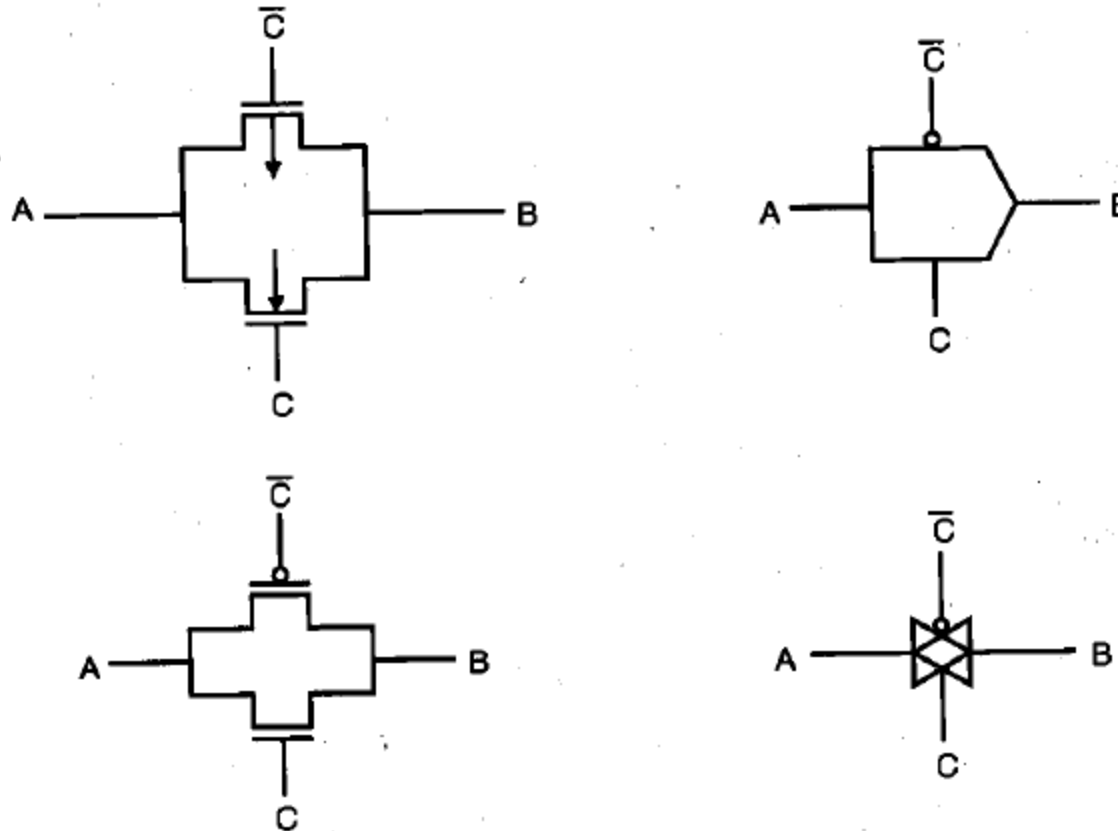


Figure 17. Four different representations of the CMOS transmission gate (TG).

DC analysis of the CMOS transmission gate

- Let the input node (A) is connected to a constant logic-high voltage, $V_{in} = V_{DD}$ as shown in Fig. 18.
- The control signal is also logic-high, thus ensuring that both transistors are turned on.
- The output node (B) may be connected to a capacitor, which represents capacitive loading of the subsequent logic stages driven by the transmission gate.

DC analysis of the CMOS transmission gate

- We will now investigate the input-output current-voltage relationship of the CMOS TG as a function of the output voltage V_{out} .
- It can be seen from Fig. 18 that the drain-to-source and the gate-to-source voltages of the nMOS transistor are

$$\begin{aligned} V_{DS,n} &= V_{DD} - V_{out} \\ V_{GS,n} &= V_{DD} - V_{out} \end{aligned} \quad \dots(40)$$

DC analysis of the CMOS transmission gate

- Thus, the nMOS transistor will be turned off for $V_{out} > V_{DD} - V_{T,n}$ and will operate in the saturation mode for $V_{out} < V_{DD} - V_{T,n}$.
- The V_{DS} and V_{GS} voltages of the pMOS transistor are

$$\begin{aligned} V_{DS,p} &= V_{out} - V_{DD} \\ V_{GS,p} &= -V_{DD} \end{aligned} \quad \dots(41)$$

DC analysis of the CMOS transmission gate

- Consequently, the pMOS transistor is in saturation for $V_{out} < IV_{T,p}$, and it operates in the linear region for $V_{out} > IV_{T,p}$.
- The total current flowing through the transmission gate is the sum of the nMOS drain current and the pMOS drain current.

$$I_D = I_{DS,n} + I_{SD,p} \quad \dots(42)$$

DC analysis of the CMOS

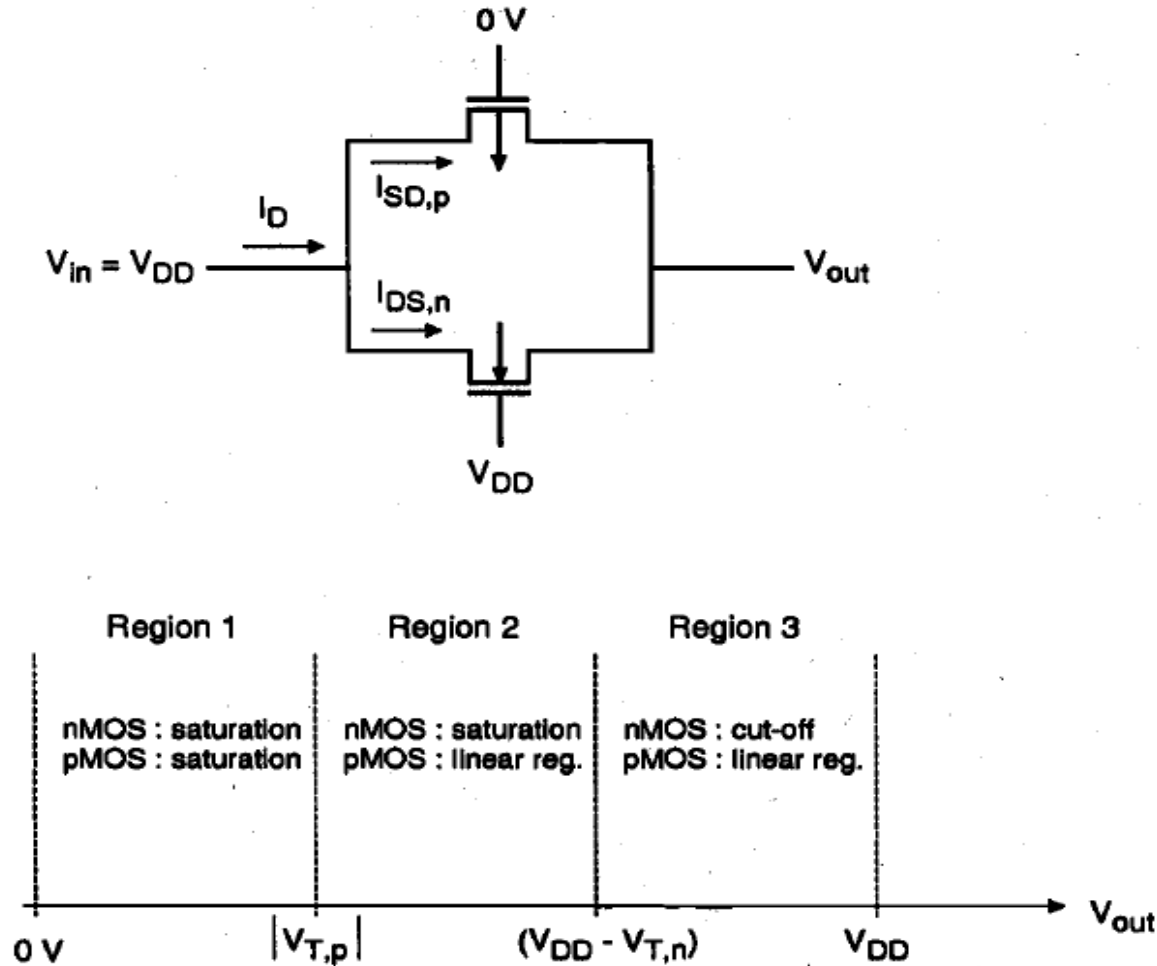


Figure 18. Bias conditions and operating regions of the CMOS transmission gate, shown as functions of the output voltage.

Equivalent Resistance

- The equivalent resistance for each transistor in this structure, can be represented as follows.

$$R_{eq,n} = \frac{V_{DD} - V_{out}}{I_{DS,n}} \quad \dots(43)$$
$$R_{eq,p} = \frac{V_{DD} - V_{out}}{I_{SD,p}}$$

Equivalent Resistance

- Now, we will calculate the equivalent resistance values for the three operating regions of the transmission gate.
- **Region 1**
- Here, the output voltage is smaller than the absolute value of the pMOS transistor threshold voltage, i.e., $V_{out} < |V_{T,p}|$.

Equivalent Resistance

- According to Fig. 18, both transistors are in saturation.
- We obtain the equivalent resistance of both devices as

$$R_{eq,n} = \frac{2(V_{DD} - V_{out})}{k_n (V_{DD} - V_{out} - V_{T,n})^2} \quad \dots(44)$$

$$R_{eq,p} = \frac{2(V_{DD} - V_{out})}{k_p (V_{DD} - |V_{T,p}|)^2} \quad \dots(45)$$

Equivalent Resistance

- **Region 2**

- In this region, $|V_{T,p}| < V_{out} < (V_{DD} - V_{T,n})$.
- Thus, the pMOS transistor now operates in the linear region, while the nMOS transistor continues to operate in saturation.

$$R_{eq,n} = \frac{2(V_{DD} - V_{out})}{k_n (V_{DD} - V_{out} - V_{T,n})^2} \dots(46)$$

Equivalent Resistance

$$\begin{aligned} R_{eq,p} &= \frac{2(V_{DD} - V_{out})}{k_p \left[2(V_{DD} - |V_{T,p}|)(V_{DD} - V_{out}) - (V_{DD} - V_{out})^2 \right]} \\ &= \frac{2}{k_p \left[2(V_{DD} - |V_{T,p}|) - (V_{DD} - V_{out}) \right]} \end{aligned} \quad \dots(47)$$

Equivalent Resistance

- **Region 3**
- Here, the output voltage is $V_{out} > (V_{DD} - V_{T,n})$.
- Consequently, the nMOS transistor will be turned off, which results in an open-circuit equivalent.
- The pMOS transistor will continue to operate in the linear region.

$$R_{eq,p} = \frac{2}{k_p \left[2(V_{DD} - |V_{T,p}|) - (V_{DD} - V_{out}) \right]} \quad \dots(48)$$

Equivalent Resistance

- Combining the equivalent resistance values found for the three operating regions, we can now plot the total resistance of the CMOS transmission gate as a function of the output voltage V_{out} as shown in Fig. 19.

Equivalent Resistance

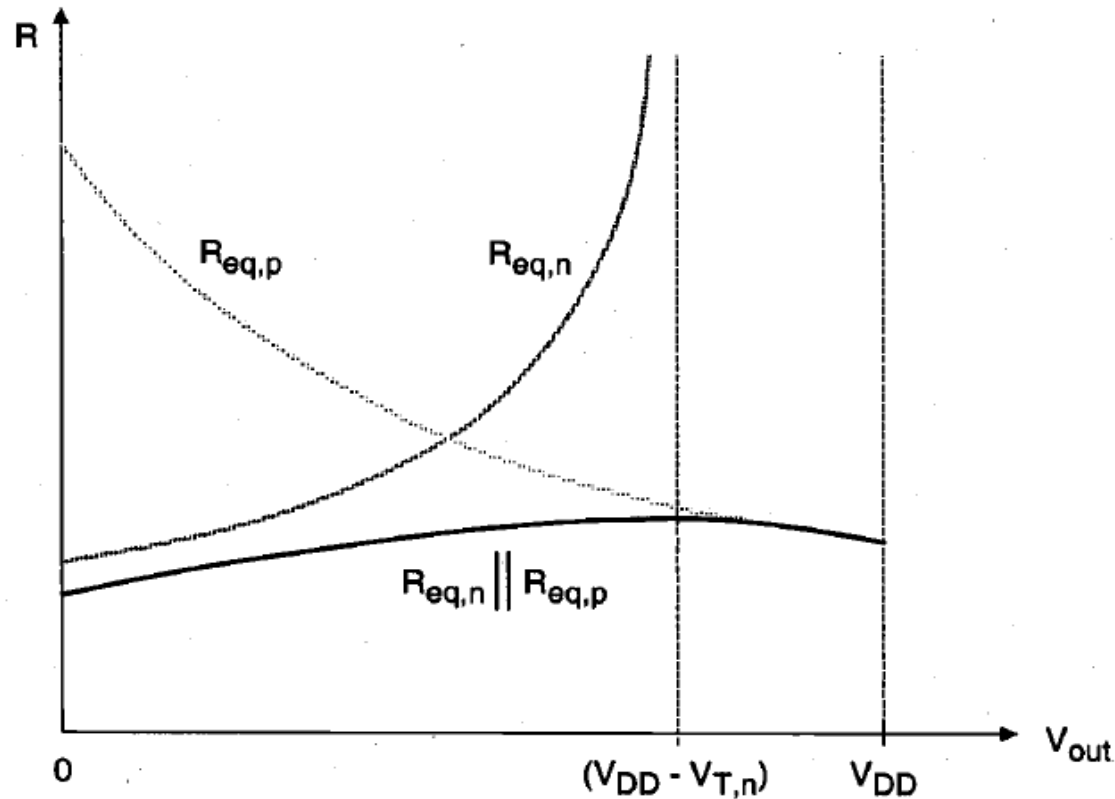


Figure 19. Equivalent resistance of the CMOS transmission gate, plotted as a function of the output voltage.

Equivalent Resistance

- It can be seen that the total equivalent resistance of the TG remains relatively constant, i.e., its value is almost independent of the output voltage, whereas the individual equivalent resistances of both the nMOS and the pMOS transistors are strongly dependent on V_{out} .

Equivalent Resistance

- This property of the CMOS TG is naturally quite desirable.
- A CMOS pass gate which is turned on by a logic-high control signal can be replaced by its simple equivalent resistance for dynamic analysis, as shown in Fig. 20.

Equivalent Resistance

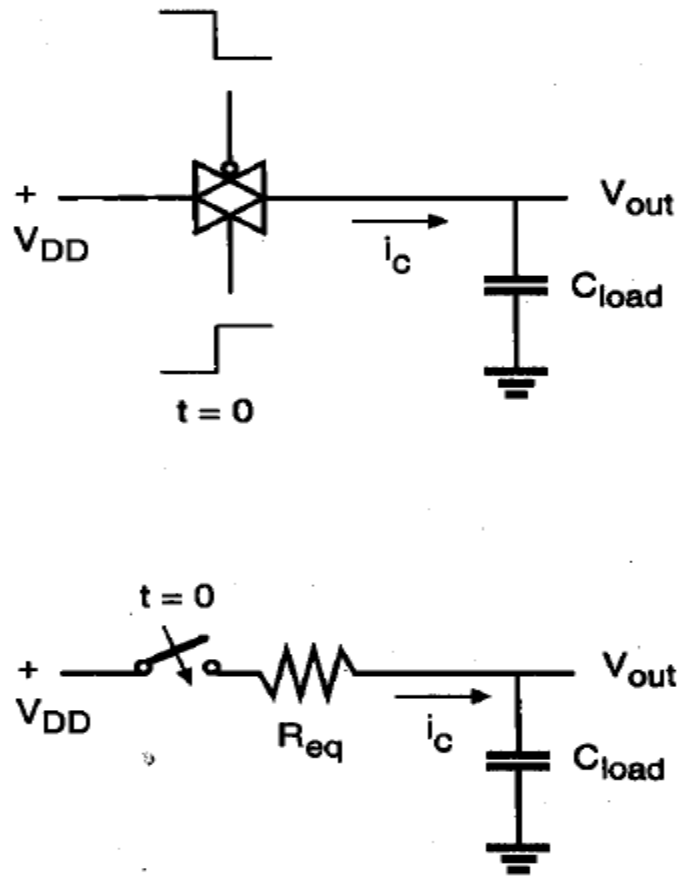


Figure 20. Replacing the CMOS TG with its resistor equivalent for transient analysis.

CMOS Transmission Gates (Pass Gates)

- The implementation of CMOS transmission gates in logic circuit design usually results in compact circuit structures which may even require a smaller number of transistors than their standard CMOS counterparts.
- The control signal and its complement must be available simultaneously for TG applications.

CMOS Transmission Gates (Pass Gates)

- Figure 21 shows a two-input multiplexor circuit consisting of two CMOS transmission gates.
- The operation of the multiplexor can be understood quite easily: If the control input S is logic-high, then the bottom TG will conduct, and the output will be equal to the input B .
- If the control signal is low, the bottom TG will turn off and the top TG will connect the input A to the output node.

CMOS Transmission Gates (Pass Gates)

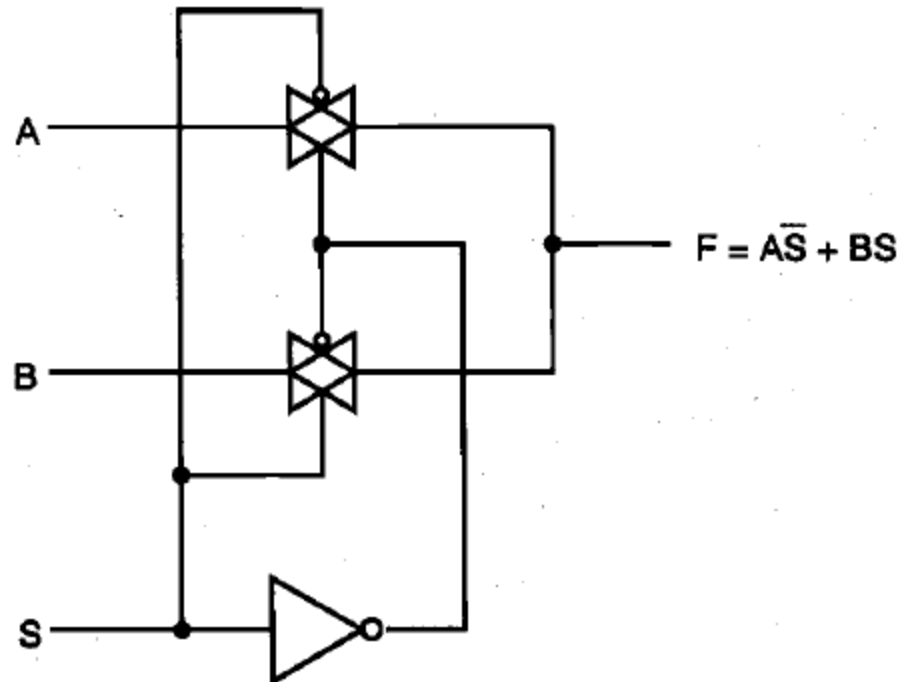


Figure 21. Two-input multiplexor circuit implemented using two CMOS TGs.

CMOS Transmission Gates (Pass Gates)

- Figure 22 shows an eight-transistor implementation of the logic XOR function, using two CMOS TGs and two CMOS inverters.
- The same function can also be implemented using only six transistors, as shown in Fig. 23.

CMOS Transmission Gates (Pass Gates)

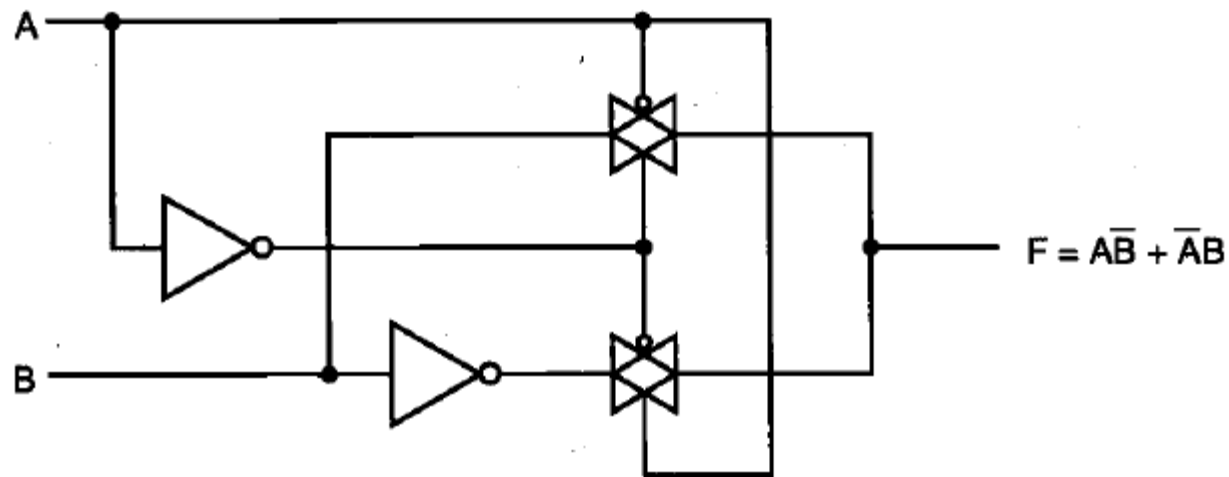


Figure 22. Eight-transistor CMOS TG implementation of the XOR function.

CMOS Transmission Gates (Pass Gates)

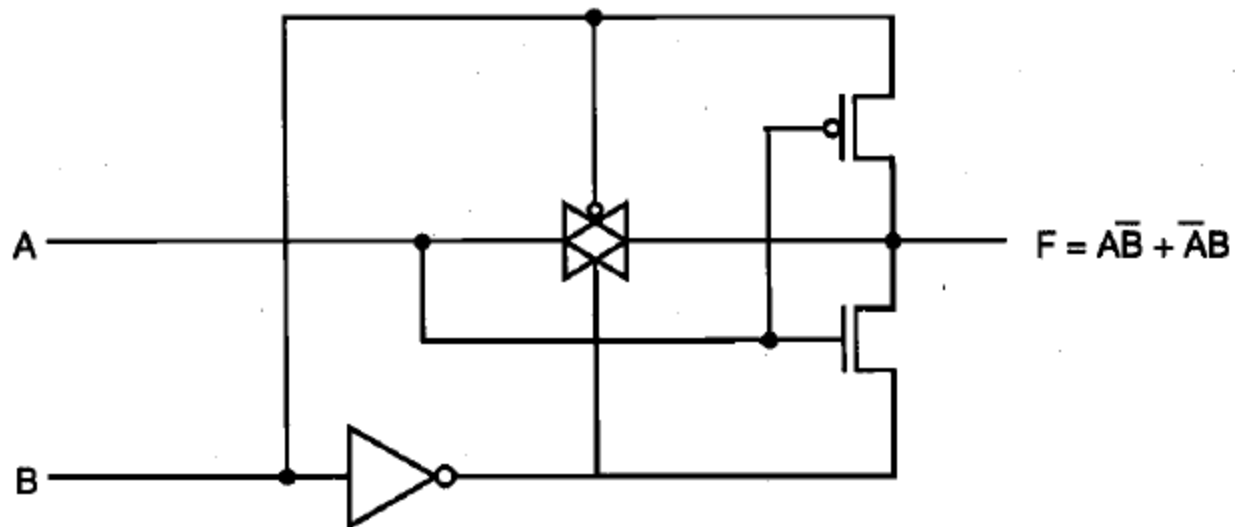


Figure 23. Six-transistor CMOS TG implementation of the XOR function

CMOS Transmission Gates (Pass Gates)

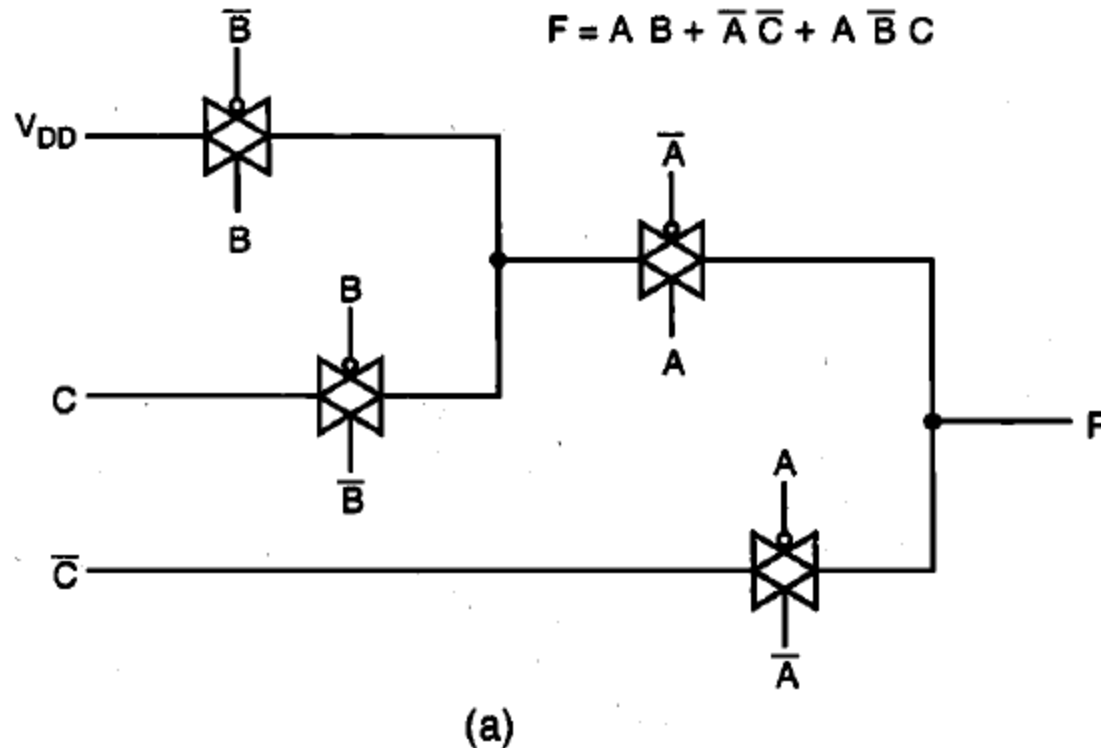
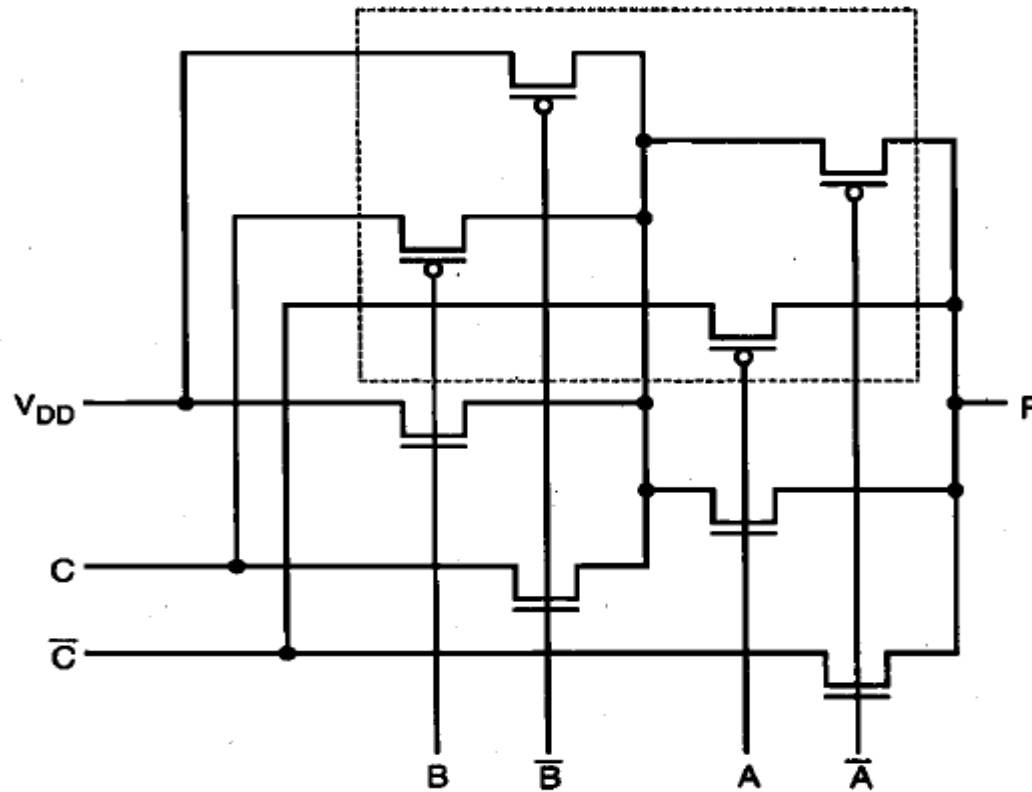


Figure 24(a) CMOS TG realization of a three-variable Boolean function. (b) All pMOS transistors can be placed into one n-well to save area.

CMOS Transmission Gates (Pass Gates)



(b)

Figure 24(a) CMOS TG realization of a three-variable Boolean function. (b) All pMOS transistors can be placed into one n-well to save area.

Complementary Pass-Transistor Logic (CPL)

- The complexity of full-CMOS pass-gate logic circuits can be reduced dramatically by adopting another circuit concept, called Complementary Pass-transistor Logic (CPL).
- The main idea behind CPL is to use a purely nMOS pass-transistor network for the logic operations, instead of a CMOS TG network.

Complementary Pass-Transistor Logic (CPL)

- All inputs are applied in complementary form, i.e., every input signal and its inverse must be provided; the circuit also produces complementary outputs, to be used by subsequent CPL stages.
- Thus, the CPL circuit essentially consists of complementary inputs, an nMOS pass transistor logic network to generate complementary outputs, and CMOS output inverters to restore the output signals.
- The circuit diagrams of a CPL NOR2 and a CPL NAND2 are shown in Fig. 25.

Complementary Pass-Transistor Logic (CPL)

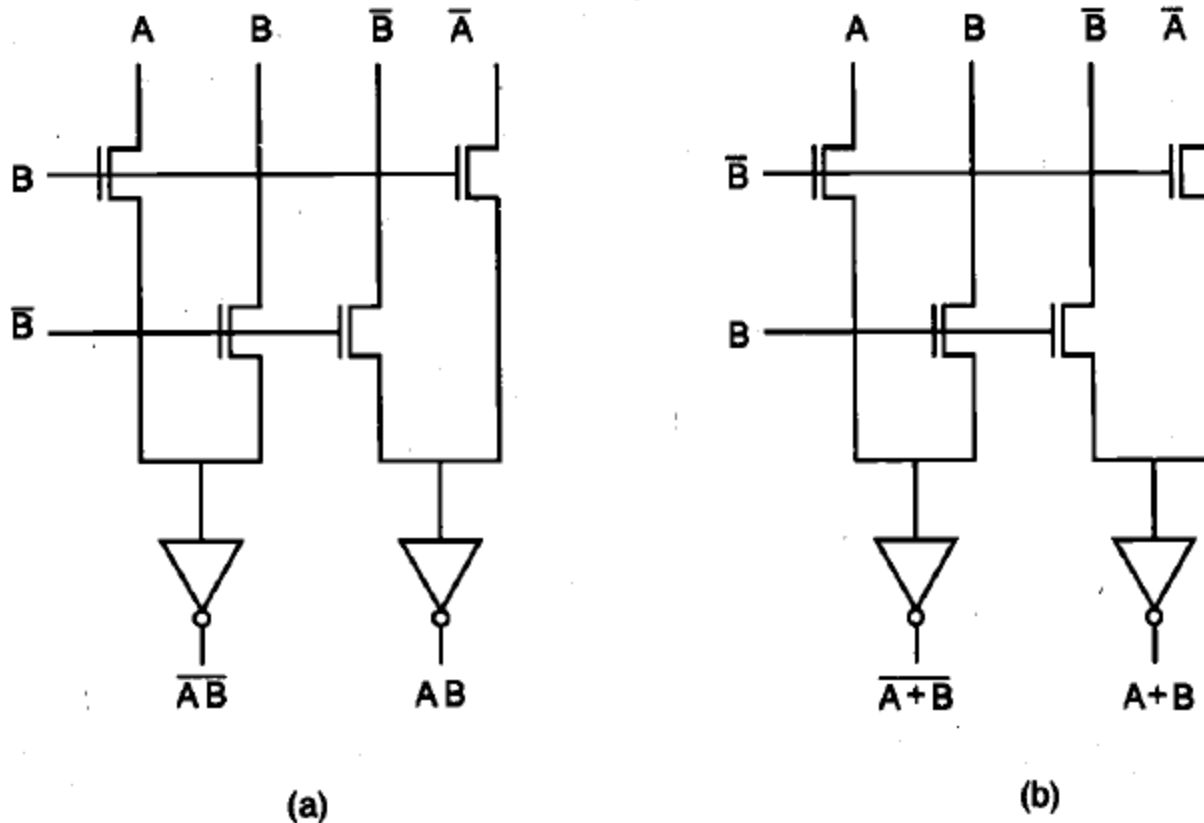


Figure 25. Circuit diagram of (a) CPL NAND2 gate and (b) CPL NOR2 gate.

Complementary Pass-Transistor Logic (CPL)

- Regarding the transistor count, CPL circuits do not always offer a marked advantage over conventional CMOS.
- The NAND2 and NOR2 circuits shown in Fig. 25 each consist of 8 transistors.

DESIGN FOR TESTABILITY

Introduction

- The task of determining whether fabricated chips are fully functional is highly complex and can be very time-consuming.
- However, when faulty chips pass an improperly designed test, they can cause system failures and enormous difficulty in system debugging.
- It is known that the debugging cost increases by about tenfold from chip level to board level, and also from board level to system level.

Introduction

- Thus, it is of great importance to detect faults as early as possible.
- As the number of transistors integrated into a single chip increases, the task of chip testing to ensure correct functionality becomes increasingly more difficult.
- However, in a production environment, many chips must be tested within a short time for timely delivery to customers.

Introduction

- To overcome such difficult issues, design for testability has become ever more critical.
- Here, we discuss types of faults, the corresponding fault models, design of testable circuits, and self-testing circuits.

Fault Types and Models

- Chip testing, in the conventional sense, is usually multi-purpose and attempts to detect faults in fabrication, design, and failures due to stressful operating conditions, namely the reliability problems.
- Input test vectors are devised and applied to the device under test (DUT) or circuit under test (CUT) as its stimuli.
- Then the measured outputs are compared with the expected correct responses to determine whether DUT is good (go) or bad (no-go).

Examples of physical defects

- Defects in silicon substrate
- Photolithographic defects
- Mask contamination and scratches
- Process variations and abnormalities
- Oxide defects

Electrical Faults

- The physical defects can cause electrical faults;
- Shorts (bridging faults)
- Opens
- Transistor stuck-on, stuck-open
- Resistive shorts and opens
- Excessive change in threshold voltage
- Excessive steady-state currents

Logical Faults

- The electrical faults in turn can be translated into logical faults. The logical faults include:
- Logical stuck-at-0 or stuck-at-1
- Slower transition (delay fault)
- AND-bridging, OR-bridging

Fault Types and Models

- The relationships between physical defects, electrical faults, and logical faults can be explained using a simple NOR2 gate as shown in Fig. 1.
- A metallic blob (physical defect) between the common drain terminal in the n-diffusion region and the ground bus line shown in Fig. 1 (a) can be modeled as a resistive short between the output node Z and the ground as

Fault Types and Models

- shown in Fig. 1(b), and also by a stuck-at-0 (s-a-0) fault of output Z when the resistance is low or a pull-up delay fault when the resistance is high, as shown in Fig. 1 (c).

Fault Types and Models

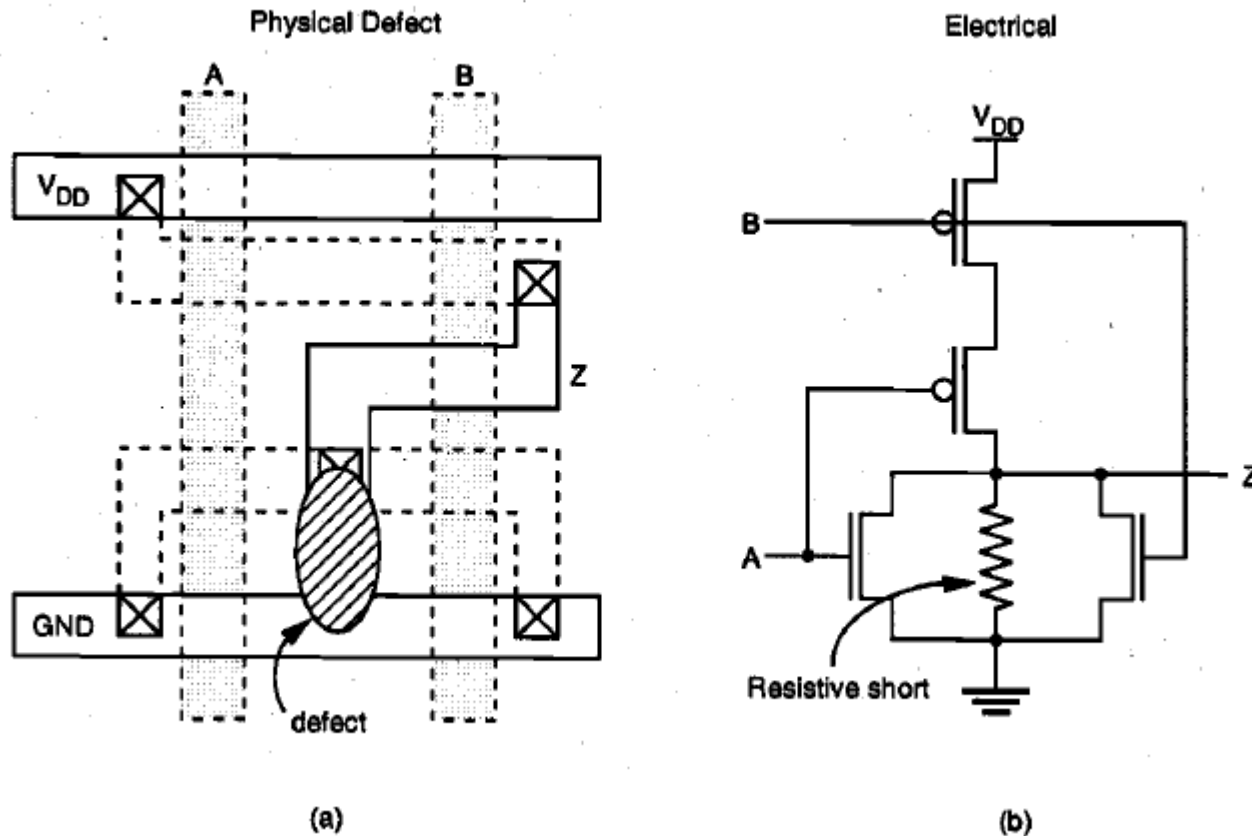


Figure 1. (a) Physical defect in NOR2 fabrication, (b) its electrical fault model; and (c) its logical fault models.

Fault Types and Models

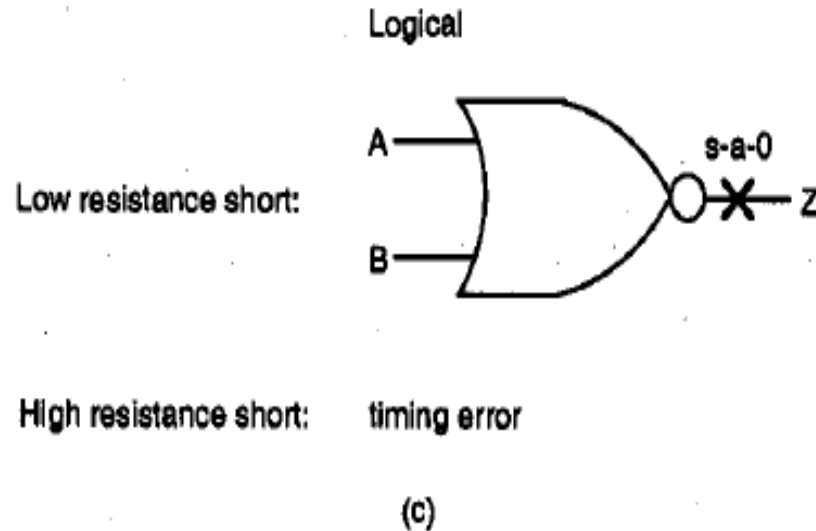


Figure 1. (a) Physical defect in NOR2 fabrication, (b) its electrical fault model; and (c) its logical fault models.

Controllability and Observability

- The controllability of a circuit is a measure of the ease (or difficulty) with which the controller (test engineer) can establish a specific signal value at each node by setting values at the circuit input terminals.
- The observability is a measure of the ease (or difficulty) with which one can determine the signal value at any logic node in the circuit by controlling its primary input and observing the primary output.

Controllability and Observability

- Here the term primary refers to the I/O boundary of the circuit under test.
- The degree of controllability and observability and, thus, the degree of testability of a circuit, can be measured with respect to whether test vectors are generated deterministically or randomly.

Controllability and Observability

- There are deterministic procedures for test generation for combinational circuits, such as the D-algorithm which uses a recursive search procedure advancing one gate at a time and backtracking, if necessary, until all the faults are detected.
- The D-algorithm requires a large amount of computer time.

Controllability and Observability

- To overcome such shortcomings, many improved algorithms such as Path-Oriented DEcision Making (PODEM) and FAN-out-oriented test generation (FAN) have been introduced, Sequential circuit test generation is several orders of magnitude more difficult than these algorithms.
- To ease the task of ATG, design-for-test (DFT) techniques are routinely employed.

Controllability and Observability

- Let us now consider the simple circuit in Fig. 2 consisting of four simple logic gates.
- To detect any defect on line 8, the primary inputs A and B must be set to logic 1.
- However, such a setting forces line 7 to logic 1. Thus, any stuck-at-1(s-a-1) fault on line 7 cannot be tested at the primary output, although in the absence of such a fault, the logic value on line 7 can be fully controllable through primary inputs B, C, and D.

Controllability and Observability

- Therefore, this circuit is not fully testable.
- The main cause of this difficulty in this circuit is the fact that input B fans out to lines 5 and 6, and then after the OR3 gate, both line signals are combined in the AND3 gate.
- Such a fanout is called reconvergent fanout.
- Reconvergent fanouts make the testing of the circuit much more difficult.

Controllability and Observability

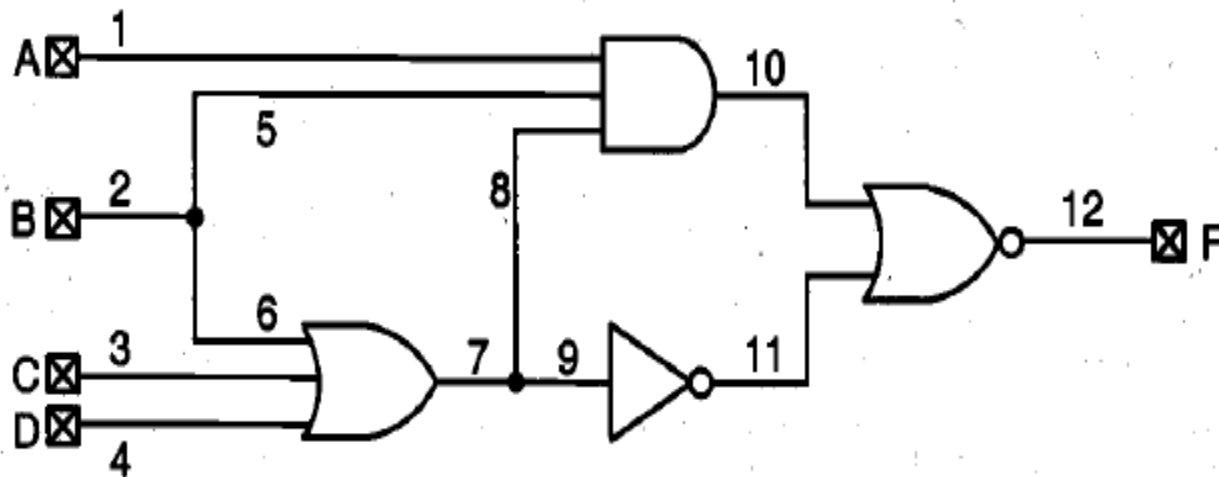


Figure 2. A simple circuit consisting of four gates with four primary inputs and one primary output.

Controllability and Observability

- If a large number of input vectors are required to set a particular node value to 1 or 0 (fault excitation) and to propagate an error at the node to an output (fault effect propagation), then the testability is low.
- The circuits with poor controllability include those with feedbacks, decoders, and clock generators.

Controllability and Observability

- The circuits with poor observability include sequential circuits with long feedback loops and circuits with reconvergent fanouts, redundant nodes, and embedded memories such as RAM, ROM, and PLA.

Ad Hoc Testable Design Techniques

- One way to increase the testability is to make nodes more accessible at some cost by physically inserting more access circuits to the original design.
- Some of the ad hoc testable design techniques are explained here.

Partition-and-Mux Technique

- Since the sequence of many serial gates, functional blocks, or large circuits are difficult to test, such circuits can be partitioned and multiplexors (muxes) can be inserted such that some of the primary inputs can be fed to partitioned parts through multiplexers with accessible control signals.
- With this design technique, the number of accessible nodes can be increased and the number of test patterns can be reduced.

Partition-and-Mux Technique

- A case in point would be the 32-bit counter.
- Dividing this counter into two 16-bit parts would reduce the testing time in principle by a factor of 2^{15} .
- However, circuit partitioning and addition of multiplexers may increase the chip area and circuit delay.
- This practice is not unique and is similar to the divide-and-conquer approach to large, complex problems (Fig. 3).

Partition-and-Mux Technique

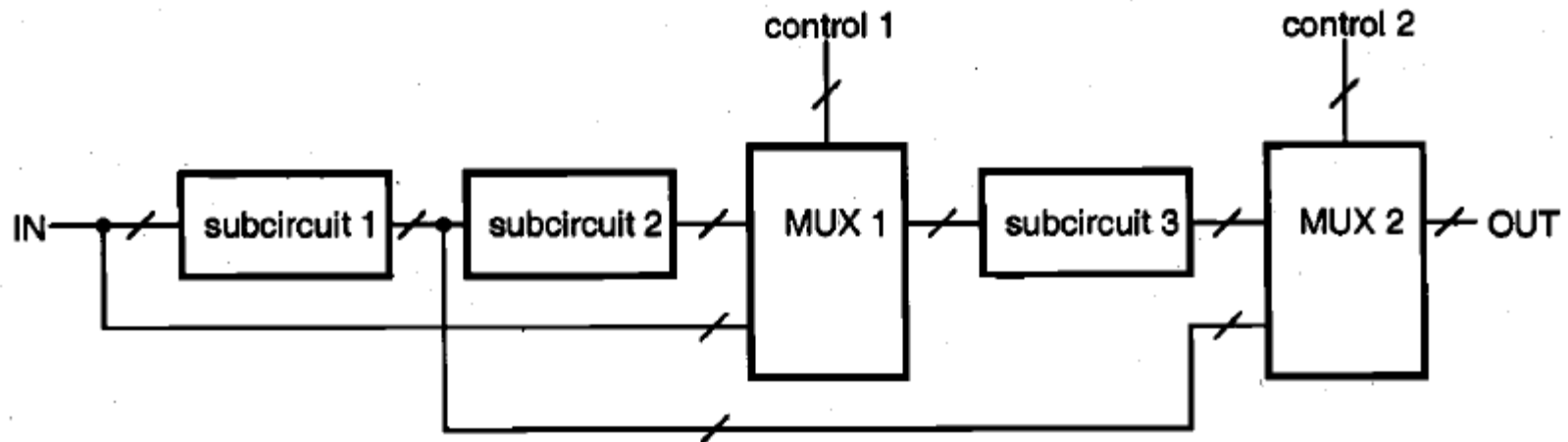


Figure 3. Partition-and-mux method for large circuits.

Initialize Sequential Circuit

- When the sequential circuit is powered up, its initial state can be a random, unknown state.
- In this case, it is not possible to start the test sequence correctly.
- The state of a sequential circuit can be brought to a known state through initialization.
- In many designs, the initialization can be easily done by connecting asynchronous preset or clear-input signals from primary or controllable inputs to flip-flops or latches.

Disable Internal Oscillators and Clocks

- To avoid synchronization problems during testing, internal oscillators and clocks should be disabled.
- For example, rather than connecting the circuit directly to the on-chip oscillator, the clock signal can be ORed with a disabling signal followed by an insertion of a testing signal as shown in Fig. 4.

Disable Internal Oscillators and Clocks

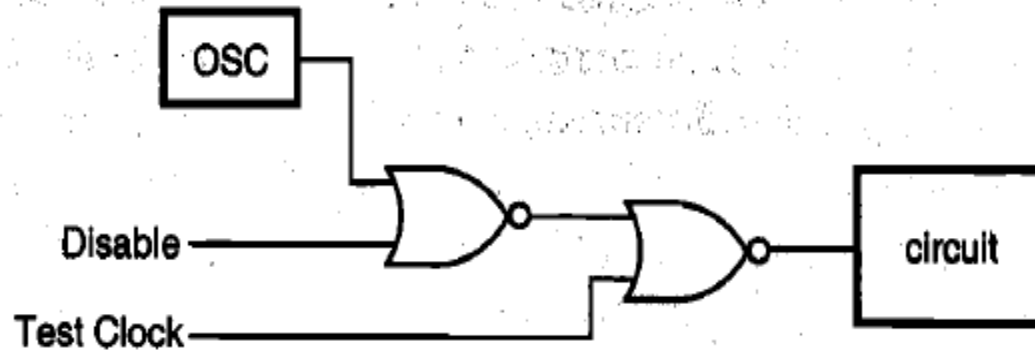


Figure 4. Avoid synchronization problems-via disabling of the oscillator.

Avoid Asynchronous Logic and Redundant Logic

- The enhancement of testability requires serious tradeoffs.
- The speed of an asynchronous logic circuit can be faster than that of the synchronous logic circuit counterpart.
- However, the design and test of an asynchronous logic circuit are more difficult than for a synchronous logic circuit, and its state transition times are difficult to predict.

Avoid Asynchronous Logic and Redundant Logic

- Also, the operation of an asynchronous logic circuit is sensitive to input test patterns, often causing race problems and hazards of having momentary signal values opposite to the expected values.
- Sometimes, designed-in logic redundancy is used to mask a static hazard condition for reliability.

Avoid Asynchronous Logic and Redundant Logic

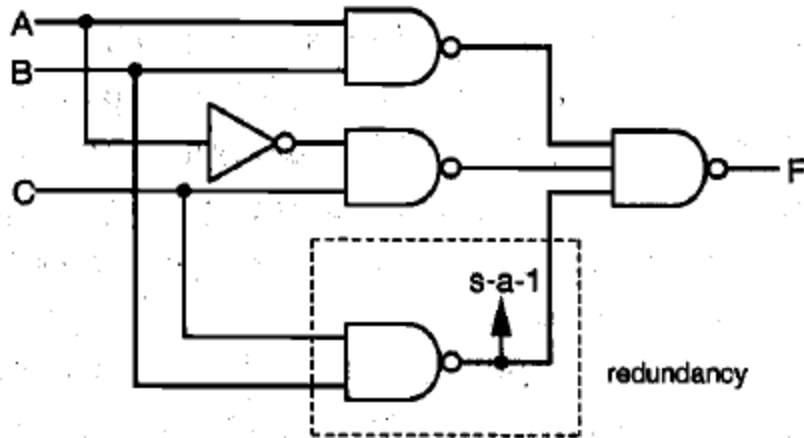
- However, the redundant node cannot be observed since the primary output value cannot be made dependent on the value of the redundant node.
- Hence, certain faults on the redundant node cannot be tested or detected.
- Figure 5 shows that the bottom NAND2 gate is redundant and the stuck-at-1 fault on its output line cannot be detected.

Avoid Asynchronous Logic and Redundant Logic

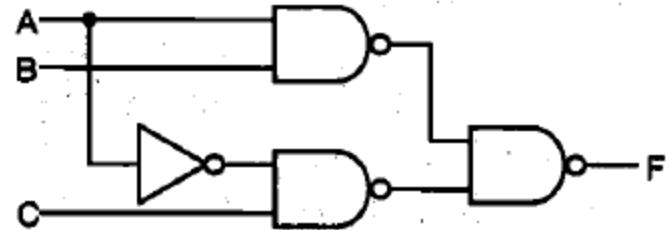
- If a fault is undetectable, the associated line or gate can be removed without changing the logic function.

Avoid Asynchronous Logic and Redundant Logic

$$F = AB + BC + \bar{A}C$$
$$= AB + \bar{A}C$$



(a)



(b)

Figure 5. (a) A redundant logic gate example. (b) Equivalent gate with redundancy removed

Avoid Asynchronous Logic and Redundant Logic

- Although it is unessential to test redundant nodes when they are designed in as backup parts, either to enhance the circuit reliability or to increase the fabrication yield, the use of redundant circuits can make test generation much more complex and difficult.
- In fact, test generators, especially the random or deterministic test generators, would not be able to recognize such design intent.
- Some redundancy in circuits may be unintentional due to lack of design efficiency.

Avoid Delay-Dependent Logic

- Chains of inverters can be used to design in delay times and use AND operation of their outputs along with inputs to generate pulses, as shown in Fig. 6.

Avoid Delay-Dependent Logic

- Most automatic test pattern generation (ATPG) programs do not include logic delays to minimize the complexity of the program.
- As a result, such delay-dependent logic is viewed as redundant combinational logic, and the output of the reconvergent gate is always set to logic 0, which is not correct.
- Thus, the use of delay-dependent logic should be avoided in design for testability.

Avoid Delay-Dependent Logic

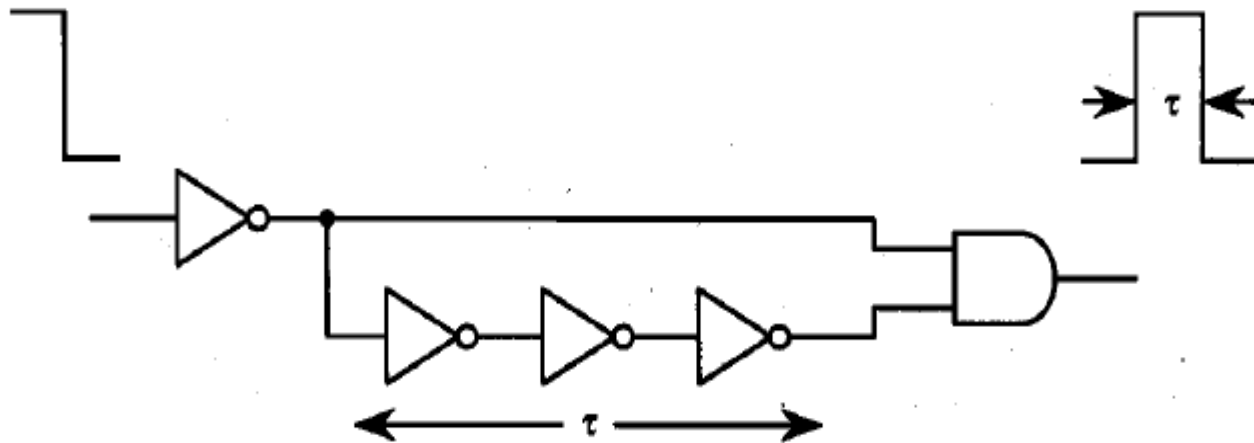


Figure 6. A pulse-generation circuit using a delay chain of three inverters.

Scan-Based Techniques

- As discussed earlier, the controllability and observability can be enhanced by providing more accessible logic nodes with use of additional primary input lines and multiplexors.
- However, the use of additional I/O pins can be costly not only for chip fabrication but also for packaging.

Scan-Based Techniques

- A popular alternative is to use scan registers with both shift and parallel load capabilities.
- The scan design technique is a structured approach to design sequential circuits for testability.
- The storage cells in registers are used as observation points, control points, or both.
- By using the scan design techniques, the testing of a sequential circuit is reduced to the problem of testing a combinational circuit.

Scan-Based Techniques

- In general, a sequential circuit consists of a combinational circuit and some storage elements. In the scan-based design, the storage elements are connected to form a long serial shift register, the so-called scan path, by using multiplexors and a mode (test/ normal) control signal, as shown in Fig. 7.

Scan-Based Techniques

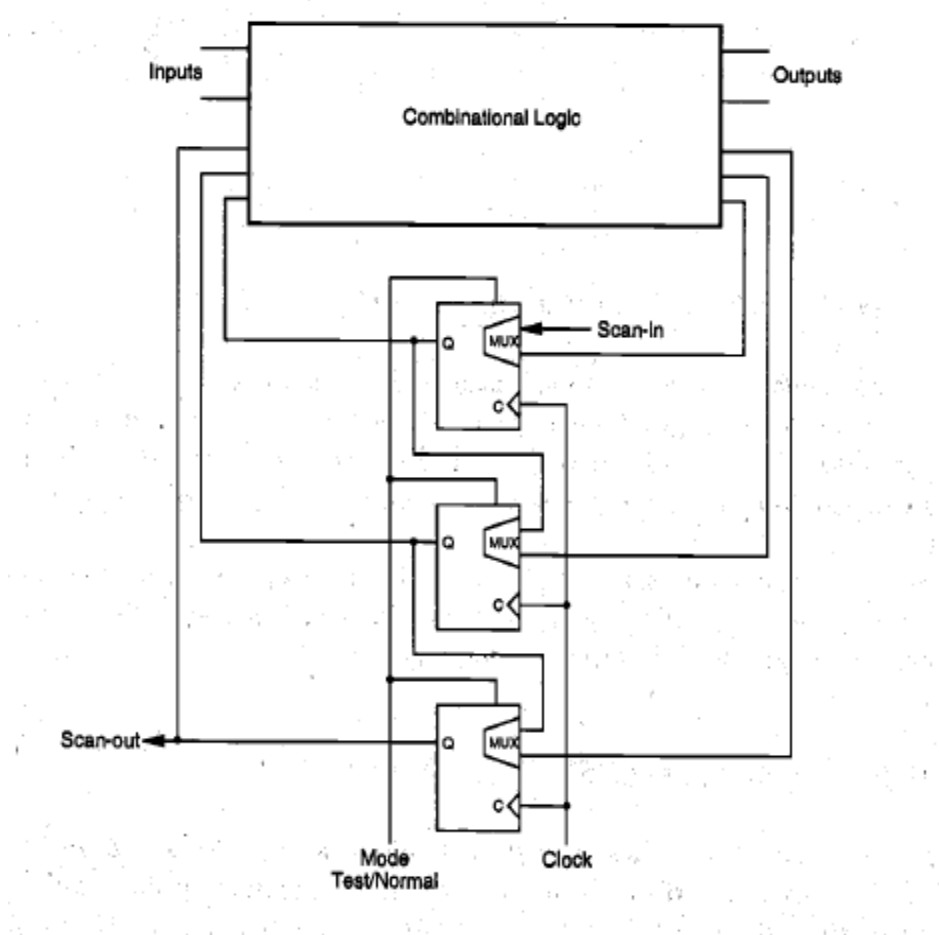


Figure 7. The general structure of scan-based design

Scan-Based Techniques

- In the test mode, the scan-in signal is clocked into the scan path, and the output of the last stage latch is scanned out.
- In the normal mode, the scan-in path is disabled and the circuit functions as a sequential circuit.
- The testing sequence is as follows:

Scan-Based Techniques

- Step 1: Set the mode to test and, let latches accept data from scan-in input,
- Step 2: Verify the scan path by shifting in and out the test data.
- Step 3: Scan in (shift in) the desired state vector into the shift register.
- Step 4: Apply the test pattern to the primary input pins.
- Step 5: Set the mode to normal and observe the primary outputs of the circuit after sufficient time for propagation.

Scan-Based Techniques

- Step 6: Assert the circuit clock, for one machine cycle to capture the outputs of the combinational logic into the registers.
- Step 7: Return to test mode; scan out the contents of the registers, and at the same time scan in the next pattern.
- Step 8: Repeat steps 3-7 until all test patterns are applied.

Scan-Based Techniques

- The storage cells in scan design can be implemented using edge-triggered D flip flops, master-slave flip-flops, or level-sensitive latches controlled by complementary clock signals to ensure race-free operation.
- Figure 8 shows a scan-based design of an edge-triggered D flip-flop.
- In large high-speed circuits, optimizing a single clock signal for skews, etc., both for normal operation and for shift operation, is difficult.

Scan-Based Techniques

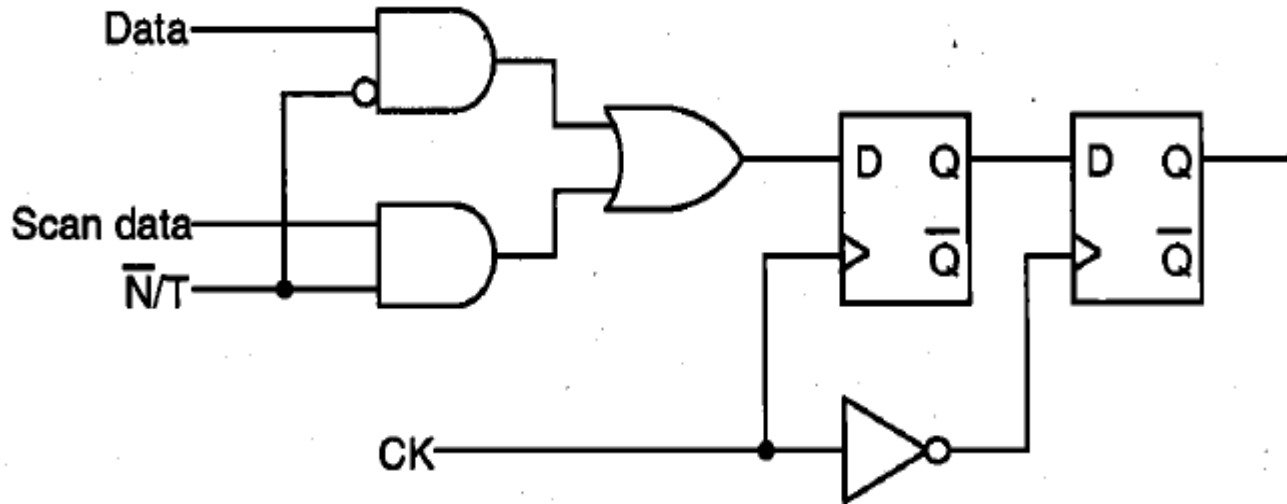


Figure 8. Scan-based design of an edge-triggered D flip-flop.

Scan-Based Techniques

- The boundary scan test method is also used for testing printed circuit boards (PCBs) and multichip modules (MCMs) carrying multiple chips.
- Shift registers are placed in each chip close to I/O pins in order to form a chain around the board for testing.
- With successful implementation of the boundary scan method, a simpler tester can be used for PCB testing.

Scan-Based Techniques

- On the negative side, scan design uses more complex latches, flip-flops, I/O pins, and interconnect wires and, thus, requires more chip area.
- The testing time per test pattern is also increased due to shift time in long registers.

Built-In Self Test (BIST) Techniques

- In built-in self test (BIST) design, parts of the circuit are used to test the circuit itself.
- Online BIST is used to perform the test under normal operation, whereas off-line BIST is used to perform the test off-line.
- The essential circuit modules required for BIST include;
- Pseudo random pattern generator (PRPG)
- Output response analyzer (ORA)

Built-In Self Test (BIST) Techniques

- The roles of these two modules are illustrated in Fig. 9.
- The implementation of both PRPG and ORA can be done with Linear Feedback Shift Registers (LFSRs).

Built-In Self Test (BIST) Techniques

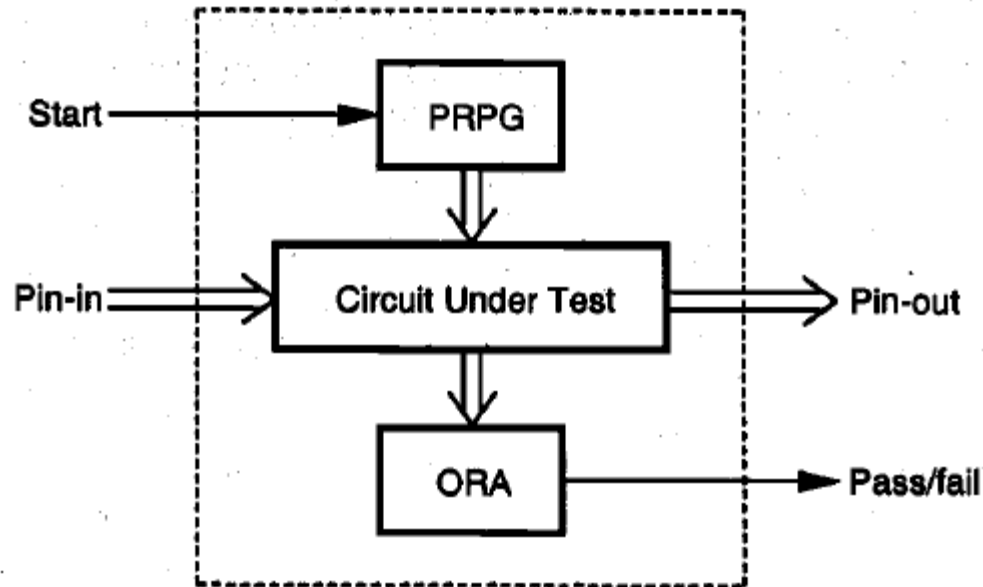


Figure 9. A procedure for BIST.

Pseudo Random Pattern Generator

- To test the circuit, test patterns first have to be generated either by using a pseudo random pattern generator, a weighted test generator, an adaptive test generator, or other means.
- A pseudo random test generator circuit can use an LFSR, as shown in Fig. 10.

Pseudo Random Pattern Generator

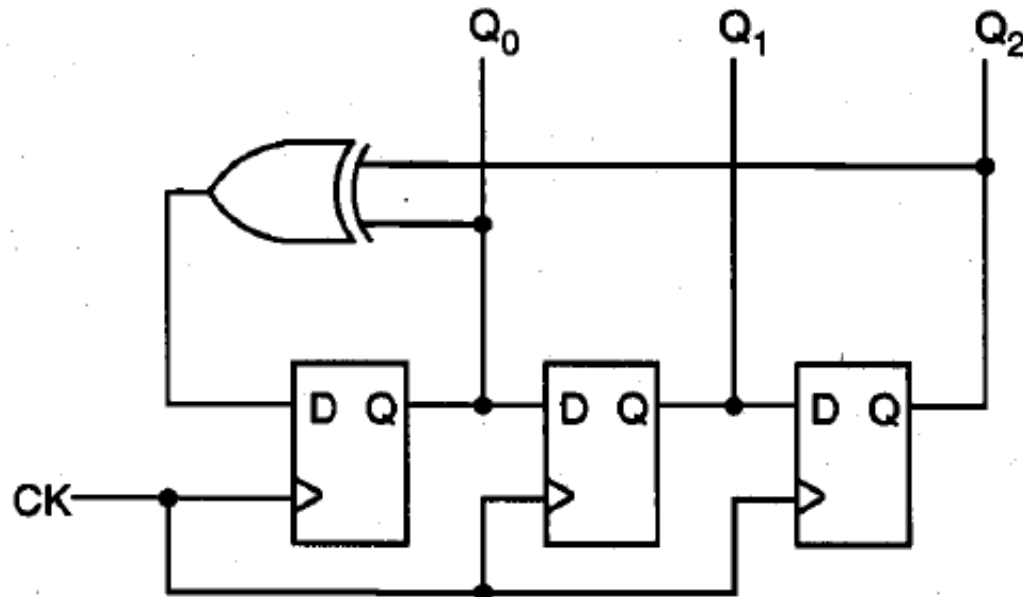


Figure 10. A pseudo-random sequence generator using LFSR

Linear Feedback Shift Register as an ORA

- To reduce the chip area penalty, data compression schemes are used to compare the compacted test responses instead of the entire raw test data.
- One of the popular data compression schemes is the signature analysis, which is based on the concept of cyclic redundancy checking.
- It uses polynomial division, which divides the polynomial representation of the test output data by a characteristic polynomial and then

Linear Feedback Shift Register as an ORA

- finds the remainder as the signature.
- The signature is then compared with the expected signature to determine whether the device under test is faulty.
- It is known that compression can cause some loss of fault coverage.

Linear Feedback Shift Register as an ORA

- It is possible that the output of a faulty circuit can match the output of the fault-free circuit; thus, the fault can go undetected in the signature analysis. Such a phenomenon is called aliasing.

Linear Feedback Shift Register as an ORA

- In its simplest form, the signature generator consists of a single-input linear feedback shift register (LFSR), as shown in Fig. 11, in which all the latches are edge-triggered.
- In this case, the signature is the content of this register after the last input bit has been sampled.
- The input sequence $\{a_n\}$ is represented by polynomial $G(x)$ and the output sequence by $Q(x)$.

Linear Feedback Shift Register as an ORA

- It can be shown that $G(x) = Q(x) P(x) + R(x)$, where $P(x)$ is the characteristic polynomial of LFSR and $R(x)$ is the remainder, the degree of which is lower than that of $P(x)$.
- For the simple case in Fig. 11, the characteristic polynomial is

$$P(x) = 1 + x^2 + x^4 + x^5$$

Linear Feedback Shift Register as an ORA

- For the 8-bit input sequence { 1 1 1 1 0 1 0 1 } the corresponding input polynomial is

$$G(x) = x^7 + x^6 + x^5 + x^4 + x^2 + 1$$

Linear Feedback Shift Register as an ORA

- and the remainder term becomes $R(x) = x^4 + x^2$, which corresponds to the register contents of {0 0 1 0 11}

Linear Feedback Shift Register as an ORA

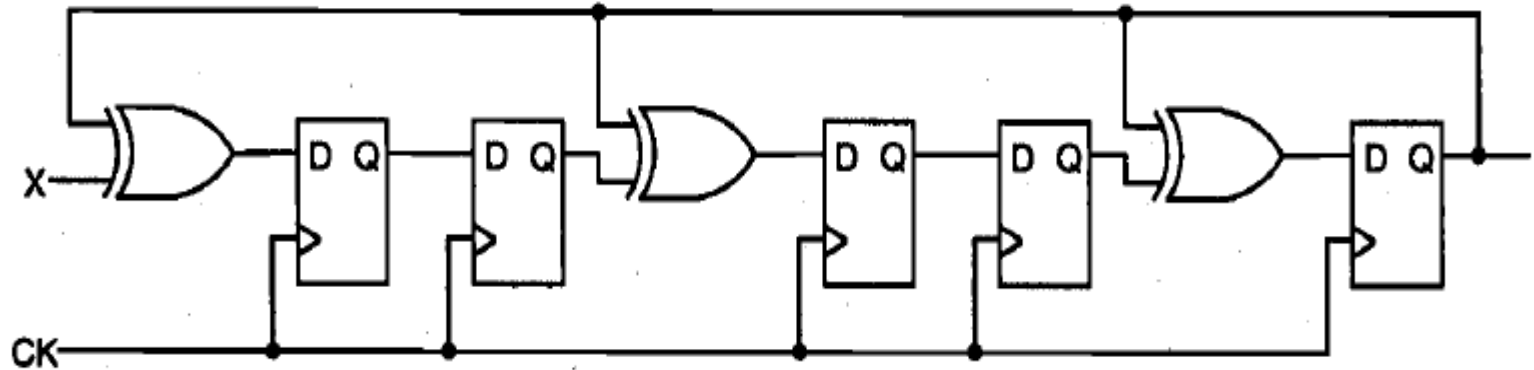


Figure 11. Polynomial division using LFSR for signature analysis

Output Response Analyzer

- The on-chip storage of a fault dictionary containing all test inputs with the corresponding outputs is prohibitively expensive in terms of the chip area.
- A simple alternative method is to compare the outputs of two identical circuits for the same input, with one of them regarded as reference.

Output Response Analyzer

- However, if both circuits have the same faults, their outputs can still match.
- Such faults cannot be detected with this technique, although, the probability of two identical circuits having exactly the same faults would be very low.

Output Response Analyzer

- In addition to the above circuits for built-in self test, self-checking design techniques
- can be used to detect faults autonomously during on-line operation. Usually a checker
- circuit is inserted such that the checker generates and sends out a signal when on-line
- faults occur. The distribution of checkers throughout a very large digital circuit or

Output Response Analyzer

- system can provide prompt detection of the fault location by tracing the checker which sent the fault signal.
- The use of self-checking circuits simplifies the development of software diagnostic programs.

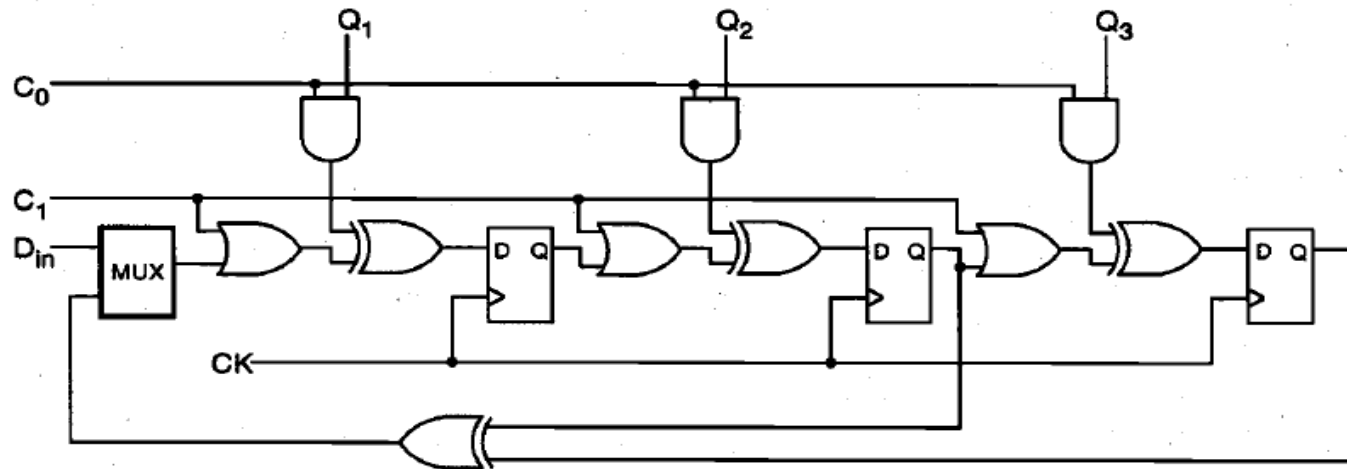
Output Response Analyzer

- However, some additional hardware is required, and the checker itself needs to have self-checking capability.
- When self-checking capability of the checker itself is required, a single-output checker is not sufficient since that output may have a stuck-at fault, thus preventing the detection of actual faults in the circuit under test.
- A checker with a pair of outputs can be used instead to overcome this problem.

Built-In Logic Block Observer

- The built-in logic block observer (BILBO) register is a form of ORA which can be used in each cluster of partitioned registers.
- A basic BILBO circuit is shown in Fig. 12, which allows four different modes controlled by C_0 and C_1 signals.

Built-In Logic Block Observer



C_0	C_1	Mode
0	0	linear shift
1	0	signature analysis
1	1	data (complemented) latch
0	1	reset

Figure 12. 3-bit built-in logic observer (BILBO) example.

Built-In Logic Block Observer

- The BILBO operation allows monitoring of circuit operation through exclusive- ORing into LFSR at multiple points, which corresponds to the signature analyzer with multiple inputs.

Current Monitoring I_{DDQ} Test

- An often-used technique for testing fabrication defects is the I_{DDQ} test.
- Under a bridging fault, the static currents drawn from the power supply in CMOS circuits can be noticeably high, well beyond the expected range of leakage currents.
- For example, if the drain node of the pMOS transistor in a CMOS inverter is shorted to the power supply rail due to a bridging fault, its I_{DDQ} current can be very high even when the input is high.

Current Monitoring I_{DDQ} Test

- It can also detect other fabrication defects not easily detected by other test methods, including:
 - Gate oxide short
 - Channel punch-through
 - p-n diode leakage
 - Transmission-gate defect

Current Monitoring I_{DDQ} Test

- The I_{DDQ} test consists of applying the test vector and then monitoring the current drawn from the power supply rail in DC steady state.
- Although this test requires more testing time, the fault detection capability is greatly improved with small circuit overhead required to monitor the I_{DDQ} in various parts of DUT.

Current Monitoring I_{DDQ} Test

- While stuck-at tests require both fault sensitization and fault effect propagation, the I_{DDQ} test requires only fault sensitization.
- However, its performance in open drain and open gate test is less effective. The I_{DDQ} fault coverage is relatively easy to obtain and may potentially offer a full-chip coverage capability for large designs.

Current Monitoring I_{DDQ} Test

- The design guidelines for I_{DDQ} testability are as follows:
- Low static current states, e.g., full CMOS is preferred
- No active pull-ups or pull-downs
- No internal drive conflicts, e.g., drivers share a bus
- No floating nodes in the circuit
- No degraded voltages, e.g., must have $V_{OH} = V_{DD}$ and $V_{OL} = 0$.

DYNAMIC LOGIC CIRCUITS

Dynamic Logic Circuits

- A wide range of static combinational and sequential logic circuits was introduced in the previous chapters.
- Static logic circuits allow versatile implementation of logic functions based on static, or steady-state, behavior of simple nMOS or CMOS structures.

Dynamic Logic Circuits

- In high-density, high-performance digital implementations where reduction of circuit delay and silicon area is a major objective, dynamic logic circuits offer several significant advantages over static logic circuits.
- The operation of all dynamic logic gates depends on temporary (transient) storage of charge in parasitic node capacitances, instead of relying on steady-state circuit behavior.

Example-1

- The example-1 presents the operation of a dynamic D-latch circuit, which essentially consists of two inverters connected in cascade.
- This simple circuit illustrates most of the basic operational concepts involved in dynamic circuit design.
- Consider the dynamic D-latch circuit shown Fig. 1.
- The circuit consists of two cascaded inverters and one nMOS pass transistor driving the input of the primary inverter stage.

Example-1

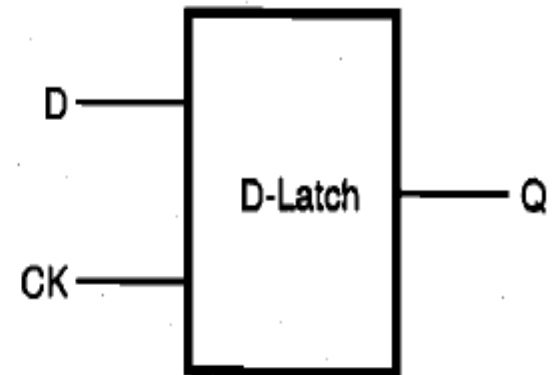
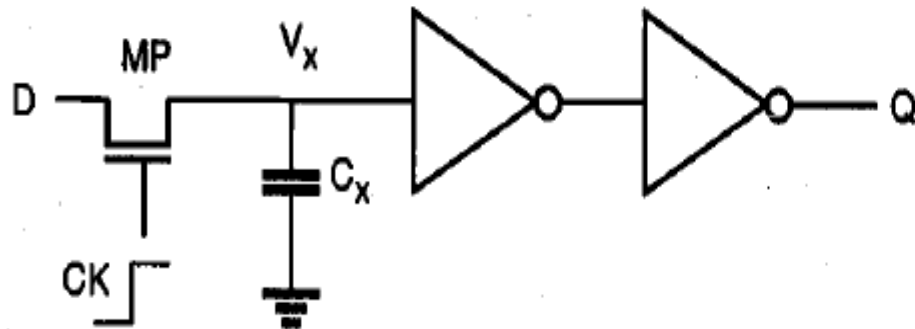


Figure 1

Example-1

- We will see that the parasitic input capacitance C_x of the primary inverter stage plays an important role in the dynamic operation of this circuit.
- The input pass transistor is being driven by the external periodic clock signal, as follows:

Example-1

- When the clock is high ($CK = 1$), the pass transistor turns on. The capacitor C_x is either charged up, or charged down through the pass transistor MP, depending on the input (D) voltage level. The output (Q) assumes the same logic level as the input.
- When the clock is low ($CK = 0$), the pass transistor MP turns off, and the capacitor C_x is isolated from the input D. Since there is no current path from the intermediate node X to either VDD or ground, the amount of charge stored in C_x during the previous cycle determines the output voltage level Q.

Example-1

- The "hold" operation during the inactive clock cycle is accomplished by temporarily storing charge in the parasitic capacitance C_x .
- Correct operation of the circuit critically depends on how long a sufficient amount of charge can be retained at node X, before the output state changes due to charge leakage. Therefore, the capacitive intermediate node X is also called a soft node.

Example-1

- Now, we will examine the circuit operation in more detail.
- Assume that the dynamic D-latch circuit is being operated with a power supply voltage of $V_{DD} = 5V$, and that the VTC of both inverters are identical, with

$$V_{OL} = 0 \text{ V}$$

$$V_{IL} = 2.1 \text{ V}$$

$$V_{IH} = 2.9 \text{ V}$$

$$V_{OH} = 5.0 \text{ V}$$

Example-1

- The threshold voltage of the pass transistor MP is given as $V_{T,n} = 0.8 \text{ V}$.
- During the active clock phase ($CK = 1$), assume that the input is equal to logic " 1," i.e., $V_{in} = V_{OH} = 5 \text{ V}$.
- The pass transistor MP is conducting during this phase, and the parasitic intermediate node capacitance C_x , is charged up to a logic-high level.

Example-1

- We recall that the nMOS pass transistor is a poor conductor for logic "1," and its output voltage V_x will be lower than V_{OH} , by one threshold voltage: $V_x = 5.0 - 0.8 = 4.2$ V. Still, this voltage is clearly higher than the V_{IH} of the first inverter, thus, the output voltage of the first inverter will be very close to $V_{OL} = 0$ V. Consequently, the output level Q of the secondary inverter becomes a logic '1,' $V_Q = V_{DD}$.

Example-1

- Next, the clock signal goes to zero, and the pass transistor turns off.
- Initially, the logic-high level at node X is preserved through charge storage in C_x .
- Thus, the output level Q also remains at logic " 1. " However, the voltage V_x eventually starts to drop from its original level of 4.2 V because of charge leakage from the soft node.

Example-1

- It can easily be seen that in order to keep the output node Q at logic "1," the voltage level at the intermediate node X cannot be allowed to drop lower than $V_{IH} = 2.9 \text{ V}$ (once V_i falls below this level, the input of the first inverter cannot be interpreted as a logic "1").
- Thus, the inactive clock phase during which the clock signal is equal to zero can, at most, be as long as it takes for the intermediate voltage V_x to drop from 4.2 V to 2.9 V, due to charge leakage.

Example-1

- To avoid an erroneous output, the charge stored in C_x must be restored, or refreshed, to its original level before V_x reaches 2.9 V.

Example-1

- This example shows that the simple dynamic-charge storage principle employed in the D-latch circuit is quite feasible for preserving an output state during the inactive clock phase, assuming that the leakage current responsible for draining the capacitance C_x , are relatively small.

Basic Principles of Pass Transistor Circuits

- The fundamental building block of nMOS dynamic logic circuits, consisting of an nMOS pass transistor driving the gate of another nMOS transistor, is shown in Fig. 2.
- The two possible operations when the clock signal is active ($CK = 1$) are the logic "1" transfer (charging up the capacitance C_x to a logic-high level) and the logic "0" transfer (charging down the capacitance C_x to a logic-low level).
- In either case, the output of the depletion-load nMOS inverter assumes a logic-low or a logic-high level, depending on the voltage V_x .

Basic Principles of Pass Transistor Circuits

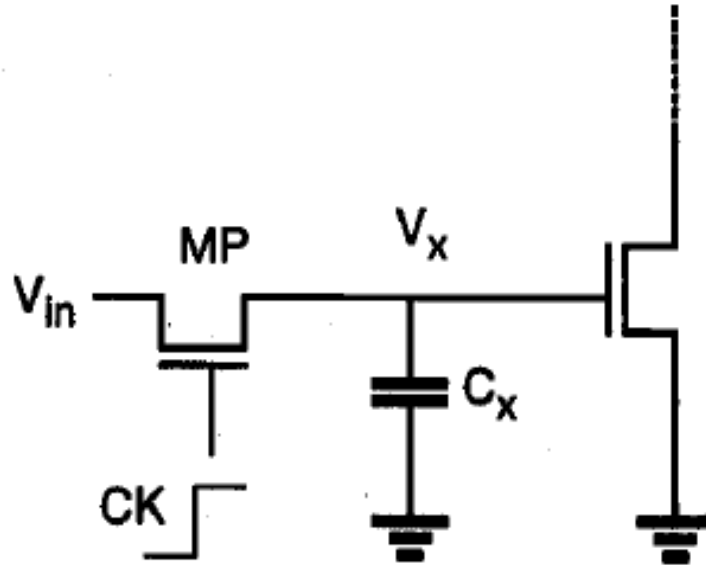


Figure 2. The basic building block for nMOS dynamic logic, which consists of an nMOS pass transistor driving the gate of another nMOS transistor.

Basic Principles of Pass Transistor Circuits

- The pass transistor MP provides the only current path to the intermediate capacitive node (soft node) X.
- When the clock signal becomes inactive ($CK = 0$), the pass transistor ceases to conduct and the charge stored in the parasitic capacitor C_x continues to determine the output level of the inverter.

Logic "1" Transfer

- Assume that the soft node voltage is equal to 0 initially, i.e., $V_x(t = 0) = 0 \text{ V}$.
- A logic "1" level is applied to the input terminal, which corresponds to $V_{in} = V_{OH} = V_{DD}$.
- Now, the clock signal at the gate of the pass transistor goes from 0 to V_{DD} at $t = 0$.

Logic "1" Transfer

- It can be seen that the pass transistor MP starts to conduct as soon as the clock signal becomes active and that MP will operate in saturation throughout this cycle since $V_{DS} = V_{GS}$.
- Consequently, $V_{DS} > V_{GS} - V_{T,n}$.
- The circuit to be analyzed for the logic 1 " transfer event can be simplified into an equivalent circuit as shown in Fig. 3.

Logic "1" Transfer

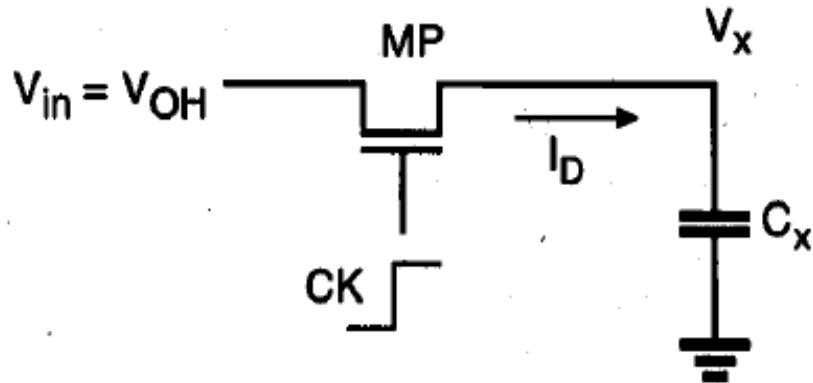


Figure 3. Equivalent circuit for the logic " 1 " transfer event

Logic "1" Transfer

- The pass transistor MP operating in the saturation region starts to charge up the capacitor C_x , thus

$$C_x \frac{dV_x}{dt} = \frac{k_n}{2} (V_{DD} - V_x - V_{T,n})^2 \quad \text{.....(1)}$$

- The threshold voltage of the pass transistor is actually subject to substrate bias effect and depends on the voltage level V_x .

Logic "1" Transfer

- To simplify the analysis, we will neglect the substrate bias effect at this point.
- Integrating (Eqn. 1), we get,

$$\int_0^t dt = \frac{2C_x}{k_n} \int_0^{V_x} \frac{dV_x}{(V_{DD} - V_x - V_{T,n})^2} \quad \text{.....(2)}$$
$$= \frac{2C_x}{k_n} \left(\frac{1}{(V_{DD} - V_x - V_{T,n})} \right) \Bigg|_0^{V_x}$$

$$t = \frac{2C_x}{k_n} \left[\left(\frac{1}{V_{DD} - V_x - V_{T,n}} \right) - \left(\frac{1}{V_{DD} - V_{T,n}} \right) \right] \quad \text{.....(3)}$$

Logic "1" Transfer

- This equation can be solved for $V_x(t)$, as follows.

$$V_x(t) = (V_{DD} - V_{T,n}) \frac{\left(\frac{k_n (V_{DD} - V_{T,n})}{2C_x} \right) t}{1 + \left(\frac{k_n (V_{DD} - V_{T,n})}{2C_x} \right) t} \dots\dots\dots(4)$$

Logic "1" Transfer

- The variation of the node voltage V according to (Eqn. 4) is plotted as a function of time in Fig. 4.
- The voltage rises from its initial value of 0 V and approaches a limit value for large t , but it cannot exceed its limit value of $V_{\max} = (V_{DD} - V_{T,n})$.
- The pass transistor will turn off for $V_x = V_{\max}$, since at this point, its gate-to-source voltage will be equal to its threshold voltage.

Logic "1" Transfer

- Therefore, the voltage at node X can never attain the full power supply voltage level of V_{DD} during the logic "1" transfer.
- The actual value of the maximum possible voltage V_{max} at node X can be found by taking into account the substrate bias effect for MP.

$$\begin{aligned} V_{max} &= V_x \Big|_{t \rightarrow \infty} = V_{DD} - V_{T,n} \\ &= V_{DD} - V_{T0,n} - \gamma \left(\sqrt{|2\phi_F| + V_{max}} - \sqrt{|2\phi_F|} \right) \end{aligned} \quad \text{.....(5)}$$

Logic "1" Transfer

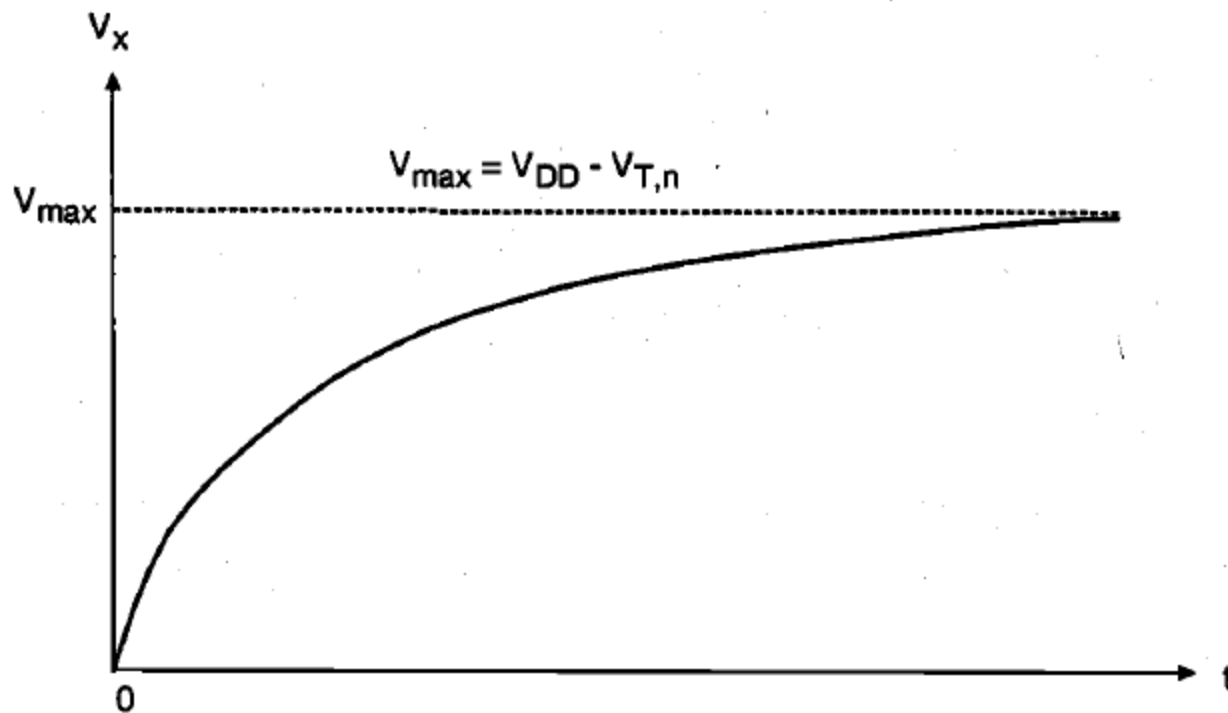


Figure 4. Variation of V_x as a function of time during logic "1" transfer

Logic "1" Transfer

- Thus, the voltage V_x which is obtained at node X following a logic "1" transfer can be considerably lower than V_{DD} .

Logic "1" Transfer

- As an example, let us consider the following case in which a logic "1" at the input node ($V_{in} = V_{DD}$) is being transferred through a chain of cascaded pass transistors (Fig. 5).
- For simple analysis, we assume that initially all internal node voltages V_1 through V_4 , are zero.

Logic "1" Transfer

- The first pass transistor M_1 operates in saturation with $V_{DS1} > V_{GS1} - V_{T,n1}$.
- Therefore, the voltage at node 1 cannot limit value $V_{\max1} = (V_{DD} - V_{T,n1})$.
- Now, assuming that the pass transistors in this circuit are identical, the second pass transistor M_2 operates at the saturation boundary.

Logic "1" Transfer

- As a result, the voltage at node 2 will be equal to $V_{\max 2} = (V_{DD} - V_{T,n2})$.
- It can easily be seen that with $V_{T,n1} = V_{T,n2} = V_{T,n3} \dots$, the node voltage at the end of the pass transistor will become one threshold voltage lower than V_{DD} , regardless of the number of pass transistors in the chain.

Logic "1" Transfer

- It can be observed that the steady-state internal node voltages in this circuit are always one threshold voltage below V_{DD} , regardless of the initial voltages.

Logic "1" Transfer

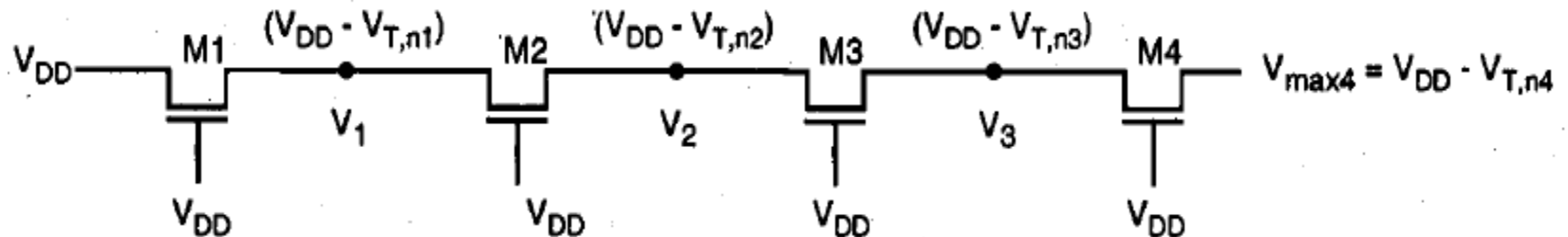


Figure 5. Node voltages in a pass-transistor chain during the logic " 1 " transfer

Logic "1" Transfer

- Now consider a different case in which the output of each pass transistor drives the gate of another pass transistor, as depicted in Fig. 6.

Logic "1" Transfer

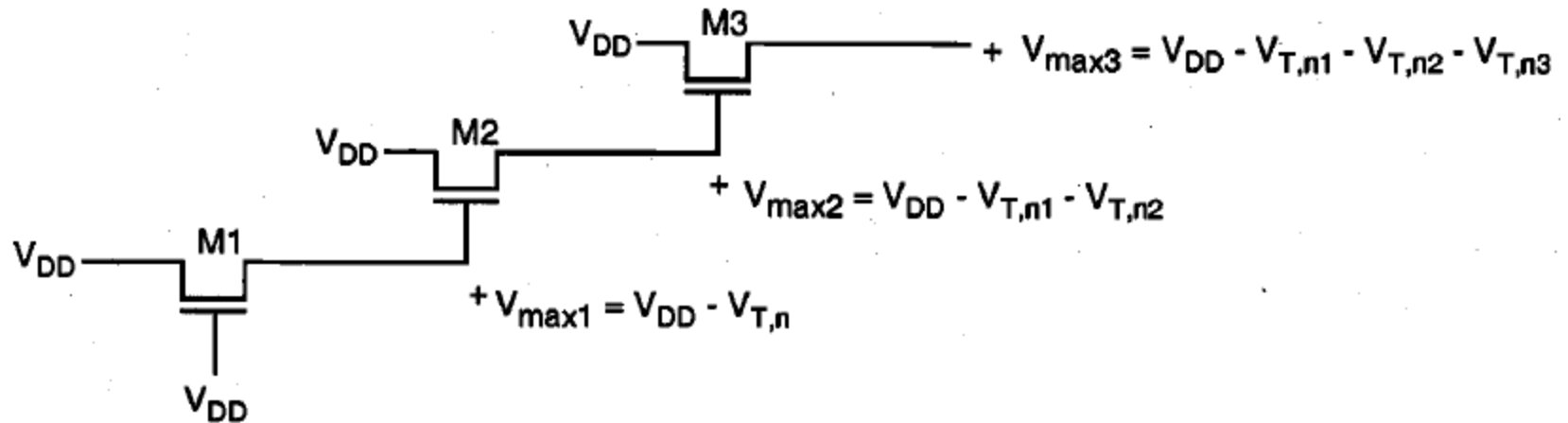


Figure 6. Node voltages during the logic " 1 " transfer, when each pass transistor is driving another pass transistor

Logic "1" Transfer

- It can be seen that in this case, each stage causes a significant loss of voltage level.
- The amount of voltage drop at each stage can be approximated more realistically by taking into account the corresponding substrate bias effect, which is different in all stages.

$$\begin{aligned} V_{T,n1} &= V_{T0,n} - \gamma \left(\sqrt{|2\phi_F| + V_{max1}} - \sqrt{|2\phi_F|} \right) \\ V_{T,n2} &= V_{T0,n} - \gamma \left(\sqrt{|2\phi_F| + V_{max2}} - \sqrt{|2\phi_F|} \right) \\ &\vdots \end{aligned} \quad \text{.....(6)}$$

Logic "0" Transfer

- Assume that the soft-node voltage V_x is equal to a logic " 1 " level initially, i.e., $V_x(t = 0) = V_{\max} = (V_{DD} - V_{T,n})$.
- A logic "0" level is applied to the input terminal, which corresponds to $V_{in} = 0$ V.
- Now, the clock signal at the gate of the pass transistor goes from 0 to V_{DD} at $t = 0$.
- The pass transistor MP starts to conduct as soon as the clock signal becomes active, and the direction of drain current flow through MP will be opposite to that during the charge-up (logic " 1 " transfer) event.

Logic “0” Transfer

- This means that the intermediate node X will now correspond to the drain terminal of MP and that the input node will correspond to its source terminal.
- With $V_{GS} = V_{DD}$ and $V_{DS} = V_{max}$, it can be seen that the pass transistor operates in the linear region throughout this cycle, since $V_{DS} < V_{GS} - V_{T,n}$.

Logic "0" Transfer

- The circuit to be analyzed for the logic "0" transfer event can be simplified into an equivalent circuit as shown in Fig. 7.
- As in the logic " 1 " transfer case, the depletion load nMOS inverter does not affect this event.

Logic "0" Transfer

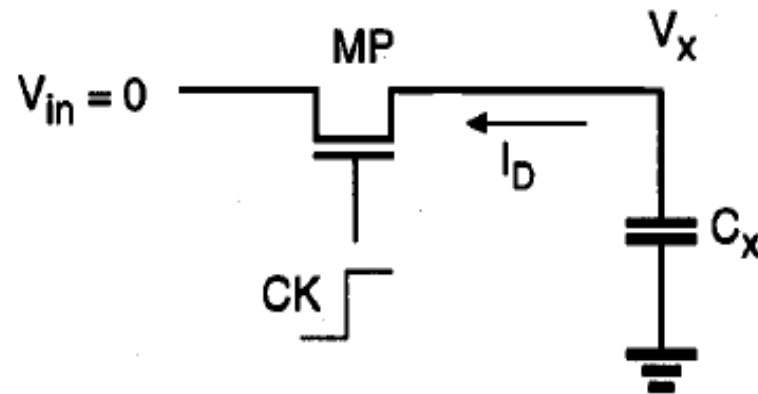


Figure 7. Equivalent circuit for the logic "0" transfer event

Logic "0" Transfer

- The pass transistor MP operating in the linear region discharges the parasitic capacitor C_x , as follows:

$$-C_x \frac{dV_x}{dt} = \frac{k_n}{2} (2(V_{DD} - V_{T,n})V_x - V_x^2) \quad \dots\dots\dots(7)$$

$$dt = -\frac{2C_x}{k_n} \cdot \frac{dV_x}{2(V_{DD} - V_{T,n})V_x - V_x^2} \quad \dots\dots\dots(8)$$

Logic "0" Transfer

- Note that the source voltage of the nMOS pass transistor is equal to 0 V during this event; hence, there is no substrate bias effect for MP ($V_{T,n} = V_{T0,n}$).
- But the initial condition: $V_x(t = 0) = (V_{DD} - V_{T,n})$ contains the threshold voltage with substrate bias effect, because the voltage V_x is set during the preceding logic "1" transfer event.

Logic "0" Transfer

- To simplify the expressions, we will use $V_{T,n}$ in the following. Integrating both sides of (Eqn. 8) yields

$$\int_0^t dt = -\frac{2C_x}{k_n} \int_{V_{DD}-V_{T,n}}^{V_x} \left(\frac{\frac{1}{2(V_{DD}-V_{T,n})}}{2(V_{DD}-V_{T,n})-V_x} + \frac{\frac{1}{2(V_{DD}-V_{T,n})}}{V_x} \right) dV_x \quad \dots\dots\dots(9)$$

$$t = \frac{C_x}{k_n(V_{DD}-V_{T,n})} \left[\ln \left(\frac{2(V_{DD}-V_{T,n})-V_x}{V_x} \right) \right] \bigg|_{V_{DD}-V_{T,n}}^{V_x} \quad \dots\dots\dots(10)$$

Logic "0" Transfer

- Finally, the fall-time expression for the node voltage V_x can be obtained as

$$t = \frac{C_x}{k_n (V_{DD} - V_{T,n})} \ln \left(\frac{2(V_{DD} - V_{T,n}) - V_x}{V_x} \right) \dots\dots\dots(11)$$

Logic "0" Transfer

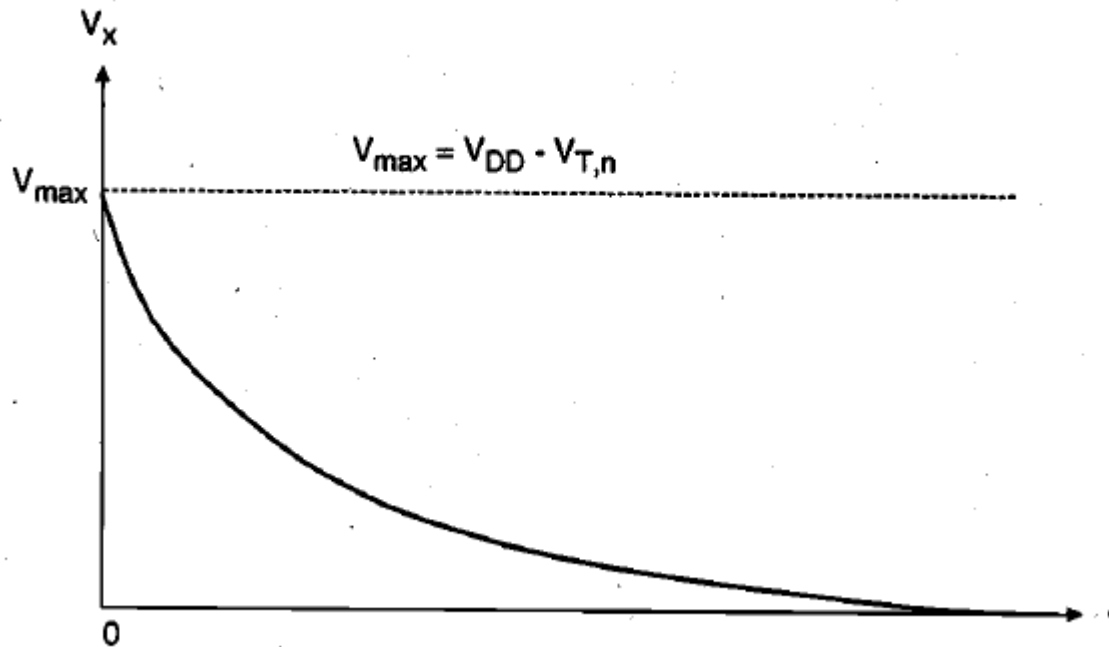


Figure 8. Variation of V_x as a function of time during logic "0" transfer.

Logic “0” Transfer

- The fall time (τ_{fall}) for the soft-node voltage V_x can be calculated from (Eqn. 11) as follows.
- First, define the two time points $t_{90\%}$ and $t_{10\%}$ as the times at which the node voltage is equal to $0.9 V_{\text{max}}$ and $0.1 V_{\text{max}}$, respectively.
- These two time points can easily be found by using (Eqn. 11).

Logic "0" Transfer

$$\begin{aligned} t_{90\%} &= \frac{C_x}{k_n (V_{DD} - V_{T,n})} \ln \left(\frac{(2 - 0.9)(V_{DD} - V_{T,n})}{0.9(V_{DD} - V_{T,n})} \right) \\ &= \frac{C_x}{k_n (V_{DD} - V_{T,n})} \ln \left(\frac{1.1}{0.9} \right) \end{aligned} \quad \text{.....(12)}$$

$$t_{10\%} = \frac{C_x}{k_n (V_{DD} - V_{T,n})} \ln \left(\frac{1.9}{0.1} \right) \quad \text{.....(13)}$$

Logic "0" Transfer

- The fall time of the soft-node voltage V_x is by definition the difference between $t_{10\%}$ and $t_{90\%}$, which is found as

$$\begin{aligned}\tau_{fall} &= t_{10\%} - t_{90\%} \\ &= \frac{C_x}{k_n (V_{DD} - V_{T,n})} [\ln(19) - \ln(1.22)] \\ &= 2.74 \frac{C_x}{k_n (V_{DD} - V_{T,n})} \quad \dots\dots\dots(14)\end{aligned}$$

Logic "0" Transfer

- Until this point, we have examined the transient charge-up and charge-down events which are responsible for logic "1" transfer and logic "0" transfer during the active clock phase, i.e., when $CK = 1$.
- Now we will turn our attention to the storage of logic levels at the soft node X during the inactive clock cycle, i.e., when $CK = 0$.

Charge Storage and Charge Leakage

- To analyze the events during the inactive clock phase in more detail, consider the scenario shown in Fig. 9.
- Let us assume that a logic-high voltage level has been transferred to the soft node during the active clock phase and that now both the input voltage V_{in} and the clock are equal to 0 V.

Charge Storage and Charge Leakage

- The charge stored in C_x will gradually leak away, primarily due to the leakage currents associated with the pass transistor.
- The gate current of the inverter driver transistor is negligible for all practical purposes

Charge Storage and Charge Leakage

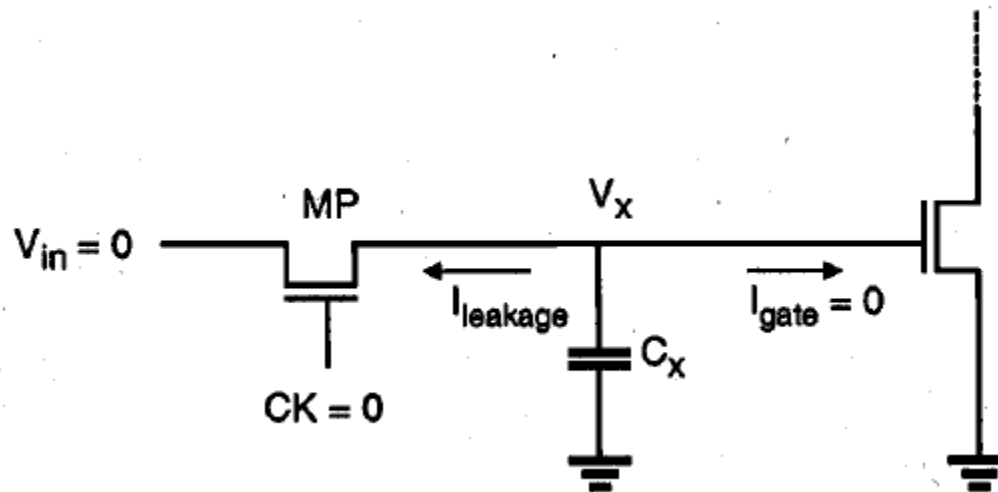


Figure 9. Charge leakage from the soft node

Charge Storage and Charge Leakage

- Figure 10 shows a simplified cross-section of the nMOS pass transistor, together with the lumped node capacitance C_x .
- We see that the leakage current responsible for draining the soft-node capacitance over time has two main components, namely, the subthreshold channel current and the reverse conduction current of the drain-substrate junction.

$$I_{leakage} = I_{subthreshold(MP)} + I_{reverse(MP)} \quad \dots\dots\dots(15)$$

Charge Storage and Charge Leakage

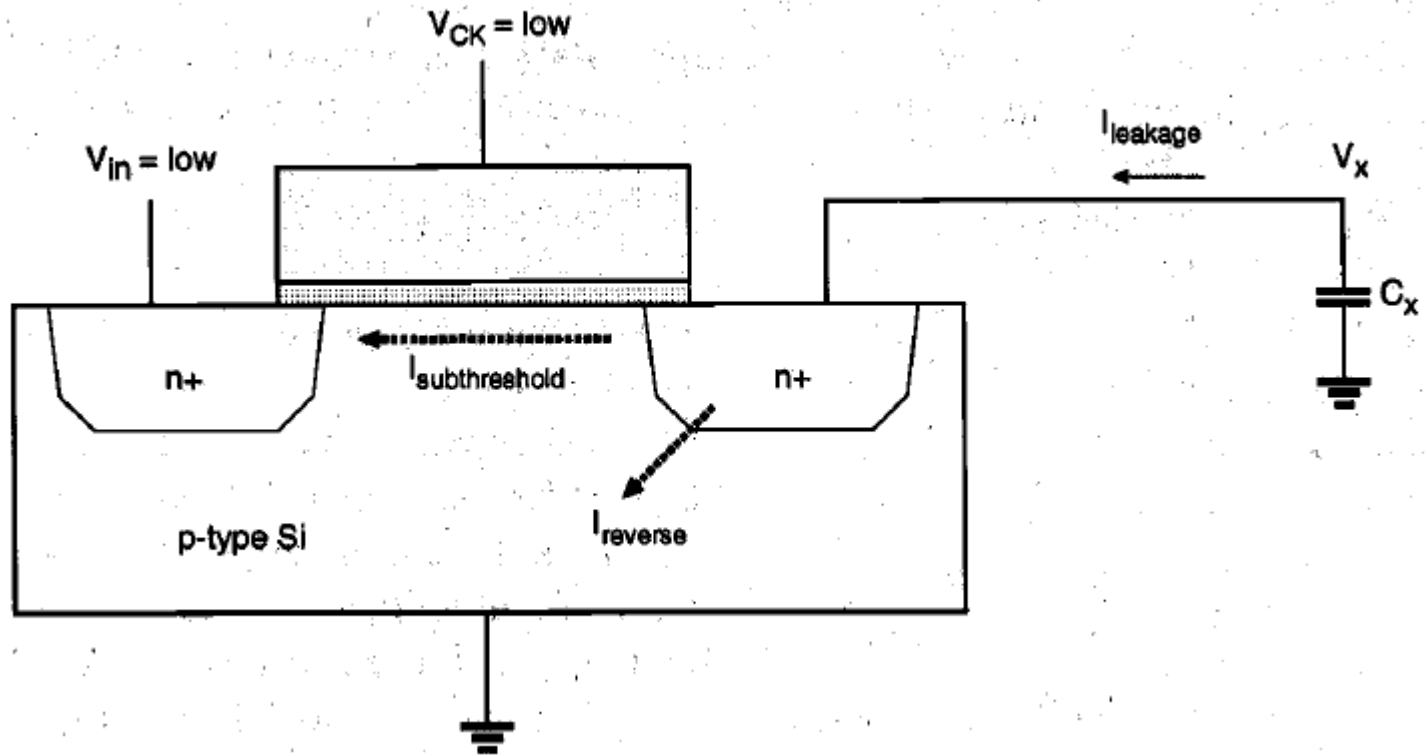


Figure 10. Simplified cross-section of the nMOS pass transistor, showing the leakage current components responsible for draining the soft-node capacitance C'' .

Charge Storage and Charge Leakage

- Other components of C_x , which are primarily due to oxide-related parasitics, can be considered constants.
- In our analysis, these constant capacitance components will be represented by C_{in} , (Fig. 11).

Charge Storage and Charge Leakage

- Thus, we have to express the total charge stored in the soft node as the sum of two main components, as follows.

$$Q = Q_j(V_x) + Q_{in} \quad \text{where} \quad Q_{in} = C_{in} \cdot V_x$$

$$C_{in} = C_{gb} + C_{poly} + C_{metal} \quad \text{.....(16)}$$

Charge Storage and Charge Leakage

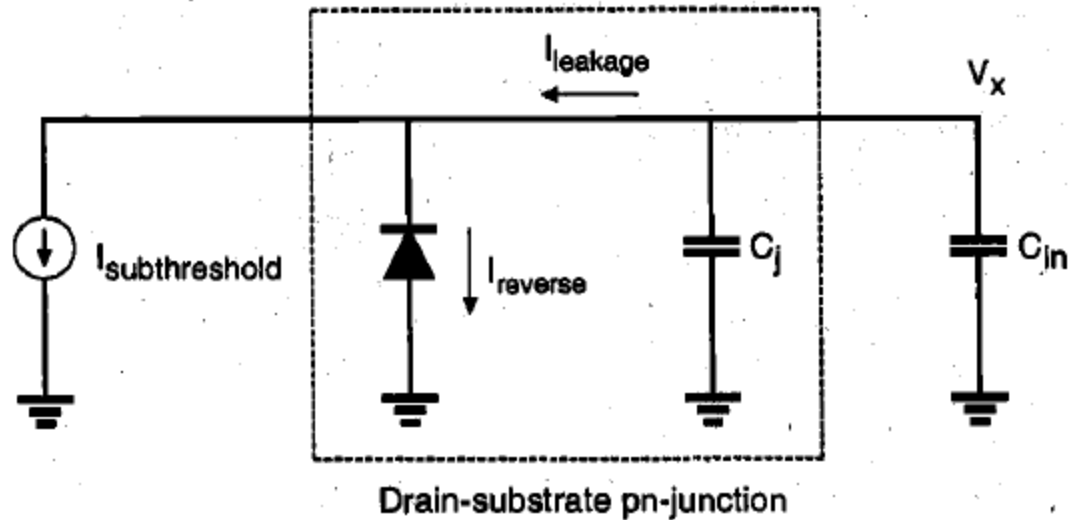


Figure 11. Equivalent circuit used for analyzing the charge leakage process.

Charge Storage and Charge Leakage

- The total leakage current can be expressed as the time derivative of the total soft-node charge Q .

$$\begin{aligned} I_{leakage} &= \frac{dQ}{dt} \\ &= \frac{dQ_j(V_x)}{dt} + \frac{dQ_{in}}{dt} \\ &= \frac{dQ_j(V_x)}{dV_x} \frac{dV_x}{dt} + C_{in} \frac{dV_x}{dt} \end{aligned} \quad \text{.....(17)}$$

Charge Storage and Charge Leakage

- where

$$\frac{dQ_j(V_x)}{dV_x} = C_j(V_x) = \frac{A \cdot C_{j0}}{\sqrt{1 + \frac{V_x}{\phi_0}}} = A \cdot \sqrt{\frac{q \epsilon_{Si} N_A}{2(\phi_0 + V_x)}} \quad \dots\dots\dots(18)$$

$$\phi_0 = \frac{kT}{q} \ln \left(\frac{N_D \cdot N_A}{n_i^2} \right) \quad \dots\dots\dots(19)$$

$$C_{j0} = \sqrt{\frac{q \epsilon_{Si} N_A N_D}{2(N_A + N_D) \phi_0}} \approx \sqrt{\frac{q \epsilon_{Si} N_A}{2 \phi_0}} \quad \dots\dots\dots(20)$$

Charge Storage and Charge Leakage

- To estimate the actual charge leakage time from the soft node, we have to solve the differential equation given in (Eqn .17), taking into account the voltage-dependent capacitance components and the nonlinear leakage currents.
- For a quick estimate of the worstcase leakage behavior, on the other hand, the problem can be further simplified.

Charge Storage and Charge Leakage

- Assume that the minimum combined soft-node capacitance is given as

$$C_{x,min} = C_{gb} + C_{poly} + C_{metal} + C_{db,min} \quad \text{.....(21)}$$

Charge Storage and Charge Leakage

- where $C_{db,min}$ represents the minimum junction capacitance, obtained under the bias condition $V_x = V_{max}$.
- Now we define the worst-case holding time (t_{hold}) as the shortest time required for the soft-node voltage to drop from its initial logic-high value to the logic threshold voltage due to leakage. Once the soft-node voltage reaches the logic threshold, the logic stage being driven by this node will lose its previously held state.

Charge Storage and Charge Leakage

$$t_{hold} = \frac{\Delta Q_{critical,min}}{I_{leakage,max}} \quad \dots\dots(22)$$

$$\Delta Q_{critical,min} = C_{x,min} \left(V_{max} - \frac{V_{DD}}{2} \right) \quad \dots\dots(23)$$

Synchronous Dynamic Circuit Techniques

- Here, we will investigate different examples of synchronous dynamic circuits implemented using depletion-load nMOS, enhancement-load nMOS, and CMOS building blocks.

Dynamic Pass Transistor Circuits

- Consider the generalized view of a multi-stage synchronous circuit shown in Fig. 12.
- The circuit consists of cascaded combinational logic stages, which are interconnected through nMOS pass transistors.
- All inputs of each combinational logic block are driven by a single clock signal.
- Individual input capacitances are not shown in this figure for simplicity, but the operation of the circuit depends on temporary charge storage in the parasitic input capacitances.

Dynamic Pass Transistor Circuits

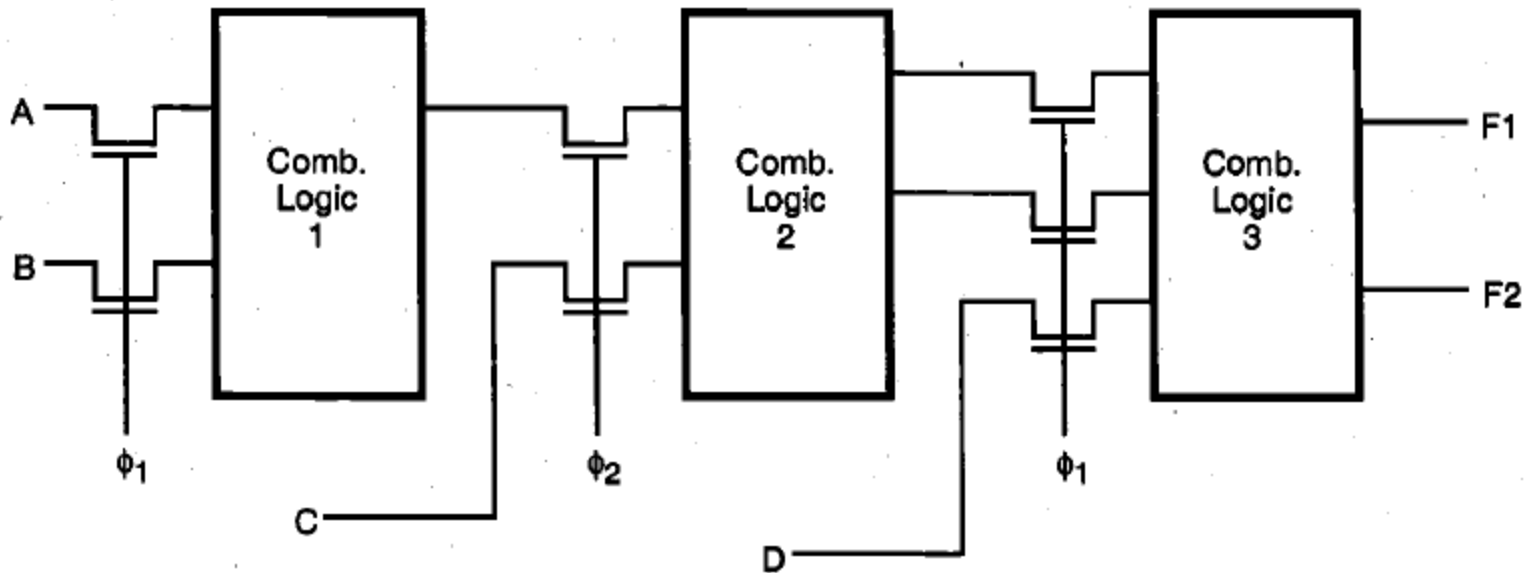


Figure 12. Multi-stage pass transistor logic driven by two nonoverlapping clocks.

Dynamic Pass Transistor Circuits

- To drive the pass transistors in this system, two nonoverlapping clock signals, ϕ_1 and ϕ_2 , are used.
- The nonoverlapping property of the two clock signals guarantees that at any given time point, only one of the two clock signals can be active, as illustrated in Fig. 13.

Dynamic Pass Transistor Circuits

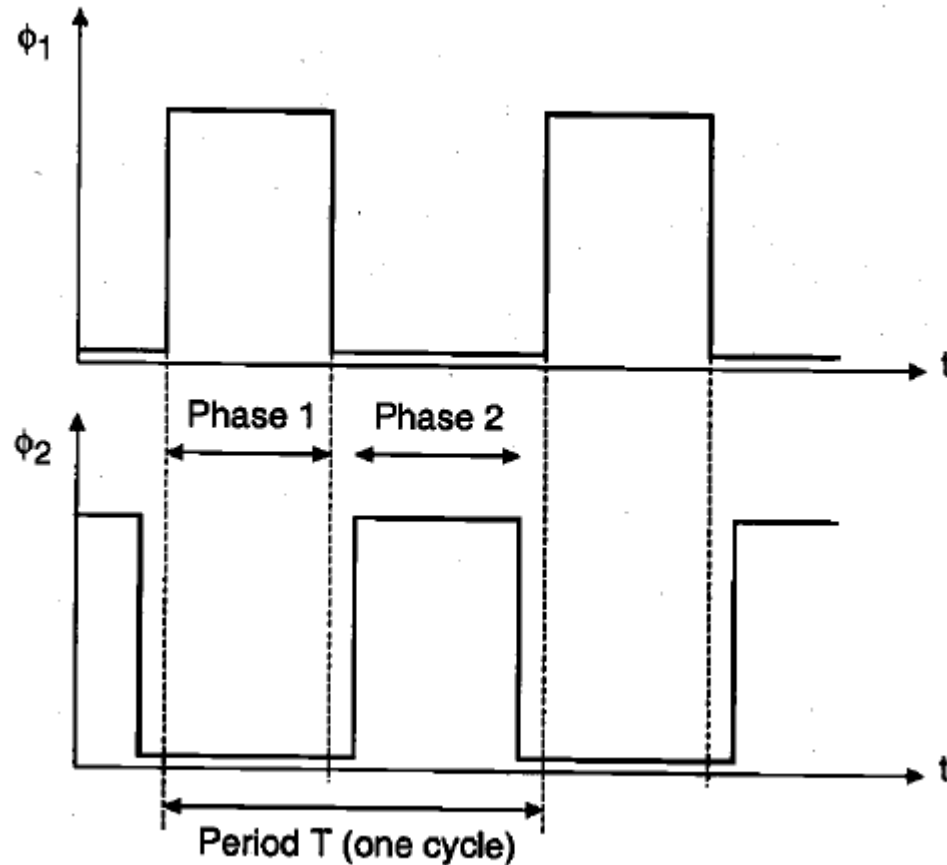


Figure 13. Nonoverlapping clock signals used for two-phase synchronous operation

Dynamic Pass Transistor Circuits

- When clock ϕ_1 , is active, the input levels of Stage 1 (and also of Stage 3) are applied through the pass transistors, while the input capacitances of Stage 2 retain their previously set logic levels. During the next phase, when clock ϕ_2 is active, the input levels of Stage 2 will be applied through the pass transistors, while the input capacitances of Stage 1 and Stage 3 retain their logic levels.

Dynamic Pass Transistor Circuits

- This allows us to incorporate the simple dynamic memory function at each stage input, and at the same time, to facilitate synchronous operation by controlling the signal flow in the circuit using the two periodic clock signals.
- This signal timing scheme is also called two-phase clocking and is one of the most widely used timing strategies.

Dynamic Pass Transistor Circuits

- Figure 14 shows a depletion-load dynamic shift register circuit, in which the input data are inverted once and transferred, or shifted into the next stage during each clock phase.

Dynamic Pass Transistor Circuits

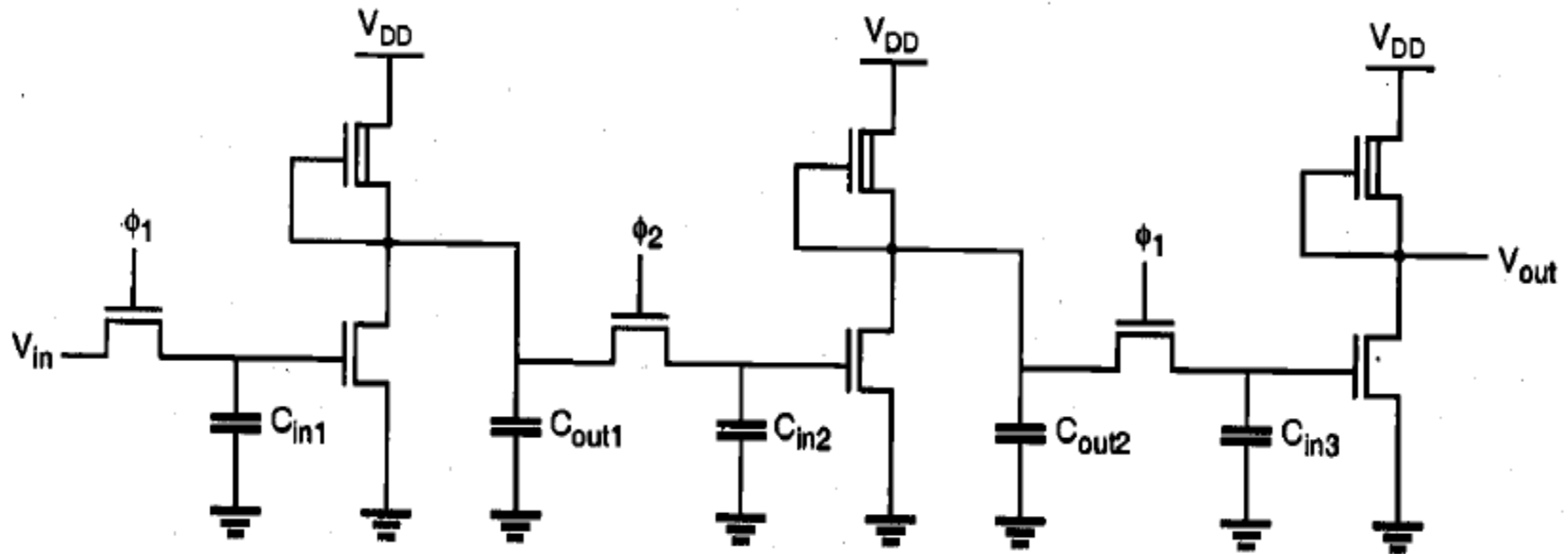


Figure 14. Three stages of a depletion-load nMOS dynamic shift register circuit driven with two-phase clocking.

Dynamic Pass Transistor Circuits

- The operation of the shift register circuit is as follows. During the active phase of ϕ_1 the input voltage level V_{in} is transferred into the input capacitance C_{in1} .
- Thus, the valid output voltage level of the first stage is determined as the inverse of the current input during this cycle.

Dynamic Pass Transistor Circuits

- When ϕ_2 becomes active during the next phase, the output voltage level of the first stage is transferred into the second stage input capacitance C_{in2} , and the valid output voltage level of the second stage is determined.
- During the active ϕ_2 phase, the first-stage input capacitance continues to retain its previous level via charge storage.

Dynamic Pass Transistor Circuits

- When ϕ_1 becomes active again, the original data bit written into the register during the previous cycle is transferred into the third stage, and the first stage can now accept the next data bit.

Dynamic Pass Transistor Circuits

- The same operation principle used in the simple shift register circuit can easily be extended to synchronous complex logic.
- Figures 15 and 16 show a two-stage circuit example implemented using depletion-load nMOS complex logic gates.

Dynamic Pass Transistor Circuits

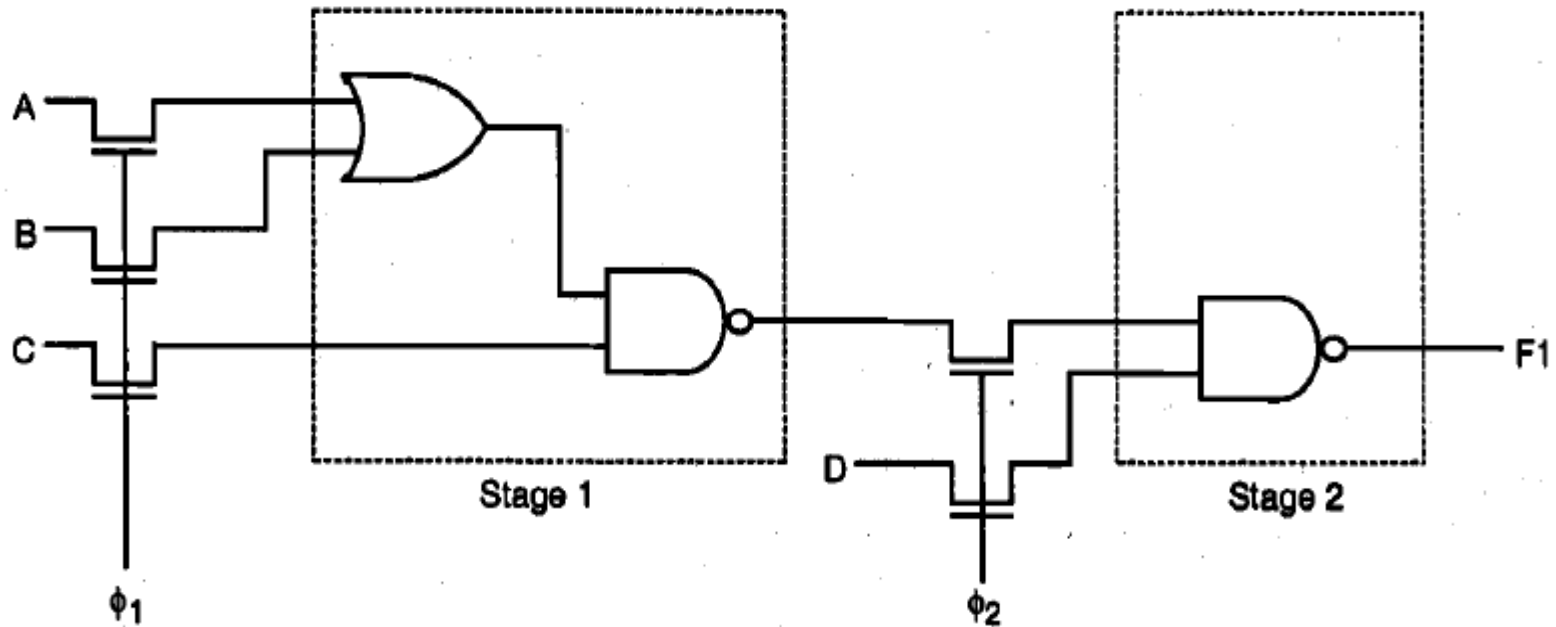


Figure 15. A two-stage synchronous complex logic circuit example

Dynamic Pass Transistor Circuits

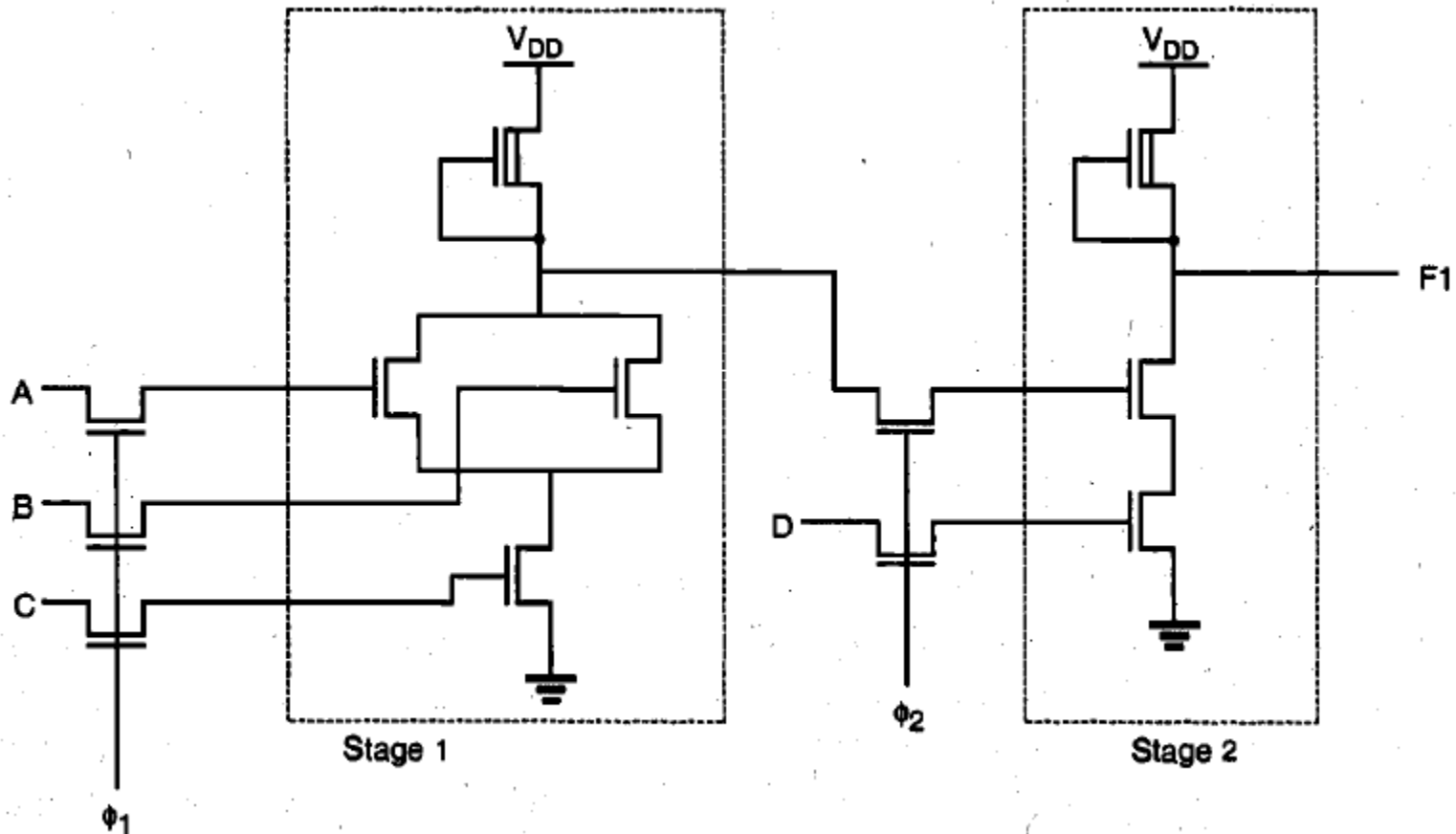


Figure 16. Depletion-load nMOS implementation of synchronous complex logic

Dynamic Pass Transistor Circuits

- In a complex logic circuit such as the one shown in Fig. 16, we see that the signal propagation delay of each stage may be different.
- Thus, in order to guarantee that correct logic levels are propagated during each active clock cycle, the half-period length of the clock signal must be longer than the largest single-stage signal propagation delay found in the circuit.

Dynamic Pass Transistor Circuits

- Now let us consider a different implementation of the simple shift register circuit, using enhancement-load nMOS inverters.
- One important difference is that, instead of biasing the load transistors with a constant gate voltage, we apply the clock signal to the gate of the load transistor.

Dynamic Pass Transistor Circuits

- It can be shown that the power dissipation and the silicon area can be reduced significantly by using this dynamic (clocked) load approach.
- Two variants of the dynamic enhancement-load shift register will be examined in the following, both of which are driven by two non-overlapping clock signals.
- Figure 17 shows the first implementation, where in each stage the input pass transistor and the load transistor are driven by opposite clock phases, ϕ_1 and ϕ_2 .

Dynamic Pass Transistor Circuits

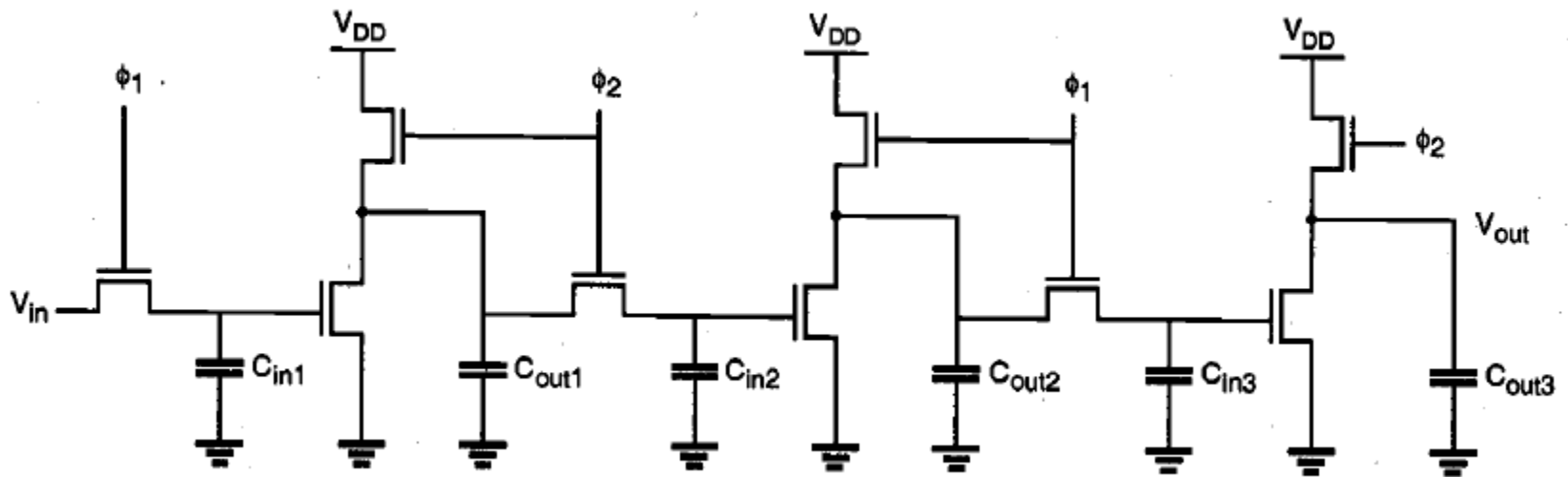


Figure 17. Enhancement-load dynamic shift register (ratioed logic).

Dynamic Pass Transistor Circuits

- When ϕ_1 is active, the input voltage level V_{in} is transferred into the first-stage input capacitance C_{in1} through the pass transistor.
- In this phase, the enhancement-type nMOS load transistor of the first-stage inverter is not active yet.
- During the next phase (active ϕ_2), the load transistor is turned on.

Dynamic Pass Transistor Circuits

- Since the input logic level is still being preserved in C_{in1} , the output of the first inverter stage attains its valid logic level.
- At the same time, the input pass transistor of the second stage is also turned on, which allows this newly determined output level to be transferred into the input capacitance C_{in2} of the second stage.

Dynamic Pass Transistor Circuits

- When clock ϕ_1 becomes active again, the valid output level across C_{out2} is determined, and transferred into C_{in3} .
- Also, a new input level can be accepted (pipelined) into C_{in} , during this phase.

Dynamic Pass Transistor Circuits

- The basic operation principle can be extended to arbitrary complex logic, as shown in Fig. 18.
- Since the power supply current flows only when the load devices are activated by the clock signal, the overall power consumption of dynamic enhancement-load logic is generally lower than for depletion-load nMOS logic.

Dynamic Pass Transistor Circuits

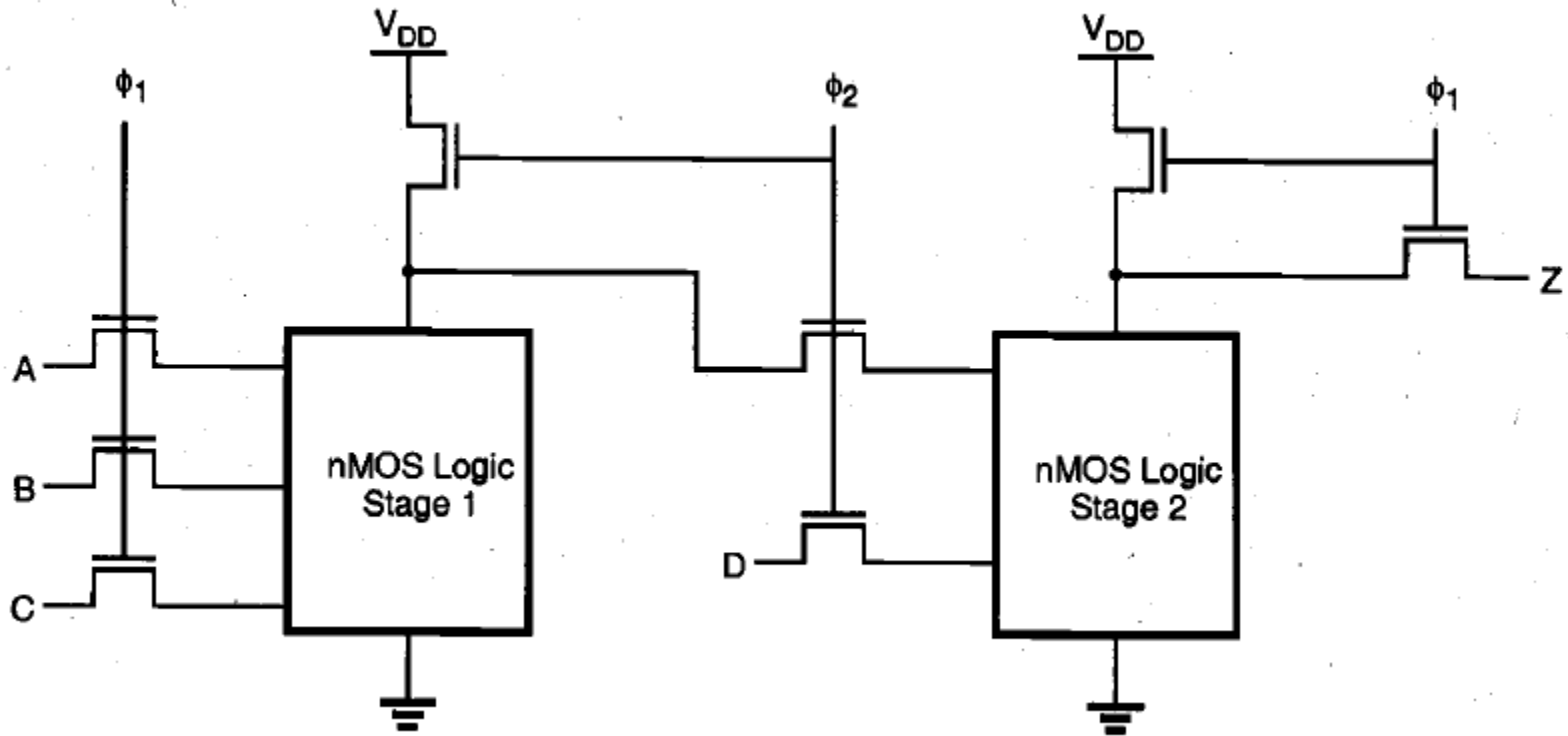


Figure 18. General circuit structure of ratioed synchronous dynamic logic

Dynamic Pass Transistor Circuits

- Next, let us consider the second dynamic enhancement-load shift register implementation where, in each stage, the input pass transistor and the load transistor are driven by the same clock phase (Fig. 19).

Dynamic Pass Transistor Circuits

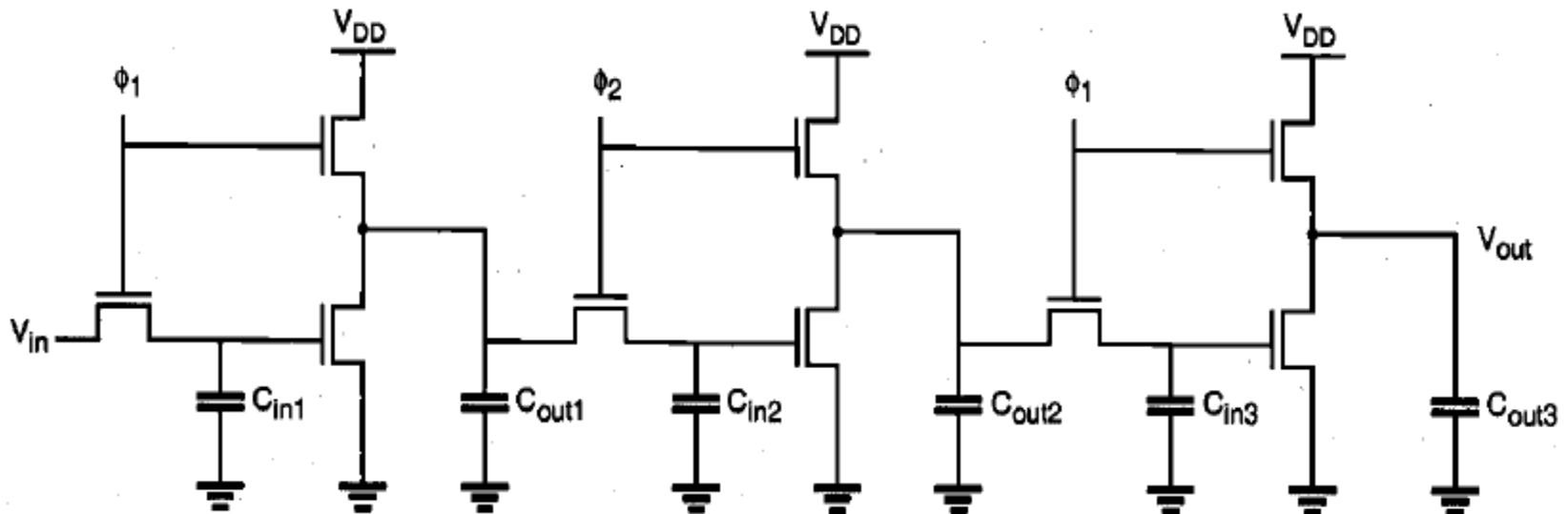


Figure 19. Enhancement-load dynamic shift register (ratioless logic).

Dynamic Pass Transistor Circuits

- When ϕ_1 is active, the input voltage level V_{in} is transferred into the first-stage input capacitance C_{in1} , through the pass transistor.
- Note that at the same time, the enhancement type nMOS load transistor of the first-stage inverter is active.
- Therefore, the output of the first inverter stage attains its valid logic level.

Dynamic Pass Transistor Circuits

- During the next phase (active ϕ_2), the input pass transistor of the next stage is turned on, and the logic level is transferred onto the next stage.
- Here, we have to consider two cases, as follows.

Dynamic Pass Transistor Circuits

- If the output level across C_{out1} is logic-high at the end of the active ϕ_1 phase, this voltage level is transferred to C_{in2} via charge sharing over the pass transistor during the active ϕ_2 phase.
- Note that the logic-high level at the output node is subject to threshold voltage drop, i.e., it is one threshold voltage lower than the power supply voltage.

Dynamic Pass Transistor Circuits

- To correctly transfer a logic-high level after charge sharing, the ratio of the capacitors ($C_{\text{out}}/C_{\text{in}}$) must be made large enough during circuit design.
- If, on the other hand, the output level of the first stage is logic-low at the end of the active ϕ_1 phase, then the output capacitor C_{out1} will be completely drained to a voltage of $V_{\text{OL}} = 0 \text{ V}$ when ϕ_1 turns off.

Dynamic Pass Transistor Circuits

- This can be achieved because a logic-high level is being stored in the input capacitance C_{in1} in this case, which forces the driver transistor to remain in conduction.
- The logic-low level of $V_{OL} = 0 \text{ V}$ is also transferred into the next stage via the pass transistor during the active ϕ_2 phase.

Dynamic Pass Transistor Circuits

- When clock ϕ_1 , becomes active again, the valid output level across C_{out2} is determined and transferred into C_{in3} .
- Also, a new input level can be accepted into C_{in1} , during this phase.
- Since the valid logic-low level of $V_{OL} = 0\text{ V}$ can be achieved regardless of the driver-to-load ratio, this circuit arrangement is called ratioless dynamic logic.
- The basic operation principle can be extended to arbitrary complex logic, as shown in Fig. 20.

Dynamic Pass Transistor Circuits

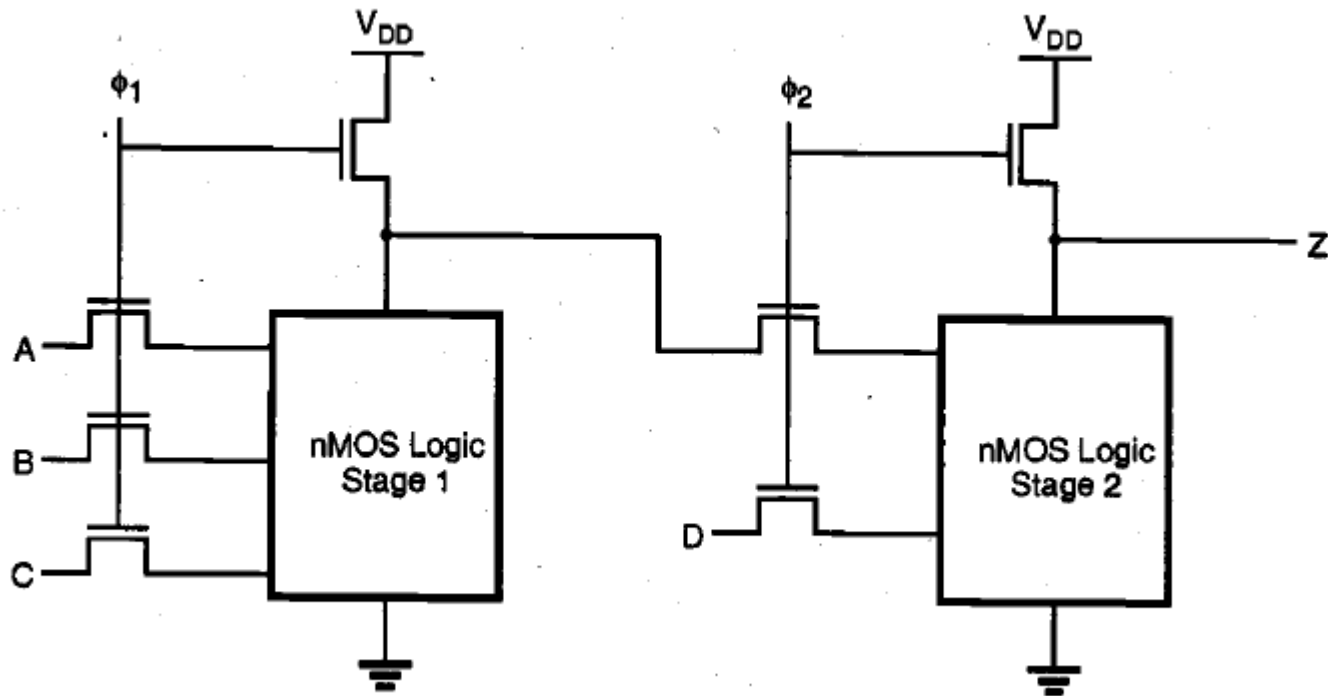


Figure 20. General circuit structure of ratioless synchronous dynamic logic.

CMOS Transmission Gate Logic

- Here, static CMOS gates are used for implementing the logic blocks, and CMOS transmission gates are used for transferring the output levels of one stage to the inputs of the next stage (Fig. 21).
- Each transmission gate is controlled by the clock signal and its complement.
- As a result, two-phase clocking in CMOS transmission gate logic requires that a total of four clock signals are generated and routed throughout the circuit.

CMOS Transmission Gate Logic

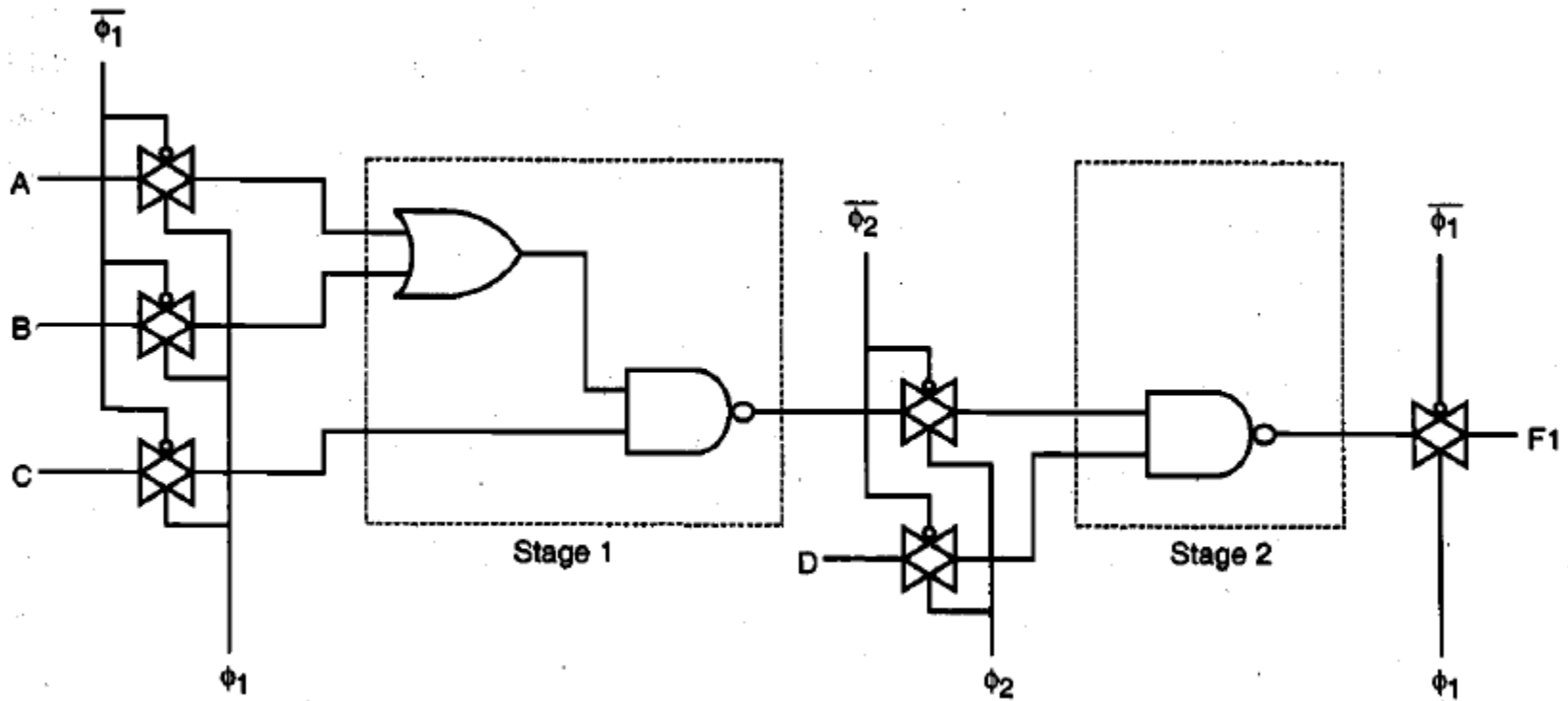


Figure 21. Typical example of dynamic CMOS transmission gate logic.

CMOS Transmission Gate Logic

- To illustrate the basic operation principles, the fundamental building block of a dynamic CMOS transmission gate shift register is shown in Fig. 22.
- It consists of a CMOS inverter, which is driven by a CMOS transmission gate.
- During the active clock phase ($CK = 1$), the input voltage V_{in} is transferred onto the parasitic input capacitance C_x via the transmission gate.

CMOS Transmission Gate Logic

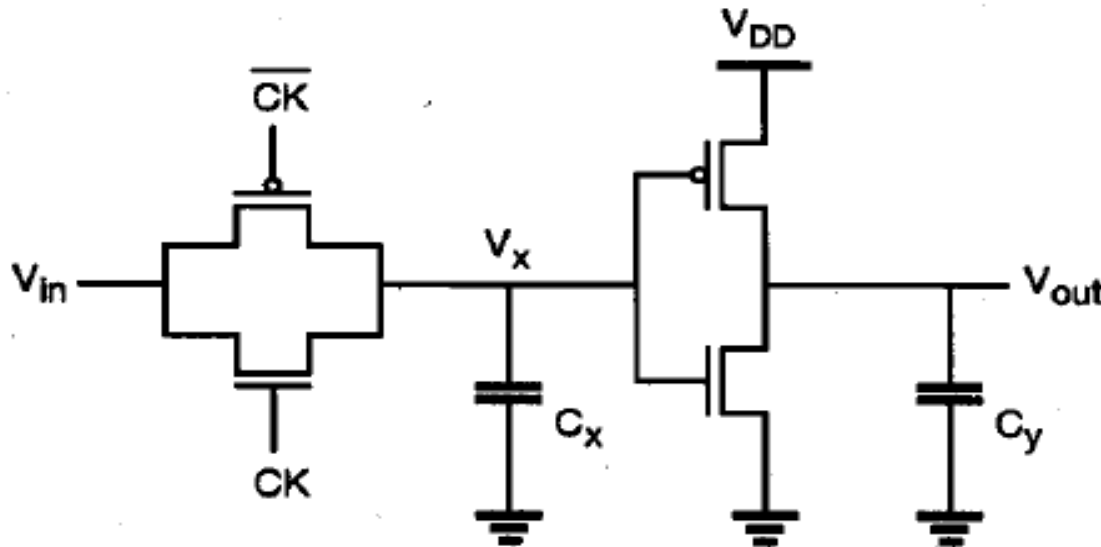


Figure 22. Basic building block of a CMOS transmission gate dynamic shift register.

CMOS Transmission Gate Logic

- The low on-resistance of the CMOS transmission gate usually results in a smaller transfer time compared to those for nMOS-only switches.
- Also, there is no threshold voltage drop across the CMOS transmission gate.
- When the clock signal becomes inactive, the CMOS transmission gate turns off and the voltage level across C_x can be preserved until the next cycle.

CMOS Transmission Gate Logic

- Figure 23 shows a single-phase CMOS shift register, which is built by cascading identical units as in Fig. 22 and by driving each stage alternately with the clock signal and its complement.

CMOS Transmission Gate Logic

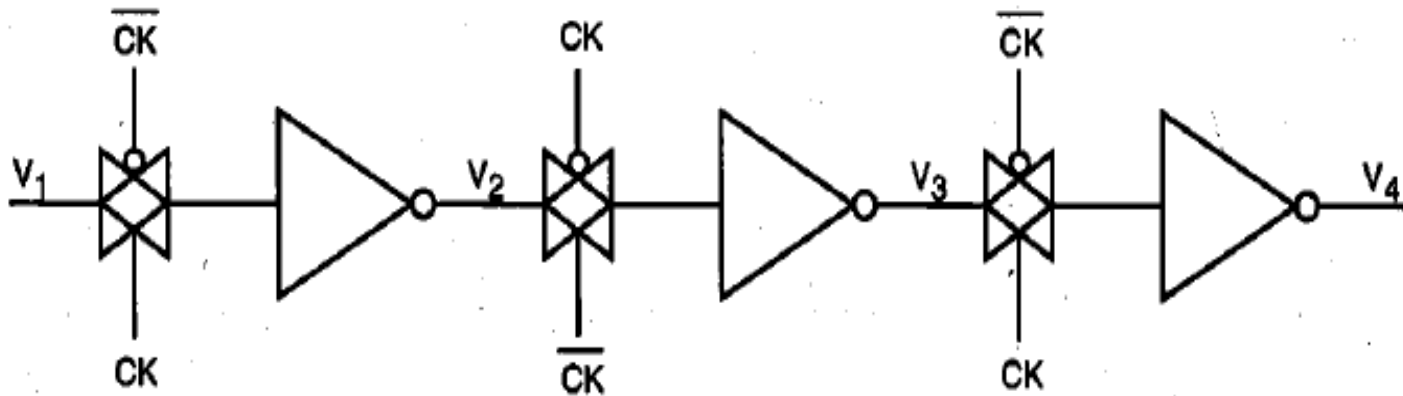


Figure 23. Single-phase CMOS transmission gate dynamic shift register.

CMOS Transmission Gate Logic

- Ideally, the transmission gates of the odd-numbered stages would conduct during the active clock phase (when $CK = 1$), while the transmission gates of the even-numbered stages are off, so that the cascaded inverter stages in the chain are alternately isolated.
- This would ensure that inputs are permitted in alternating half cycles.

CMOS Transmission Gate Logic

- In practice, the clock signal and its complement do not constitute a truly nonoverlapping signal pair, since the clock voltage waveform has finite rise and fall times.
- Also, the clock skew between CK and $\overline{CK}$ may be unavoidable because one of the signals is generated by inverting the other.
- Therefore, true two-phase clocking with two nonoverlapping clock signals (ϕ_1 and ϕ_2) and their complements is usually preferred over single-phase clocking in dynamic CMOS transmission gate logic.

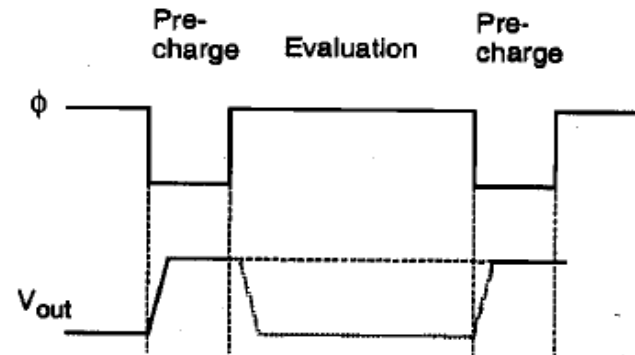
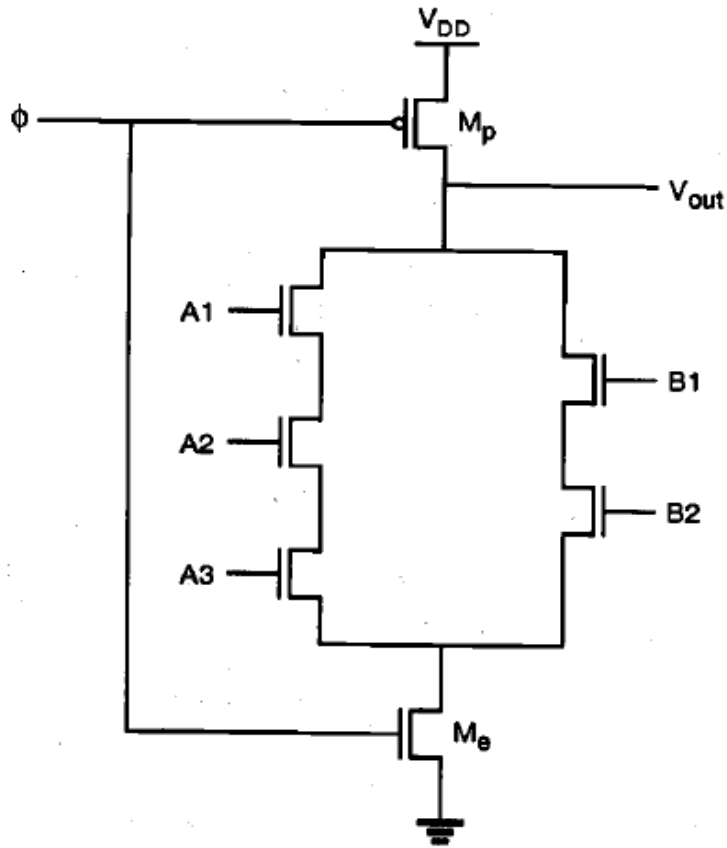
Dynamic CMOS Logic (Precharge-Evaluate Logic)

- Now, we will introduce a dynamic CMOS circuit technique which allows us to significantly reduce the number of transistors used to implement any logic function.
- The circuit operation is based on first precharging the output node capacitance and evaluating the output level according to the applied inputs.
- Both of these operations are scheduled by a single clock signal, which drives one nMOS and one pMOS transistor in each dynamic stage.

Dynamic CMOS Logic (Precharge-Evaluate Logic)

- A dynamic CMOS logic gate which implements the function $F = \overline{(A_1 A_2 A_3 + B_1 B_2)}$ is shown in Fig. 24.
- When the clock signal is low (precharge phase), the pMOS precharge transistor M_p is conducting, while the complementary nMOS transistor M_e is off.
- The parasitic output capacitance of the circuit is charged up through the conducting pMOS transistor to a logic-high level of $V_{out} = V_{DD}$.
- The input voltages are also applied during this phase, but they have no influence yet upon the output level since M_e is turned off.

Dynamic CMOS Logic (Precharge-Evaluate Logic)



Dynamic CMOS Logic (Precharge-Evaluate Logic)

- When the clock signal becomes high (evaluate phase), the precharge transistor M_p turns off and M_e turns on.
- The output node voltage may now remain at the logic- high level or drop to a logic low, depending on the input voltage levels.
- If the input signals create a conducting path between the output node and the ground, the output capacitance will discharge toward $V_{OL} = 0 \text{ V}$.

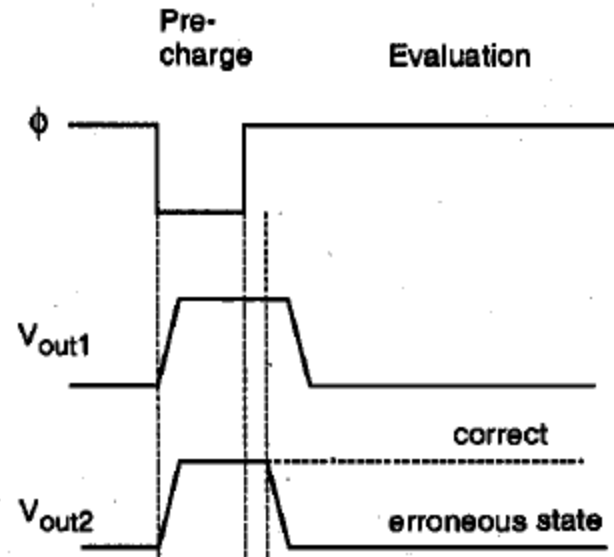
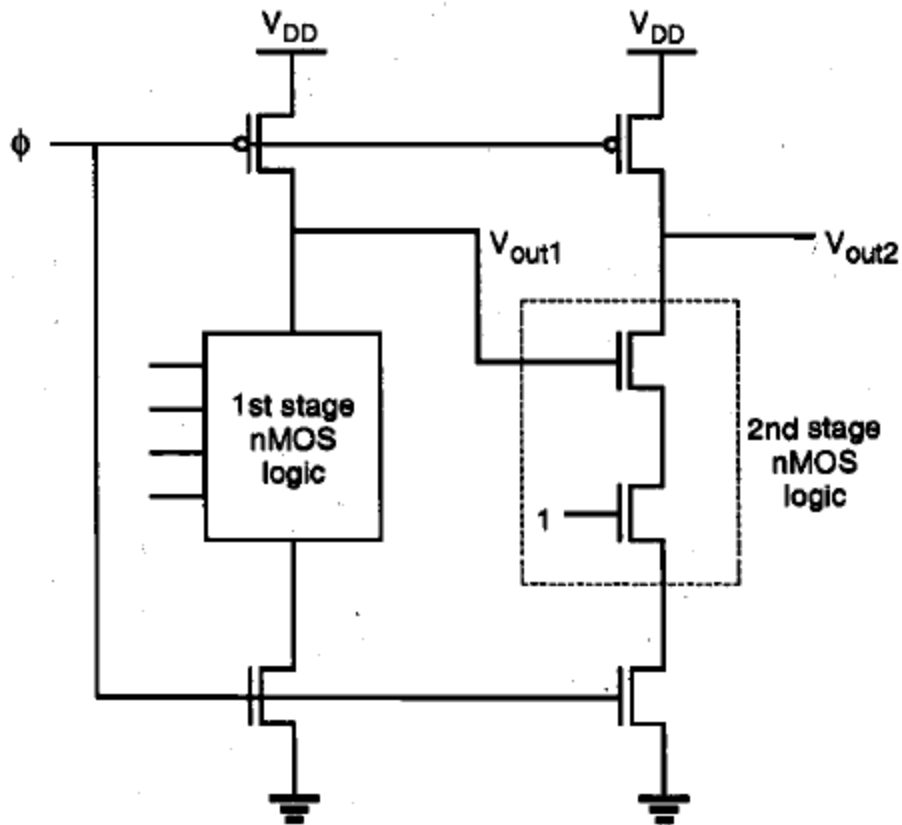
Dynamic CMOS Logic (Precharge-Evaluate Logic)

- The final discharged output level depends on the time span of the evaluation phase.
- Otherwise, V_{out} remains at V_{DD} .

Dynamic CMOS Logic (Precharge-Evaluate Logic)

- The operation of the single-stage dynamic CMOS logic gate is quite straightforward.
- For practical multi-stage applications, however, the dynamic CMOS gate presents a significant problem.
- To examine this fundamental limitation, consider the two-stage cascaded structure shown in Fig. 25.

Dynamic CMOS Logic (Precharge-Evaluate Logic)



Dynamic CMOS Logic (Precharge-Evaluate Logic)

- Here, the output of the first dynamic CMOS stage drives one of the inputs of the second dynamic CMOS stage, which is assumed to be a two input NAND gate for simplicity.

Dynamic CMOS Logic (Precharge-Evaluate Logic)

- During the precharge phase, both output voltages V_{out1} and V_{out2} are pulled up by the respective pMOS precharge devices.
- Also, the external inputs are applied during this phase.
- The input variables of the first stage are assumed to be such that the output V_{out1} will drop to logic "0" during the evaluation phase.
- On the other hand, the external input of the second-stage NAND2 gate is assumed to be a logic "1," as shown in Fig. 25.

Dynamic CMOS Logic (Precharge-Evaluate Logic)

- When the evaluation phase begins, both output voltages V_{out1} and V_{out2} are logic-high.
- The output of the first stage (V_{out1}) eventually drops to its correct logic level after a certain time delay.
- However, since the evaluation in the second stage is done concurrently, starting with the high value of V_{out1} at the beginning of the evaluation phase, the output voltage V_{out2} at the end of the evaluation phase will be erroneously low.

Dynamic CMOS Logic (Precharge-Evaluate Logic)

- Although the first stage output subsequently assumes its correct output value once the stored charge is drained, the correction of the second-stage output is not possible.

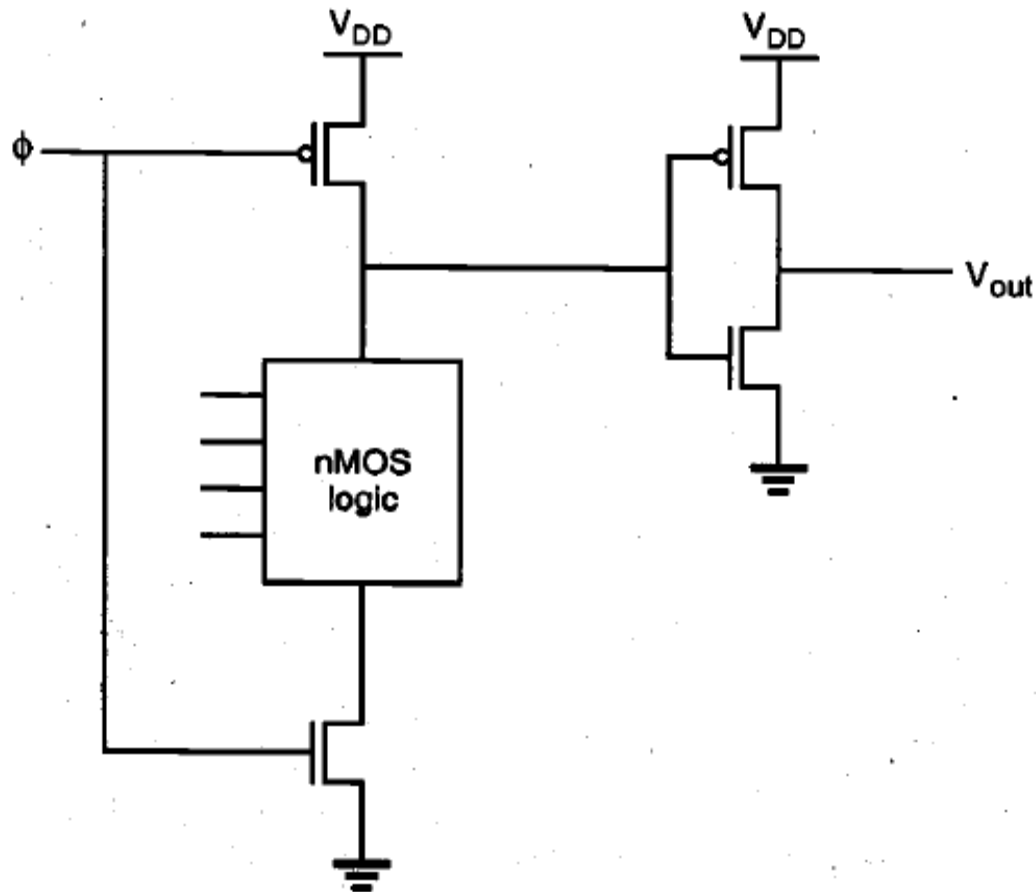
High-Performance Dynamic CMOS Circuits

- Here the ultimate goal is to achieve reliable, high-speed, compact circuits using the least complicated clocking scheme possible.

Domino CMOS Logic

- Consider the generalized circuit diagram of a domino CMOS logic gate shown in Fig. 26.
- A dynamic CMOS logic stage, such as the one shown in Fig. 24, is cascaded with a static CMOS inverter stage.
- The addition of the inverter allows us to operate a number of such structures in cascade.

Domino CMOS Logic



Domino CMOS Logic

- During the precharge phase (when $CK = 0$), the output node of the dynamic CMOS stage is precharged to a high logic level, and the output of the CMOS inverter (buffer) becomes low.
- When the clock signal rises at the beginning of the evaluation phase, there are two possibilities: The output node of the dynamic CMOS stage is either discharged to a low level through the nMOS circuitry (1 to 0 transition), or it remains high.

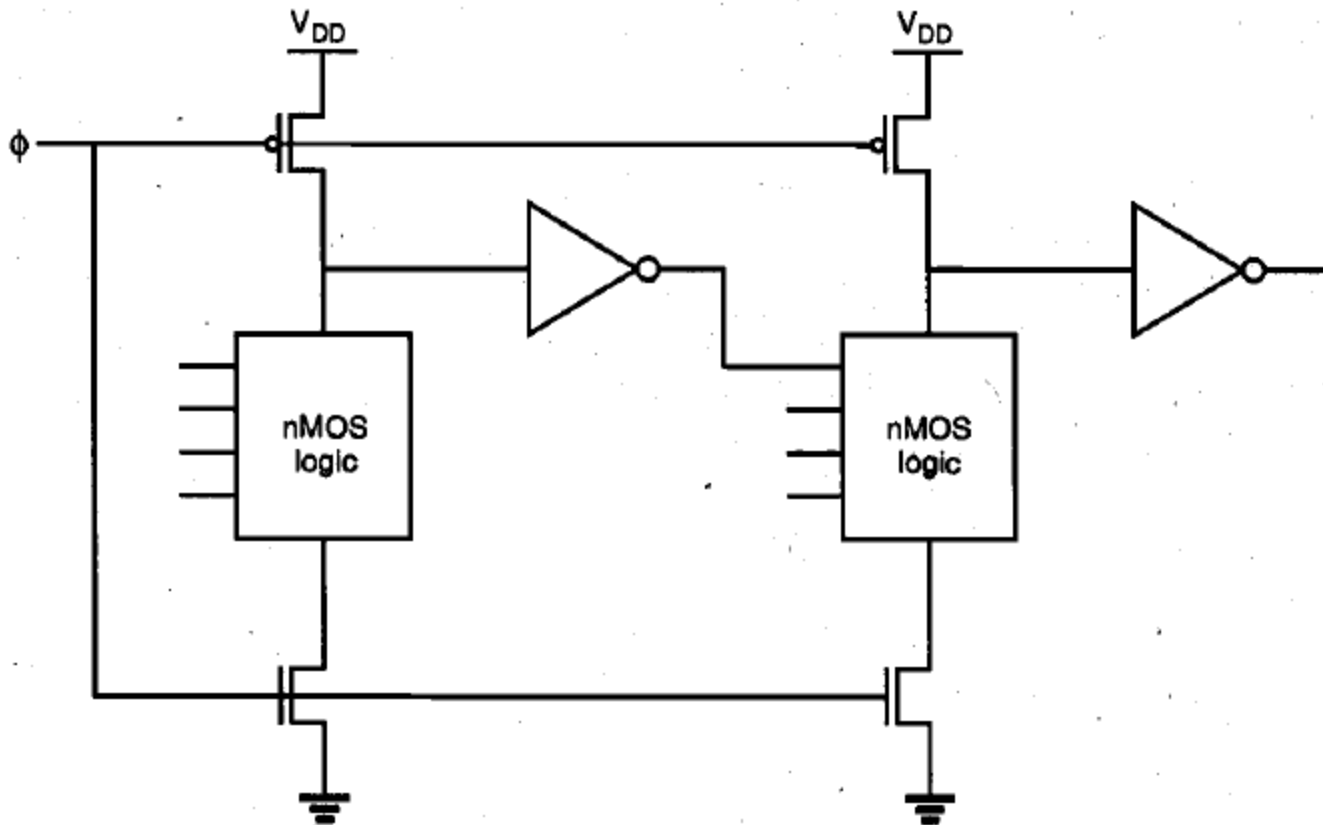
Domino CMOS Logic

- Consequently, the inverter output voltage can also make at most one transition during the evaluation phase, from 0 to 1.
- Regardless of the input voltages applied to the dynamic CMOS stage, it is not possible for the buffer output to make a 1 to 0 transition during the evaluation phase.

Domino CMOS Logic

- The problem in cascading conventional dynamic CMOS stages occurs when one or more inputs of a stage make a 1 to 0 transition during the evaluation phase, as illustrated in Fig. 25.
- If we build a system by cascading domino CMOS logic gates as shown in Fig. 27, all input transistors in subsequent logic blocks will be turned off during the precharge phase, since all buffer outputs are equal to 0.

Domino CMOS Logic



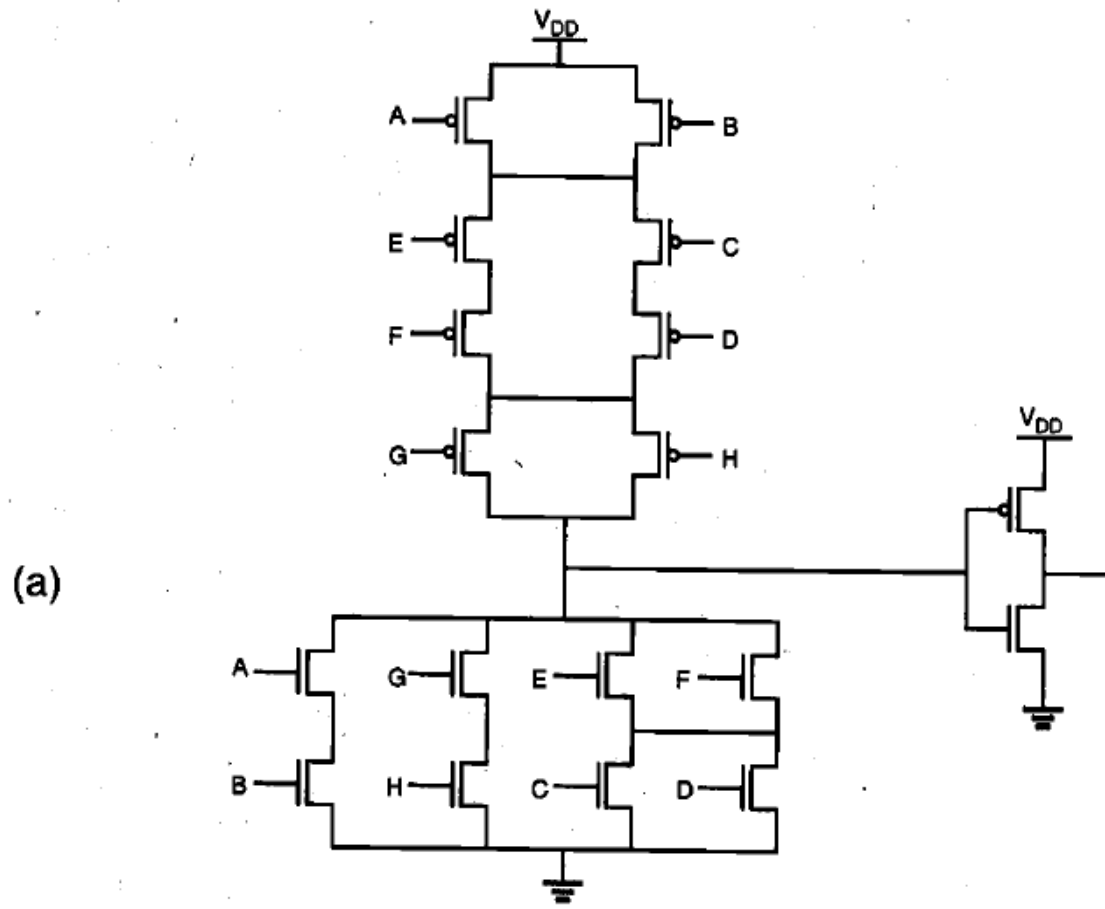
Domino CMOS Logic

- During the evaluation phase, each buffer output can make at most one transition (from 0 to 1), and thus each input of all subsequent logic stages can also make at most one (0 to 1) transition.
- In a cascade structure consisting of several such stages, the evaluation of each stage ripples the next stage evaluation, similar to a chain of dominos falling one after the other. The structure is hence called domino CMOS logic.

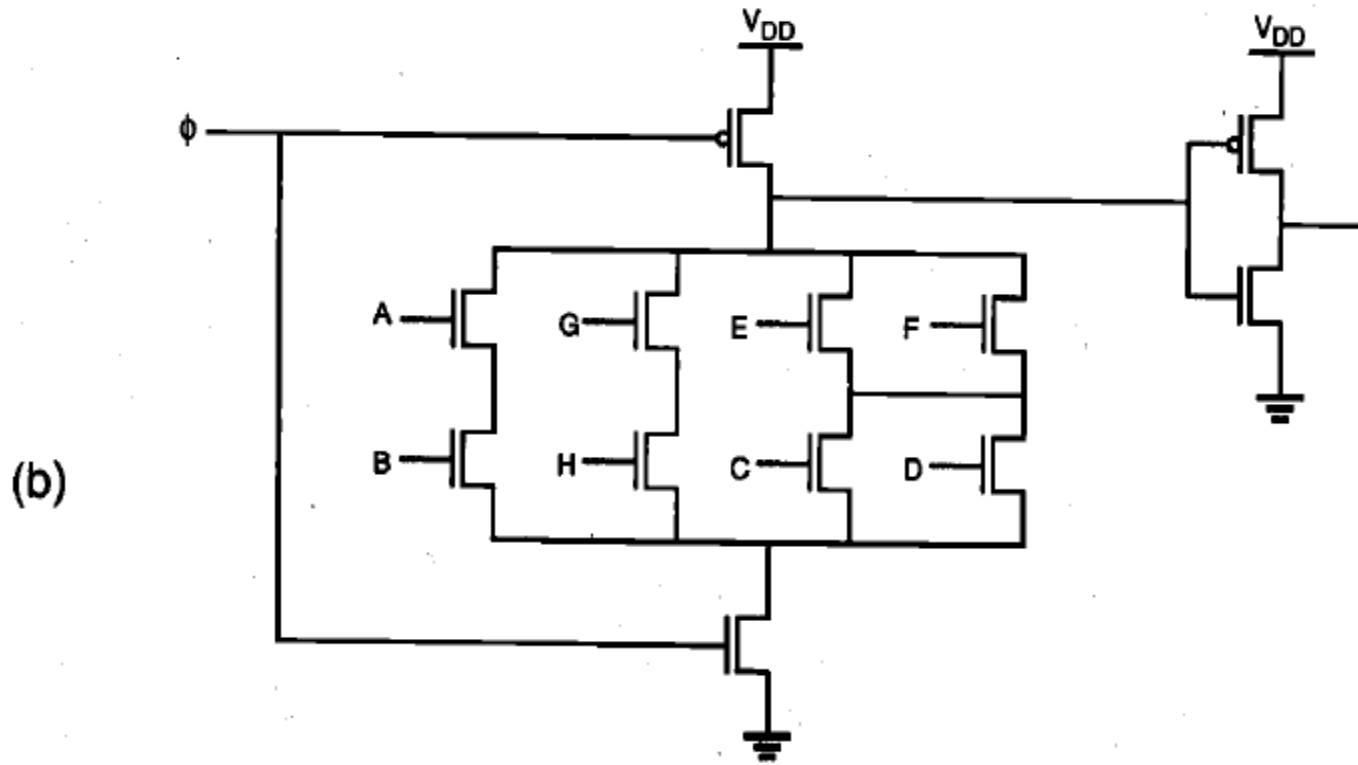
Domino CMOS Logic

- Domino CMOS logic gates allow a significant reduction in the number of transistors required to realize any complex Boolean function.
- The implementation of the 8-input Boolean function, $Z = AB + (C + D)(E + F) + GH$, using standard CMOS and domino CMOS, is shown in Fig. 28, where the reduction of circuit complexity is achieved.

Domino CMOS Logic



Domino CMOS Logic



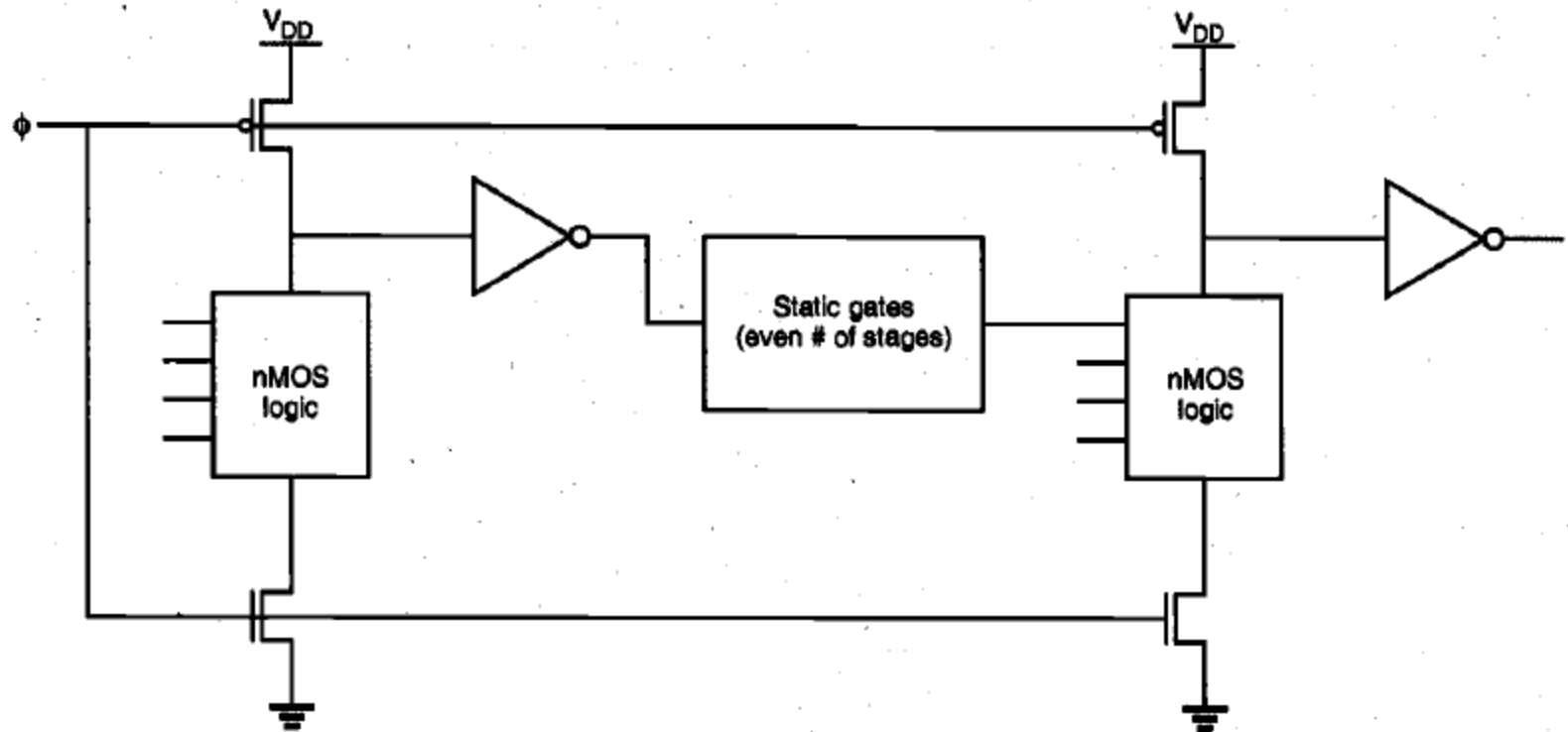
Domino CMOS Logic

- The distribution of the clock signal within the system is quite straightforward, since a single clock can be used to precharge and evaluate any number of cascaded stages, as long as the signal propagation delay from the first stage to the last stage does not exceed the time span of the evaluation phase.

Domino CMOS Logic

- Also, conventional static CMOS logic gates can be used together with domino CMOS gates in a cascaded configuration (Fig. 29).
- The limitation is that the number of inverting static logic stages in cascade must be even, so that the inputs of the next domino CMOS stage experience only 0 to 1 transitions during the evaluation.

Domino CMOS Logic



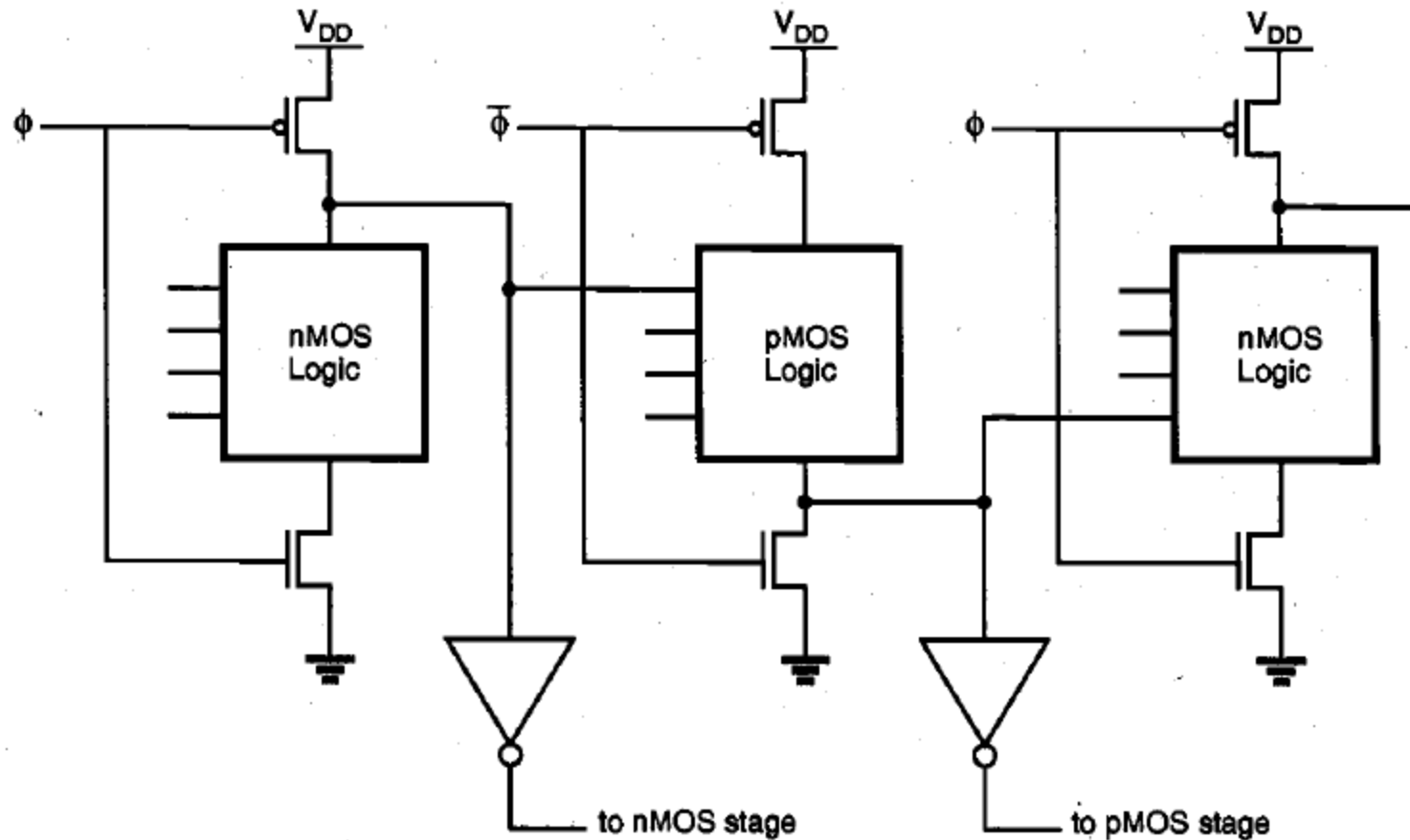
NORA CMOS Logic (NP-Domino Logic)

- In domino CMOS logic gates, all logic operations are performed by the nMOS transistors acting as pull-down networks, while the role of pMOS transistors is limited to precharging the dynamic nodes.
- As an alternative and a complement to nMOS-based domino CMOS logic, we can construct dynamic logic stages using pMOS transistors as well.

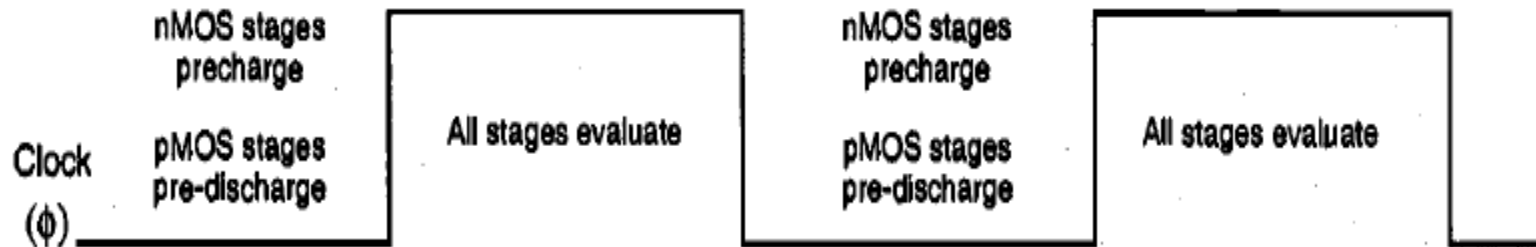
NORA CMOS Logic (NP-Domino Logic)

- Consider the circuit shown in Fig. 30, with alternating nMOS and pMOS logic stages.

NORA CMOS Logic (NP-Domino Logic)



NORA CMOS Logic (NP-Domino Logic)



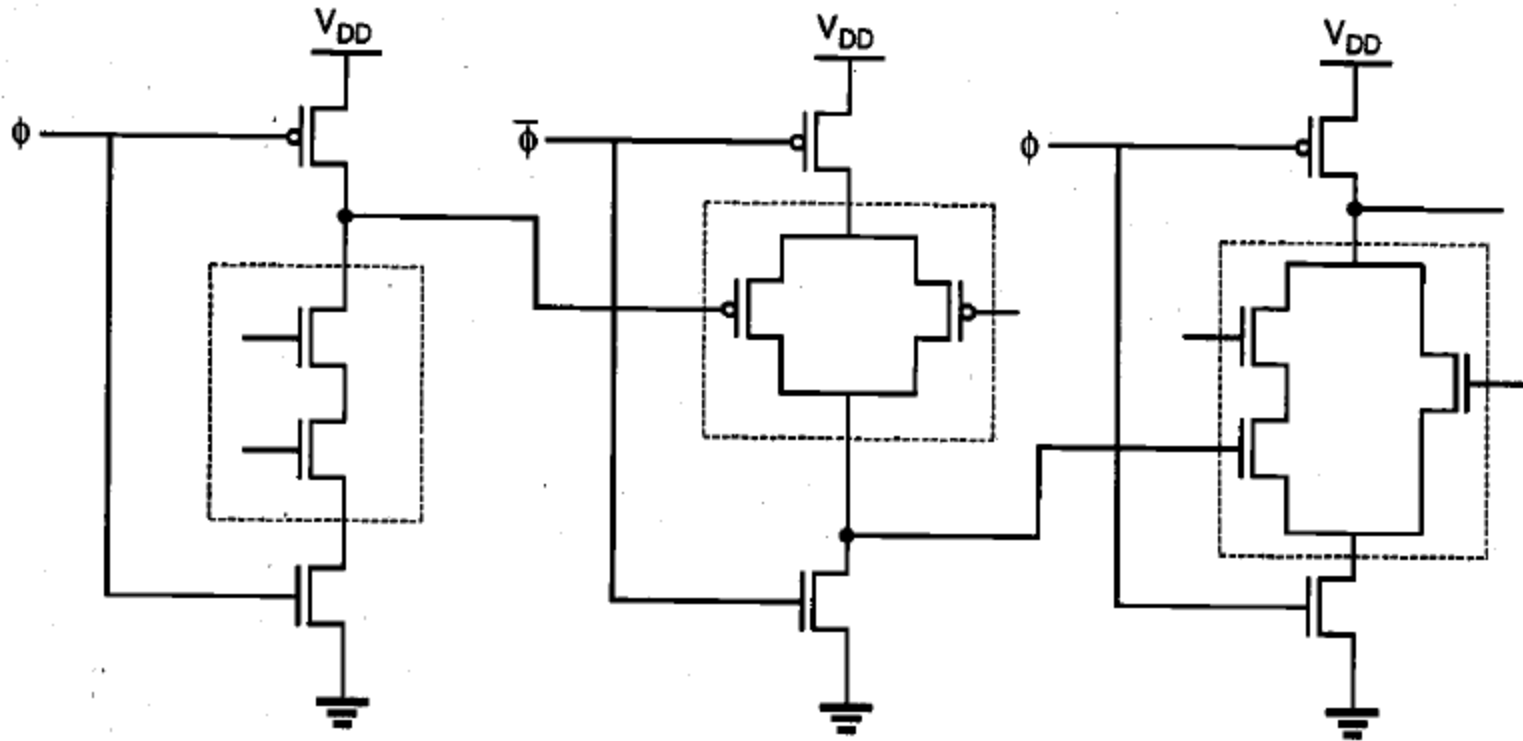
NORA CMOS Logic (NP-Domino Logic)

- The precharge-and-evaluate timing of nMOS logic stages is accomplished by the clock signal ϕ , whereas the pMOS logic stages are controlled by the inverted clock signal, $\bar{\phi}$.
- The operation of the NORA CMOS circuit is as follows:
- When the clock signal is low, the output nodes of nMOS logic blocks are precharged to V_{DD} through the pMOS precharge transistors,

NORA CMOS Logic (NP-Domino Logic)

- whereas the output nodes of pMOS logic blocks are pre-discharged to 0 V through the nMOS discharge transistors, driven by $\bar{\phi}$.
- When the clock signal makes a low-to-high transition (note that the inverted clock signal $\bar{\phi}$ makes a high-to-low transition simultaneously), all cascaded nMOS and pMOS logic stages evaluate one after the other, much like the domino CMOS examined earlier.
- A simple NORA CMOS circuit example is shown in Fig. 31.

NORA CMOS Logic (NP-Domino Logic)



NORA CMOS Logic (NP-Domino Logic)

- The advantage of NORA CMOS logic is that a static CMOS inverter is not required at the output of every dynamic logic stage.
- Instead, direct coupling of logic blocks is feasible by alternating nMOS and pMOS logic blocks.
- NORA logic is also compatible with domino CMOS logic.

NORA CMOS Logic (NP-Domino Logic)

- Outputs of NORA nMOS logic blocks can be inverted, and then applied to the input of a domino CMOS block, which is also driven by the clock signal ϕ .
- Similarly, the buffered output of a domino CMOS stage can be applied directly to the input of a NORA nMOS stage.

NORA CMOS Logic (NP-Domino Logic)

- The second important advantage of NORA CMOS logic is that it allows pipelined system architecture.
- Consider the circuit shown in Fig. 32(a), which consists of an nMOS-pMOS logic sequence similar to the one shown in Fig. 30, and a clocked CMOS (C²MOS) output buffer.

NORA CMOS Logic (NP-Domino Logic)

- It can easily be seen that all stages of this circuit perform the precharge-discharge operation when the clock is low, and all stages of the circuit evaluate output levels when the clock is high.
- Therefore, we will call this circuit a ϕ -section, meaning that evaluation occurs during active ϕ .

NORA CMOS Logic (NP-Domino Logic)

- Now consider the circuit shown in Fig. 32(b), which is essentially the same circuit shown in Fig. 32(a), with the only difference being that the signals ϕ and $\bar{\phi}$ have been interchanged.
- In this circuit, all logic stages perform the precharge-discharge operation when the clock is high, and all stages evaluate output levels when the clock is low.
- Therefore, we will call this circuit a $\bar{\phi}$ section, meaning that evaluation occurs during active $\bar{\phi}$.

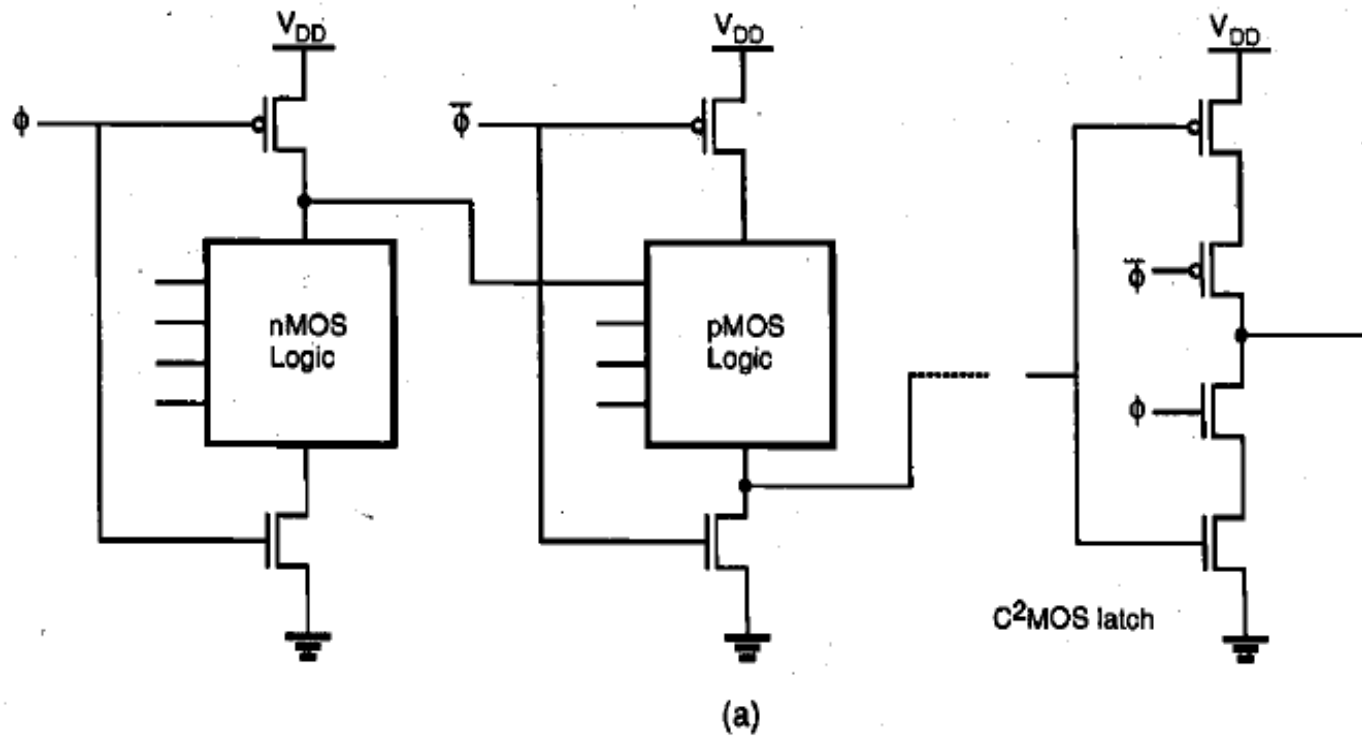
NORA CMOS Logic (NP-Domino Logic)

- A pipelined system can be constructed by simply cascading alternating ϕ and $\bar{\phi}$ sections, as shown in Fig. 32(c).
- Note that each of the sections may consist of several logic stages, and that all logic stages in one section are evaluated during the same clock cycle.
- When the clock is low, the ϕ -sections in the pipelined system undergo the precharge cycle, while the $\bar{\phi}$ sections undergo evaluation.

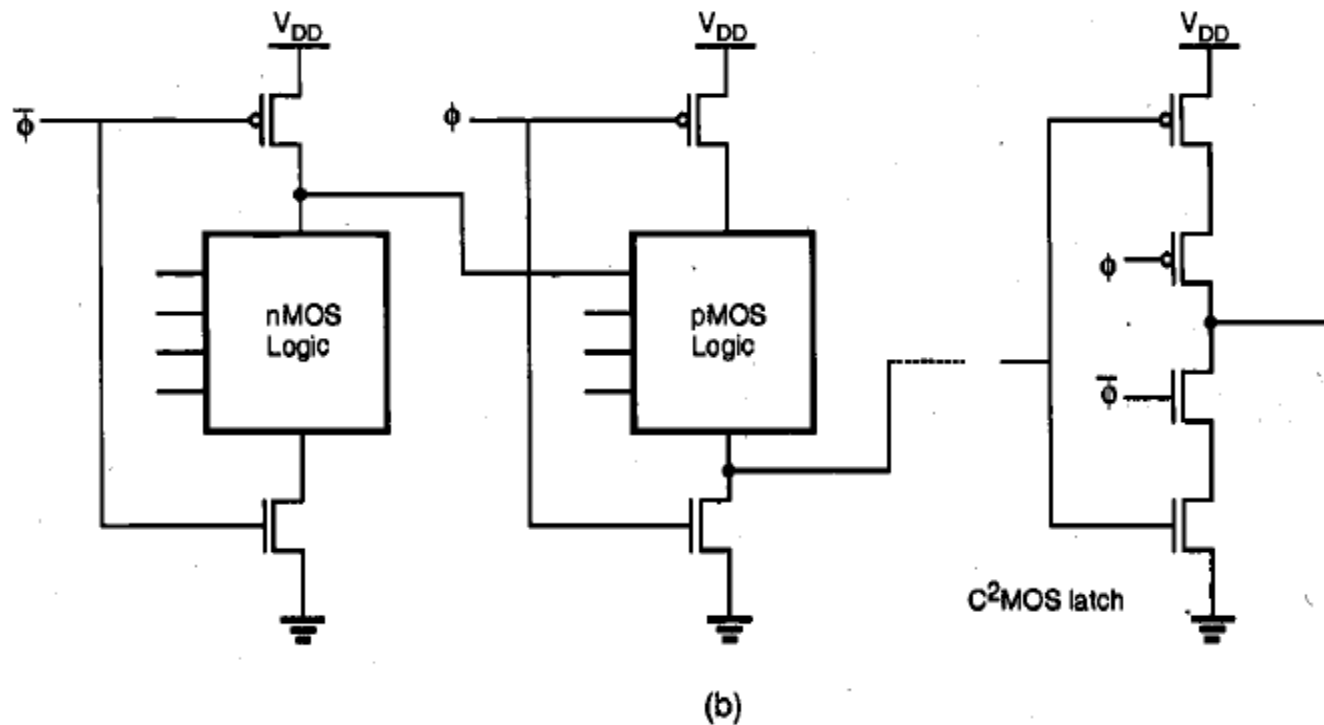
NORA CMOS Logic (NP-Domino Logic)

- When the clock signal changes from low to high, the ϕ -sections start the evaluation cycle, while the $\bar{\phi}$ sections undergo precharge.
- Thus, consecutive sets of input data can be processed in alternating sections of the pipelined system.

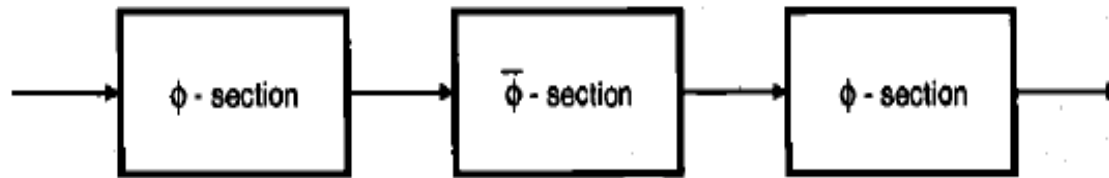
NORA CMOS Logic (NP-Domino Logic)



NORA CMOS Logic (NP-Domino Logic)



NORA CMOS Logic (NP-Domino Logic)



(c)

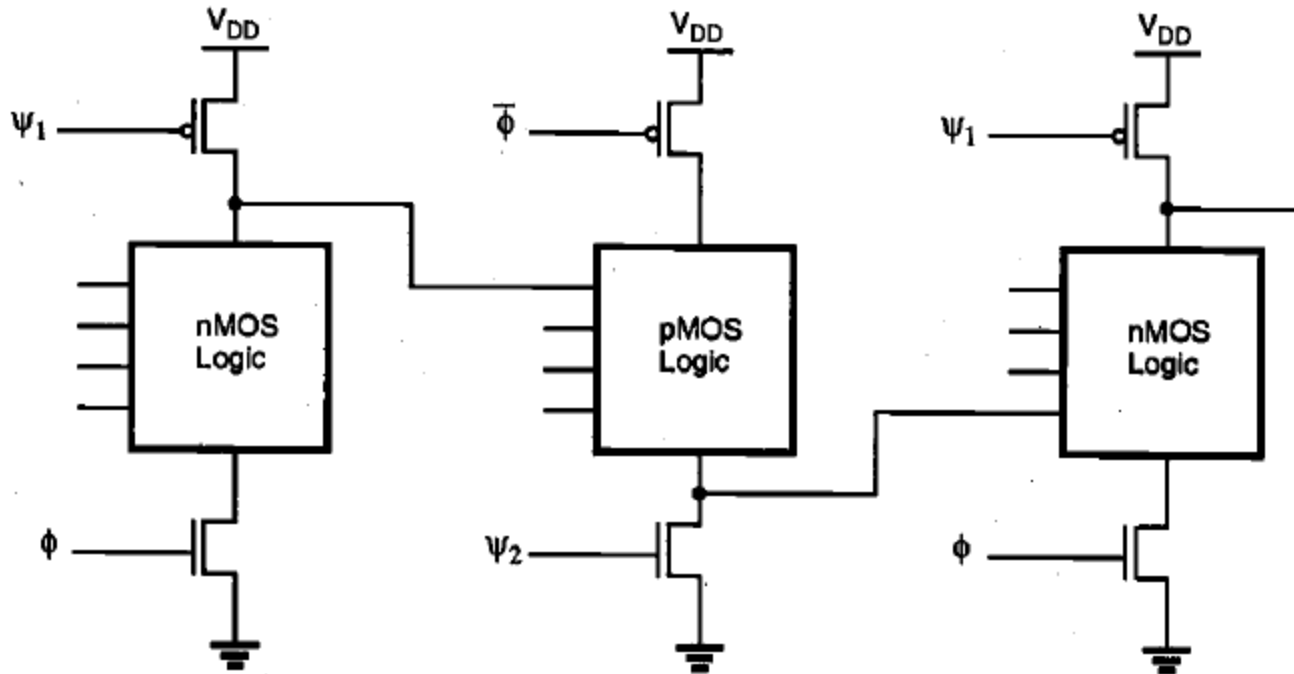
Zipper CMOS Circuits

- The basic circuit architecture of Zipper CMOS is essentially identical to NORA CMOS, with the exception of the clock signals.
- The Zipper CMOS clock scheme requires the generation of slightly different clock signals for the precharge (discharge) transistors and for the pull-down (pull-up) transistors.
- In particular, the clock signals which drive the pMOS precharge and nMOS discharge transistors allow these transistors to remain in

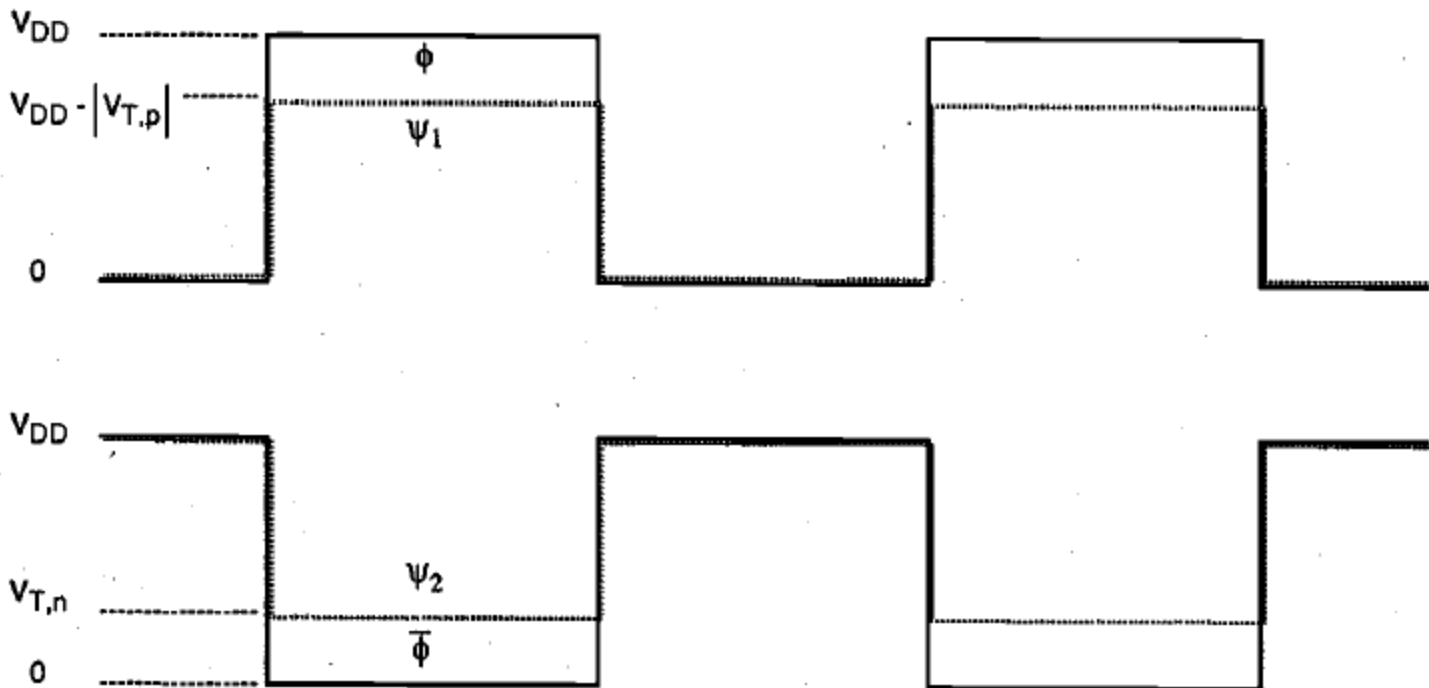
Zipper CMOS Circuits

- weak conduction or near cut-off during the evaluation phase, thus compensating for the charge leakage and charge-sharing problems.
- The generalized circuit diagram and the clock signals of the Zipper CMOS architecture are shown in Fig. 33.

Zipper CMOS Circuits



Zipper CMOS Circuits



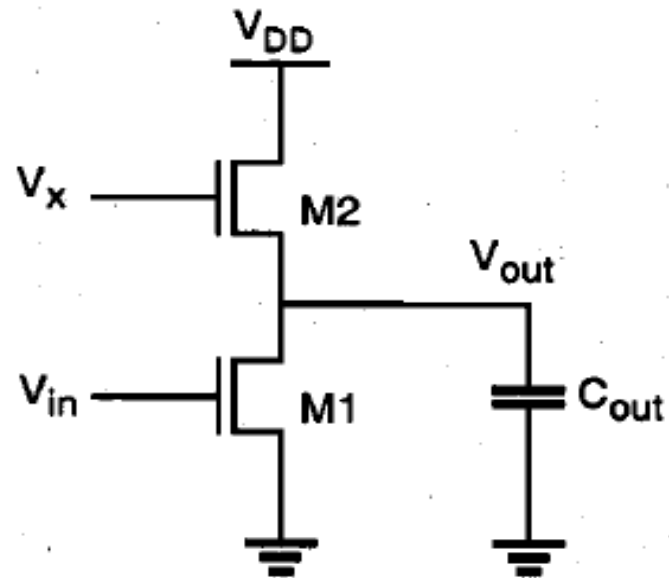
Voltage Bootstrapping

- In this section, we will briefly examine a very useful dynamic circuit technique for overcoming threshold voltage drops in digital circuits, which is called voltage bootstrapping.
- We have already seen that output voltage levels may suffer from threshold voltage drops in several circuit structures, such as pass transistor gates or enhancement load inverters and logic gates.

Voltage Bootstrapping

- Dynamic voltage bootstrapping techniques offer a simple yet effective way to overcome threshold voltage drops which occur in most situations.
- Consider the circuit shown in Fig. 34, where the voltage V_x is equal to or smaller than the power supply voltage, $V_x \leq V_{DD}$.
- Consequently, the enhancement-type nMOS transistor M_2 will operate in saturation.

Voltage Bootstrapping



Voltage Bootstrapping

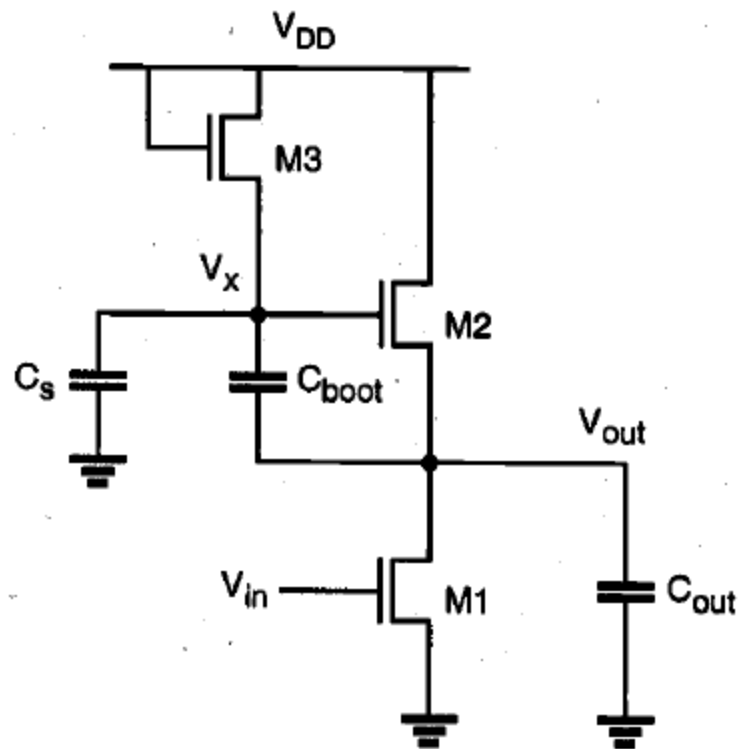
- When the input voltage V_{in} is low, the maximum value that the output voltage can attain is limited by

$$V_{out}(max) = V_x - V_{T2}(V_{out})$$

Voltage Bootstrapping

- To overcome the threshold voltage drop and to obtain a full logic-high level (V_{DD}) at the output node, the voltage V_x must be increased.
- Now consider the circuit shown in Fig. 35, where a third transistor M_3 has been added to the circuit.
- The two capacitors C_s and C_{boot} , seen in the circuit diagram represent the capacitances which dynamically couple the voltage V_x

Voltage Bootstrapping



Voltage Bootstrapping

- during switching, so that the threshold voltage drop can be overcome at the output node.

$$V_x \geq V_{DD} + V_{T2}(V_{out})$$

- Initially assume that the input voltage V_{in} is logic-high, so that M_1 and M_2 have a nonzero drain current and that the output voltage is low.

Voltage Bootstrapping

- At this point, M_1 is in the linear region and M_2 is in saturation.
- Since $I_{D3} = 0$, the initial condition for the voltage V_x can be found as

$$V_x = V_{DD} - V_{T3}(V_x)$$

Voltage Bootstrapping

- Now, assume that the input switches from its logic-high level to 0 V at $t = 0$. **As a**
- result, the driver transistor M1 will turn off and the output voltage **V_o will start to rise.**
- This change in the output voltage level will now be coupled to V through the bootstrap
- capacitor, C_{boot} . Let i_{Cboot} represent the transient current flowing through the capacitor

Voltage Bootstrapping

- *C_{boot} during this charge-up event, and let i be the current through C_s . Assuming that the*
- two current components are approximately equal, we obtain

Voltage Bootstrapping

- It can be seen from (9.29) that the increase in the output voltage V , during this switching
- event will generate a proportional increase in the voltage level V . Integrating both sides
- of (9.29), we obtain

Voltage Bootstrapping

- which proves that voltage
- bootstrapping can significantly boost the voltage level V . Now
- remember that in order to overcome the threshold voltage drop at the output, the minimum
- required voltage level V is

Voltage Bootstrapping

- Note that C_s is essentially the sum of the parasitic source-to-substrate capacitance of M3
- and the gate-to-substrate
- capacitance of M2. To obtain a sufficiently large bootstrap
- capacitance *C_{boot} in comparison to C_S an extra "dummy" transistor is typically added to*
- the circuit, as shown in Fig. 9.13.

Voltage Bootstrapping

- Since its drain and source terminals are connected together, the dummy transistor
- simply acts as an MOS capacitor between V_I and V_{ov} . *Although this circuit arrangement*
- contains two additional transistors to achieve voltage bootstrapping, the resulting circuit performance
- improvement is usually well worth the extra silicon area used for the
- bootstrapping devices.

MOS INVERTERS

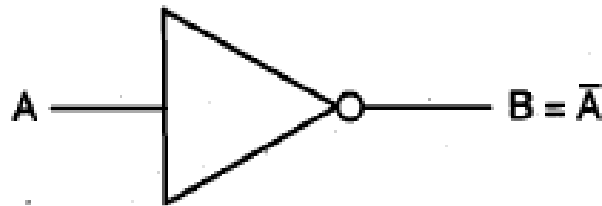
VLSI DESIGN

MOS INVERTERS

- For any IC technology used in digital circuit design, the basic circuit element is the logic inverter.
- Once the operation and characterization of the inverter circuits are thoroughly understood, the results can be extended to the design of the logic gates and other more complex circuits.
- Here, the DC (static) characteristics of various MOS inverter circuits are examined.

IDEAL INVERTER

- The logic symbol and the truth table of the ideal inverter are shown in Fig. 1.



Symbol

A	B
0	1
1	0

Truth Table

Figure 1: Logic symbol and truth table of the inverter

IDEAL INVERTER

- The Boolean (or logic) value of "1" can be represented by a high voltage of V_{DD} , and the Boolean (or logic) value of "0" can be represented by a low voltage of 0.
- The DC voltage transfer characteristic (VTC) of the ideal inverter circuit is shown in Fig. 2.

IDEAL INVERTER

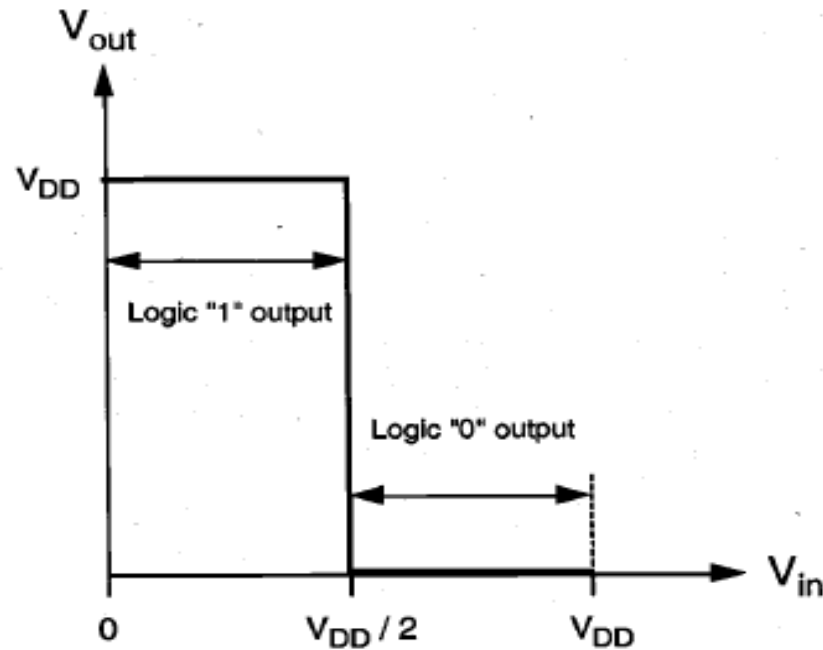


Figure 2. Voltage transfer characteristic (VTC) of the ideal inverter.

IDEAL INVERTER

- The voltage V_{th} is called the inverter threshold voltage.
- It is observed that for any input voltage between 0 and $V_{th} = V_{DD}/2$, the output voltage is equal to V_{DD} (logic "1").
- The output switches from V_{DD} to 0 when the input is equal to V_{th} . For any input voltage between V_{th} and V_{DD} , the output voltage assumes a value of 0 (logic "0").

IDEAL INVERTER

- Thus, an input voltage $0 < V_{in} < V_{th}$ is interpreted by this ideal inverter as a logic "0," while an input voltage $V_{th} < V_{in} < V_{DD}$ is interpreted as a logic "1."

nMOS INVERTER

- Figure 3 shows the generalized circuit structure of an nMOS inverter.
- The input voltage of the inverter circuit is also the gate-to-source voltage of the nMOS transistor ($V_{in} = V_{GS}$), while the output voltage of the circuit is equal to the drain-to-source voltage ($V_{out} = V_{DS}$).
- The source-to-substrate voltage is $V_{SB} = 0$.

nMOS INVERTER

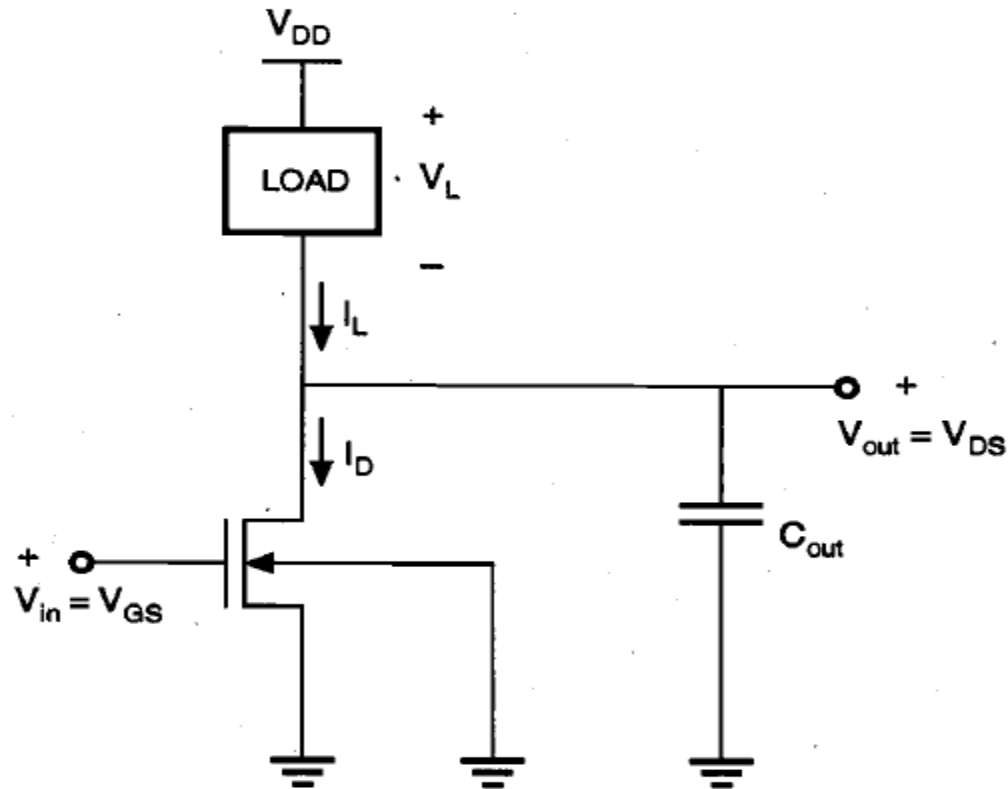


Fig. 3: General circuit structure of an nMOS inverter

nMOS INVERTER

- In this generalized representation, the load device is represented as a two-terminal circuit element with terminal current I_L and terminal voltage $V_L(I_L)$.
- One terminal of the load device is connected to the drain of the n-channel MOSFET, while the other terminal is connected to V_{DD} , the power supply voltage.

nMOS INVERTER

- The output terminal of the inverter shown in Fig. 3 is connected to the input of another MOS inverter.
- Consequently, the next circuit seen by the output node can be represented as a lumped capacitance, C_{out} .
- Since the DC gate current of an MOS transistor is negligible for all practical purposes, there will be no current flow into or out of the input and output terminals of the inverter in DC steady state.

Voltage Transfer Characteristic (VTC)

- Applying Kirchhoff's Current Law (KCL) to this simple circuit, we see that the load current is always equal to the nMOS drain current.

$$I_D(V_{in}, V_{out}) = I_L(V_L) \quad \text{.....(1)}$$

- The VTC of a realistic nMOS inverter is shown in Fig. 4.

Voltage Transfer Characteristic (VTC)

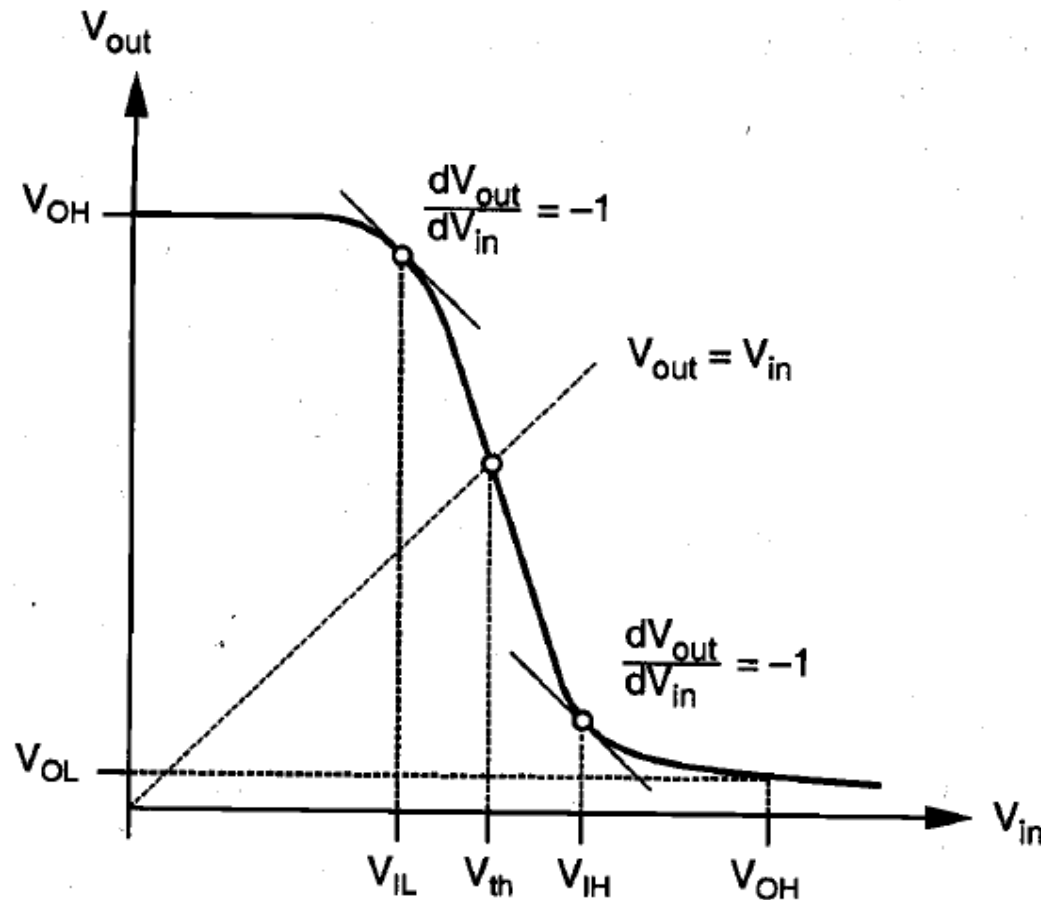


Figure 4. Typical voltage transfer characteristic (VTC) of a realistic nMOS inverter

Voltage Transfer Characteristic (VTC)

- The general shape of the VTC in Fig. 4 is qualitatively similar to that of the ideal inverter transfer characteristic shown in Fig. 2, but with several significant differences.
- For very low input voltage levels, the output voltage V_{out} is equal to the high value of V_{OH} (output high voltage).
- In this case, the driver nMOS transistor is in cut-off, and hence, does not conduct any current.

Voltage Transfer Characteristic (VTC)

- Consequently, the voltage drop across the load device is very small in magnitude, and the output voltage level is high.
- As the input voltage V_{in} increases, the driver transistor starts conducting a certain drain current, and the output voltage eventually starts to decrease.
- This drop in the output voltage level does not occur abruptly, such as the vertical drop assumed for the ideal inverter VTC, but rather gradually and with a finite slope.

Voltage Transfer Characteristic (VTC)

- Two critical voltage points on this curve are identified, where the slope of the $V_{out}(V_{in})$ characteristic becomes equal to -1, i.e.,

$$\frac{dV_{out}}{dV_{in}} = -1 \quad \text{.....(2)}$$

- The smaller input voltage value satisfying this condition is called the input low voltage V_{IL} and the larger input voltage satisfying this condition is called the input high voltage V_{IH} .
- Both of these voltages play significant roles in determining the noise margins of the inverter circuit.

Voltage Transfer Characteristic (VTC)

- As the input voltage is further increased, the output voltage continues to drop and reaches a value of V_{OL} (output low voltage) when the input voltage is equal to V_{OH} .
- The inverter threshold voltage V_{th} , which is considered as the transition voltage, is defined as the point where $V_{in} = V_{out}$, on the VTC.

Critical Voltages

- V_{OH} : Maximum output voltage when the output level is logic "1"
- V_{OL} : Minimum output voltage when the output level is logic "0"
- V_{IL} : Maximum input voltage which can be interpreted as logic "0"
- V_{IH} : Minimum input voltage which can be interpreted as logic "1"

Noise Immunity and Noise Margins

- The ability of an inverter to interpret an input signal within a voltage range as either a logic "0" or as a logic "1" allows digital circuits to operate with a certain tolerance to external signal perturbations.
- This tolerance to variations in the signal level is especially valuable in environments in which circuit noise can significantly corrupt the signals.

Noise Immunity and Noise Margins

- Here the circuit noise corresponds to unwanted signals that are coupled to some part of the circuit from neighboring lines (usually interconnection lines) by capacitive or inductive coupling.

Noise Immunity and Noise Margins

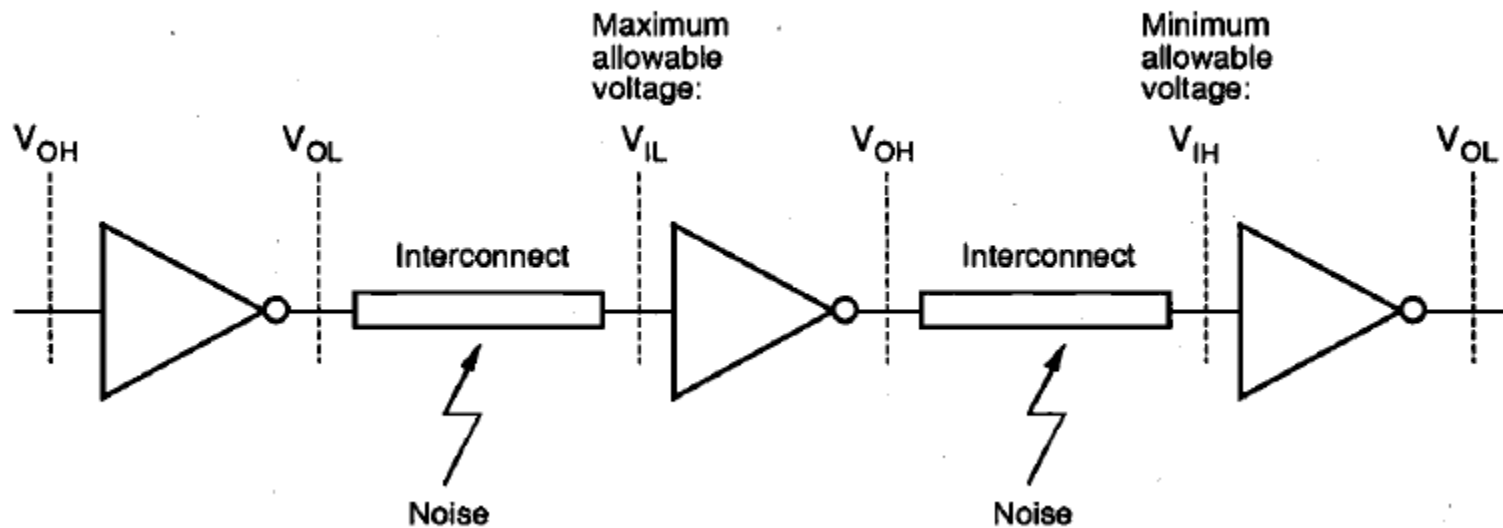


Figure 5: Propagation of digital signals under the influence of noise.

Noise Immunity and Noise Margins

- To illustrate the effect of noise on the circuit reliability, let us consider the circuit consisting of three cascaded inverters, as shown in Fig. 5.
- Assume that all inverters are identical, and that the input voltage of the first inverter is equal to V_{OH} , i.e., a logic "1."
- The output voltage of the first inverter will be equal to V_{OL} corresponding to a logic "0" level.

Noise Immunity and Noise Margins

- Now, this output signal is being transmitted to the next inverter input via an interconnect, which could be a metal or polysilicon line connecting the two gates.
- Since on-chip interconnects are generally prone to signal noise, the output signal of the first inverter will be perturbed during transmission.

Noise Immunity and Noise Margins

- Consequently, the voltage level at the input of the second inverter will be either larger or smaller than V_{OL} .
- If the input voltage of the second inverter is smaller than V_{OL} this signal will be interpreted correctly as a logic "0" input by the second inverter.
- On the other hand, if the input voltage becomes larger than V_{IL} as a result of noise, then it may not be interpreted correctly by the inverter.

Noise Immunity and Noise Margins

- Thus, we conclude that V_{IL} is the maximum allowable voltage at the input of the second inverter, which is low enough to ensure a logic "1" output.
- Now consider the signal transmission from the output of the second inverter to the input of the third inverter, assuming that the second inverter produces an output voltage level of V_{OH} .

Noise Immunity and Noise Margins

- As in the previous case, this output signal will be perturbed because of noise interference, and the voltage level at the input of the third inverter will be different from V_{OH} .
- If the input voltage of the third inverter is larger than V_{OH} , this signal will be interpreted correctly as a logic " 1 " input by the third inverter.

Noise Immunity and Noise Margins

- If the voltage level drops below V_{IH} due to noise, however, the input cannot be interpreted as a logic "1."
- Consequently, V_{IH} is the minimum allowable voltage at the input of the third inverter which is high enough to ensure a logic "0" output.

Noise Immunity and Noise Margins

- These observations lead us to the definition of noise tolerances for digital circuits, called noise margins and denoted by NM.
- The noise immunity of the circuit increases with NM.
- Two noise margins will be defined: the noise margin for low signal levels (NM_L) and the noise margin for high signal levels (NM_H).

$$NM_L = V_{IL} - V_{OL} \quad \dots\dots(3)$$

$$NM_H = V_{OH} - V_{IH} \quad \dots\dots(4)$$

Noise Immunity and Noise Margins

- Figure 6 shows a graphical illustration of the noise margins.
- Here, the shaded areas indicate the valid regions of the input and output voltages, and the noise margins are shown as the amount of variation in the signal levels that can be allowed while the signal is transmitted from the output of one gate to the input of the next gate.

Noise Immunity and Noise Margins

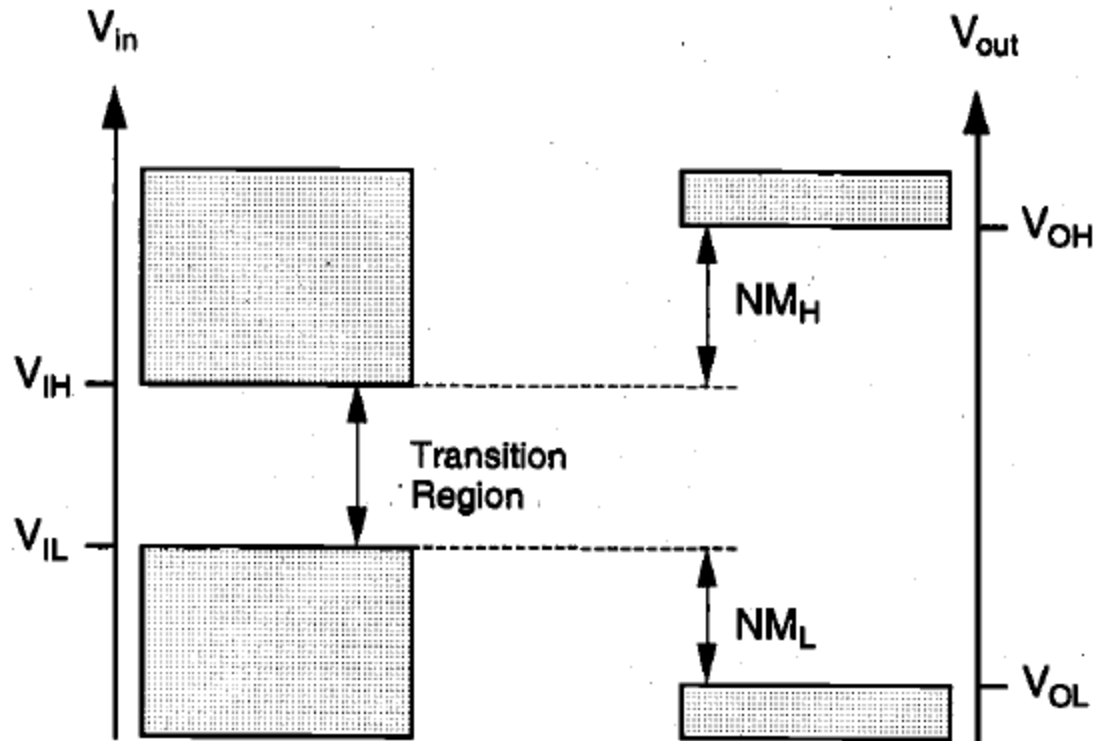


Fig. 6: Definition of noise margins NM_L and NM_H . Note that the shaded regions indicate valid high and low levels for the input and output signals

Noise Immunity and Noise Margins

- The selection of the two critical voltage points, V_{IL} and V_{IH} , using the slope condition can now be justified based on noise considerations.
- We know that the output voltage V_{out} , of an inverter under noise-free, steady-state conditions is a nonlinear function of the input voltage, V_{in} .

Noise Immunity and Noise Margins

- The shape of this function is described by the voltage transfer characteristic (VTC).

$$V_{out} = f(V_{in}) \quad \text{.....(5)}$$

- If the input signal is perturbed from its nominal value because of external influences, such as noise, output voltage will also deviate from its nominal level.

Noise Immunity and Noise Margins

- Here, ΔV_{noise} represents the voltage perturbation that is assumed to affect the input voltage of the inverter.

$$V_{\text{out}}' = f(V_{\text{in}} + \Delta V_{\text{noise}}) \quad \dots\dots(6)$$

- By using a simple first-order Taylor series expansion the perturbed output voltage V_{out}' can be expressed;

$$V_{\text{out}}' = f(V_{\text{in}}) + \frac{dV_{\text{out}}}{dV_{\text{in}}} \cdot \Delta V_{\text{noise}} + \text{higher order terms (neglected)} \quad \dots\dots(7)$$

Noise Immunity and Noise Margins

- $f(V_{in})$ represents the nominal (non-perturbed) output (dV_{out}/dV_{in}) represents the voltage gain of the inverter at the nominal input voltage value V_{in}

$$\text{Perturbed Output} = \text{Nominal Output} + \text{Gain} \times \text{External Perturbation} \quad \dots\dots(8)$$

Noise Immunity and Noise Margins

- If the magnitude of the voltage gain at the nominal input voltage V_{in} is smaller than unity then the input perturbation is not amplified and, consequently, the output perturbation remains relatively small.
- Otherwise, with the voltage gain larger than unity a small perturbation in the input voltage level will cause a rather large perturbation in the output voltage.

Noise Immunity and Noise Margins

- There is a voltage range between V_{IL} and V_{IH} (Fig. 6), corresponding to input voltage values that may not be processed correctly either as a logic "0" input or as a logic "1" input by the inverter.
- This region is called the uncertain region or, alternatively, the transition region.

Noise Immunity and Noise Margins

- Ideally, the slope of the voltage transfer characteristic should be very large between V_{IL} and V_{IH} , because a narrow uncertain (transition) region obviously allows for larger noise margins.
- We will see that reducing the width of the uncertain region is one of the most important design objectives.

Power and Area Considerations

- The DC power dissipation of an inverter circuit can be calculated as the product of its power supply voltage and the amount of current drawn from the power supply during steady state.

$$P_{DC} = V_{DD} \cdot I_{DC} \quad \text{.....(9)}$$

Power and Area Considerations

- The DC current drawn by the inverter circuit may vary depending on the input and output voltage levels.
- Assuming that the input voltage level corresponds to logic "0" during 50% of the operation time and to logic " 1 " during the other 50%, the overall DC power consumption of the circuit can be estimated as follows:

$$P_{DC} = \frac{V_{DD}}{2} \cdot [I_{DC}(V_{in} = low) + I_{DC}(V_{in} = high)] \quad \dots\dots\dots(10)$$

Resistive-Load Inverter

- The basic structure of the resistive-load inverter circuit is shown in Fig. 7.
- An enhancement-type nMOS transistor acts as the driver device.
- The load consists of a simple linear resistor, R_L .
- The power supply voltage of this circuit is V_{DD} .
- Since the following analysis concentrates on the static behavior of the circuit, the output load capacitance is not shown in this figure.

Resistive-Load Inverter

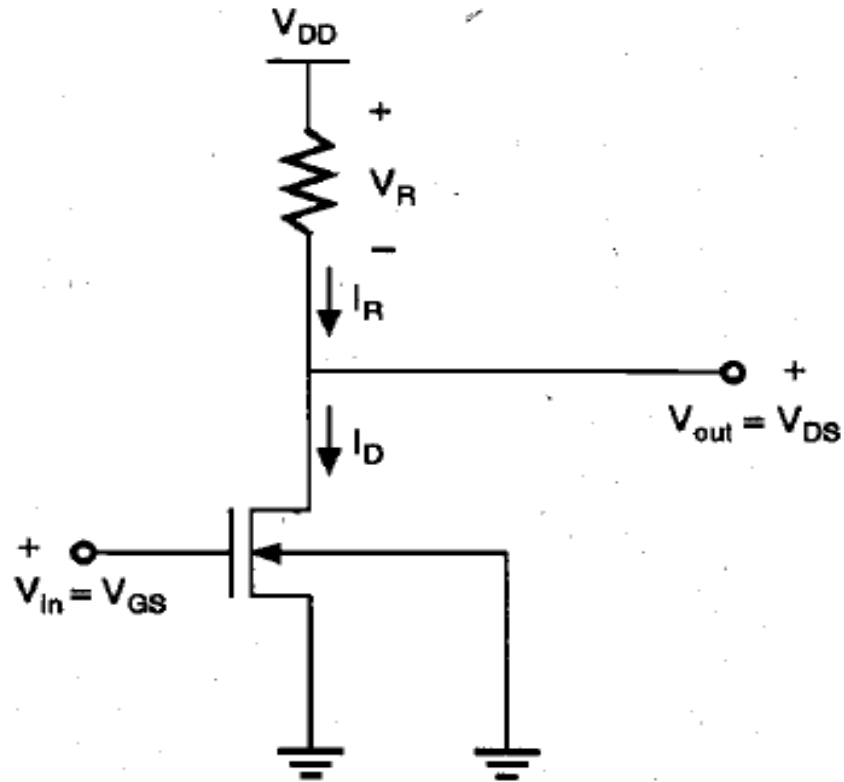


Figure 7. Resistive-load inverter circuit

Resistive-Load Inverter

- The drain current I_D of the driver MOSFET is equal to the load current I_R in DC steady-state operation.
- To simplify the calculations, the channel-length modulation effect is neglected in the following, i.e., $\lambda = 0$.
- Also, the source and the substrate terminals of the driver transistor are both connected to the ground; hence, $V_{SB} = 0$.

Resistive-Load Inverter

- Consequently, the threshold voltage of the driver transistor is always equal to V_{T0} .
- For input voltages smaller than the threshold voltage V_{T0} , the transistor is in cut-off, and does not conduct any drain current.
- Since the voltage drop across the load resistor is equal to zero, the output voltage must be equal to the power supply voltage, V_{DD} .

Resistive-Load Inverter

- As the input voltage is increased beyond V_{T0} , the driver transistor starts conducting a nonzero drain current.
- The driver MOSFET is initially in saturation, since its drain-to-source voltage ($V_{DS} = V_{out}$) is larger than $(V_{in} - V_{T0})$.
- Thus,

$$I_R = \frac{k_n}{2} \cdot (V_{in} - V_{T0})^2 \quad \text{.....(11)}$$

Resistive-Load Inverter

- With increasing input voltage, the drain current of the driver also increases, and the output voltage V_{out} , starts to drop.
- Eventually, for input voltages larger than $V_{out} + V_{T0}$, the driver transistor enters the linear operation region.
- At larger input voltages, the transistor remains in linear mode, as the output voltage continues to decrease.

$$I_R = \frac{k_n}{2} \cdot [2 \cdot (V_{in} - V_{T0}) \cdot V_{out} - V_{out}^2] \quad \text{.....(12)}$$

Resistive-Load Inverter

- The various operating regions of the driver transistor and the corresponding input-output conditions are listed in the following table.

Input Voltage Range	Operating Mode
$V_{in} < V_{T0}$	cut-off
$V_{T0} \leq V_{in} < V_{out} + V_{T0}$	saturation
$V_{in} \geq V_{out} + V_{T0}$	linear

Resistive-Load Inverter

- Figure 8 shows the voltage transfer characteristic of a typical resistive-load inverter circuit, indicating the operating modes of the driver transistor and the critical voltage points on the VTC.
- Now, let us start with the calculation of the five critical voltage points, which determine the steady-state input-output behavior of the inverter.

Resistive-Load Inverter

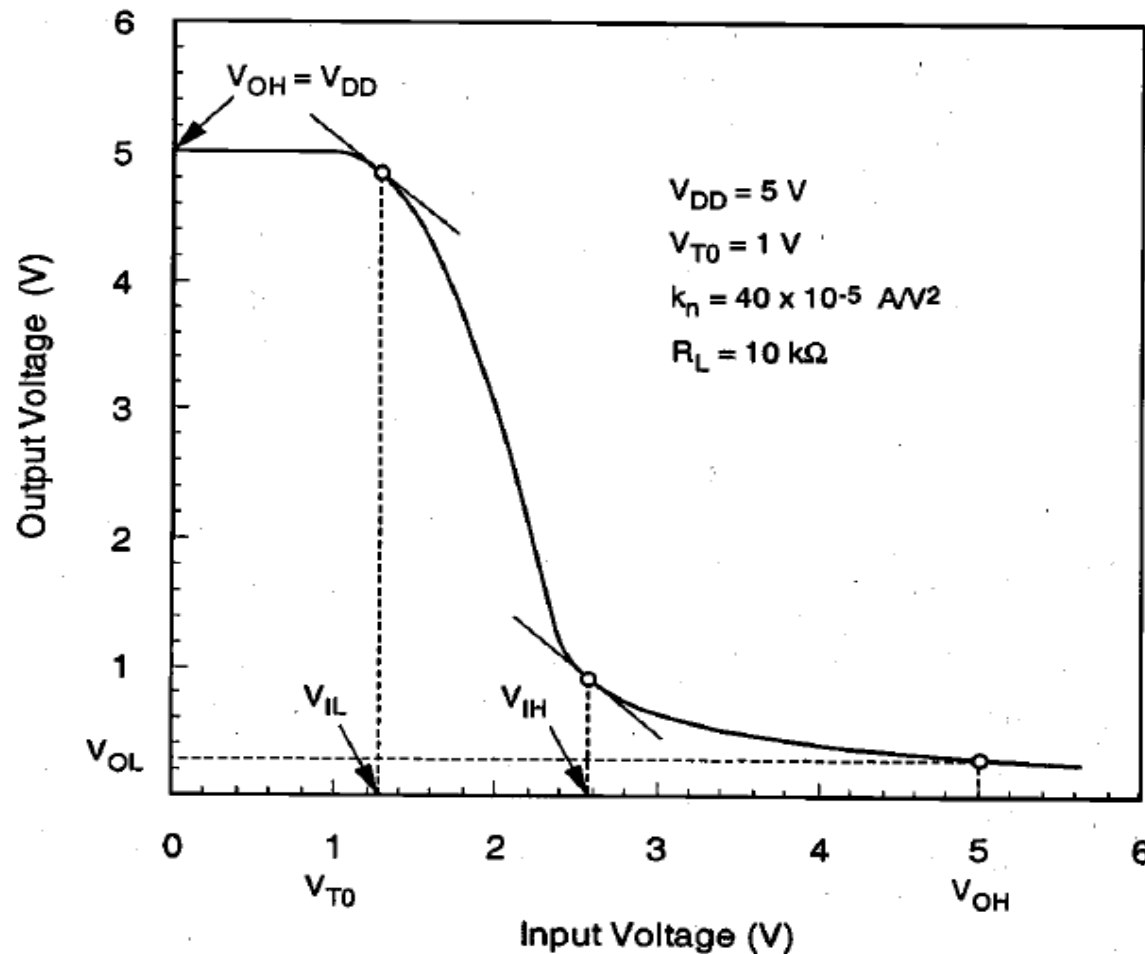


Fig. 8: Typical VTC of a resistive-load inverter circuit

Calculation of V_{OH}

- The output voltage V_{out} is given by;

$$V_{out} = V_{DD} - R_L \cdot I_R \quad \text{.....(13)}$$

- When the input voltage V_{in} is low, i.e., smaller than the threshold voltage of the driver MOSFET, the driver transistor is cut-off.
- Since the drain current of the driver transistor is equal to the load current, $I_R = I_D = 0$, the output voltage of the inverter is:

$$V_{OH} = V_{DD} \quad \text{.....(14)}$$

Calculation of V_{OL}

- To calculate the output low voltage V_{OL} let us assume that the input voltage is equal to V_{OH} i.e., $V_{in} = V_{OH} = V_{DD}$.
- Since $V_{in} - V_{T0} > V_{out}$ in this case, the driver transistor operates in the linear region.
- Also the load current I_R is;

$$I_R = \frac{V_{DD} - V_{out}}{R_L} \quad \text{.....(15)}$$

Calculation of V_{OL}

- Using KCL for the output node, i.e., $I_R = I_D$ we can write the following equation:

$$\frac{V_{DD} - V_{OL}}{R_L} = \frac{k_n}{2} \cdot [2 \cdot (V_{DD} - V_{T0}) \cdot V_{OL} - V_{OL}^2] \quad \dots\dots(16)$$

- This equation yields a simple quadratic in V_{OL} which is solved to find the value of the output low voltage.

$$V_{OL}^2 - 2 \cdot \left(V_{DD} - V_{T0} + \frac{1}{k_n R_L} \right) \cdot V_{OL} + \frac{2}{k_n R_L} \cdot V_{DD} = 0 \quad \dots\dots(17)$$

Calculation of V_{OL}

- There are two possible solutions of (Eqn. 17), we must choose the one that is physically correct, i.e., the value of the output low voltage must be between 0 and V_{DD} .
- It is observed that the product ($k_n R_L$) is one of the important design parameters that determine the value of V_{OL} .

$$V_{OL} = V_{DD} - V_{T0} + \frac{1}{k_n R_L} - \sqrt{\left(V_{DD} - V_{T0} + \frac{1}{k_n R_L}\right)^2 - \frac{2V_{DD}}{k_n R_L}} \quad \dots\dots(18)$$

Calculation of V_{IL}

- By definition, V_{IL} is the smaller of the two input voltage values at which the slope of the VTC becomes equal to (-1) , i.e., $dV_{out}/dV_{in} = -1$.
- Fig. 8 shows that when the input is equal to V_{IL} , the output voltage (V_{out}) is only slightly smaller than V_{OH} .
- Consequently, $V_{out} > V_{in} - V_{T0}$, and the driver transistor operates in saturation.

Calculation of V_{IL}

- By writing the KCL for the output node.

$$\frac{V_{DD} - V_{out}}{R_L} = \frac{k_n}{2} \cdot (V_{in} - V_{T0})^2 \quad \text{.....(19)}$$

- To satisfy the derivative condition, we differentiate both sides of (Eqn. 19) with respect to V_{in} , which results in the following equation:

$$-\frac{1}{R_L} \cdot \frac{dV_{out}}{dV_{in}} = k_n \cdot (V_{in} - V_{T0}) \quad \text{.....(20)}$$

Calculation of V_{IL}

- Since the derivative of the output voltage with respect to the input voltage is equal to (-1) at V_{IL} , we can substitute $dV_{out} / dV_{in} = -1$ in (Eqn. 20)

$$-\frac{1}{R_L} \cdot (-1) = k_n \cdot (V_{IL} - V_{T0}) \quad \dots\dots(21)$$

- Now

$$V_{IL} = V_{T0} + \frac{1}{k_n R_L} \quad \dots\dots(22)$$

Calculation of V_{IL}

- The value of the output voltage when the input is equal to V_{IL} can also be found by substituting (Eqn. 22) into (Eqn. 19), as follows:

$$\begin{aligned} V_{out}(V_{in} = V_{IL}) &= V_{DD} - \frac{k_n R_L}{2} \cdot \left(V_{T0} + \frac{1}{k_n R_L} - V_{T0} \right)^2 \\ &= V_{DD} - \frac{1}{2k_n R_L} \end{aligned} \quad \text{.....(23)}$$

Calculation of V_{IH}

- V_{IH} is the larger of the two voltage points on VTC at which the slope is equal to (-1).
- It can be seen from Fig. 8 that when the input voltage is equal to V_{IH} , the output voltage V_{out} is only slightly larger than the output low voltage V_{OL} .
- Hence, $V_{out} < V_{in} - V_{T0}$, and the driver transistor operates in the linear region.

Calculation of V_{IH}

- The KCL equation for the output node is given as;

$$\frac{V_{DD} - V_{out}}{R_L} = \frac{k_n}{2} \cdot [2 \cdot (V_{in} - V_{T0}) \cdot V_{out} - V_{out}^2] \quad \dots\dots(24)$$

- Differentiating both sides of (Eqn. 24) with respect to V_{in} ;

$$-\frac{1}{R_L} \cdot \frac{dV_{out}}{dV_{in}} = \frac{k_n}{2} \cdot \left[2 \cdot (V_{in} - V_{T0}) \cdot \frac{dV_{out}}{dV_{in}} + 2 V_{out} - 2 V_{out} \cdot \frac{dV_{out}}{dV_{in}} \right] \quad \dots\dots(25)$$

Calculation of V_{IH}

- Next, we can substitute $dV_{out}/dV_{in} = -1$ into (Eqn. 25), since the slope of the VTC is equal to (-1) also at $V_{in} = V_{IH}$.

$$-\frac{1}{R_L} \cdot (-1) = k_n \cdot \left[(V_{IH} - V_{T0}) \cdot (-1) + 2V_{out} \right] \quad \dots\dots(26)$$

- Solving (Eqn. 26) for V_{IH} yields the following expression.

$$V_{IH} = V_{T0} + 2V_{out} - \frac{1}{k_n R_L} \quad \dots\dots(27)$$

Calculation of V_{IH}

- By using Eqn. 27 into Eqn. 24 we obtain

$$\frac{V_{DD} - V_{out}}{R_L} = \frac{k_n}{2} \cdot \left[2 \cdot \left(V_{T0} + 2V_{out} - \frac{1}{k_n R_L} - V_{T0} \right) \cdot V_{out} - V_{out}^2 \right] \dots\dots\dots(28)$$

$$V_{out} (V_{in} = V_{IH}) = \sqrt{\frac{2}{3} \cdot \frac{V_{DD}}{k_n R_L}} \dots\dots\dots(29)$$

Calculation of V_{IH}

- Finally, V_{IH} can be found by substituting (Eqn. 29) into (Eqn. 27), as follows.

$$V_{IH} = V_{T0} + \sqrt{\frac{8}{3} \cdot \frac{V_{DD}}{k_n R_L}} - \frac{1}{k_n R_L} \dots\dots(30)$$

Resistive-Load Inverter

- The four critical voltage points V_{OL} , V_{OH} , V_{IL} , V_{IH} are used to determine the noise margins, NM_L and NM_H .
- The inverter threshold voltage can be found simply by substituting $V_{in} = V_{out} = V_{th}$ into (Eqn. *5.19), and by solving the resulting quadratic for V_{th} .

Role of the term $(k_n R_L)$

- The term $(k_n R_L)$ plays an important role in determining the shape of the voltage transfer characteristic.
- For larger $(k_n R_L)$ values, the output low voltage V_{OL} becomes smaller and that the shape of the VTC approaches that of the ideal inverter, with very large transition slope as shown in Fig 9.
- Achieving larger $(k_n R_L)$ values in a design, however, may involve other trade-offs with the area and the power consumption of the circuit.

Role of the term $(k_n R_L)$

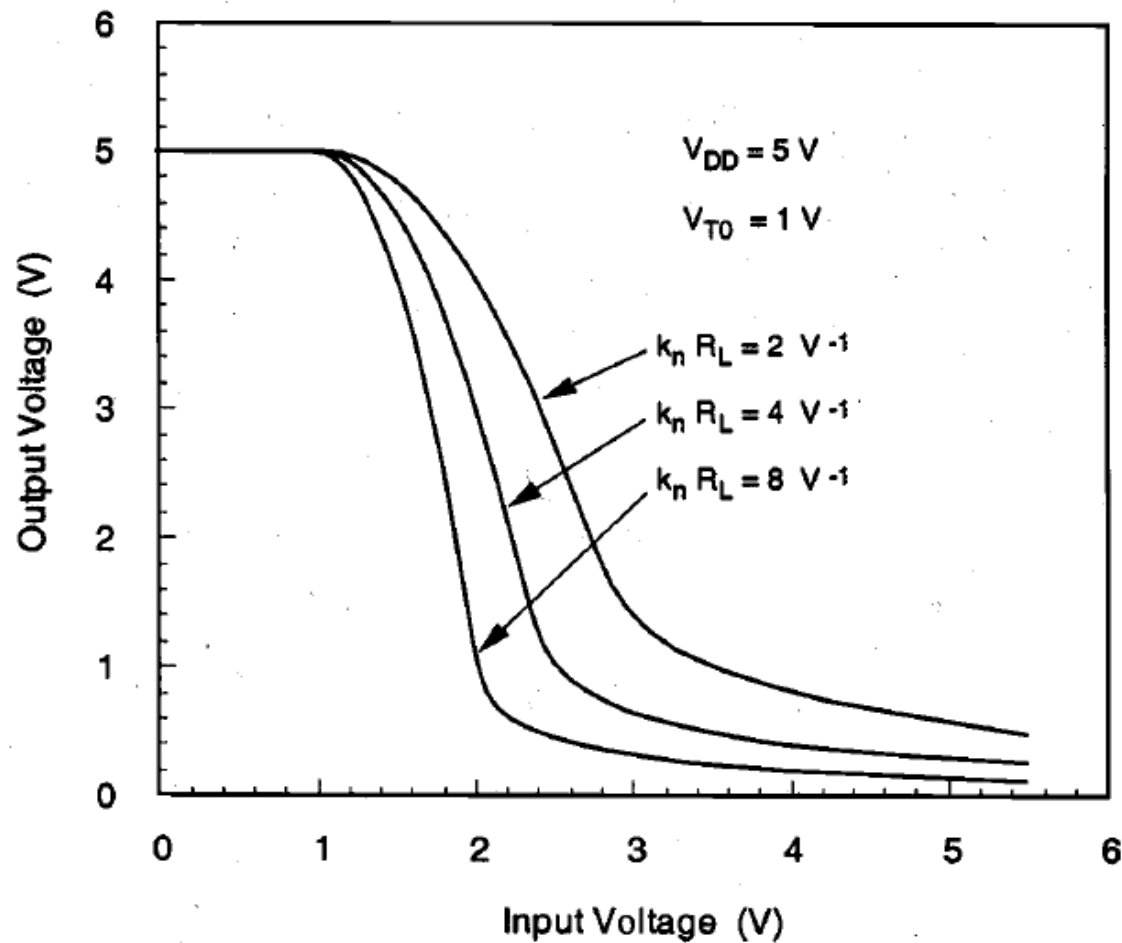


Fig. 9

Power Consumption and Chip Area

- The average DC power consumption of the resistive-load inverter circuit is found by considering two cases.
- (i) $V_{in} = V_{OL}$ (low) and (ii) $V_{in} = V_{OH}$ (high).
- When the input voltage is equal to V_{OL} , the driver transistor is in cut-off. Consequently, there is no steady-state current flow in the circuit ($I_D = I_R = 0$), and the DC power dissipation is equal to zero.

Power Consumption and Chip Area

- When the input voltage is equal to V_{OH} , the output voltage is equal to V_{OL} , the current drawn from the power supply can be found as;

$$I_D = I_R = \frac{V_{DD} - V_{OL}}{R_L} \quad \text{.....(31)}$$

- Assuming that the input voltage is "low" during 50% of the operation time, and "high" during the remaining 50%, the average DC power consumption of the inverter can be estimated as follows:

$$P_{DC}(\text{average}) = \frac{V_{DD}}{2} \cdot \frac{V_{DD} - V_{OL}}{R_L} \quad \text{.....(32)}$$

Problem 1

- Consider the following inverter design problem: Given $V_{DD} = 5V$, $k_n' = 30 \mu A/V^2$, and $V_{T0} = 1 V$, design a resistive-load inverter circuit with $V_{OL} = 0.2 V$. Specifically, determine the (W/L) ratio of the driver transistor and the value of the load resistor R_L that achieve the required V_{OL} .

Solution

- The driver transistor is operating in the linear region when the output voltage is equal to V_{OL} and the input voltage is equal to $V_{OH} = V_{DD}$.

$$\frac{V_{DD} - V_{OL}}{R_L} = \frac{k_n'}{2} \cdot \frac{W}{L} \cdot [2 \cdot (V_{OH} - V_{T0}) \cdot V_{OL} - V_{OL}^2]$$

Solution

- Assuming $V_{OL} = 0.2$ V and using the given values for the power supply voltage, the driver threshold voltage and the driver transconductance k_n' , we obtain the following equation:

$$\frac{5 - 0.20}{R_L} = \frac{30 \times 10^{-6}}{2} \cdot \frac{W}{L} \cdot (2 \cdot 4 \cdot 0.20 - 0.20^2)$$

Solution

- This equation can be rewritten as:

$$\frac{W}{L} \cdot R_L = 2.05 \times 10^5 \, \Omega$$

Solution

- The designer has a choice of different (W/L) and R_L values, all of which satisfy the given design specification, $V_{OL} = 0.2 \text{ V}$.
- The selection of the pair of values, to use for (W/L) and R_L in the final design ultimately depends on other considerations, such as the power consumption of the circuit and the silicon area.

Solution

- The table below lists some of the design possibilities, along with the average DC power consumption estimated for each design.

$\left(\frac{W}{L}\right)$ – Ratio	Load Resistor R_L [k Ω]	DC Power Consumption $P_{DC,average}$ [μ W]
1	205.0	58.5
2	102.5	117.1
3	68.4	175.4
4	51.3	233.9
5	41.0	292.7
6	34.2	350.8

Solution

- It is seen that the power consumption increases significantly as the value of the load resistor R_L is decreased, and the (W/L) ratio is increased.
- If lowering the DC power consumption is the overriding concern, we may choose a small (W/L) ratio and a large load resistor.
- On the other hand, if the fabrication of the large load resistor requires a large silicon area, a clear trade-off exists between the DC power dissipation and the area occupied by the inverter circuit.

Resistive-Load Inverter

- The chip area occupied by the resistive-load inverter circuit depends on two parameters, the (W/L) ratio of the driver transistor and the value of the resistor R_L .
- The area of the driver transistor can be approximated by the gate area, $(W \times L)$.
- Assuming that the gate length L is kept at its smallest possible value for the given technology, the gate area will be proportional to the (W/L) ratio of the transistor.

Problem 2

- Consider a resistive-load inverter circuit with $V_{DD} = 5\text{ V}$, $k_n' = 20\text{ }\mu\text{A/V}^2$, $V_{T0} = 0.8\text{ V}$, $R_L = 200\text{ k}\Omega$, and $W/L = 2$. Calculate the critical voltages ($V_{OL}, V_{OH}, V_{IL}, V_{IH}$) on the VTC and find the noise margins of the circuit.

Solution

- When the input voltage is low, i.e., when the driver nMOS transistor is cut-off, the output high voltage can be found as;

$$V_{OH} = V_{DD} = 5 \text{ V}$$

- In this resistive-load inverter example, the transconductance of the driver transistor is $k_n = k_n'(W/L) = 40 \mu\text{A}/\text{V}^2$ and, hence, $(k_n R_L) = 8\text{V}^{-1}$

Solution

- The output low voltage V_{OL} is calculated;

$$\begin{aligned} V_{OL} &= V_{DD} - V_{T0} + \frac{1}{k_n R_L} - \sqrt{\left(V_{DD} - V_{T0} + \frac{1}{k_n R_L} \right)^2 - \frac{2V_{DD}}{k_n R_L}} \\ &= 5 - 0.8 + \frac{1}{8} - \sqrt{\left(5 - 0.8 + \frac{1}{8} \right)^2 - \frac{2 \cdot 5}{8}} \\ &= 0.147 \text{ V} \end{aligned}$$

Solution

- The critical voltage V_{IL} is obtained;

$$V_{IL} = V_{T0} + \frac{1}{k_n R_L} = 0.8 + \frac{1}{8} = 0.925 \text{ V}$$

- Finally, the critical voltage V_{IH} can be calculated;

$$V_{IH} = V_{T0} + \sqrt{\frac{8}{3} \cdot \frac{V_{DD}}{k_n R_L}} - \frac{1}{k_n R_L} = 0.8 + \sqrt{\frac{8}{3} \cdot \frac{5}{8}} - \frac{1}{8} = 1.97 \text{ V}$$

Solution

- Now, the noise margins can be found,

$$NM_L = V_{IL} - V_{OL} = 0.93 - 0.15 = 0.78 \text{ V}$$

$$NM_H = V_{OH} - V_{IH} = 5.0 - 1.97 = 3.03 \text{ V}$$

Solution

- At this point, we can comment on the quality of this particular inverter design for DC operation.
- The noise margin NM_L found here is quite low, and it may eventually lead to misinterpretation of input signal levels.
- For better noise immunity, the noise margin for "low" signals should be at least about 25% of the power supply voltage V_{DD} , i.e., about 1.25 V.

Inverters with n-Type MOSFET Load

- The simple resistive-load inverter circuit examined in the previous section is not a suitable candidate for most digital VLSI system applications, because of the large area occupied by the load resistor.
- In this section, inverter circuits are introduced which use an nMOS transistor as the active load device, instead of the linear load resistor.

Inverters with n-Type MOSFET Load

- The main advantage of using a MOSFET as the load device is that the silicon area occupied by the transistor is usually smaller than that occupied by a comparable resistive load.
- Moreover, inverter circuits with active loads can be designed to have better overall performance compared to that of passive-load inverters.

Enhancement-Load nMOS Inverter

- The circuit configurations of two inverters with enhancement-type load devices are shown in Fig. 10.
- Depending on the bias voltage applied to its gate terminal, the load transistor can be operated either in the saturation region or in the linear region.

Enhancement-Load nMOS Inverter

- The saturated enhancement-load inverter shown in Fig. 10(a) requires a single voltage supply and a relatively simple fabrication process, yet the V_{OH} level is limited to

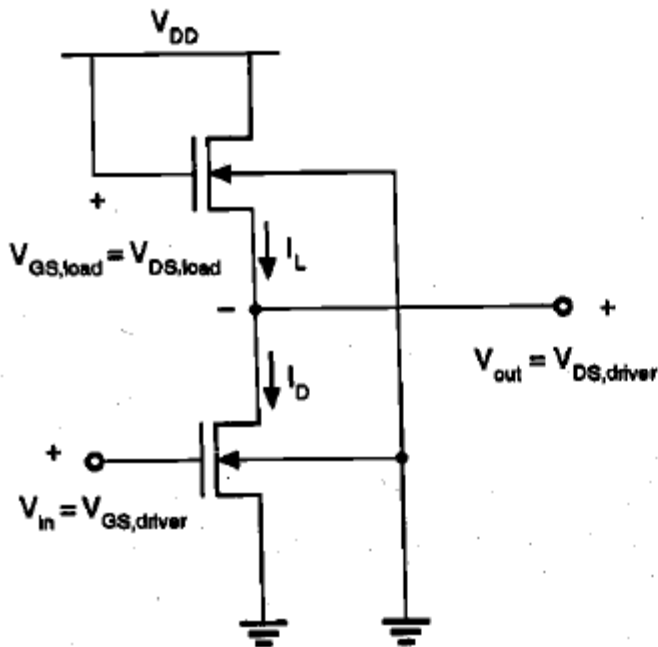
$$V_{DD} - V_{T,load}.$$

- The load device of the inverter circuit shown in Fig. 10(b), on the other hand, is always biased in the linear region.
- Thus, the V_{OH} level is equal to V_{DD} , resulting in higher noise margins compared to saturated enhancement-load inverter.

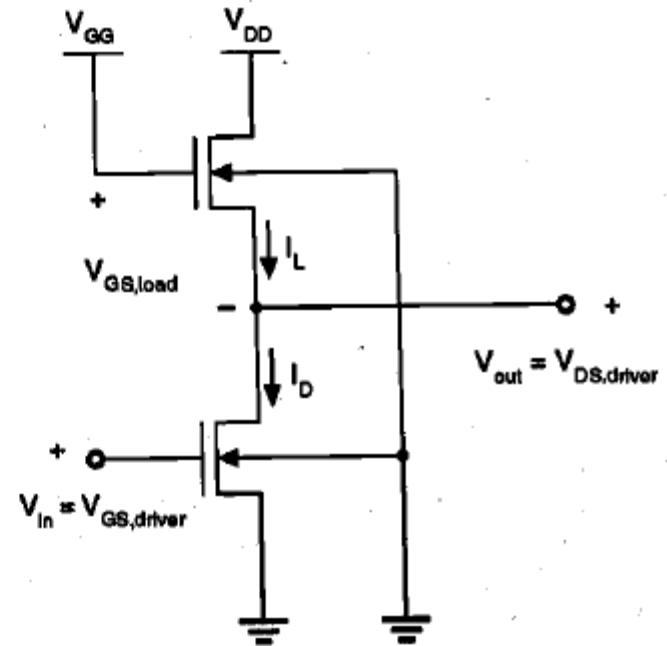
Drawbacks of Enhancement-Load nMOS Inverter

- The most significant drawback of this configuration is the use of two separate power supply voltages.
- In addition, both types of inverter circuits shown in Fig. 10 suffer from relatively high stand-by (DC) power dissipation.
- Hence, enhancement-load nMOS inverters are not used in any large-scale digital applications.

Enhancement-Load nMOS Inverter



(a)



(b)

Figure 10. (a) Inverter circuit with saturated enhancement-type nMOS load (b) Inverter with linear enhancement-type load

Depletion-Load nMOS Inverter

- Several of the disadvantages of the enhancement-type load inverter can be avoided by using a depletion-type nMOS transistor as the load device.
- Although the fabrication process for a depletion-type nMOS load is slightly more complicated and requires additional processing steps, the immediate advantages of implementing this circuit configuration are:
 - (i) sharp VTC transition and better noise margins,
 - (ii) single power supply, and (iii) smaller overall layout area.

Depletion-Load nMOS Inverter

- The circuit diagram of the depletion-load inverter circuit is shown in Fig. 11(a), and a simplified view of the circuit consisting of a nonlinear load resistor and a nonideal switch (driver) is shown in Fig. 11(b).
- The driver device is an enhancement-type nMOS transistor, with $V_{T0,driver} > 0$, whereas the load is a depletion-type nMOS transistor, with $V_{T0,load} < 0$.

Depletion-Load nMOS Inverter

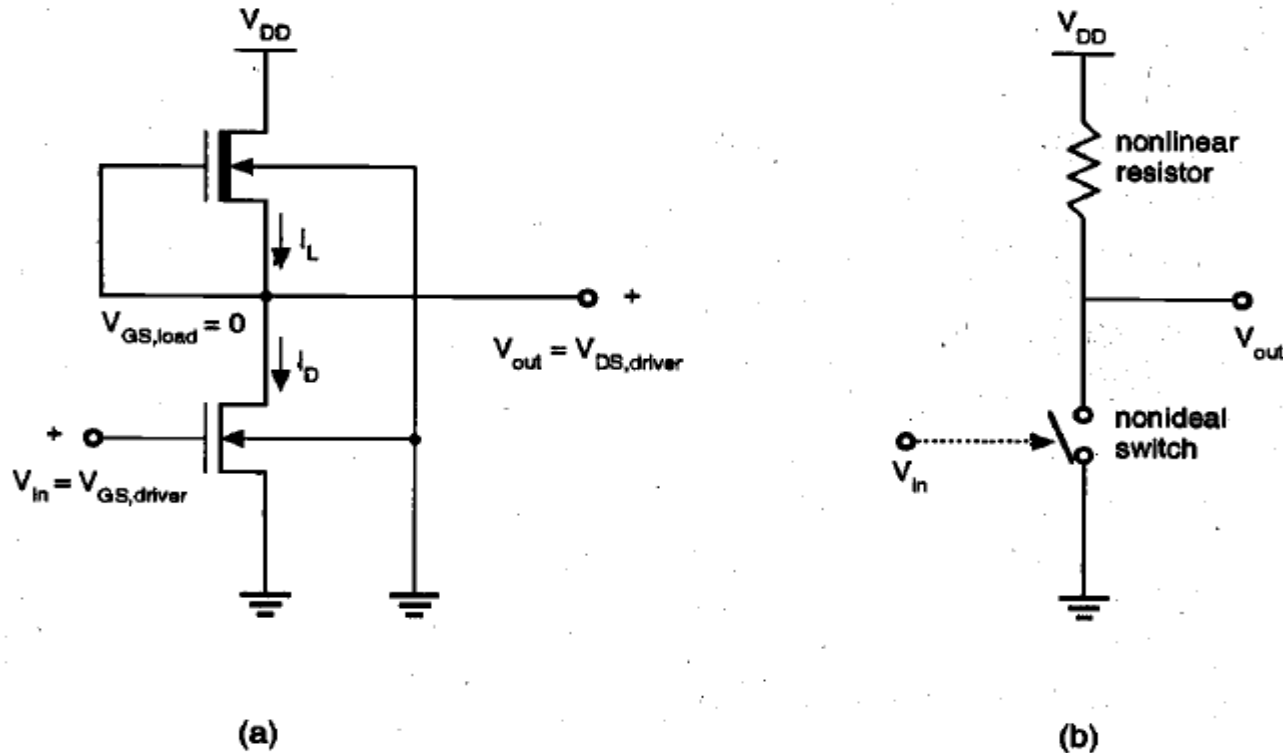


Figure 11. (a) Inverter circuit with depletion-type nMOS load. (b) Simplified equivalent circuit consisting of a nonlinear load resistor and a nonideal switch controlled by the input.

Depletion-Load nMOS Inverter

- The current-voltage equations to be used for the depletion-type load transistor are identical to those of the enhancement-type device, with the exception of the negative threshold voltage .
- The gate and the source nodes of the load transistor are connected, hence, $V_{GS,load} = 0$ always.

Depletion-Load nMOS Inverter

- Since the threshold voltage of the depletion type load is negative, the condition $V_{GS,load} > V_{T,load}$ is satisfied, and the load device always has a conducting channel regardless of the input and output voltage levels.
- Also both the driver transistor and the load transistor are built on the same p-type substrate, which is connected to the ground.

Depletion-Load nMOS Inverter

- Consequently, the load device is subject to the substrate-bias effect, so that its threshold voltage is a function of its source-to-substrate voltage, $V_{SB,load} = V_{out}$.

$$V_{T,load} = V_{T0,load} + \gamma \left(\sqrt{|2\phi_F| + V_{out}} - \sqrt{|2\phi_F|} \right) \dots\dots(33)$$

Depletion-Load nMOS Inverter

- The operating mode of the load transistor is determined by the output voltage level.
- When the output voltage is small, i.e., when $V_{\text{out}} < V_{\text{DD}} + V_{\text{T,load}}$, the load transistor is in saturation.
- Note that this condition corresponds to $V_{\text{DS,load}} > V_{\text{GS,load}} - V_{\text{T,load}}$.

Depletion-Load nMOS Inverter

- Then, the load current is given by the following equation.

$$I_{D,load} = \frac{k_{n,load}}{2} \cdot \left[-V_{T,load}(V_{out}) \right]^2 = \frac{k_{n,load}}{2} \cdot \left| V_{T,load}(V_{out}) \right|^2 \quad \dots\dots(34)$$

- For larger output voltage levels, i.e., for $V_{out} > V_{DD} + V_{T,load}$, the depletion-type load transistor operates in the linear region. The load current in this case is

$$I_{D,load} = \frac{k_{n,load}}{2} \cdot \left[2 \left| V_{T,load}(V_{out}) \right| \cdot (V_{DD} - V_{out}) - (V_{DD} - V_{out})^2 \right] \quad \dots\dots(35)$$

Depletion-Load nMOS Inverter

- The voltage transfer characteristic (VTC) of this inverter can be constructed by setting
- $I_{D,driver} = I_{D,load}$, $V_{GS,driver} = V_{in}$, and $V_{DS,driver} = V_{out}$.
- Figure 12 shows the VTC of a typical depletion-load inverter, with $k_{n,driver}' = k_{n,load}'$

Depletion-Load nMOS Inverter

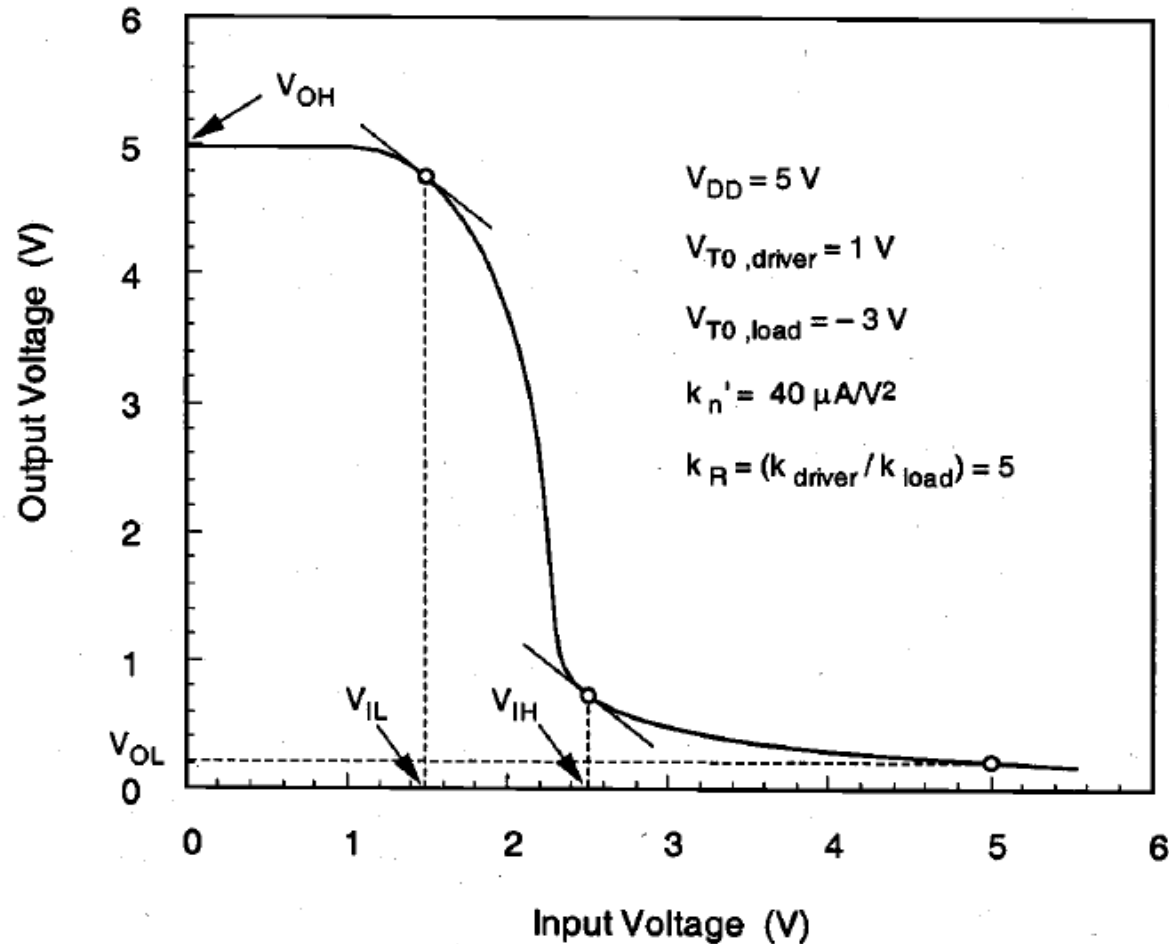


Figure 12. Typical VTC of a depletion-load inverter circuit

Depletion-Load nMOS Inverter

- Next, we will consider the critical voltage points V_{OH} , V_{OL} , V_{IL} and V_{IH} for this inverter circuit.
- The operating regions and the voltage levels of the driver and the load transistors at these critical points are listed below.

V_{in}	V_{out}	Driver operating region	Load operating region
V_{OL}	V_{OH}	cut-off	linear
V_{IL}	$\approx V_{OH}$	saturation	linear
V_{IH}	small	linear	<i>saturation</i>
V_{OH}	V_{OL}	linear	saturation

Calculation of V_{OH}

- When the input voltage V_{in} is smaller than the driver threshold voltage V_{T0} , the driver transistor is turned off and does not conduct any drain current.
- Consequently, the load device, which operates in the linear region, also has zero drain current.
- Substituting V_{OH} for V_{out} in (Eqn. 35), and letting the load current $I_{D,load} = 0$, we obtain

Calculation of V_{OH}

$$I_{D,load} = \frac{k_{n,load}}{2} \cdot \left[2|V_{T,load}(V_{OH})| \cdot (V_{DD} - V_{OH}) - (V_{DD} - V_{OH})^2 \right] = 0 \quad \dots\dots(36)$$

- The only valid solution in the linear region is
 $V_{OH} = V_{DD}$

Calculation of V_{OL}

- To calculate the output low voltage V_{OL} , we assume that the input voltage V_{in} of the inverter is equal to $V_{OH} = V_{DD}$.
- Note that in this case, the driver transistor operates in the linear region while the depletion-type load is in saturation.

$$\frac{k_{driver}}{2} \cdot [2 \cdot (V_{OH} - V_{T0}) \cdot V_{OL} - V_{OL}^2] = \frac{k_{load}}{2} \cdot [-V_{T,load}(V_{OL})]^2 \quad \dots\dots(37)$$

Calculation of V_{OL}

- This second-order equation in V_{OL} can be solved by temporarily neglecting the dependence of $V_{T,load}$ on V_{OL} , as follows.

$$V_{OL} = V_{OH} - V_{T0} - \sqrt{(V_{OH} - V_{T0})^2 - \left(\frac{k_{load}}{k_{driver}}\right) \cdot |V_{T,load}(V_{OL})|^2} \quad \dots\dots(38)$$

Calculation of V_{IL}

- By definition, the slope of the VTC is equal to (-1), i.e., $dV_{out}/dV_{in} = -1$ when the input voltage is $V_{in} = V_{IL}$.
- Note that in this case, the driver transistor operates in saturation while the load transistor operates in the linear region.

Calculation of V_{IL}

- Applying KCL for the output node, we obtain the following current equation:

$$\frac{k_{driver}}{2} \cdot (V_{in} - V_{T0})^2 = \frac{k_{load}}{2} \cdot \left[2|V_{T,load}(V_{out})| \cdot (V_{DD} - V_{out}) - (V_{DD} - V_{out})^2 \right] \dots\dots(39)$$

- To satisfy the derivative condition at V_{IL} we differentiate both sides of (Eqn. 39) with respect to V_{in}

Calculation of V_{IL}

$$k_{driver} \cdot (V_{in} - V_{T0}) = \frac{k_{load}}{2} \cdot \left[2|V_{T,load}(V_{out})| \left(-\frac{dV_{out}}{dV_{in}} \right) \right. \\ \left. + 2(V_{DD} - V_{out}) \left(-\frac{dV_{T,load}}{dV_{in}} \right) - 2(V_{DD} - V_{out}) \left(-\frac{dV_{out}}{dV_{in}} \right) \right] \\ \dots(40)$$

Calculation of V_{IL}

- In general, we can assume that the term $(dV_{T,load}/dV_{in})$ is negligible with respect to the others.
- Substituting V_{IL} for V_{in} , and letting $dV_{out}/dV_{in} = -1$, we obtain V_{IL} as a function of the output voltage V_{out}

$$V_{IL} = V_{T0} + \left(\frac{k_{load}}{k_{driver}} \right) \cdot [V_{out} - V_{DD} + |V_{T,load}(V_{out})|] \quad \dots(41)$$

Calculation of V_{IH}

- V_{IH} is the larger of the two voltage points on the VTC at which the slope is equal to (-1).
- Since the output voltage corresponding to this operating point is relatively small, the driver transistor is in the linear region and the load transistor is in saturation.

$$\frac{k_{driver}}{2} \cdot [2 \cdot (V_{in} - V_{T0}) \cdot V_{out} - V_{out}^2] = \frac{k_{load}}{2} \cdot [-V_{T,load}(V_{out})]^2 \quad \dots\dots(42)$$

Calculation of V_{IH}

- Differentiating both sides of (Eqn. 42) with respect to V_{in} , we obtain:

$$\begin{aligned} k_{driver} \cdot \left[V_{out} + (V_{in} - V_{T0}) \left(\frac{dV_{out}}{dV_{in}} \right) - V_{out} \left(\frac{dV_{out}}{dV_{in}} \right) \right] \\ = k_{load} \cdot \left[-V_{T,load}(V_{out}) \right] \cdot \left(\frac{dV_{T,load}}{dV_{out}} \right) \cdot \left(\frac{dV_{out}}{dV_{in}} \right) \end{aligned}$$

.....(43)

Calculation of V_{IH}

- Now, substitute $dV_{out} / dV_{in} = -1$ into (Eqn. 43), and solve for $V_{in} = V_{IH}$.

$$V_{IH} = V_{T0} + 2V_{out} + \left(\frac{k_{load}}{k_{driver}} \right) \cdot \left[-V_{T,load}(V_{out}) \right] \cdot \left(\frac{dV_{T,load}}{dV_{out}} \right) \quad \dots\dots(44)$$

Calculation of V_{IH}

- The derivative of the load threshold voltage with respect to the output voltage cannot be neglected in this case.

$$\frac{dV_{T,load}}{dV_{out}} = \frac{\gamma}{2\sqrt{|2\phi_F| + V_{out}}} \quad \text{.....(45)}$$

Depletion-Load nMOS Inverter

- It is seen from this discussion that the critical voltage points, the general shape of the inverter VTC, and ultimately, the noise margins, are determined essentially by the threshold voltages of the driver and the load devices, and by the driver-to-load ratio ($k_{\text{driver}}/k_{\text{load}}$).
- Figure 13 shows the VTCs of depletion-load inverter circuits with different driver-to-load ratios $k_R = (k_{\text{driver}}/k_{\text{load}})$.

Depletion-Load nMOS Inverter

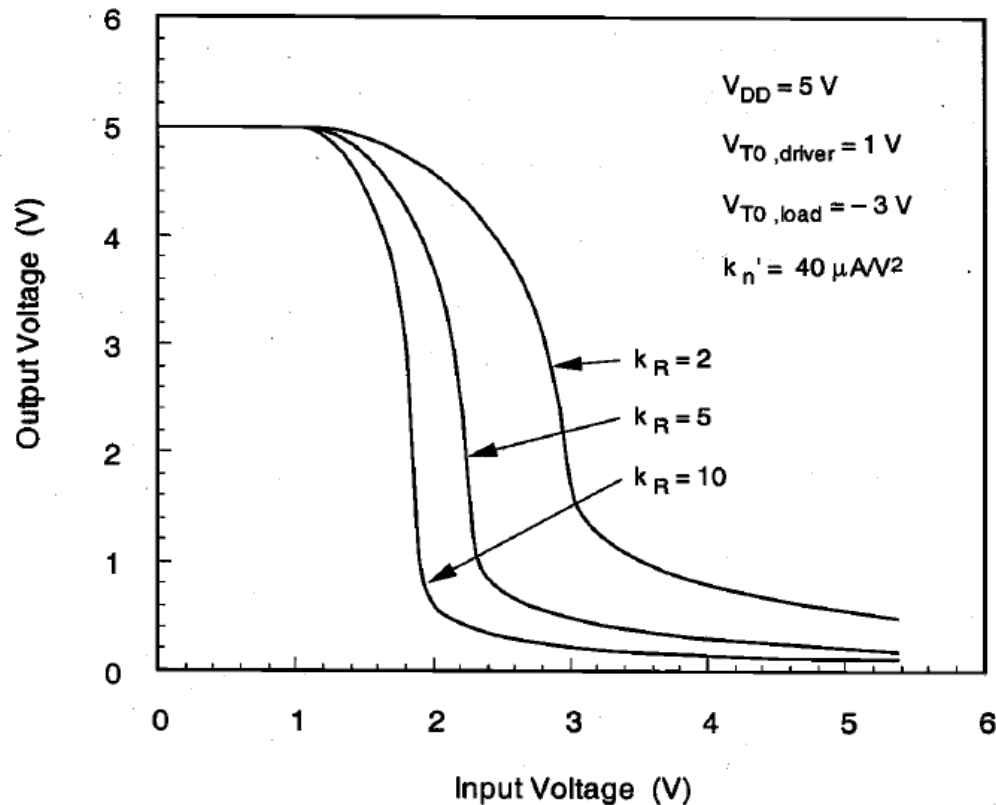


Figure 13. Voltage transfer characteristics of depletion-load inverters, with different driver-to-load ratios.

Depletion-Load nMOS Inverter

- One important observation is that, unlike in the enhancement-load inverter case, a sharp VTC transition and larger noise margins can be obtained with relatively small driver-to-load ratios.
- Thus, the total area occupied by a depletion-load inverter circuit with an acceptable circuit performance is expected to be much smaller than the area occupied by a comparable resistive-load or enhancement-load inverter.

Design of Depletion-Load Inverters

- Based on the VTC analysis given in the previous section, we can now consider the design of depletion-load inverters to satisfy certain DC performance criteria.
- The designable parameters in an inverter circuit are: (i) the power supply voltage V_{DD} , (ii) the threshold voltages of the driver and the load transistors, and (iii) the (W/L) ratios of the driver and the load transistors.

Design of Depletion-Load Inverters

- Equation (37) can be rearranged to calculate the driver-to-load ratio that achieves a target V_{OL} value, assuming that the power supply voltage and the threshold voltage values are set previously by independent design and processing constraints.

$$k_R = \frac{k_{driver}}{k_{load}} = \frac{|V_{T,load}(V_{OL})|^2}{2(V_{OH} - V_{T0})V_{OL} - V_{OL}^2} \quad \dots\dots(46)$$

Design of Depletion-Load Inverters

- Here, the driver-to-load ratio is given by;

$$k_R = \frac{k'_{n,driver} \cdot \left(\frac{W}{L}\right)_{driver}}{k'_{n,load} \cdot \left(\frac{W}{L}\right)_{load}} \dots\dots(47)$$

Design of Depletion-Load Inverters

- Since the channel doping densities and, consequently, the channel electron mobilities of the enhancement-type driver transistor and the depletion-type load transistor are not equal, we shall expect that $k_{n,driver}' \neq k_{n,load}'$
- Only if $k_{n,driver}' \approx k_{n,load}'$ can the driver-to-load ratio be reduced to

$$k_R = \frac{\left(\frac{W}{L}\right)_{driver}}{\left(\frac{W}{L}\right)_{load}} \dots\dots(48)$$

Design of Depletion-Load Inverters

- As a result, one can propose a number of designs with different (W/L) ratios for the driver (and for the load) device, each of which satisfies the driver-to-load ratio condition stated before.

Power and Area Considerations

- When the input voltage is low, i.e., when the driver transistor is cut-off and $V_{\text{out}} = V_{\text{OH}} = V_{\text{DD}}$, there is no significant current flow through the driver and the load transistors.
- Consequently, the inverter does not dissipate DC power under this condition.
- When the input is at the high state with $V_{\text{in}} \approx V_{\text{DD}}$ and $V_{\text{out}} = V_{\text{OL}}$, both the driver and the load transistors conduct a significant current, given by

Power and Area Considerations

$$\begin{aligned} I_{DC}(V_{in} = V_{DD}) &= \frac{k_{load}}{2} \cdot \left[-V_{T,load}(V_{OL}) \right]^2 \\ &= \frac{k_{driver}}{2} \cdot \left[2 \cdot (V_{OH} - V_{T0}) \cdot V_{OL} - V_{OL}^2 \right] \end{aligned}$$

.....(49)

Power and Area Considerations

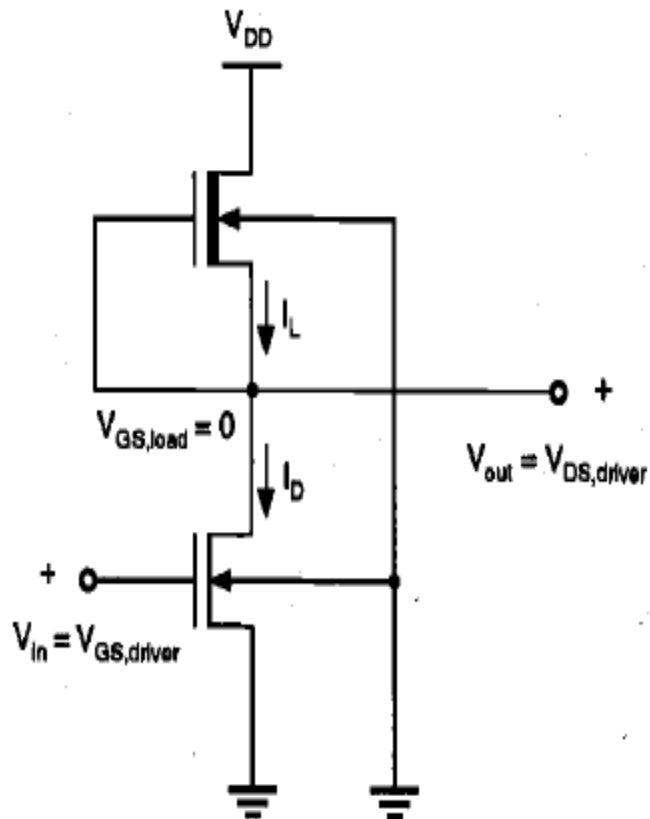
- Assuming that the input voltage level is low during 50% of the operation time and high during the other 50%, the overall average DC power consumption of this circuit can be estimated as follows:

$$P_{DC} = \frac{V_{DD}}{2} \cdot \frac{k_{load}}{2} \cdot \left[-V_{T,load}(V_{OL}) \right]^2 \quad \dots\dots(50)$$

Problem 3

- Calculate the critical voltages ($V_{OL}, V_{OH}, V_{IL}, V_{IH}$) and find the noise margins of the following depletion-load inverter circuit:

Problem 3



$$V_{DD} = 5 \text{ V}$$

$$V_{T0,driver} = 1.0 \text{ V}$$

$$V_{T0,load} = -3.0 \text{ V}$$

$$(W/L)_{driver} = 2, (W/L)_{load} = 1/3$$

$$k_{n,driver}' = k_{n,load}' = 25 \mu\text{A/V}^2$$

$$\gamma = 0.4 \text{ V}^{1/2}$$

$$\phi_F = -0.3 \text{ V}$$

Solution

- First, the output high voltage is simply found according to (Eqn. 36) as $V_{OH} = V_{DD} = 5 \text{ V}$.
- To calculate the output low voltage V_{OL} , (Eqn. 33) and (Eqn. 38) is solved simultaneously, using numerical iterations.
- We start the iterations by assuming that the output voltage is equal to zero; thus, letting $V_{T,\text{load}} = V_{T0,\text{load}} = -3 \text{ V}$. Solving (Eqn. 38) with this assumption yields a first-order estimate for V_{OL} .

Solution

$$\begin{aligned} V_{OL} &= V_{OH} - V_{T0} - \sqrt{(V_{OH} - V_{T0})^2 - \left(\frac{k_{load}}{k_{driver}}\right) \cdot |V_{T,load}(V_{OL})|^2} \\ &= 5 - 1 - \sqrt{(5 - 1)^2 - \left(\frac{1}{6}\right) |3|^2} = 0.192 \text{ V} \end{aligned}$$

Solution

- Now, the threshold voltage of the depletion-load device can be updated by substituting this output voltage into (Eqn. 33).

$$\begin{aligned} V_{T,load} &= V_{T0,load} + \gamma \left(\sqrt{|2\phi_F| + V_{OL}} - \sqrt{|2\phi_F|} \right) \\ &= -3 + 0.4 \left(\sqrt{0.6 + 0.2} - \sqrt{0.6} \right) = -2.95 \text{ V} \end{aligned}$$

Solution

- Using this new value for $V_{T,load}$ we now recalculate V_{OL} , again according to (Eqn. 38).

$$V_{OL} = 0.186 \text{ V}$$

$$V_{T,load} = -2.95 \text{ V}$$

- At this point, we can stop the iteration process since the threshold voltage of the load device has not changed in the two significant digits after the decimal point.
- Continuing the iteration would not produce a perceptible improvement of V_{OL} .

Solution

- The calculation of V_{IL} involves simultaneous solution of (Eqn. 33), (Eqn. 39), and (Eqn. 41), using numerical iterations.
- When the input voltage is equal to V_{IL} , we expect that the output voltage is slightly lower than the, output high voltage, V_{OH} .
- As a first-order approximation, assume that $V_{out} = V_{OH} = 5 \text{ V}$ for $V_{in} = V_{IL}$.

Solution

- Then, the threshold voltage of the load device can be estimated as $V_{T,load}(V_{out} = 5 \text{ V}) = -2.36 \text{ V}$.
- Substituting this value into (5.41) gives V_{IL} as a function of the output voltage V_{out} .

$$\begin{aligned} V_{IL}(V_{out}) &= V_{T0} + \frac{k_{load}}{k_{driver}} \cdot [V_{out} - V_{DD} + |V_{T,load}(V_{out})|] \\ &= 1 + \left(\frac{1}{6}\right)(V_{out} - 5 + 2.36) = 0.167 V_{out} + 0.56 \end{aligned}$$

Solution

- This expression can be rearranged as;

$$V_{out} = 6 V_{IL} - 3.35$$

- Now, substitute this into the KCL equation (5.39) to obtain the following quadratic equation for V_{IL} :

$$\frac{k_{driver}}{2} \cdot (V_{IL} - V_{T0})^2 = \frac{k_{load}}{2} \cdot \left[2 |V_{T,load}(V_{out})| \cdot (V_{DD} - 6 V_{IL} + 3.35) - (V_{DD} - 6 V_{IL} + 3.35)^2 \right]$$

Solution

$$2 \cdot (V_{IL} - 1)^2 = \frac{1}{3} \cdot \left[2 \cdot 2.36 \cdot (5 - 6V_{IL} + 3.35) - (5 - 6V_{IL} + 3.35)^2 \right]$$

- The solution of this second-order equation yields two possible values for V_{IL} .

$$V_{IL} = \begin{cases} 0.98 \text{ V} \\ \underline{\underline{1.36 \text{ V}}} \end{cases}$$

Solution

- Note that V_{IL} must be larger than the threshold voltage V_t of the driver transistor, hence, $V_{IL} = 1.36 \text{ V}$ is the physically correct solution.
- The output voltage level at this point can also be found as

$$V_{out} = 6 \cdot 1.36 - 3.35 = 4.81 \text{ V}$$

Solution

- which significantly improves our initial assumption of $V_{\text{out}} = 5 \text{ V}$.
- At this point, the threshold voltage of the load transistor must be recalculated, in order to update its value.
- Substituting $V_{\text{out}} = 4.81 \text{ V}$ into (Eqn. 33) yields $V_{\text{T,load}} = -2.38 \text{ V}$.
- We observe that this value is only slightly higher (by 20 mV) than the threshold voltage value used in the previous calculations.

Solution

- For practical purposes, we can terminate the numerical iteration at this stage and accept $V_{IL} = 1.36 \text{ V}$ as a fairly accurate estimate.
- To calculate V_{IH} , we first have to find the numerical value of $(dV_{T,load} / V_{out})$ using (Eqn. 45).
- When the input voltage is equal to V_{IH} , the output voltage is expected to be relatively low.

Solution

- As a first-order approximation, assume that the output voltage level is $V_{\text{out}} = V_{\text{OL}} \approx 0.2 \text{ V}$ when $V_{\text{in}} = V_{\text{IH}}$.
- The threshold voltage of the load device can also be estimated as $V_{T,\text{load}}(V_{\text{out}} = 0.2 \text{ V}) = -2.95 \text{ V}$. Thus,

$$\frac{dV_{T,\text{load}}}{dV_{\text{out}}} = \frac{\gamma}{2\sqrt{|2\phi_F| + V_{\text{out}}}} = \frac{0.4}{2\sqrt{0.6 + 0.2}} = 0.22$$

Solution

- This value can now be used in (Eqn. 44) to find V_{IH} as a function of the output voltage.

$$\begin{aligned} V_{IH}(V_{out}) &= V_{T0} + 2V_{out} + \frac{k_{load}}{k_{driver}} \cdot [-V_{T,load}(V_{out})] \cdot \left(\frac{dV_{T,load}}{dV_{out}} \right) \\ &= 1 + 2V_{out} + \left(\frac{1}{6} \right) \cdot 2.95 \cdot 0.22 = 2V_{out} + 1.1 \end{aligned}$$

- This expression is rearranged as:

$$V_{out} = 0.5V_{IH} - 0.55$$

Solution

- Next, substitute V_{out} , in the KCL equation (Eqn. 42), to obtain

$$2 \cdot \left[2 \cdot (V_{IH} - 1) \cdot (0.5 V_{IH} - 0.55) - (0.5 V_{IH} - 0.55)^2 \right] = \frac{1}{3} \cdot (2.95)^2$$

- The solution of this simple quadratic equation yields two values for V_{IH} .

$$V_{IH} = \begin{cases} -0.35 \text{ V} \\ \underline{\underline{2.43 \text{ V}}} \end{cases}$$

Solution

- where $V_{IH} = 2.43 \text{ V}$ is the physically correct solution. The output voltage level at this point is calculated as

$$V_{out} = 0.5 \cdot 2.43 - 0.55 = 0.67 \text{ V}$$

- With this updated output voltage value, we can now reevaluate the load threshold voltage as $V_{T,load} (V_{out} = 0.67 \text{ V}) = -2.9 \text{ V}$, and the $(dV_{T,load} / V_{out})$ value as

$$\frac{dV_{T,load}}{dV_{out}} = 0.18$$

Solution

- Note both of these values are fairly close to those used at the beginning of this iteration process.
- Repeating the iterative calculation will provide only a marginal improvement of accuracy, thus, we may accept $V_{IH} = 2.43 \text{ V}$ as a good estimate.

Solution

- In conclusion, the noise margins for high signal levels and for low signal levels can be found as follows:

$$NM_H = V_{OH} - V_{IH} = 2.57 \text{ V}$$

$$NM_L = V_{IL} - V_{OL} = 1.17 \text{ V}$$

CMOS Inverter

- In the nMOS inverter, the input signal is always applied to the gate of the driver transistor, and the operation of the inverter is controlled primarily by switching the driver.
- Now, we will turn our attention to a radically different inverter structure, which consists of an enhancement-type nMOS transistor and an enhancement-type pMOS transistor, operating in complementary mode as shown in Fig. 14.

CMOS Inverter

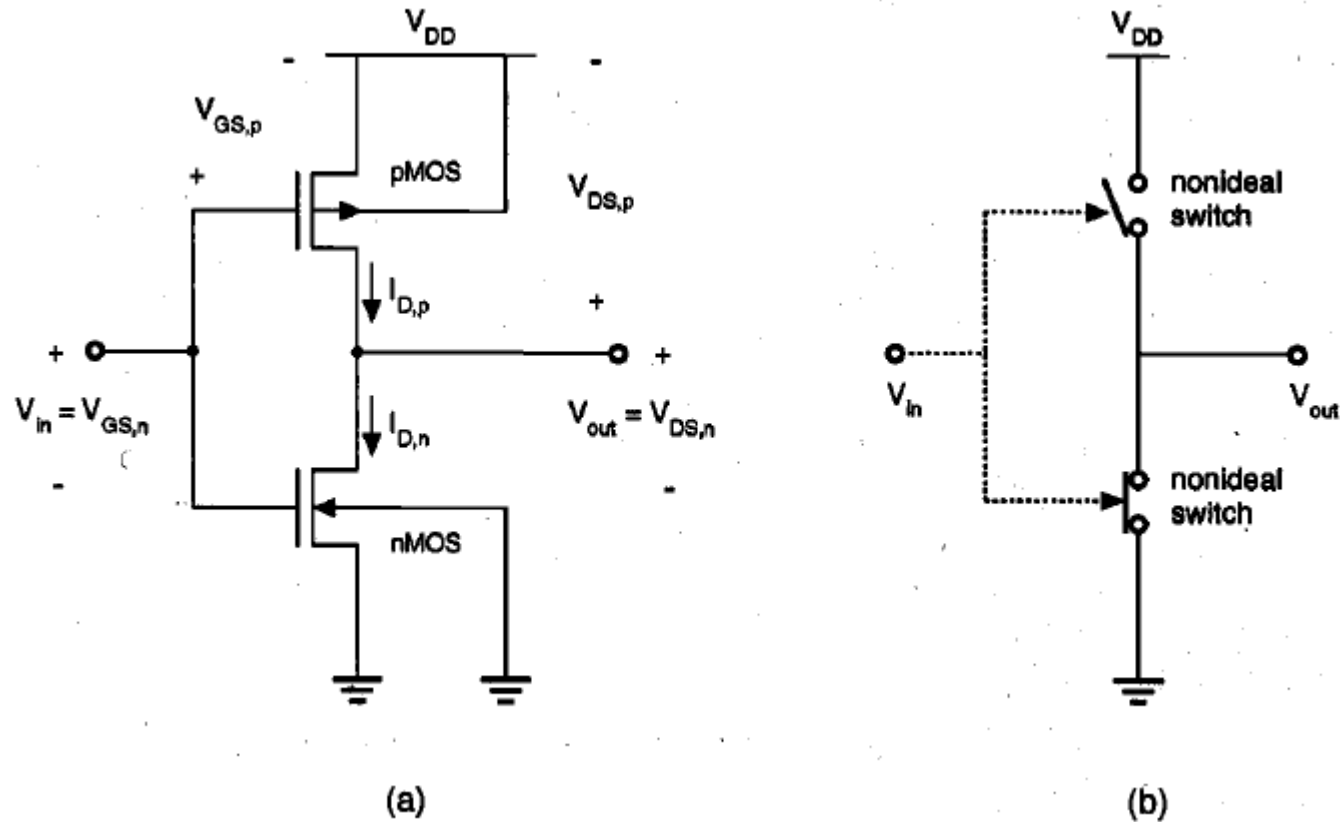


Figure 14 (a) CMOS inverter circuit. (b) Simplified view of the CMOS inverter, consisting of two complementary nonideal switches.

CMOS Inverter

- This configuration is called Complementary MOS (CMOS).
- The circuit topology is complementary push-pull in the sense that for high input, the nMOS transistor drives (pulls down) the output node while the pMOS transistor acts as the load, and for low input the pMOS transistor drives (pulls up) the output node while the nMOS transistor acts as the load.
- Consequently, both devices contribute equally to the circuit operation characteristics.

CMOS Inverter

- The CMOS inverter has two important advantages over the other inverter configurations.
- The first and the most important advantage is that the steady-state power dissipation of the CMOS inverter circuit is virtually negligible whereas in all other inverter structures examined so far, a nonzero steady-state current is drawn from the power source when the driver transistor is turned on, which results in a significant DC power consumption.

CMOS Inverter

- The other advantages of the CMOS configuration are that the voltage transfer characteristic (VTC) exhibits a full output voltage swing between 0 V and V_{DD} , and that the VTC transition is usually very sharp.
- Thus, the VTC of the CMOS inverter resembles that of an ideal inverter.

Circuit Operation

- In Fig. 14, the input voltage is connected to the gate terminals of both the nMOS and the pMOS transistors.
- Thus, both transistors are driven directly by the input signal, V_{in} .
- The substrate of the nMOS transistor is connected to the ground, while the substrate of the pMOS transistor is connected to the power supply voltage, V_{DD} , in order to reverse-bias the source and drain junctions.

Circuit Operation

- Since $V_{SB} = 0$ for both devices, there will be no substrate-bias effect for either device.

$$\begin{aligned} V_{GS,n} &= V_{in} \\ V_{DS,n} &= V_{out} \end{aligned} \quad \text{.....(51)}$$

$$\begin{aligned} V_{GS,p} &= -(V_{DD} - V_{in}) \\ V_{DS,p} &= -(V_{DD} - V_{out}) \end{aligned} \quad \text{.....(52)}$$

Circuit Operation

- When the input voltage is smaller than the nMOS threshold voltage, i.e., when $V_{in} < V_{T0,n}$, the nMOS transistor is cut-off. At the same time, the pMOS transistor is on, operating in the linear region.
- Since the drain currents of both transistors are approximately equal to zero (except for small leakage currents), i.e.,

$$I_{D,n} = I_{D,p} = 0 \quad \text{.....(53)}$$

Circuit Operation

- the drain-to-source voltage of the pMOS transistor is also equal to zero, and the output voltage V_{OH} is equal to the power supply voltage.

$$V_{out} = V_{OH} = V_{DD} \quad \text{.....(54)}$$

Circuit Operation

- On the other hand, when the input voltage exceeds $(V_{DD} + V_{T0,p})$ the pMOS transistor is turned off. In this case, the nMOS transistor is operating in the linear region, but its drain-to-source voltage is equal to zero because condition (Eqn. 53) is satisfied.
- Consequently, the output voltage of the circuit is

$$V_{out} = V_{OL} = 0 \quad \dots\dots(55)$$

Circuit Operation

- Next, we examine the operating modes of the nMOS and the pMOS transistors as functions of the input and output voltages.
- The nMOS transistor operates in saturation if $V_{in} > V_{T0,n}$ and if the following condition is satisfied.

$$V_{DS,n} \geq V_{GS,n} - V_{T0,n} \quad \Leftrightarrow \quad V_{out} \geq V_{in} - V_{T0,n} \quad \dots\dots(56)$$

Circuit Operation

- The pMOS transistor operates in saturation if $V_{in} < (V_{DD} + V_{T0,p})$, and if:

$$V_{DS,p} \leq V_{GS,p} - V_{T0,p} \quad \Leftrightarrow \quad V_{out} \leq V_{in} - V_{T0,p} \quad \dots\dots(57)$$

- Both of these conditions for device saturation are illustrated graphically as shaded areas in Fig. 15.

Circuit Operation

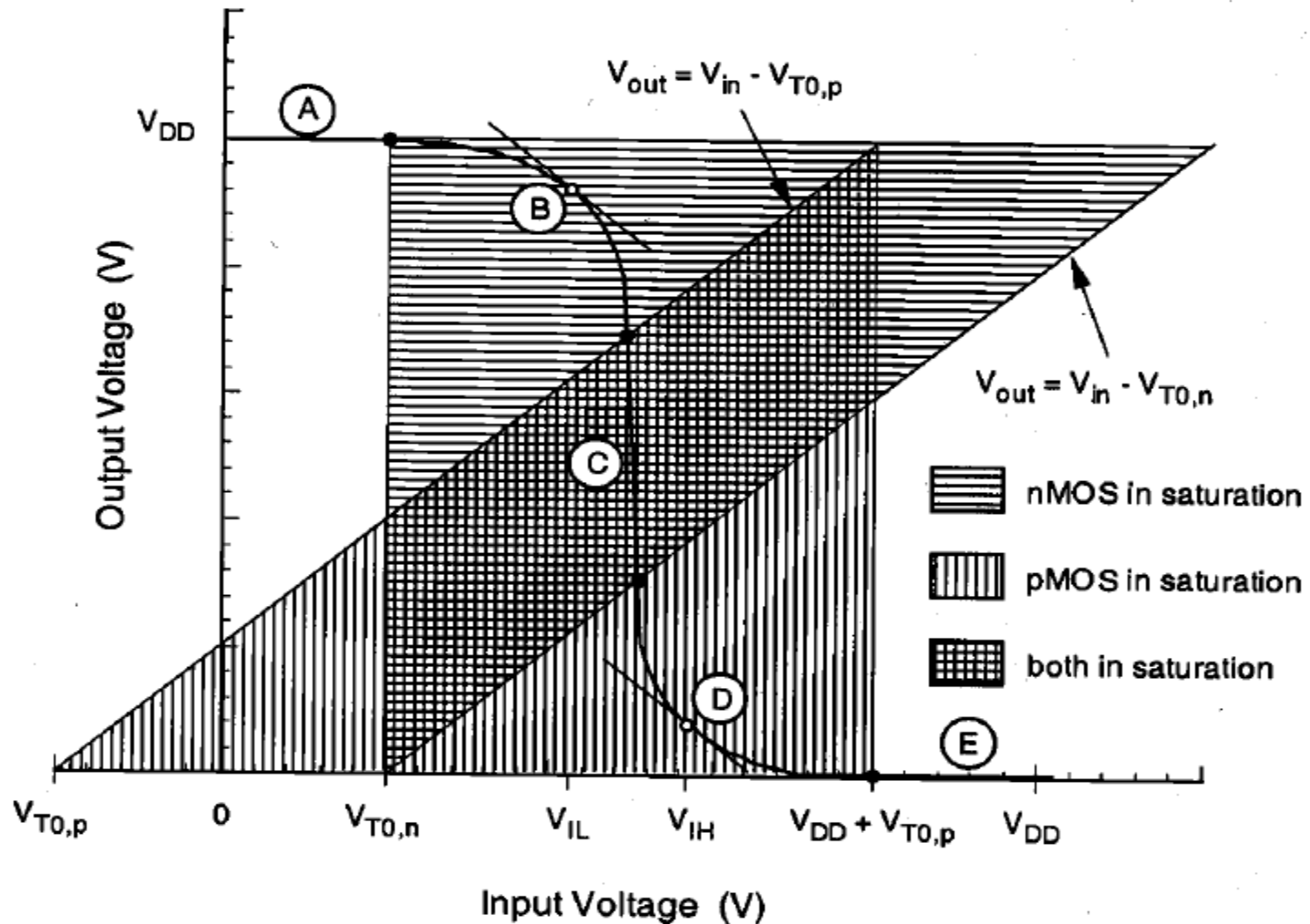


Fig. 15: Operating Regions of the nMOS pMOS Transistor

Circuit Operation

- The table below lists the operating regions and the corresponding critical input and output voltage levels.

Region	V_{in}	V_{out}	nMOS	pMOS
A	$< V_{T0,n}$	V_{OH}	cut-off	linear
B	V_{IL}	high $\approx V_{OH}$	saturation	linear
C	V_{th}	V_{th}	saturation	saturation
D	V_{IH}	low $\approx V_{OL}$	linear	saturation
E	$> (V_{DD} + V_{T0,p})$	V_{OL}	linear	cut-off

Circuit Operation

- In Region A, where $V_{in} < V_{T0,n}$, the nMOS transistor is cut-off and the output voltage is equal to $V_{OH} = V_{DD}$.
- As the input voltage is increased beyond $V_{T0,n}$ (into Region B), the nMOS transistor starts conducting in saturation mode and the output voltage begins to decrease. Also the critical voltage V_{IL} which corresponds to $(dV_{out}/dV_{in}) = -1$ is located within Region B.

Circuit Operation

- As the output voltage further decreases, the pMOS transistor enters saturation at the boundary of Region C.
- It is seen from Fig. 15 that the inverter threshold voltage, where $V_{in} = V_{out}$, is located in Region C.
- When the output voltage V_{out} falls below $(V_{in} - V_{T0,n})$, the nMOS transistor starts to operate in linear mode.

Circuit Operation

- This corresponds to Region D in Fig. 15, where the critical voltage point V_{IH} with $(dV_{out}/dV_{in}) = -1$ is also located.
- Finally, in Region E, with the input voltage $V_{in} > (V_{DD} + V_{T0,p})$, the pMOS transistor is cut-off, and the output voltage is $V_{OL} = 0$.

Circuit Operation

- The steady-state input-output voltage characteristics of the CMOS inverter can be visualized by considering the interaction of individual nMOS and pMOS transistor characteristics in the current-voltage space.
- The drain current $I_{D,n}$ of the nMOS transistor is a function of the voltages $V_{GS,n}$ and $V_{DS,n}$.
- Hence, the nMOS drain current is also a function of the inverter input and output voltages V_{in} and V_{out} according to (Eqn. 51).

$$I_{D,n} = f(V_{in}, V_{out})$$

Circuit Operation

- Similarly, the drain current $I_{D,p}$ of the pMOS transistor is also a function of the inverter input and output voltages V_{in} and V_{out} , according to (Eqn. 52).

$$I_{D,p} = f(V_{in}, V_{out})$$

Circuit Operation

- In a CMOS inverter operating in steady-state, the drain current of the nMOS transistor is always equal to the drain current of the pMOS transistor, according to KCL.

$$I_{D,n} = I_{D,p}$$

Critical Voltage Points

- The critical voltage points of the CMOS inverter on the VTC is now represented.
- It has already been established that $V_{OH} = V_{DD}$ and $V_{OL} = 0$ for this inverter.

Calculation of V_{IL}

- By definition, the slope of the VTC is equal to (-1), i.e., $dV_{out}/dV_{in} = -1$ when the input voltage is $V_{in} = V_{IL}$.
- In this case, the nMOS transistor operates in saturation while the pMOS transistor operates in the linear region.
- From $I_{D,n} = I_{D,p}$ we obtain the following current equation:

$$\frac{k_n}{2} \cdot (V_{GS,n} - V_{T0,n})^2 = \frac{k_p}{2} \cdot \left[2 \cdot (V_{GS,p} - V_{T0,p}) \cdot V_{DS,p} - V_{DS,p}^2 \right] \quad \dots\dots(58)$$

Calculation of V_{IL}

- Using equations (Eqn. 51) and (Eqn. 52), this expression can be rewritten as;

$$\frac{k_n}{2} \cdot (V_{in} - V_{T0,n})^2 = \frac{k_p}{2} \cdot \left[2 \cdot (V_{in} - V_{DD} - V_{T0,p}) \cdot (V_{out} - V_{DD}) - (V_{out} - V_{DD})^2 \right] \dots\dots(59)$$

Calculation of V_{IL}

- To satisfy the derivative condition at V_{IL} , differentiate both sides of (Eqn. 59) with respect to V_{in} .

$$\begin{aligned} k_n \cdot (V_{in} - V_{T0,n}) &= k_p \cdot \left[(V_{in} - V_{DD} - V_{T0,p}) \cdot \left(\frac{dV_{out}}{dV_{in}} \right) + (V_{out} - V_{DD}) \right. \\ &\quad \left. - (V_{out} - V_{DD}) \cdot \left(\frac{dV_{out}}{dV_{in}} \right) \right] \end{aligned} \quad \text{.....(60)}$$

Calculation of V_{IL}

- Substituting $V_{in} = V_{IL}$ and $(dV_{out}/dV_{in}) = -1$ in (Eqn. 60), we obtain;

$$k_n \cdot (V_{IL} - V_{T0,n}) = k_p \cdot (2V_{out} - V_{IL} + V_{T0,p} - V_{DD})$$

.....(61)

Calculation of V_{IL}

- The critical voltage V_{IL} can now be found as a function of the output voltage V_{out} , as follows:

$$V_{IL} = \frac{2V_{out} + V_{T0,p} - V_{DD} + k_R V_{T0,n}}{1 + k_R} \quad \text{.....(62)}$$

- where k_R is defined as

$$k_R = \frac{k_n}{k_p}$$

Calculation of V_{IH}

- When the input voltage is equal to V_{IH} , the nMOS transistor operates in the linear region, and the pMOS transistor operates in saturation.
- Applying KCL to the output node, we obtain;

$$\frac{k_n}{2} \cdot \left[2 \cdot (V_{GS,n} - V_{T0,n}) \cdot V_{DS,n} - V_{DS,n}^2 \right] = \frac{k_p}{2} \cdot (V_{GS,p} - V_{T0,p})^2 \quad \dots\dots(63)$$

Calculation of V_{IH}

- Using equations (Eqn. 51) and (Eqn. 52), this expression can be rewritten as;

$$\frac{k_n}{2} \cdot [2 \cdot (V_{in} - V_{T0,n}) \cdot V_{out} - V_{out}^2] = \frac{k_p}{2} \cdot (V_{in} - V_{DD} - V_{T0,p})^2$$

.....(64)

Calculation of V_{IH}

- Now, differentiating both sides of (Eqn. 64) with respect to V_{in} .

$$\begin{aligned} k_n \cdot \left[\left(V_{in} - V_{T0,n} \right) \cdot \left(\frac{dV_{out}}{dV_{in}} \right) + V_{out} - V_{out} \cdot \left(\frac{dV_{out}}{dV_{in}} \right) \right] \\ = k_p \cdot \left(V_{in} - V_{DD} - V_{T0,p} \right) \end{aligned}$$

.....(65)

Calculation of V_{IH}

- Substituting, $V_{in} = V_{IH}$ and $(dV_{out}/dV_{in}) = -1$ in (Eqn. 65), we obtain;

$$k_n \cdot (-V_{IH} + V_{T0,n} + 2V_{out}) = k_p \cdot (V_{IH} - V_{DD} - V_{T0,p}) \quad \text{.....(66)}$$

Calculation of V_{IH}

- The critical voltage V_{IH} can now be found as a function of V_{out} , as follows:

$$V_{IH} = \frac{V_{DD} + V_{T0,p} + k_R \cdot (2V_{out} + V_{T0,n})}{1 + k_R} \quad \dots\dots(67)$$

- This equation is solved simultaneously with the KCL equation (64) to obtain the numerical values of V_{IH} and V_{out} .

Calculation of V_{th}

- The inverter threshold voltage is defined as $V_{th} = V_{in} = V_{out}$.
- Since the CMOS inverter exhibits large noise margins and a very sharp VTC transition, the inverter threshold voltage emerges as an important parameter characterizing the DC performance of the inverter.

Calculation of V_{th}

- For $V_{in} = V_{out}$, both transistors are expected to be in saturation mode;
- Hence, we can write the following KCL equation.

$$\frac{k_n}{2} \cdot (V_{GS,n} - V_{T0,n})^2 = \frac{k_p}{2} \cdot (V_{GS,p} - V_{T0,p})^2 \quad \dots\dots(68)$$

Calculation of V_{th}

- Replacing $V_{GS,n}$ and $V_{GS,p}$ in (Eqn. 68) according to (Eqn. 51) and (Eqn. 52), we obtain;

$$\frac{k_n}{2} \cdot (V_{in} - V_{T0,n})^2 = \frac{k_p}{2} \cdot (V_{in} - V_{DD} - V_{T0,p})^2 \quad \dots\dots(69)$$

Calculation of V_{th}

- The correct solution for V_{in} for this equation is

$$V_{in} \cdot \left(1 + \sqrt{\frac{k_p}{k_n}} \right) = V_{T0,n} + \sqrt{\frac{k_p}{k_n}} \cdot (V_{DD} + V_{T0,p}) \quad \dots\dots(70)$$

Calculation of V_{th}

- Finally, the inverter threshold (switching threshold) voltage V_{th} is found as;

$$V_{th} = \frac{V_{T0,n} + \sqrt{\frac{1}{k_R}} \cdot (V_{DD} + V_{T0,p})}{\left(1 + \sqrt{\frac{1}{k_R}}\right)} \quad \text{.....(71)}$$

Calculation of V_{th}

- It has already been established that the CMOS inverter does not draw any significant current from the power source, except for small leakage and subthreshold currents, when the input voltage is either smaller than $V_{T0,n}$, or larger than $(V_{DD} + V_{T0,p})$.
- The nMOS and the pMOS transistors conduct a nonzero current, on the other hand, during low-to-high and high-to-low transitions, i.e., in Regions B, C, and D.

Calculation of V_{th}

- It can be shown that the current being drawn from the power source during transition reaches its peak value when $V = V_{th}$.
- In other words, the maximum current is drawn when both transistors are operating in saturation mode.
- Figure 16 shows the voltage transfer characteristic of a typical CMOS inverter circuit and the power supply current, as a function of the input voltage.

VTC of a CMOS Inverter

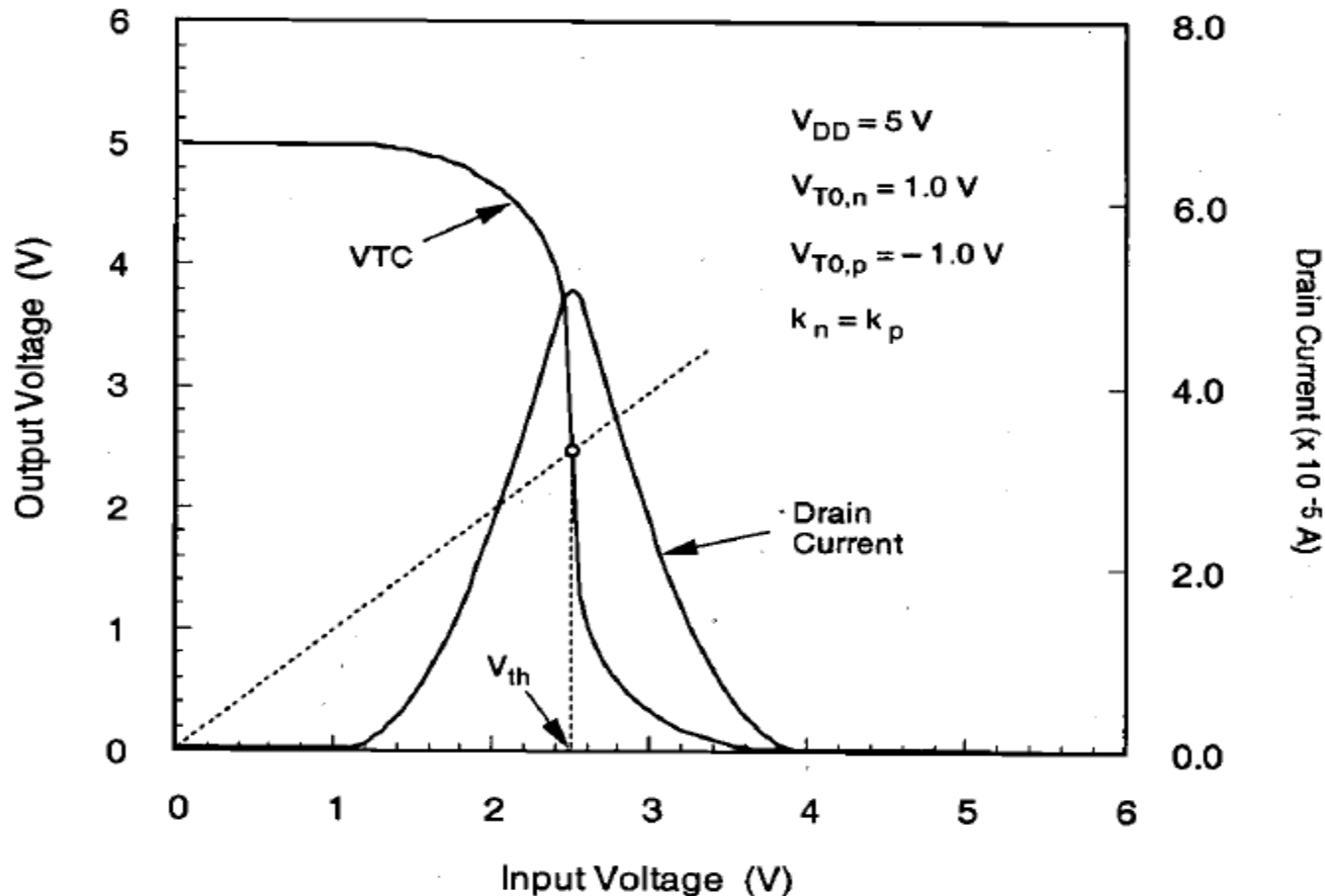


Figure 16. Typical VTC and the power supply current of a CMOS inverter circuit.

Design of CMOS Inverters

- The inverter threshold voltage V_{th} was identified as one of the most important parameters that characterize the steady-state input-output behavior of the CMOS inverter circuit.
- Given the power supply voltage V_{DD} , the nMOS and the pMOS transistor threshold voltages, and the desired inverter threshold voltage V_{th} , the corresponding ratio k_R can be found as follows. Reorganizing (Eqn. 71) yields

Design of CMOS Inverters

$$\sqrt{\frac{1}{k_R}} = \frac{V_{th} - V_{T0,n}}{V_{DD} + V_{T0,p} - V_{th}} \quad \text{.....(72)}$$

- Now k_R is solved that is required to achieve the given V_{th} .

$$k_R = \frac{k_n}{k_p} = \left(\frac{V_{DD} + V_{T0,p} - V_{th}}{V_{th} - V_{T0,n}} \right)^2 \quad \text{.....(73)}$$

Design of CMOS Inverters

- Recalling that the switching threshold voltage of an ideal inverter is defined as;

$$V_{th,ideal} = \frac{1}{2} \cdot V_{DD} \quad \dots\dots(74)$$

- Substituting (Eqn. 74) in (Eqn. 73) gives

$$\left(\frac{k_n}{k_p} \right)_{ideal} = \left(\frac{0.5 V_{DD} + V_{T0,p}}{0.5 V_{DD} - V_{T0,n}} \right)^2 \quad \dots\dots(75)$$

Design of CMOS Inverters

- For a near-ideal CMOS VTC that satisfies the condition (Eqn. 74).
- Since the operations of the nMOS and the pMOS transistors of the CMOS inverter are fully complementary, we can achieve completely symmetric input-output characteristics by setting the threshold voltages as $V_{T0} = V_{T0,n} = |V_{T0,p}|$. This reduces (Eqn. 75) to:

$$\left(\frac{k_n}{k_p} \right)_{\text{symmetric inverter}} = 1 \quad \dots\dots(76)$$

Design of CMOS Inverters

- The ratio k_R is defined as;

$$\frac{k_n}{k_p} = \frac{\mu_n C_{ox} \cdot \left(\frac{W}{L}\right)_n}{\mu_p C_{ox} \cdot \left(\frac{W}{L}\right)_p} = \frac{\mu_n \cdot \left(\frac{W}{L}\right)_n}{\mu_p \cdot \left(\frac{W}{L}\right)_p} \quad \dots\dots(77)$$

Design of CMOS Inverters

- Assuming that the gate oxide thickness t_{ox} , and hence, the gate oxide capacitance C_{ox} have the same value for both nMOS and pMOS transistors.
- The unity-ratio condition (Eqn. 76) for the ideal symmetric inverter requires that

$$\frac{\left(\frac{W}{L}\right)_n}{\left(\frac{W}{L}\right)_p} = \frac{\mu_p}{\mu_n} \approx \frac{230 \text{ cm}^2/\text{V}\cdot\text{s}}{580 \text{ cm}^2/\text{V}\cdot\text{s}} \quad \dots\dots(78)$$

$$\left(\frac{W}{L}\right)_p \approx 2.5 \left(\frac{W}{L}\right)_n \quad \dots\dots(79)$$

Design of CMOS Inverters

- The VTCs of three CMOS inverter circuits with different k_R ratios are shown in Fig. 17.
- It can be seen clearly that the inverter threshold voltage V_{th} shifts to lower values with increasing k_R ratio.

Design of CMOS Inverters

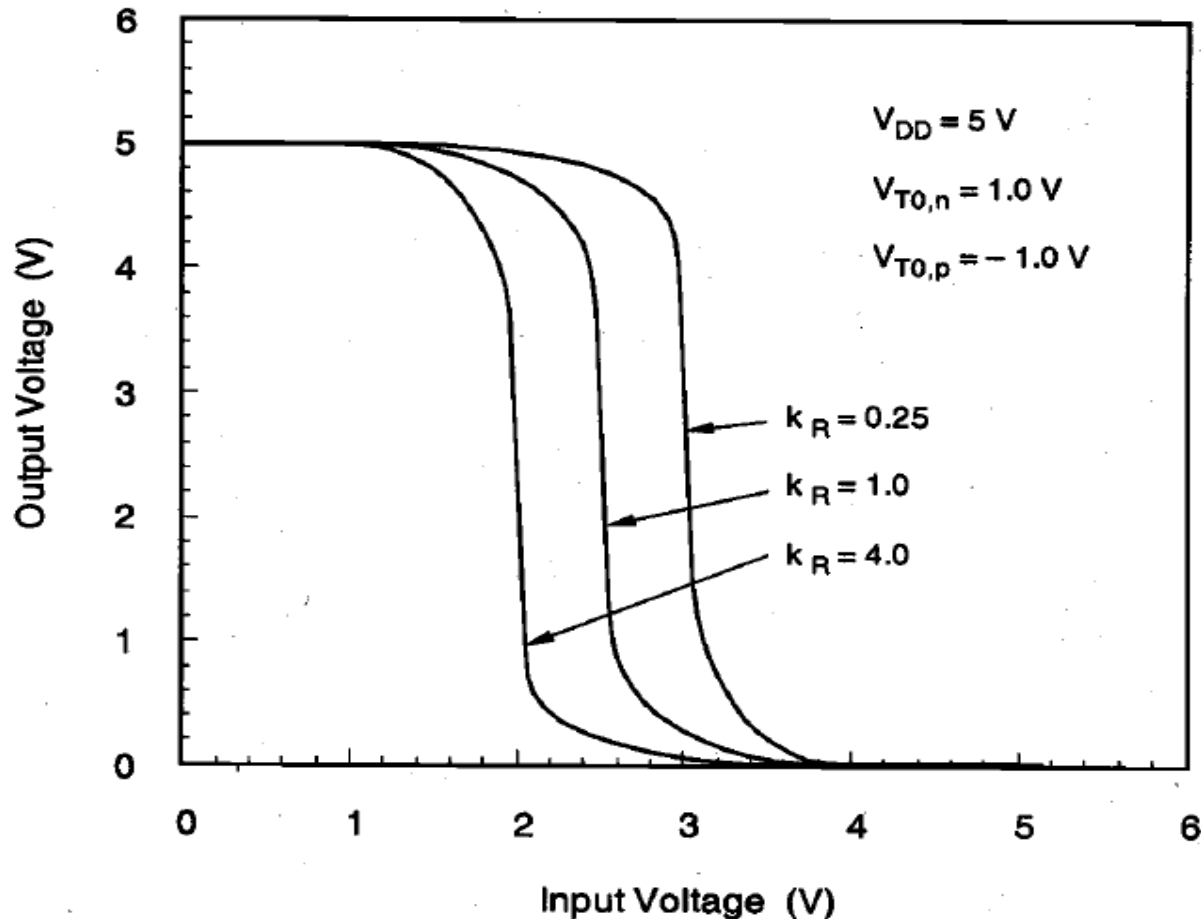


Fig. 17: Voltage transfer characteristics of three CMOS inverters, with different nMOS –to-pMOS ratios

Design of CMOS Inverters

- For a symmetric CMOS inverter with $V_{T0,n} = |V_{T0,p}|$ and $k_R = 1$, the critical voltage V_{IL} can be found, using (Eqn. 62), as follows:

$$V_{IL} = \frac{1}{8} \cdot (3V_{DD} + 2V_{T0,n}) \quad \text{.....(80)}$$

- Also the critical voltage V_{IH} can be found as;

$$V_{IH} = \frac{1}{8} \cdot (5V_{DD} - 2V_{T0,n}) \quad \text{.....(81)}$$

Design of CMOS Inverters

- The sum of V_{IL} and V_{IH} is always equal to V_{DD} in a symmetric inverter.

$$V_{IL} + V_{IH} = V_{DD} \quad \text{.....(82)}$$

- The noise margins NM_L and NM_H for this symmetric CMOS inverter are now calculated using (Eqn. 3) and (Eqn. 4).

$$\begin{aligned} NM_L &= V_{IL} - V_{OL} = V_{IL} \\ NM_H &= V_{OH} - V_{IH} = V_{DD} - V_{IH} \end{aligned} \quad \text{.....(83)}$$

- which are equal to each other, and also to V_{IL} .

$$NM_L = NM_H = V_{IL} \quad \text{.....(84)}$$

Problem 4

- Consider a CMOS inverter circuit with the following parameters:

$$\begin{aligned}V_{DD} &= 3.3 \text{ V} \\V_{T0,n} &= 0.6 \text{ V} \\V_{T0,p} &= -0.7 \text{ V} \\k_n &= 200 \mu\text{A/V}^2 \\k_p &= 80 \mu\text{A/V}^2\end{aligned}$$

- Calculate the noise margins of the circuit. Notice that the CMOS inverter being considered here has $k_R = 2.5$ and $V_{T0,n} \neq |V_{T0,p}|$; hence, it is not a symmetric inverter.

Solution

- First, the output low voltage V_{OL} and the output high voltage V_{OH} are found, using (Eqn. 54) and (Eqn. 55), as $V_{OL} = 0$ and $V_{OH} = 5$ V.
- To calculate V_{IL} in terms of the output voltage, we use (Eqn. 62).

$$\begin{aligned} V_{IL} &= \frac{2V_{out} + V_{T0,p} - V_{DD} + k_R V_{T0,n}}{1 + k_R} \\ &= \frac{2V_{out} - 0.7 - 3.3 + 1.5}{1 + 2.5} = 0.57V_{out} - 0.71 \end{aligned}$$

Solution

- Now substitute this expression into the KCL equation (Eqn. 59).

$$2.5(0.57V_{out} - 0.71 - 0.6)^2 = 2(0.57V_{out} - 0.71 - 3.3 + 0.7)(V_{out} - 3.3) - (V_{out} - 3.3)^2$$

- This expression yields a second-order polynomial in V_{out} as follows:

$$0.66V_{out}^2 + 0.05V_{out} - 6.65 = 0$$

Solution

- Only one root of this quadratic equation corresponds to a physically correct solution for V_{out} (i.e., $V_{\text{out}} > 0$).

$$V_{\text{out}} = 3.14 \text{ V}$$

- From this value, we can calculate the critical voltage V_{IL} as:

$$V_{\text{IL}} = 0.57 \cdot 3.14 - 0.71 = \underline{\underline{1.08 \text{ V}}}$$

Solution

- To calculate V_{IH} in terms of the output voltage, use (Eqn. 67):

$$\begin{aligned} V_{IH} &= \frac{V_{DD} + V_{T0,p} + k_R \cdot (2V_{out} + V_{T0,n})}{1 + k_R} \\ &= \frac{3.3 - 0.7 + 2.5(2V_{out} + 0.6)}{1 + 2.5} = 1.43V_{out} + 1.17 \end{aligned}$$

Solution

- Next, substitute this expression into the KCL equation (Eqn. 64) to obtain a second-order polynomial in V_{out}

$$2.5[2(1.43V_{out} + 1.17 - 0.6)V_{out} - V_{out}^2] = (1.43V_{out} - 1.43)^2$$

$$2.61V_{out}^2 + 6.94V_{out} - 2.04 = 0$$

Solution

- Again, only one root of this quadratic equation corresponds to the physically correct solution for V_{out} , at this operating point, i.e., when $V_{in} = V_{IH}$.

$$V_{out} = 0.27 \text{ V}$$

- From this value, we can calculate the critical voltage V_{IH} as:

$$V_{IH} = 1.43 \cdot 0.37 + 1.17 = \underline{\underline{1.55 \text{ V}}}$$

Solution

- Finally, we find the noise margins for low voltage levels and for high voltage levels using

$$NM_L = V_{IL} - V_{OL} = 1.08 \text{ V}$$

$$NM_H = V_{OH} - V_{IH} = 1.75 \text{ V}$$

Supply Voltage Scaling in CMOS Inverters

- Here, we will briefly examine the effects of supply voltage scaling, i.e., reduction of V_{DD} , upon the static voltage transfer characteristics of CMOS inverters.
- The overall power dissipation of any digital circuit is a strong function of the supply voltage V_{DD} .

Supply Voltage Scaling in CMOS Inverters

- Neglecting second-order effects such as subthreshold conduction, it can be seen that the CMOS inverter will continue to operate correctly with a supply voltage which is as low as the following limit value.

$$V_{DD}^{min} = V_{T0,n} + |V_{T0,p}| \quad \dots\dots(85)$$

Supply Voltage Scaling in CMOS Inverters

- Figure 19 shows the voltage transfer characteristics of a CMOS inverter, obtained with different supply voltage levels.
- The exact shape of the VTC near the limit value is essentially determined by subthreshold conduction properties of the nMOS and pMOS transistors, yet it is clear that the circuit operates as an inverter over a large range of supply voltages levels.

Supply Voltage Scaling in CMOS

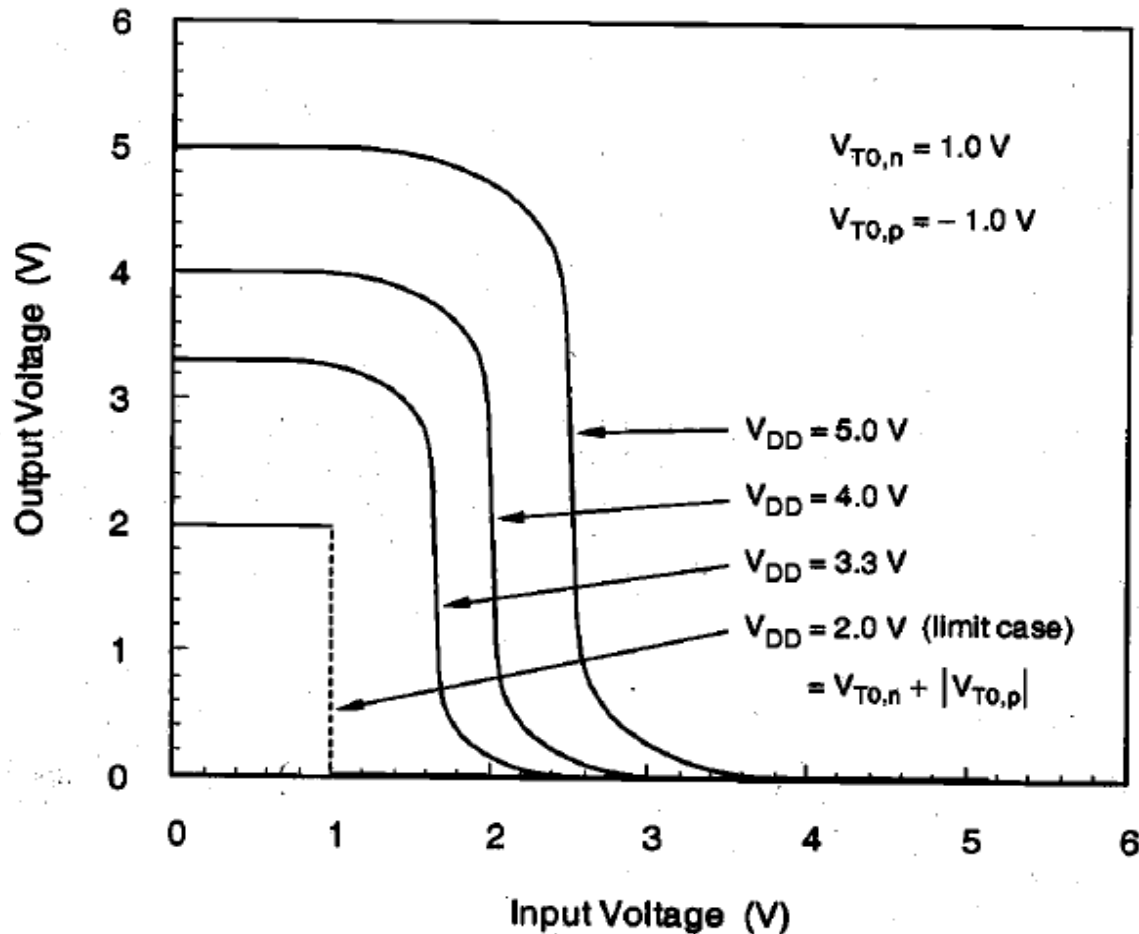


Figure 19: Voltage transfer characteristics of a CMOS inverter, obtained with different power supply voltage levels.

Power and Area Considerations

- Since the CMOS inverter does not draw any significant current from the power source in both of its steady-state operating points ($V_{\text{out}} = V_{\text{OH}}$ and $V_{\text{out}} = V_{\text{OL}}$), the DC power dissipation of this circuit is almost negligible.
- The drain current that flows through the nMOS and the pMOS transistors in both cases is essentially limited to the reverse leakage current of the source and drain pn-junctions, and in short-channel MOSFETs, the relatively small subthreshold current.

Power and Area Considerations

- This unique property of the CMOS inverter was already identified as one of the most important advantages of this configuration.
- In many applications requiring a low overall power consumption, CMOS is preferred over other circuit alternatives for this reason.
- It must be noted, however, that the CMOS inverter does conduct a significant amount of current during a switching event, i.e., when the output voltage changes from a low to high state, or from a high to low state.

Power and Area Considerations

- Consequently, CMOS logic circuits tend to occupy more area than comparable nMOS logic circuits, which apparently affects the integration density of pure-CMOS logic.
- The actual integration density of nMOS logic, on the other hand, is limited by power dissipation and heat generation problems.

Power and Area Considerations

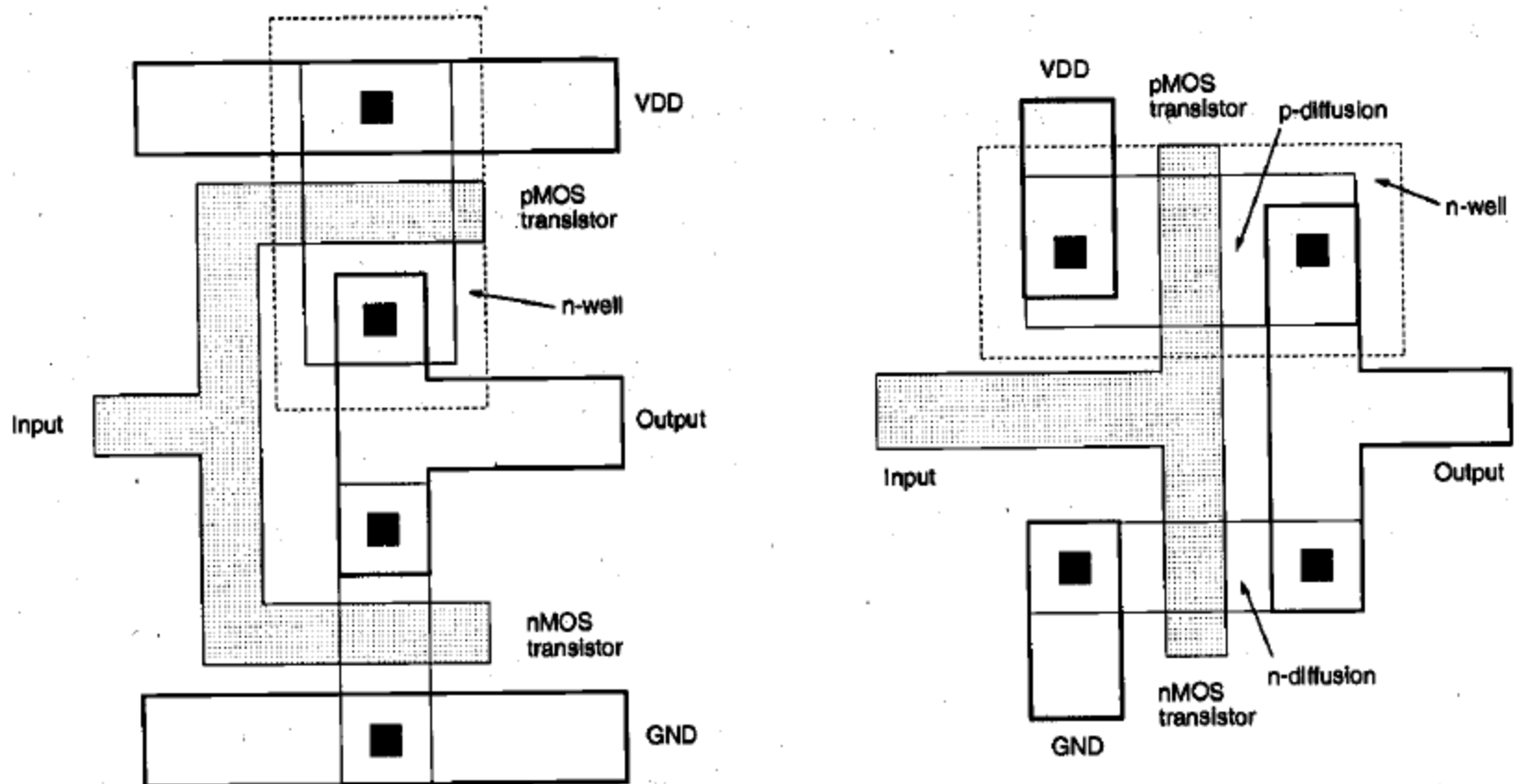
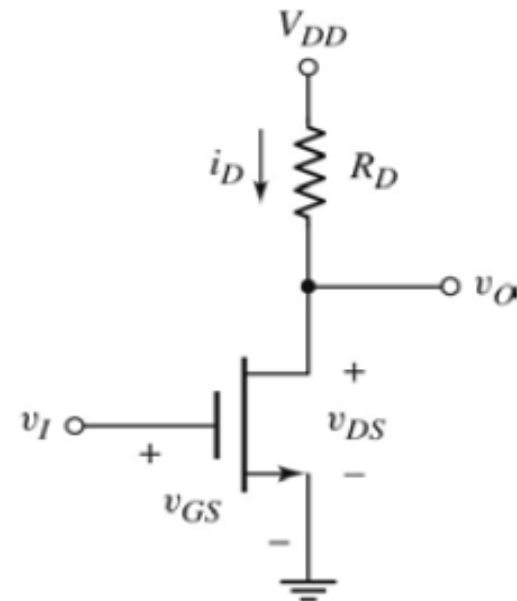


Figure 20: Two sample layouts of CMOS inverter circuits (for p-type substrate).

Problem 5

Example

For the NMOS inverter shown in Fig. $V_{DD} = 3\text{ V}$. Assume transistor parameters of $K'_n = 60\text{ }\mu\text{A/V}^2$, $W/L = 5$, and $V_{TN} = 0.5\text{ V}$. (a) Find the value of R_D such that $v_o = 0.1\text{ V}$ when $v_I = 3\text{ V}$. (b) Using the results of part (a) determine the transition point for the driver transistor



Solution

EX16.2

(a)

$$v_o = V_{DD} - I_D R_D$$

$$v_o = 3 - \left(\frac{k'_n}{2}\right)\left(\frac{W}{L}\right)[2(3 - 0.5)v_o - v_o^2]R_D$$

$$v_o = 0.1$$

$$0.1 = 3 - \left(\frac{0.06}{2}\right)(5)[(5)(0.1) - (0.1)^2]R_D$$

$$0.1 = 3 - 0.0735R_D$$

$$R_D = 39.5 \text{ K}$$

(b)

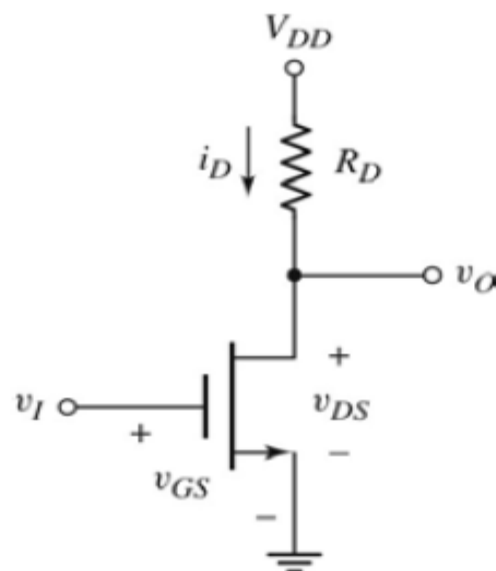
$$\left(\frac{0.06}{2}\right)(5)(39.5)(V_{th} - 0.5)^2 + (V_{th} - 0.5) - 3 = 0$$

$$5.925(V_{th} - 0.5)^2 + (V_{th} - 0.5) - 3 = 0$$

$$(V_{th} - 0.5) = V_{Ot} = \frac{-1 \pm \sqrt{1 + 4(5.925)(3)}}{2(5.925)}$$

$$V_{Ot} = 0.632 \text{ V}$$

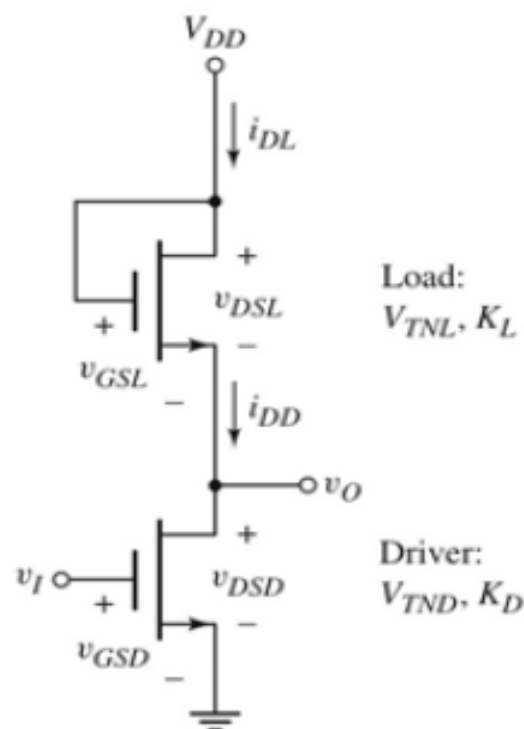
$$V_{th} = 1.132 \text{ V}$$



Problem 6

Example

The enhancement-load NMOS inverter shown in Fig. is biased at $V_{DD} = 3$ V. The transistor parameters are $V_{TND} = V_{TNL} = 0.4$ V, $k'_n = 60$ mA/V², $(W/L)_D = 16$ and $(W/L)_L = 2$. (a) Find v_o when (i) $v_I = 0$, (ii) $v_I = 2.6$, (b) Calculate the power dissipated in the inverter when $v_I = 2.6$ V.



Solution

EX16.3

(a)

(i)

$$v_o = V_{DD} - V_{TNL} = 3 - 0.4$$

$$v_o = 2.6 \text{ V}$$

(ii)

$$\left(\frac{W}{L}\right)_D [2(v_I - 0.4)v_o - v_o^2] = \left(\frac{W}{L}\right)_L [V_{DD} - v_o - 0.4]^2$$

$$16[2(2.6 - 0.4)v_o - v_o^2] = 2[3 - v_o - 0.4]^2$$

$$35.2v_o - 8v_o^2 = 6.76 - 5.2v_o + v_o^2$$

$$9v_o^2 - 40.4v_o + 6.76 = 0$$

$$v_o = \frac{40.4 \pm \sqrt{1632.16 - 4(9)(6.76)}}{2(9)}$$

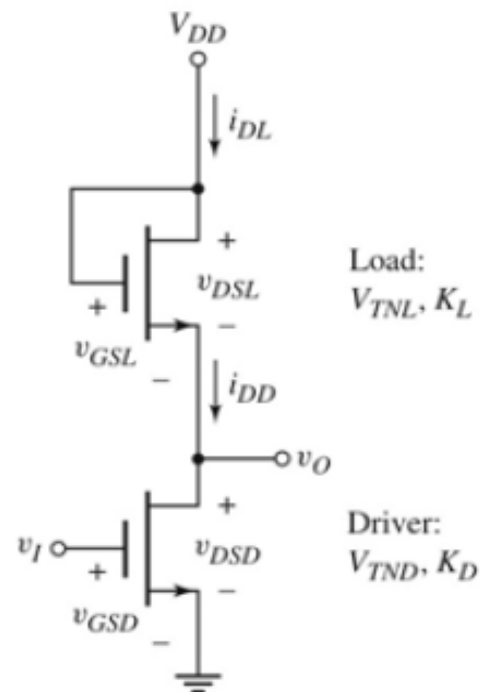
$$v_o = 0.174 \text{ V}$$

(b)

$$i_D = \left(\frac{60}{2}\right)(2)[3 - 0.174 - 0.4]^2$$

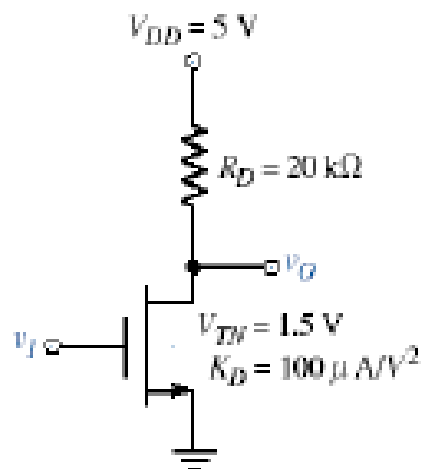
$$i_D = 353.1 \mu\text{A}$$

$$P = i_D \cdot V_{DD} = 1.06 \text{ mW}$$

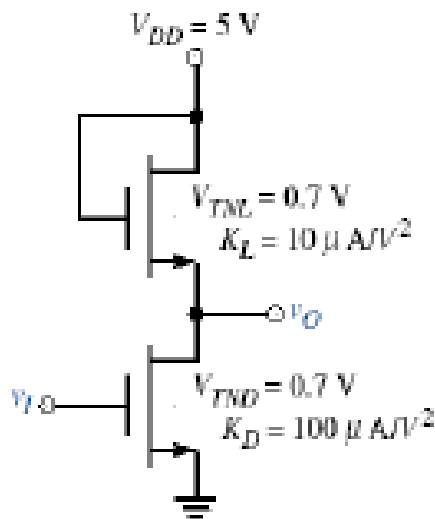


Problem 7

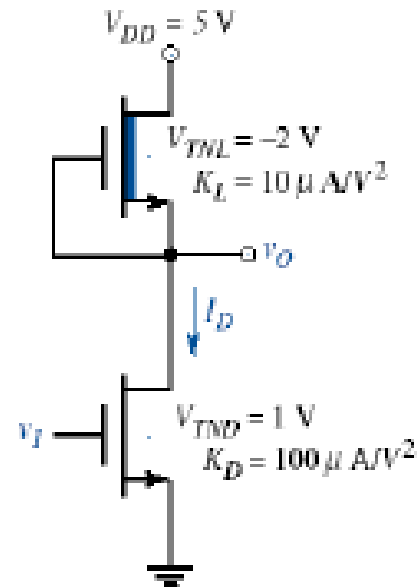
16.14 Calculate the power dissipated in each inverter circuit in Figure P16.14 for the following input conditions: (a) Inverter a: (i) $v_I = 0.5 \text{ V}$, (ii) $v_I = 5 \text{ V}$; (b) Inverter b: (i) $v_I = 0.25 \text{ V}$, (ii) $v_I = 4.3 \text{ V}$; (c) Inverter c: (i) $v_I = 0.03 \text{ V}$, (ii) $v_I = 5 \text{ V}$.



(a)



(b)



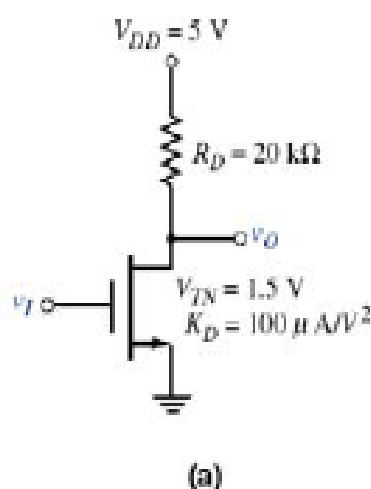
(c)

Solution

Example 16.14

P1098

16.14 Calculate the power dissipated in each inverter circuit in Figure P16.14 for the following input conditions: (a) Inverter a: (i) $v_I = 0.5$ V, (ii) $v_I = 5$ V; (b) Inverter b: (i) $v_I = 0.25$ V, (ii) $v_I = 4.3$ V; (c) Inverter c: (i) $v_I = 0.03$ V, (ii) $v_I = 5$ V.



(i) $v_I = 0.5$ V $\Rightarrow i_D = 0 \Rightarrow \underline{P = 0}$

(ii) $v_I = 5$ V, From Equation (16.12),

nonsaturation region

$$v_O = V_{DD} - K_n R_D [2(v_I - V_{TN})v_O - v_O^2]$$

$$v_O = 5 - (0.1)(20)[2(5 - 1.5)v_O - v_O^2]$$

$$2v_O^2 - 15v_O + 5 = 0$$

$$v_O = \frac{15 \pm \sqrt{(15)^2 - 4(2)(5)}}{2(2)} \Rightarrow \underline{v_O = 0.35 \text{ V}}$$

$$i_D = \frac{5 - 0.35}{20} = 0.2325 \text{ mA}$$

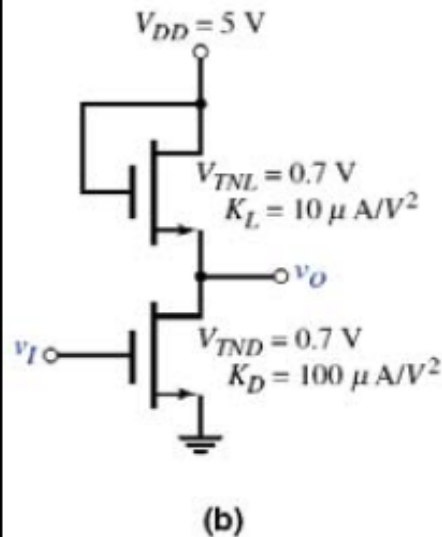
$$P = i_D \cdot V_{DD} = (0.2325)(5) \Rightarrow \underline{P = 1.16 \text{ mW}}$$

Solution

Example 16.14

P1098

16.14 Calculate the power dissipated in each inverter circuit in Figure P16.14 for the following input conditions: (a) Inverter a: (i) $v_I = 0.5 \text{ V}$, (ii) $v_I = 5 \text{ V}$; (b) Inverter b: (i) $v_I = 0.25 \text{ V}$, (ii) $v_I = 4.3 \text{ V}$; (c) Inverter c: (i) $v_I = 0.03 \text{ V}$, (ii) $v_I = 5 \text{ V}$.



(i) $v_I = 0.25 \text{ V} \Rightarrow i_D = 0 \Rightarrow \underline{P = 0}$

(ii) $v_I = 4.3 \text{ V}$, From Equation (16.23),

input voltage greater than V_{th}

$$K_D[2(v_I - V_{TND})v_O - v_O^2] = K_L(V_{DD} - v_O - V_{TNL})^2$$

$$100[2(4.3 - 0.7)v_O - v_O^2] = 10[5 - v_O - 0.7]^2$$

$$10[7.2v_O - v_O^2] = 18.49 - 8.6v_O + v_O^2$$

$$11v_O^2 - 80.6v_O + 18.49 = 0$$

$$v_O = \frac{80.6 \pm \sqrt{(80.6)^2 - 4(11)(18.49)}}{2(11)} = \underline{0.237 \text{ V}}$$

$$i_D = 10[5 - 0.237 - 0.7]^2 = 165 \mu\text{A}$$

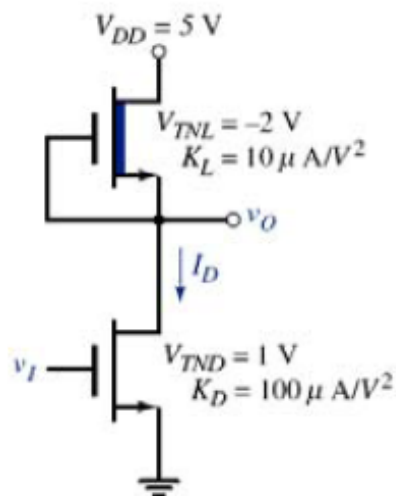
$$P = i_D \cdot V_{DD} = (165)(5) \Rightarrow \underline{P = 825 \mu\text{W}}$$

Solution

Example 16.14

P1098

16.14 Calculate the power dissipated in each inverter circuit in Figure P16.14 for the following input conditions: (a) Inverter a: (i) $v_I = 0.5 \text{ V}$, (ii) $v_I = 5 \text{ V}$; (b) Inverter b: (i) $v_I = 0.25 \text{ V}$, (ii) $v_I = 4.3 \text{ V}$; (c) Inverter c: (i) $v_I = 0.03 \text{ V}$, (ii) $v_I = 5 \text{ V}$.



(c)

(i) $v_I = 0.03 \text{ V} \Rightarrow i_D = 0 \Rightarrow \underline{P = 0}$

(ii) $v_I = 5 \text{ V}$

$$i_D = K_L (-V_{TNL})^2 = (10)[-(-2)]^2 = 40 \mu\text{A}$$

$$P = i_D \cdot V_{DD} = (40)(5) \Rightarrow \underline{P = 200 \mu\text{W}}$$

Summary of Transistor Operation

NMOS

Nonsaturation region ($v_{DS} < v_{DS}(\text{sat})$)

$$i_D = K_n[2(v_{GS} - V_{TN})v_{DS} - v_{DS}^2]$$

Saturation region ($v_{DS} > v_{DS}(\text{sat})$)

$$i_D = K_n(v_{GS} - V_{TN})^2$$

Transition point

$$v_{DS}(\text{sat}) = v_{GS} - V_{TN}$$

Enhancement mode

$$V_{TN} > 0$$

Depletion mode

$$V_{TN} < 0$$

PMOS

Nonsaturation region ($v_{SD} < v_{SD}(\text{sat})$)

$$i_D = K_p[2(v_{SG} + V_{TP})v_{SD} - v_{SD}^2]$$

Saturation region ($v_{SD} > v_{SD}(\text{sat})$)

$$i_D = K_p(v_{SG} + V_{TP})^2$$

Transition point

$$v_{SD}(\text{sat}) = v_{SG} + V_{TP}$$

Enhancement mode

$$V_{TP} < 0$$

Depletion mode

$$V_{TP} > 0$$

MOS INVERTERS

**SWITCHING
CHARACTERISTICS
AND
INTERCONNECT
EFFECTS**

Introduction

- Here, we will investigate the dynamic (time-domain) behavior of the inverter circuits.
- The cascade connection of two CMOS inverter circuits are considered as shown in Fig. 1.
- The parasitic capacitances associated with each MOSFET are illustrated individually.

Introduction

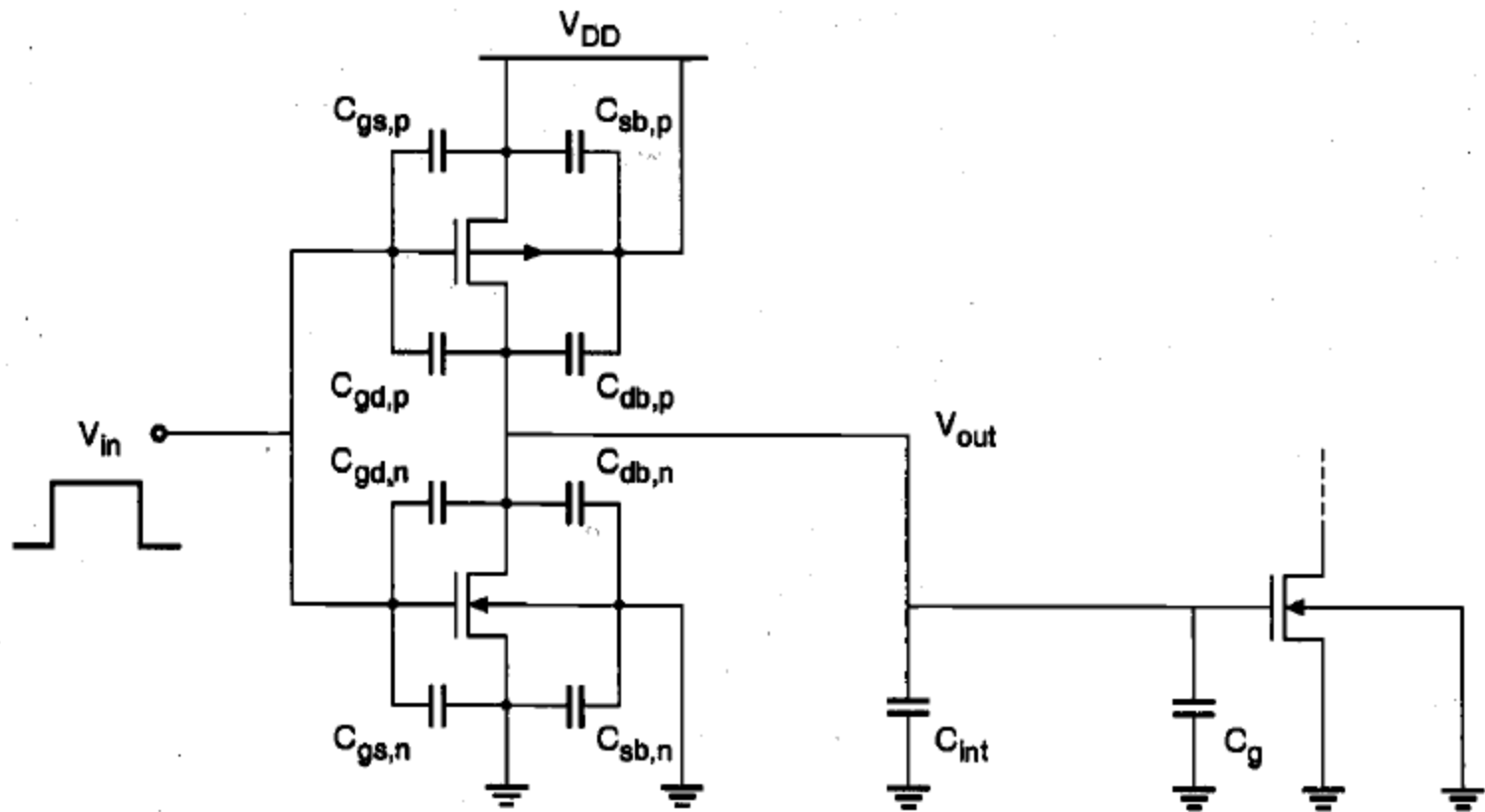


Figure 1. Cascaded CMOS inverter stages

Introduction

- In addition to other capacitances already discussed here, we also consider the lumped interconnect capacitance C_{int} , which represents the parasitic capacitance contribution of the metal or polysilicon connection between the two inverters.
- It is assumed that a pulse waveform is applied to the input of the first-stage inverter.
- We wish to analyze the time-domain behavior of the first-stage output, V_{out} .

Introduction

- The problem of analyzing the output voltage waveform is fairly complicated, even for this relatively simple circuit, because a number of nonlinear, voltage-dependent capacitances are involved.
- To simplify the problem, we first combine the capacitances seen in Fig. 1 into an equivalent lumped linear capacitance, connected between the output node of the inverter and the ground.

Introduction

- This combined capacitance at the output node will be called the load capacitance, C_{load}

$$C_{load} = C_{gd,n} + C_{gd,p} + C_{db,n} + C_{db,p} + C_{int} + C_g \quad \text{.....(1)}$$

- Note that some of the parasitic capacitance components shown in Fig. 1 do not appear in this lumped capacitance expression.

Introduction

- In particular, $C_{sb,n}$ and $C_{sb,p}$ have no effect on the transient behavior of the circuit since the source-to-substrate voltages of both transistors are always equal to zero.
- The capacitances $C_{gs,n}$ and $C_{gs,p}$ are also not included in (Eqn. 1) because they are connected between the input node and the ground (or the power supply).
- The capacitance terms $C_{db,n}$ and $C_{db,p}$ in (Eqn. 1) are the equivalent junction capacitances.

Introduction

- The first-stage CMOS inverter is shown with the single lumped output load capacitance C_{load} in Fig. 2.
- The delay times calculated using C_{load} may slightly overestimate the actual inverter delay, but this is not considered a significant deficiency in a first-order approximation.

Introduction

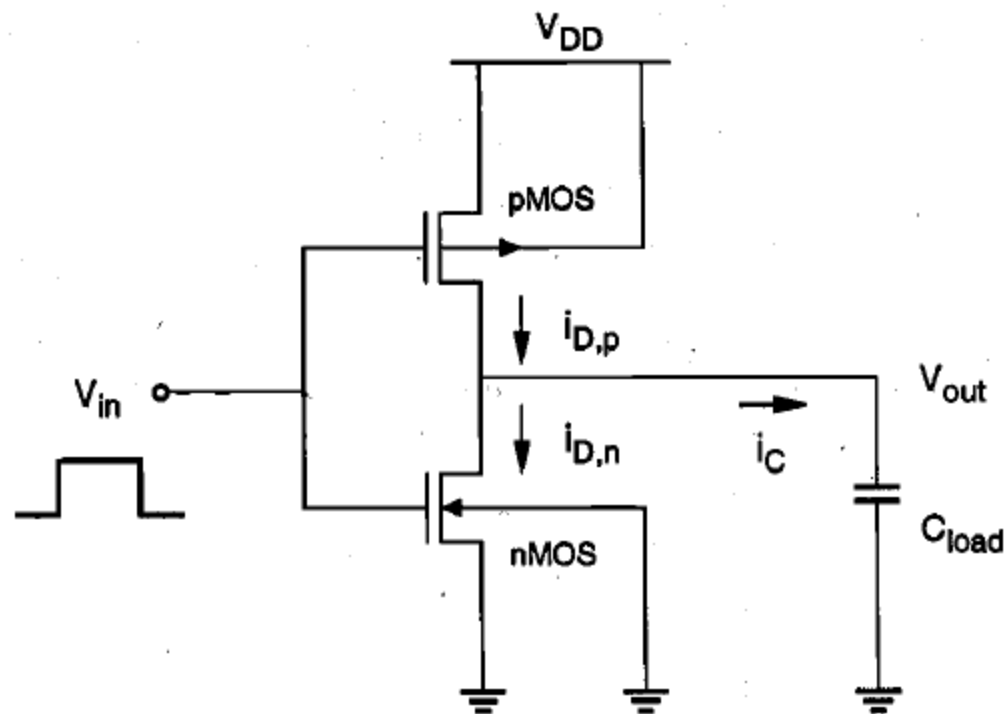


Figure 2. First-stage CMOS inverter with lumped output load capacitance

Delay-Time Definitions

- The input and output voltage waveforms of a typical inverter circuit are shown in Fig. 3.
- The propagation delay times τ_{PHL} and τ_{PLH} determine the input-to-output signal delay during the high-to-low and low-to-high transitions of the output, respectively.
- By definition, τ_{PHL} is the time delay between the $V_{50\%}$ -transition of the rising input voltage and the $V_{50\%}$ -transition of the falling output voltage.

Delay-Time Definitions

- Similarly, τ_{PLH} is defined as the time delay between the $V_{50\%}$ -transition of the falling input voltage and the $V_{50\%}$ -transition of the rising output voltage.
- The input voltage waveform is usually assumed to be an ideal step pulse with zero rise and fall times for simplification.
- Under this assumption, τ_{PHL} becomes the time required for the output voltage to fall from

Delay-Time Definitions

- V_{OH} to the $V_{50\%}$ level, and τ_{PLH} becomes the time required for the output voltage to rise from V_{OL} to the $V_{50\%}$ level.
- The voltage point $V_{50\%}$ is defined as follows.

$$V_{50\%} = V_{OL} + \frac{1}{2}(V_{OH} - V_{OL}) = \frac{1}{2}(V_{OL} + V_{OH}) \quad \dots\dots(2)$$

Delay-Time Definitions

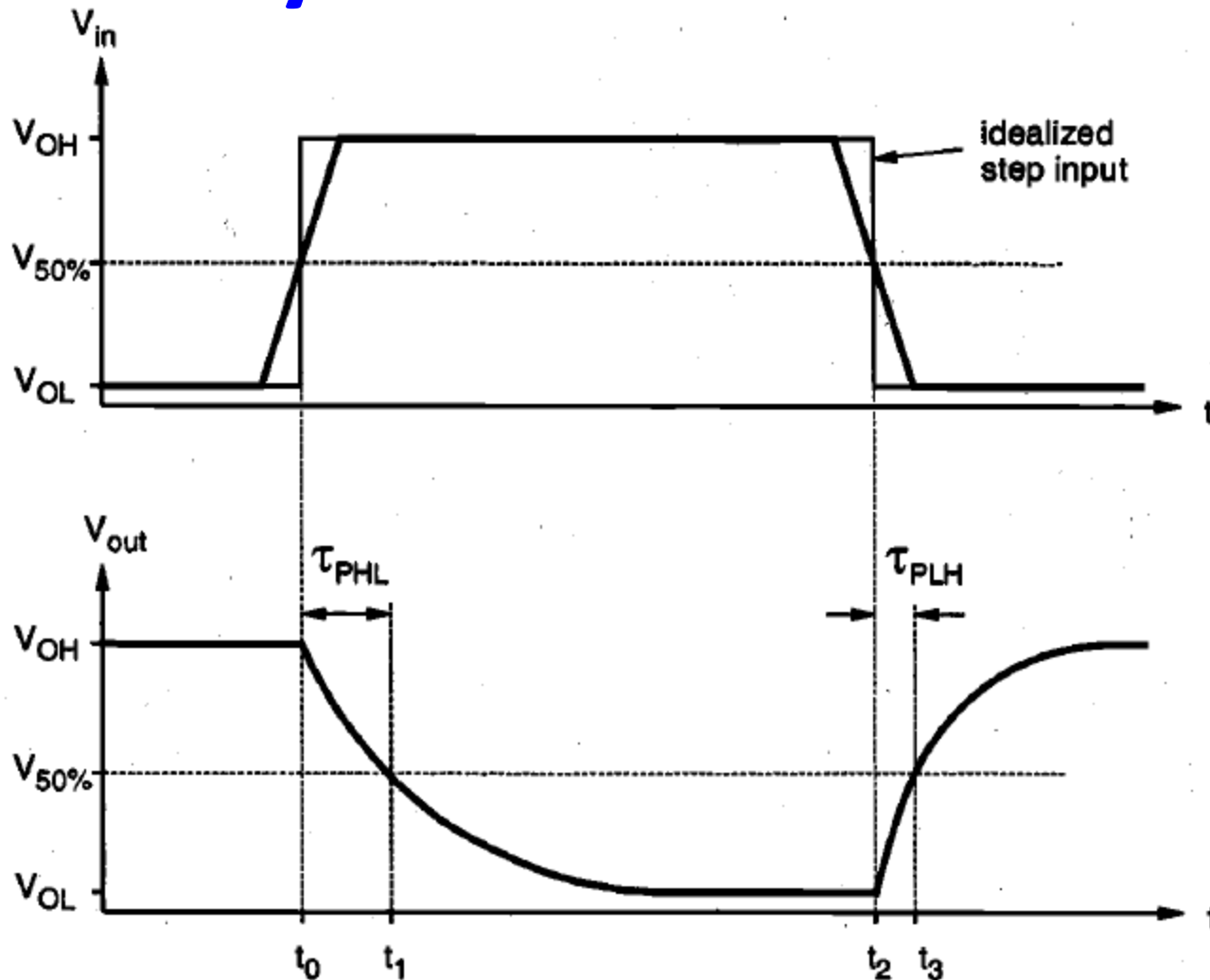


Figure 3. Input and output voltage waveforms of a typical inverter, and the definitions of propagation delay times. The input voltage waveform is idealized as a step pulse for simplicity.

Delay-Time Definitions

- The propagation delay times τ_{PHL} and τ_{PLH} are found from Fig. 3 as

$$\begin{aligned}\tau_{PHL} &= t_1 - t_0 \\ \tau_{PLH} &= t_3 - t_2\end{aligned}\quad \text{.....(3)}$$

- The average propagation delay τ_p of the inverter characterizes the average time required for the input signal to propagate through the inverter.

$$\tau_p = \frac{\tau_{PHL} + \tau_{PLH}}{2} \quad \text{.....(4)}$$

Delay-Time Definitions

- Fig. 4 is referred for the definition of output voltage rise and fall times.
- The rise time τ_{rise} is defined here as the time required for the output voltage to rise from the $V_{10\%}$ level to $V_{90\%}$ level.
- Similarly, the fall time τ_{fall} is defined here as the time required for the output voltage to drop from the $V_{90\%}$ level to $V_{10\%}$ level.

Delay-Time Definitions

- The voltage levels $V_{10\%}$ and $V_{90\%}$ are defined as;

$$V_{10\%} = V_{OL} + 0.1 \cdot (V_{OH} - V_{OL}) \quad \dots\dots(5)$$

$$V_{90\%} = V_{OL} + 0.9 \cdot (V_{OH} - V_{OL}) \quad \dots\dots(6)$$

- Thus, t times are found from Fig. 4 as follows.

$$\tau_{fall} = t_B - t_A \quad \dots\dots(7)$$

$$\tau_{rise} = t_D - t_C$$

Delay-Time Definitions

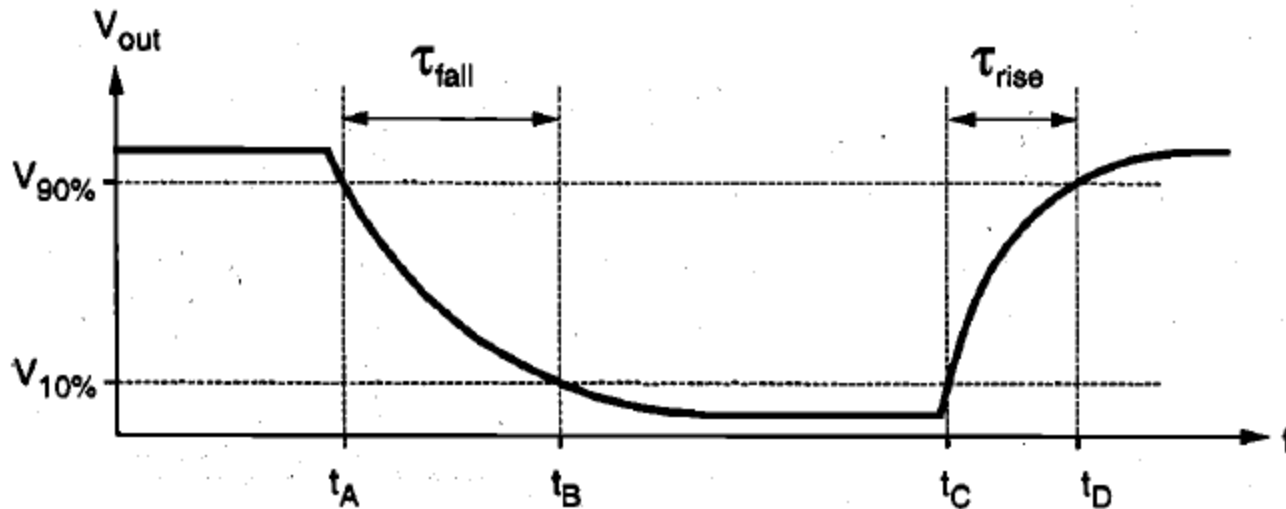


Figure 4. Output voltage rise and fall times.

Calculation of Delay Times

- If the capacitance current during an output transition is approximated by a constant average current I_{avg} the delay times are found as;

$$\tau_{PHL} = \frac{C_{load} \cdot \Delta V_{HL}}{I_{avg, HL}} = \frac{C_{load} \cdot (V_{OH} - V_{50\%})}{I_{avg, HL}} \quad \dots\dots(8)$$

$$\tau_{PLH} = \frac{C_{load} \cdot \Delta V_{LH}}{I_{avg, LH}} = \frac{C_{load} \cdot (V_{50\%} - V_{OL})}{I_{avg, LH}} \quad \dots\dots(9)$$

Calculation of Delay Times

- The average current during high-to-low transition can be calculated by using the current values at the beginning and the end of the transition.

$$I_{avg,HL} = \frac{1}{2} \left[i_C(V_{in} = V_{OH}, V_{out} = V_{OH}) + i_C(V_{in} = V_{OH}, V_{out} = V_{50\%}) \right] \dots\dots(10)$$

- Similarly, the average capacitance current during low-to-high transition is

$$I_{avg,LH} = \frac{1}{2} \left[i_C(V_{in} = V_{OL}, V_{out} = V_{50\%}) + i_C(V_{in} = V_{OL}, V_{out} = V_{OL}) \right] \dots\dots(11)$$

Calculation of Delay Times

- While the average-current method is relatively simple and requires minimal calculation, it neglects the variations of the capacitance current between the beginning and end points of the transition.
- Therefore, we do not expect the average-current method to provide a very accurate estimate of the delay times. Still, this approach can provide rough, first-order estimates of the charge-up and charge-down delay times.

Calculation of Delay Times

- The propagation delay times can be found more accurately by solving the state equation of the output node in the time domain.
- The differential equation associated with the output node is given below.
- The capacitance current is also a function of the output voltage

$$C_{load} \frac{dV_{out}}{dt} = i_C = i_{D,p} - i_{D,n} \quad \dots\dots(12)$$

Calculation of Delay Times

- First, we consider the rising-input case for a CMOS inverter.
- Initially, the output voltage is assumed to be equal to V_{OH} .
- When the input voltage switches from low (V_{OL}) to high (V_{OH}), the nMOS transistor is turned on and it starts to discharge the load capacitance.
- At the same time, the pMOS transistor is switched off; thus,

Calculation of Delay Times

$$i_{D,p} \approx 0 \quad \text{.....(13)}$$

- Now

$$C_{load} \frac{dV_{out}}{dt} = -i_{D,n} \quad \text{.....(14)}$$

- Therefore, (Eqn. 14) can be used to calculate the chargedown time not only in CMOS inverters, but also in almost all common types of inverter circuits.

Calculation of Delay Times

- The, input and output voltage waveforms during this high-to-low transition are illustrated in Fig. 6.
- When the nMOS transistor starts conducting, it initially operates in the saturation region. When the output voltage falls below $(V_{DD} - V_{T,n})$, the nMOS transistor starts to conduct in the linear region.
- These two operating regions are also shown in Fig. 6.

Calculation of Delay Times

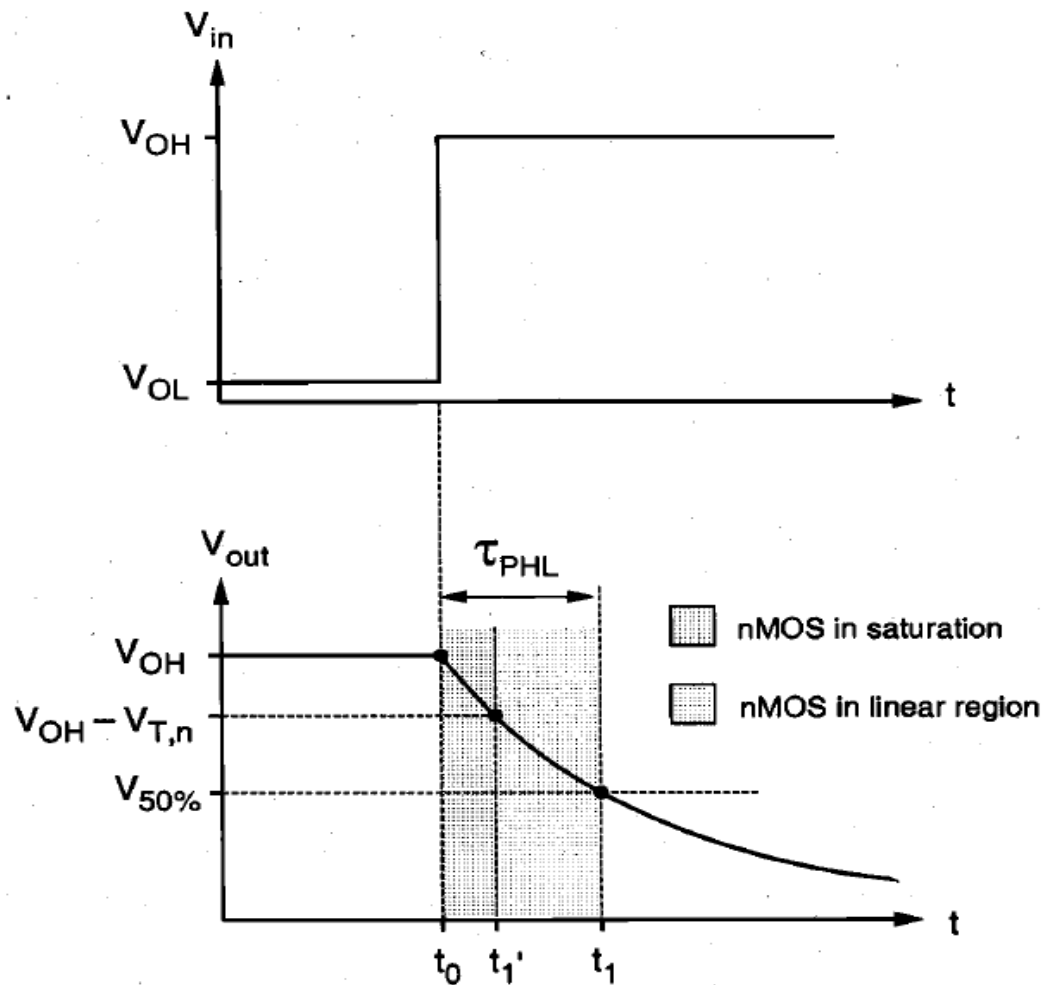


Figure 6. Input and output voltage waveforms during high-to-low transition.

Calculation of Delay Times

- First, consider the nMOS transistor operating in saturation

$$\begin{aligned} i_{D,n} &= \frac{k_n}{2} (V_{in} - V_{T,n})^2 \\ &= \frac{k_n}{2} (V_{OH} - V_{T,n})^2, \quad \text{for } V_{OH} - V_{T,n} < V_{out} \leq V_{OH} \quad \dots\dots(15) \end{aligned}$$

- Since the saturation current is practically independent of the output voltage (neglecting channel-length modulation), the solution of (Eqn. 14) in the time interval between t_0 and t_1' can be found as

Calculation of Delay Times

$$\int_{t=t_0}^{t=t_1'} dt = -C_{load} \int_{V_{out}=V_{OH}}^{V_{out}=V_{OH}-V_{T,n}} \left(\frac{1}{i_{D,n}} \right) dV_{out}$$

$$= -\frac{2C_{load}}{k_n (V_{OH} - V_{T,n})^2} \int_{V_{out}=V_{OH}}^{V_{out}=V_{OH}-V_{T,n}} dV_{out} \quad \dots\dots(16)$$

• OR

$$t_1' - t_0 = \frac{2C_{load} V_{T,n}}{k_n (V_{OH} - V_{T,n})^2} \quad \dots\dots(17)$$

Calculation of Delay Times

- At $t = t_1'$, the output voltage will be equal to $(V_{DD} - V_{T,n})$ and the transistor will be at the saturation-linear region boundary.
- Next, consider the nMOS transistor operating in the linear region.

$$\begin{aligned} i_{D,n} &= \frac{k_n}{2} \left[2(V_{in} - V_{T,n})V_{out} - V_{out}^2 \right] \\ &= \frac{k_n}{2} \left[2(V_{OH} - V_{T,n})V_{out} - V_{out}^2 \right], \quad \text{for } V_{out} \leq V_{OH} - V_{T,n} \quad \dots\dots(18) \end{aligned}$$

Calculation of Delay Times

- The solution of (Eqn. 14) in the time interval between t_1' and t_1 can be found as

$$\int_{t=t_1'}^{t=t_1} dt = -C_{load} \int_{V_{out}=V_{OH}-V_{T,n}}^{V_{out}=V_{50\%}} \left(\frac{1}{i_{D,n}} \right) dV_{out}$$

$$= -2 C_{load} \int_{V_{out}=V_{OH}-V_{T,n}}^{V_{out}=V_{50\%}} \left(\frac{1}{k_n \left[2(V_{OH}-V_{T,n})V_{out} - V_{out}^2 \right]} \right) dV_{out} \quad \dots\dots(19)$$

Calculation of Delay Times

- Evaluating this integral yields

$$t_1 - t_1' = -\frac{2C_{load}}{k_n} \frac{1}{2(V_{OH} - V_{T,n})} \ln \left(\frac{V_{out}}{2(V_{OH} - V_{T,n}) - V_{out}} \right) \bigg|_{V_{out}=V_{OH}-V_{T,n}}^{V_{out}=V_{50\%}} \dots\dots(20)$$

$$t_1 - t_1' = \frac{C_{load}}{k_n (V_{OH} - V_{T,n})} \ln \left(\frac{2(V_{OH} - V_{T,n}) - V_{50\%}}{V_{50\%}} \right) \dots\dots(21)$$

Calculation of Delay Times

- Finally, the propagation delay time for high-to-low output transition (τ_{PHL}) can be found by combining (Eqn. 17) and (Eqn. 21):

$$\tau_{PHL} = \frac{C_{load}}{k_n (V_{OH} - V_{T,n})} \left[\frac{2V_{T,n}}{V_{OH} - V_{T,n}} + \ln \left(\frac{4(V_{OH} - V_{T,n})}{V_{OH} + V_{OL}} - 1 \right) \right] \quad \text{.....(22a)}$$

Calculation of Delay Times

- For $V_{OH} = V_{DD}$ and $V_{OL} = 0$, as is the case for the CMOS inverter, (Eqn. 22a) becomes:

$$\tau_{PHL} = \frac{C_{load}}{k_n (V_{DD} - V_{T,n})} \left[\frac{2V_{T,n}}{V_{DD} - V_{T,n}} + \ln \left(\frac{4(V_{DD} - V_{T,n})}{V_{DD}} - 1 \right) \right] \dots\dots(22b)$$

Problem 1

- Consider the CMOS inverter circuit shown in Fig. 2, with $V_{DD} = 3.3$ V.
- The I-V characteristics of the nMOS transistor are specified as follows: when $V_{GS} = 3.3$ V, the drain current reaches its saturation level $I_{sat} = 2$ mA for $V_{DS} \geq 2.5$ V.
- Assume that the input signal applied to the gate is a step pulse that switches instantaneously from 0 V to 3.3 V.

Problem 1

- Using the data above, calculate the delay time necessary for the output to fall from its initial value of 3.3 V to 1.65 V, assuming an output load capacitance of 300 fF.

Solution

- For the solution, consider the simplified pull-down circuit shown in Fig. 5.
- We will assume that the nMOS transistor operates in saturation from $t = 0$ to $t = t_1' = t_{\text{sat}}$ and that it will operate in the linear region from $t = t_1' = t_{\text{sat}}$ to $t = t_2 = t_{\text{delay}}$.
- We can also deduce from the I-V characteristics that $V_{T,n} = 0.8 \text{ V}$, since the nMOS transistor enters saturation when $V_{DS} > V_{GS} - V_{T,n}$.

Solution

- The voltage V_{GS} is equal to 3.3 V for $t \geq 0$.
- The current equation for the saturation region can be written as

$$C \frac{dV_{out}}{dt} = -I_D = -I_{sat} = -\frac{1}{2} k_n (V_{OH} - V_{T,n})^2$$

- We can calculate the amount of time in which the nMOS transistor operates in saturation (t_{sat}), by integrating this equation.

Solution

$$\int_{t=0}^{t=t_{sat}} dt = - \int_{V_{out}=3.3}^{V_{out}=2.5} \frac{C}{I_D} dV_{out}$$

$$t_{sat} = \frac{V_{T,n} C}{I_{sat}} = \frac{0.8 \text{ V} \cdot 300 \text{ fF}}{2 \text{ mA}} = 120 [\text{ps}]$$

Solution

- The transconductance k_n of the nMOS transistor can be found as follows:

$$k_n = \frac{2I_{sat}}{(V_{OH} - V_{T,n})^2} = \frac{2 \times 2 \times 10^{-3}}{(3.3 - 0.8)^2} = 0.640 \times 10^{-3} \text{ [A / V}^2\text{]}$$

- Now, the current equation for the linear operating region is

$$C \frac{dV_{out}}{dt} = -I_D = -\frac{1}{2} k_n [2(V_{OH} - V_{T,n})V_{out} - V_{out}^2]$$

Solution

- Integrating this differential equation between the two voltage boundary conditions yields the time in which the nMOS transistor operates in the linear region during this transition.

Solution

$$\int_{t=t_{sat}}^{t=t_{delay}} dt = -2C \int_{V_{out}=2.5}^{V_{out}=1.65} \frac{dV_{out}}{k_n [2(V_{OH} - V_{T,n})V_{out} - V_{out}^2]}$$

$$t_{delay} - t_{sat} = -\frac{C}{k_n (V_{OH} - V_{T,n})} \ln \left(\frac{V_{out}}{2(V_{OH} - V_{T,n}) - V_{out}} \right) \bigg|_{V_{out}=2.5}^{V_{out}=1.65}$$

$$= \frac{C}{k_n (V_{OH} - V_{T,n})} \left[\ln \left(\frac{2(V_{OH} - V_{T,n}) - V_{1.65}}{V_{1.65}} \right) - \ln \left(\frac{2(V_{OH} - V_{T,n}) - V_{2.5}}{V_{2.5}} \right) \right]$$

$$= \frac{0.3 \times 10^{-12}}{0.640 \times 10^{-3}} \frac{1}{(3.3 - 0.8)} \left[\ln \left(\frac{5 - 1.65}{1.65} \right) - \ln \left(\frac{5 - 2.5}{2.5} \right) \right] = 133 [\text{ps}]$$

Solution

- Thus, the total delay time is found to be

$$t_{\text{delay}} = 120 + 133 = 253 [\text{ps}]$$

- Note that t_{delay} corresponds to the propagation delay time τ_{PHL} for falling output.

Problem 2

- For the CMOS inverter shown in Fig. 2 with a power supply voltage of $V_{DD} = 5\text{ V}$, determine the fall time τ_{fall} which is defined as the time elapsed between the time point at which $V_{\text{out}} = V_{90\%} = 4.5\text{ V}$ and the time point at which $V_{\text{out}} = V_{10\%} = 0.5\text{ V}$.
- Use both the average-current method and the differential equation method for calculating τ_{fall} .

Problem 2

- The output load capacitance is 1 pF.
- The nMOS transistor parameters are given as

$$\begin{aligned}\mu_n C_{ox} &= 20 \mu\text{A}/\text{V}^2 \\ (W/L)_n &= 10 \\ V_{T,n} &= 1.0 \text{ V}\end{aligned}$$

Solution

- Using a simple expression similar to (Eqn. 10), we can determine the average capacitor current during the charge-down event described above.

$$\begin{aligned} I_{avg} &= \frac{1}{2} \left[I(V_{in} = 5 \text{ V}, V_{out} = 4.5 \text{ V}) + I(V_{in} = 5 \text{ V}, V_{out} = 0.5 \text{ V}) \right] \\ &= \frac{1}{2} \left[\frac{1}{2} k_n (V_{in} - V_{T,n})^2 + \frac{1}{2} k_n (2(V_{in} - V_{T,n})V_{out} - V_{out}^2) \right] \\ &= \frac{1}{2} \cdot \frac{1}{2} \cdot 20 \times 10^{-6} \cdot 10 \left[(5 - 1)^2 + (2(5 - 1)0.5 - 0.5^2) \right] = 0.9875 \text{ [mA]} \end{aligned}$$

Solution

- The fall time is then found as

$$\tau_{fall} = \frac{C \cdot \Delta V}{I_{avg}} = \frac{1 \times 10^{-12} (4.5 - 0.5)}{0.9875 \times 10^{-3}} = 4.05 \times 10^{-9} \text{ [s]} = 4.05 \text{ [ns]}$$

- Now, we will recalculate the fall time using the differential equation approach. The nMOS transistor operates in the saturation region for $4.0 \text{ V} < V_{out} < 4.5 \text{ V}$.

Solution

- Writing the current equation for the saturation region, we obtain

$$C \frac{dV_{out}}{dt} = -\frac{1}{2} k_n (V_{in} - V_{T,n})^2, \text{ where } k_n = \mu_n C_{ox} \left(\frac{W}{L} \right)_n$$

$$\frac{dV_{out}}{dt} = \frac{-20 \times 10^{-6} \cdot 10 \cdot (5-1)^2}{2 \cdot 1 \times 10^{-12}} = -1.6 \times 10^9 \text{ [V/s]}$$

Solution

- Integrating this simple expression yields the time during which the nMOS transistor operates in saturation

$$\int_{t=0}^{t=t_{sat}} dt = -\frac{1}{1.6 \times 10^9} \int_{V_{out}=4.5}^{V_{out}=4} dV_{out}$$

$$t_{sat} = \frac{0.5}{1.6 \times 10^9} = 0.3125 \times 10^{-9} [s] = 0.3125 [ns]$$

Solution

- The nMOS transistor operates in the linear region for $0.5 \text{ V} < V < 4.0 \text{ V}$.
- The current equation for this operating region is written as follows:

$$C \frac{dV_{out}}{dt} = -\frac{1}{2} k_n \left[2(V_{in} - V_{T,n})V_{out} - V_{out}^2 \right]$$

Solution

- Integrating this equation, we obtain the delay component during which the nMOS transistor operates in the linear region.

$$\int_{t=t_{sat}}^{t=t_{delay}} dt = -2C \int_{V_{out}=4}^{V_{out}=0.5} \frac{dV_{out}}{k_n [2(V_{in} - V_{T,n})V_{out} - V_{out}^2]}$$

$$\begin{aligned} \tau_{fall} - t_{sat} &= \frac{C}{k_n (V_{in} - V_{T,n})} \left[\ln \left(\frac{2(V_{in} - V_{T,n}) - V_{0.5}}{V_{0.5}} \right) - \ln \left(\frac{2(V_{in} - V_{T,n}) - V_{4.0}}{V_{4.0}} \right) \right] \\ &= \frac{1 \times 10^{-12}}{20 \times 10^{-6} \cdot 10 \cdot 4} \left[\ln \left(\frac{8 - 0.5}{0.5} \right) - \ln \left(\frac{8 - 4}{4} \right) \right] = 3.385 \times 10^{-9} \text{ [s]} = 3.385 \text{ [ns]} \end{aligned}$$

Solution

- Thus, the fall time of the CMOS inverter is found as follows:

$$\tau_{fall} = 3.6975 \text{ [ns]}$$

Calculation of Delay Times

- In a CMOS inverter, the charge-up event of the output load capacitance for falling input transition is completely analogous to the charge-down event for rising input.
- When the input voltage switches from high (V_{OH}) to low (V_{OL}) the nMOS transistor is cut off, and the load capacitance is being charged up through the pMOS transistor.

Calculation of Delay Times

- Following a very similar derivation procedure, the propagation delay time τ_{PLH} can be found as

$$\tau_{PLH} = \frac{C_{load}}{k_p (V_{OH} - V_{OL} - |V_{T,p}|)} \left[\frac{2|V_{T,p}|}{V_{OH} - V_{OL} - |V_{T,p}|} + \ln \left(\frac{2(V_{OH} - V_{OL} - |V_{T,p}|)}{V_{OH} - V_{50\%}} - 1 \right) \right] \dots\dots(23a)$$

Calculation of Delay Times

- For $V_{OH} = V_{DD}$ and $V_{OL} = 0$, (Eqn. 23a) becomes

$$\tau_{PLH} = \frac{C_{load}}{k_p (V_{DD} - |V_{T,p}|)} \left[\frac{2|V_{T,p}|}{V_{DD} - |V_{T,p}|} + \ln \left(\frac{4(V_{DD} - |V_{T,p}|)}{V_{DD}} - 1 \right) \right] \dots\dots(23b)$$

Calculation of Delay Times

- Comparing the delay expression (Eqn. 23b) with (Eqn. 22b), we can see that the sufficient conditions for balanced propagation delays, i.e., for $\tau_{\text{PHL}} = \tau_{\text{PLH}}$ in a CMOS inverter are:

$$V_{T,n} = |V_{T,p}| \quad \text{and}$$

$$k_n = k_p \quad (\text{or } W_p/W_n = \mu_p/\mu_n)$$

Calculation of Delay Times

- The calculation of τ_{PLH} in different types of inverters depends on the load device and its operation, yet the analysis procedure is very similar to the CMOS case.
- We will consider the nMOS depletion-load inverter case as an example in the following.
- When the input voltage switches from high to low, the enhancement-type nMOS driver transistor is turned off.

Calculation of Delay Times

- The output load capacitance is then being charged up through the depletion-type load transistor. The differential equation describing this event is

$$C_{load} \frac{dV_{out}}{dt} = i_{D,load}(V_{out}) \quad \dots\dots(24)$$

Calculation of Delay Times

- The load device is initially in saturation, and enters the linear region when the output voltage rises above $(V_{DD} + V_{T,load})$, where
- $V_{T,load} < 0$.

$$i_{D,load} = \frac{k_{n,load}}{2} (|V_{T,load}|)^2, \text{ for } V_{out} \leq V_{DD} - |V_{T,load}| \quad \dots\dots(25)$$

$$i_{D,load} = \frac{k_{n,load}}{2} \left[2|V_{T,load}|(V_{DD} - V_{out}) - (V_{DD} - V_{out})^2 \right] \\ \text{for } V_{out} > V_{DD} - |V_{T,load}| \quad \dots\dots(26)$$

Calculation of Delay Times

- The delay time τ_{PLH} can be found as follows:

$$\tau_{PLH} = C_{load} \left[\int_{V_{out}=V_{OL}}^{V_{out}=V_{DD}-|V_{T,load}|} \left(\frac{dV_{out}}{i_{D,load}(sat)} \right) + \int_{V_{out}=V_{DD}-|V_{T,load}|}^{V_{out}=V_{50\%}} \left(\frac{dV_{out}}{i_{D,load}(linear)} \right) \right] \dots\dots(27)$$

Calculation of Delay Times

$$\tau_{PLH} = \frac{C_{load}}{k_{n,load} |V_{T,load}|} \left[\frac{2(V_{DD} - |V_{T,load}| - V_{OL})}{|V_{T,load}|} + \ln \left(\frac{2|V_{T,load}| - (V_{DD} - V_{50\%})}{V_{DD} - V_{50\%}} \right) \right] \dots\dots(28)$$

Calculation of Delay Times

- All of the delay time derivations in this section were made under the simplifying assumption that the input signal waveform is a step pulse with zero rise and fall times.
- Now, we consider the case where the input voltage waveform is not an ideal (step) pulse waveform, but has finite rise and fall times, τ_r and τ_f .

Calculation of Delay Times

- The exact calculation of the output voltage delay times is more complicated under this more realistic assumption, since both the nMOS transistor and the pMOS transistor conduct current during the charge-up and charge-down events.

Calculation of Delay Times

- To simplify the estimation of the actual propagation delays, we can utilize the propagation delay times calculated under the step-input assumption, using the following empirical expressions:

$$\tau_{PHL}(actual) = \sqrt{\tau_{PHL}^2(step\ input) + \left(\frac{\tau_r}{2}\right)^2} \quad \dots\dots(29)$$

$$\tau_{PLH}(actual) = \sqrt{\tau_{PLH}^2(step\ input) + \left(\frac{\tau_f}{2}\right)^2} \quad \dots\dots(30)$$

Problem 3

- A company in Urbana, IL called Prairie Technology has access to a CMOS fabrication process with the device parameters listed below.

$$\mu_n C_{ox} = 120 \mu\text{A}/\text{V}^2$$

$$\mu_p C_{ox} = 60 \mu\text{A}/\text{V}^2$$

$$L = 0.6 \mu\text{m} \text{ for both nMOS and pMOS devices}$$

$$V_{T0,n} = 0.8 \text{ V}$$

$$V_{T0,p} = -1.0 \text{ V}$$

$$W_{min} = 1.2 \mu\text{m}$$

Problem 3

- Design a CMOS inverter by determining the channel widths W_n and W_p of the nMOS and pMOS transistors, to meet the following performance specifications.
- $V_{th} = 1.5 \text{ V}$ for $V_{DD} = 3 \text{ V}$,
- Propagation delay times $\tau_{PHL}^* < 0.2 \text{ ns}$ and $\tau_{PLH}^* < 0.15 \text{ ns}$,
- A falling delay of 0.35 ns for an output transition from 2 V to 0.5 V , assuming a combined output load capacitance of 300fF and ideal step input.

Solution

- First, the minimum (W/L) ratios of the nMOS and pMOS transistors which are dictated by the propagation delay constraints can be found as follows.

Solution

$$\begin{aligned}\left(\frac{W_n}{L_n}\right) &= \frac{C_{load}}{\tau_{PHL}^* \mu_n C_{ox} (V_{DD} - V_{T,n})} \left[\frac{2V_{T,n}}{V_{DD} - V_{T,n}} + \ln \left(\frac{4(V_{DD} - V_{T,n})}{V_{DD}} - 1 \right) \right] \\ &= \frac{300 \times 10^{-15}}{0.2 \times 10^{-9} \cdot 120 \times 10^{-6} \cdot (3 - 0.8)} \left[\frac{2 \cdot 0.8}{3 - 0.8} + \ln \left(\frac{4(3 - 0.8)}{3} - 1 \right) \right] = 7.9\end{aligned}$$

$$\begin{aligned}\left(\frac{W_p}{L_p}\right) &= \frac{C_{load}}{\tau_{PLH}^* \mu_p C_{ox} (V_{DD} - |V_{T,p}|)} \left[\frac{2|V_{T,p}|}{V_{DD} - |V_{T,p}|} + \ln \left(\frac{4(V_{DD} - |V_{T,p}|)}{V_{DD}} - 1 \right) \right] \\ &= \frac{300 \times 10^{-15}}{0.15 \times 10^{-9} \cdot 60 \times 10^{-6} \cdot (3 - 1)} \left[\frac{2 \cdot 1}{3 - 1} + \ln \left(\frac{4(3 - 1)}{3} - 1 \right) \right] = 25.2\end{aligned}$$

Solution

- During the falling output transition (from 2 V to 0.5 V) described above, the nMOS transistor of the CMOS inverter will operate entirely in the linear region.
- The current equation of the nMOS transistor in this region is

$$C_{load} \frac{dV_{out}}{dt} = -\frac{1}{2} \mu_n C_{ox} \frac{W_n}{L_n} \left[2(V_{OH} - V_{T0,n})V_{out} - V_{out}^2 \right]$$

Solution

- By integrating this expression, we obtain the following relationship.

$$t_{delay} = 0.35 \times 10^{-9} = -2C_{load} \int_{V_{out}=2}^{V_{out}=0.5} \frac{dV_{out}}{\mu_n C_{ox} \frac{W_n}{L_n} [2(V_{OH} - V_{T0,n})V_{out} - V_{out}^2]}$$

$$t_{delay} = \frac{-C_{load}}{\mu_n C_{ox} \frac{W_n}{L_n}} \frac{1}{(V_{OH} - V_{T0,n})} \ln \left(\frac{V_{out}}{2(V_{OH} - V_{T0,n}) - V_{out}} \right) \Bigg|_2^{0.5}$$

$$t_{delay} = \frac{-C_{load}}{\mu_n C_{ox} \left(\frac{W_n}{L_n} \right)} \frac{1}{(3-0.8)} \left[\ln \left(\frac{0.5}{2(3-0.8)-0.5} \right) - \ln \left(\frac{2}{2(3-0.8)-2} \right) \right]$$

$$0.35 \times 10^{-9} = \frac{-300 \times 10^{-15}}{120 \times 10^{-6} \left(\frac{W_n}{L_n} \right) 2.2} [-2.054 + 0.182]$$

Solution

- Now we solve this equation for the nMOS transistor (W/L) ratio:

$$\left(\frac{W_n}{L_n} \right) = 6.1$$

Solution

- Notice that this ratio is smaller than the (W/L) -ratio found from the propagation delay constraint.
- Thus, we take the larger ratio which will satisfy both timing constraints, and determine the size of the nMOS transistor as $W_n = 4.7 \mu\text{m}$, for the given $L_n = 0.6 \mu\text{m}$.

Solution

- Next, the logic threshold constraint of $V_{th} = 1.5$ V will help determine the pMOS transistor dimensions.

$$V_{th} = \frac{V_{T0,n} + \sqrt{\frac{1}{k_R}} (V_{DD} + V_{T0,p})}{1 + \sqrt{\frac{1}{k_R}}} = 1.5$$

Solution

- We find that the ratio k_R which satisfies this design constraint is equal to 0.51.
- This value can now be used to calculate the (W/L)-ratio of the pMOS transistor.

Calculation of Delay Times

- The total capacitive load of the inverter can now be expressed as

$$C_{load} = \alpha_0 + \alpha_n W_n + \alpha_p W_p \quad \text{.....(31)}$$

Calculation of Delay Times

- Where

$$\alpha_0 = 2 D_{drain} (C_{jsw,n} K_{eq,n} + C_{jsw,p} K_{eq,p}) + C_{int} + C_g \quad \dots\dots(32)$$

$$\alpha_n = K_{eq,n} (C_{j0,n} D_{drain} + 2 C_{jsw,n}) \quad \dots\dots(33)$$

$$\alpha_p = K_{eq,p} (C_{j0,p} D_{drain} + 2 C_{jsw,p}) \quad \dots\dots(34)$$

Calculation of Delay Times

- Using (Eqn. 31), the propagation delay expressions (Eqn. 22b) and (Eqn. 23b) for falling and rising output transitions are rewritten as

$$\tau_{PHL} = \left(\frac{\alpha_0 + \alpha_n W_n + \alpha_p W_p}{W_n} \right) \times \left(\frac{L_n}{\mu_n C_{ox} (V_{DD} - V_{T,n})} \right) \times \left[\frac{2 V_{T,n}}{V_{DD} - V_{T,n}} + \ln \left(\frac{4(V_{DD} - V_{T,n})}{V_{DD}} - 1 \right) \right] \dots\dots(35)$$

$$\tau_{PLH} = \left(\frac{\alpha_0 + \alpha_n W_n + \alpha_p W_p}{W_p} \right) \times \left(\frac{L_p}{\mu_p C_{ox} (V_{DD} - |V_{T,p}|)} \right) \times \left[\frac{2 |V_{T,p}|}{V_{DD} - |V_{T,p}|} + \ln \left(\frac{4(V_{DD} - |V_{T,p}|)}{V_{DD}} - 1 \right) \right] \dots\dots(36)$$

Calculation of Delay Times

- The channel lengths L_n and L_p are usually fixed and equal to each other.
- Also, the ratio between the channel widths W_n and W_p is usually dictated by other design constraints such as noise margins and the logic inversion threshold.
- Let this transistor aspect ratio be defined as

$$R \text{ (aspect ratio)} \equiv \frac{W_p}{W_n} \quad \dots\dots(37)$$

Calculation of Delay Times

- Now, the propagation delay τ_{PHL} for the falling output transition can be expressed as a function of the single design parameter W_n , and τ_{PLH} can be expressed as a function of the single design parameter W_p .

$$\tau_{PHL} = \Gamma_n \left(\frac{\alpha_0 + (\alpha_n + R \alpha_p) W_n}{W_n} \right) \quad \dots\dots(38)$$

$$\tau_{PLH} = \Gamma_p \left(\frac{\alpha_0 + \left(\frac{\alpha_n}{R} + \alpha_p \right) W_p}{W_p} \right) \quad \dots\dots(39)$$

Calculation of Delay Times

- where Γ_n and Γ_p are defined as;

$$\Gamma_n = \left(\frac{L_n}{\mu_n C_{ox} (V_{DD} - V_{T,n})} \right) \times \left[\frac{2V_{T,n}}{V_{DD} - V_{T,n}} + \ln \left(\frac{4(V_{DD} - V_{T,n})}{V_{DD}} - 1 \right) \right] \dots\dots(40)$$

$$\Gamma_p = \left(\frac{L_p}{\mu_p C_{ox} (V_{DD} - |V_{T,p}|)} \right) \times \left[\frac{2|V_{T,p}|}{V_{DD} - |V_{T,p}|} + \ln \left(\frac{4(V_{DD} - |V_{T,p}|)}{V_{DD}} - 1 \right) \right] \dots\dots(41)$$

Calculation of Delay Times

- From (Eqn. 38) and (Eqn. 39), the limit delay values can be found as;

$$\tau_{PHL}^{limit} = \Gamma_n (\alpha_n + R \alpha_p) \quad \dots\dots(42)$$

$$\tau_{PLH}^{limit} = \Gamma_p \left(\frac{\alpha_n}{R} + \alpha_p \right) \quad \dots\dots(43)$$

CMOS Ring Oscillator Circuit

- The cascade connection of three identical CMOS inverters are considered, as shown in Fig. 7, where the output node of the third inverter is connected to the input node of the first inverter. As such, the three inverters form a voltage feedback loop.

CMOS Ring Oscillator Circuit

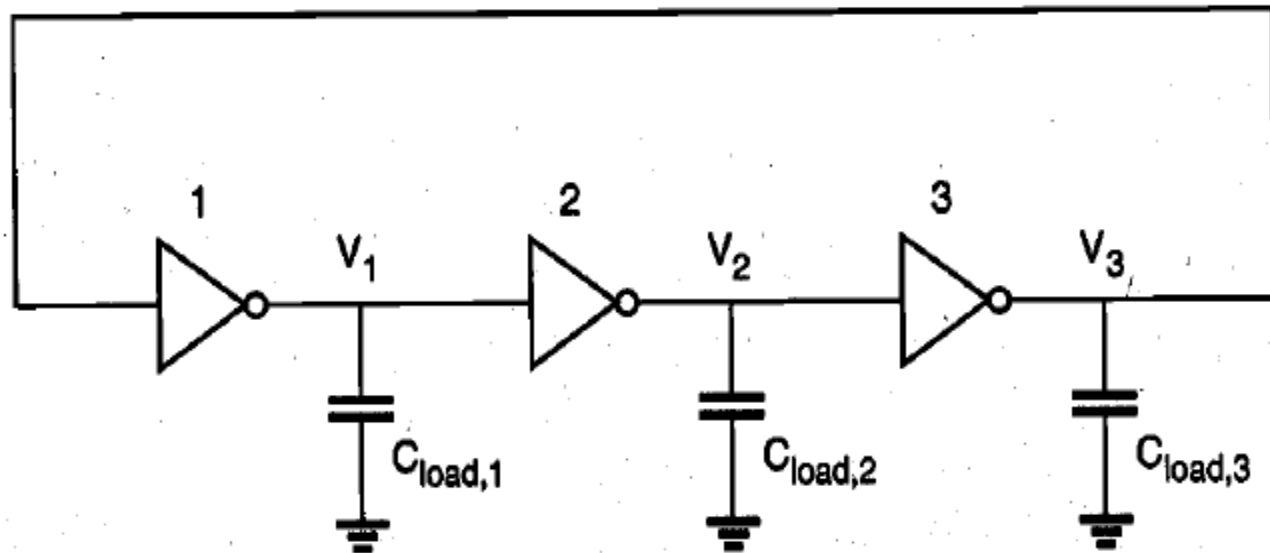


Figure 7. Three-stage ring oscillator circuit consisting of identical inverters.

CMOS Ring Oscillator Circuit

- In the three-stage circuit, the oscillation period T of any of the inverter output voltages can be expressed as the sum of six propagation delay times (Fig. 8).
- Since the three inverters in the closed-loop cascade connection are assumed to be identical, and since the output load capacitances are equal to each other ($C_{load1} = C_{load2} = C_{load3}$);

CMOS Ring Oscillator Circuit

- we can express the oscillation period T in terms of the average propagation delay τ_p , as

$$\begin{aligned} T &= \tau_{PHL1} + \tau_{PLH1} + \tau_{PHL2} + \tau_{PLH2} + \tau_{PHL3} + \tau_{PLH3} \\ &= 2\tau_p + 2\tau_p + 2\tau_p \\ &= 3 \cdot 2\tau_p = 6\tau_p \end{aligned}$$

.....(44)

CMOS Ring Oscillator Circuit

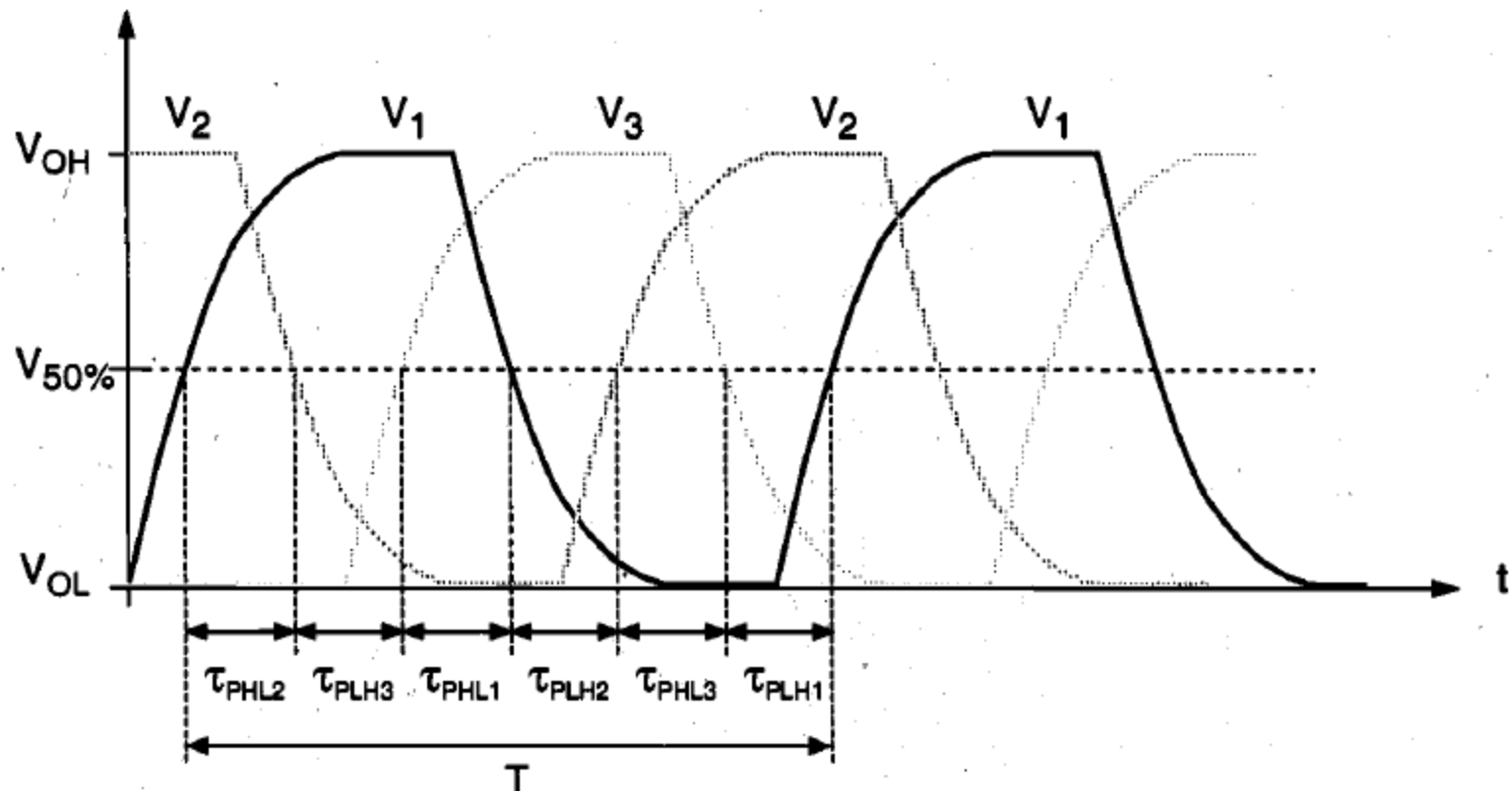


Figure 8. Typical voltage waveforms of the three inverters shown in Fig. 7.

CMOS Ring Oscillator Circuit

- Generalizing this relationship for an arbitrary odd number (n) of cascade-connected inverters, we obtain the oscillation frequency (f) is found to be a very simple function of the average propagation delay of an inverter stage.

$$f = \frac{1}{T} = \frac{1}{2 \cdot n \cdot \tau_P} \quad \text{.....(45)}$$

- And
$$\tau_P = \frac{1}{2 \cdot n \cdot f} \quad \text{.....(46)}$$

CMOS Ring Oscillator Circuit

- Typically, the number n is made much larger than just 3 or 5, in order to keep the oscillation frequency of the circuit within an easily measurable range.
- The ring oscillator circuit can be used as a very simple pulse generator, where the output waveform is utilized as a simple master clock signal generated on-chip.

Interconnect Capacitance Estimation

- In a large-scale integrated circuit, the parasitic interconnect capacitances are among the most difficult parameters to estimate accurately.
- Each interconnection line (wire) is a three-dimensional structure in metal and/or polysilicon, with significant variations of shape, thickness, and vertical distance from the ground plane (substrate).

Interconnect Capacitance Estimation

- Also, each interconnect line is typically surrounded by a number of other lines, either on the same level or on different levels.
- Figure 9 shows a simplified view of six interconnections on three different levels, running in close proximity of each other.

Interconnect Capacitance Estimation

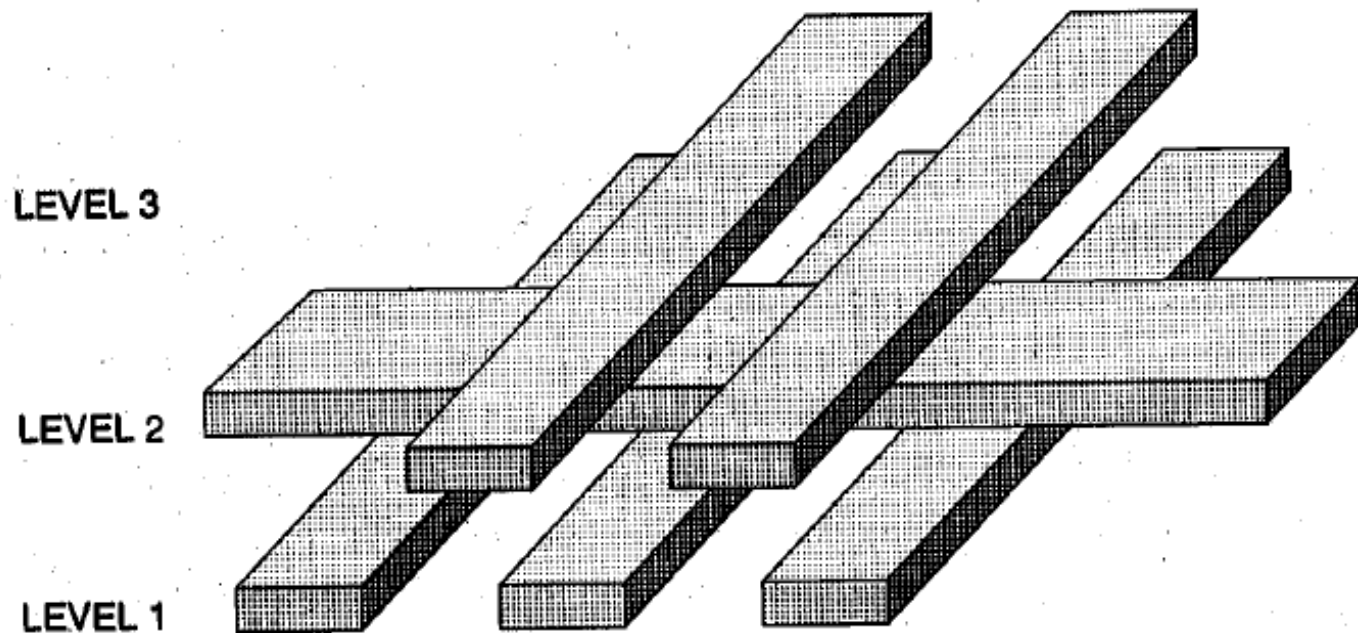


Figure 9. An example of six interconnect lines running on three different levels.

Interconnect Capacitance Estimation

- First, consider the section of a single interconnect which is shown in Fig. 10.
- It is assumed that this wire segment has a length of (l) in the current direction, a width of (w) and a thickness of (t).
- Moreover, we assume that the interconnect segment runs parallel to the chip surface and is separated from the ground plane by a dielectric (oxide) layer of height (h).

Interconnect Capacitance Estimation

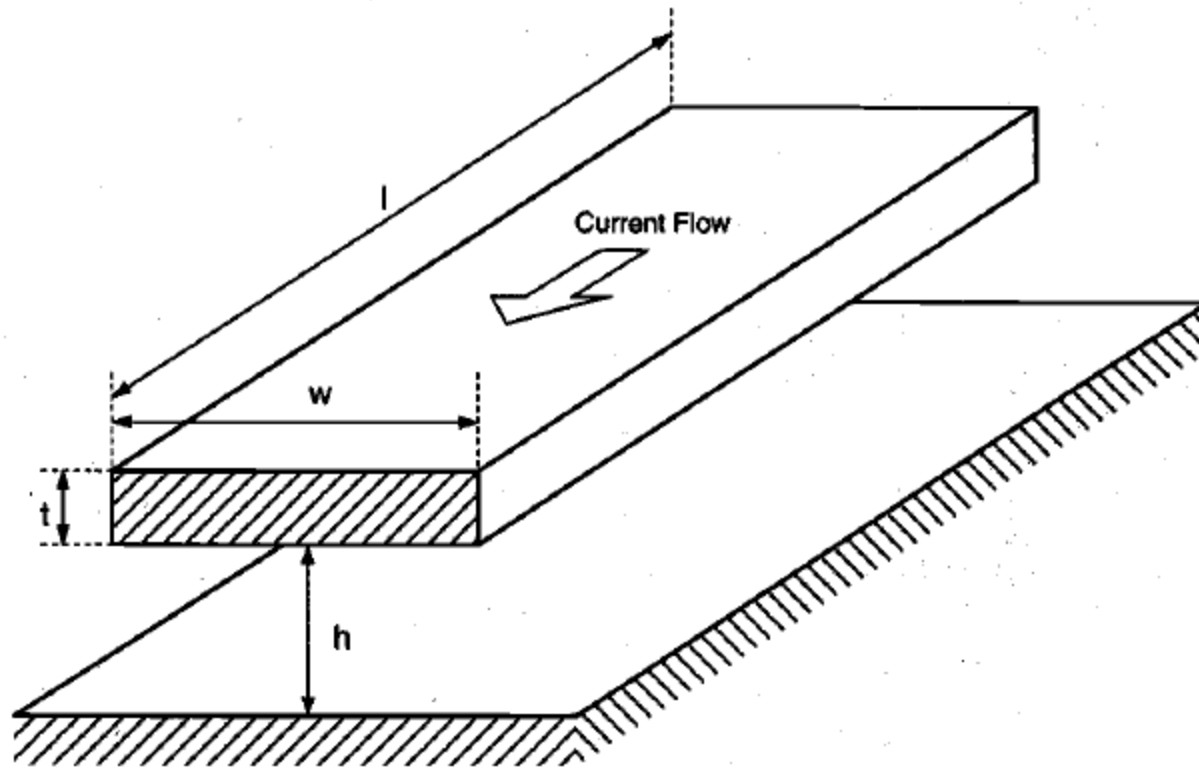


Figure 10: Interconnect segment running parallel to the surface, which is used for parasitic resistance and capacitance estimations.

Interconnect Capacitance Estimation

- Using the basic geometry given in Fig. 10, one can calculate the parallel-plate capacitance C_{pp} of the interconnect segment.
- However, in interconnect lines where the wire thickness (t) is comparable in magnitude to the ground-plane distance (h), fringing electric fields significantly increase the total parasitic capacitance.

Interconnect Capacitance Estimation

- The following two cases are considered for two different ranges of width line width (w).

$$C = \epsilon \left[\frac{\left(w - \frac{t}{2}\right)}{h} + \frac{2\pi}{\ln\left(1 + \frac{2h}{t} + \sqrt{\frac{2h}{t}\left(\frac{2h}{t} + 2\right)}\right)} \right] \quad \text{for } w \geq \frac{t}{2} \quad \dots\dots(47)$$

$$C = \epsilon \left[\frac{w}{h} + \frac{\pi\left(1 - 0.0543 \cdot \frac{t}{2h}\right)}{\ln\left(1 + \frac{2h}{t} + \sqrt{\frac{2h}{t}\left(\frac{2h}{t} + 2\right)}\right)} + 1.47 \right] \quad \text{for } w < \frac{t}{2} \quad \dots\dots(48)$$

Interconnect Resistance Estimation

- The parasitic resistance of a metal or polysilicon line can also have a significant influence on the signal propagation delay over that line.
- The resistance of a line depends on the type of material used (e.g., polysilicon, aluminum, or gold), the dimensions of the line and finally, the number and locations of the contacts on that line.

Interconnect Resistance Estimation

- Considering the interconnection line shown in Fig. 10, the total resistance in the indicated current direction can be found as;

$$R_{wire} = \rho \cdot \frac{l}{w \cdot t} = R_{sheet} \left(\frac{l}{w} \right) \quad \text{.....(49)}$$

- where (ρ) represents the characteristic resistivity of the interconnect material, and R_{sheet} represents the sheet resistivity of the line, in (Ω/square).

Interconnect Resistance Estimation

$$R_{sheet} = \frac{\rho}{t} \quad \text{.....(50)}$$

- Using (Eqn. 49), we can estimate the total parasitic resistance of a wire segment based on its geometry.

Calculation of Interconnect Delay

RC Delay Models

- The simplest model which can be used to represent the resistive and capacitive parasitics of the interconnect line consists of one lumped resistance and one lumped capacitance (Fig. 11(a)).
- Assuming that the capacitance is discharged initially, and assuming that the input signal is a rising step pulse at time $t = 0$, the output voltage waveform of this simple RC circuit is found as

Calculation of Interconnect Delay

RC Delay Models

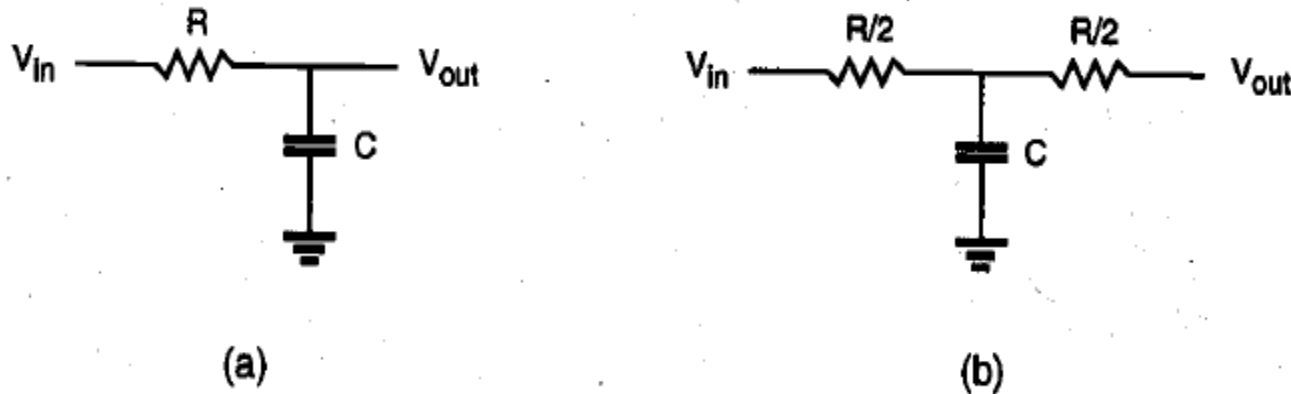


Figure 11. (a) Simple lumped RC model of an interconnect line, where R and C represent the total line resistance and capacitance, respectively. (b) The T-model of the same line.

Calculation of Interconnect Delay

RC Delay Models

$$V_{out}(t) = V_{DD} \left(1 - e^{-\frac{t}{RC}} \right) \quad \dots\dots(51)$$

- The rising output voltage reaches the 50%-point at $t = \tau_{PLH}$, thus,

$$V_{50\%} = V_{DD} \left(1 - e^{-\frac{\tau_{PLH}}{RC}} \right) \quad \dots\dots(52)$$

- and the propagation delay for the simple lumped RC network is found as

$$\tau_{PLH} \approx 0.69 RC \quad \dots\dots(53)$$

Calculation of Interconnect Delay

RC Delay Models

- Unfortunately, this simple lumped RC network model provides a very rough approximation of the actual transient behavior of the interconnect line. The accuracy of the simple lumped RC model can be significantly improved by dividing the total line resistance into two equal parts (the T-model), as shown in Fig. 11(b).

Calculation of Interconnect Delay

RC Delay Models

- The transient behavior of an interconnect line can be more accurately represented using the RC ladder network, shown in Fig. 12.
- Here, each RC-segment consists of a series resistance (R/N), and a capacitance (C/N) connected between the node and the ground.
- It can be expected that the accuracy of this model increases with increasing N , where the transient behavior approaches that of a distributed RC line for very large values of N .

Calculation of Interconnect Delay

RC Delay Models

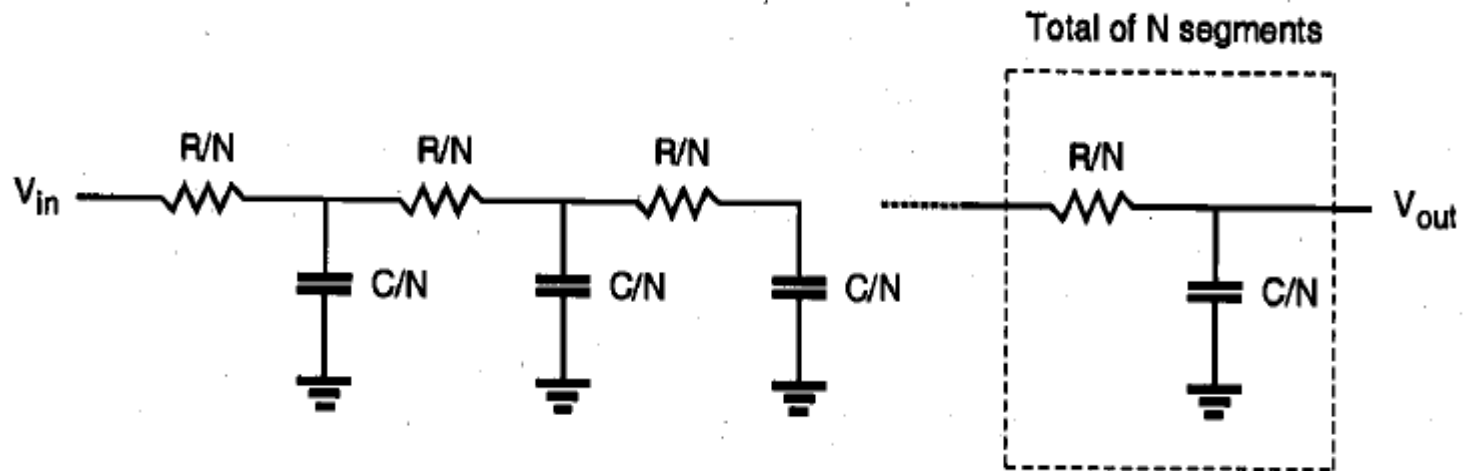


Figure 12. Distributed RC ladder network model consisting of N equal segments.

The Elmore Delay

- Consider a general RC tree network, as shown in Fig. 13.
 - (i) there are no resistor loops in this circuit,
 - (ii) all of the capacitors in an RC tree are connected between a node and the ground,
 - (iii) there is one input node in the circuit.
- Also there is a unique resistive path, from the input node to any other node in the circuit.

The Elmore Delay

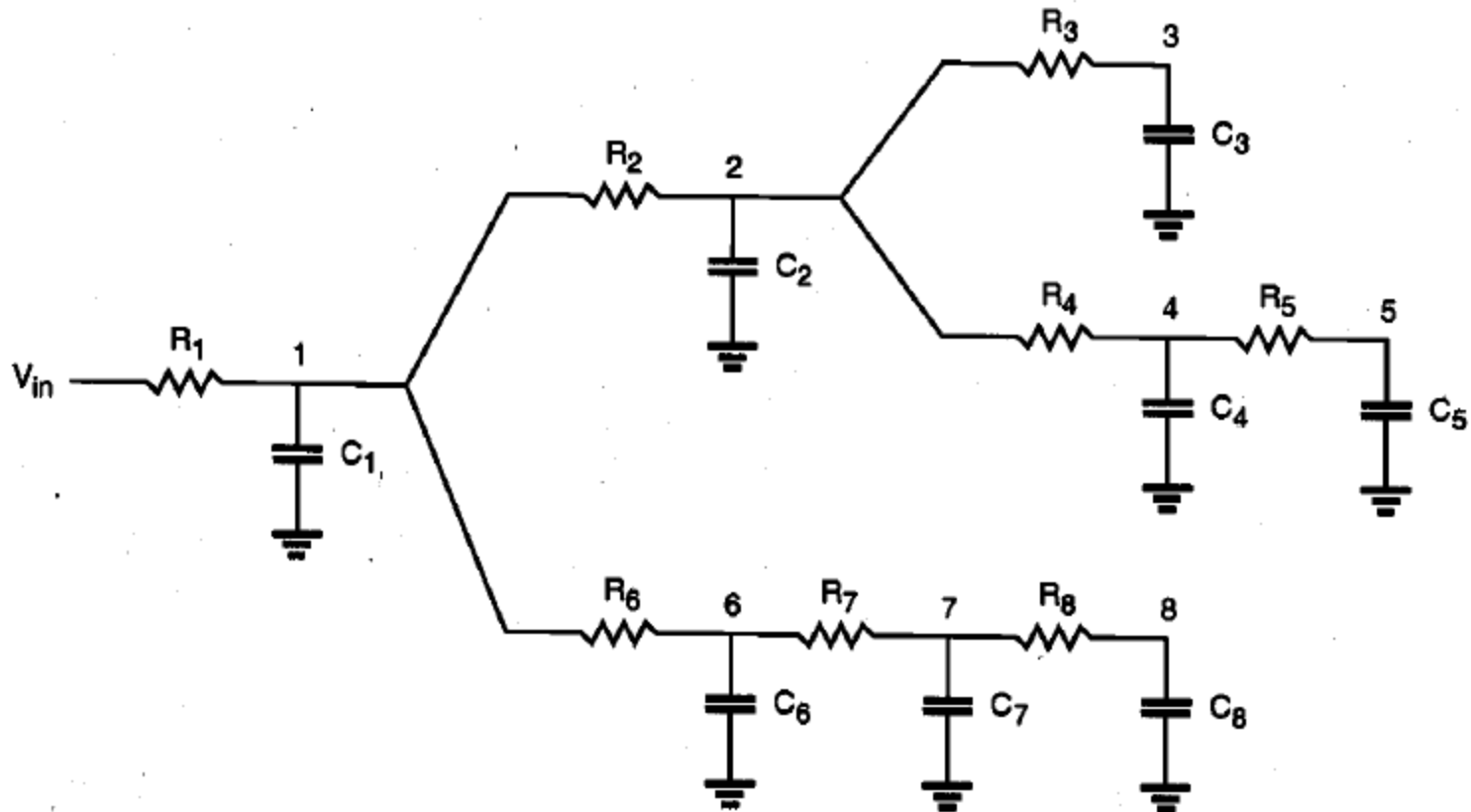


Figure 13. A general RC tree network consisting of several branches.

The Elmore Delay

- Assuming that the input signal is a step pulse at time $t = 0$, the Elmore delay at node i of this RC tree is given by the following expression.

$$\tau_{Di} = \sum_{j=1}^N C_j \sum_{\substack{\text{for all} \\ k \in P_{ij}}} R_k \quad \text{.....(54)}$$

The Elmore Delay

- For an example, the Elmore delay at node 7 can be found according to (Eqn. 54) as

$$\begin{aligned}\tau_{D7} = & R_1 C_1 + R_1 C_2 + R_1 C_3 + R_1 C_4 + R_1 C_5 + (R_1 + R_6) C_6 \\ & + (R_1 + R_6 + R_7) C_7 + (R_1 + R_6 + R_7) C_8\end{aligned}\quad \text{.....(55)}$$

- Similarly, the Elmore delay at node 5 can be calculated as

$$\begin{aligned}\tau_{D5} = & R_1 C_1 + (R_1 + R_2) C_2 + (R_1 + R_2) C_3 + (R_1 + R_2 + R_4) C_4 \\ & + (R_1 + R_2 + R_4 + R_5) C_5 + R_1 C_6 + R_1 C_7 + R_1 C_8\end{aligned}\quad \text{.....(56)}$$

The Elmore Delay

- As a special case of the general RC tree network, consider now the simple RC ladder network shown in Fig. 14.
- Here, the entire network consists of one single branch, and the Elmore delay from the input to the output (node N) is found according to (Eqn. 54) as

$$\tau_{DN} = \sum_{j=1}^N C_j \sum_{k=1}^j R_k \quad \text{.....(57)}$$

The Elmore Delay

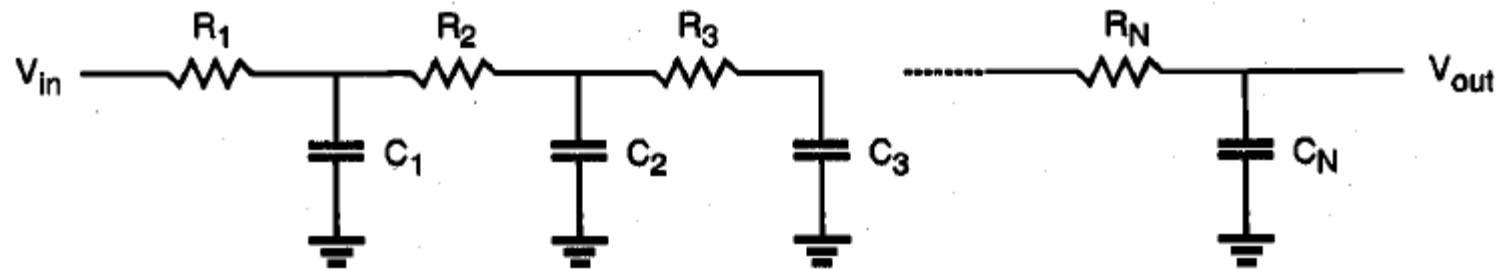


Figure 14 . Simple RC ladder network consisting of one branch.

The Elmore Delay

- If we further assume a uniform RC ladder network, consisting of identical elements (R/N) and (C/N) as shown in Fig. 12, then the Elmore delay from the input to the output node becomes

$$\begin{aligned}\tau_{DN} &= \sum_{j=1}^N \left(\frac{C}{N} \right) \sum_{k=1}^j \left(\frac{R}{N} \right) \\ &= \left(\frac{C}{N} \right) \left(\frac{R}{N} \right) \left(\frac{N(N+1)}{2} \right) = RC \left(\frac{N+1}{2N} \right) \quad \text{.....(58)}\end{aligned}$$

The Elmore Delay

- For very large N (distributed RC line behavior), this delay expression reduces to

$$\tau_{DN} = \frac{RC}{2} \quad \text{for } N \rightarrow \infty \quad \text{.....(59)}$$

Switching Power Dissipation of CMOS Inverters

- Consider the simple CMOS inverter circuit shown in Fig. 15.
- We will assume that the input voltage is an ideal step waveform with negligible rise and fall times.
- When the input voltage switches from low to high, the pMOS transistor in the circuit is turned off, and the nMOS transistor starts conducting.

Switching Power Dissipation of CMOS Inverters

- During this phase, the output load capacitance C_{load} is being discharged through the nMOS transistor.
- Thus, the capacitor current equals the instantaneous drain current of the nMOS transistor.

Switching Power Dissipation of CMOS Inverters

- When the input voltage switches from high to low, the nMOS transistor in the circuit is turned off, and the pMOS transistor starts conducting.
- During this phase, the output load capacitance C_{load} is being charged up through the pMOS transistor;
- therefore, the capacitor current equals the instantaneous drain current of the pMOS transistor.

Switching Power Dissipation of CMOS Inverters

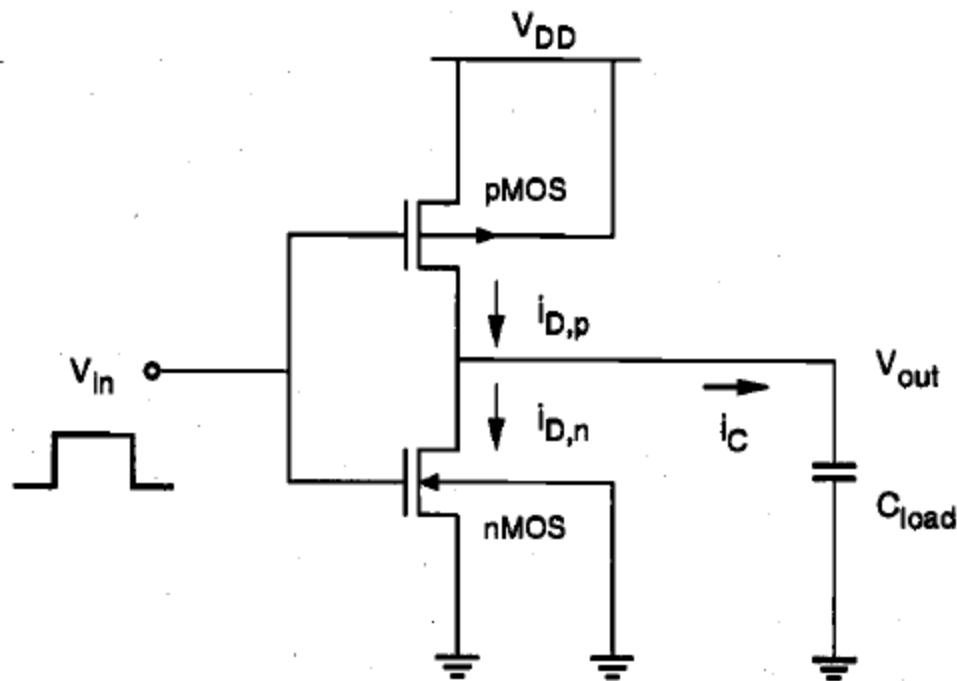


Figure 15. CMOS inverter used in the dynamic power-dissipation analysis.

Switching Power Dissipation of CMOS Inverters

- Assuming periodic input and output waveforms, the average power dissipated by any device over one period can be found as follows:

$$P_{avg} = \frac{1}{T} \int_0^T v(t) \cdot i(t) dt \quad \text{.....(60)}$$

Switching Power Dissipation of CMOS Inverters

- Since during switching, the nMOS transistor and the pMOS transistor in a CMOS inverter conduct current for one-half period each, the average power dissipation of the CMOS inverter can be calculated as the power required to charge up and charge down the output load capacitance.

$$P_{avg} = \frac{1}{T} \left[\int_0^{T/2} V_{out} \left(-C_{load} \frac{dV_{out}}{dt} \right) dt + \int_{T/2}^T (V_{DD} - V_{out}) \left(C_{load} \frac{dV_{out}}{dt} \right) dt \right]$$

.....(61)

Switching Power Dissipation of CMOS Inverters

- Evaluating the integrals in (Eqn. 61), we obtain

$$P_{avg} = \frac{1}{T} \left[\left(-C_{load} \frac{V_{out}^2}{2} \right) \Big|_0^{T/2} + \left(V_{DD} \cdot V_{out} \cdot C_{load} - \frac{1}{2} C_{load} V_{out}^2 \right) \Big|_{T/2}^T \right] \dots\dots(62)$$

$$P_{avg} = \frac{1}{T} C_{load} V_{DD}^2 \dots\dots(63)$$

- Noting that $f = 1/T$, this expression can also be written as:

$$P_{avg} = C_{load} \cdot V_{DD}^2 \cdot f \dots\dots(64)$$

Generalized CMOS logic circuit

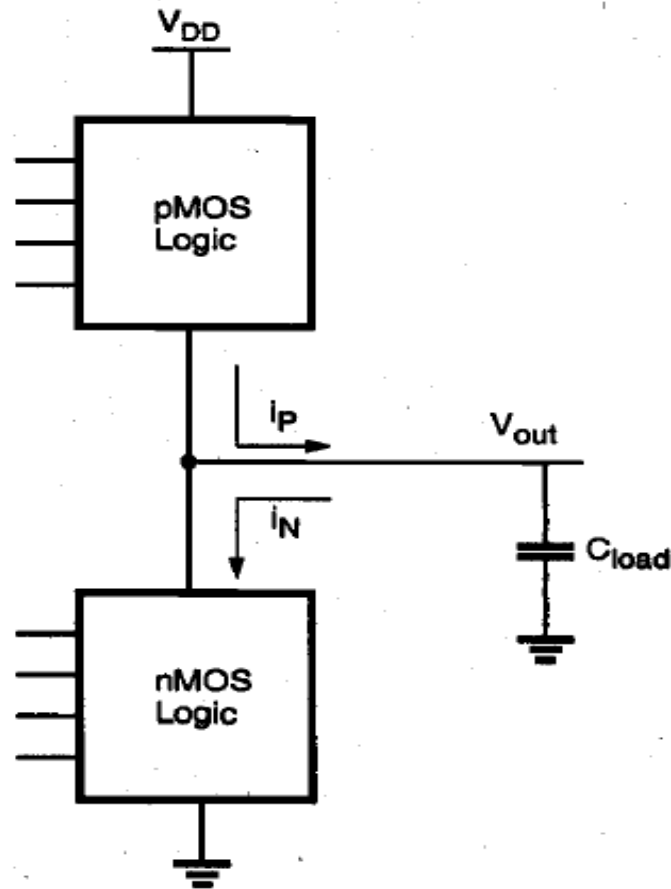


Figure 16. Generalized CMOS logic circuit.

Power Meter Simulation

- Using a simple simulation model called the power meter, we can estimate the average power dissipation of an arbitrary device or circuit driven by a periodic input, with transient circuit simulation.
- Consider the circuit structure shown in Fig. 17, in which a zero-volt independent voltage source is connected in series with the power supply voltage source V_{DD} of the device or circuit in question.

Power Meter Simulation

- Consequently, the instantaneous power supply current $i_{DD}(t)$ which is being drawn by the circuit will also pass through the zero-volt voltage source, $i_s(t) = i_{DD}(t)$.

Power Meter Simulation

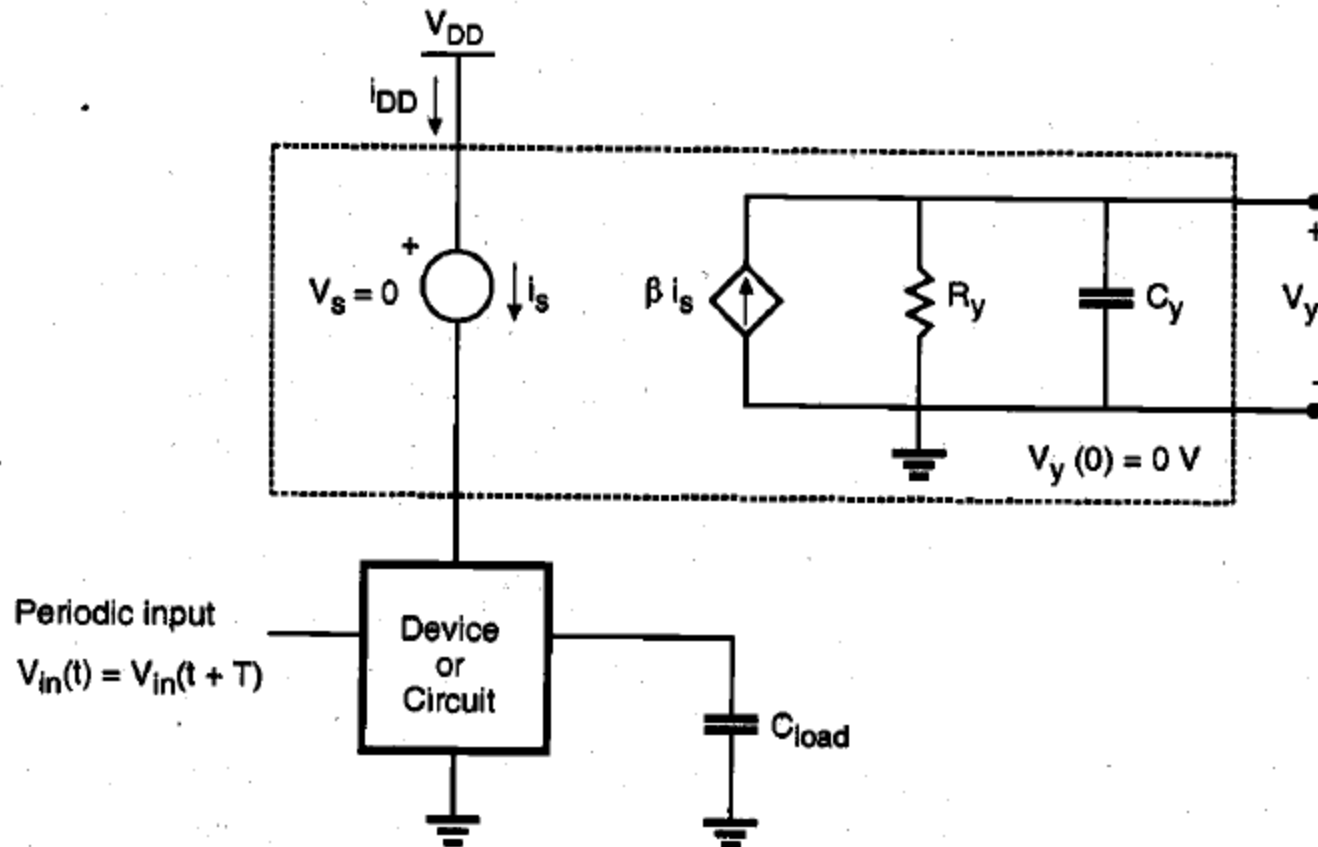


Figure 17. The power meter circuit used for the simulation of average dynamic power dissipation of an arbitrary device or circuit

Power Meter Simulation

- The current equation for the common node of the power meter circuit can be written as follows:

$$C_y \frac{dV_y}{dt} = \beta i_S - \frac{V_y}{R_y} \quad \text{.....(65)}$$

Power Meter Simulation

- The initial condition of the node voltage V_y is set as $V_y(0) = 0$ V. Then, the time-domain solution of $V_y(t)$ can be found by integrating (Eqn. 65).

$$V_y(t) = \frac{\beta}{C_y} \int_0^t e^{-\frac{1}{R_y C_y}(t-\tau)} i_{DD}(\tau) d\tau \quad \dots\dots(66)$$

Power Meter Simulation

- Assuming $R_y C_y \gg T$, the voltage value $V_y(T)$ at the end of one period can be approximated as follows

$$V_y(T) \approx \frac{\beta}{C_y} \int_0^T i_{DD}(\tau) d\tau \quad \text{.....(67)}$$

- If the constant coefficient value of the current-controlled current source is set to be

$$\beta = V_{DD} \frac{C_y}{T} \quad \text{.....(68)}$$

Power Meter Simulation

- The voltage value $V_y(T)$ at the end of one period will be found by transient simulation as:

$$V_y(T) = V_{DD} \cdot \frac{1}{T} \int_0^T i_{DD}(\tau) d\tau \quad \text{.....(69)}$$

- The right-hand side of (Eqn. 69) corresponds to the average power drawn from the power supply source over one period.
- Thus, the value of the node voltage V_y at $t = T$ gives the average power dissipation.

Power-Delay Product

- The power-delay product (PDP) is a fundamental parameter which is often used for measuring the quality and the performance of a CMOS process and gate design.
- As a physical quantity, the power-delay product can be interpreted as the average energy required for a gate to switch its output voltage from low to high and from high to low.

Power-Delay Product

- In a CMOS logic gate, energy is dissipated
 - (i) by the pMOS network while the output load capacitance C_{load} is being charged up from 0 to V_{DD} , and
 - (ii) by the nMOS network while the output load capacitance is being charged down from V_{DD} to 0.

Power-Delay Product

- Following a simple analysis procedure which is very similar to the one used for deriving the average dynamic power dissipation (Eqn. 63) in CMOS logic gates and ignoring the short-circuit and leakage currents, the amount of energy required to switch the output can be found as

$$PDP = C_{load} V_{DD}^2 \quad \text{.....(70)}$$

Power-Delay Product

- The energy described by (Eqn. 70) is mainly dissipated as heat when the nMOS and pMOS transistors conduct current during switching.
- Thus, from a design point-of-view, it is desirable to minimize the power-delay product.
- Since the PDP is a function of the output load capacitance and the power supply voltage, the designer should try to keep both C_{load} and V_{DD} as small as possible when designing a CMOS logic gate.

Power-Delay Product

- The power delay product is also defined as

$$PDP = 2 P_{avg}^* \tau_P \quad \text{.....(71)}$$

Power-Delay Product

- where P_{avg}^* is the average switching power dissipation at maximum operating frequency and τ_p is the average propagation delay, as defined in (Eqn. 4).
- The factor of 2 in (Eqn. 71) accounts for two transitions of the output, from low to high and from high to low.

Power-Delay Product

- Using (Eqn. 63) and (Eqn. 4), this expression can be rewritten as

$$\begin{aligned} PDP &= 2 \left(C_{load} V_{DD}^2 f_{max} \right) \tau_P \\ &= 2 \left[C_{load} V_{DD}^2 \left(\frac{1}{\tau_{PHL} + \tau_{PLH}} \right) \right] \left(\frac{\tau_{PHL} + \tau_{PLH}}{2} \right) \\ &= C_{load} V_{DD}^2 \end{aligned} \quad \text{.....(72)}$$

MOS TRANSISTOR

VLSI DESIGN

PART-1

MOS TRANSISTOR

- The MOSFET is the fundamental building block of MOS and CMOS digital integrated circuits.
- Compared to BJT, the MOS transistor occupies relatively smaller silicon area and fewer processing steps in its fabrication.
- The technological advantages with its simple operation have helped make the MOS transistor widely used switching device in LSI and VLSI circuits.

MOS Structure

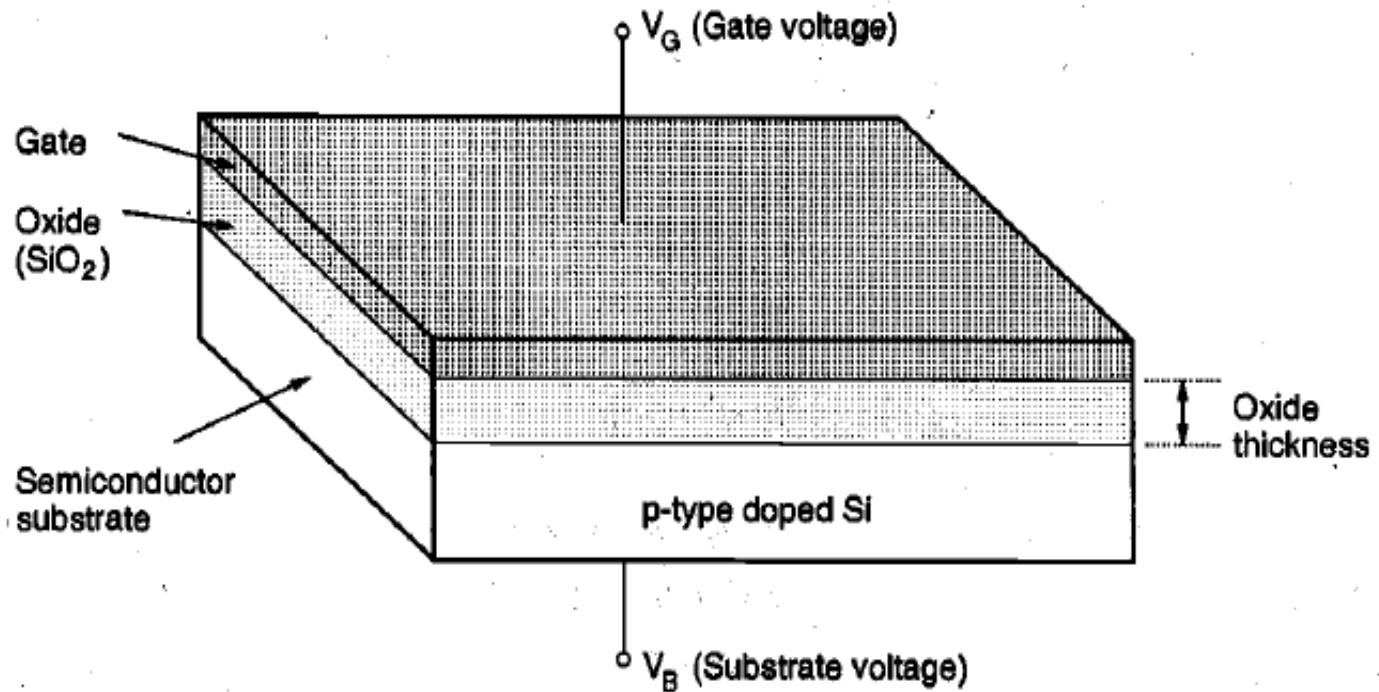


Fig. 1: MOS Structure

MOS Structure

- A simple two terminal MOS structure is shown in Fig. 1 of the previous slide.
- The electrical behavior of the MOS structure is described here.
- The structure consists of three layers:
- The metal gate electrode, the insulating oxide (SiO_2) layer, and the p-type bulk semiconductor (Si), called the substrate.

MOS Structure

- The MOS structure forms a capacitor, with the gate and the substrate acting as the two terminals (plates) and the oxide layer as the dielectric.
- The thickness of the silicon dioxide layer is usually between 10 nm and 50 nm.
- The carrier concentration and its local distribution within the semiconductor substrate can now be manipulated by the external voltages applied to the gate and substrate terminals.

MOS Structure

- A basic understanding of the bias conditions for establishing different carrier concentrations in the substrate will also provide valuable insight into the operating conditions of more complicated MOSFET structures.

MOS Structure

- The basic electrical properties of the semiconductor (Si) substrate, which acts as one of/the electrodes of the MOS capacitor is considered.
- The equilibrium concentrations of mobile carriers in a semiconductor always obey the **Mass Action Law** given by;

$$n.p=n_i^2 \text{ (1)}$$

MOS Structure

- Here, n and p denote the mobile carrier concentrations of electrons and holes, respectively, and n_i denotes the intrinsic carrier concentration of silicon, which is a function of the temperature T .
- At room temperature, i.e., $T = 300\text{K}$, n_i is approximately equal to $1.45 \times 10^{10} \text{ cm}^{-3}$.

Review: Charge Carriers in Semiconductors

- The energy band diagram of a semiconductor, shown in Fig. 2, displaying the conduction and valence bands and the bandgap.
- The electrons in the valence band are bound to the atoms of the crystal. They need to have enough extra energy to go across the forbidden bandgap to get into the energy levels of the conduction band.

Review: Charge Carriers in Semiconductors

- To quantify electron concentration, we count the number of electrons in the conduction band per a unit volume, typically a cm^3 . This is shown by the symbol n . When the system is in equilibrium, it is shown by n_0 .

Review: Charge Carriers in Semiconductors

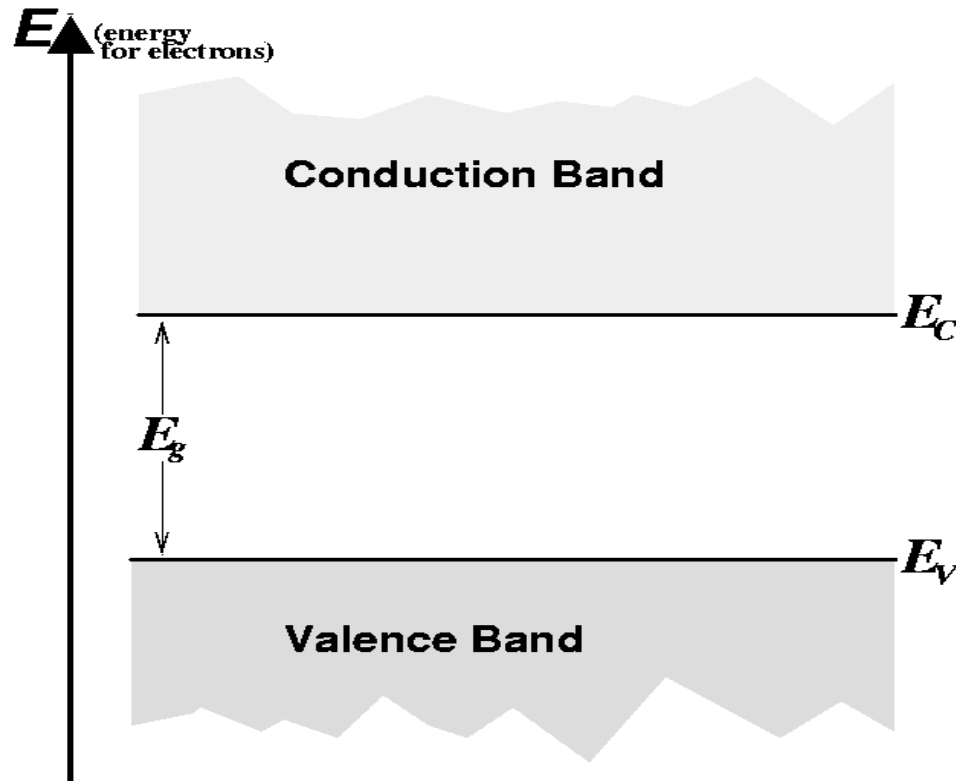


Fig. 2: Energy Band Diagram of Semiconductor

Review: Charge Carriers in Semiconductors

- When an electron leaves the valence band to go to the conduction band, it is essentially breaking free of its bond, leaving behind an empty state available for another electron in the valence band energy levels.
- This empty state is called a "hole" and effectively behaves like a positively charged particle.

Review: Charge Carriers in Semiconductors

- The number of holes in the valence band per unit volume is called "hole concentration" and shown by the symbol p .
- In equilibrium, we use p_0 .

Mass Action Law

- In an intrinsic semiconductor, the electron concentration and the hole concentrations are both equal to the intrinsic concentration, i.e. $n = p = n_i$.
- If this were not true, the material would not be electrically neutral.
- Adding an n-type impurity to the semiconductor increases n and decreases p . Similarly, adding a p-type impurity increases p and decreases n .

Review: Intrinsic and Extrinsic Materials

- An intrinsic semiconductor is ideally a perfect crystal. When an electron in an intrinsic semiconductor gets enough energy, it can go to the conduction band and leave behind a hole.
- This process is called "electronhole pair (EHP) creation". For the intrinsic material, since electrons and holes are always created in pairs,

$$n = p = n_i \quad \text{.....(2)}$$

Review: Intrinsic and Extrinsic Materials

- where n_i is the symbol for "intrinsic carrier concentration at room temperature, relatively few electrons have enough thermal energy to make this jump.
- So the intrinsic carrier concentration in silicon at room temperature is approximately;

$$n_i = 1.45 \times 10^{10} \left[\frac{1}{\text{cm}^3} \right] \text{ silicon, } 300^\circ \text{K} \quad \text{..... (3)}$$

Review: Intrinsic and Extrinsic Materials

- Extrinsic semiconductors are made by introducing different atoms, called dopant atoms, into the crystal.
- We can get two types of extrinsic material:
 1. n-type:
 - The dopant atoms added to the semiconductor crystal in this case are donor atoms.

Review: Intrinsic and Extrinsic Materials

- When these atoms are included in the silicon crystal, one of the electrons in this shell can easily jump to the conduction band, leaving a positively charged atom behind.
- This process is sometimes called “activation” or “ionization” of the donor atoms.

Review: Intrinsic and Extrinsic Materials

- The positively charged donor atom that is left behind after ionization is immobile and does not contribute to conduction.
- The electron leaving the atom by ionization does, and is counted in the electron concentration n .
- Because the activation energy is low, at room temperature almost all of the donor atoms included in the crystal will give an electron to the conduction band.

Review: Intrinsic and Extrinsic Materials

- So if N_D is the donor concentration, for an n-type material at equilibrium:

$$n_0 \approx N_D \left[\frac{1}{cm^3} \right] \dots\dots\dots (4)$$

Review: Intrinsic and Extrinsic Materials

2. p-type:

- The dopant atoms in this case are acceptor atoms.
- For silicon, we can use boron (B), Aluminum (Al) and Gallium (Ga) as acceptors. These are column III elements, with three electrons in their outermost shell.

Review: Intrinsic and Extrinsic Materials

- When these atoms are included in the silicon crystal, one of the electrons in the silicon valence band can easily jump to the valence shell of one of the acceptor atoms, leaving a hole behind and making the acceptor atom negatively charged.

Review: Intrinsic and Extrinsic Materials

- The negatively charged acceptor atom after an electron joins its valence shell is immobile and does not contribute to conduction.
- The hole left behind by that electron does, and is counted in the hole concentration p .
- Because the activation energy is low, at room temperature almost all of the acceptor atoms included in the crystal will accept an electron from the valence band.

Review: Intrinsic and Extrinsic Materials

- So if N_A is the acceptor concentration, for a p-type material at equilibrium:

$$p_0 \approx N_A \left[\frac{1}{\text{cm}^3} \right] \dots\dots\dots(5)$$

Review: Law of Mass Action

- For both intrinsic and extrinsic materials, at equilibrium:

$$n_0 p_0 = n_i^2 \Rightarrow n_0 = \frac{n_i^2}{p_0}, p_0 = \frac{n_i^2}{n_0} \quad \text{.....(6)}$$

- So for an n-type material with only donor atoms:

$$n_0 = N_D, p_0 = \frac{n_i^2}{n_0} = \frac{n_i^2}{N_D} \quad \text{.....(7)}$$

Review: Law of Mass Action

- And for a p-type material with only acceptor atoms:

$$p_0 = N_A, \quad n_0 = \frac{n_i^2}{p_0} = \frac{n_i^2}{N_A} \quad \text{.....(8)}$$

Mass Action Law

- Let N_D be the number of donor atoms per cm^3 and N_A the number of acceptor atoms per cm^3 .
- We assume that all donor atoms and all acceptor atoms are ionized so that there are N_D bound positive charges per cm^3 and N_A bound negative charges per cm^3 .

Mass Action Law

- Each donor ion has a charge $+q$ and each acceptor ion has a charge $-q$. The total number of negative charges per cm^3 is equal to the number n of free electrons per cm^3 plus the number N_A of bound acceptor atoms per cm^3 , i.e. $n + N_A$.
- Similarly, the number of positive charges per cm^3 is equal to the number p of holes per cm^3 plus the number N_D of bound donor atoms per cm^3 , i.e. $p + N_D$.

Mass Action Law

- Because the semiconductor is electrically neutral, the number of positive charges must equal the number of negative charges. This gives the condition;

$$n + N_A = p + N_D \quad \text{.....(9)}$$

- In an n-type semiconductor, $N_A = 0$ and $p \ll n$ so that the above equation can be solved for n to obtain

$$n = p + N_D \simeq N_D \quad \text{.....(10)}$$

Mass Action Law

- The approximation in this equation can be used to solve for the hole concentration p to obtain

$$p \simeq \frac{n_i^2}{N_D} \quad \text{.....(11)}$$

- Similarly, in a p-type semiconductor, we can write

$$p = n + N_A \simeq N_A \quad \text{.....(12)}$$

$$n \simeq \frac{n_i^2}{N_A} \quad \text{.....(13)}$$

MOS Structure

- Assuming that the substrate is uniformly doped with an acceptor (e.g., Boron) concentration N_A , the equilibrium electron and hole concentrations in the p-type substrate are approximated by

$$n_{po} \equiv \frac{n_i^2}{N_A} \quad \text{.....(14)}$$

$$p_{po} \equiv N_A \quad \text{.....(15)}$$

MOS Structure

- The doping concentration N_A is typically on the order of 10^{15} to 10^{16} cm^{-3} ; thus, it is much greater than the intrinsic carrier concentration n_i .

Review: The Fermi-Dirac Distribution Function and Fermi Level

- The conduction band in a piece of semiconductor consists of many available, allowed, empty energy levels.
- When calculating how many electrons will fill these levels and thus be counted in n , contributing to conductivity, we consider two factors:

Review: The Fermi-Dirac Distribution Function and Fermi Level

- How many energy levels are there within a given range of energy, in the conduction band,
- How likely is it that each level will be populated by an electron.

Review: The Fermi-Dirac Distribution Function and Fermi Level

- The likelihood in the second item is given by a probability function called the Fermi-Dirac distribution function.
- $f(E)$ is the probability that a level with energy E will be filled by an electron, and the expression is:

$$f(E) = \frac{1}{1 + \exp((E - E_F)/(k_B T))} \quad \text{.....(16)}$$

Review: The Fermi-Dirac Distribution Function and Fermi Level

- where k_B is Boltzmann's constant, 8.62×10^{-5} [eV/K], and T is the temperature in degrees Kelvin.
- E_F is called the Fermi energy or Fermi level. It is determined as the energy point where the probability of occupancy by an electron is exactly 50%, or 0.5:

$$f(E_F) = \frac{1}{1 + \exp((E_F - E_F)/(k_B T))} = \frac{1}{1 + \exp(0)} = \frac{1}{2} \quad \dots (17)$$

Review:Fermi Level in Intrinsic and Extrinsic Semiconductors

- In an intrinsic semiconductor, $n = p$. If we use the band-symmetry approximation, which assumes that there are equal number of states in equal-sized energy bands at the edges of the conduction and valence bands, $n = p$ implies that there is an equal chance of finding an electron at the conduction band edge as there is of finding a hole at the valence band edge:

Review:Fermi Level in Intrinsic and Extrinsic Semiconductors

- $f(E_C) = 1 - f(E_V)$
- For an intrinsic semiconductor, the Fermi level E_F must be in the middle of the bandgap as shown in Fig. 3 and this level is called the “intrinsic Fermi level” and shown by E_i :

$$E_i = E_C - E_g/2 = E_V + E_g/2 \quad \text{..... (18)}$$

Where E_g is the bandgap energy.

Energy Band Diagram of Intrinsic Semiconductor

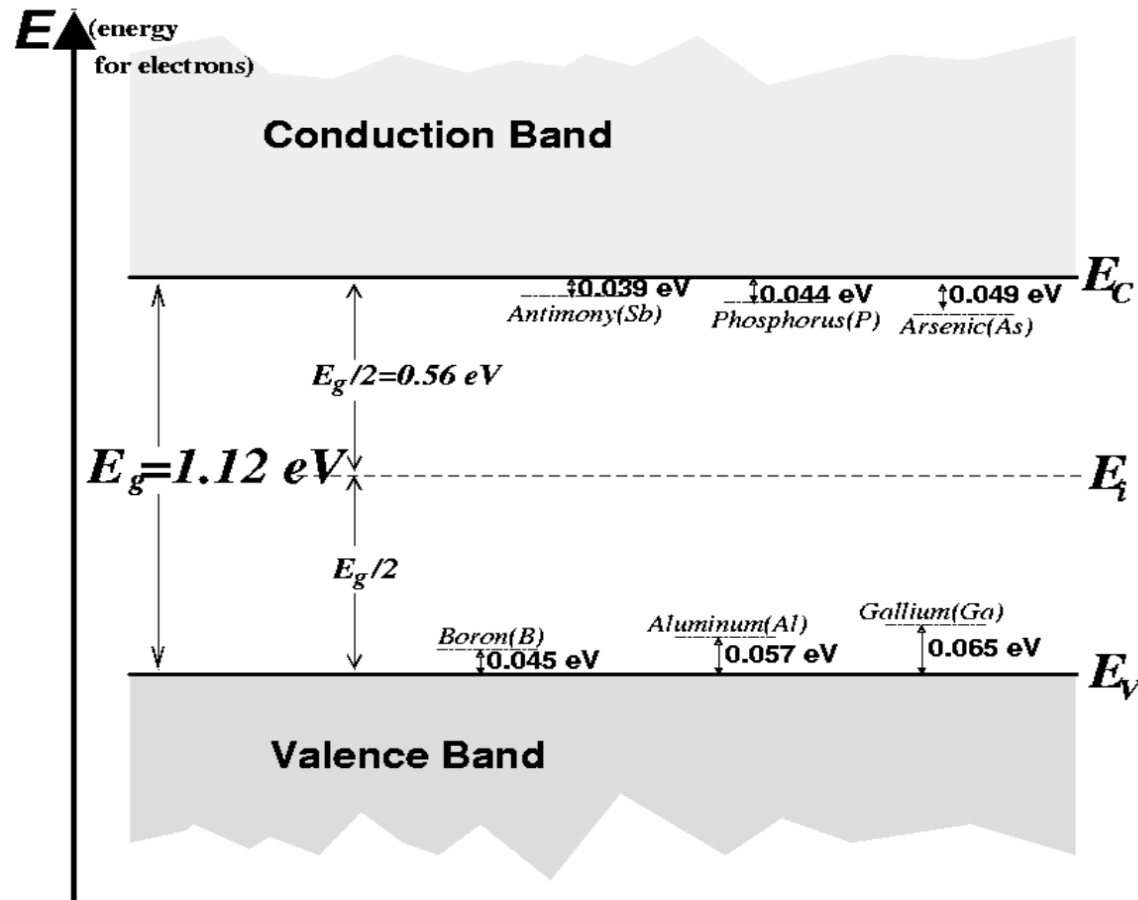


Fig. 3: Energy Band Diagram of Intrinsic Semiconductor

Review:Fermi Level in Intrinsic and Extrinsic Semiconductors

- For an n-type semiconductor, there are more electrons in the conduction band than there are holes in the valence band.
- This also implies that the probability of finding an electron near the conduction band edge is larger than the probability of finding a hole at the valence band edge.

Review:Fermi Level in Intrinsic and Extrinsic Semiconductors

- Therefore, the Fermi level is closer to the conduction band in an n-type semiconductor.
- The Fermi level can be explained in a similar manner for a p-type semiconductor.

Material type	Relationships
n-type	$E_F > E_i$ $E_C - E_F = \frac{E_g}{2} - (E_F - E_i)$
p-type	$E_F < E_i$ $E_F - E_V = \frac{E_g}{2} - (E_i - E_F)$

Energy Band Diagram of Extrinsic Semiconductors

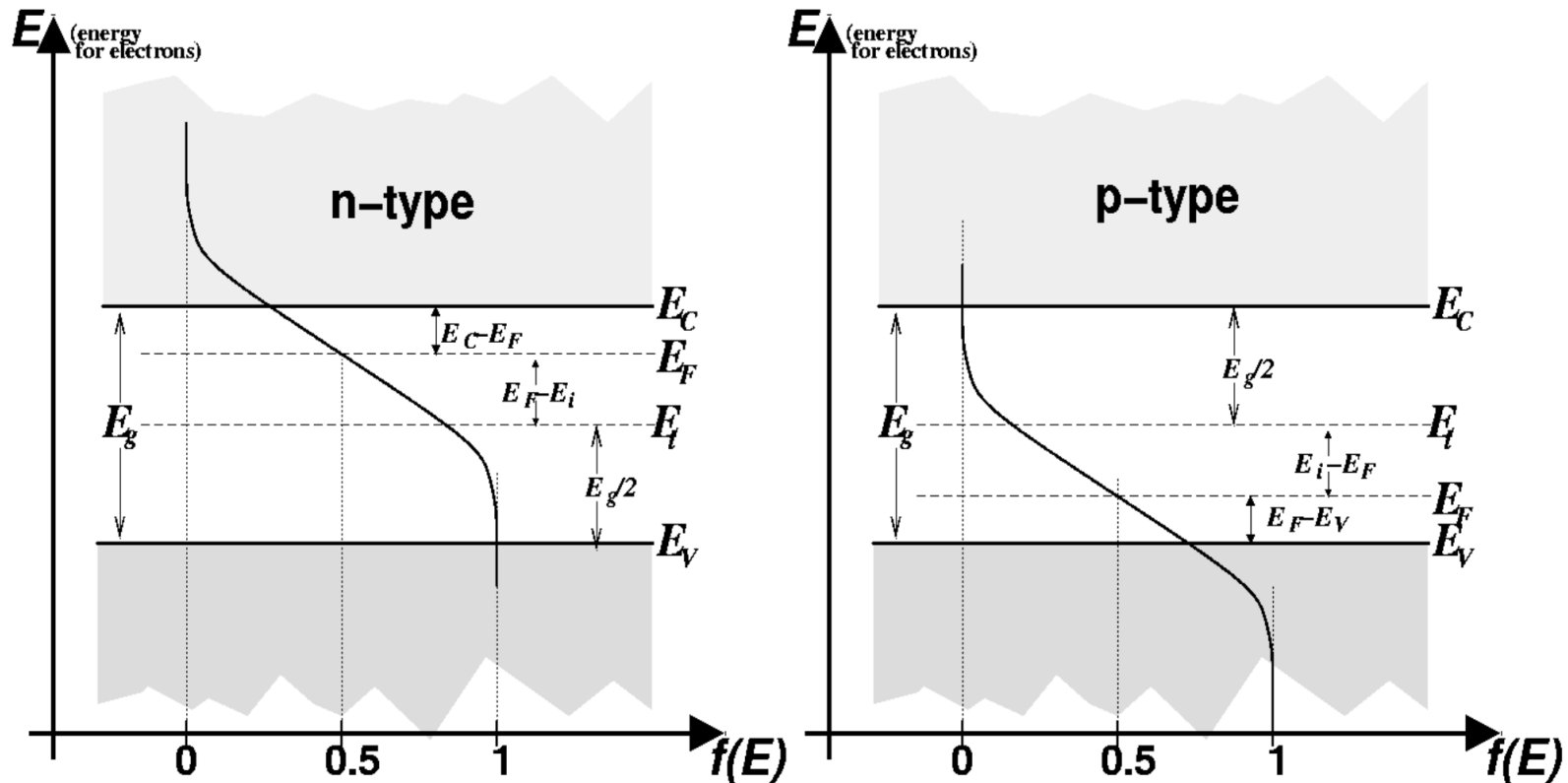


Fig. 4: Energy Band Diagram of Extrinsic Semiconductor

Review: Effective Density of States and the Carrier Concentrations

- The many energy states are there available for electrons in the conduction band, is described by a density-of-states function, $N(E)$.
- The expression $N(E)dE$ gives the number of states in the energy range $[E, E + dE]$.
- To find the total number of electrons in the conduction band, we multiply this density of states with the probability that each energy level will have an electron occupying it ($f(E)$) and integrate across the conduction band.

Review: Effective Density of States and the Carrier Concentrations

- However, assuming that the Fermi level is far away enough from the conduction band edge ($|E_C - E_F| > 4k_B T$), we can consider all the available energy levels in the conduction band to be gathered right at the band edge.
- In doing so, we are defining an effective density of states, shown with N_C for the conduction band and N_V for the valence band.

Review: Effective Density of States and the Carrier Concentrations

- Then, in the conduction band, all we have to do to find the number of electrons per unit volume is to count this number of states at the band edge per unit volume (N_C) and multiply it with the probability that a state at that level will be filled ($f(E_C)$):

$$n_0 = N_C f(E_C) = N_C \frac{1}{1 + \exp((E_C - E_F)/(k_B T))} \dots\dots(19)$$

Review: Effective Density of States and the Carrier Concentrations

- If the assumption $|E_C - E_F| > 4k_B T$ holds, we can also approximate the Fermi-Dirac distribution function with the Boltzmann distribution function:

$$f(E_C) = \frac{1}{1 + \exp((E_C - E_F)/(k_B T))} \approx \exp(-(E_C - E_F)/(k_B T)) \quad \dots(20)$$

Review: Effective Density of States and the Carrier Concentrations

- In an n-type semiconductor, if we know the doping level N_D , we know we can say $n_0 = N_D$; therefore, using the Boltzmann approximation:

$$n_0 = N_D = N_C \exp\left(-\frac{E_C - E_F}{k_B T}\right) \quad \text{.....(21)}$$

- Similarly for a p-type semiconductor

$$p_0 = N_V(1 - f(E_V)) = N_V \exp\left(-\frac{E_F - E_V}{k_B T}\right), \quad \text{.....(22)}$$

Energy Band Diagram of Metal

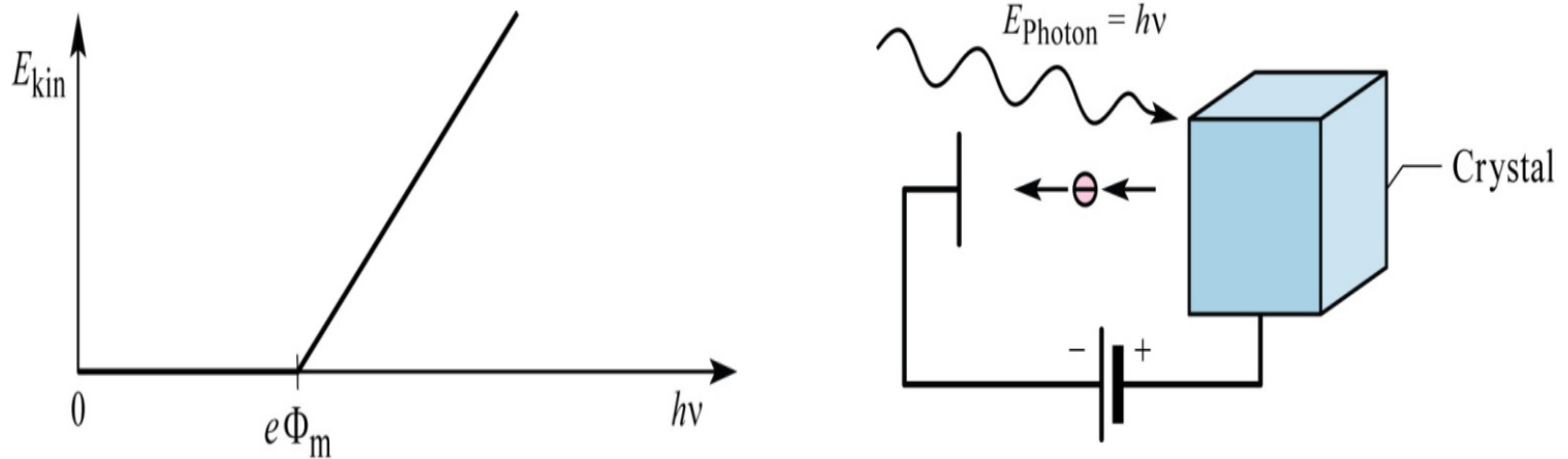


Fig. 5: Experimental finding for metal

Experimental finding: $E_{\text{kin}} = h\nu - e\Phi_m$

$e\Phi_m$ = “threshold energy” required to remove one electron from a solid

Φ_m = *Work function*

dimension $\{\Phi_m\} = \text{V}$

Energy Band Diagram of Metal

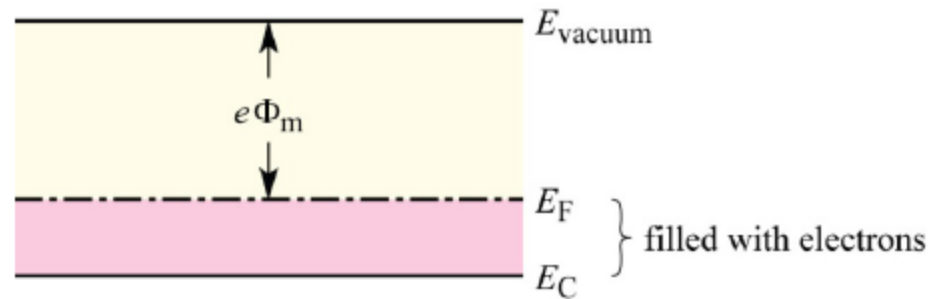


Fig. 6: Band diagram for a metal

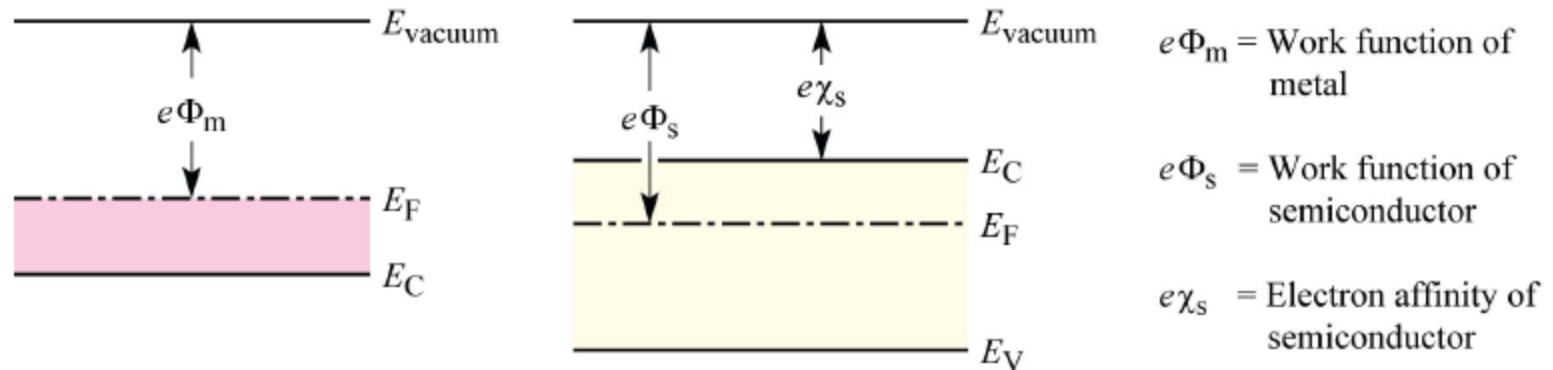


Fig. 7: Band diagram for a metal and semiconductor

Energy Band Diagram

- The energy band diagram of the p-type substrate is shown in Fig. 8.
- The band-gap between the conduction band and the valence band for silicon is approximately 1.1 eV.
- The location of the equilibrium Fermi level E_F within the band-gap is determined by the doping type and the doping concentration in the silicon substrate.

Energy Band Diagram

- The Fermi potential ϕ_F , which is a function of temperature and doping, denotes the difference between the intrinsic Fermi level E_i and the Fermi level E_F .

$$\phi_F = \frac{E_F - E_i}{q} \quad \text{.....(23)}$$

Energy Band Diagram

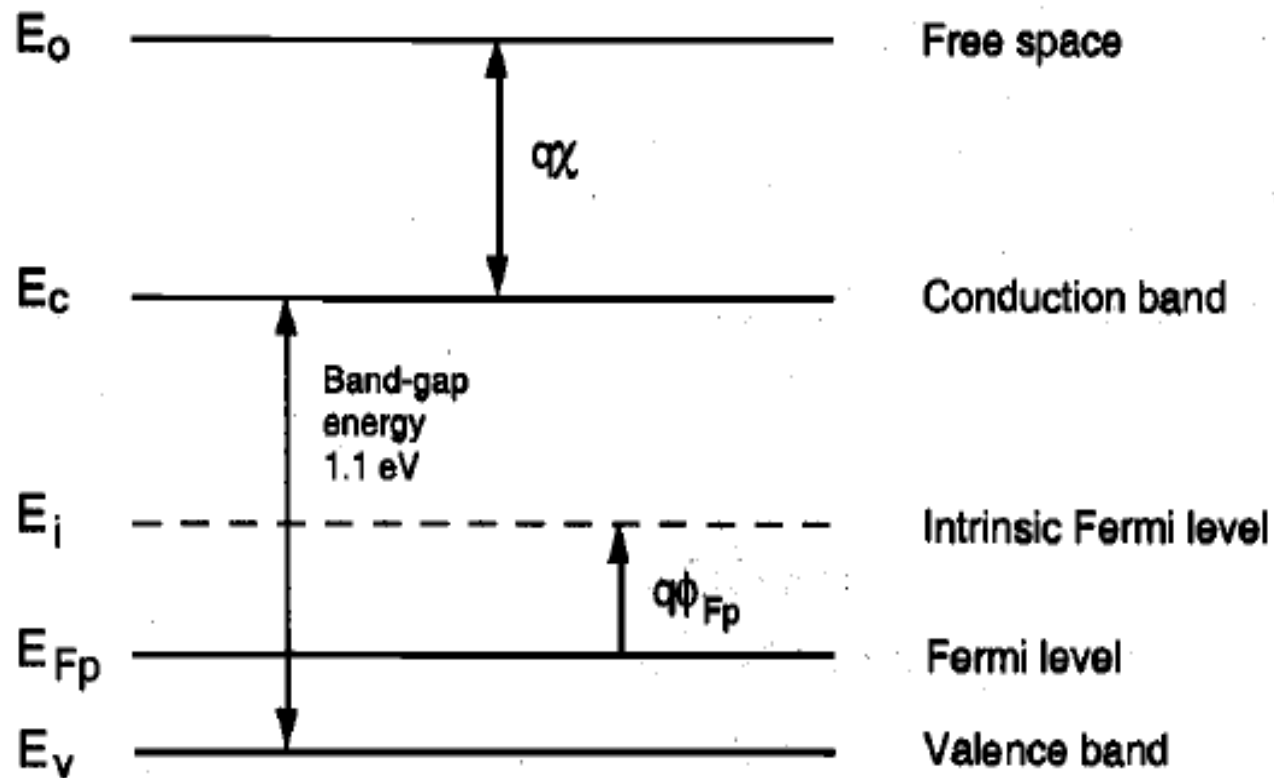


Fig. 8: Energy band diagram of p-type silicon substrate

Energy Band Diagram

- For a p-type semiconductor, the Fermi potential can be approximated by

$$\phi_{Fp} = \frac{kT}{q} \ln \frac{n_i}{N_A} \dots\dots\dots(24)$$

- whereas for an n-type semiconductor (doped with a donor concentration N_D), the Fermi potential is given by

$$\phi_{Fn} = \frac{kT}{q} \ln \frac{N_D}{n_i} \dots\dots\dots(25)$$

Energy Band Diagram

- Where, k denotes the Boltzmann constant and q denotes the unit (electron) charge.
- The Fermi potential result in a positive Fermi potential for n-type material, and a negative Fermi potential for p-type material.
- The electron affinity of silicon, which is the potential difference between the conduction band level and the vacuum (free-space) level, is denoted by $q\chi$.

Energy Band Diagram

- The energy required for an electron to move from the Fermi level into freespace is called the work function $q\Phi_s$, and is given by;

$$q\Phi_s = q\chi + (E_c - E_F) \quad \text{.....(26)}$$

Energy Band Diagram

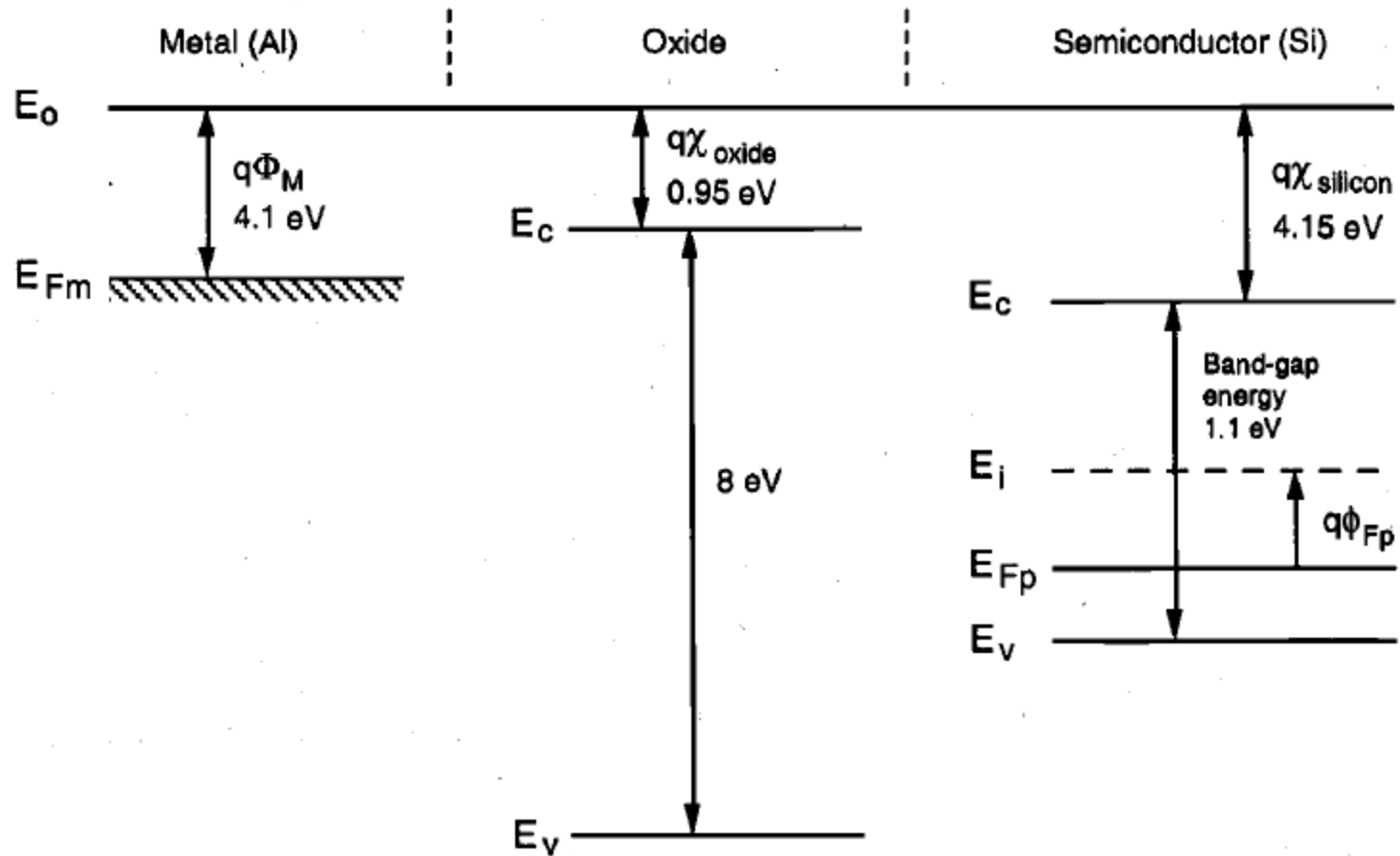


Fig. 9: Energy Band diagram of components that make up the MOS system

Energy Band Diagram

- The insulating silicon dioxide layer between the silicon substrate and the gate has a large band-gap of about 8 eV and an electron affinity of about 0.95 eV.
- On the other hand, the work function $q\Phi_M$ of an aluminum gate is about 4.1 eV.
- Fig. 9 shows the energy band diagrams of metal, oxide, and semiconductor layers in a MOS system as three separate components.

Energy Band Diagram

- Now consider that the three components of the ideal MOS system are brought into physical contact.
- The Fermi levels of all three materials must line up, as they form the MOS capacitor shown in Fig. 1.
- Because of the work-function difference between the metal and the semiconductor, a voltage drop occurs across the MOS system.

Energy Band Diagram

- Part of this built-in voltage drop occurs across the insulating oxide layer.
- The rest of the voltage drop (potential difference) occurs at the silicon surface next to the silicon-oxide interface, forcing the energy bands of silicon to bend in this region.
- The resulting combined energy band diagram of the MOS system is shown in Fig. 10.

Energy Band Diagram

- The equilibrium Fermi levels of the semiconductor (Si) substrate and the metal gate are at the same potential.
- The bulk Fermi level is not significantly affected by the band bending, whereas the surface Fermi level moves closer to the intrinsic Fermi (mid-gap) level.
- The Fermi potential at the surface, also called surface potential ϕ_s , is smaller in magnitude than the bulk Fermi potential ϕ_F .

Energy Band Diagram

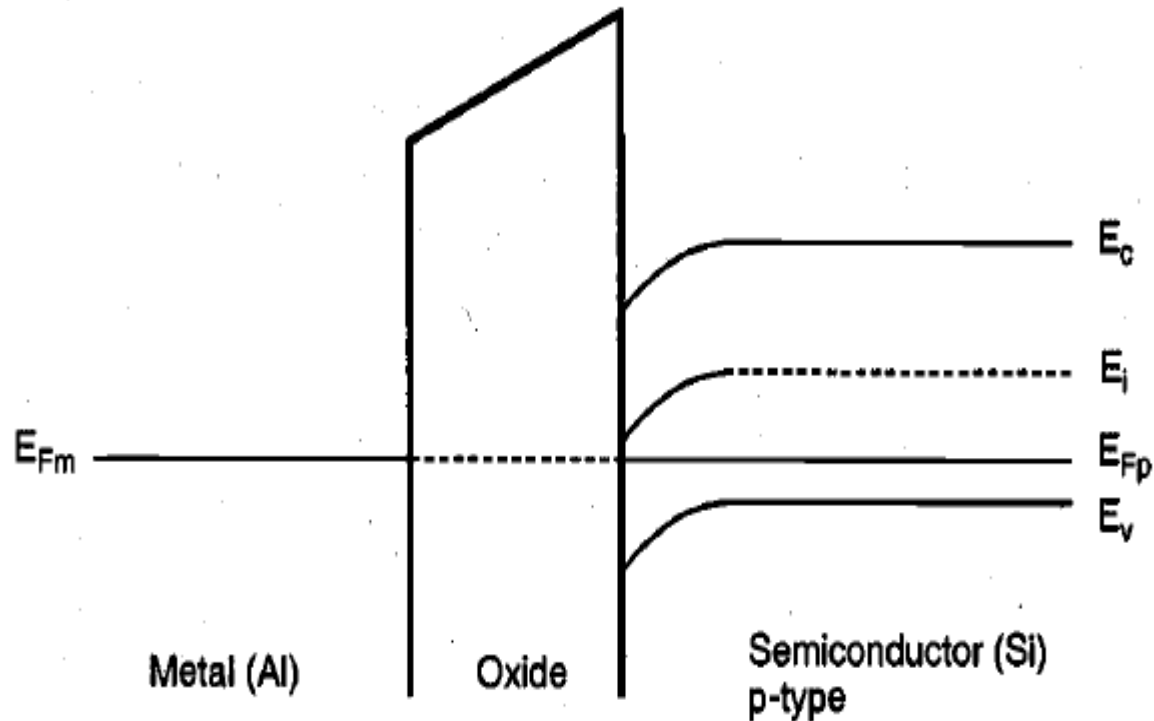


Fig. 10: Energy Band diagram of the combined MOS system

Problem 1

- Example 1

- Consider the MOS structure that consists of a p-type doped silicon substrate, a silicon dioxide layer, and a metal (aluminum) gate. The equilibrium Fermi potential of the doped silicon substrate is given as $q\phi_{Fp} = 0.2$ eV. Using the electron affinity for silicon and the work function for aluminum given in Fig. 9, calculate the built-in potential difference across the MOS system. Assume that the MOS system contains no other charges in the oxide or on the silicon-oxide interface.

MOS System under External Bias

- The electrical behavior of MOS structure under externally applied bias voltages is explained.
- Let us assume that the substrate voltage is set at $V_B=0$ and let the gate voltage V_G be the controlling parameter.
- Depending on V_G , three different operating regions can be observed for the MOS system.

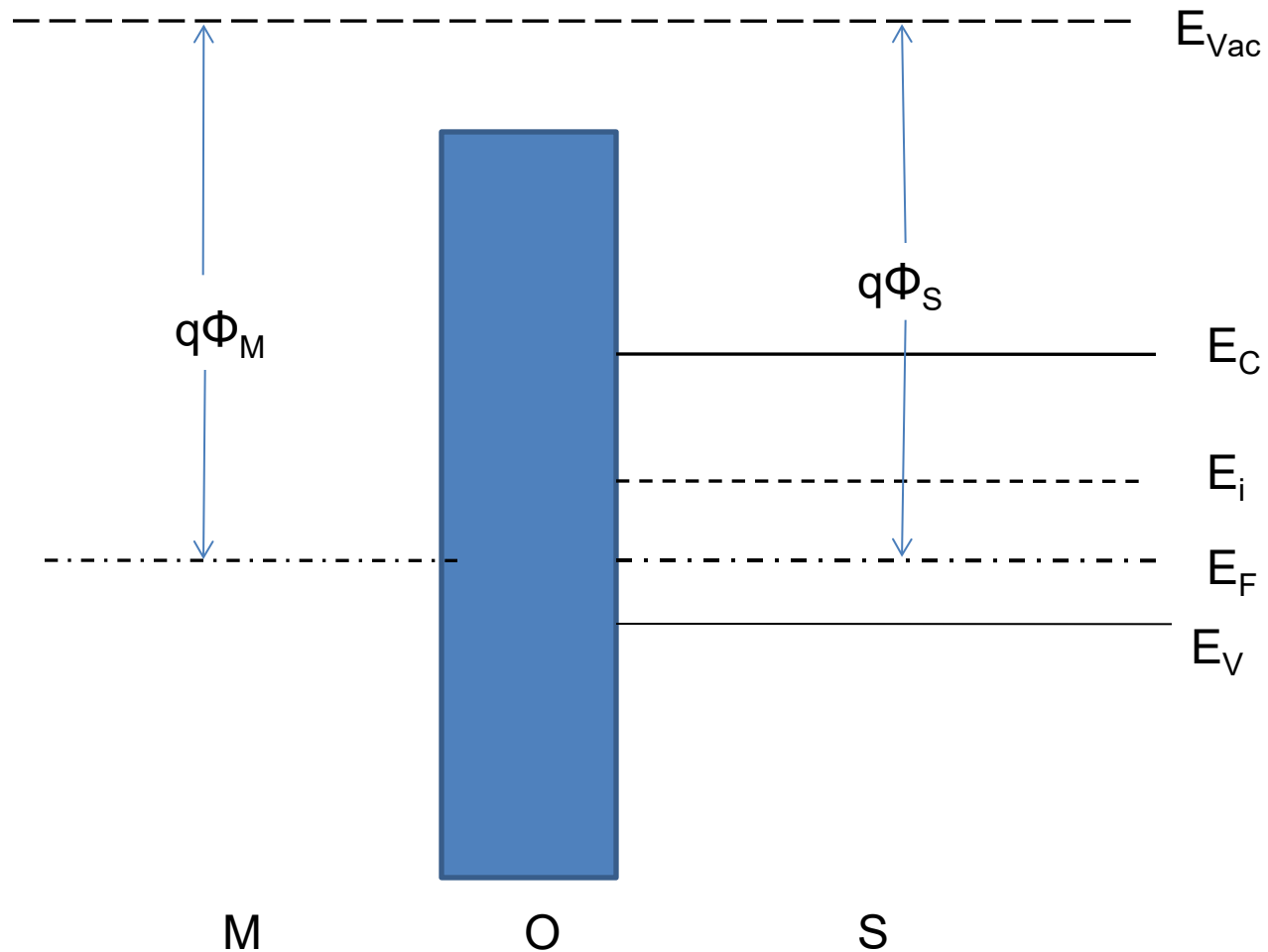
MOS System under External Bias

- Accumulation
- Depletion
- Inversion

The Ideal MOS Capacitor

- MOS capacitor consists of;
- M Metal with work function Φ_M
- O The oxide is SiO_2 ; has a large $E_g > 5\text{eV}$
- S Semiconductor is Si with work function Φ_S
- For ideal case $\Phi_M = \Phi_S$

The Ideal MOS Capacitor



Flat Band Phenomenon

- Here the energy band diagram for a special bias condition called the **flat-band condition** is represented.
- **Flat band is the condition where the energy band (E_c and E_v) of the substrate is flat at the Si–SiO₂ interface as shown in Fig. 11.**
- This condition is achieved by applying a negative voltage to the gate in Fig. 10, thus raising the band diagram on the left-hand side.

Flat Band Phenomenon

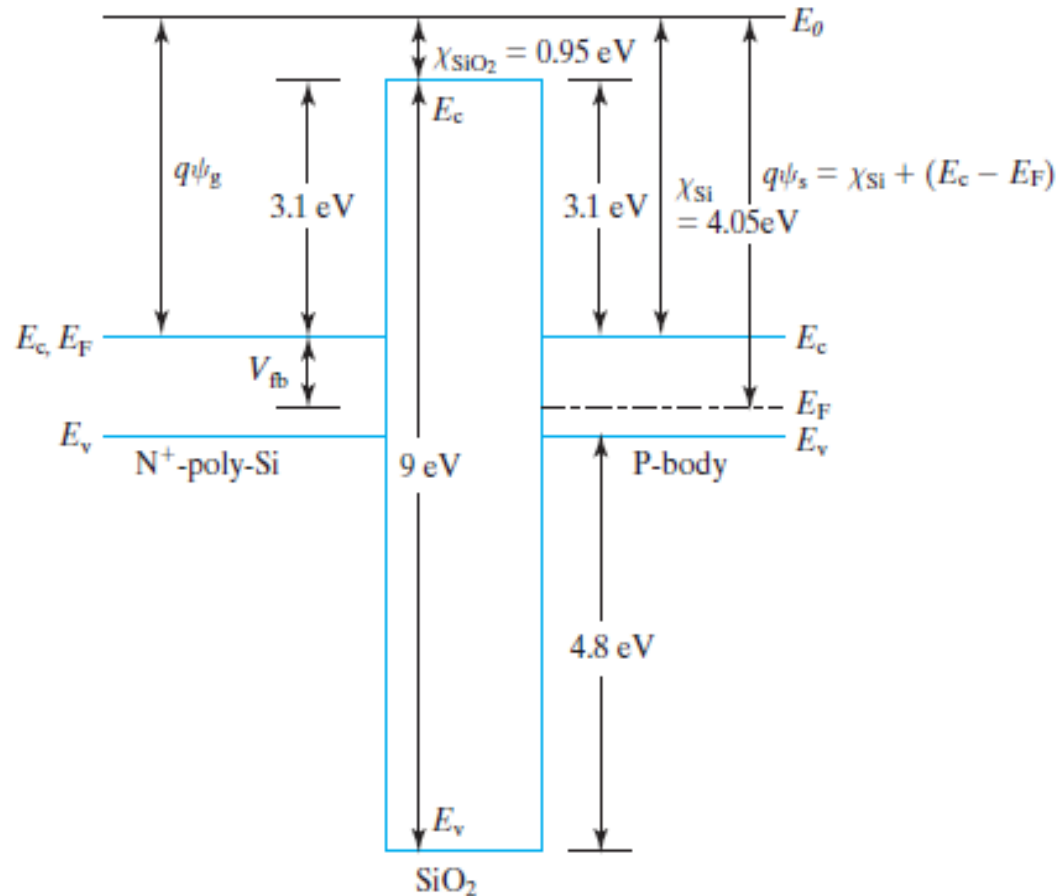


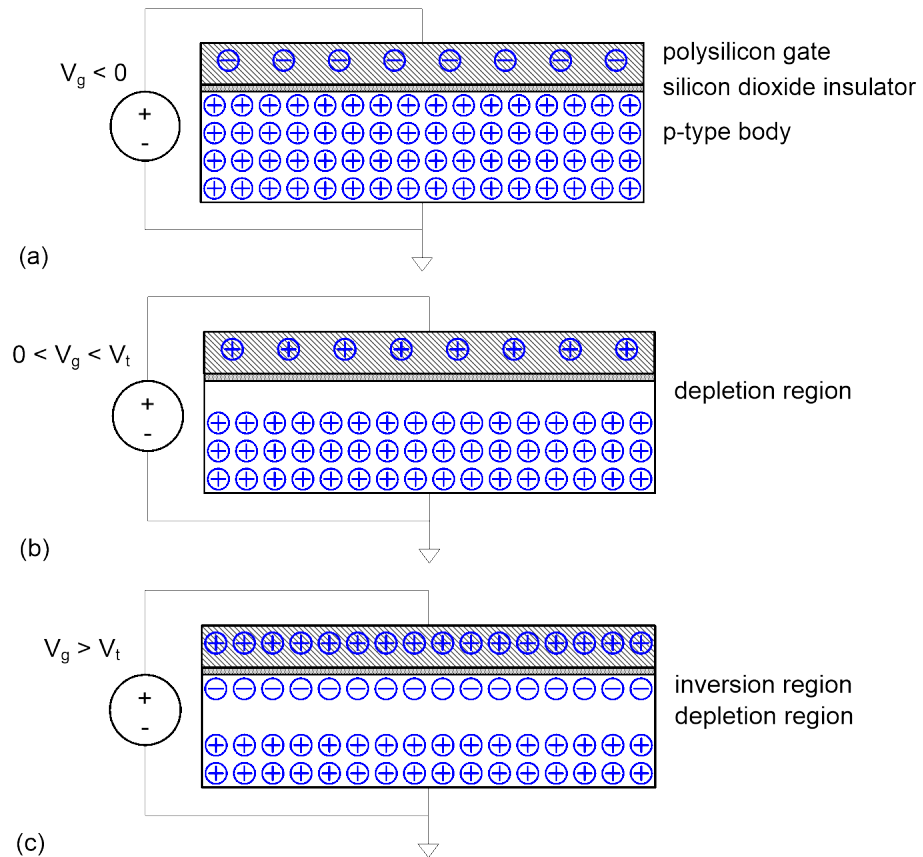
Fig 11: Energy Band Diagram for Flat Band

Flat Band Phenomenon

- When the band is flat in the body as in Fig. 11, the surface electric field in the substrate is zero.
- Therefore the electric field in the oxide is also zero, i.e., E_c and E_v of SiO_2 are flat, too.

MOS Capacitor

- Gate and body form MOS capacitor
- Operating modes
 - Accumulation
 - Depletion
 - Inversion



Accumulation

- If a negative voltage V_G is applied to the gate electrode, the holes in the p-type substrate are attracted to the semiconductor-oxide interface.
- The majority carrier concentration near the surface becomes larger than the equilibrium hole concentration in the substrate.
- Hence, this condition is called carrier accumulation on the surface as shown in Fig. 12.

Accumulation

- In this case, the oxide electric field is directed towards the gate electrode.
- The negative surface potential also causes the energy bands to bend upward near the surface.
- While the hole density near the surface increases as a result of the applied negative gate bias, the electron (minority carrier) concentration decreases as the negatively charged electrons are pushed deeper into the substrate.

Accumulation

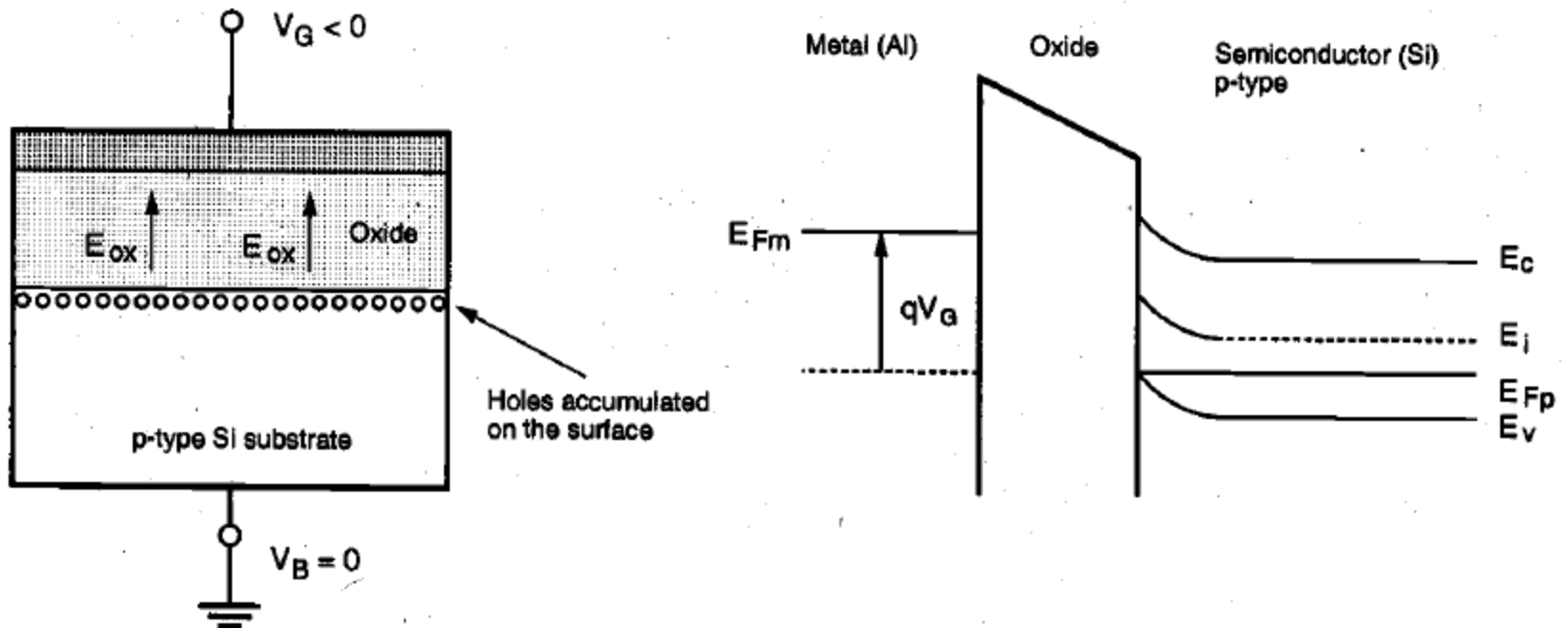


Fig 12 :The cross-sectional view and the energy band diagram of the MOS structure operating in accumulation region.

Depletion

- Now consider the next case in which a small positive gate bias V_G is applied to the gate of a MOS structure whose substrate is grounded produces E penetrating into the substrate.
- Since the substrate bias is zero, the oxide electric field will be directed towards the substrate.
- For a p-type substrate, E repels majority holes from the surface, creating a depletion region as shown in Fig. 13.

Depletion

- Some minority electrons are attracted to the surface, but at low values of V_G their numbers are not sufficient to cause much effect.
- The positive surface potential causes the energy bands to bend downward near the surface, as shown in Fig. 14.

Depletion

- The majority carriers, i.e., the holes in the substrate, will be repelled back into the substrate as a result of the positive gate bias, and these holes will leave negatively charged fixed acceptor ions behind.
- Thus, a depletion region is created near the surface.
- Under this bias condition, the region near the semiconductor-oxide interface is nearly devoid of all mobile carriers.

Depletion

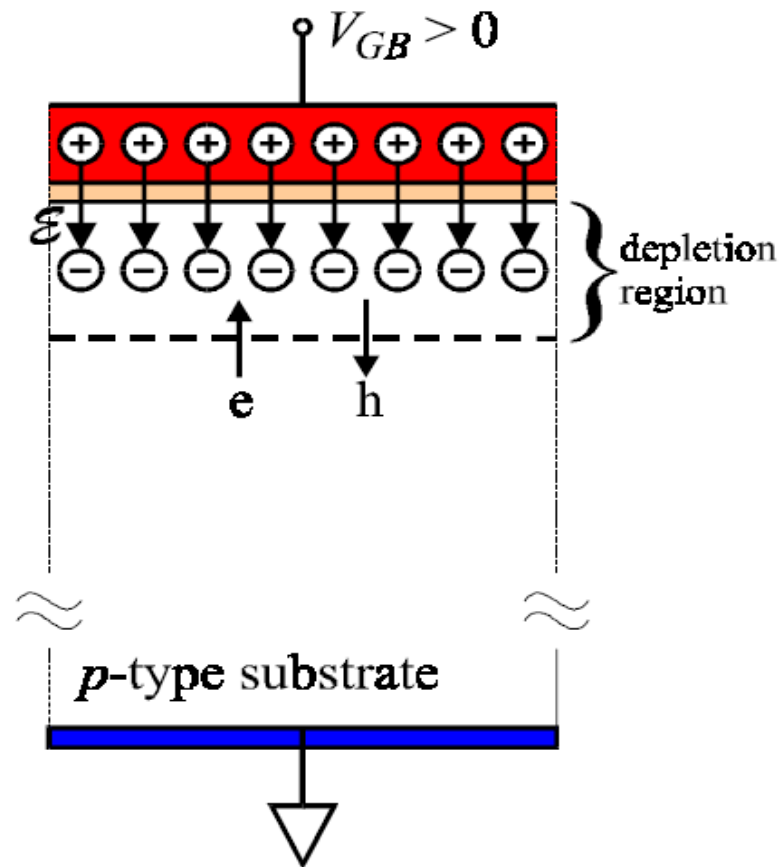


Fig 13: Depletion Region

Depletion

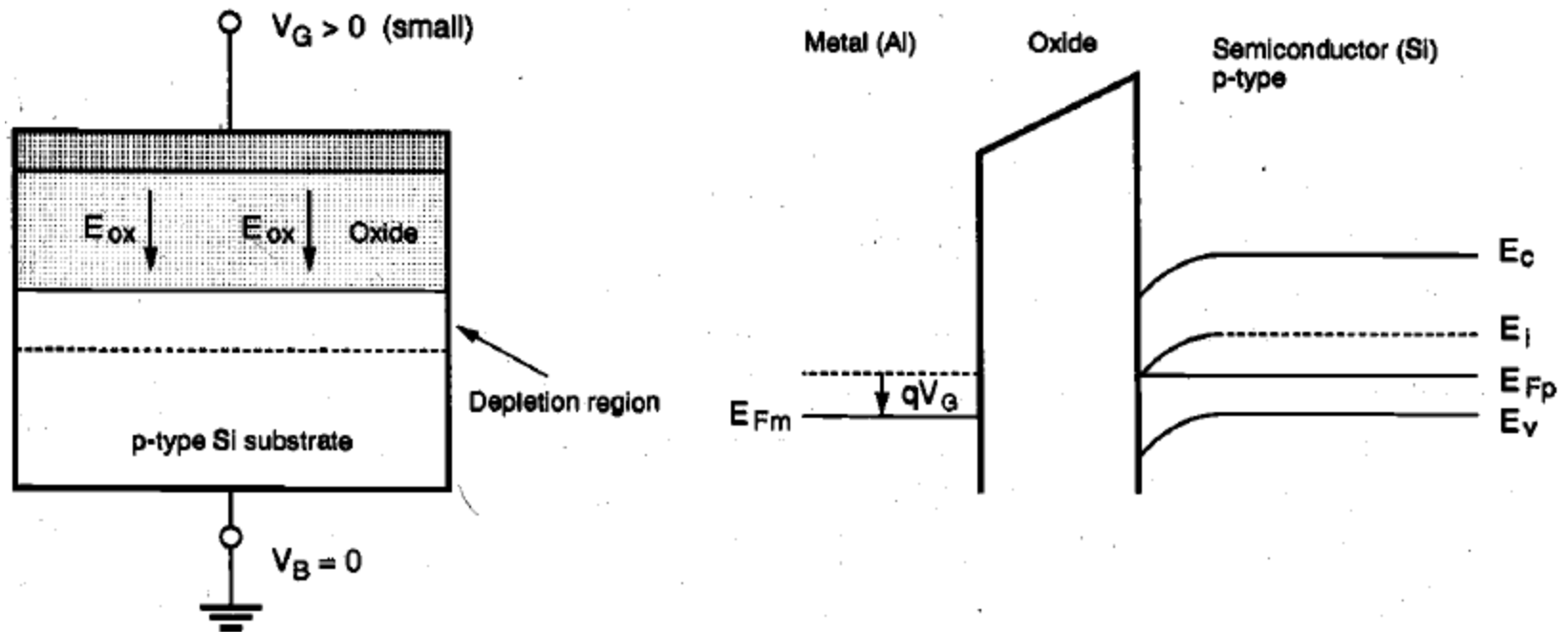


Fig 14: The cross-sectional view and the energy band diagram of the MOS structure operating in depletion mode, under small gate bias.

Depletion

- The thickness x_d of this depletion region on the surface can easily be found as a function of the surface potential ϕ_s .
- Assume that the mobile hole charge in a thin horizontal layer parallel to the surface is;

$$dQ = -q \cdot N_A \cdot dx \quad \text{.....(27)}$$

Where N_A : Acceptor concentration.

Depletion

- The change in surface potential required to displace this charge sheet dQ by a distance x_d away from the surface can be found by using the Poisson equation.

$$d\phi_s = -x \cdot \frac{dQ}{\epsilon_{Si}} = \frac{q \cdot N_A \cdot x}{\epsilon_{Si}} dx \quad \text{.....(28)}$$

Where ϵ_{Si} : Permittivity of the substrate.

Depletion

- Integrating (Eqn. 28) along the vertical dimension (perpendicular to the surface) yields

$$\int_{\phi_F}^{\phi_s} d\phi_s = \int_0^{x_d} \frac{q \cdot N_A \cdot x}{\epsilon_{Si}} dx \quad \text{.....(29)}$$

$$\phi_s - \phi_F = \frac{q \cdot N_A \cdot x_d^2}{2 \epsilon_{Si}} \quad \text{.....(30)}$$

- Thus, the depth of the depletion region is

$$x_d = \sqrt{\frac{2 \epsilon_{Si} \cdot |\phi_s - \phi_F|}{q \cdot N_A}} \quad \text{.....(31)}$$

Depletion

- The depletion region charge density, which consists solely of fixed acceptor ions in this region, is given by the following expression;

$$Q = -q \cdot N_A \cdot x_d = -\sqrt{2q \cdot N_A \cdot \epsilon_{Si} \cdot |\phi_s - \phi_F|} \quad \text{.....(32)}$$

- The amount of this depletion region charge plays a very important role in the analysis of threshold voltage.

Inversion

- To complete the qualitative overview of different bias conditions and their effects upon the MOS system, consider next a further increase in the positive gate bias.
- As a result of the increasing surface potential, the downward bending of the energy bands will increase .

Inversion

- The mid-gap energy level E_i becomes smaller than the Fermi level E_{FP} on the surface, which means that the substrate semiconductor in this region becomes n-type.
- Within this thin layer, the electron density is larger than the majority hole density, since the positive gate potential attracts additional minority carriers(electrons) from the bulk substrate to the surface as shown in Fig. 16.

Inversion

- The n-type region created near the surface by the positive gate bias is called the inversion layer, and this condition is called surface inversion.
- It will be seen that the thin inversion layer on the surface with a large mobile electron concentration can be utilized for conducting current between two terminals of the MOS transistor (Fig. 15).

Inversion

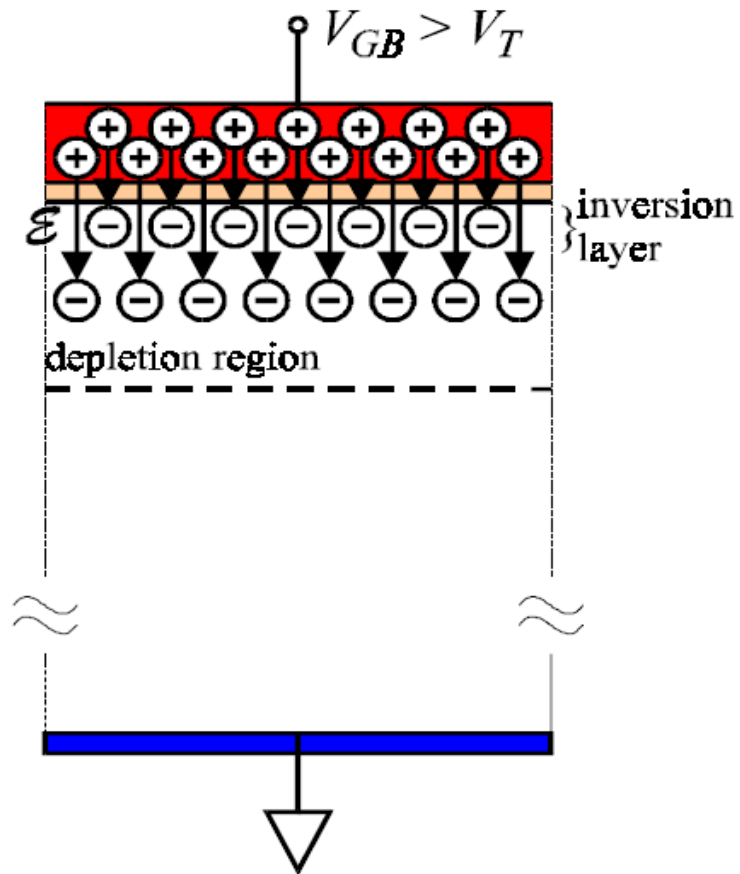


Fig 15: Inversion Layer

Inversion

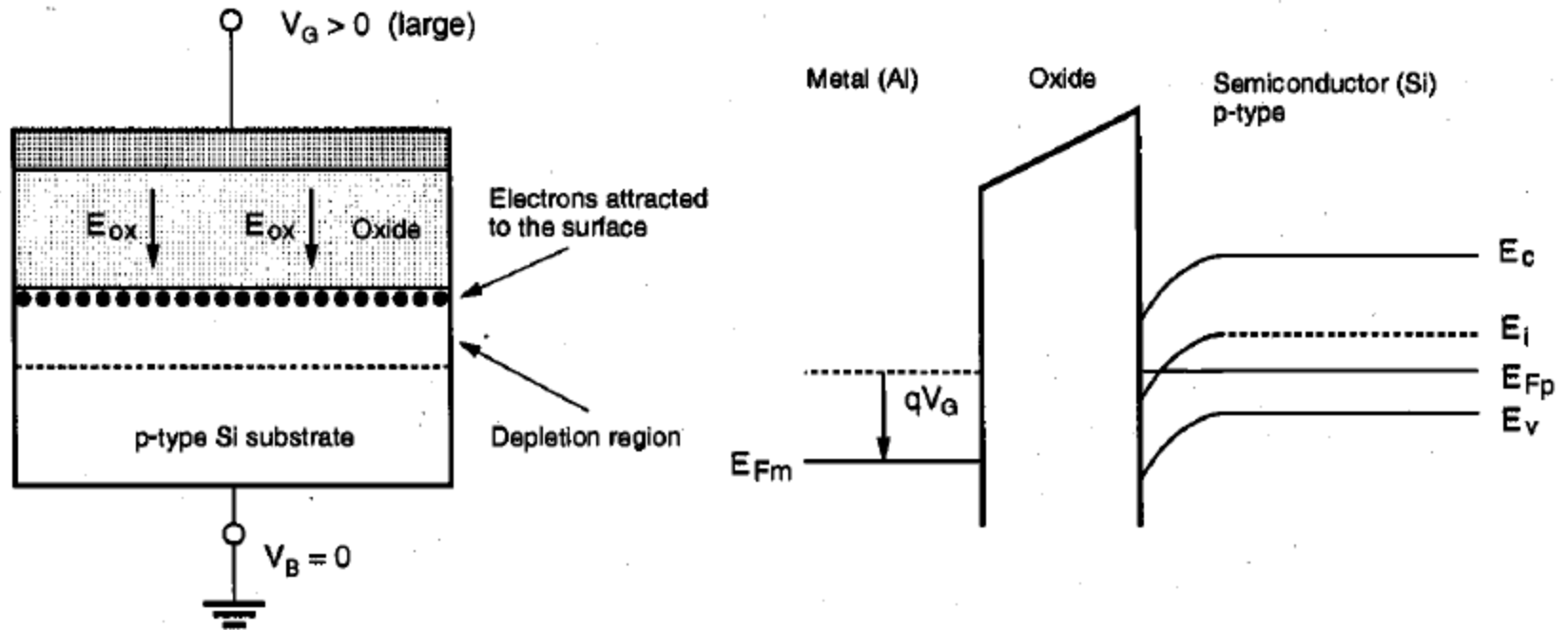


Fig 16: The cross-sectional view and the energy band diagram of the MOS structure in surface inversion, under larger gate bias voltage.

Inversion

- As a practical definition, the surface is said to be inverted when the density of mobile electrons on the surface becomes equal to the density of holes in the bulk (p-type) substrate.
- This condition requires that the surface potential has the same magnitude, but the reverse polarity, as the bulk Fermi potential ϕ_F .

Inversion

- Once the surface is inverted, any further increase in the gate voltage leads to an increase of mobile electron concentration on the surface, but not to an increase of the depletion depth.
- Thus, the depletion region depth achieved at the onset of surface inversion is also equal to the maximum depletion depth, x_{dm} , which remains constant for higher gate voltages.

Inversion

- Using the inversion condition $\phi_s = -\phi_F$, the maximum depletion region depth at the onset of surface inversion can be found from (Eqn. 31) as follows:

$$x_{dm} = \sqrt{\frac{2 \cdot \epsilon_{Si} \cdot |2\phi_F|}{q \cdot N_A}} \quad \text{.....(33)}$$

Inversion

- At large V_G , a dense inversion layer of electrons forms under the surface.
- Further increases in V_G only change the density of the inversion layer.
- The potential at which the inversion layer dominates the substrate behavior is the threshold voltage V_T .
- This inversion layer will form the conductive channel between the source and drain of the MOSFET.

MOS TRANSISTOR

VLSI DESIGN

PART-2

Structure of MOS Transistor

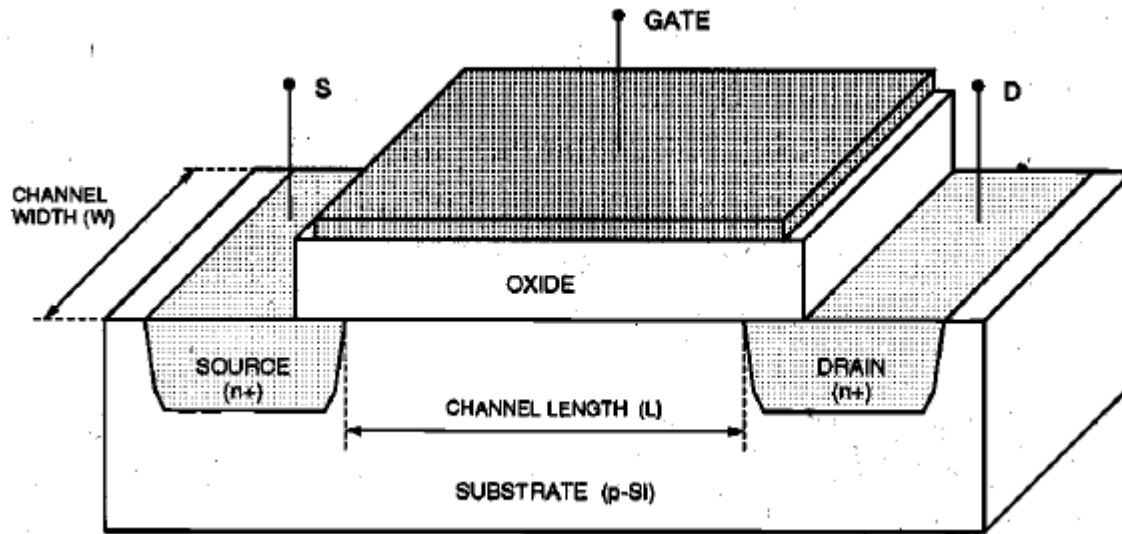


Fig. 16: The physical structure of an n-channel enhancement-type MOSFET

Structure of MOS Transistor

- The basic structure of an n-channel MOSFET is shown in Fig. 16.
- This four-terminal consists of a p-type substrate, in which two n+ diffusion regions, the drain and the source are formed.
- The surface of the substrate region between the drain and the source is covered with a thin oxide layer, and the metal (or polysilicon) gate is deposited on top dielectric.

Structure of MOS Transistor

- The midsection of the device can easily be recognized as the basic MOS structure which was examined in the previous sections.
- The two n⁺ regions will be current conducting terminals of this device.
- The device structure is completely symmetrical with respect to the drain and source regions.

Structure of MOS Transistor

- A conducting channel will eventually be formed through applied gate voltage in the section of the device between the drain and the source diffusion regions.
- The distance between the drain and source diffusion regions is the channel length L , and the lateral extent of the channel (perpendicular to the length dimension) is the channel width W .

Structure of MOS Transistor

- Both the channel length and the channel width are important parameters which can be used to control some of the electrical properties of the MOSFET.
- The thickness of the oxide layer covering the channel region, t_{ox} , is also an important parameter.

Enhancement & Depletion MOSFET

- A MOS transistor which has no conducting channel region at zero gate bias is called an **enhancement-type (or enhancement-mode) MOSFET**.
- If a conducting channel already exists at zero gate bias, on the other hand, the device is called a **depletion-type (or depletion-mode) MOSFET**.

Abbreviations for MOSFET

- G:Gate, D:Drain, S:Source, B:Substrate(Body).
- In an n-channel MOSFET source is defined as the n^+ region which has a lower potential than the other n^+ region, the drain.
- V_{GS} : Gate-to-Source Voltage.
- V_{DS} : Drain-to-Source Voltage.
- V_{BS} : Substrate-to-Source Voltage.

Principle of n-channel e-MOSFET

- The simple operation principle of this device is: control the current conduction between the source and the drain, using the electric field generated by the gate voltage as a control variable.
- Since the current flow in the channel is also controlled by the drain-to-source voltage and by the substrate voltage, the current can be considered a function of these external terminal voltages.

Bias Condition for nMOS(Enhancement Type)

- The simplest bias condition that can be applied to the n-channel enhancement-type MOSFET is shown in Fig. 17.
- The source, the drain, and the substrate terminals are all connected to ground.
- A positive gate-to-source voltage V_{GS} is then applied to the gate in order to create the conducting channel underneath the gate.

Bias Condition for nMOS(Enhancement Type)

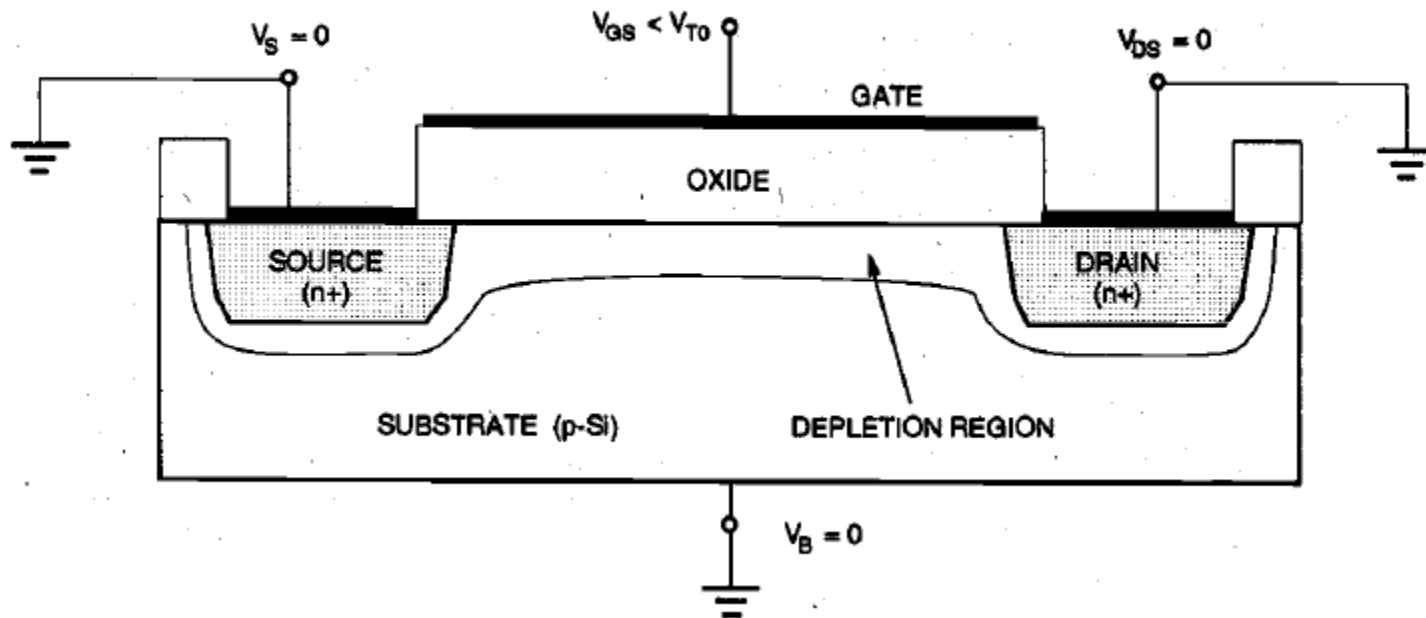


Fig. 17: Formation of a depletion region in an n-channel enhancement-type MOSFET

Bias Condition for nMOS(Enhancement Type)

- With this bias arrangement, the channel region between the source and the drain diffusions behaves exactly the same as for the simple MOS structure already explained in the MOS Section under external bias.
- For small gate voltage levels, the majority carriers (holes) are repelled back into the substrate, and the surface of the p-type substrate is depleted.

Bias Condition for nMOS(Enhancement Type)

- Since the surface is devoid of any mobile carriers, current conduction between the source and the drain is not possible.
- Now the gate-to-source voltage is further increased. As soon as the surface potential in the channel region reaches $-\phi_{Fp}$, surface inversion will be established, and a conducting n-type layer will form between the source and the drain diffusion regions (Fig. 18).

Bias Condition for nMOS(Enhancement Type)

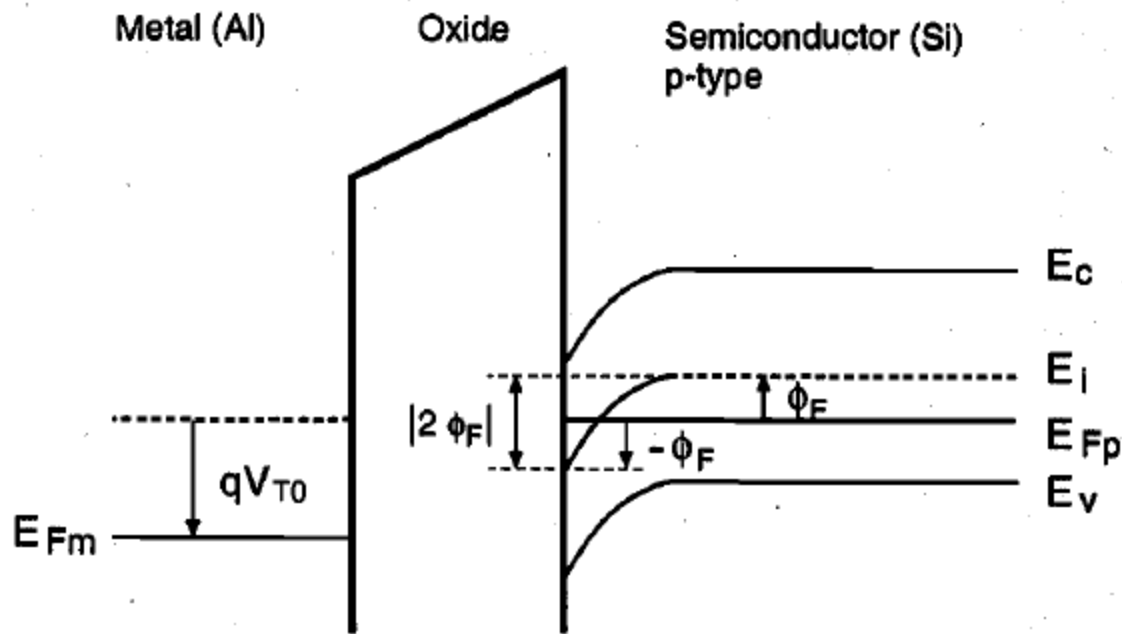


Fig 18: Band diagram of the MOS structure underneath the gate, at surface inversion.
Notice the band bending by $|2\phi_F|$ at the surface.

Bias Condition for nMOS(Enhancement Type)

- This channel now provides an electrical connection between the two n+ regions, and it allows current flow, as long as there is a potential difference between the source and the drain terminal voltages (Fig. 19).

Bias Condition for nMOS(Enhancement Type)

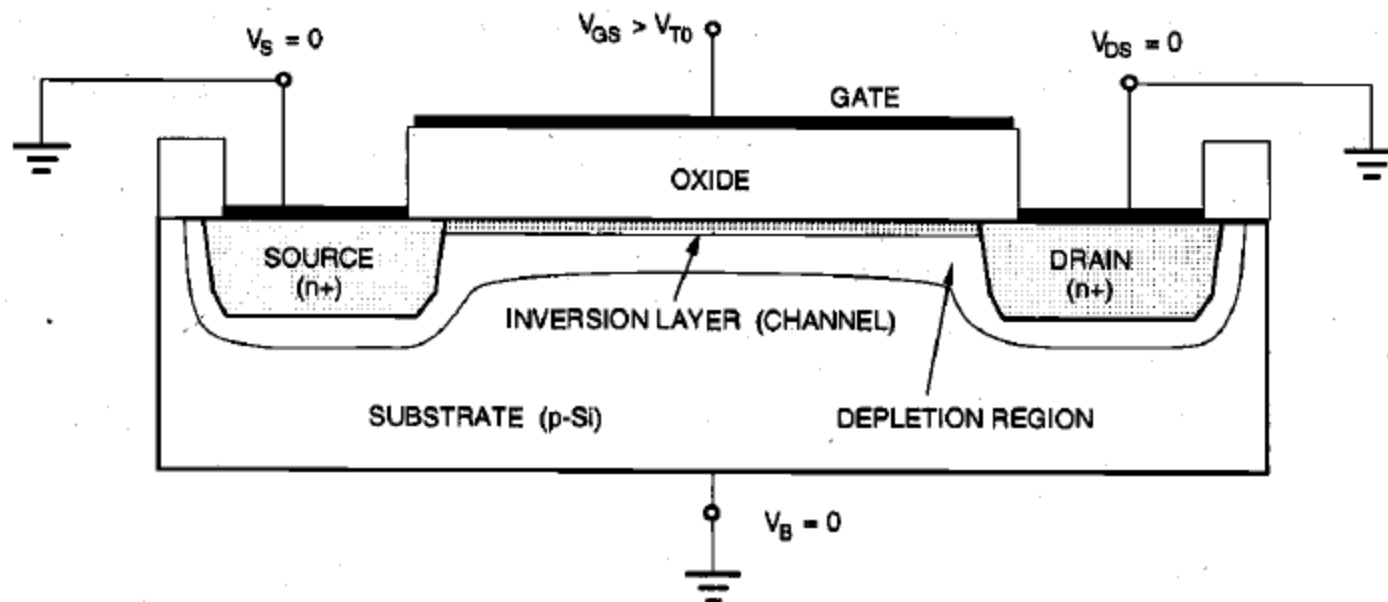


Fig. 19: Formation of an inversion layer (channel) in an n-channel enhancement-type MOSFET

Bias Condition for nMOS(Enhancement Type)

- The value of the gate-to-source voltage V_{GS} needed to cause surface inversion (to create the conducting channel) is called the threshold voltage V_{T0} .
- Any gate-to-source voltage smaller than V_{T0} is not sufficient to establish an inversion layer.
- Thus current conduction condition; $V_{GS} > V_{T0}$.

Bias Condition for nMOS(Enhancement Type)

- Also it has been observed that increasing the gate-to-source voltage above and beyond the threshold voltage will not affect the surface potential and the depletion region depth.
- Both quantities will remain approximately constant and equal to their values attained at the onset of surface inversion.

The Threshold Voltage

- The physical parameters affecting the threshold voltage of a MOS structure will be examined by considering the various components of V_{TO} .
- For all practical purposes, four physical components of the threshold voltage are identified:

The Threshold Voltage

- (i) the work function difference between the gate and the channel.
 - (ii) the gate voltage component to change the surface potential.
 - (iii) the gate voltage component to offset the depletion region charge.
- and
- (iv) the voltage component to offset the fixed charges in the gate oxide and in the silicon-oxide interface.

The Threshold Voltage

- (i) The work function difference Φ_{GC} between the gate and the channel reflects the built-in potential of the MOS system, which consists of the p-type substrate, the thin silicon dioxide layer, and the gate electrode.
- Depending on the gate material, the work function difference is:

$$\Phi_{GC} = \phi_F(\text{substrate}) - \phi_M \quad \text{for metal gate} \quad \dots(34)$$

$$\Phi_{GC} = \phi_F(\text{substrate}) - \phi_F(\text{gate}) \quad \text{for polysilicon gate} \quad \dots(35)$$

The Threshold Voltage

- This first component of the threshold voltage accounts for part of the voltage drop across the MOS system that is built-in.
- (ii) Now, the externally applied gate voltage must be changed to achieve surface inversion, i.e., to change the surface potential by $-2\phi_F$.
- This will be the second component of the threshold voltage.

The Threshold Voltage

- (iii) Another component of the applied gate voltage is necessary to offset the depletion region charge, which is due to the fixed acceptor ions located in the depletion region near the surface.
- The depletion region charge density at surface inversion ($\phi_s = -\phi_F$) using (Eqn. 32),

The Threshold Voltage

$$Q_{B0} = -\sqrt{2q \cdot N_A \cdot \epsilon_{Si} \cdot |-2\phi_F|} \quad \text{.....(36)}$$

- The substrate (body) is biased at a different voltage level than the source, which is at ground potential (reference).
- Thus the depletion region charge density can be expressed as a function of the source-to-substrate voltage V_{SB} as in (Eqn. 37).

The Threshold Voltage

$$Q_B = -\sqrt{2q \cdot N_A \cdot \epsilon_{Si} \cdot |-2\phi_F + V_{SB}|} \quad \text{.....(37)}$$

- The component that offsets the depletion region charge is then equal to $-Q_B/C_{ox}$,
- Where C_{ox} is the gate oxide capacitance per unit area represented by (Eqn. 38).

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} \quad \text{.....(38)}$$

The Threshold Voltage

- (iv) There always exists a fixed positive charge density Q_{ox} at the interface between the gate oxide and the silicon substrate, due to impurities and/or lattice imperfections at the interface.
- The gate voltage component that is necessary to offset this positive charge at the interface is $-Q_{ox}/C_{ox}$.

The Threshold Voltage

- Now, we can combine all of these voltage components to find the threshold voltage.
- For zero substrate bias, the threshold voltage V_{T0} is expressed as follows:

$$V_{T0} = \Phi_{GC} - 2\phi_F - \frac{Q_{B0}}{C_{ox}} - \frac{Q_{ox}}{C_{ox}} \quad \text{.....(39)}$$

The Threshold Voltage

- For nonzero substrate bias, on the other hand, the depletion charge density term must be modified to reflect the influence of V_{SB} upon that charge, resulting in the following **generalized threshold voltage expression**.

$$V_T = \Phi_{GC} - 2\phi_F - \frac{Q_B}{C_{ox}} - \frac{Q_{ox}}{C_{ox}} \quad \text{.....(40)}$$

The Threshold Voltage

- The (Eqn. 40) can be rewritten as;

$$V_T = \Phi_{GC} - 2\phi_F - \frac{Q_{B0}}{C_{ox}} - \frac{Q_{ox}}{C_{ox}} - \frac{Q_B - Q_{B0}}{C_{ox}} = V_{T0} - \frac{Q_B - Q_{B0}}{C_{ox}} \quad \text{.....(41)}$$

- The threshold voltage differs from V_{T0} by an additive term.
- This substrate-bias term is a simple function of the material constants and of the source- to-substrate voltage V_{SB} .

The Threshold Voltage

$$\frac{Q_B - Q_{B0}}{C_{ox}} = -\frac{\sqrt{2q \cdot N_A \cdot \epsilon_{Si}}}{C_{ox}} \cdot \left(\sqrt{|-2\phi_F + V_{SB}|} - \sqrt{|2\phi_F|} \right) \quad \dots\dots(42)$$

- Thus, the most general expression of the threshold voltage V_{T0} can be found as follows:

$$V_T = V_{T0} + \gamma \cdot \left(\sqrt{|-2\phi_F + V_{SB}|} - \sqrt{|2\phi_F|} \right) \quad \dots\dots(43)$$

- where the parameter γ is the substrate-bias (or body-effect) coefficient expressed in (Eqn. 44).

The Threshold Voltage

$$\gamma = \frac{\sqrt{2q \cdot N_A \cdot \epsilon_{Si}}}{C_{ox}} \quad \text{.....(44)}$$

- The threshold voltage expression given in (Eqn. 43) can be used both for n-channel and p-channel MOS transistors.
- One must be careful, however, since some of the terms and coefficients in this equation have different polarities for the n-channel (nMOS) case and for the p-channel (pMOS) case.

The Threshold Voltage

- The reason for this polarity difference is that the substrate semiconductor is p-type in an n-channel MOSFET and n-type in a p-channel MOSFET.
- The substrate Fermi potential ϕ_F is negative in nMOS, positive in pMOS.
- The depletion region charge densities Q_{BO} and Q_B are negative in nMOS, positive in pMOS.

The Threshold Voltage

- The substrate bias coefficient γ is positive in nMOS, negative in pMOS.
- The substrate bias voltage V_{SB} is positive in nMOS, negative in pMOS.
- Typically, the threshold voltage of an enhancement-type n-channel MOSFET is a positive quantity, whereas the threshold voltage of a p-channel MOSFET is negative.

Physical and Materials Constants

Boltzmann's constant	k	1.38×10^{-23}	J/K
Electron charge	q	1.6×10^{-19}	C
Thermal voltage	kT/q	0.026 (at T= 300 K)	V
Energy gap of silicon (Si)	E_g	1.12 (at T = 300 K)	eV

Physical and Materials Constants

Intrinsic carrier concentration of silicon (Si)	n_i	1.45×10^{10} (at $T = 300\text{ K}$)	cm^{-3}
Dielectric constant of vacuum	ϵ_0	8.85×10^{-14}	F/cm
Dielectric constant of silicon (Si)	ϵ_{Si}	$11.7 \times \epsilon_0$	F/cm
Dielectric constant of silicon dioxide (SiO ₂)	ϵ_{ox}	$3.97 \times \epsilon_0$	F/cm

Commonly Used Prefixes for Units

giga	G	10^9
mega	M	10^6
kilo	k	10^3
milli	m	10^{-3}
micro	μ	10^{-6}
nano	n	10^{-9}
pico	p	10^{-12}
femto	f	10^{-15}

Problem 2

- Calculate the threshold voltage V_{TO} at $V_{SB} = 0$, for a polysilicon gate n-channel MOS transistor, with the following parameters: substrate doping density $N_A = 10^{16} \text{ cm}^{-3}$, polysilicon gate doping density $N_D = 2 \times 10^{20} \text{ cm}^{-3}$, gate oxide thickness $t_{ox} = 500 \text{ \AA}$, and oxide-interface fixed charge density $N_{ox} = 4 \times 10^{10} \text{ cm}^{-2}$.

Assignment 2

- Question 1
- Consider the following p-channel MOSFET process:
- Substrate doping $N_D = 10^{15} \text{ cm}^{-3}$, polysilicon gate doping density $N_D = 10^{20} \text{ cm}^{-3}$, gate oxide thickness $t_{\text{ox}} = 650 \text{ \AA}$,
- and oxide-interface charge density $N_{\text{ox}} = 2 \times 10^{10} \text{ cm}^{-2}$.

Assignment 2

- Use $\epsilon_{si} = 11.7\epsilon_0$ and $\epsilon_{ox} = 3.97\epsilon_0$ for the dielectric coefficients of silicon and silicon-dioxide, respectively.
- (a) Calculate the threshold voltage V_{T0} for $V_{SB} = 0$.
- (b) Determine the type and the amount of channel ion implantation which are necessary to achieve a threshold voltage of $V_{T0} = -2V$.

Problem 3

- Consider the n-channel MOSFET process given in Problem 2. In several digital circuit applications, the condition $V_{SB} = 0$ cannot be guaranteed for all transistors.
- We will examine in this example how a nonzero source-to-substrate voltage V_{SB} affects the threshold voltage of the MOS transistor.

Problem 4

- Find x_{dm} for ideal MOS capacitors with $N_A = 10^{16} \text{ cm}^{-3}$ and $N_A = 10^{18} \text{ cm}^{-3}$. Assume the following values: $n_i = 10^{10} \text{ cm}^{-3}$, $\epsilon_r = 9$.

Electric Fields in the MOSFET

- Two distinct electric field distributions exist in the MOSFET structure
 - The transverse field is caused by the potential difference between the conductive gate and the substrate. This field supports the substrate depletion region and inversion layer.
 - The lateral field arises due to a non-zero source to drain potential, and is (in the simple model) the main mechanism for current flow in the MOSFET (Fig. 20).

Electric Fields in the MOSFET

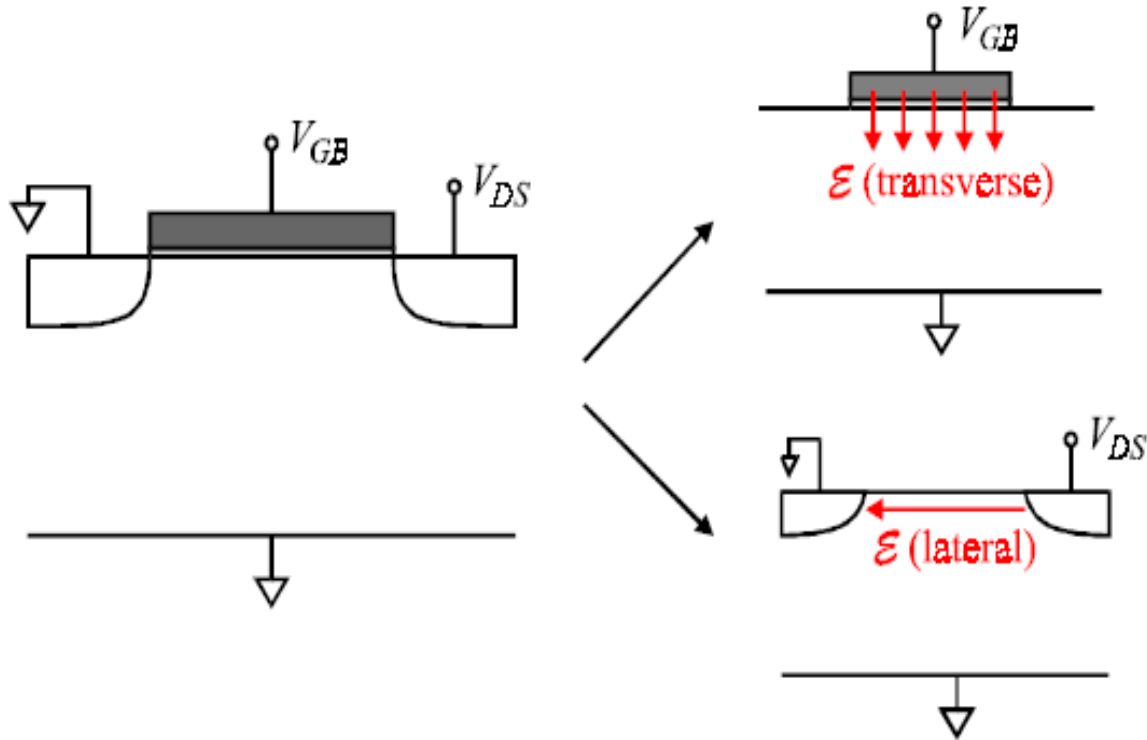


Fig. 20: Electric field in MOSFET

n-Channel MOSFET With $V_{GS} < V_T$

- With $V_{GS} < V_T$, there is no inversion layer present under the surface;
- At $V_{DS} = 0$, the source and drain depletion regions are symmetrical.
- A positive V_{DS} reverse biases the drain substrate junction, hence the depletion region around the drain widens, and since the drain is adjacent to the gate edge, the depletion region widens in the channel.

n-Channel MOSFET With $V_{GS} < V_T$

- No current flows even for $V_{DS} > 0$, since there is no conductive channel between the source and drain for $V_{GS} < V_T$.

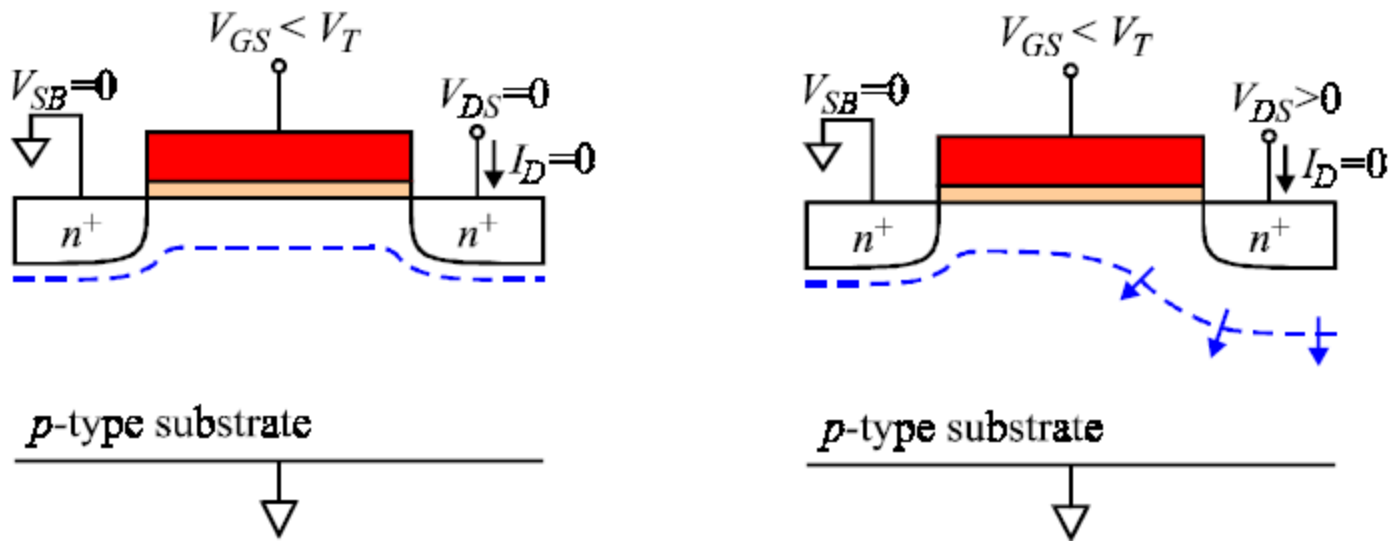


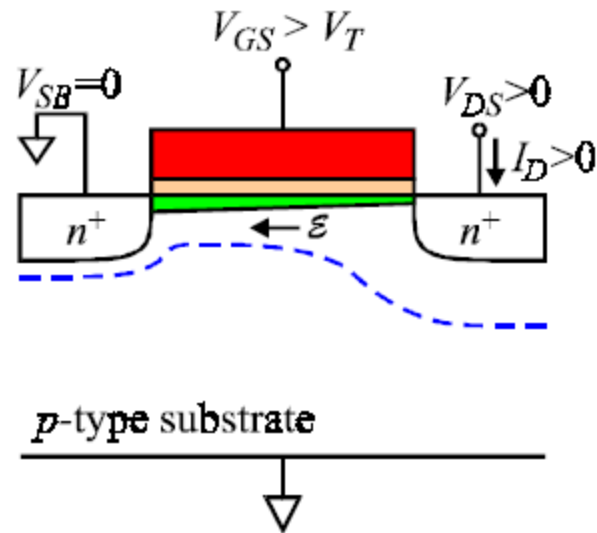
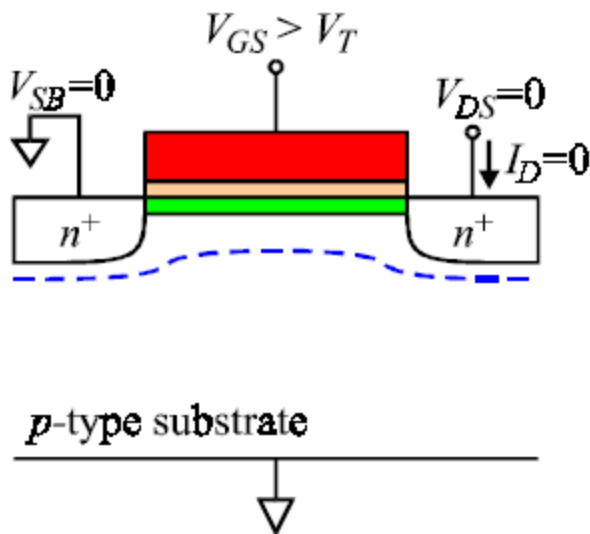
Fig . 21 :The Depletion Region

n-Channel MOSFET With $V_{GS} > V_T$ small V_{DS}

- With $V_{GS} > V_T$, a conductive channel forms under the surface - a nonzero transverse field is present.
- I_D is zero for $V_{DS} = 0$ since no lateral field is present.
- For $V_{DS} > 0$, transverse E is present and current flows.

n-Channel MOSFET With $V_{GS} > V_T$ small V_{DS}

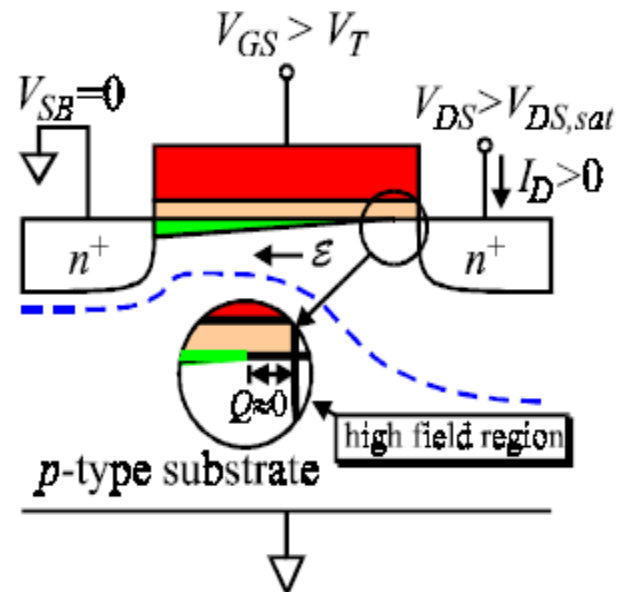
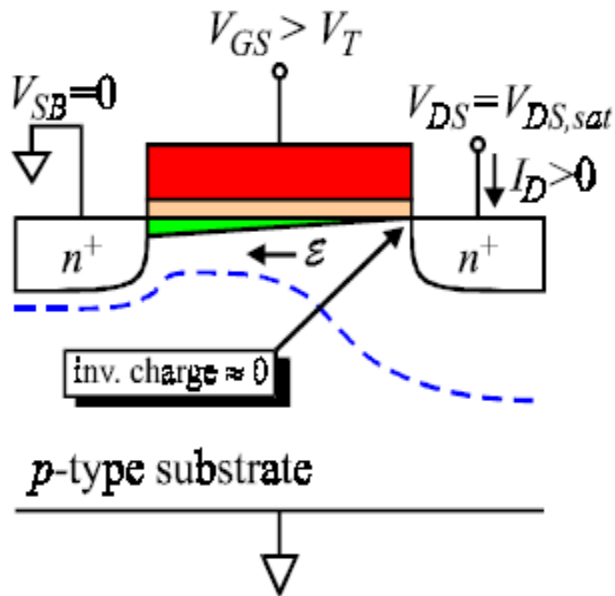
- The increased reverse bias on the drain substrate junction in contact with the inversion layer causes inversion layer density to decrease.



n-Channel MOSFET With $V_{GS} > V_T$ large V_{DS}

- The point at which the inversion layer density becomes very small (essentially zero) at the drain end is termed pinch-off.
- The value of V_{DS} at pinchoff is denoted $V_{DS,sat}$.
- Past pinchoff, further increases in lateral electric field are absorbed by the creation of a narrow high field region with low carrier density.
- ($J_n = qn\mu_n E$, so if n is small E is large)

n-Channel MOSFET With $V_{GS} > V_T$ large V_{DS}



MOSFET Regions of Operation

- There are three regions of operation in the MOSFET:
 - When $V_{GS} < V_T$, no conductive channel is present and $I_D = 0$, the **cutoff region**.
 - If $V_{GS} > V_T$ and $V_{DS} < V_{DS,sat}$, the device is in the **triode region or linear region** of operation.
- Increasing V_{DS} increases the lateral field in the channel, and hence the current.

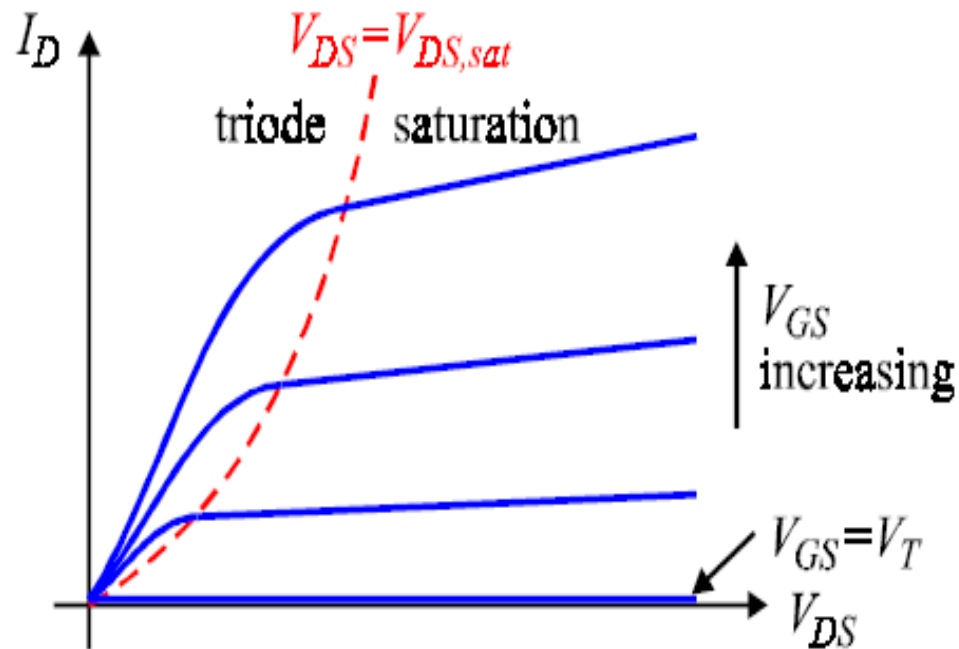
MOSFET Regions of Operation

- Increasing V_{GS} increases the transverse field and hence the inversion layer density, which also increases the current.
- If $V_{GS} > V_T$ and $V_{DS} > V_{DS,sat}$, the device is in the **saturation region** of operation.
- Since the drain end channel density has become small, the current is much less dependent on V_{DS} , but is still dependent on V_{GS} , since increased V_{GS} still increases the inversion layer density.

MOSFET I_D - V_{DS} Characteristic

- For $V_{GS} < V_T$, $I_D = 0$
- As V_{DS} increases at a fixed V_{GS} , I_D increases in the triode region due to the increased lateral field, but at a decreasing rate since the inversion layer density is decreasing.
- Once pinchoff is reached, further V_{DS} increases only increase I_D due to the formation of the high field region.
- The device starts in triode, and moves into saturation at higher V_{DS} .

MOSFET I_D - V_{DS} Characteristic



Gradual Channel Approximation

- The current-flow analysis, can be explained by considering the cross-sectional view of the n channel MOSFET operating in the linear mode, as shown in Fig. 22.
- Here, the source and the substrate terminals are connected to ground, i.e., $V_S = V_B = 0$.
- The gate-to-source voltage (V_{GS}) and the drain-to-source voltage (V_{DS}) are the external parameters controlling the drain (channel) current I_D .

Gradual Channel Approximation

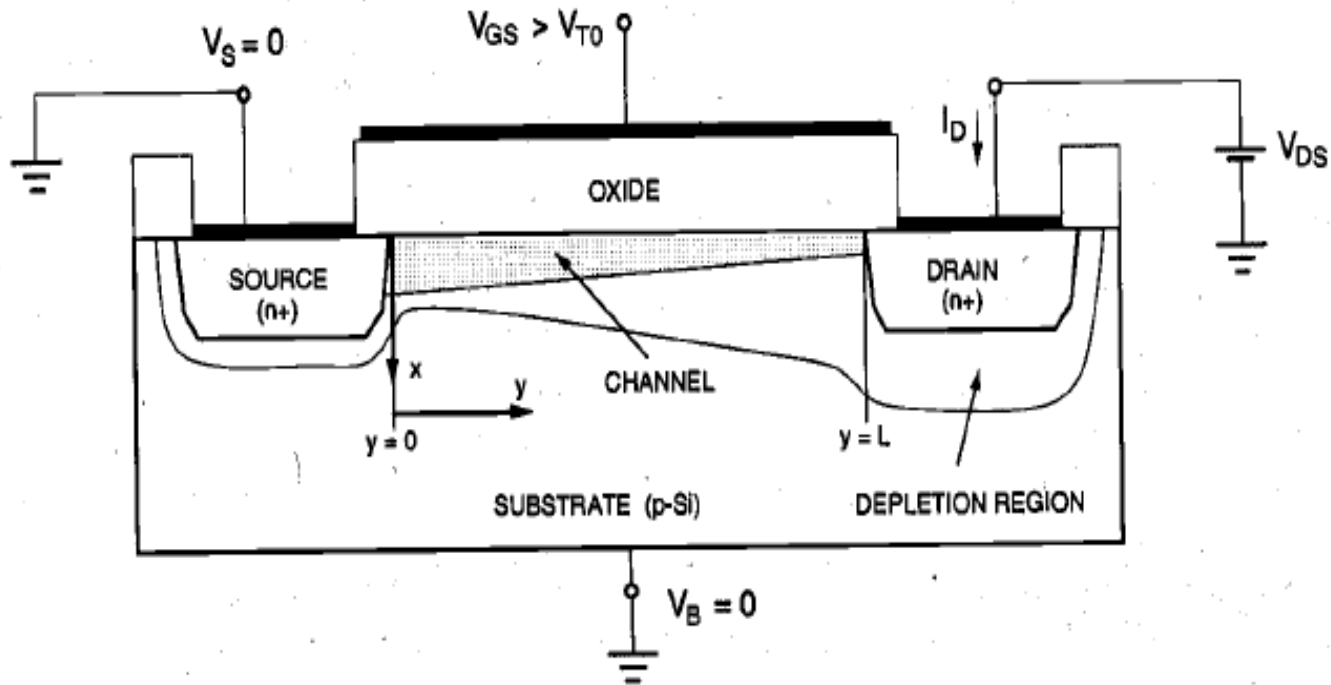


Fig. 22: Cross-sectional view of an n-channel transistor, operating in linear region.

Gradual Channel Approximation

- The gate-to-source voltage is set to be larger than the threshold voltage V_{T0} to create a conducting inversion layer between the source and the drain.
- The coordinate system for this structure is defined such that the x-direction is perpendicular to the surface, pointing down into the substrate, and the y-direction is parallel to the surface.

Gradual Channel Approximation

- The y -coordinate origin ($y = 0$) is at the source end of the channel.
- The channel voltage with respect to the source will be denoted by $V_c(y)$.
- Let us assume that the threshold voltage V_{T0} is constant along the entire channel region, between $y = 0$ and $y = L$.

Gradual Channel Approximation

- In reality, the threshold voltage changes along the channel since the channel voltage is not constant.
- Next, let us assume that the electric field component E_y along the y-coordinate is dominant compared to the electric field component E_x along the x-coordinate.
- This assumption allows to reduce the current-flow problem in the channel to the y dimension only.

Gradual Channel Approximation

- The boundary conditions for the channel voltage V_c are:

$$\begin{aligned} V_c(y=0) &= V_s = 0 \\ V_c(y=L) &= V_{DS} \end{aligned} \quad \text{.....(45)}$$

Also, it is assumed that the entire channel region between the source and the drain is inverted, i.e.,

Gradual Channel Approximation

$$\begin{aligned} V_{GS} &\geq V_{T0} \\ V_{GD} = V_{GS} - V_{DS} &\geq V_{T0} \end{aligned} \quad \text{.....(46)}$$

- The channel current (drain current) I_D is due to the electrons in the channel region traveling from the source to the drain under the influence of the lateral electric field component E_y .

Gradual Channel Approximation

- Since the current flow in the channel is primarily governed by the lateral drift of the mobile electron charge in the surface inversion layer, Let $Q_I(y)$ be the total mobile electron charge in the surface inversion layer.
- This charge can be expressed as a function of the gate-to-source voltage V_{GS} and of the channel voltage $V_c(y)$ as follows:

$$Q_I(y) = -C_{ox} \cdot [V_{GS} - V_c(y) - V_{T0}] \quad \text{.....(47)}$$

Gradual Channel Approximation

- Figure 23 shows the spatial geometry of the surface inversion layer and indicates its significant dimensions.
- The thickness of the inversion layer tapers off from the source to the drain, since the gate-to-channel voltage causing surface inversion is smaller at the drain end.

Gradual Channel Approximation

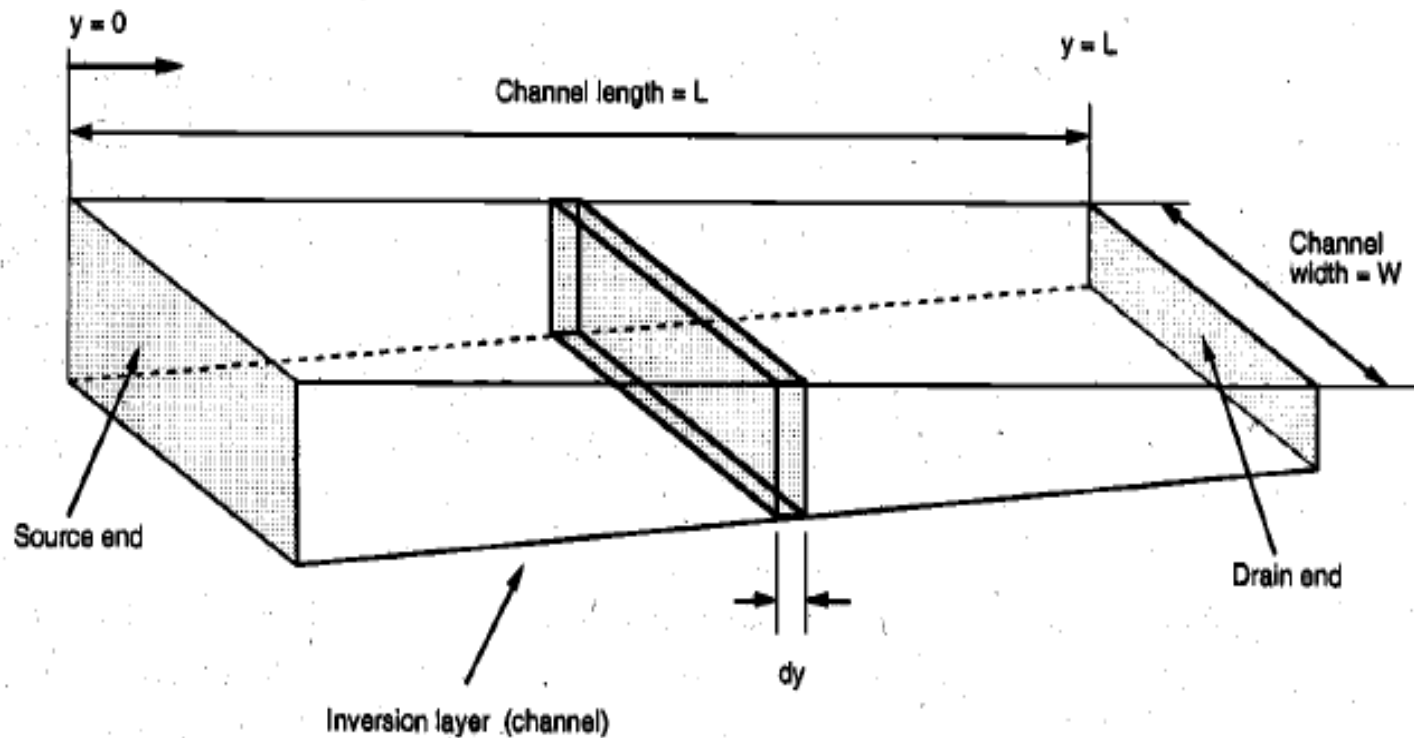


Fig. 23: Simplified geometry of the surface inversion layer (channel region).

Gradual Channel Approximation

- The incremental resistance dR of the differential channel segment is shown in Fig 23.
- Assuming that all mobile electrons in the inversion layer have a constant surface mobility μ_n , the incremental resistance can be expressed as follows.

$$dR = -\frac{dy}{W \cdot \mu_n \cdot Q_I(y)} \quad \text{.....(48)}$$

Gradual Channel Approximation

- The minus sign is due to the negative polarity of the inversion layer charge Q_i .
- Applying Ohm's law for this segment yields the voltage drop along the incremental segment dy , in the y direction;

$$dV_c = I_D \cdot dR = -\frac{I_D}{W \cdot \mu_n \cdot Q_i(y)} \cdot dy \quad \text{.....(49)}$$

Gradual Channel Approximation

- The Eqn. 49 can now be integrated along the channel, i.e., from $y = 0$ to $y = L$, using the boundary conditions given in (Eqn. 45).

$$\int_0^L I_D \cdot dy = -W \cdot \mu_n \int_0^{V_{DS}} Q_I(y) \cdot dV_c \quad \text{.....(50)}$$

- The integral on the right-hand side is evaluated by replacing $Q_I(y)$ with (Eqn. 47).

Gradual Channel Approximation

- Thus,

$$I_D \cdot L = W \cdot \mu_n \cdot C_{ox} \int_0^{V_{DS}} (V_{GS} - V_c - V_{T0}) \cdot dV_c \quad \text{.....(51)}$$

- Assuming that the channel voltage V_c , is the only variable in (Eqn. 51) that depends on the position y , the drain current is found as follows.

Gradual Channel Approximation

$$I_D = \frac{\mu_n \cdot C_{ox}}{2} \cdot \frac{W}{L} \cdot \left[2 \cdot (V_{GS} - V_{T0}) V_{DS} - V_{DS}^2 \right] \quad \text{.....(52)}$$

- Equation (52) represents the drain current I_D as a simple second-order function of the two external voltages, V_{GS} and V_{DS} .
- This current equation can also be rewritten as;

$$I_D = \frac{k'}{2} \cdot \frac{W}{L} \cdot \left[2 \cdot (V_{GS} - V_{T0}) V_{DS} - V_{DS}^2 \right] \quad \text{.....(53)}$$

Gradual Channel Approximation

- Or

$$I_D = \frac{k}{2} \cdot \left[2 \cdot (V_{GS} - V_{T0}) V_{DS} - V_{DS}^2 \right] \quad \text{.....(54)}$$

- where the parameters k and k' are defined as;

$$k' = \mu_n \cdot C_{ox} \quad \text{.....(55)}$$

$$k = k' \cdot \frac{W}{L} \quad \text{.....(56)}$$

Gradual Channel Approximation

- The drain current equation given in (53) is the simplest analytical approximation for the MOSFET current-voltage relationship.
- In addition to the process dependent constants k' and V_{T0} , the current-voltage relationship is also affected by the device dimensions, W and L .
- In fact, the ratio of W/L is one of the most important design parameters in MOS digital circuit design.

Problem 5

- For an n-channel MOS transistor with $\mu_n = 600 \text{ cm}^2/\text{V.s}$, $C_{ox} = 7 \cdot 10^{-8} \text{ F/cm}^2$, $W = 20 \text{ }\mu\text{m}$, $L = 2 \text{ }\mu\text{m}$ and $V_{T0} = 1.0 \text{ V}$, examine the relationship between the drain current and the terminal voltages.

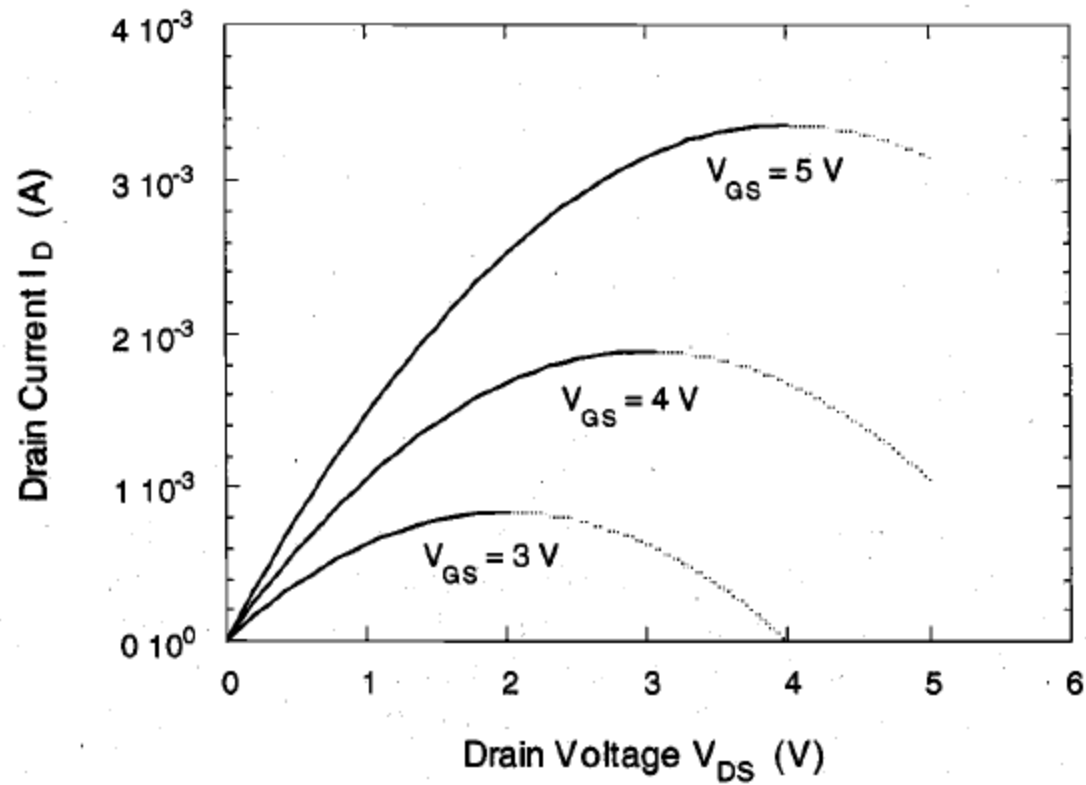
Solution

- First, calculate the parameter k and then I_D using Eqn. 54 as follows :

$$k = \mu_n \cdot C_{ox} \cdot \frac{W}{L} = 600 \text{ cm}^2/\text{V} \cdot \text{s} \times 7 \cdot 10^{-8} \text{ F/cm}^2 \times \frac{20 \text{ } \mu\text{m}}{20 \text{ } \mu\text{m}} = 0.42 \text{ mA/V}^2$$

$$I_D = 0.21 \text{ mA/V}^2 \left[2 \cdot (V_{GS} - 1.0) \cdot V_{DS} - V_{DS}^2 \right]$$

Solution



Solution

- The drain current-drain voltage curves shown above reach their peak value for
- $V_{DS} = V_{GS} - V_{T0}$.
- Beyond this maximum, each curve exhibits a negative differential conductance, which is not observed in actual MOSFET current-voltage measurements (section shown by the dashed lines).

Solution

- The drain current equation (52) has been derived under the following voltage assumptions,
- $V_{GS} \geq V_{T0}$
- $V_{GD} = V_{GS} - V_{DS} \geq V_{T0}$
- which guarantee that the entire channel region between the source and the drain is inverted.

Solution

- This condition corresponds to the linear operating mode for the MOSFET.
- Hence, the current equation (52) is MOS Transistor valid only for the linear mode operation.

Saturation Mode of Operation

- Beyond the linear region boundary, i.e., for V_{DS} values larger than $V_{GS} - V_{T0}$, the MOS transistor will be assumed to be in saturation.
- A different current-voltage expression will be necessary for the MOSFET operating in this region.

Saturation Mode of Operation

- Problem 5 shows that the current equation (52) is not valid beyond the linear region/saturation region boundary, i.e., for
- $V_{DS} \geq V_{DSAT} = V_{GS} - V_{TO}$ (57)
- This saturation drain current can be obtained simply by substituting Eqn. (57) for V_{DS} in Eqn. (52).

Saturation Mode of Operation

$$\begin{aligned} I_D(sat) &= \frac{\mu_n \cdot C_{ox}}{2} \cdot \frac{W}{L} \cdot \left[2 \cdot (V_{GS} - V_{T0}) \cdot (V_{GS} - V_{T0}) - (V_{GS} - V_{T0})^2 \right] \\ &= \frac{\mu_n \cdot C_{ox}}{2} \cdot \frac{W}{L} \cdot (V_{GS} - V_{T0})^2 \end{aligned} \quad \dots(58)$$

- Thus, the drain current I_D becomes a function only of the gate-to-source voltage V_{GS} , beyond the saturation boundary.

MOSFET I_D - V_{DS} Characteristic

- Figure 24 shows the typical drain current versus drain voltage characteristics of an n-channel MOSFET, as described by the current equations (52) and (58).
- The parabolic boundary between the linear and the saturation regions is indicated here by the dashed line.

MOSFET I_D - V_{DS} Characteristic

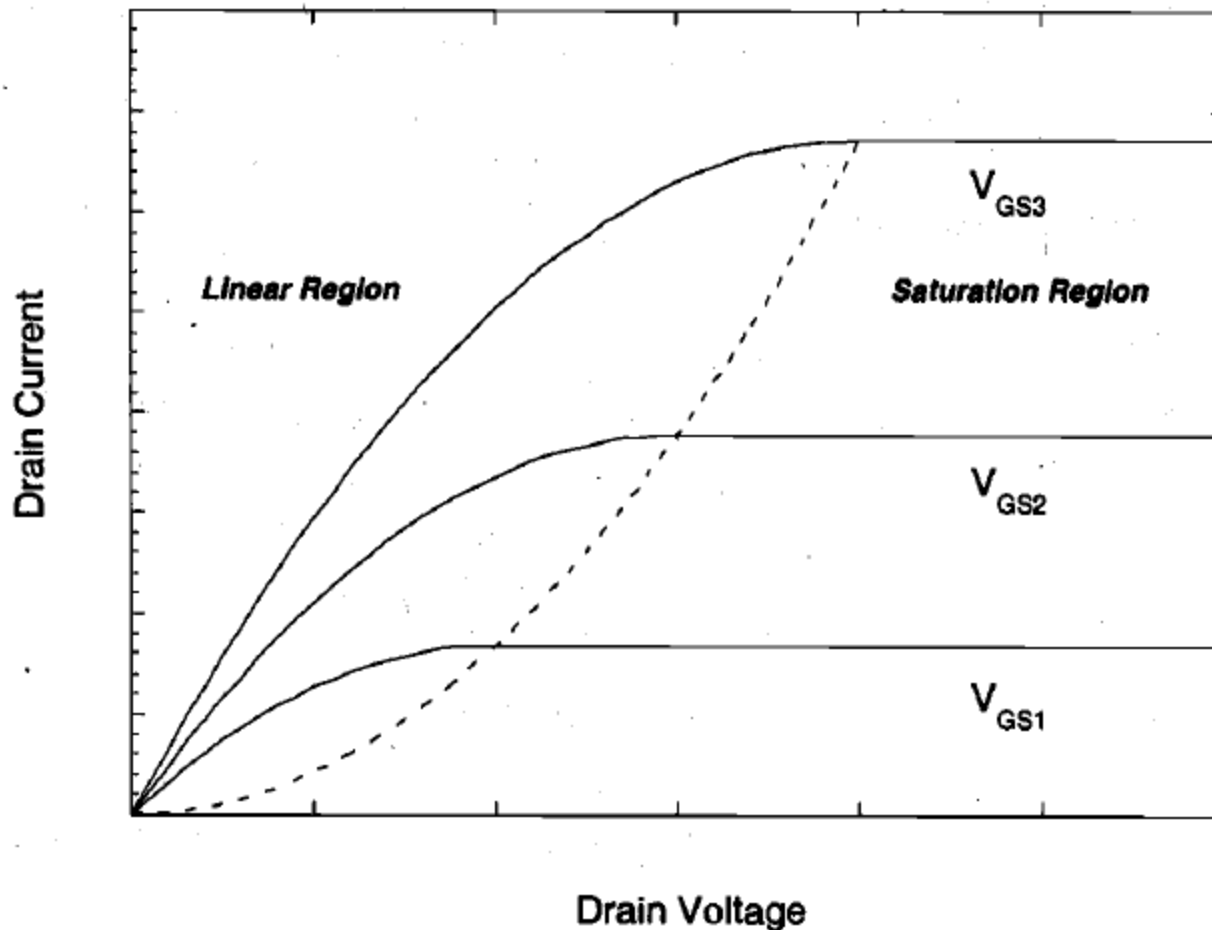


Fig. 24: Basic current-voltage characteristics of an n-channel MOS transistor.

MOSFET I_D - V_{GS} Characteristic

- The current-voltage characteristics of the MOS transistor can also be visualized by plotting the drain current as a function of the gate voltage, as shown in Fig. 25.
- The current is obviously equal to zero for any gate voltage smaller than the threshold voltage V_{T0} .

MOSFET I_D - V_{GS} Characteristic

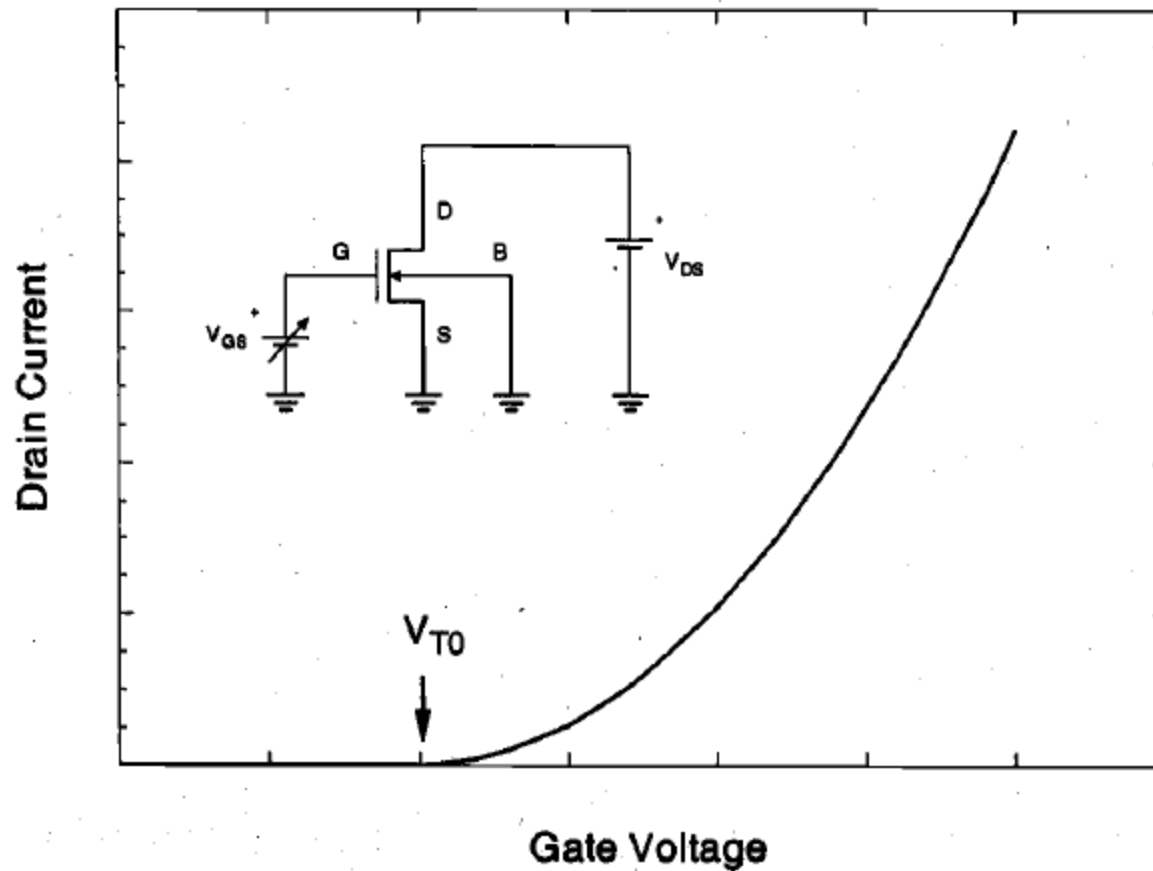
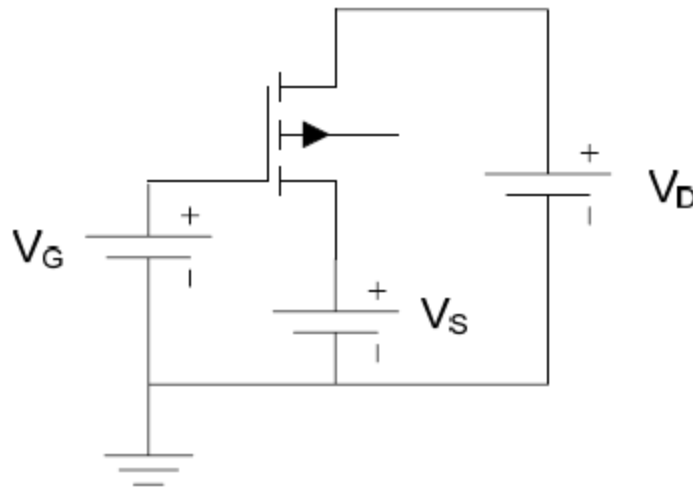


Fig. 25

Problem 6

- Determine the regions of operation for the following different NMOS transistor configurations. Assume $V_{TH} = 0.4V$

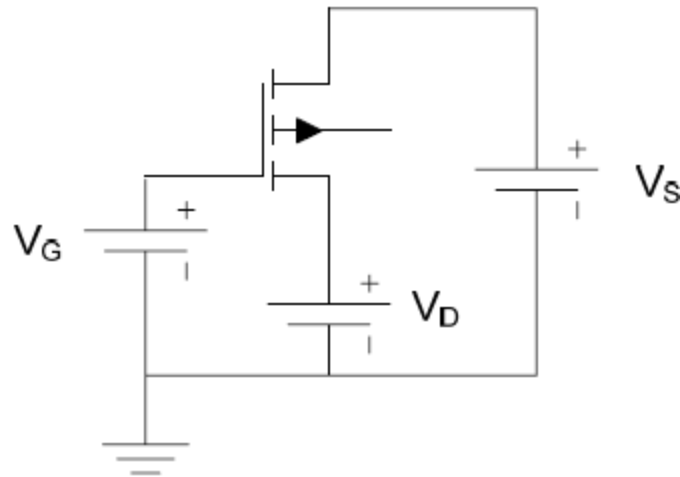


Solution

- a) $V_G = 1V$, $V_S = 0V$, $V_D = 0.5V$. Region: Triode, $V_{DS} = 0.5V < 0.6V = (V_{GS} - V_{TH})$
- b) $V_G = 0.7V$, $V_S = -0.4V$, $V_D = 0V$. Region: Triode, $V_{DS} = 0.4V < 0.7V = (V_{GS} - V_{TH})$
- c) $V_G = 0.3V$, $V_S = -1V$, $V_D = 0.2V$. Region: Saturation, $V_{DS} = 1.2V > 0.9V = (V_{GS} - V_{TH})$

Problem 7

- Assume the transistor is now a PMOS transistor, find the regions of operations. Assume $V_{TH} = -0.4V$

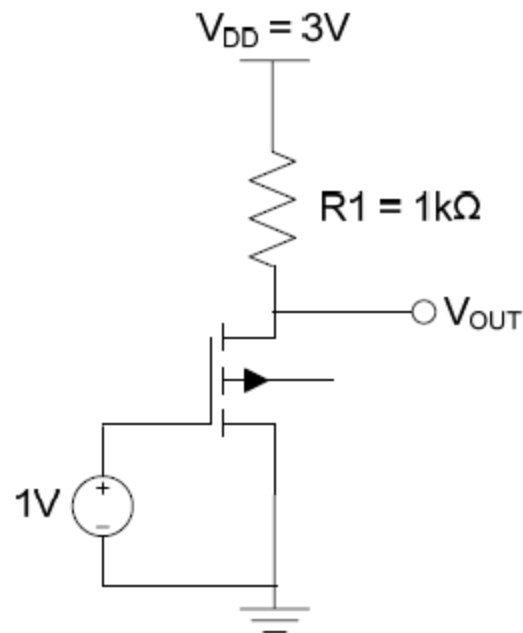


Solution

- a) $V_G = 0V$, $V_S = 1V$, $V_D = 0.5V$. Region: Triode, $V_{SD} = 0.5V < 0.6V = (V_{SG} + V_{TH})$
- b) $V_G = 0.2V$, $V_S = 0.5V$, $V_D = -0.3V$. Region: Cutoff, $V_{SG} = 0.3V < 0.4V = -V_{TH}$
- c) $V_G = 0V$, $V_S = 0.6V$, $V_D = 0.2V$. Region: Saturation, $V_{SD} = 0.4V > 0.2V = (V_{SG} + V_{TH})$

Problem 8

- Consider the circuit in figure. What is the smallest we can size the NMOS transistor with 90nm technology such that the out voltage $V_{OUT} = 2V$? Assume $V_{TH} = 0.4V$, and $\mu_n C_{ox} = 2.22 \times 10^{-4} \text{ F/V} \cdot \text{s}$.



Solution

$$I_D = \frac{V_{DD} - V_{OUT}}{R_1} = \frac{3V - 2V}{1000\Omega} = 1mA$$

$$I_{D,sat} = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_{TH})^2$$

$$1mA = \frac{1}{2} 2.22 \times 10^{-4} \frac{F}{cm^2} \frac{W}{L} (1V - 0.4V)^2$$

$$\frac{W}{L} = 25$$

$$L_{\min} = 90nm$$

$$W_{\min} = 2.25\mu m$$

Channel Length Modulation

- The mechanisms of channel pinch-off and current flow in saturation mode is examined here.
- Using Eqn. (47), the inversion layer charge at the source end of the channel is;

$$Q_l(y=0) = -C_{ox} \cdot (V_{GS} - V_{T0}) \quad \text{.....(59)}$$

Channel Length Modulation

- and the inversion layer charge at the drain end of the channel is;

$$Q_I(y = L) = -C_{ox} \cdot (V_{GS} - V_{T0} - V_{DS}) \quad \text{.....(60)}$$

- when the drain-to-source voltage reaches V_{DSAT} , the inversion layer charge at the drain end

$$V_{DS} = V_{DSAT} = V_{GS} - V_{T0} \quad \text{.....(61)}$$

Channel Length Modulation

- becomes zero, according to Eqn. (60).
- In reality, the channel charge does not become exactly equal to zero, but it indeed becomes very small

$$Q_I(y = L) \approx 0 \quad \text{.....(62)}$$

- Thus, under the bias condition given in (61), the channel is pinched off at the drain end, i.e., at $y = L$.

Channel Length Modulation

- If the drain-to-source voltage V_{DS} is increased even further beyond the saturation edge so that $V_{DS} > V_{DSAT}$, an even larger portion of the channel becomes pinched-off.
- Consequently, the effective channel length is reduced to

$$L' = L - \Delta L \quad \text{.....(63)}$$

- where ΔL is the length of the channel segment with $Q_i = 0$ as shown in Fig. 26.

Channel Length Modulation

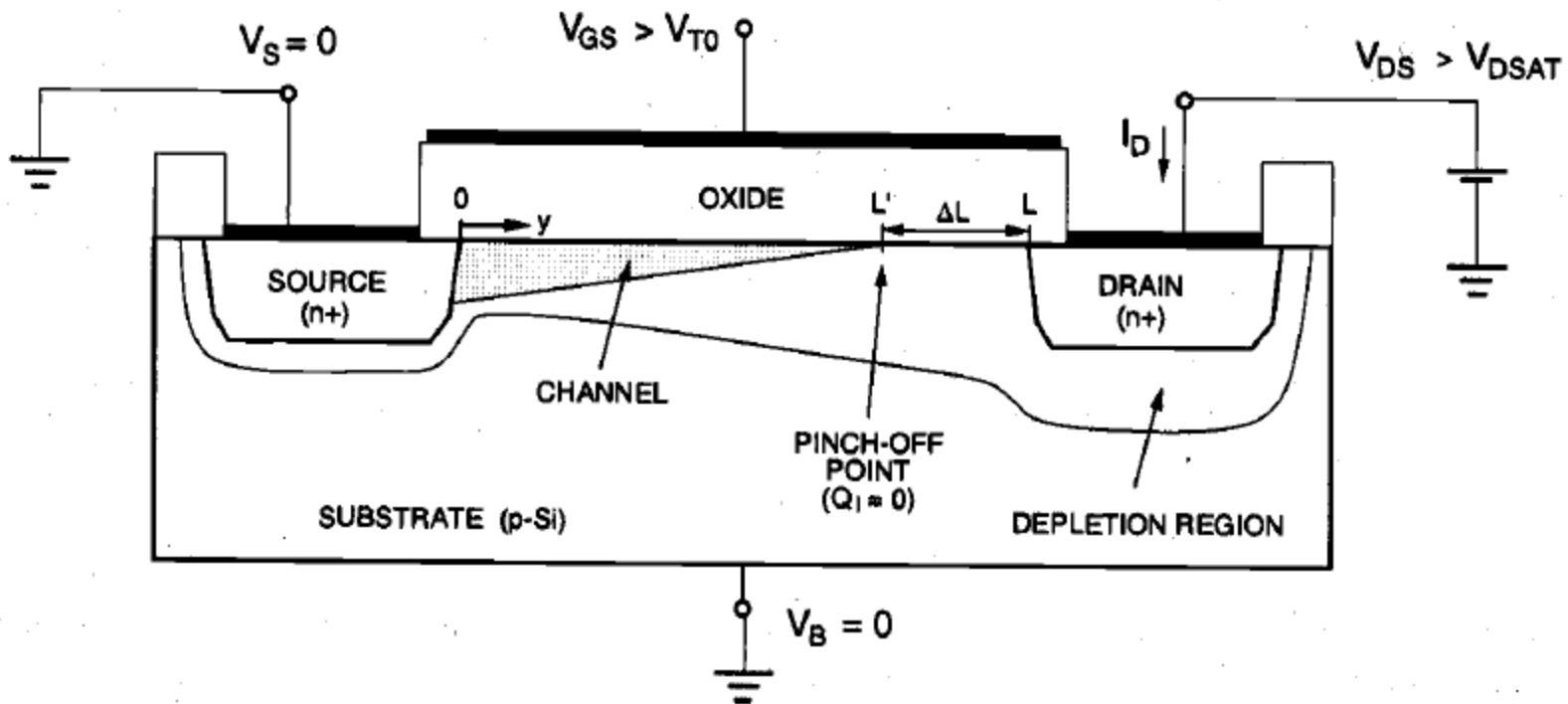


Fig. 26: Channel length modulation in an n-channel MOSFET operation in saturation mode

Channel Length Modulation

- Since $Q_i(y) = 0$ for $L' < y < L$, the channel voltage at the pinch-off point remains equal to V_{DSAT} , i.e.,

$$V_c(y = L') = V_{DSAT} \quad \text{.....(64)}$$

- As seen in Fig. 26, the inverted portion of the surface by a shortened channel can be represented, with a channel-end voltage of V_{DSAT} .

Channel Length Modulation

- The gradual channel approximation is valid in this region; thus, the channel current can be found using Eqn. (58).

$$I_D(sat) = \frac{\mu_n \cdot C_{ox}}{2} \cdot \frac{W}{L'} \cdot (V_{GS} - V_{T0})^2 \quad \text{.....(65)}$$

- This current equation corresponds to a MOSFET with effective channel length L' , operating in saturation.

Channel Length Modulation

- Thus, Eqn. (65) accounts for the actual shortening of the channel, also called channel length modulation.
- Since $L' < L$, the saturation current calculated by using Eqn. (65) will be larger than that found by using Eqn. (58) under the same bias conditions.

Channel Length Modulation

- As L' decreases with increasing V_{DS} , the saturation mode current $I_{D(sat)}$ will also increase with V_{DS} .
- By approximating the effective channel length $L' = L - \Delta L$ as a function of the drain bias voltage, we can modify (65) to reflect this drain voltage dependence as follows.

$$I_{D(sat)} = \left(\frac{1}{1 - \frac{\Delta L}{L}} \right) \cdot \frac{\mu_n \cdot C_{ox}}{2} \cdot \frac{W}{L} \cdot (V_{GS} - V_{T0})^2 \quad \text{.....(66)}$$

Channel Length Modulation

- The first term of this saturation current expression accounts for the channel modulation effect, while the rest of this expression is identical to (58).
- It can be shown that the channel length shortening ΔL is actually proportional to the square root of $(V_{DS} - V_{DSAT})$.

$$\Delta L \propto \sqrt{V_{DS} - V_{DSAT}} \quad \text{.....(67)}$$

Channel Length Modulation

- To simplify the analysis even further, we will use the following empirical relation between ΔL and the drain-to-source voltage instead:

$$1 - \frac{\Delta L}{L} \approx 1 - \lambda \cdot V_{DS} \quad \text{.....(68)}$$

- Here, λ is an empirical model parameter, and is called the channel length modulation coefficient.

Channel Length Modulation

- Assuming that $\lambda V_{DS} \ll 1$, the saturation current given in (Eqn. 66) can now be written as:

$$I_D(sat) = \frac{\mu_n \cdot C_{ox}}{2} \cdot \frac{W}{L} \cdot (V_{GS} - V_{T0})^2 \cdot (1 + \lambda \cdot V_{DS}) \quad \text{.....(69)}$$

- The slope of the current voltage curve in the saturation region is determined by the channel length modulation coefficient λ as shown in Fig. 27.

Current-voltage characteristics of an n-channel MOS transistor, including the channel length modulation effect

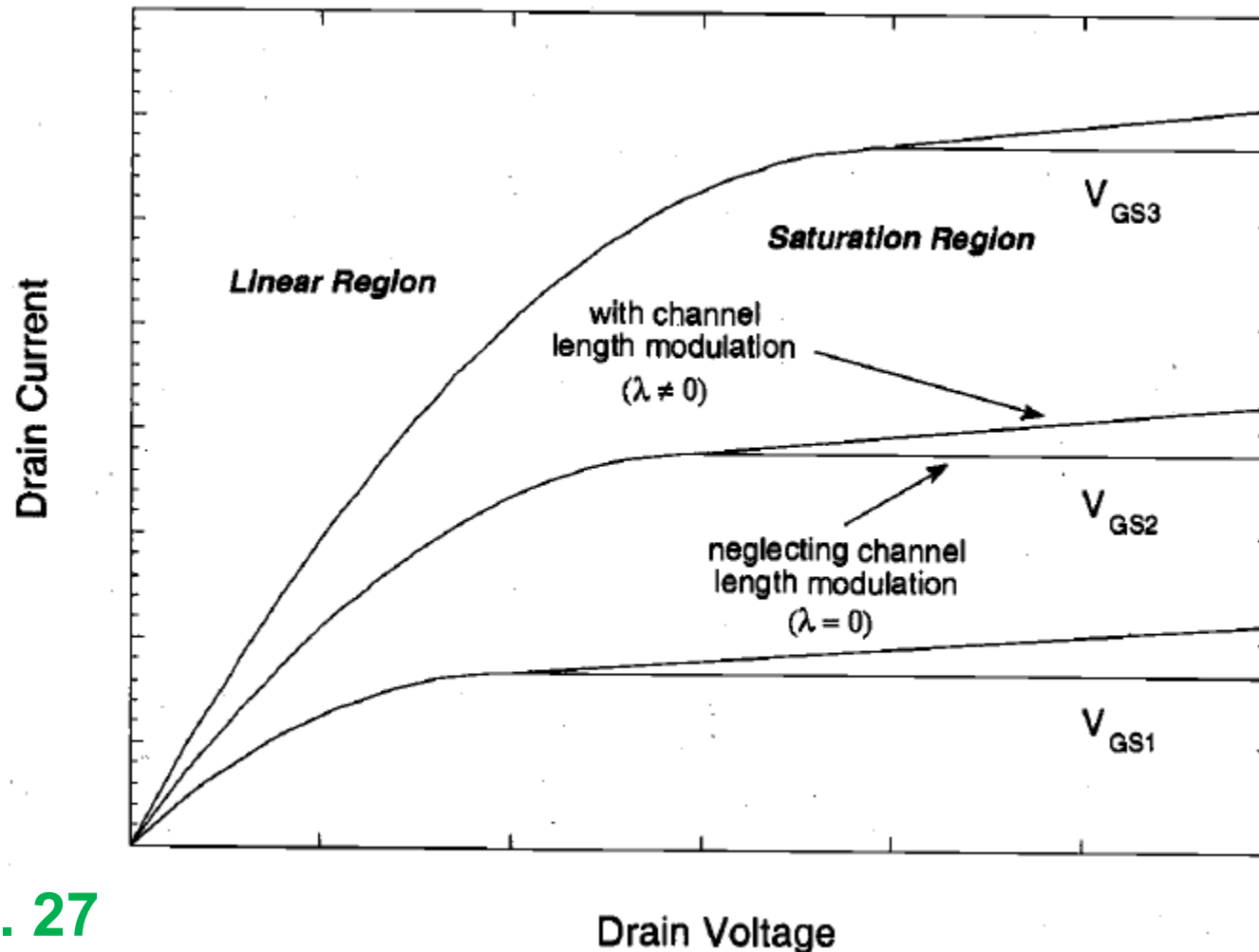


Fig. 27

Substrate Bias Effect

- The derivation of linear-mode and saturation-mode current-voltage characteristics in the previous sections has been done under the assumption that the substrate potential is equal to the source potential, i.e., $V_{SB} = 0$.
- Consequently, the zero-substrate bias threshold voltage V_{TO} has been used in the current equations.

Substrate Bias Effect

- In many digital circuit applications, the source potential of an nMOS transistor can be larger than the substrate potential, which results in a positive source-to-substrate voltage $V_{SB} > 0$.
- The general expression for the threshold voltage V_T , which already includes the substrate bias term can be expressed as;

$$V_T(V_{SB}) = V_{T0} + \gamma \cdot \left(\sqrt{|2\phi_F| + V_{SB}} - \sqrt{|2\phi_F|} \right) \dots\dots(70)$$

Substrate Bias Effect

- The threshold voltage terms in linear-mode and saturation-mode current equations with the more general $V_T(V_{SB})$ term can be represented as;

$$I_D(lin) = \frac{\mu_n \cdot C_{ox}}{2} \cdot \frac{W}{L} \cdot [2 \cdot (V_{GS} - V_T(V_{SB}))V_{DS} - V_{DS}^2] \quad \dots(71)$$

$$I_D(sat) = \frac{\mu_n \cdot C_{ox}}{2} \cdot \frac{W}{L} \cdot (V_{GS} - V_T(V_{SB}))^2 \cdot (1 + \lambda \cdot V_{DS}) \quad \dots(72)$$

Substrate Bias Effect

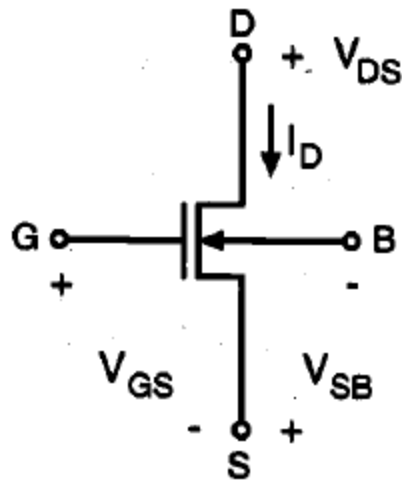
- The substrate-bias effect can significantly change the value of the threshold voltage and, hence, the current capability of the MOSFET.
- The first-order characterization of the drain (channel) current as a nonlinear function of the terminal voltages can be expressed as follows;

$$I_D = f(V_{GS}, V_{DS}, V_{BS}) \quad \text{.....(73)}$$

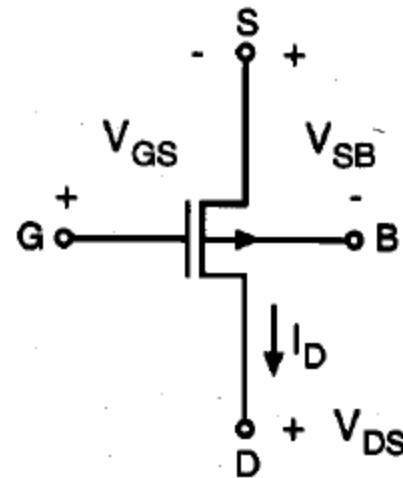
Substrate Bias Effect

- Figure 28 shows the polarities of applied terminal voltages and the drain current directions for nMOS and pMOS transistors.
- The threshold voltage V_T and the terminal voltages V_{GS} , V_{DS} , and V_{SB} are all negative for the pMOS transistor.
- The parameter μ_p denotes the surface hole mobility in the pMOSFET.

Polarities of voltages of nMOS & pMOS



n-channel MOSFET



p-channel MOSFET

Fig. 28

Current-voltage equations of the n-channel MOSFET

$$I_D = 0, \text{ for } V_{GS} < V_T$$

$$I_D(\text{lin}) = \frac{\mu_n \cdot C_{ox}}{2} \cdot \frac{W}{L} \cdot [2 \cdot (V_{GS} - V_T) V_{DS} - V_{DS}^2] \quad \text{for } V_{GS} \geq V_T$$

and $V_{DS} < V_{GS} - V_T$

$$I_D(\text{sat}) = \frac{\mu_n \cdot C_{ox}}{2} \cdot \frac{W}{L} \cdot (V_{GS} - V_T)^2 \cdot (1 + \lambda \cdot V_{DS}) \quad \text{for } V_{GS} \geq V_T$$

and $V_{DS} \geq V_{GS} - V_T$

Current-voltage equations of the p-channel MOSFET

$$I_D = 0, \quad \text{for } V_{GS} > V_T$$

$$I_D(\text{lin}) = \frac{\mu_p \cdot C_{ox}}{2} \cdot \frac{W}{L} \cdot \left[2 \cdot (V_{GS} - V_T) V_{DS} - V_{DS}^2 \right] \quad \text{for } V_{GS} \leq V_T$$

and $V_{DS} > V_{GS} - V_T$

$$I_D(\text{sat}) = \frac{\mu_p \cdot C_{ox}}{2} \cdot \frac{W}{L} \cdot (V_{GS} - V_T)^2 \cdot (1 + \lambda \cdot V_{DS}) \quad \text{for } V_{GS} \leq V_T$$

and $V_{DS} \leq V_{GS} - V_T$

Measurement of Parameters

- To exploit the simplicity of the equations and to achieve the maximum possible accuracy in calculations, the parameters appearing in the current equations must be determined carefully, through experimental measurements.
- The model parameters that are used in the previous equations are the zero-bias threshold voltage V_{T0} , the substrate-bias coefficient γ , the channel length modulation coefficient λ ,

Measurement of Parameters

- and the following transconductance parameters:

$$k_n = \mu_n \cdot C_{ox} \cdot \frac{W}{L} \quad \text{.....(74)}$$

$$k_p = \mu_p \cdot C_{ox} \cdot \frac{W}{L} \quad \text{.....(75)}$$

Measurement of Parameters

- Let us consider the test circuit setup shown in Fig. 29 for an nMOS transistor.
- The source-to-substrate voltage V_{SB} is set at a constant value, and the drain current is measured for different values of the gate-to-source voltage V_{GS} .
- Since the drain and the gate of the transistor are at the same potential, $V_{DS} = V_{GS}$. Hence, the saturation condition $V_{DS} > V_{GS} - V_T$ is

Measurement of Parameters

- always satisfied, i.e., the nMOS transistor shown in Fig. 29(a) operates in saturation mode.
- Neglecting the channel length modulation effect for simplicity, the drain current is described by

$$I_D(sat) = \frac{k_n}{2} \cdot (V_{GS} - V_{T0})^2 \quad \text{.....(76)}$$

Measurement of Parameters

- Now, the square root of the drain current can be written as a linear function of the gate-to-source voltage.

$$\sqrt{I_D} = \sqrt{\frac{k_n}{2}} \cdot (V_{GS} - V_{T0}) \quad \text{.....(77)}$$

Measurement of Parameters

- If the square root of the measured drain current values is plotted against the gate-to-source voltage, the slope and the voltage-axis intercept of the resulting curve(s) can determine the parameters k_n , V_{T0} and γ .
- Figure 29(b) shows the measured drain current vs. gate voltage curves, obtained for different values of substrate bias.
- The slope of each curve is equal to the square root of $(k_n/2)$. *Thus, the transconductance parameter k_n can simply be calculated from this slope.*

Measurement of Parameters

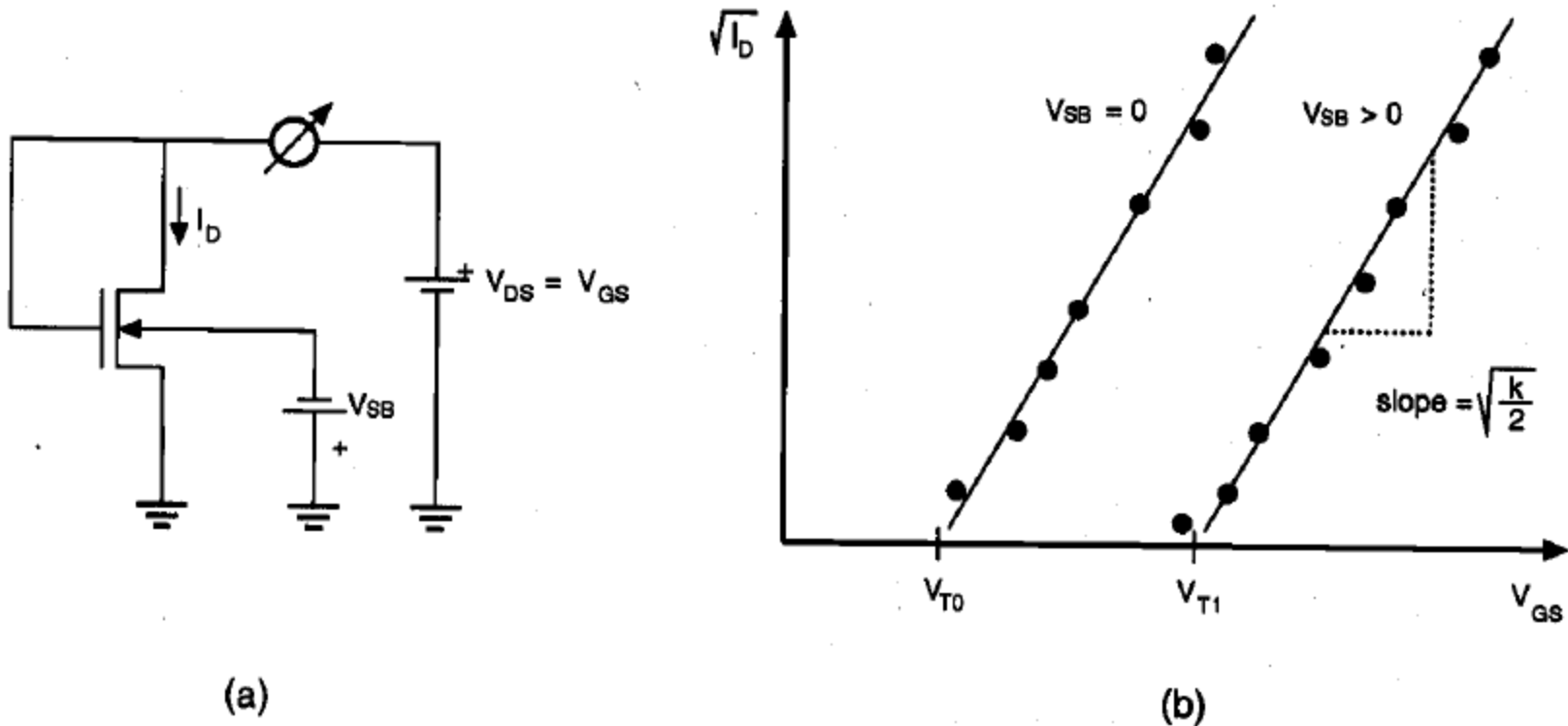


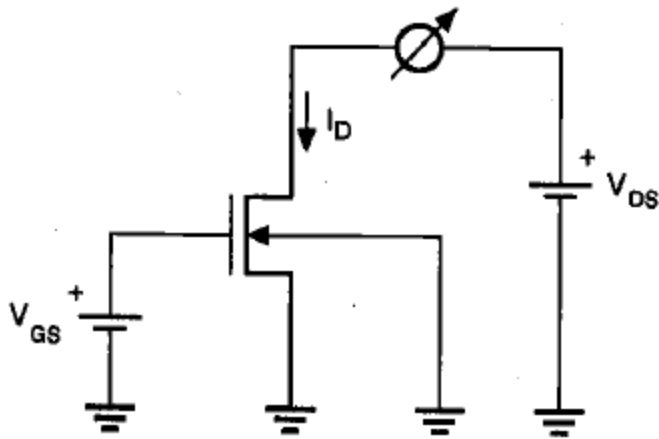
Figure 29. (a) Test circuit arrangement and (b) measured data for experimental determination of the parameters k_n , V_{T0} and γ .

Measurement of Parameters

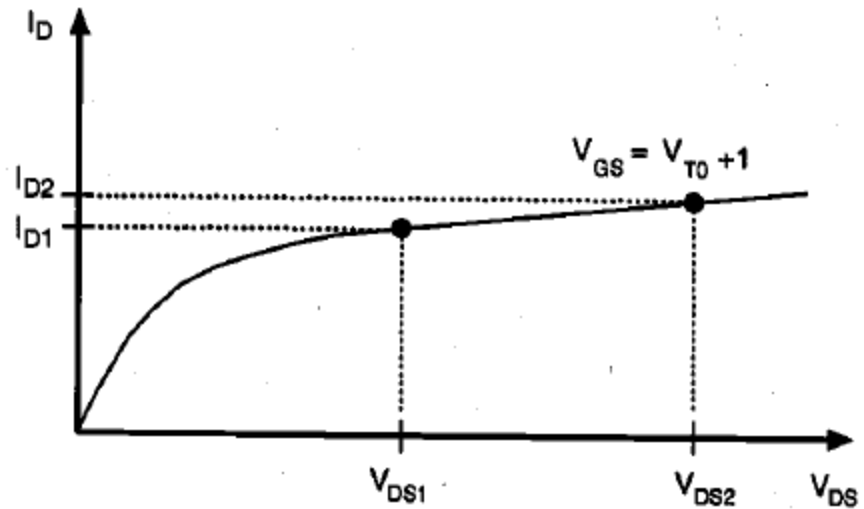
- Next, consider the extrapolated threshold voltage values, obtained from voltage axis intercepts at nonzero substrate bias voltage.
- Using one of the available V_{SB} values, the substrate bias coefficient can be found from

$$\gamma = \frac{V_T(V_{SB}) - V_{T0}}{\sqrt{|2\phi_F| + V_{SB}} - \sqrt{|2\phi_F|}} \quad \text{.....(78)}$$

Measurement of Parameters



(a)



(b)

Figure 30. (a) Test circuit arrangement and (b) measured data for experimental determination of the channel length modulation coefficient λ .

Measurement of Parameters

- The experimental measurement of the channel length modulation coefficient λ requires a different test circuit setup, as shown in Fig. 30 (a).
- The gate-to-source voltage V_{GS} is set to $V_{T0} + 1$.
- The drain-to-source voltage is chosen sufficiently large ($V_{DS} > V_{GS} - V_{T0}$) that the transistor operates in the saturation mode.

Measurement of Parameters

- The drain current in the saturation mode is given by

$$I_D(sat) = \frac{k_n}{2} \cdot (V_{GS} - V_{T0})^2 \cdot (1 + \lambda \cdot V_{DS}) \quad \dots\dots(79)$$

- Since $V_{GS} = V_{T0} + 1$, the ratio of the measured drain current values I_{D1} , and I_{D2} is

$$\frac{I_{D2}}{I_{D1}} = \frac{1 + \lambda \cdot V_{DS2}}{1 + \lambda \cdot V_{DS1}} \quad \dots\dots\dots(80)$$

Measurement of Parameters

- which can be used to calculate the channel length modulation coefficient λ .
- This is equivalent to calculating the slope of the drain current versus drain voltage curve in the saturation region, as shown in Fig. 30(b). Specifically, the slope is $(\lambda k_n/2)$.

Problem 9

- Measured voltage and current data for a MOSFET are given below.
- Determine the type of the device, and calculate the parameters k_n , V_{T0} , and γ .
- Assume $\phi_F = -0.3$ V.

Problem 9

V_{GS} (V)	V_{DS} (V)	V_{SB} (V)	I_D (μ A)
3	3	0	97
4	4	0	235
5	5	0	433
3	3	3	59
4	4	3	173
5	5	3	347

Solution

- First, the MOS transistor is on ($I_D > 0$) for $V_{GS} > 0$ and $V_{DS} > 0$.
- Thus, the transistor must be an n-channel MOSFET.
- Assume that the transistor is enhancement-type and, therefore, operating in saturation mode for $V_{GS} = V_{DS}$.
- Neglecting the channel length modulation effect, the saturation mode current is written as

Solution

$$I_D = \frac{k_n}{2} \cdot (V_{GS} - V_T)^2 \quad \Leftrightarrow \quad \sqrt{I_D} = \sqrt{\frac{k_n}{2}} \cdot (V_{GS} - V_T)$$

- Let (V_{GS1}, I_{D1}) and (V_{GS2}, I_{D2}) be any two current-voltage pairs obtained from the table.
- Then, the square-root of the transconductance parameter k_n can be calculated.

$$\sqrt{\frac{k_n}{2}} = \frac{\sqrt{I_{D1}} - \sqrt{I_{D2}}}{V_{GS1} - V_{GS2}} = \frac{\sqrt{433 \mu\text{A}} - \sqrt{97 \mu\text{A}}}{5 \text{ V} - 3 \text{ V}} = 5.48 \times 10^{-3} \text{ A}^{1/2}/\text{V}$$

Solution

- Thus, the transconductance parameter of this n-channel MOSFET is:

$$k_n = 2 \cdot (5.48 \times 10^{-3})^2 = 60 \times 10^{-6} \text{ A/V}^2 = 60 \mu\text{A/V}^2$$

- The extrapolated threshold voltage V_{T0} at zero substrate bias can be found by calculating the x-axis intercept of the square-root of (I_D) versus V_{GS} curve.

$$V_{T0} = V_{GS} - \sqrt{\frac{2 \cdot I_D}{k_n}} = 1.2 \text{ V}$$

Solution

- To find the substrate bias coefficient γ , we must first determine the threshold voltage V_T at the source-to-substrate voltage of 3 V.
- Using one of the current-voltage data pairs corresponding to $V_{SB} = 3$ V, V_T can be calculated as follows:

$$V_T(V_{SB} = 3 \text{ V}) = V_{GS} - \sqrt{\frac{2 \cdot I_D}{k_n}} = 4 \text{ V} - \sqrt{\frac{2 \cdot 173 \mu\text{A}}{60 \mu\text{A}/\text{V}^2}} = 1.6 \text{ V}$$

Solution

- Finally, the substrate bias coefficient is found as:

$$\gamma = \frac{V_T(V_{SB} = 3 \text{ V}) - V_{T0}}{\sqrt{|2\phi_F| + V_{SB}} - \sqrt{|2\phi_F|}} = \frac{1.6 \text{ V} - 1.2 \text{ V}}{\sqrt{0.6 \text{ V} + 3 \text{ V}} - \sqrt{0.6 \text{ V}}} = 0.36 \text{ V}^{1/2}$$

MOSFET Scaling and Small-Geometry Effects

- The design of high-density chips in MOS VLSI (Very Large Scale Integration) technology requires that the packing density of MOSFETs used in the circuits is as high as possible and, consequently, that the sizes of the transistors are as small as possible.
- The reduction of the size, i.e., the dimensions of MOSFETs, is commonly referred to as **scaling**.

MOSFET Scaling and Small-Geometry Effects

- There are two basic types of size-reduction strategies: full scaling (also called constant-field scaling) and constant voltage scaling.
- Scaling of MOS transistors is concerned with systematic reduction of overall dimensions of the devices as allowed by the available technology, while preserving the geometric ratios found in the larger devices.

MOSFET Scaling and Small-Geometry Effects

- The proportional scaling of all devices in a circuit would certainly result in a reduction of the total silicon area occupied by the circuit, thereby increasing the overall functional density of the chip.
- The device scaling can be explained by introducing a constant scaling factor $S > 1$.
- All horizontal and vertical dimensions of the large-size transistor are divided by this scaling factor to obtain the scaled device.

MOSFET Scaling and Small-Geometry Effects

- The extent of scaling that is achievable is determined by the fabrication technology and more specifically, by the minimum feature size.
- It is seen that a new generation of manufacturing technology replaces the previous one about every two or three years, and the down-scaling factor S of the minimum feature size from one generation to the next is about 1.2 to 1.5.

MOSFET Scaling and Small-Geometry Effects

- The proportional scaling of all three dimensions are considered by the same scaling factor S .
- Figure 31 shows the reduction of key dimensions on a typical MOSFET, together with the corresponding increase of the doping densities.

MOSFET Scaling and Small-Geometry Effects

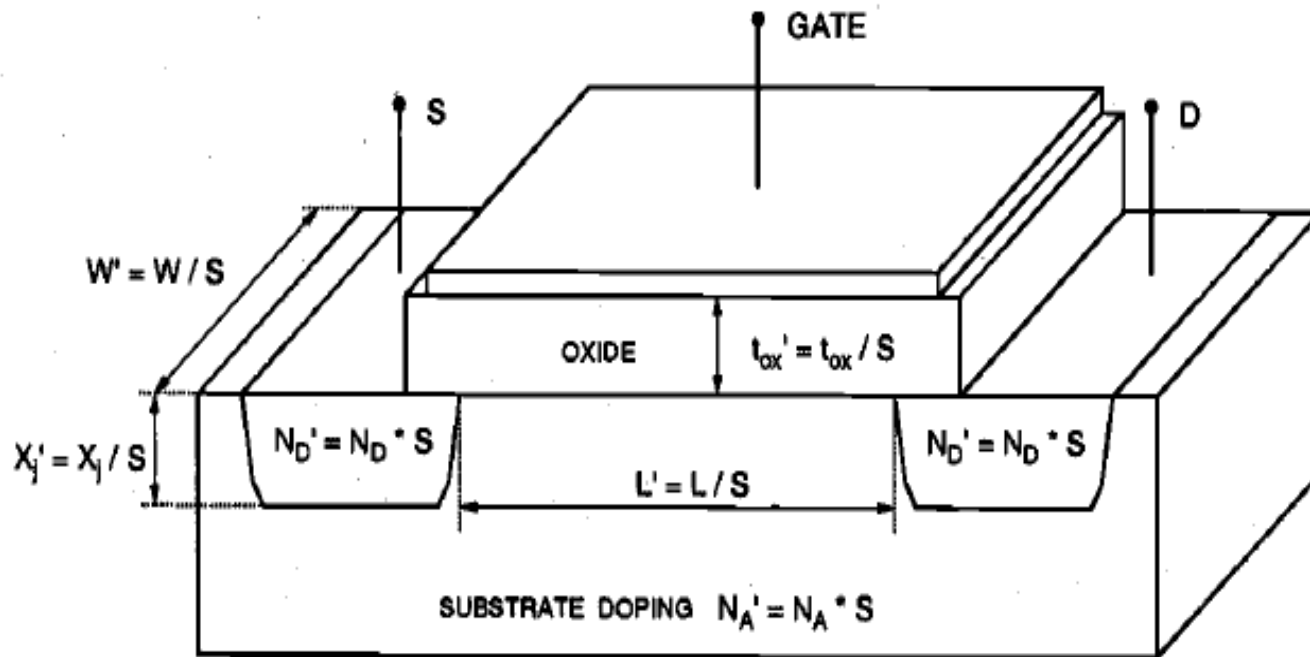


Fig. 31

MOSFET Scaling and Small-Geometry Effects

- The primed quantities in Fig. 31 indicate the scaled dimensions and doping densities.
- It is easy to recognize that the scaling of all dimensions by a factor of $S > 1$ leads to the reduction of the area occupied by the transistor by a factor of S^2 .

Full Scaling (Constant-Field Scaling)

- This scaling option attempts to preserve the magnitude of internal electric fields in the MOSFET, while the dimensions are scaled down by a factor of S .
- To achieve this, all potentials must be scaled down proportionally, by the same scaling factor.

Full Scaling (Constant-Field Scaling)

- This potential scaling also affects the threshold voltage V_{T0} .
- The scaling factors for all significant dimensions, potentials, and doping densities of the MOS transistor are listed as follows.

Full Scaling (Constant-Field Scaling)

Quantity	Before Scaling	After Scaling
Channel length	L	$L' = L / S$
Channel width	W	$W' = W / S$
Gate oxide thickness	t_{ox}	$t_{ox}' = t_{ox} / S$
Junction depth	x_j	$x_j' = x_j / S$
Power supply voltage	V_{DD}	$V_{DD}' = V_{DD} / S$
Threshold voltage	V_{T0}	$V_{T0}' = V_{T0} / S$
Doping densities	N_A	$N_A' = S \cdot N_A$
	N_D	$N_D' = S \cdot N_D$

Full Scaling (Constant-Field Scaling)

- The influence of full scaling described here upon the current-voltage characteristics of the MOS transistor is now considered.
- It will be assumed that the surface mobility μ_n is not significantly affected by the scaled doping density.

Full Scaling (Constant-Field Scaling)

- The gate oxide capacitance per unit area is changed as follows.

$$C_{ox}' = \frac{\epsilon_{ox}}{t_{ox}'} = S \cdot \frac{\epsilon_{ox}}{t_{ox}} = S \cdot C_{ox} \quad \text{.....(81)}$$

Full Scaling (Constant-Field Scaling)

- The aspect ratio W/L of the MOSFET will remain unchanged under scaling.
- Consequently, the transconductance parameter k_n will also be scaled by a factor of S .
- Since all terminal voltages are scaled down by the factor S as well, the linear-mode drain current of the scaled MOSFET can now be found as:

Full Scaling (Constant-Field Scaling)

$$\begin{aligned} I_D'(lin) &= \frac{k_n'}{2} \cdot [2 \cdot (V_{GS}' - V_T') \cdot V_{DS}' - V_{DS}'^2] \\ &= \frac{S \cdot k_n}{2} \cdot \frac{1}{S^2} \cdot [2 \cdot (V_{GS} - V_T) \cdot V_{DS} - V_{DS}^2] = \frac{I_D(lin)}{S} \quad \dots(82) \end{aligned}$$

Full Scaling (Constant-Field Scaling)

- Similarly, the saturation-mode drain current is also reduced by the same scaling factor.

$$I_D'(sat) = \frac{k_n'}{2} \cdot (V_{GS}' - V_T')^2 = \frac{S \cdot k_n}{2} \cdot \frac{1}{S^2} \cdot (V_{GS} - V_T)^2 = \frac{I_D(sat)}{S} \quad (83)$$

Full Scaling (Constant-Field Scaling)

- Now consider the power dissipation of the MOSFET (before scaling) can be found as.

$$P = I_D \cdot V_{DS} \quad \text{.....(84)}$$

- After scaling it becomes;

$$P' = I_D' \cdot V_{DS}' = \frac{1}{S^2} \cdot I_D \cdot V_{DS} = \frac{P}{S^2} \quad \text{.....(85)}$$

Full Scaling (Constant-Field Scaling)

- This significant reduction of the power dissipation is one of the most attractive features of full scaling.
- The changes in key device characteristics as a result of full (constant-field) scaling is summarized as follows:

Full Scaling (Constant-Field Scaling)

Quantity	Before Scaling	After Scaling
Oxide capacitance	C_{ox}	$C_{ox}' = S \cdot C_{ox}$
Drain current	I_D	$I_D' = I_D / S$
Power dissipation	P	$P' = P / S^2$
Power density	$P / Area$	$P' / Area' = P / Area$

Constant-Voltage Scaling

- While the full scaling strategy dictates that the power supply voltage and all terminal voltages be scaled down proportionally with the device dimensions, the scaling of voltages may not be very practical in many cases.
- In particular, the peripheral and interface circuitry may require certain voltage levels for all input and output voltages, which in turn would necessitate multiple power supply voltages and complicated level shifter arrangements.

Constant-Voltage Scaling

- For these reasons, constant-voltage scaling is usually preferred over full scaling.
- In constant-voltage scaling, all dimensions of the MOSFET are reduced by a factor of S , as in full scaling. The power supply voltage and the terminal voltages, on the other hand, remain unchanged.
- The doping densities must be increased by a factor of S^2 in order to preserve the charge-field relations.

Constant-Voltage Scaling

- The constant-voltage scaling of key dimensions, voltages, and densities is shown as shown in the Table of the next slide.
- The gate oxide capacitance per unit area C_{ox} is increased by a factor of S , which means that the transconductance parameter is also increased by S .

Constant-Voltage Scaling

Quantity	Before Scaling	After Scaling
Dimensions	W, L, t_{ox}, x_j	reduced by S ($W' = W / S, \dots$)
Voltages	V_{DD}, V_T	remain unchanged
Doping densities	N_A, N_D	increased by S^2 ($N_A' = S^2 \cdot N_A, \dots$)

Constant-Voltage Scaling

- Since the terminal voltages remain unchanged, the linear mode drain current and the saturation current of the scaled MOSFET can be written as:

$$\begin{aligned} I_D'(lin) &= \frac{k_n'}{2} \cdot [2 \cdot (V_{GS}' - V_T') \cdot V_{DS}' - V_{DS}'^2] \\ &= \frac{S \cdot k_n}{2} \cdot [2 \cdot (V_{GS} - V_T) \cdot V_{DS} - V_{DS}^2] = S \cdot I_D(lin) \end{aligned} \quad \dots(86)$$

$$I_D'(sat) = \frac{k_n'}{2} \cdot (V_{GS}' - V_T')^2 = \frac{S \cdot k_n}{2} \cdot (V_{GS} - V_T)^2 = S \cdot I_D(sat) \quad \dots(87)$$

Constant-Voltage Scaling

- The power dissipation of the MOSFET increases by a factor of S.

$$P' = I_D' \cdot V_{DS}' = (S \cdot I_D) \cdot V_{DS} = S \cdot P \quad \text{.....(88)}$$

Constant-Voltage Scaling

Quantity	Before Scaling	After Scaling
Oxide capacitance	C_{ox}	$C_{ox}' = S \cdot C_{ox}$
Drain current	I_D	$I_D' = S \cdot I_D$
Power dissipation	P	$P' = S \cdot P$
Power density	$P / Area$	$P' / Area' = S^3 \cdot (P / Area)$

Advantages of Constant-Voltage Scaling

- To summarize, constant-voltage scaling may be preferred over full (constant-field) scaling in many practical cases because of the external voltage-level constraints.

Drawbacks of Constant-Voltage Scaling

- It must be recognized, however, that constant-voltage scaling increases the drain current density and the power density by a factor of S^3 .
- This large increase in current and power densities may eventually cause serious reliability problems for the scaled transistor, such as electromigration, hot-carrier degradation, oxide breakdown, and electrical over-stress.

small-geometry effects

Short-Channel Effects

- A MOS transistor is called a short-channel device if the effective length is given as L_{eff} is approximately equal to the source and drain junction depths given as x_j and the type of the device is a MOSFET short channel device.

$$L_{\text{eff}} \approx x_j \quad \text{.....(89)}$$

Short-Channel Effects

- The short-channel effects are attributed to two physical phenomena:
- (i) the limitations imposed on electron drift characteristics in the channel.
- (ii) the modification of the threshold voltage due to the shortening channel length.

Short-Channel Effects

- The lateral electric field E_y along the channel increases, as the effective channel length L_{eff} is decreased.
- The electron drift velocity v_d in the channel is proportional to the electric field for lower field values.
- This drift velocity tends to saturate at high channel electric fields.

Short-Channel Effects

- For channel electric fields of $E_y = 10^5$ V/cm and higher, the electron drift velocity in the channel reaches a saturation value of about $v_d(\text{sat}) = 10^7$ cm/s.
- This affects the drain current as;

$$I_D(\text{sat}) = W \cdot v_d(\text{sat}) \cdot \int_0^{L_{\text{eff}}} q \cdot n(x) dx = W \cdot v_d(\text{sat}) \cdot |Q_I| \quad \text{.....(90)}$$

Short-Channel Effects

- Since the channel-end voltage is equal to V_{DSAT} , the saturation current can be found as follows:

$$I_D(sat) = W \cdot v_d(sat) \cdot C_{ox} \cdot V_{DSAT} \quad \text{.....(91)}$$

Short-Channel Effects

- In short-channel MOS transistors, the carrier velocity in the channel is also a function of the normal (vertical) electric-field component E_x .
- Since the vertical field influences the scattering of carriers (collisions suffered by the carriers) in the surface region, the surface mobility is reduced with respect to the bulk mobility.

Short-Channel Effects

- The dependence of the surface electron mobility on the vertical electric field can be expressed by the following empirical formula:

$$\mu_n(eff) = \frac{\mu_{no}}{1 + \Theta \cdot E_x} = \frac{\mu_{no}}{1 + \frac{\Theta \epsilon_{ox}}{t_{ox} \epsilon_{Si}} \cdot (V_{GS} - V_c(y))} \dots\dots(92)$$

- where μ_{no} is the low-field surface electron mobility and Θ is an empirical factor.

Short-Channel Effects

- For a simple estimation of field-related mobility reduction, (Eqn. 92) can be approximated by

$$\mu_n(eff) = \frac{\mu_{no}}{1 + \eta \cdot (V_{GS} - V_T)} \quad \text{.....(93)}$$

- where η is an empirical coefficient.

Modification of Threshold Voltage

- Now, let us consider the modification of the threshold voltage due to short-channel effects.
- The threshold voltage expression(in Eqn. 43) was derived for a long-channel MOSFET. Specifically, the channel depletion region was assumed to be created only by the applied gate voltage, and the depletion regions associated with the drain and source pn-junctions were neglected.

Modification of Threshold Voltage

- The shape of this gate-induced bulk (channel) depletion region was assumed to be rectangular, extending from the source to the drain.
- In short-channel MOS transistors, the n^+ drain and source diffusion regions in the p-type substrate induce a significant amount of depletion charge.

Modification of Threshold Voltage

- Consequently, the long-channel threshold voltage expression derived earlier overestimates, the depletion charge supported by the gate voltage.
- So the threshold voltage value found by using (Eqn. 43) is therefore larger than the actual threshold voltage of the short-channel MOSFET.

Modification of Threshold Voltage

- Figure 32 shows the simplified geometry of the gate-induced bulk depletion region and the pn-junction depletion regions in a short-channel MOS transistor whereas Fig. 33 represents a close-up view of the drain diffusion edge.
- Here the bulk depletion region is assumed to have an asymmetric trapezoidal shape, instead of a rectangular shape, to represent accurately the gate-induced charge.

Modification of Threshold Voltage

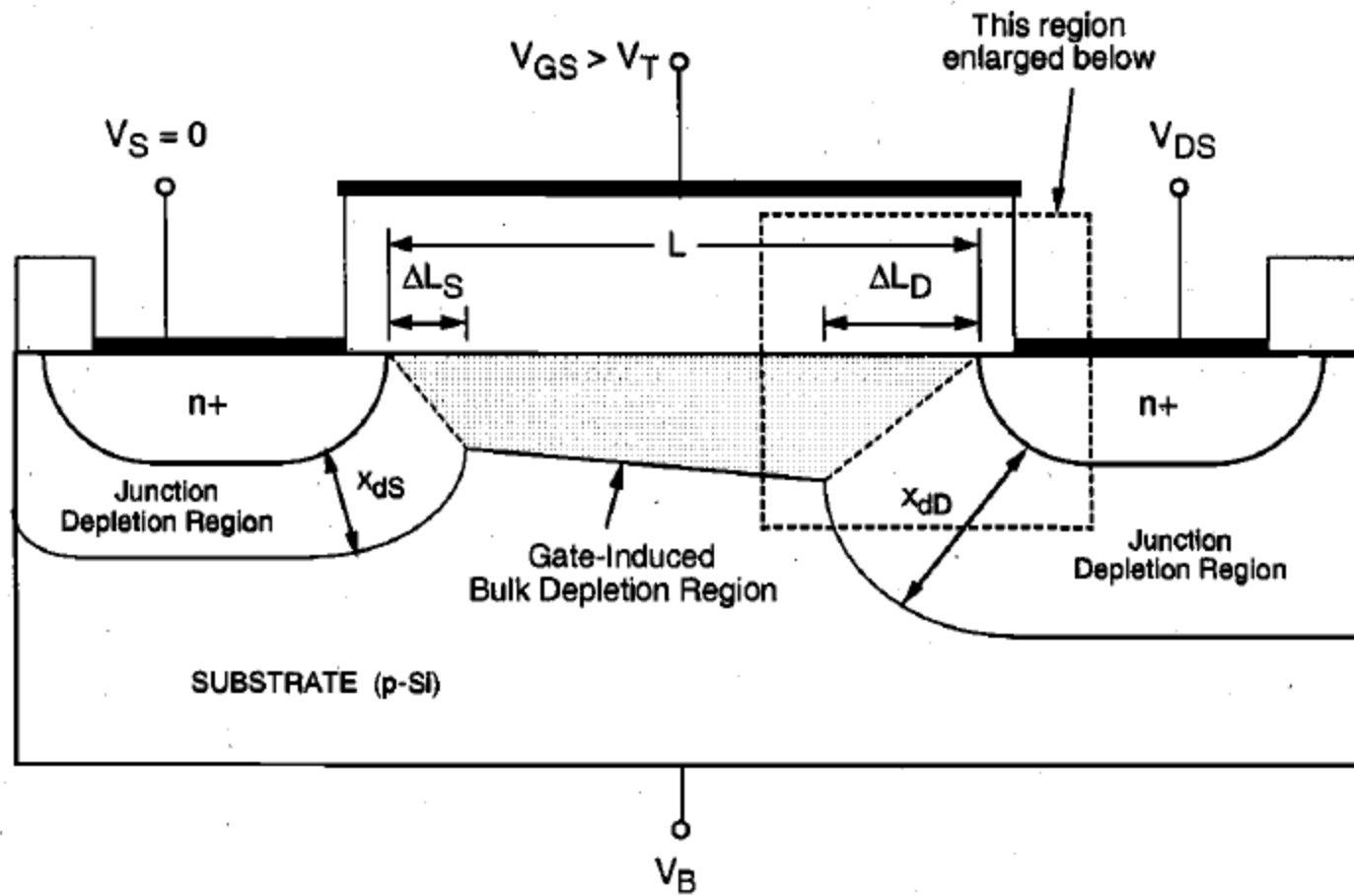


Fig. 32: Simplified geometry of MOSFET Channel Region

Modification of Threshold Voltage

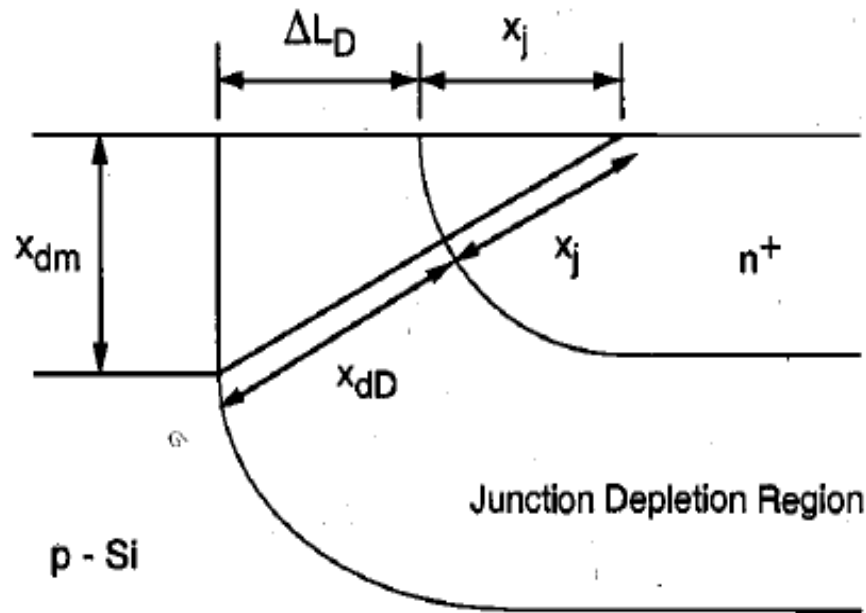


Fig. 33: a close-up view of the drain diffusion edge.

Modification of Threshold Voltage

- The drain depletion region is larger than the source depletion region because the positive drain-to-source voltage reverse-biases the drain-substrate junction.
- Following the modification of the bulk charge term, the threshold voltage of the short-channel MOSFET can be written as;

$$V_{T0}(\text{short channel}) = V_{T0} - \Delta V_{T0} \quad \text{.....(94)}$$

Modification of Threshold Voltage

- where V_{T0} is the zero-bias threshold voltage calculated using the conventional long channel formula (Eqn. 43) and ΔV_{T0} is the threshold voltage shift (reduction) due to the short-channel effect.
- The reduction term actually represents the amount of charge differential between a rectangular depletion region and a trapezoidal depletion region.

Modification of Threshold Voltage

- Let ΔL_S and ΔL_D represent the lateral extent of the depletion regions associated with the source junction and the drain junction, respectively.
- Then, the bulk depletion region charge contained within the trapezoidal region is;

$$Q_{B0} = -\left(1 - \frac{\Delta L_S + \Delta L_D}{2L}\right) \cdot \sqrt{2 \cdot q \cdot \epsilon_{Si} \cdot N_A \cdot |2\phi_F|} \quad \text{.....(95)}$$

Modification of Threshold Voltage

- ΔL_S and ΔL_D , can be calculated using the simplified geometry shown in Fig. 33.
- Here, x_{dS} and x_{dD} represent the depth of the pn-junction depletion regions associated with the source and the drain, respectively.
- The edges of the source and drain diffusion regions are represented by quarter-circular arcs, each with a radius equal to the junction depth, x_j .

Modification of Threshold Voltage

- The vertical extent of the bulk depletion region into the substrate is represented by x_{dm} . The junction depletion region depths can be approximated by;

$$x_{dS} = \sqrt{\frac{2 \cdot \epsilon_{Si}}{q \cdot N_A} \cdot \phi_0} \quad \text{.....(96)}$$

$$x_{dD} = \sqrt{\frac{2 \cdot \epsilon_{Si}}{q \cdot N_A} \cdot (\phi_0 + V_{DS})} \quad \text{.....(97)}$$

Modification of Threshold Voltage

- With the junction built in voltage

$$\phi_0 = \frac{kT}{q} \cdot \ln\left(\frac{N_D \cdot N_A}{n_i^2}\right) \dots\dots\dots(98)$$

- By applying Pythagoras theorem;

$$(x_j + x_{dD})^2 = x_{dm}^2 + (x_j + \Delta L_D)^2 \dots\dots(99)$$

Modification of Threshold Voltage

$$\Delta L_D^2 + 2 \cdot x_j \cdot \Delta L_D + x_{dm}^2 - x_{dD}^2 - 2 \cdot x_j \cdot x_{dD} = 0 \quad \dots(100)$$

$$\Delta L_D = -x_j + \sqrt{x_j^2 - (x_{dm}^2 - x_{dD}^2) + 2x_j x_{dD}} \cong x_j \cdot \left(\sqrt{1 + \frac{2x_{dD}}{x_j}} - 1 \right) \quad \dots(101)$$

Modification of Threshold Voltage

- Similarly

$$\Delta L_S \cong x_j \cdot \left(\sqrt{1 + \frac{2x_{dS}}{x_j}} - 1 \right) \quad \dots(102)$$

- Now, the amount of threshold voltage reduction ΔV_{T0} due to short-channel effects can be found as:

$$\Delta V_{T0} = \frac{1}{C_{ox}} \cdot \sqrt{2q\epsilon_{Si} N_A |2\phi_F|} \cdot \frac{x_j}{2L} \cdot \left[\left(\sqrt{1 + \frac{2x_{dS}}{x_j}} - 1 \right) + \left(\sqrt{1 + \frac{2x_{dD}}{x_j}} - 1 \right) \right] \quad \dots\dots\dots(103)$$

Modification of Threshold Voltage

- The threshold voltage shift term is proportional to (x_j / L) .
- As a result, this term becomes more prominent for MOS transistors with shorter channel lengths, and it approaches zero for long-channel MOSFETs where $L \gg x$.

Problem 10

- Consider an n-channel MOS process with the following parameters:
- substrate doping density $N_A = 10^{16} \text{ cm}^{-3}$, polysilicon gate doping density $N_D (\text{gate}) = 2 \times 10^{20} \text{ cm}^{-3}$, gate oxide thickness $t_{\text{ox}} = 50 \text{ nm}$, oxide-interface fixed charge density $N_{\text{ox}} = 4 \times 10^{10} \text{ cm}^{-2}$, and source and drain diffusion doping density $N_D = 10^{17} \text{ cm}^{-3}$.

Problem 10

- In addition, the channel region is implanted with p-type impurities (impurity concentration $N_i = 2 \times 10^{11} \text{ cm}^{-2}$) to adjust the threshold voltage.
- The junction depth of the source and drain diffusion regions is $x_j = 1.0 \text{ } \mu\text{m}$.
- Plot the variation of the zero-bias threshold voltage V_{T0} as a function of the channel length (assume that $V_{DS} = V_{SB} = 0$). Also find V_{T0} for $L = 0.7 \text{ } \mu\text{m}$, $V_{DS} = 5 \text{ V}$, and $V_{SB} = 0$.

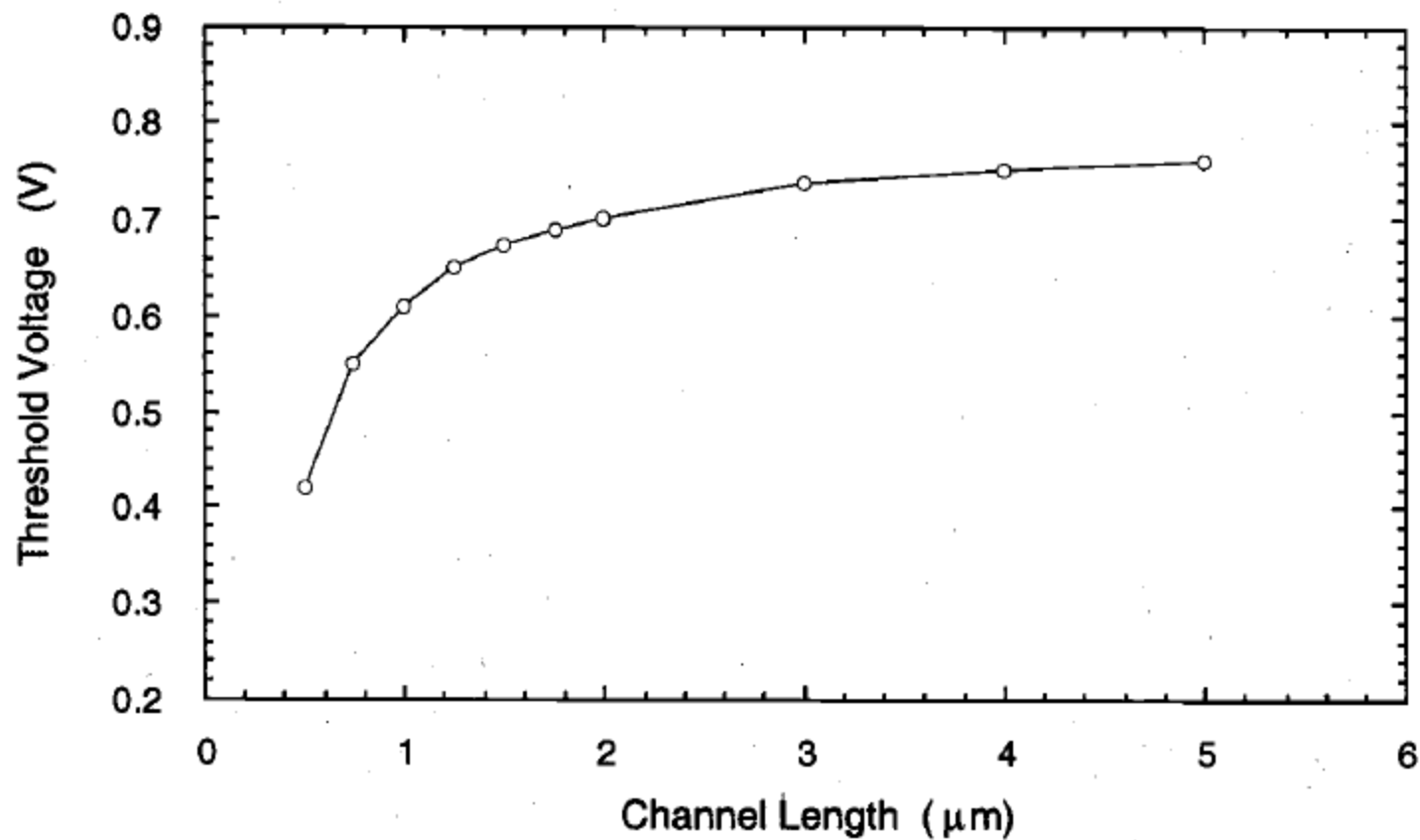
Solution

- First, find out the zero-bias threshold voltage using the conventional formula. The threshold voltage without the channel implant was already calculated for the same process parameters in Problem 2.
- The additional p-type channel implant will increase the threshold voltage by an amount of qN_i / C_{ox} .

Solution

- The following plot shows the variation of the threshold voltage with the channel length.
- The threshold voltage decreases by as much as 50% for channel lengths in the submicronrange, while it approaches the value of 0.8 V for larger channel lengths.

Solution



Narrow-Channel Effects

- MOS transistors that have channel widths W on the same order of magnitude as the maximum depletion region thickness x_{dm} are defined as narrow-channel devices.
- Here the threshold voltage of such device is larger than that predicted by conventional threshold voltage;

$$V_{T0}(\text{narrow channel}) = V_{T0} + \Delta V_{T0} \quad \text{.....(104)}$$

Narrow-Channel Effects

- The additional contribution to the threshold voltage due to narrow-channel effects can be modeled as follows:

$$\Delta V_{T0} = \frac{1}{C_{ox}} \cdot \sqrt{2 q \epsilon_{Si} N_A |2 \phi_F|} \cdot \frac{K \cdot x_{dm}}{W} \quad \text{.....(105)}$$

Narrow-Channel Effects

- where K is an empirical parameter depending on the shape of the fringe depletion region.
- Assuming that the depletion region edges are modeled by quarter-circular arcs, for example, the parameter K can be found as;

$$K = \frac{\pi}{2} \quad \dots(106)$$

Narrow-Channel Effects

- Finally, for minimum-geometry MOSFETs which have a small channel length and a small channel width, the threshold voltage variations due to short- and narrow-channel effects may tend to cancel each other out.

MOSFET Capacitances

- In order to examine the transient (AC) response of MOSFETs and digital circuits consisting of MOSFETs, here the nature and the amount of parasitic capacitances associated with the MOS transistor are described.
- A simple approximations for the on-chip MOSFET capacitances that can be used in most hand calculations is developed.

MOSFET Capacitances

- These capacitance models are sufficiently accurate to represent the crucial characteristics of MOSFET charge-voltage behavior, and the equations are all based on fundamental semiconductor device theory.

MOSFET Capacitances

- Figure 34 shows the cross-sectional view and the top view (mask view) of an n-channel MOSFET.
- Until now, we concentrated on the cross-sectional view of the device, since we were primarily concerned with the flow of carriers within the MOSFET.
- To study the parasitic device capacitances, the top view of the MOSFET is required.

MOSFET Capacitances

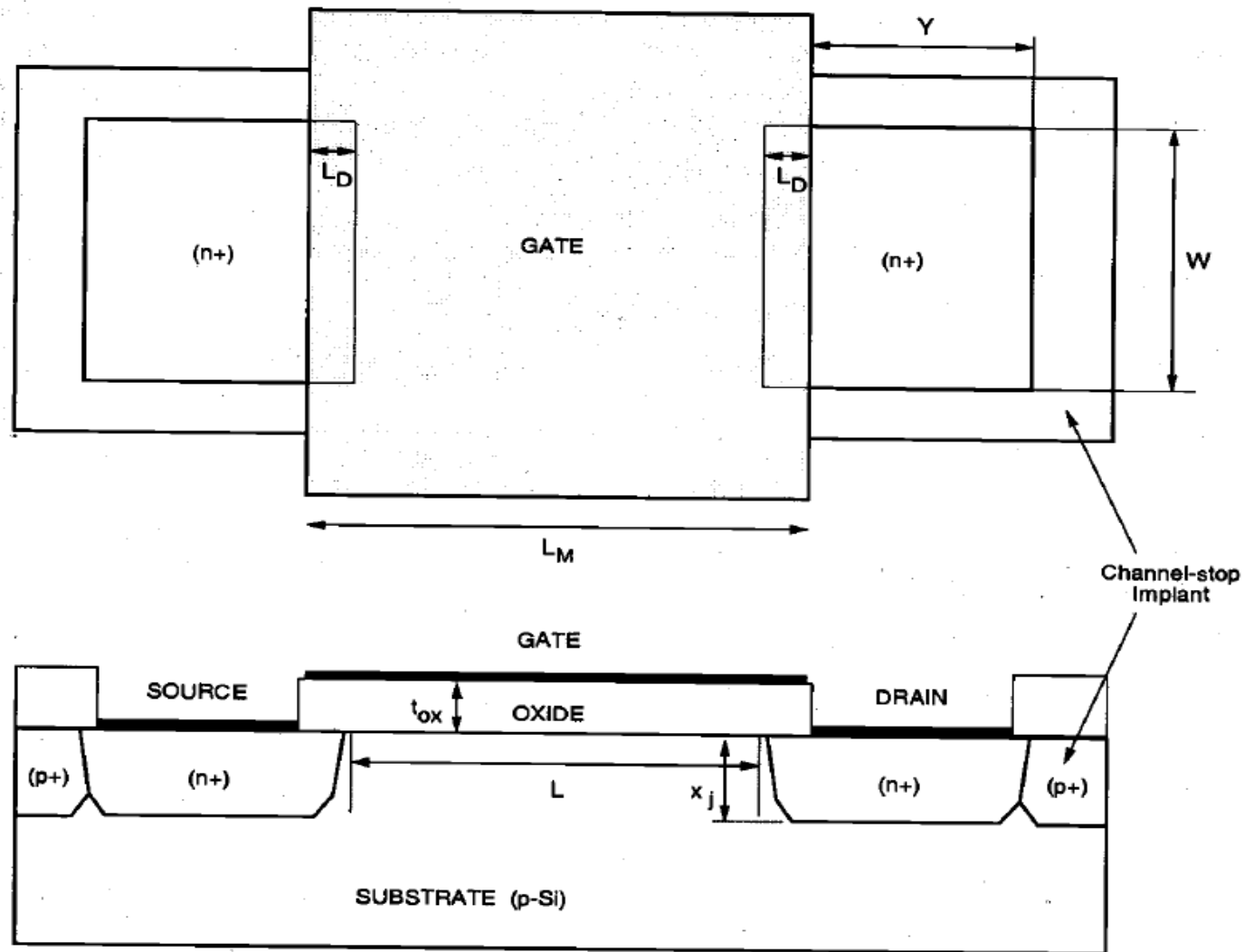


Figure 34. Cross-sectional view and top view (mask view) of a typical n-channel MOSFET.

MOSFET Capacitances

- In Figure 34, the mask length (drawn length) of the gate is indicated by L_M , and the actual channel length is indicated by L . The extent of both the gate-source and the gate-drain overlap are L_D ; thus, the channel length is given by;

$$L = L_M - 2 \cdot L_D \quad \text{.....(107)}$$

MOSFET Capacitances

- The source and drain overlap region lengths are usually equal to each other because of the symmetry of the MOSFET structure.
- Typically, L_D is on the order of $0.1 \mu\text{m}$. Both the source and the drain diffusion regions have a width of W .
- The typical diffusion region length is denoted by Y .

MOSFET Capacitances

- Lumped representation of the parasitic MOSFET capacitances are represented in Figure 35.
- Based on their physical origins, the parasitic device capacitances can be classified into two major groups: **oxide-related capacitances and junction capacitances.**

MOSFET Capacitances

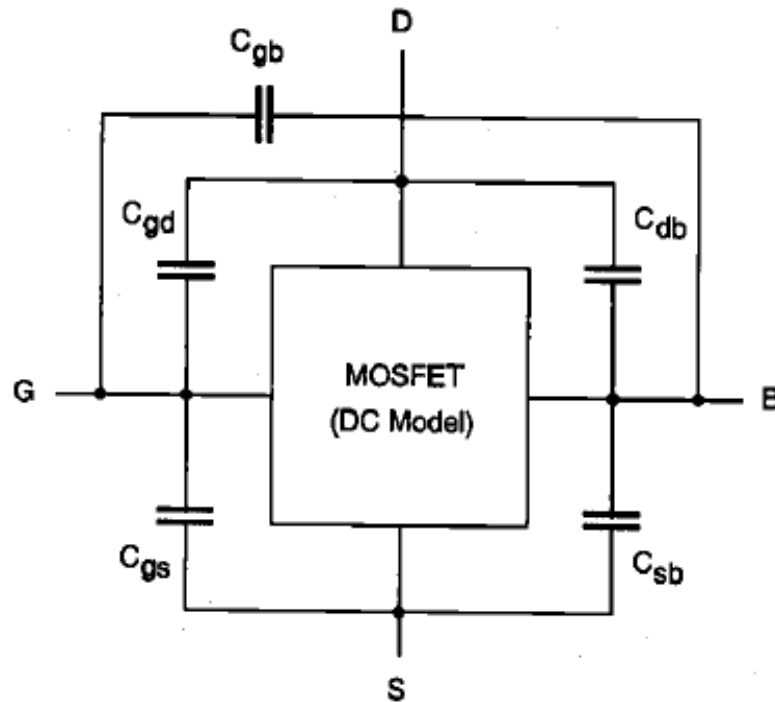


Fig. 35: Lumped representation of the parasitic MOSFET capacitances.

Oxide-related Capacitances

- It was shown earlier that the gate electrode overlaps both the source region and the drain region at the edges.
- The two overlap capacitances that arise as a result of this structural arrangement are called C_{GD} (overlap) and C_{GS} (overlap), respectively.
- Assuming that both the source and the drain diffusion regions have the same width W , the overlap capacitances can be found as;

Oxide-related Capacitances

$$C_{GS}(\text{overlap}) = C_{ox} \cdot W \cdot L_D \quad \dots(108)$$

$$C_{GD}(\text{overlap}) = C_{ox} \cdot W \cdot L_D$$

$$C_{ox} = \frac{\epsilon_{ox}}{t_{ox}} \quad \dots(109)$$

- both of these overlap capacitances do not depend on the bias conditions, i.e., they are voltage-independent.

Oxide-related Capacitances

- Now let us consider the capacitances which result from the interaction between the gate voltage and the channel charge.
- Since the channel region is connected to the source, the drain, and the substrate, we can identify three capacitances between the gate and these regions, i.e., C_{gs} , C_{gd} and C_{gb} respectively.
- The gate-to-channel capacitance is distributed and voltage-dependent.

C_{gs}, C_{gd}, C_{gb} in cut-off mode

- In cut-off mode (Fig. 36), the surface is not inverted and there is no conducting channel that links the surface to the source and to the drain.
- Therefore, the gate-to-source and the gate-to-drain capacitances are both equal to zero:
- $C_{gs} = C_{gd} = 0$ (110)
- The gate-to-substrate capacitance can be approximated by

$$C_{gb} = C_{ox} \cdot W \cdot L \quad \text{.....(111)}$$

C_{gs}, C_{gd}, C_{gb} in cut-off mode

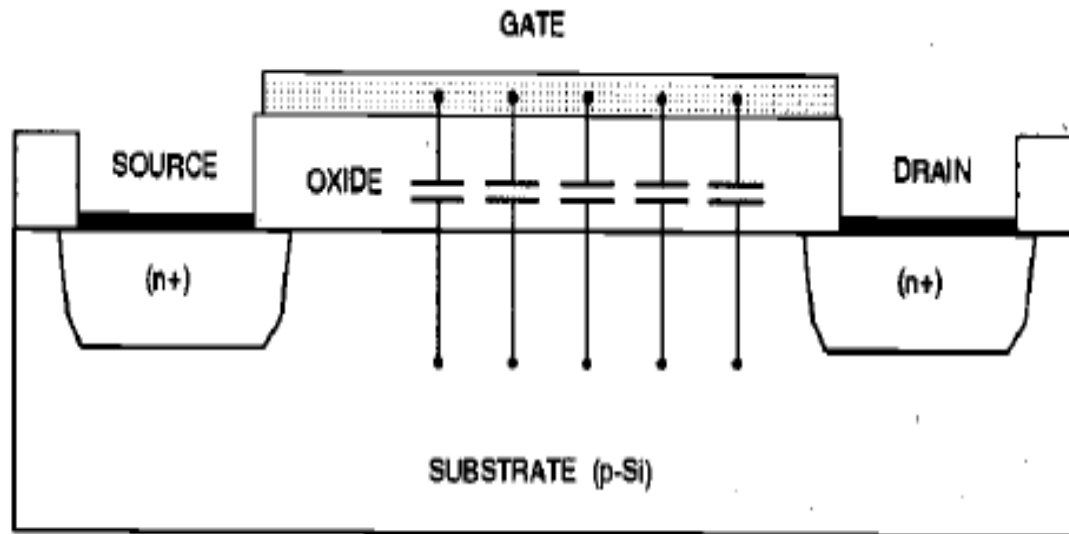


Fig. 36: Cut-off Mode

C_{gs} , C_{gd} , C_{gb} in linear-mode

- In linear-mode operation, the inverted channel extends across the MOSFET, between the source and the drain (Fig. 37).
- This conducting inversion layer on the surface shields the substrate from the gate electric field; thus, $C_{gb} = 0$.
- In this case, the distributed gate-to-channel capacitance may be viewed as being shared equally between the source and the drain.

C_{gs}, C_{gd}, C_{gb} in linear-mode

- Thus

$$C_{gs} \cong C_{gd} \cong \frac{1}{2} \cdot C_{ox} \cdot W \cdot L \quad \text{.....(112)}$$

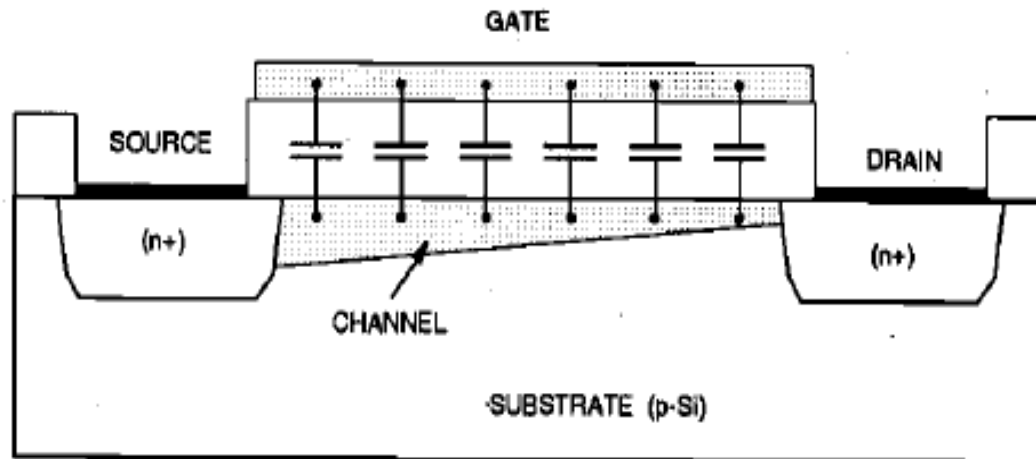


Fig. 37: Linear Mode

C_{gs} , C_{gd} , C_{gb} in Saturation mode

- When the MOSFET is operating in saturation mode, the inversion layer on the surface does not extend to the drain, but it is pinched off (Fig. 38).
- The gate-to-drain capacitance component is therefore equal to zero ($C_{gd} = 0$).
- Since the source is still linked to the conducting channel, its shielding effect also forces the gate-to-substrate capacitance to be zero, $C_{gb} = 0$.

C_{gs}, C_{gd}, C_{gb} in Saturation mode

- Finally, the distributed gate-to-channel capacitance as seen between the gate and the source can be approximated by

$$C_{gs} \cong \frac{2}{3} \cdot C_{ox} \cdot W \cdot L \quad \dots(113)$$

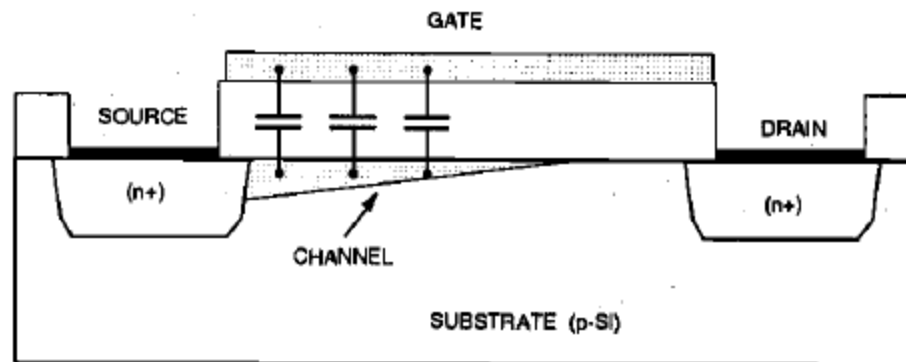


Fig. 38: Saturation Mode

Oxide-related Capacitances

- The following Table lists a summary of the approximate oxide capacitance values in three different operating modes of the MOSFET.

Capacitance	Cut-off	Linear	Saturation
$C_{gb} \text{ (total)}$	$C_{ox} WL$	0	0
$C_{gd} \text{ (total)}$	$C_{ox} WL_D$	$\frac{1}{2} C_{ox} WL + C_{ox} WL_D$	$C_{ox} WL_D$
$C_{gs} \text{ (total)}$	$C_{ox} WL_D$	$\frac{1}{2} C_{ox} WL + C_{ox} WL_D$	$\frac{2}{3} C_{ox} WL + C_{ox} WL_D$

Oxide-related Capacitances

- The variation of the distributed parasitic oxide capacitances as functions of the gate-to-source voltage is also shown in Fig. 39.

Oxide-related Capacitances

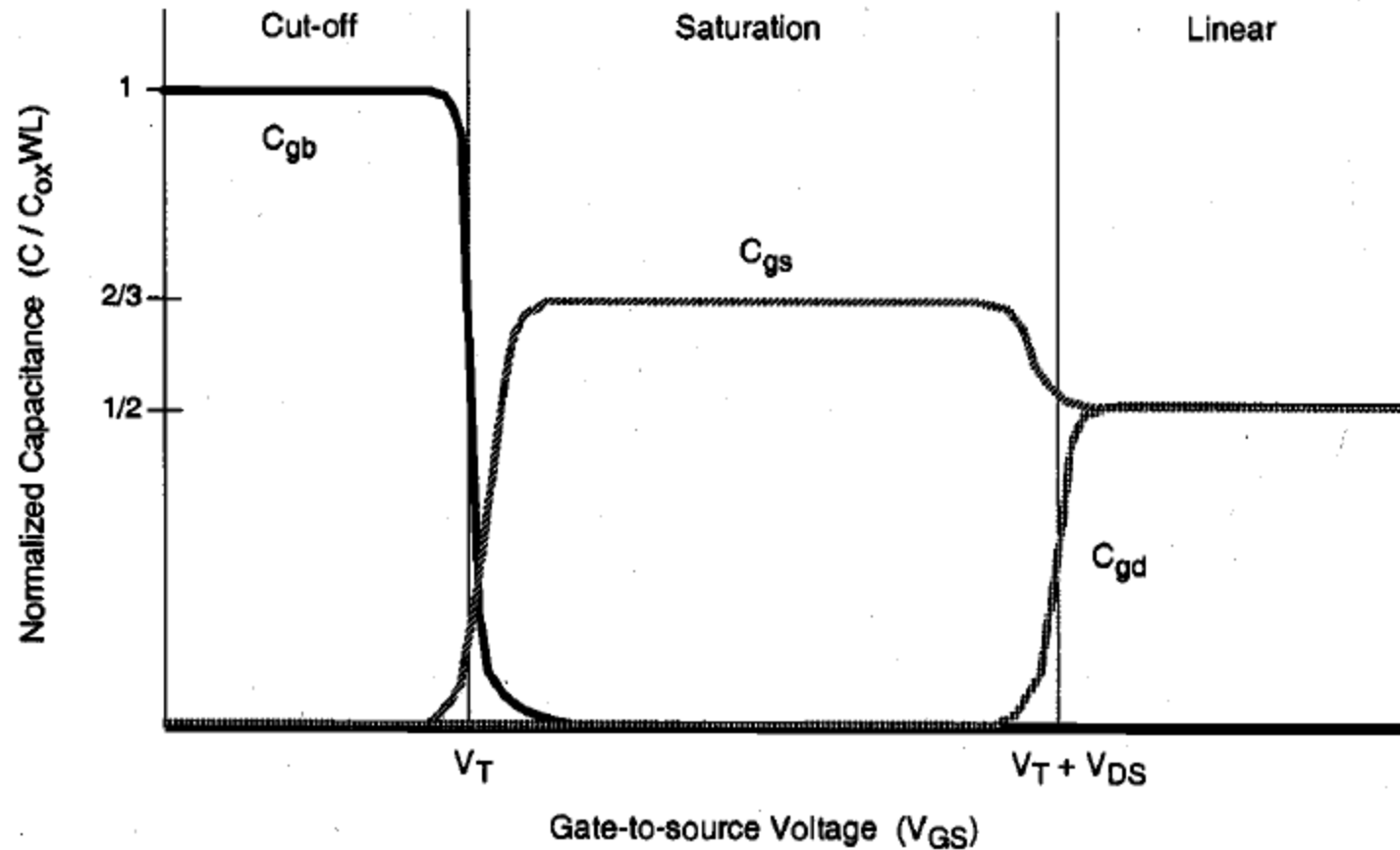


Fig. 39

Oxide-related Capacitances

- It is also observed that the sum of all three voltage dependent(distributed) gate oxide capacitances ($C_{gb} + C_{gs} + C_{gd}$) has a minimum value of $0.66 C_{ox}WL$ (in saturation mode) and a maximum value of $C_{ox}WL$ (in cut-off and linear modes).
- For simple hand calculations where all three capacitances can be considered to be connected in parallel, a constant worst-case value of $C_{ox}W (L+2L_D)$ can be used for the sum of MOSFET gate oxide capacitances.

Junction Capacitances

- Now let us consider the voltage-dependent source-substrate and drain-substrate junction capacitances, C_{sb} and C_{db} , respectively.
- Both of these capacitances are due to the depletion charge surrounding the respective source or drain diffusion regions embedded in the substrate.

Junction Capacitances

- Both of these junctions are reverse-biased under normal operating conditions of the MOSFET and that the amount of junction capacitance is a function of the applied terminal voltages.
- Figure 40 shows the simplified, partial geometry of a typical n-channel enhancement MOSFET, focusing on the n-type diffusion region within the p-type substrate.

Junction Capacitances

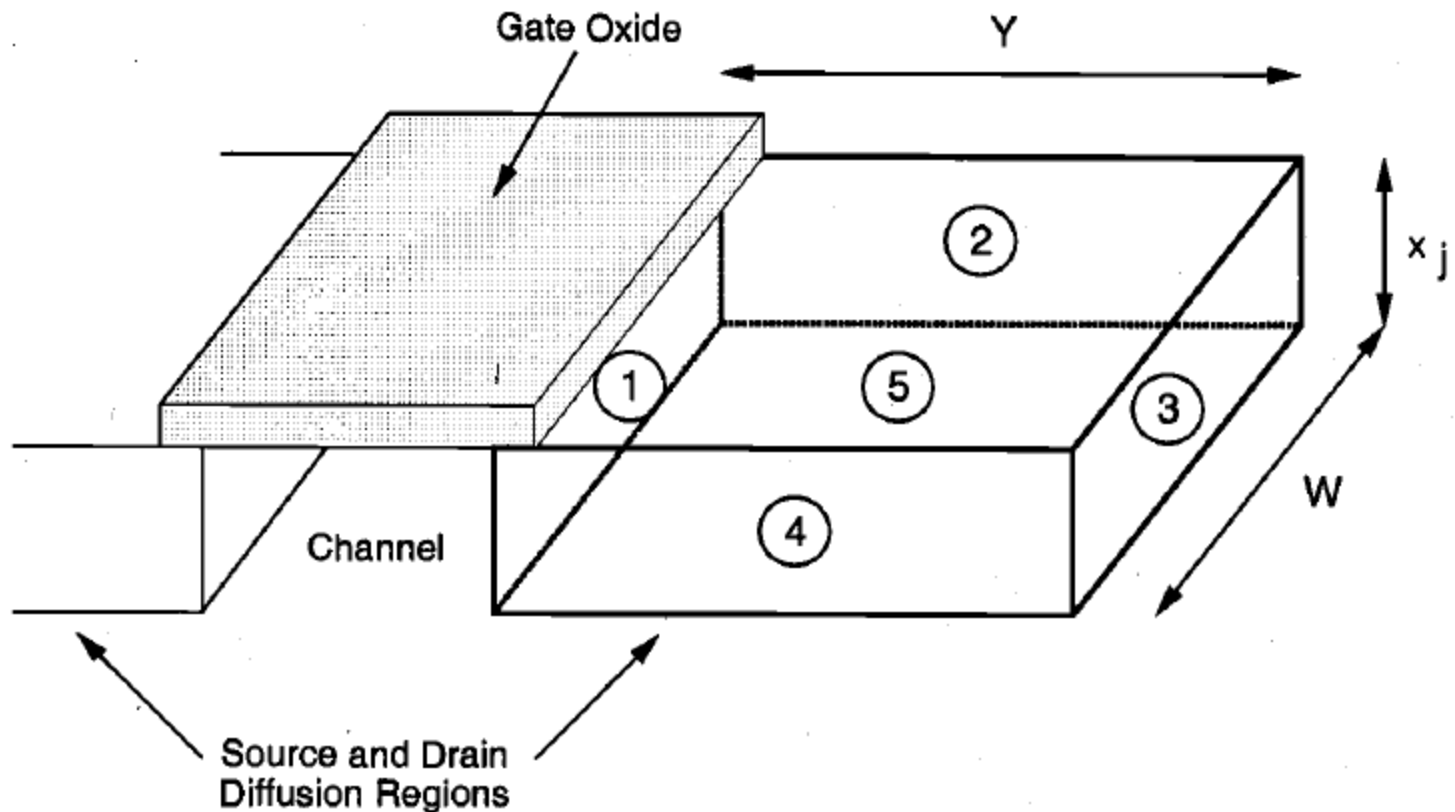


Figure 40: Three-dimensional view of the n^+ diffusion region within the p-type substrate

Junction Capacitances

- As seen in Fig. 40, the n^+ diffusion region forms a number of planar pn-junctions with the surrounding p-type substrate, indicated here with 1 through 5.
- The dimensions of the rectangular box representing the diffusion region are given as W , Y , and x_j .

Junction Capacitances

- Comparing this three-dimensional view with Fig. 34, we recognize that three of the five planar junctions shown here (2, 3, and 4) are actually surrounded by the p+ channel-stop implant.
- The junction labeled (1) is facing the channel, and the bottom junction (5) is facing the p-type substrate, which has a doping density of N_A .

Junction Capacitances

- Since the p+ channel-stop implant density is usually about $10N_A$, the junction capacitances associated with these sidewalls will be different from the other junction capacitances (as shown in the Table of next slide).

Junction Capacitances

Junction	Area	Type
1	$W \cdot x_j$	n^+ / p
2	$Y \cdot x_j$	n^+ / p^+
3	$W \cdot x_j$	n^+ / p^+
4	$Y \cdot x_j$	n^+ / p^+
5	$W \cdot Y$	n^+ / p

- To calculate the depletion capacitance of a reverse-biased abrupt pn-junction, let us consider first the depletion region thickness, x_d .

Junction Capacitances

- Assume that the n-type and p-type doping densities are given by N_D and N_A , respectively, and that the reverse bias voltage is given by V (negative), the depletion region thickness can be found as follows:

$$x_d = \sqrt{\frac{2 \cdot \epsilon_{Si}}{q} \cdot \frac{N_A + N_D}{N_A \cdot N_D} \cdot (\phi_0 - V)} \quad \dots(114)$$

Junction Capacitances

- where the built-in junction potential is calculated as;

$$\phi_0 = \frac{kT}{q} \cdot \ln\left(\frac{N_A \cdot N_D}{n_i^2}\right) \quad \dots(115)$$

- The junction is forward-biased for a positive bias voltage V , and reverse-biased for a negative bias voltage.

Junction Capacitances

- The depletion-region charge stored in this area can be written in terms of the depletion region thickness, x_d .

$$Q_j = A \cdot q \cdot \left(\frac{N_A \cdot N_D}{N_A + N_D} \right) \cdot x_d = A \sqrt{2 \cdot \epsilon_{Si} \cdot q \cdot \left(\frac{N_A \cdot N_D}{N_A + N_D} \right) \cdot (\phi_0 - V)} \quad ..(116)$$

- where, A indicates the junction area.

Junction Capacitances

- The junction capacitance associated with the depletion region is defined as;

$$C_j = \left| \frac{dQ_j}{dV} \right| \quad \text{.....(117)}$$

- By differentiating (Eqn. 117) with respect to the bias voltage V , we can now obtain the expression for the junction capacitance as follows.

$$C_j(V) = A \cdot \sqrt{\frac{\epsilon_{Si} \cdot q}{2} \cdot \left(\frac{N_A \cdot N_D}{N_A + N_D} \right)} \cdot \frac{1}{\sqrt{\phi_0 - V}} \quad \text{....(118)}$$

Junction Capacitances

- This expression can be rewritten in a more general form, to account for the junction grading.

$$C_j(V) = \frac{A \cdot C_{j0}}{\left(1 - \frac{V}{\phi_0}\right)^m} \quad \text{.....(119)}$$

Junction Capacitances

- The parameter m in (Eqn. 119) is called the grading coefficient. Its value is equal to $1/2$ for an abrupt junction profile, and $1/3$ for a linearly graded junction profile.
- For an abrupt pn-junction profile, i.e., for $m = 1/2$, the equations (118) and (119) become identical.

Junction Capacitances

- The zero-bias junction capacitance per unit area C_{j0} is defined as

$$C_{j0} = \sqrt{\frac{\epsilon_{Si} \cdot q}{2} \cdot \left(\frac{N_A \cdot N_D}{N_A + N_D} \right) \cdot \frac{1}{\phi_0}} \quad \dots(120)$$

Junction Capacitances

- The problem of estimating capacitance values under changing bias conditions can be simplified, if we calculate a large-signal average (linear) junction capacitance instead, which, by definition, is independent of the bias potential.
- This equivalent large-signal capacitance can be defined as follows:

Junction Capacitances

$$C_{eq} = \frac{\Delta Q}{\Delta V} = \frac{Q_j(V_2) - Q_j(V_1)}{V_2 - V_1} = \frac{1}{V_2 - V_1} \cdot \int_{V_1}^{V_2} C_j(V) dV \quad \dots\dots(121)$$

- Here, the reverse bias voltage across the pn-junction is assumed to change from V_1 to V_2 .
- Hence, the equivalent capacitance C_{eq} is always calculated for a transition between two known voltage levels.

Junction Capacitances

- By substituting (119) into (121), we obtain

$$C_{eq} = -\frac{A \cdot C_{j0} \cdot \phi_0}{(V_2 - V_1) \cdot (1 - m)} \cdot \left[\left(1 - \frac{V_2}{\phi_0} \right)^{1-m} - \left(1 - \frac{V_1}{\phi_0} \right)^{1-m} \right] \dots(122)$$

- For the special case of abrupt pn-junctions, equation (122) becomes

$$C_{eq} = -\frac{2 \cdot A \cdot C_{j0} \cdot \phi_0}{(V_2 - V_1)} \cdot \left[\sqrt{1 - \frac{V_2}{\phi_0}} - \sqrt{1 - \frac{V_1}{\phi_0}} \right] \dots(123)$$

Junction Capacitances

- This equation can be rewritten in a simpler form by defining a dimensionless coefficient K_{eq} as follows:

$$C_{eq} = A \cdot C_{j0} \cdot K_{eq} \quad \text{.....(124)}$$

$$K_{eq} = -\frac{2\sqrt{\phi_0}}{V_2 - V_1} \cdot (\sqrt{\phi_0 - V_2} - \sqrt{\phi_0 - V_1}) \quad \text{.....(125)}$$

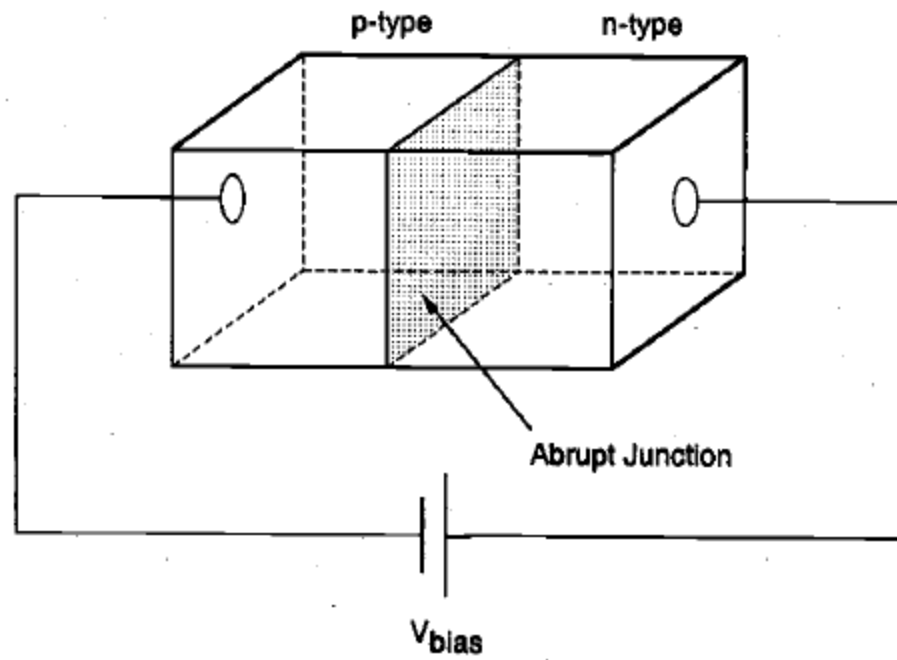
Junction Capacitances

- where K_{eq} is the voltage equivalence factor ($0 < K_{eq} < 1$). Thus, the coefficient
- K_{eq} allows to take into account the voltage-dependent variations of the junction capacitance.

Problem 11

- Consider a simple abrupt pn-junction, which is reverse-biased with a voltage V_{bias} . The doping density of the n-type region is $N_D = 10^{19} \text{ cm}^{-3}$, and the doping density of the p-type region is given as $N_A = 10^{16} \text{ cm}^{-3}$. The junction area is $A = 20 \text{ }\mu\text{m} \times 20 \text{ }\mu\text{m}$. Find C_{eq} .

Solution



Solution

$$\phi_0 = \frac{kT}{q} \cdot \ln\left(\frac{N_A \cdot N_D}{n_i^2}\right) = 0.026 \text{ V} \cdot \ln\left(\frac{10^{16} \cdot 10^{19}}{2.1 \times 10^{20}}\right) = 0.88 \text{ V}$$

$$\begin{aligned} C_{j0} &= \sqrt{\frac{\epsilon_{Si} \cdot q}{2} \cdot \left(\frac{N_A \cdot N_D}{N_A + N_D}\right) \cdot \frac{1}{\phi_0}} \\ &= \sqrt{\frac{11.7 \cdot 8.85 \times 10^{-14} \text{ F/cm} \cdot 1.6 \times 10^{-19} \text{ C}}{2} \cdot \left(\frac{10^{16} \cdot 10^{19}}{10^{16} + 10^{19}}\right) \cdot \frac{1}{0.88 \text{ V}}} \\ &= 3.1 \times 10^{-8} \text{ F/cm}^2 \end{aligned}$$

Solution

$$\begin{aligned}K_{eq} &= -\frac{2\sqrt{\phi_0}}{V_2 - V_1} \cdot (\sqrt{\phi_0 - V_2} - \sqrt{\phi_0 - V_1}) \\&= -\frac{2\sqrt{0.88}}{-5} \cdot (\sqrt{0.88 - (-5)} - \sqrt{0.88}) = 0.56\end{aligned}$$

$$C_{eq} = A \cdot C_{j0} \cdot K_{eq} = 400 \times 10^{-8} \text{ cm}^2 \cdot 3.1 \times 10^{-8} \text{ F/cm}^2 \cdot 0.56 = 69 \text{ fF}$$

Side Wall Factors

- the sidewalls of a typical MOSFET source or drain diffusion region are surrounded by a p^+ channel-stop implant, with a higher doping density than the substrate doping density N_A .
- Consequently, the sidewall zerobias capacitance C_{j0sw} as well as the sidewall voltage equivalence factor $K_{eq}(sw)$ will be different from those of the bottom junction.

Side Wall Factors

- Assuming that the sidewall doping density is given by $N_A(sw)$, the zero-bias capacitance per unit area can be found as follows:

$$C_{j0sw} = \sqrt{\frac{\epsilon_{Si} \cdot q}{2} \cdot \left(\frac{N_A(sw) \cdot N_D}{N_A(sw) + N_D} \right) \cdot \frac{1}{\phi_{0sw}}} \quad \dots(126)$$

- we can define a zerobias sidewall junction capacitance per unit length.

$$C_{jsw} = C_{j0sw} \cdot x_j \quad \dots(127)$$

Side Wall Factors

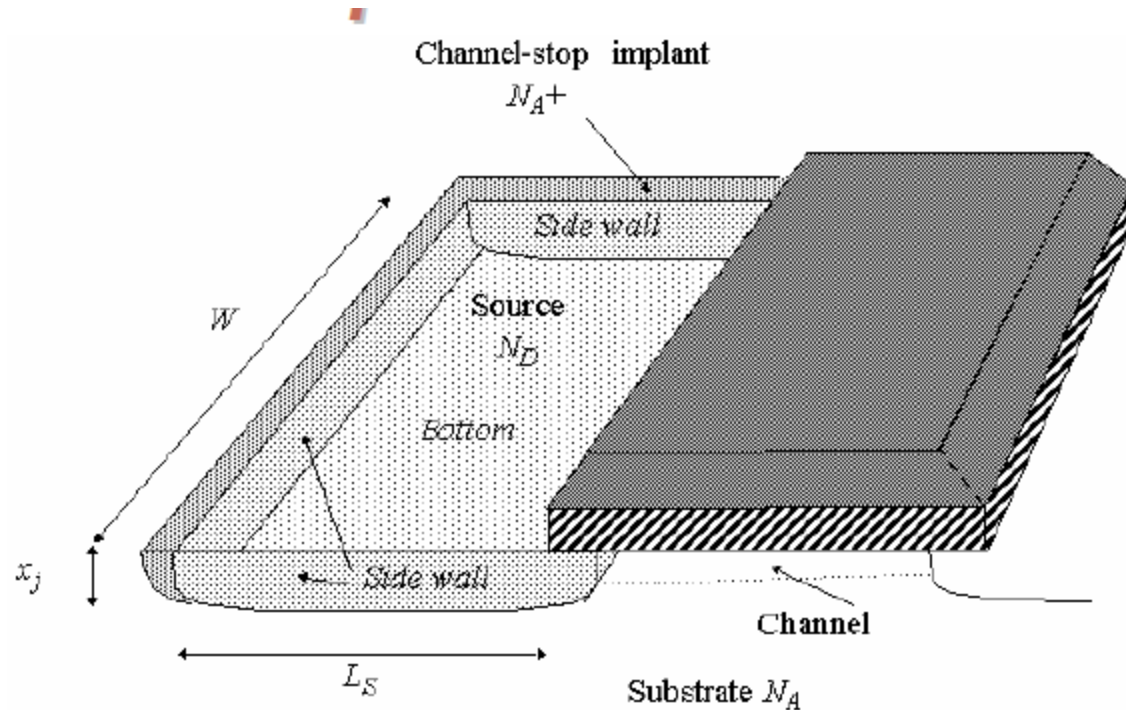
- Similarly the sidewall voltage equivalence factor $K_{eq}(sw)$ for a voltage swing between V_1 and V_2 is defined as follows:

$$K_{eq}(sw) = -\frac{2\sqrt{\phi_{0sw}}}{V_2 - V_1} \cdot (\sqrt{\phi_{0sw} - V_2} - \sqrt{\phi_{0sw} - V_1}) \quad \dots(128)$$

- The equivalent large-signal junction capacitance $C_{eq}(SW)$ for a sidewall of length (perimeter) P can be calculated as

$$C_{eq}(sw) = P \cdot C_{jsw} \cdot K_{eq}(sw) \quad \dots(129)$$

Side Wall Factors



$$\begin{aligned}
 C_{diff} &= C_{bottom} + C_{sw} = C_j \times AREA + C_{jsw} \times PERIMETER \\
 &= C_j L_S W + C_{jsw} (2L_S + W)
 \end{aligned}$$

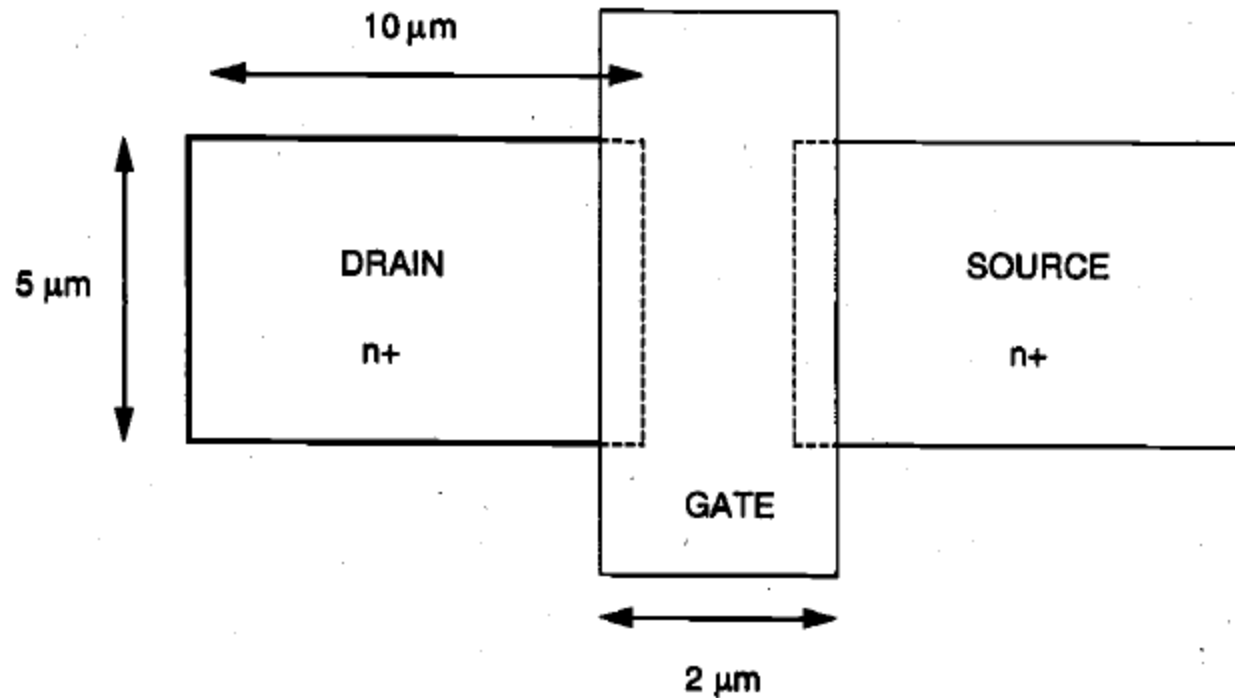
Problem 12

Consider the n-channel enhancement-type MOSFET shown below. The process parameters are given as follows:

Substrate doping	$N_A = 2 \times 10^{15} \text{ cm}^{-3}$
Source / drain doping	$N_D = 10^{19} \text{ cm}^{-3}$
Sidewall (p+) doping	$N_A(\text{sw}) = 4 \times 10^{16} \text{ cm}^{-3}$
Gate oxide thickness	$t_{ox} = 45 \text{ nm}$
Junction depth	$x_j = 1.0 \text{ } \mu\text{m}$

Problem 12

- Find C_{j0} , C_{jsw} and C_{db} .



Solution

$$\phi_0 = \frac{kT}{q} \cdot \ln\left(\frac{N_A \cdot N_D}{n_i^2}\right) = 0.026 \text{ V} \cdot \ln\left(\frac{2 \times 10^{15} \cdot 10^{19}}{2.1 \times 10^{20}}\right) = 0.837 \text{ V}$$

$$\phi_{0sw} = \frac{kT}{q} \cdot \ln\left(\frac{N_A(sw) \cdot N_D}{n_i^2}\right) = 0.026 \text{ V} \cdot \ln\left(\frac{4 \times 10^{16} \cdot 10^{19}}{2.1 \times 10^{20}}\right) = 0.915 \text{ V}$$

Solution

$$\begin{aligned}C_{j0} &= \sqrt{\frac{\epsilon_{Si} \cdot q}{2} \cdot \left(\frac{N_A \cdot N_D}{N_A + N_D} \right) \cdot \frac{1}{\phi_0}} \\&= \sqrt{\frac{11.7 \cdot 8.85 \times 10^{-14} \text{ F/cm} \cdot 1.6 \times 10^{-19} \text{ C}}{2} \cdot \left(\frac{2 \times 10^{15} \cdot 10^{19}}{2 \times 10^{15} + 10^{19}} \right) \cdot \frac{1}{0.837 \text{ V}}} \\&= 1.41 \times 10^{-8} \text{ F/cm}^2\end{aligned}$$

Solution

$$\begin{aligned}C_{j0sw} &= \sqrt{\frac{\epsilon_{Si} \cdot q}{2} \cdot \left(\frac{N_A(sw) \cdot N_D}{N_A(sw) + N_D} \right) \cdot \frac{1}{\phi_{0sw}}} \\&= \sqrt{\frac{11.7 \cdot 8.85 \times 10^{-14} \text{ F/cm} \cdot 1.6 \times 10^{-19} \text{ C}}{2} \cdot \left(\frac{4 \times 10^{16} \cdot 10^{19}}{4 \times 10^{16} + 10^{19}} \right) \cdot \frac{1}{0.915 \text{ V}}} \\&= 6.01 \times 10^{-8} \text{ F/cm}^2\end{aligned}$$

$$C_{jsw} = C_{j0sw} \cdot x_j = 6.01 \times 10^{-8} \text{ F/cm}^2 \cdot 10^{-4} \text{ cm} = 6.01 \text{ pF/cm}$$

Solution

$$K_{eq} = -\frac{2\sqrt{0.837}}{-5 - (-0.5)} \cdot \left(\sqrt{0.837 + 5} - \sqrt{0.837 + 0.5} \right) = 0.51$$

$$K_{eq}(sw) = -\frac{2\sqrt{0.915}}{-5 - (-0.5)} \cdot \left(\sqrt{0.915 + 5} - \sqrt{0.915 + 0.5} \right) = 0.53 \cong K_{eq}$$

$$A = (10 \times 5) \mu\text{m}^2 + (5 \times 1) \mu\text{m}^2 = 55 \mu\text{m}^2$$

Solution

$$\begin{aligned}\langle C_{db} \rangle &= A \cdot C_{j0} \cdot K_{eq} + P \cdot C_{jsw} \cdot K_{eq}(sw) \\ &= 55 \times 10^{-8} \text{ cm}^2 \cdot 1.41 \times 10^{-8} \text{ F/cm}^2 \cdot 0.51 \\ &\quad + 25 \times 10^{-4} \text{ cm} \cdot 6.01 \times 10^{-12} \text{ F/cm} \cdot 0.53 = 11.9 \times 10^{-15} \text{ F} = \underline{\underline{11.9 \text{ fF}}}\end{aligned}$$

SEMICONDUCTOR MEMORIES

Semiconductor Memories

- Semiconductor memory arrays capable of storing large quantities of digital information are essential to all digital systems.
- Memory circuits are generally classified according to the type of data storage and the type of data access.

Semiconductor Memories

- Read-Only Memory (ROM) circuits allow, as the name implies, only the retrieval of previously stored data and do not permit modifications of the stored information contents during normal operation.
- ROMs are non-volatile memories, i.e., the data storage function is not lost even when the power supply voltage is off.

Semiconductor Memories

- Depending on the type of data storage (data write) method, ROMs are classified as mask-programmed ROMs, Programmable ROMs (PROM), Erasable PROMs (EPROM), and Electrically Erasable PROMs (EEPROM).

Semiconductor Memories

- Read-write (R/W) memory circuits, on the other hand, must permit the modification (writing) of data bits stored in the memory array, as well as their retrieval (reading) on demand.
- This requires that the data storage function be volatile, i.e., the stored data are lost when the power supply voltage is turned off.
- The read-write memory circuit is commonly called Random Access Memory (RAM), mostly due to historical reasons.

Semiconductor Memories

- Compared to sequential-access memories such as magnetic tapes, any cell in the R/W memory array can be accessed with nearly equal access time.
- Based on the operation type of individual data storage cells, RAMs are classified into two main categories: Static RAMs (SRAM) and Dynamic RAMs (DRAM).
- Figure 1 shows an overview of the different memory types and their classifications.

Semiconductor Memories

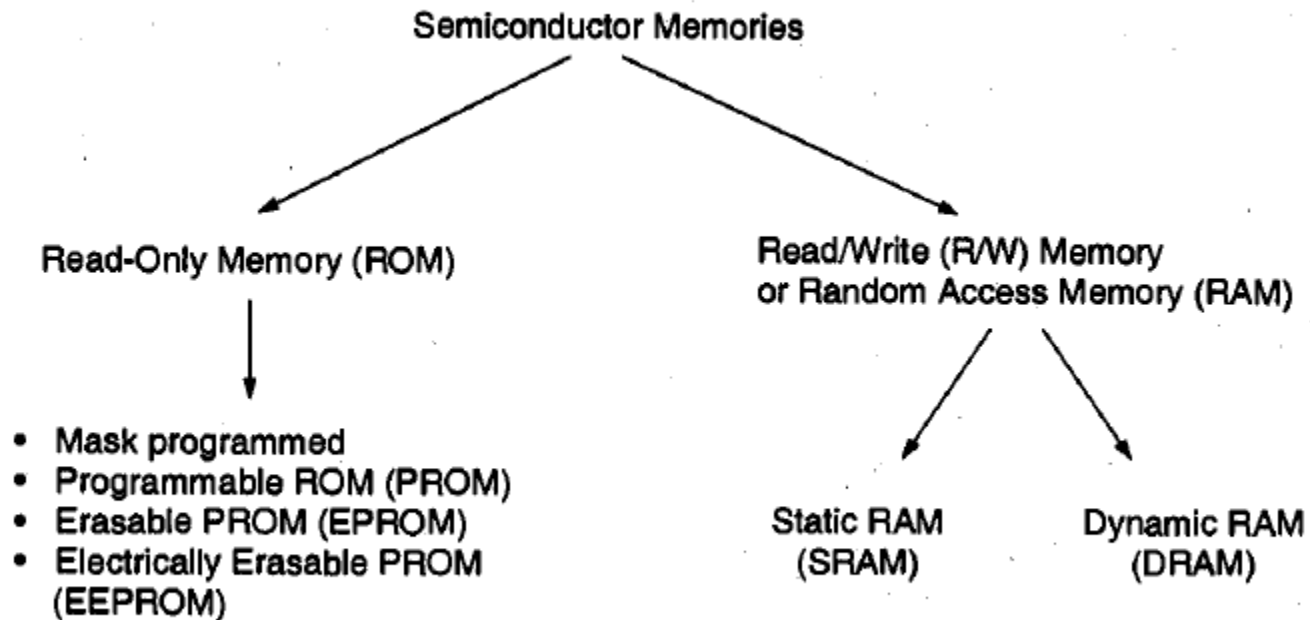


Figure 1. Overview of semiconductor memory types.

Semiconductor Memories

- A typical memory array organization is shown in Fig. 2.
- The data storage structure, or core, consists of individual memory cells arranged in an array of horizontal rows and vertical columns. Each cell is capable of storing one bit of binary information.

Semiconductor Memories

- Also, each memory cell shares a common connection with the other cells in the same row, and another common connection with the other cells in the same column.
- In this structure, there are 2^N rows, also called word lines, and 2^M columns, also called bit lines.
- Thus, the total number of memory cells in this array is $2^M \times 2^N$.

Semiconductor Memories

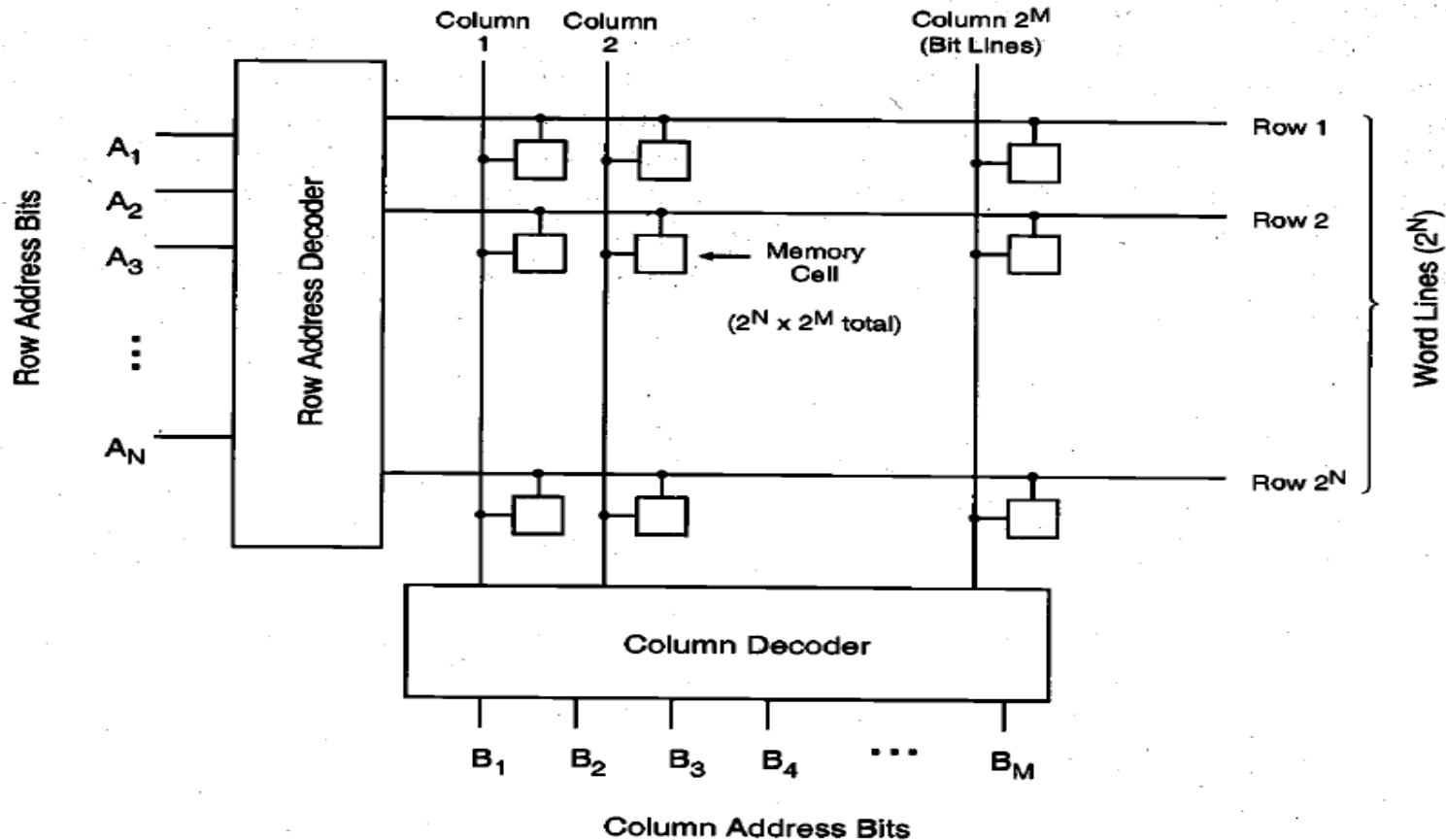


Figure 2. Typical random-access memory array organization.

Semiconductor Memories

- To access a particular memory cell, i.e., a particular data bit in this array, the corresponding bit line and the corresponding word line must be activated (selected).
- The row and column selection operations are accomplished by row and column decoders, respectively.
- The row decoder circuit selects one out of 2^N word lines according to an N bit row address, while the column decoder circuit selects one out of 2^M bit lines according to an M-bit column address.

Semiconductor Memories

- Once a memory cell or a group of memory cells are selected in this fashion, a data read and/or a data write operation may be performed on the selected single bit or multiple bits on a particular row.
- The column decoder circuit serves the double duties of selecting the particular columns and routing the corresponding data content in a selected row to the output.

Semiconductor Memories

- We can see from this simple discussion that individual memory cells can be accessed for data read and/or data write operations in random order, independent of their physical locations in the memory array.
- Thus, the array organization examined here is called a Random Access Memory (RAM) structure.

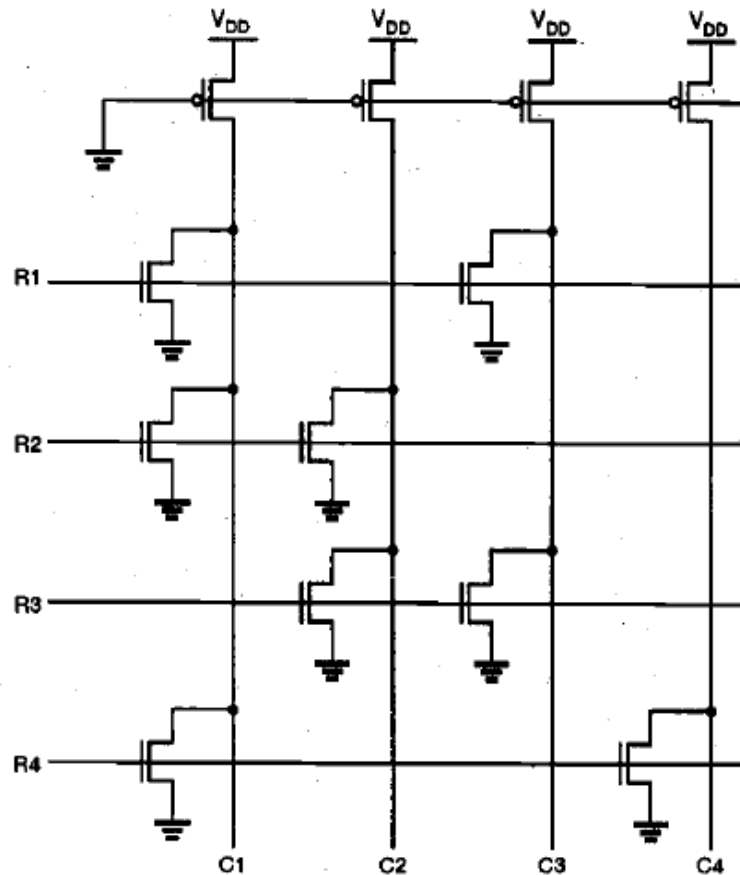
Read-Only Memory (ROM) Circuits

- The storing of binary information at a particular address location can be achieved by the presence or absence of a data path from the selected row (word line) to the selected column (bit line), which is equivalent to the presence or absence of a device at that particular location.
- Here , we will examine two different implementations for MOS ROM arrays.

Read-Only Memory (ROM) Circuits

- Consider first the 4-bit x4-bit memory array shown in Fig. 3.
- Here, each column consists of a pseudo-nMOS NOR gate driven by some of the row signals, i.e., the word lines

Read-Only Memory (ROM) Circuits



R1	R2	R3	R4	C1	C2	C3	C4
1	0	0	0	0	1	0	1
0	1	0	0	0	0	1	1
0	0	1	0	1	0	0	1
0	0	0	1	0	1	1	0

Figure 3. Example of a 4-bit x 4-bit NOR-based ROM array

Read-Only Memory (ROM) Circuits

- As described, before only one word line is activated (selected) at a time by raising its voltage to V_{DD} , while all other rows are held at a low voltage level.
- If an active transistor exists at the cross point of a column and the selected row, the column voltage is pulled down to the logic low level by that transistor.

Read-Only Memory (ROM) Circuits

- If no active transistor exists at the cross point, the column voltage is pulled high by the pMOS load device.
- Thus, a logic " 1 "-bit is stored as the absence of an active transistor, while a logic "0"-bit is stored as the presence of an active transistor at the crosspoint.

Read-Only Memory (ROM) Circuits

- In actual ROM layout, the array can be initially manufactured with nMOS transistors at every row-column intersection.
- The "1"-bits are then realized by omitting the drain or source connection, or the gate electrode of the corresponding nMOS transistors in the final metallization step.
- Figure 4 shows four nMOS transistors in a NOR ROM array, forming the intersection of two metal bit lines and two polysilicon word lines.

Read-Only Memory (ROM) Circuits

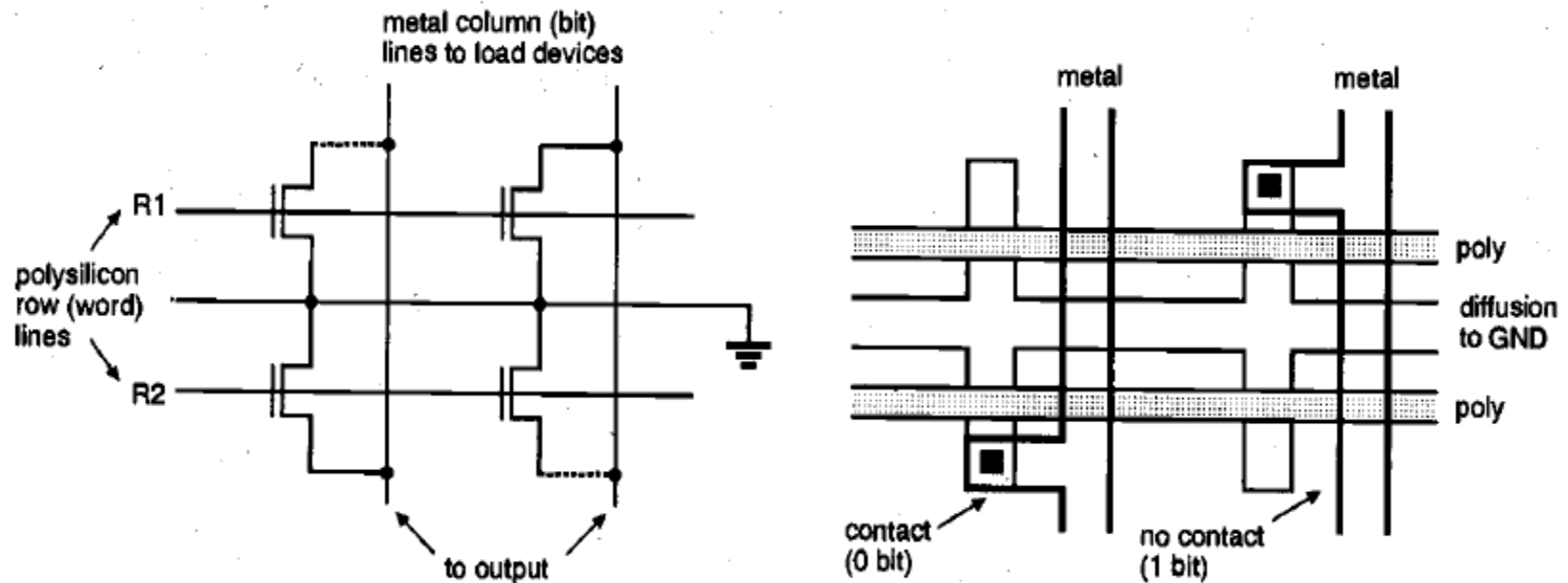


Figure 4. Layout example of a NOR ROM array

Read-Only Memory (ROM) Circuits

- To save silicon area, the transistors in every two adjacent rows are arranged to share a common ground line, also routed in n-type diffusion.
- To store a "0 "- bit at a particular address location, the drain diffusion of the corresponding transistor must be connected to the metal bit line via a metal-to-diffusion contact.
- Omission of this contact, results in a stored "1 "-bit.

Read-Only Memory (ROM) Circuits

- Next, we will examine a significantly different ROM array design, which is also called a NAND ROM (Fig. 5).
- Here, each bit line consists of a depletion-load NAND gate, driven by some of the row signals, i.e., the word lines.
- In normal operation, all word lines are held at the logic-high voltage level except for the selected line, which is pulled down to logic-low level.

Read-Only Memory (ROM) Circuits

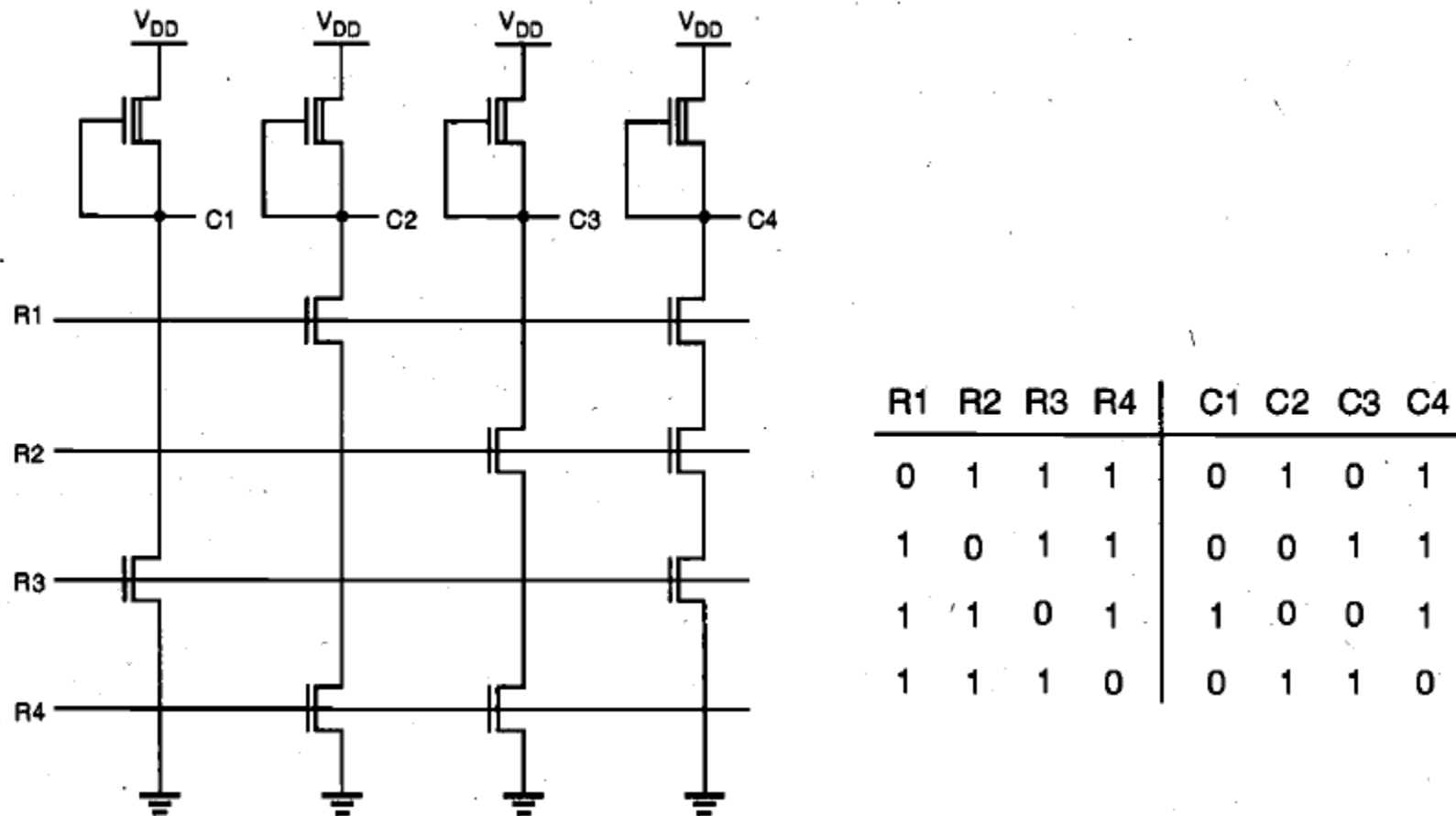


Figure 5. A 4-bit x 4-bit NAND-based ROM array

Read-Only Memory (ROM) Circuits

- If a transistor exists at the crosspoint of a column and the selected row, that transistor is turned off and the column voltage is pulled high by the load device.
- On the other hand, if no transistor exists (shorted) at that particular crosspoint, the column voltage is pulled low by the other nMOS transistors in the multi-input NAND structure.

Read-Only Memory (ROM) Circuits

- Thus, a logic "1"-bit is stored by the presence of a transistor that can be deactivated, while a logic "0"-bit is stored by a shorted or normally on transistor at the crosspoint.

Design of Row and Column Decoders

- Now we will turn our attention to the circuit structures of row and column address decoders, which select a particular memory location in the array, based on the binary row and column addresses.
- A row decoder designed to drive a NOR ROM array must, by definition, select one of the 2^N word lines by raising its voltage to V_{OH} .

Design of Row and Column Decoders

- As an example, consider the simple row address decoder shown in Fig. 6, which decodes a two-bit row address and selects one out of four word lines by raising its level.

Design of Row and Column Decoders

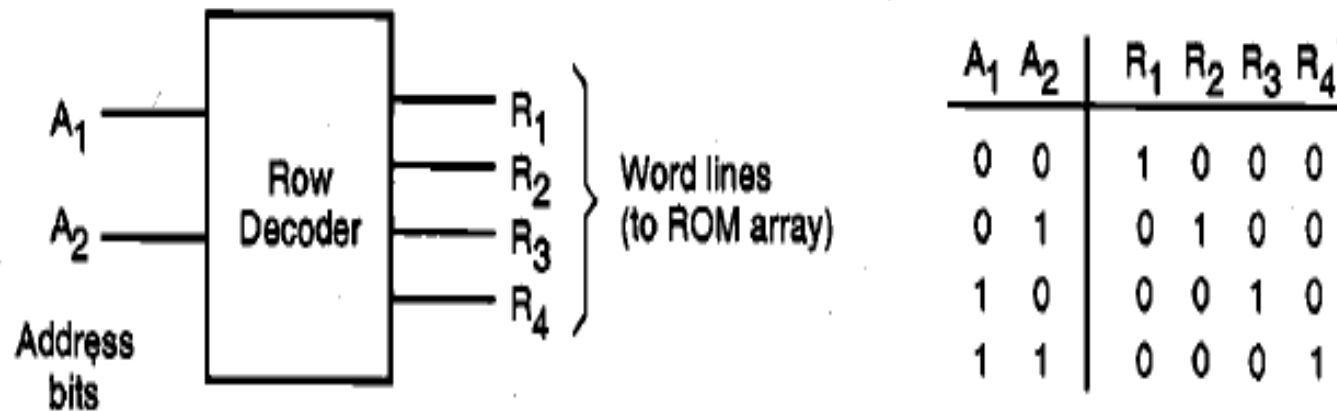


Figure 6. Row address decoder example for 2 address bits and 4 word lines.

Design of Row and Column Decoders

- A most straightforward implementation of this decoder is another NOR array, consisting of 4 rows (outputs) and 4 columns (two address bits and their complements).
- This NOR-based decoder array can be built just like the NOR ROM array, using the same selective programming approach (Fig. 7).
- The ROM array and its row decoder can thus be fabricated as two adjacent NOR arrays, as shown in Fig. 8.

Design of Row and Column Decoders

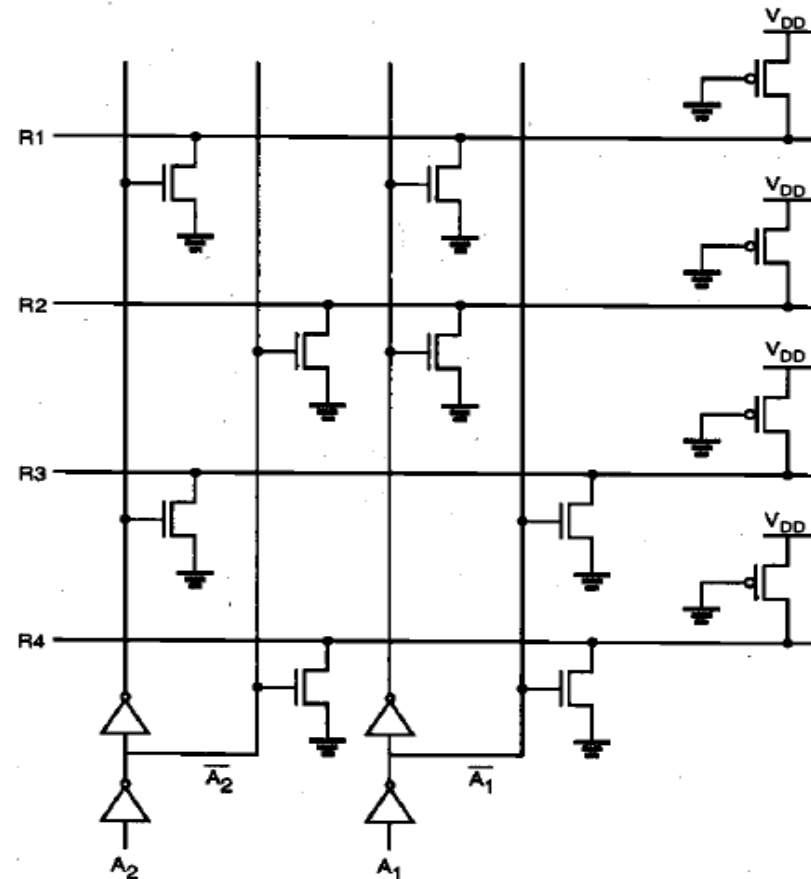


Figure 7. NOR-based row decoder circuit for 2 address bits and 4 word lines.

Design of Row and Column Decoders

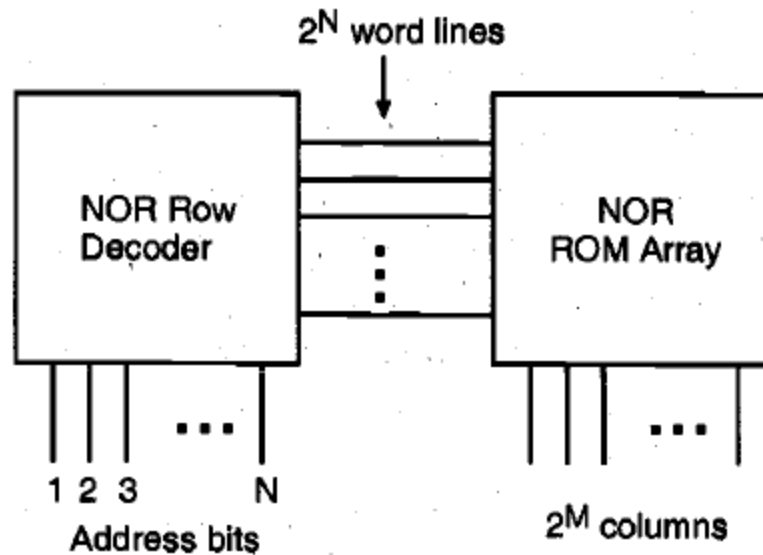


Figure 8. Implementation of the row decoder circuit and the ROM array as two adjacent NOR planes.

Design of Row and Column Decoders

- A row decoder designed to drive a NAND ROM, on the other hand, must lower the voltage level of the selected row to logic "0" while keeping all other rows at a logic-high level.
- This function can be implemented by using an N-input NAND gate for each of the row outputs.

Design of Row and Column Decoders

- The truth table of a simple address decoder for four rows and the double NAND-array implementation of the decoder and the ROM are shown in Fig. 9.
- As in the NOR ROM case, the row address decoder of the NAND ROM array can thus be realized using the same layout strategy as the memory array itself.

Design of Row and Column Decoders

A_1	A_2	R_1	R_2	R_3	R_4
0	0	0	1	1	1
0	1	1	0	1	1
1	0	1	1	0	1
1	1	1	1	1	0

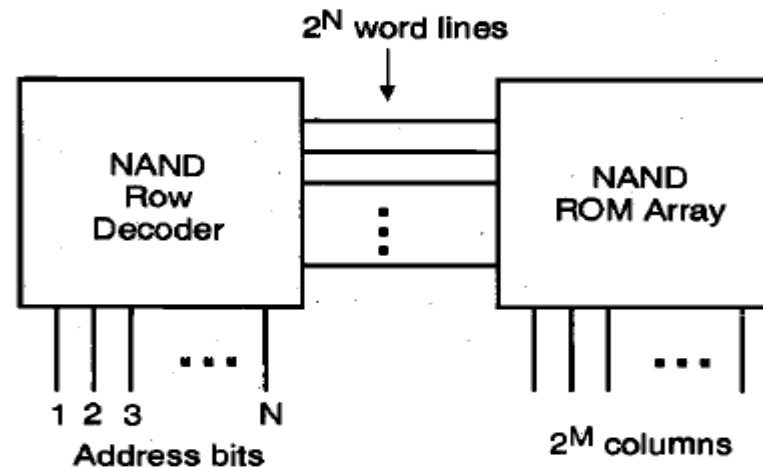


Figure 9. Truth table of a row decoder for the NAND ROM array, and implementation of the row decoder circuit and the ROM array as two adjacent NAND planes.

Design of Row and Column Decoders

- The column decoder circuitry is designed to select one out of 2^M bit lines (columns) of the ROM array according to an M-bit column address, and to route the data content of the selected bit line to the data output.
- A straightforward but costly approach would be to connect an nMOS pass transistor to each bit-line (column) output, and to selectively drive one out of 2^M pass transistors by using a NOR-based column address decoder, as shown in Fig. 10.

Design of Row and Column Decoders

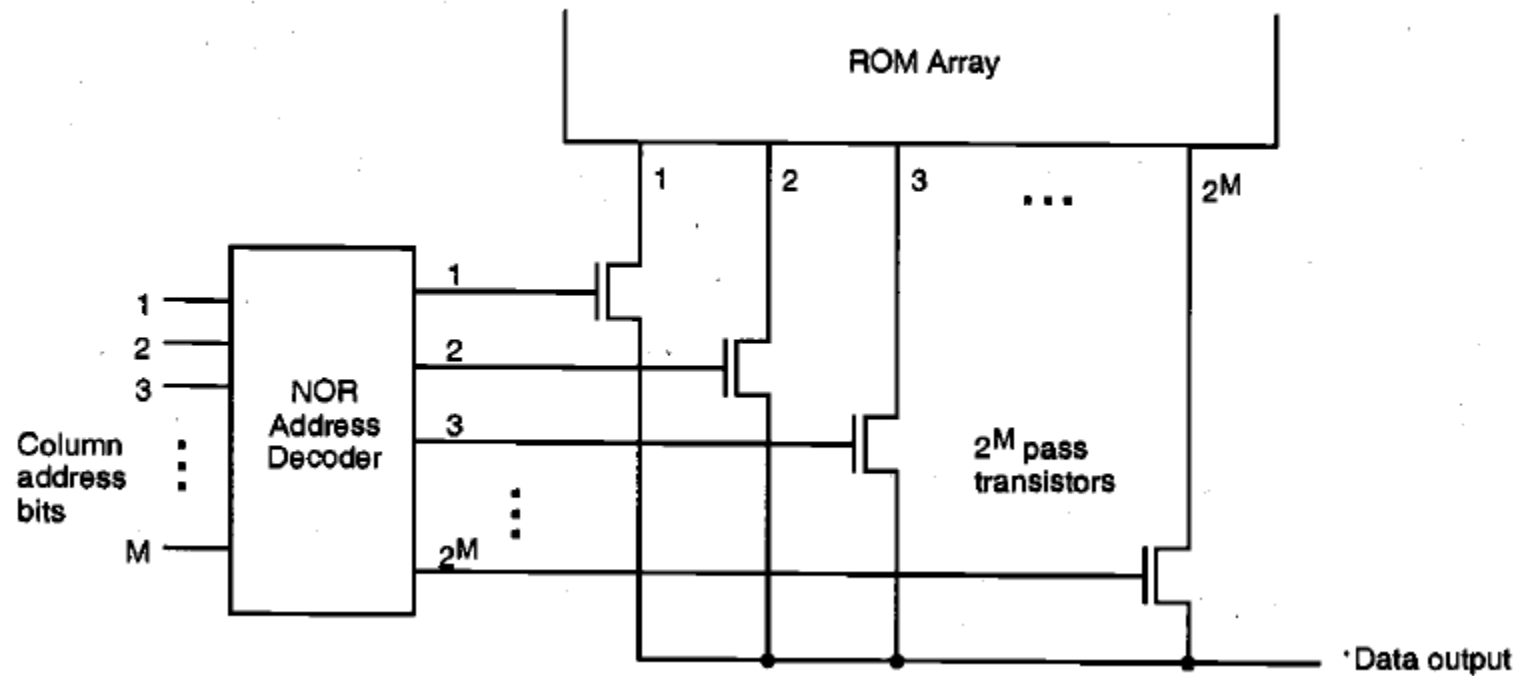


Figure 10. Bit-line (column) decoder arrangement using a NOR address decoder and nMOS pass transistors for every bit line.

Design of Row and Column Decoders

- In this arrangement, only one nMOS pass transistor is turned on at a time, depending on the column address bits applied to the decoder inputs.
- The conducting pass transistor routes the selected column signal to the data output.
- Similarly, a number of columns can be chosen at a time, and the selected columns can be routed to a parallel data output port.

Design of Row and Column Decoders

- The number of transistors required for this column decoder implementation is $2^M(M+1)$, i.e., 2^M pass transistors for each bit line and $M \cdot 2^M$ transistors for the decoder circuit.
- This number can quickly become excessive for large M , i.e., for a large number of bit lines.

Design of Row and Column Decoders

- An alternative design of the column decoder circuit is to build a binary selection tree consisting of consecutive stages, as shown in Fig. 11.
- In this case, the pass transistor network is used to select one out of every two bit lines at each stage (level), whereas the column address bits drive the gates of the nMOS pass transistors.

Design of Row and Column Decoders

- Notice that a NOR address decoder is not needed for this decoder tree structure, thereby reducing the number of transistors significantly although it requires M additional inverters ($2M$ transistors) for complementing column address bits.
- The example shown in Fig. 11 is a column decoder tree for eight bit lines, which requires three column address bits (and their complements) to select one of the eight columns.

Design of Row and Column Decoders

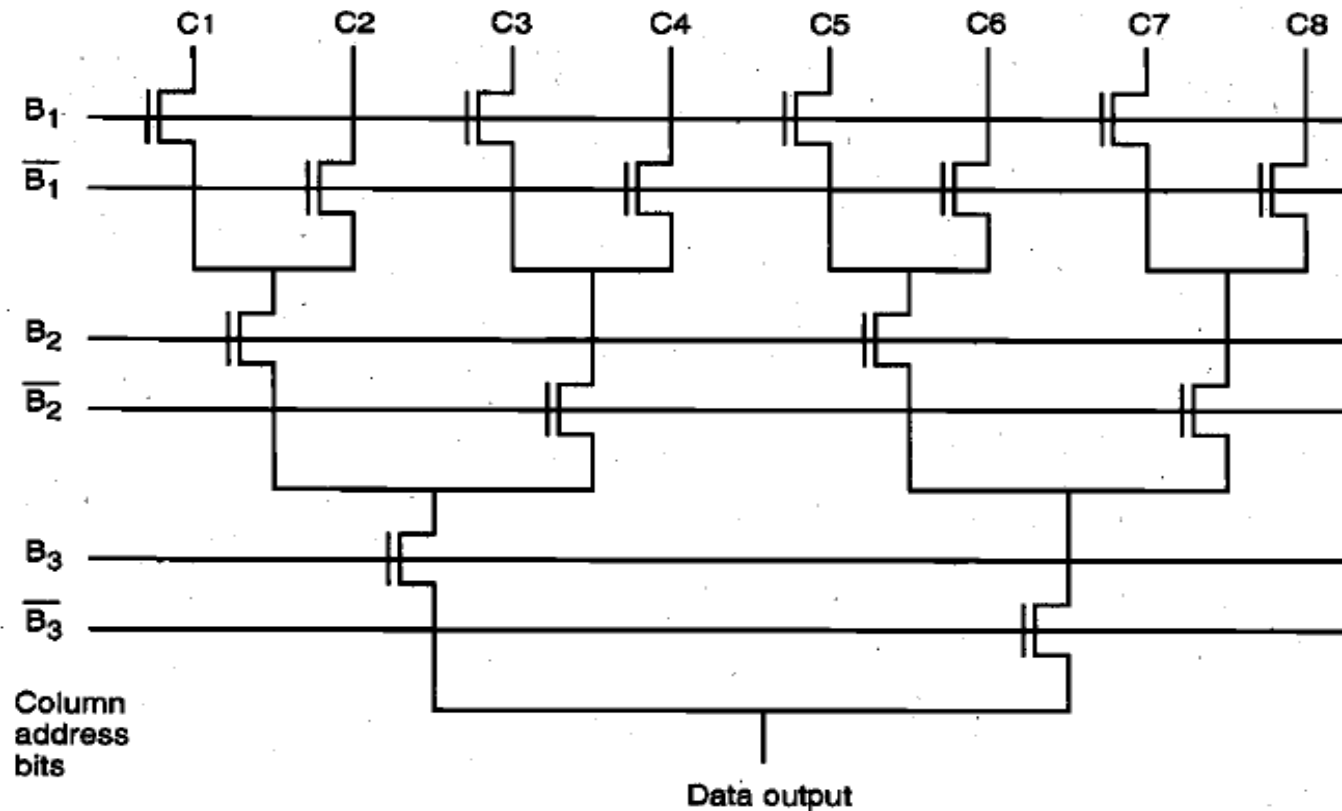


Figure 11. Column decoder circuit for eight bit lines, implemented as a binary tree decoder which is driven directly by the three column address bits.

Design of Row and Column Decoders

- One drawback of the decoder tree approach is that the number of series-connected nMOS pass transistors in the data path is equal to the number of column address bits, M .
- This situation can cause a long data access time, since the decoder delay time depends on the equivalent series resistance of the decoder branch that directs the column data to the output.

Design of Row and Column Decoders

- To overcome this constraint, column address decoders can be built as a combination of the two structures.

Static Read-Write Memory (SRAM) Circuits

- The data storage cell, i.e., the 1-bit memory cell in static RAM arrays, consists of a simple latch circuit with two stable operating points (states).
- Depending on the preserved state of the two-inverter latch circuit, the data being held in the memory cell will be interpreted either as a logic "0" or as a logic " 1."

Static Read-Write Memory (SRAM) Circuits

- To access (read and write) the data contained in the memory cell via the bit line, we need at least one switch, which is controlled by the corresponding word line, i.e., the row address selection signal (Fig.12(a)).
- Usually, two complementary access switches consisting of nMOS pass transistors are implemented to connect the 1-bit SRAM cell to the complementary bit lines (columns).

Static Read-Write Memory (SRAM) Circuits

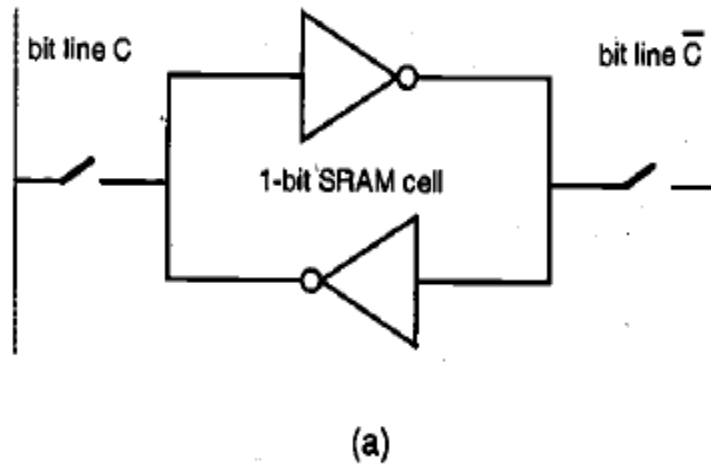


**Figure 12. Various configurations of the static RAM cell.
(a) Symbolic representation of the two-inverter latch circuit with access switches.**

Static Read-Write Memory (SRAM) Circuits

- Figure 12 (b) shows the generic structure of the MOS static RAM cell, consisting of two cross-coupled inverters and two access transistors.
- The load devices may be polysilicon resistors, depletion-type nMOS transistors, or pMOS transistors, depending on the type of the memory cell.
- The pass gates acting as data access switches are enhancement-type nMOS transistors.

Static Read-Write Memory (SRAM) Circuits

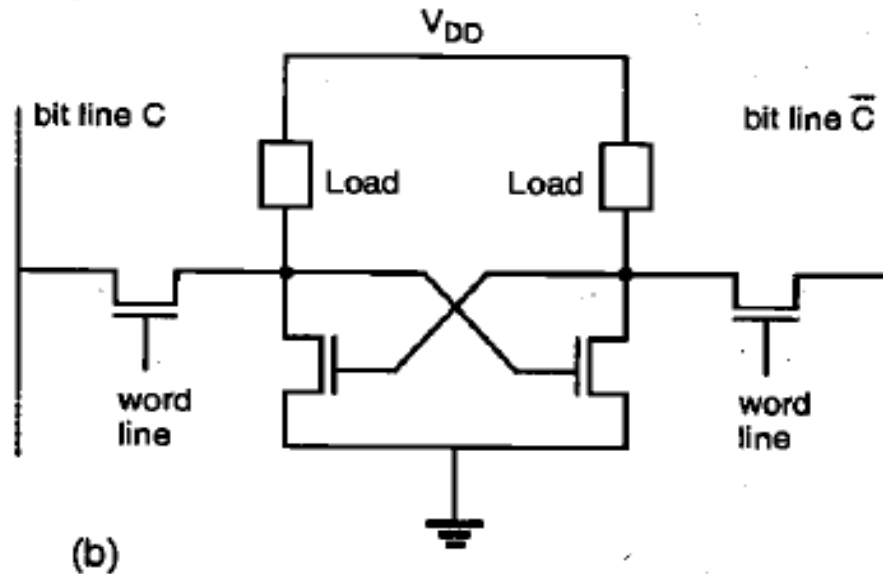


Figure 12. Various configurations of the static RAM cell
(b) Generic circuit topology of the MOS static RAM cell.

Static Read-Write Memory (SRAM) Circuits

- The use of resistive-load inverter with undoped polysilicon resistors in the latch structure typically results in a significantly more compact cell size, compared with the other alternatives (Fig. 12(c)).
- This is true since the resistors can be stacked on top of the cell (using double-polysilicon technology), thereby reducing the cell size to four transistors, as opposed to the six-transistor cell topologies.

Static Read-Write Memory (SRAM) Circuits

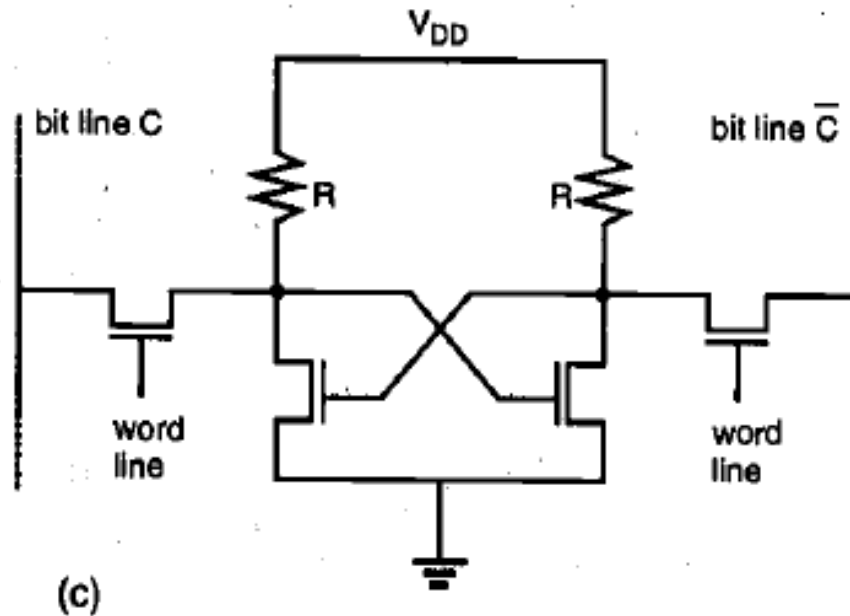


Figure 12. Various configurations of the static RAM cell.
(c) Resistive-load SRAM cell

Static Read-Write Memory (SRAM) Circuits

- If multiple polysilicon layers are available, one layer can be used for the gates of the enhancement-type nMOS transistors, while another level is used for load resistors and interconnects.

Static Read-Write Memory (SRAM) Circuits

- The six-transistor depletion-load nMOS SRAM cell shown in Fig. 12 (d) can be easily implemented with one polysilicon and one metal layer, and the cell size tends to be relatively small, especially with the use of buried metal-diffusion contacts.
- The static characteristics and the noise margins of this memory cell are typically better than those of the resistive-load cell.
- The static power consumption of the depletion-load SRAM cell, however, makes it an unsuitable candidate for high-density SRAM arrays.

Static Read-Write Memory (SRAM) Circuits

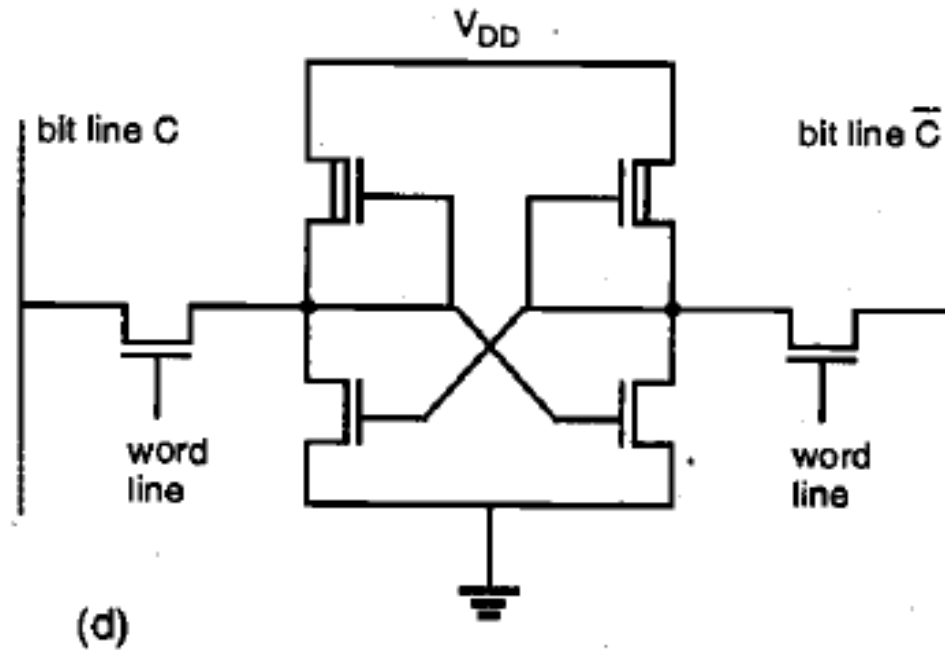


Fig. 12(d) Depletion-load nMOS SRAM cell

Static Read-Write Memory (SRAM) Circuits

- The full CMOS SRAM cell shown in Fig. 12 (e) achieves the lowest static power dissipation among the various circuit configurations presented here.
- In addition, the CMOS cell offers superior noise margins and switching speed.

Static Read-Write Memory (SRAM) Circuits

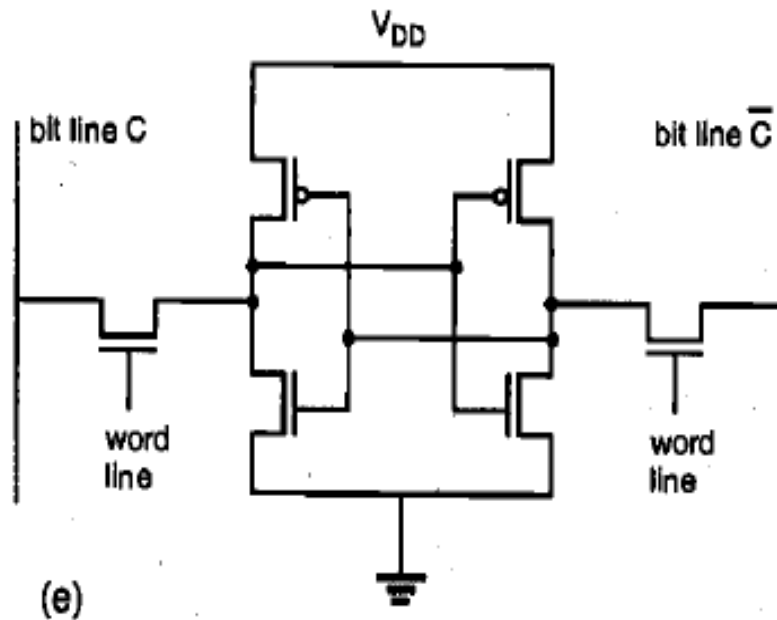


Fig. 12(e) Full CMOS SRAM cell.

SRAM Operation Principles

- Figure 13 shows a typical four-transistor resistive-load SRAM cell widely used in high-density memory arrays, consisting of a pair of cross-coupled inverters.
- The two stable operating points of this basic latch circuit are used to store a one-bit piece of information; hence, this pair of cross-coupled inverters make up the central component of the SRAM cell.

SRAM Operation Principles

- To perform read and write operations, we use two nMOS pass transistors, both of which are driven by the row select signal, RS.
- The SRAM cell shown in Fig. 13 is accessed via two bit lines or columns, instead of one.

SRAM Operation Principles

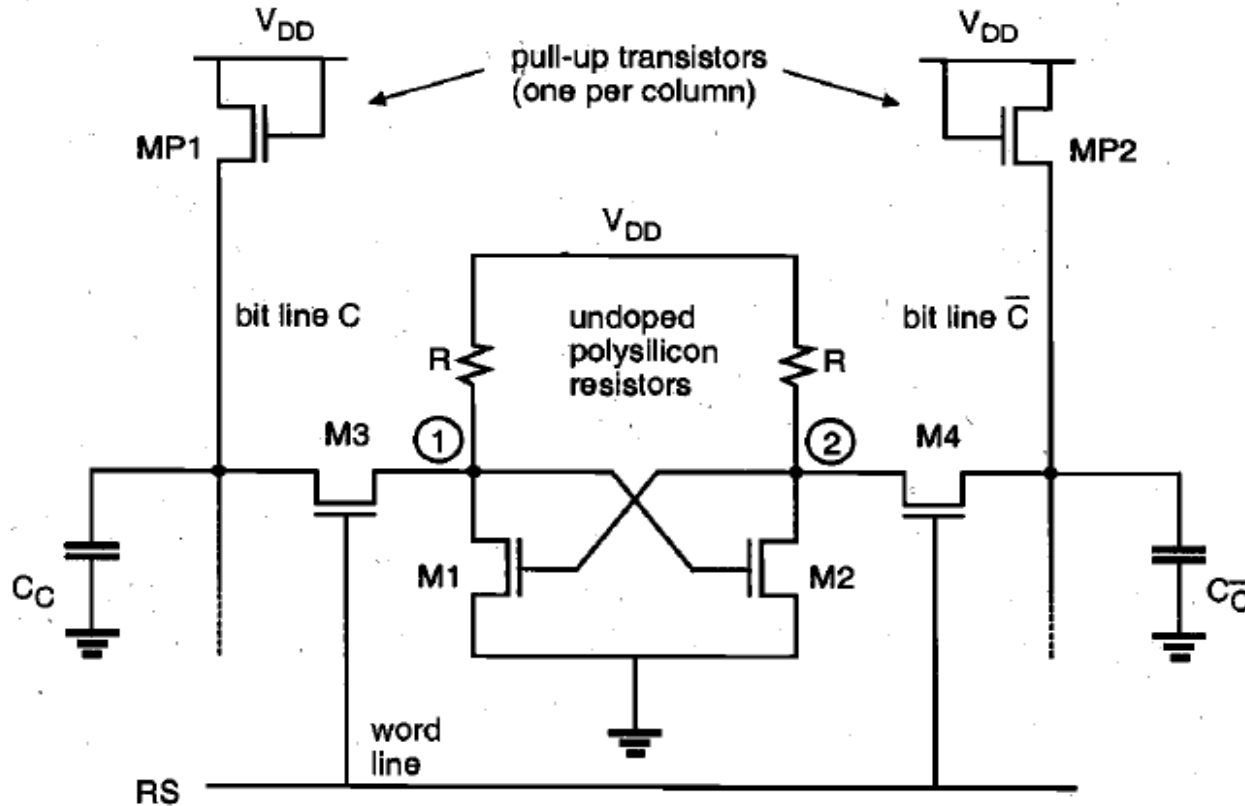


Figure 13. Basic structure of the resistive-load SRAM cell, shown with the column pull-up transistors.

SRAM Operation Principles

- When the word line (RS) is not selected, i.e., when the voltage level of line RS is equal to logic "0," the pass transistors M_3 and M_4 are turned off.
- The simple latch circuit consisting of two cross-connected inverters preserves one of its two stable operating points; hence, data is being held.
- At this point, consider the two columns, C and $\bar{C}$.

SRAM Operation Principles

- If all word lines in the SRAM array are inactive, the relatively large column capacitances are charged-up by the column pull-up transistors, MP1 and MP2.
- Since both transistors operate in saturation, the steady-state voltage level V_C for both columns is determined by the following relationship:

$$V_{DD} - V_C = V_{T0} + \gamma \left(\sqrt{|2\phi_F| + V_C} - \sqrt{|2\phi_F|} \right)$$

SRAM Operation Principles

Assuming $V_{DD} = 5\text{ V}$, $V_{T0} = 1\text{ V}$, $|2\phi_F| = 0.6\text{ V}$, and $\gamma = 0.4\text{ V}^{1/2}$, this voltage level is found to be approximately equal to 3.5 V . Note that the voltage levels of the two complementary bit lines (columns) are equal during this phase.

SRAM Operation Principles

- Now assume that we select the memory cell by raising its word line voltage to logic "1," hence, the pass transistors M_3 and M_4 are turned on.
- Once the memory cell is selected, four basic operations may be performed on this cell.

SRAM Operation Principles

(a) Write "1" operation:

- The voltage level of column $\bar{C}$ is forced to logic-low by the data-write circuitry. The driver transistor M_1 turns off. The voltage V_1 attains a logic-high level, while V_2 goes low.

SRAM Operation Principles

(b) Read "1" operation:

- The voltage of column C retains its precharge level while the voltage of column $\bar{C}$ is pulled down by M_2 and M_4 . The data-read circuitry detects the small voltage difference ($V_C > V_{\bar{C}}$) and amplifies it as a logic "1" data output.

SRAM Operation Principles

(c) Write "0" operation:

- The voltage level of column C is forced to logic-low by the data-write circuitry. The driver transistor M_2 turns off. The voltage V_2 attains a logic-high level, while V_1 goes low.

SRAM Operation Principles

(d) Read "0" operation:

- The voltage of column $\bar{C}$ retains its precharge level while the voltage of column C is pulled down by M_1 and M_3 . The data-read circuitry detects the small voltage difference ($V_C < V_{\bar{C}}$) and amplifies it as a logic "0" data output.

Full CMOS SRAM Cell

- A low-power SRAM cell may be designed simply by using cross-coupled CMOS inverters instead of the resistive-load nMOS inverters.
- In this case, the stand-by power consumption of the memory cell will be limited to the relatively small leakage currents of both CMOS inverters.
- The possible drawback of using CMOS SRAM cells, is that the cell area tends to increase in order to accommodate the n-well for the pMOS transistors and the polysilicon contacts.

Full CMOS SRAM Cell

- The circuit structure of the full CMOS static RAM cell is shown in Fig. 14, along with the pMOS column pull-up transistors on the complementary bit lines.
- The basic operation principle of the CMOS SRAM cell is identical to that of the resistive-load nMOS cell examined earlier.
- The most important advantage of this circuit topology is that the static power dissipation is even smaller; essentially, it is limited by the leakage current of the pMOS transistors.

Full CMOS SRAM Cell

- A CMOS memory cell thus draws current from the power supply only during a switching transition.
- The low standby power consumption has certainly been a driving force for the increasing prominence of high-density CMOS SRAMs.

Full CMOS SRAM Cell

- Other advantages of CMOS SRAM cells include high noise immunity due to larger noise margins, and the ability to operate at lower power supply voltages than, for example, the resistive-load SRAM cells.
- The major disadvantages of CMOS memories historically were larger cell size, the added complexity of the CMOS process, and the tendency to exhibit "latch-up" phenomena.

Full CMOS SRAM Cell

- With the widespread use of multi-layer polysilicon and multi-layer metal processes, however, the area disadvantage of the CMOS SRAM cell has been reduced significantly in recent years.
- Considering the undisputable advantages of CMOS for low-power and low-voltage operation, the added process complexity and the required latch-up prevention measures do not present a substantial barrier against the implementation of CMOS cells in high density SRAM arrays.

Full CMOS SRAM Cell

- Note that unlike the nMOS column pull-up devices used in resistive-load SRAM, the pMOS column pull-up transistors shown in Fig. 14 allow the column voltages to reach full V_{DD} level.
- To further reduce power consumption, these transistors can also be driven by a periodic precharge signal, which activates the pull-up devices to charge-up column capacitances.

CMOS SRAM Cell Design Strategy

- To determine the (W/L) ratios of the transistors in a typical CMOS SRAM cell as shown in Fig. 14, a number of design criteria must be taken into consideration.
- The two basic requirements which dictate the (W/L) ratios are: (a) the data-read operation should not destroy the stored information in the SRAM cell, and (b) the cell should allow modification of the stored information during the data-write phase.

CMOS SRAM Cell Design Strategy

- Consider the data-read operation first, assuming that a logic "0" is stored in the cell.
- The voltage levels in the CMOS SRAM cell at the beginning of the "read" operation are depicted in Fig. 15.
- Here, the transistors M_2 and M_5 are turned off, while the transistors M_1 and M_6 operate in the linear mode.

CMOS SRAM Cell Design Strategy

- Thus, the internal node voltages are $V_1 = 0$ and $V_2 = V_{DD}$ before the cell access (or pass) transistors M_3 and M_4 are turned on.
- The active transistors at the beginning of the data-read operation are highlighted in Fig. 15.

CMOS SRAM Cell Design Strategy

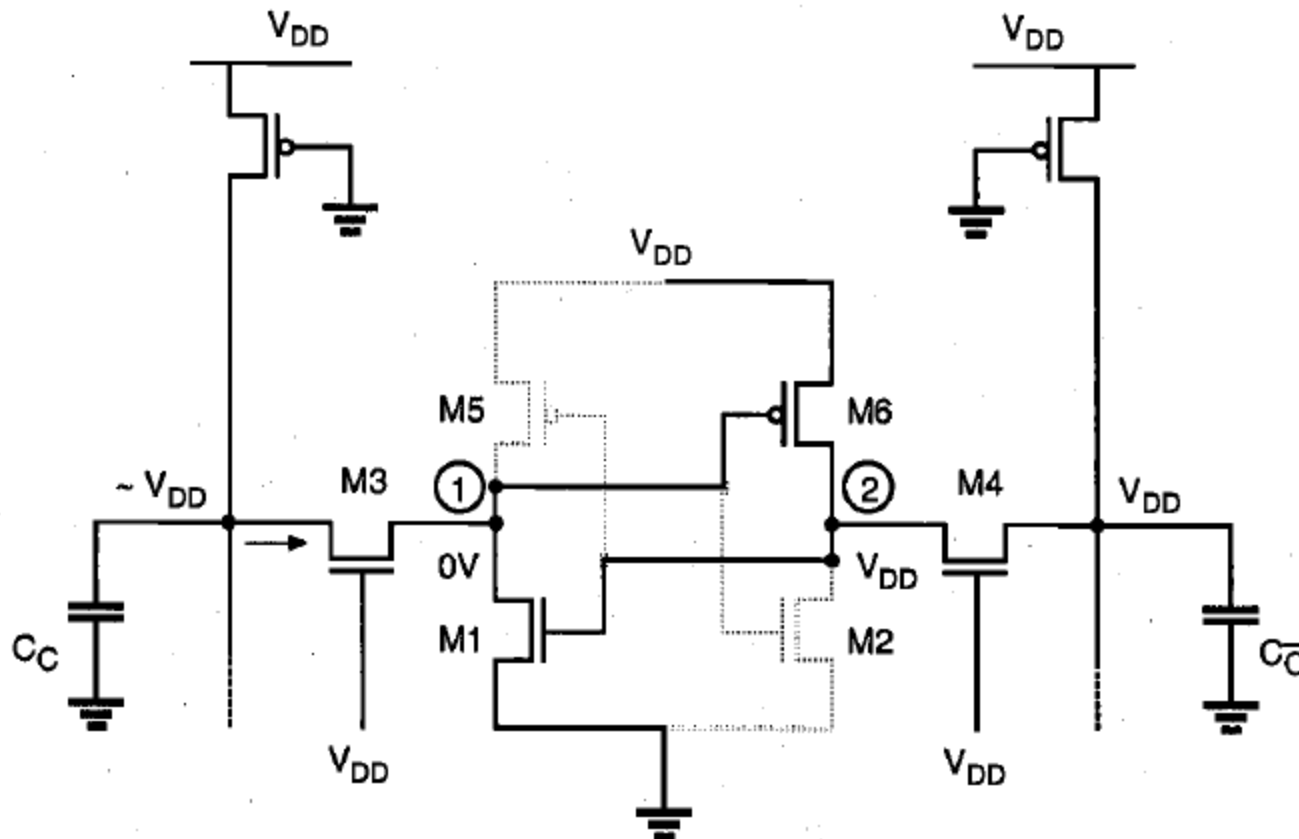


Figure 15. Voltage levels in the SRAM cell at the beginning of the "read" operation.

CMOS SRAM Cell Design Strategy

- After the pass transistors M_3 and M_4 are turned on by the row selection circuitry, the voltage level of column $\bar{C}$ will not show any significant variation since no current will flow through M_4 .
- On the other half of the cell, however, M_3 and M_1 will conduct a nonzero current and the voltage level of column C will begin to drop slightly.

CMOS SRAM Cell Design Strategy

- Note that the column capacitance C_c is typically very large; therefore, the amount of decrease in the column voltage is limited to a few hundred millivolts during the read phase.

CMOS SRAM Cell Design Strategy

- The data-read circuitry to be examined later in this chapter is responsible for detecting this small voltage drop and amplifying it as a stored "0. "
- While M_1 and M_3 are slowly discharging the column capacitance, the node voltage V , will increase from its initial value of 0 V.

CMOS SRAM Cell Design Strategy

- Especially if the (W/L) ratio of the access transistor M_3 is large compared to the (W/L) ratio of M_1 , the node voltage V_1 may exceed the threshold voltage of M_2 during this process, forcing an unintended change of the stored state.
- The key design issue for the data-read operation is then to guarantee that the voltage V_1 does not exceed the threshold voltage of M_2 , so that the transistor M_2 remains turned off during the read phase, i.e.,

CMOS SRAM Cell Design Strategy

$$V_{1,max} \leq V_{T,2} \quad \dots\dots(1)$$

- We can assume that after the access transistors are turned on, the column voltage V_c remains approximately equal to V_{DD} .
- Hence, M3 operates in saturation while M operates in the linear region

$$\frac{k_{n,3}}{2}(V_{DD} - V_1 - V_{T,n})^2 = \frac{k_{n,1}}{2}(2(V_{DD} - V_{T,n})V_1 - V_1^2) \quad \dots\dots(2)$$

CMOS SRAM Cell Design Strategy

$$\frac{k_{n,3}}{k_{n,1}} = \frac{\left(\frac{W}{L}\right)_3}{\left(\frac{W}{L}\right)_1} < \frac{2(V_{DD} - 1.5V_{T,n})V_{T,n}}{(V_{DD} - 2V_{T,n})^2} \dots\dots(3)$$

- The upper limit of the aspect ratio found above is actually more conservative, since a portion of the drain current of M_3 will also be used to charge-up the parasitic node capacitance of node (1).

CMOS SRAM Cell Design Strategy

- To summarize, the transistor M_2 will remain in cut-off during the read "0" operation if condition (Eqn. 3) is satisfied.
- A symmetrical condition also dictates the aspect ratios of M_2 and M_4 .

CMOS SRAM Cell Design Strategy

- Now consider the write "0" operation, assuming that a logic "1" is stored in the SRAM cell initially.
- Figure 16 shows the voltage levels in the CMOS SRAM cell at the beginning of the data-write operation.
- The transistors M_1 and M_6 are turned off, while the transistors M_2 and M_5 operate in the linear mode.
- Thus, the internal node voltages are $V_1 = V_{DD}$ and $V_2 = 0$ V before the cell access (or pass) transistors M_3 and M_4 are turned on.

CMOS SRAM Cell Design Strategy

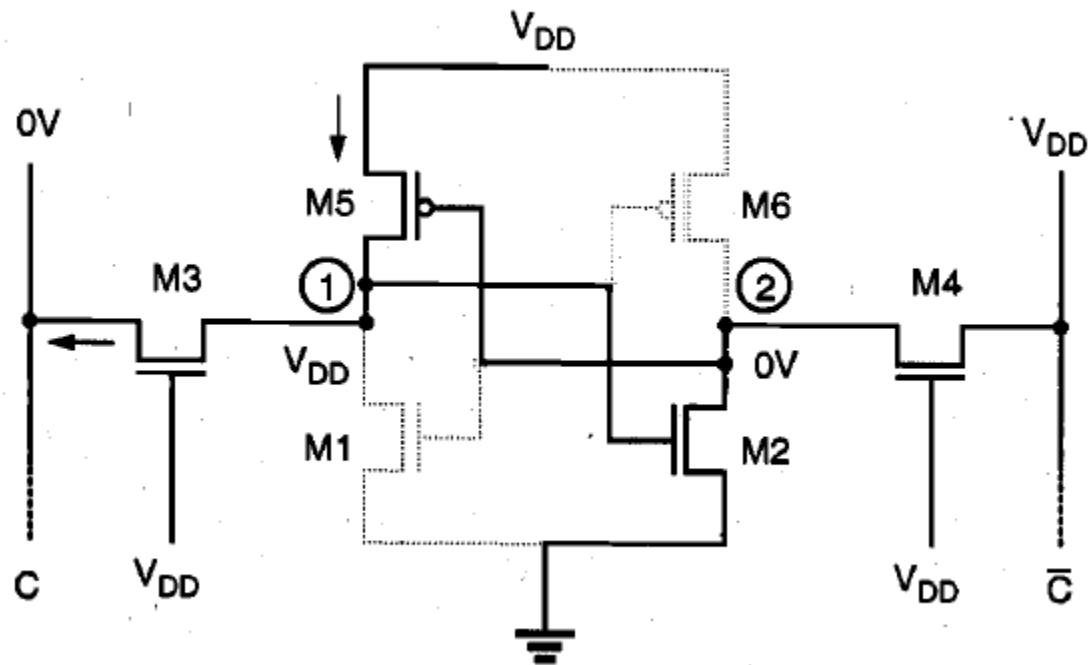


Figure 16. Voltage levels in the SRAM cell at the beginning of the "write" operation

CMOS SRAM Cell Design Strategy

- The column voltage V_c is forced to logic "0" level by the data-write circuitry; thus, we may assume that V_c is approximately equal to 0 V.
- Once the pass transistors M_3 and M_4 are turned on by the row selection circuitry, we expect that the node voltage V_2 remains below the threshold voltage of M_1 , since M_2 and M_4 are designed according to condition (Eqn. 3).

CMOS SRAM Cell Design Strategy

- Consequently, the voltage level at node (2) would not be sufficient to turn on M_1 .
- To change the stored information, i.e., to force V_1 to 0 V and V_2 to V_{DD} , the node voltage V_1 must be reduced below the threshold voltage of M2, so that M2 turns off first.
- When $V_1 = V_{T,n}$ the transistor M3 operates in the linear region while M5 operates in saturation.

CMOS SRAM Cell Design Strategy

$$\frac{k_{p,5}}{2} (0 - V_{DD} - V_{T,p})^2 = \frac{k_{n,3}}{2} (2(V_{DD} - V_{T,n})V_{T,n} - V_{T,n}^2) \quad \dots\dots(4)$$

$$\frac{k_{p,5}}{k_{n,3}} < \frac{2(V_{DD} - 1.5V_{T,n})V_{T,n}}{(V_{DD} + V_{T,p})^2}$$
$$\frac{\left(\frac{W}{L}\right)_5}{\left(\frac{W}{L}\right)_3} < \frac{\mu_n}{\mu_p} \cdot \frac{2(V_{DD} - 1.5V_{T,n})V_{T,n}}{(V_{DD} + V_{T,p})^2} \quad \dots\dots(5)$$

CMOS SRAM Cell Design Strategy

- To summarize, the transistor M2 will be forced into cut-off mode during the write "0" operation if condition (Eqn. 5) is satisfied.
- This will guarantee that M1 subsequently turns on, changing the stored information.
- Note that a symmetrical condition also dictates the aspect ratios of M6 and M4.

Dynamic Read-Write Memory (DRAM) Circuits

- All of the static RAM cells examined in the previous section consist of a two-inverter latch circuit, which is accessed for "read" and "write" operations via two pass transistors.
- Consequently, the SRAM cells require four to six transistors per bit, and four to five lines connecting to each cell, including the power and ground connections.

Dynamic Read-Write Memory (DRAM) Circuits

- To satisfy these requirements, a substantial silicon area must be reserved for each memory cell.
- In addition, most SRAM cells have non-negligible standby (static) power dissipation, with the exception of the full CMOS SRAM cell.

Dynamic Read-Write Memory (DRAM) Circuits

- As the trend for high-density RAM arrays forces the memory cell size to shrink, alternative data storage concepts must be considered to accommodate these demands.
- In a dynamic RAM cell, binary data is stored simply as charge in a capacitor, where the presence or absence of stored charge determines the value of the stored bit.

Dynamic Read-Write Memory (DRAM) Circuits

- Note that the data stored as charge in a capacitor cannot be retained indefinitely, because the leakage currents eventually remove or modify the stored charge.
- Thus, all dynamic memory cells require a periodic refreshing of the stored data, so that unwanted modifications due to leakage are prevented before they occur.

Dynamic Read-Write Memory (DRAM) Circuits

- The use of a capacitor as the primary storage device generally enables the DRAM cell to be realized on a much smaller silicon area compared to the typical SRAM cell.
- Notice that even as the binary data is stored as charge in a capacitor, the DRAM cell must have access devices, or switches, which can be activated externally for "read" and "write" operations.

Dynamic Read-Write Memory (DRAM) Circuits

- But this requirement does not significantly affect the area advantage over the SRAM cell, since the cell access circuitry is usually very simple.
- Also, no static power is dissipated for storing charge on the capacitance. Consequently, dynamic RAM arrays can achieve higher integration densities than SRAM arrays.

Dynamic Read-Write Memory (DRAM) Circuits

- Note that a DRAM array also requires additional peripheral circuitry for scheduling and performing the periodic data refresh operations.
- The hardware overhead of the refresh circuitry, however, does not overshadow the area advantages gained by the small cell size.

Dynamic Read-Write Memory (DRAM) Circuits

- Figure 17 shows some of the steps in the historical evolution of the DRAM cell.
- The four-transistor cell shown in Fig. 17(a) is the simplest and one of the earliest dynamic memory cells.
- This cell is derived from the six-transistor static RAM cell by removing the load devices. The cell has in fact two storage nodes, i.e., the parasitic oxide and diffusion capacitances of the nodes indicated in the circuit diagram.

Dynamic Read-Write Memory (DRAM) Circuits

- Since no current path is provided to the storage nodes for restoring the charge being lost to leakage, the cell must be refreshed periodically.
- It is obvious that the four-transistor dynamic RAM cell can have only a marginal area advantage over the six-transistor SRAM cell.

Dynamic Read-Write Memory (DRAM) Circuits

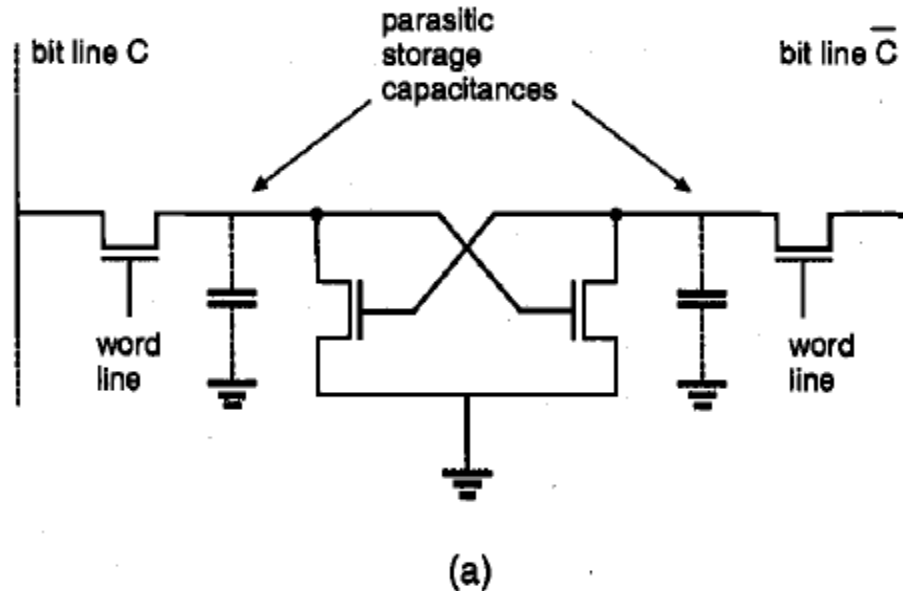


Figure 17. Various configurations of the dynamic RAM cell. (a) Four-transistor DRAM cell with two storage nodes

Dynamic Read-Write Memory (DRAM) Circuits

- The three-transistor DRAM cell shown in Fig. 17(b) was the first widely used dynamic memory cell.
- It utilizes a single transistor as the storage device (where the transistor is turned on or off depending on the charge stored on its gate capacitance), and one transistor each for the "read" and "write" access switches.

Dynamic Read-Write Memory (DRAM) Circuits

- The cell has two control and two I/O lines. Its separate read and write select lines make it relatively fast, but the four lines with their additional contacts tend to increase the cell area.

Dynamic Read-Write Memory (DRAM) Circuits

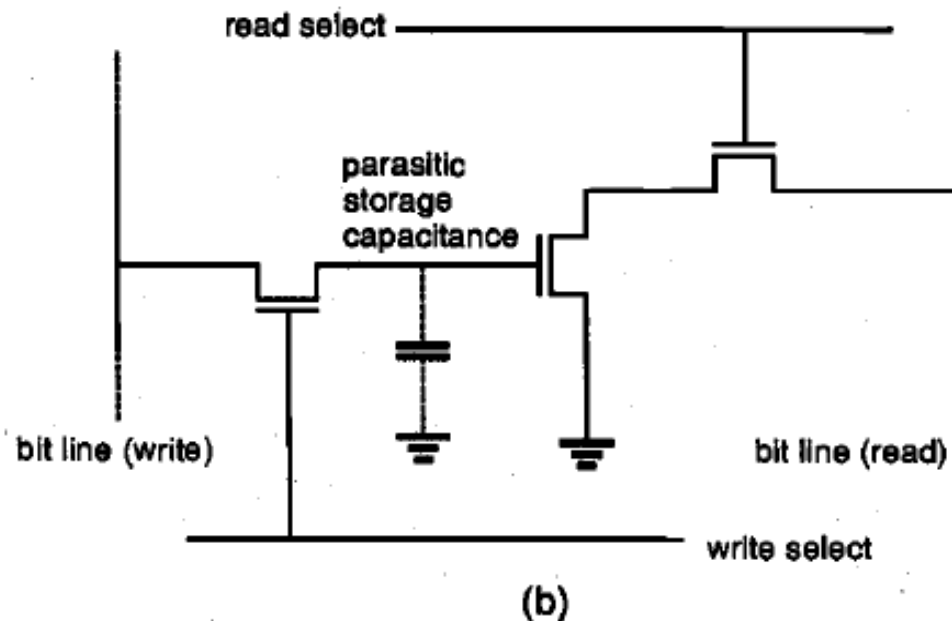


Figure 17. Various configurations of the dynamic RAM cell. (b) Three-transistor DRAM cell with two bit lines and two word lines.

Dynamic Read-Write Memory (DRAM) Circuits

- The one-transistor DRAM cell shown in Fig. 17(c) has become the industry standard dynamic RAM cell in high-density DRAM arrays.
- With only one transistor and one capacitor, it has the lowest component count and, hence, the smallest silicon area of all the dynamic memory cells.
- The cell has one read-write control line (word line) and one I/O line (bit line).

Dynamic Read-Write Memory (DRAM) Circuits

- We have to emphasize at this point that, unlike in the other dynamic memory cells shown in Figs. 17(a) and 17(b), the storage capacitance of the one transistor DRAM cell is explicit.
- This means that a separate capacitor must be manufactured for each storage cell, instead of relying on the parasitic oxide and diffusion capacitances of the transistors for data storage.

Dynamic Read-Write Memory (DRAM) Circuits

- The word line of the one-transistor DRAM cell is controlled by the row address decoder.
- Once the selected transistor is turned on, the charge stored in the capacitor can be detected and/or modified through the bit line.

Dynamic Read-Write Memory (DRAM) Circuits

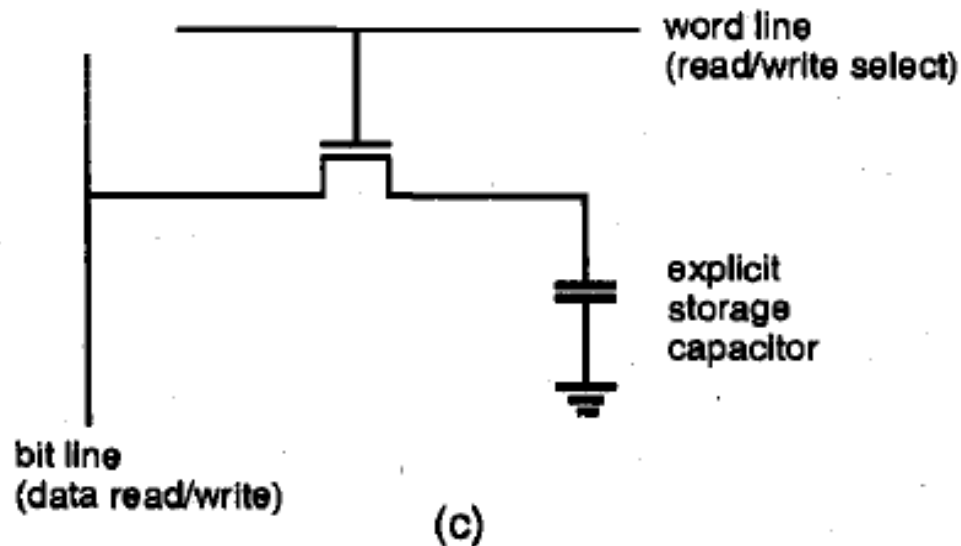


Figure 17. Various configurations of the dynamic RAM cell (c) One-transistor DRAM cell with one bit line and one word line.

Three-Transistor DRAM Cell

- The circuit diagram of a typical three-transistor dynamic RAM cell is shown in Fig. 18 as well as the column pull-up (precharge) transistors and the column read/write circuitry.
- Here, the binary information is stored in the form of charge in the parasitic node capacitance C_1 .

Three-Transistor DRAM Cell

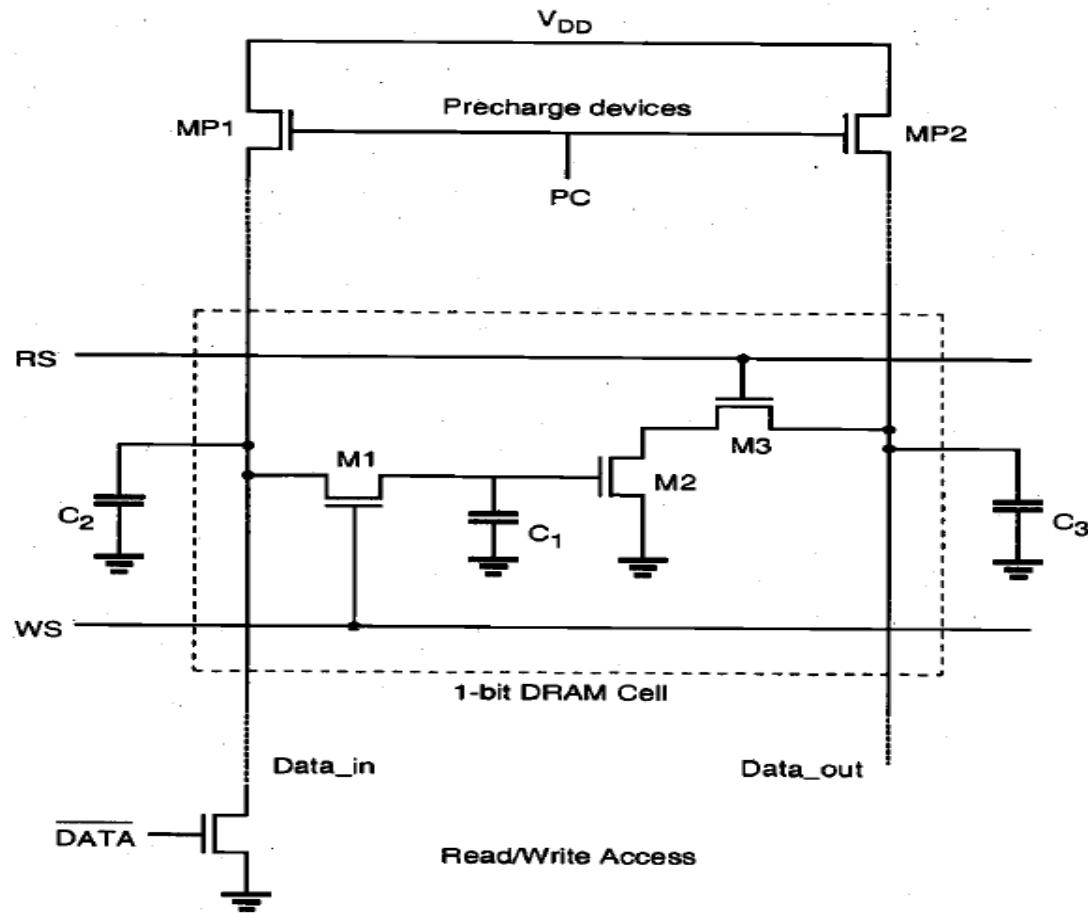


Figure 18. Three-transistor DRAM cell with the pull-up and read/write circuitry

Three-Transistor DRAM Cell

- The storage transistor M2 is turned on or off depending on the charge stored in C_1 , and the pass transistors M1 and M3 act as access switches for data read and write operations. The cell has two separate bit lines for "data read" and "data write," and two separate word lines to control the access transistors.

Three-Transistor DRAM Cell

- The operation of the three-transistor DRAM cell and its peripheral circuitry is based on a two-phase non-overlapping clock scheme.
- The precharge events are driven by ϕ_1 , whereas the "read" and "write" events are driven by ϕ_2 .
- Every "data read" and "data write" operation is preceded by a precharge cycle, which is initiated with the precharge signal PC going high.

Three-Transistor DRAM Cell

- During the precharge cycle, the column pull-up transistors are activated, and the corresponding column capacitances C_2 and C_3 are charged up to logic-high level.
- With typical enhancement type nMOS pull-up transistors ($V_{T0} \approx 1.0$ V) and a power supply voltage of 5 V, the voltage level of both columns after the precharge is approximately equal to 3.5 V.

Three-Transistor DRAM Cell

- All "data read" and "data write" operations are performed during the active ϕ_2 phase, i.e., when PC is low.
- Figure 19 depicts the typical voltage waveforms associated with the 3-T DRAM cell during a sequence of four consecutive operations: write "1," read "1," write "0," and read "0."
- The four precharge cycles shown in Fig. 19 are numbered 1, 3, 5, and 7, respectively.

Three-Transistor DRAM Cell

- Figure 20 illustrates the transient currents charging up the two columns (D_{in} and D_{out}) during a precharge cycle.
- The precharge cycle is effectively completed when both capacitance voltages reach their steady-state values.
- Note here that the two column capacitances C_2 and C_3 are at least one order of magnitude larger than the internal storage capacitance C_1 .

Three-Transistor DRAM Cell

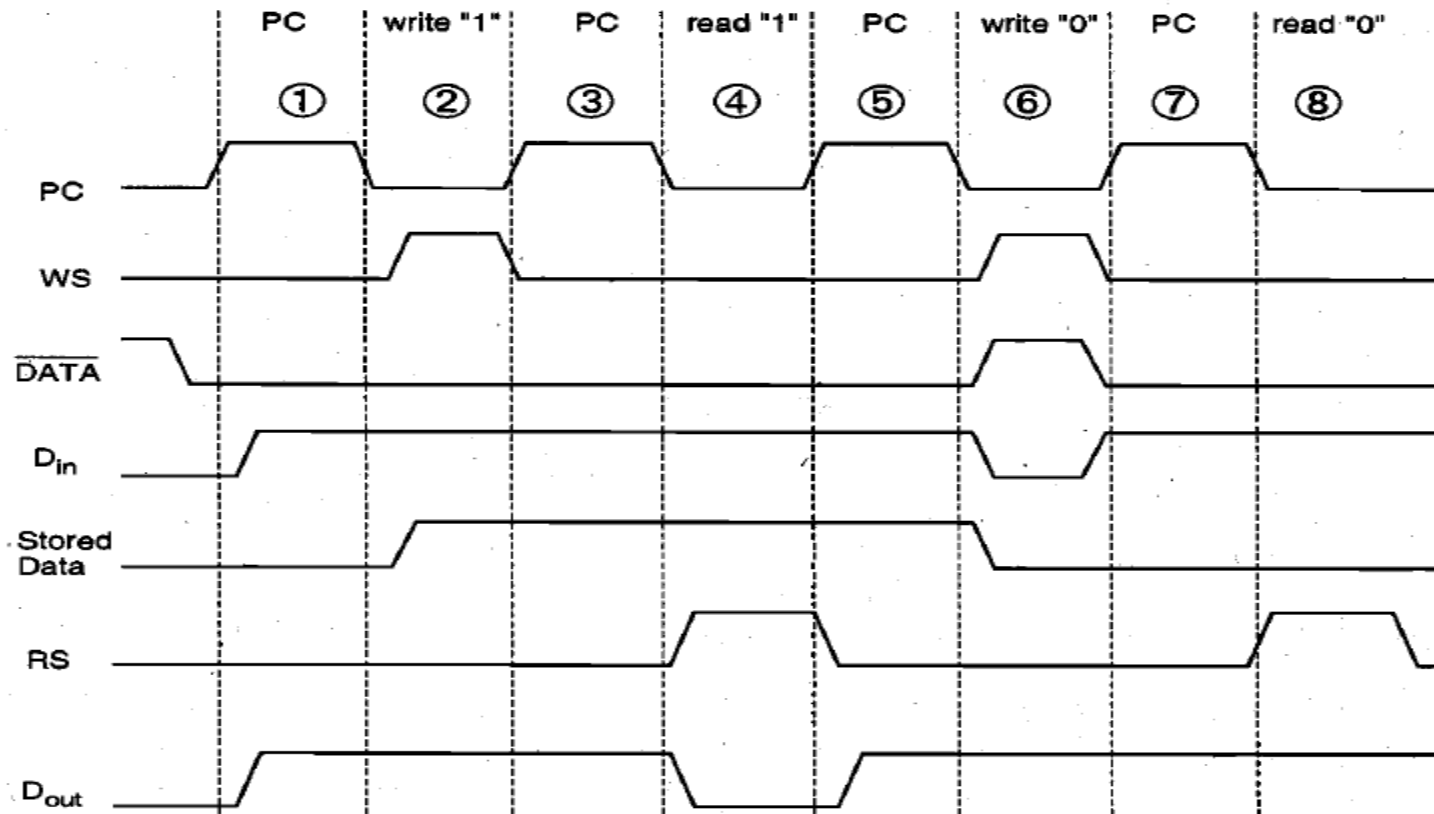


Figure 19. Typical voltage waveforms associated with the 3-T DRAM cell during four consecutive operations: write "1," read "1," write "0," and read "0."

Three-Transistor DRAM Cell

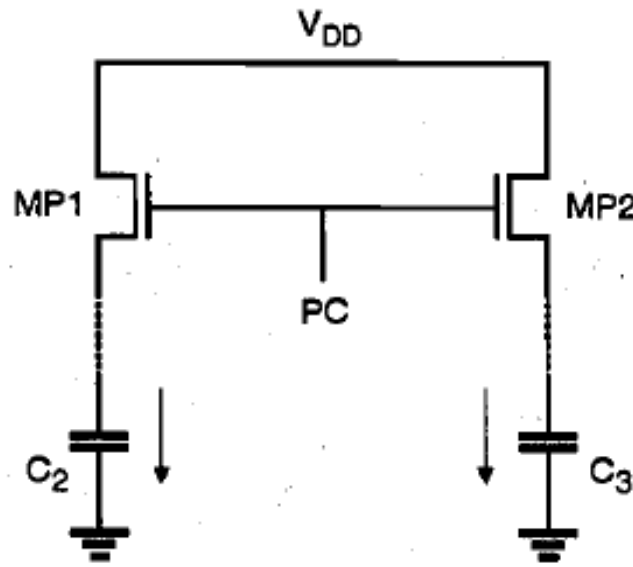


Figure 20. Column capacitances C_2 and C_3 are being charged-up through MP1 and MP2 during the precharge cycle.

Three-Transistor DRAM Cell

- For the write "1" operation, the inverse data input is at the logic-low level, because the data to be written onto the DRAM cell is logic "1."
- Consequently, the "data write" transistor MD is turned off, and the voltage level on column D_{in} remains high.
- Now, the "write select" signal WS is pulled high during the active phase of ϕ_2 .

Three-Transistor DRAM Cell

- As a result, the write access transistor M1 is turned on.
- With M1 conducting, the charge on C_2 is now shared with C_1 (Fig. 21).
- Since the capacitance C_2 is very large compared to C_1 , the storage node capacitance C_1 attains approximately the same logic-high level as the column capacitance C_2 at the end of the charge-sharing process.

Three-Transistor DRAM Cell

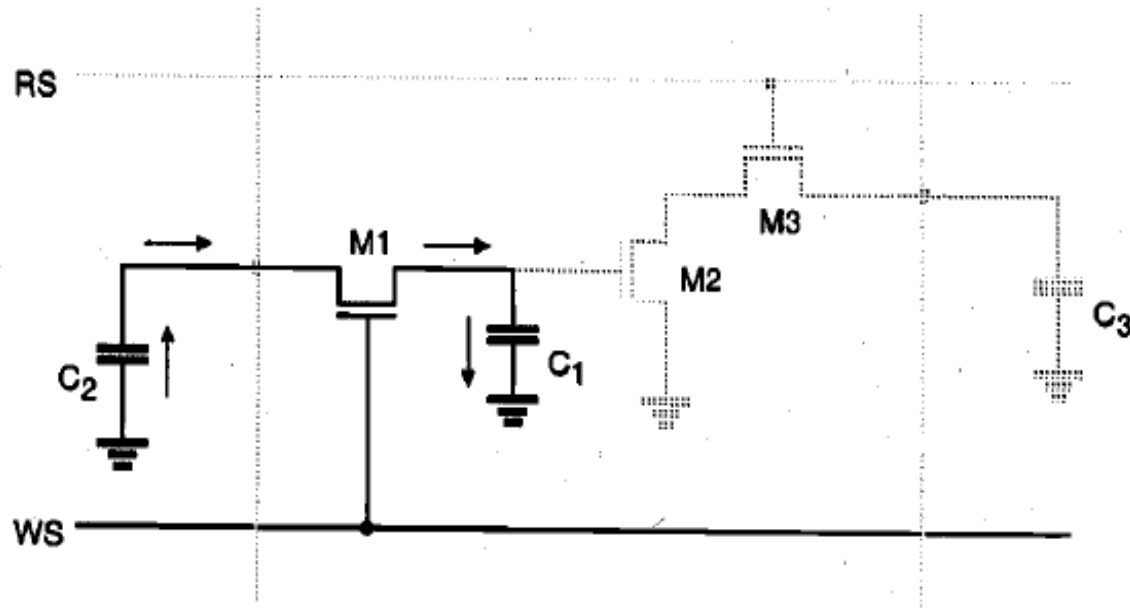


Figure 21. Charge sharing between C_2 and C_1 during the write " 1 " sequence

Three-Transistor DRAM Cell

- After the write "1" operation is completed, the write access transistor M1 is turned off.
- With the storage capacitance C1 charged-up to a logic-high level, transistor M2 is now conducting.
- In order to read this stored "1," the "read select" signal RS must be pulled high during the active phase of ϕ_2 , following a precharge cycle.

Three-Transistor DRAM Cell

- As the read access transistor M3 turns on, M2 and M3 create a conducting path between the "data read" column capacitance C_3 and the ground.
- The capacitance C_3 discharges through M2 and M3, and the falling column voltage is interpreted by the "data read" circuitry as a stored logic "1."

Three-Transistor DRAM Cell

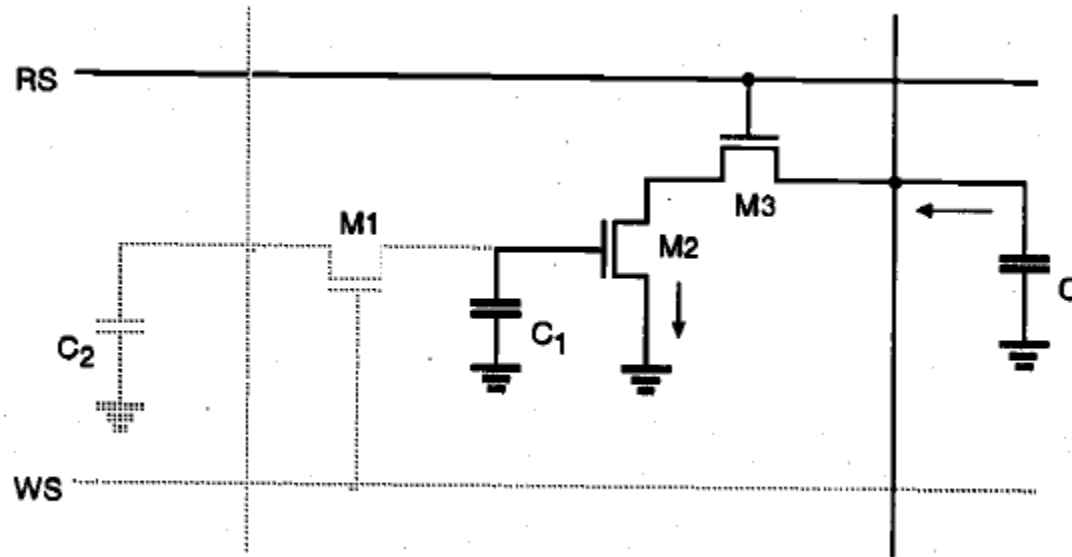


Figure 22. The column capacitance C_3 is discharged through the transistors M2 and M3 during the read "1" operation

Three-Transistor DRAM Cell

- The active portion of the DRAM cell during the read "1" cycle is shown in Fig. 22.
- Note that the 3-T DRAM cell may be read repeatedly in this fashion without disturbing the charge stored in C_1 .
- For the write "0" operation, the inverse data input is at the logic-high level, because the data to be written onto the DRAM cell is a logic "0."

Three-Transistor DRAM Cell

- Consequently, the data write transistor is turned on, and the voltage level on column D_{in} is pulled to logic "0."
- Now, the "write select" signal WS is pulled high during the active phase of ϕ_2 .
- As a result, the write access transistor M1 is turned on.
- The voltage level on C_2 , as well as that on the storage node C_1 , is pulled to logic "0" through M1 and the data write transistor, as shown in Fig. 23.

Three-Transistor DRAM Cell

- Thus, at the end of the write "0" sequence, the storage capacitance C_1 contains a very low charge, and the transistor M2 is turned off since its gate voltage is approximately equal to zero.

Three-Transistor DRAM Cell

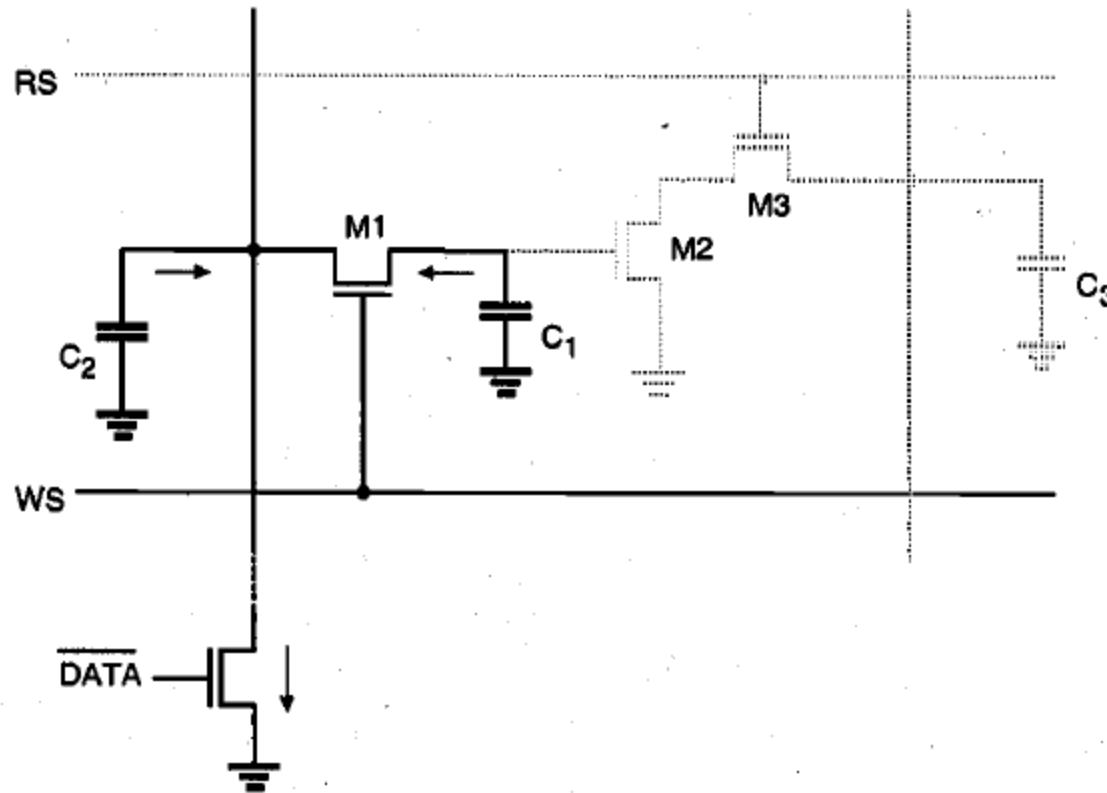


Figure 23. Both C_1 and C_2 are discharged via M1 and the data write transistor during the write "0" sequence

Three-Transistor DRAM Cell

- In order to read this stored "0," the "read select" signal RS must be pulled high during the active phase of ϕ_2 , following a precharge cycle.
- The read access transistor M3 turns on, but since M2 is off, there is no conducting path between the column capacitance C_3 and the ground (Fig. 24).

Three-Transistor DRAM Cell

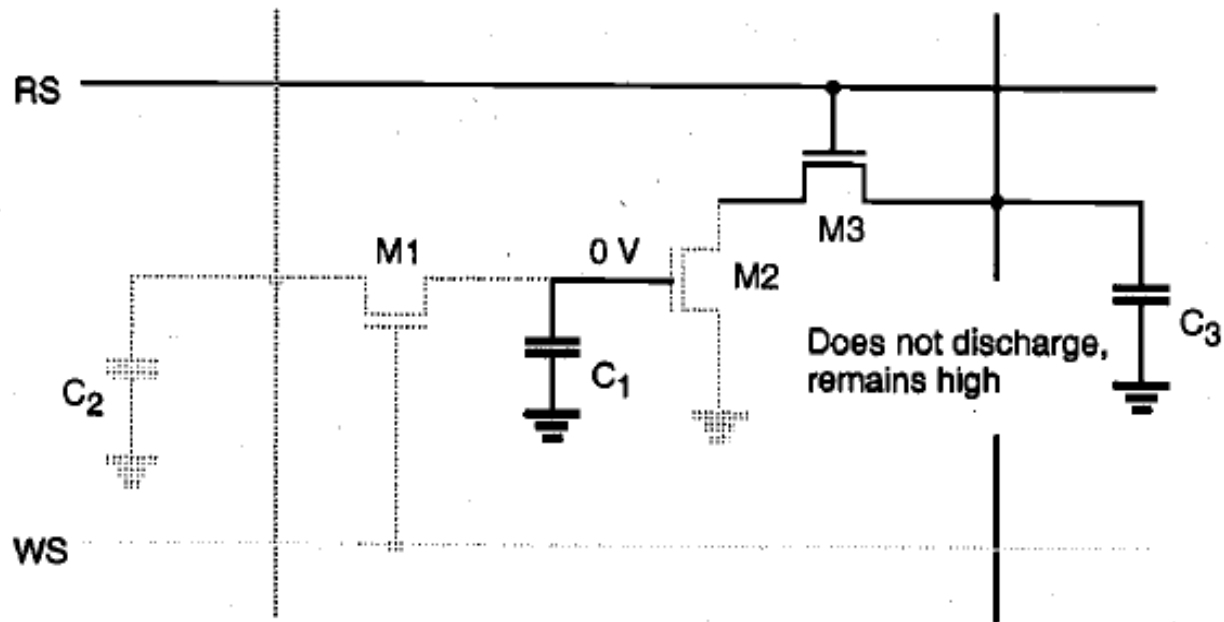


Figure 24. The column capacitance C_3 cannot discharge during the read "0" cycle.

Three-Transistor DRAM Cell

- Consequently, C_3 does not discharge, and the logic-high level on the D_{out} column is interpreted by the data read circuitry as a stored "0" bit.
- As we already pointed out in the beginning of this section, the charge stored in C_1 cannot be held indefinitely, even though the "data read" operations do not significantly disturb the stored charge.

Three-Transistor DRAM Cell

- The drain junction leakage current of the write access transistor M1 is the main reason for the gradual depletion of the stored charge on C_1 .
- In order to refresh the data stored in the DRAM cells before they are altered due to leakage, the data must be periodically read, inverted (since the data output level reflects the inverse of the stored data), and then written back into the same cell location.

Three-Transistor DRAM Cell

- This refresh operation is performed for all storage cells in the DRAM array every 2 to 4 ms.
- Note that all bits in one row can be refreshed at once, which significantly simplifies the procedure.

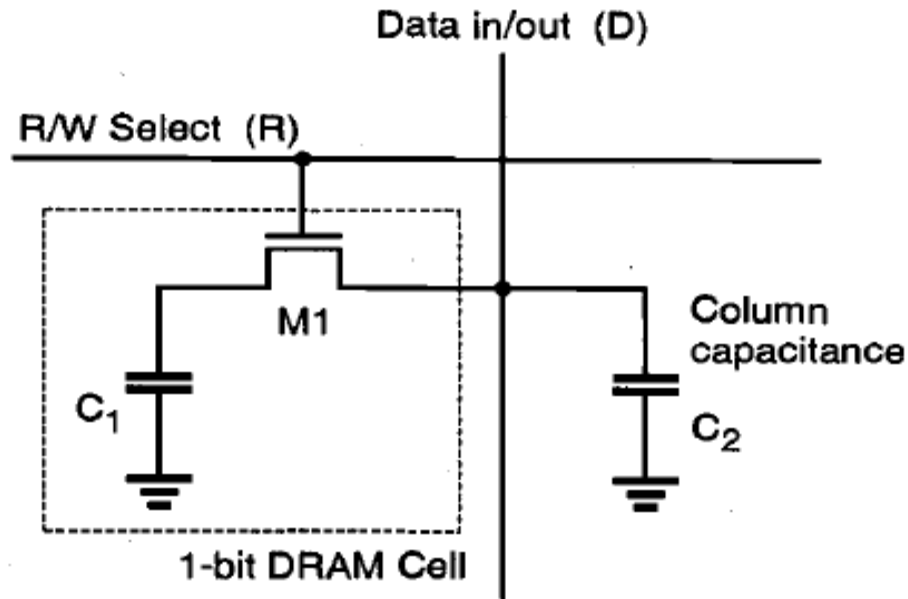
Three-Transistor DRAM Cell

- It can be seen that the three-transistor dynamic RAM cell examined here does not dissipate any static power for data storage, since there is no continuous current flow in the circuit.
- Also, the use of periodic precharge cycles instead of static pull-up further reduces the dynamic power dissipation.
- The additional peripheral circuitry required for scheduling the non-overlapping control signals and the refresh cycles does not significantly overshadow these advantages of the low-power dynamic memory.

One-Transistor DRAM Cell

- Currently, the one-transistor cell is the most widely used storage structure in the DRAM industry.
- The circuit diagram of the one-transistor (1-T) DRAM cell consisting of one explicit storage capacitor and one access transistor is shown in Fig. 25(a).
- Here, C_1 represents the storage capacitor which typically has a value of 30 to 100 fF.
- Similar to the 3-T DRAM cell, binary data are stored as the presence or absence of charge in the storage capacitor.

One-Transistor DRAM Cell



(a)

Figure 25. (a) Typical one-transistor (1-T) DRAM cell with its access lines

One-Transistor DRAM Cell

- Capacitor C_2 represents the much larger parasitic column capacitance associated with the word line.
- Charge sharing between this large capacitance and the very small storage capacitance plays a very important role in the operation of the 1-T DRAM cell.

One-Transistor DRAM Cell

- The "data write" operation on the 1-T cell is quite straightforward.
- For the write "1" operation, the bit line (D) is raised to logic " 1 " by the write circuitry, while the selected word line is pulled high by the row address decoder.

One-Transistor DRAM Cell

- The access transistor M1 turns on, allowing the storage capacitor C_1 to charge up to a logic-high level.
- For the write “0” operation, the bit line (D) is pulled to logic “0” and the word line is pulled high by the row address decoder.
- In this case, the storage capacitor C_1 discharges through the access transistor, resulting in a stored “0” bit.

One-Transistor DRAM Cell

- In order to read stored data out of a 1-T DRAM cell, on the other hand, we have to build a fairly elaborate read-refresh circuit.
- The reason for this is the fact that the "data read" operation on the one-transistor DRAM cell is by necessity a "destructive readout."
- This means that the stored data must be destroyed or lost during the read operation.

One-Transistor DRAM Cell

- Typically, the read operation starts with precharging the column capacitance C_2 .
- Then, the word line is pulled high in order to activate the access transistor M1.
- Charge sharing between C_1 and C_2 occurs and, depending on the amount of stored charge on C_1 the column voltage either increases or decreases slightly.
- Note that charge sharing inevitably destroys the stored charge on C_1 .
- Hence, we also have to refresh data every time we perform a "data read" operation.

SEQUENTIAL MOS LOGIC CIRCUITS

Sequential MOS Logic Circuit

- Figure 1(a) shows a sequential circuit consisting of a combinational circuit and a memory block in the feedback loop.
- In most cases, the regenerative behavior of sequential circuits is due to either a direct or indirect feedback connection between the output and the input.

Sequential MOS Logic Circuit

- The critical components of sequential systems are the basic regenerative circuits, which can be classified into three main groups:
- bistable circuits, monostable circuits, and astable circuits.

Sequential MOS Logic Circuit

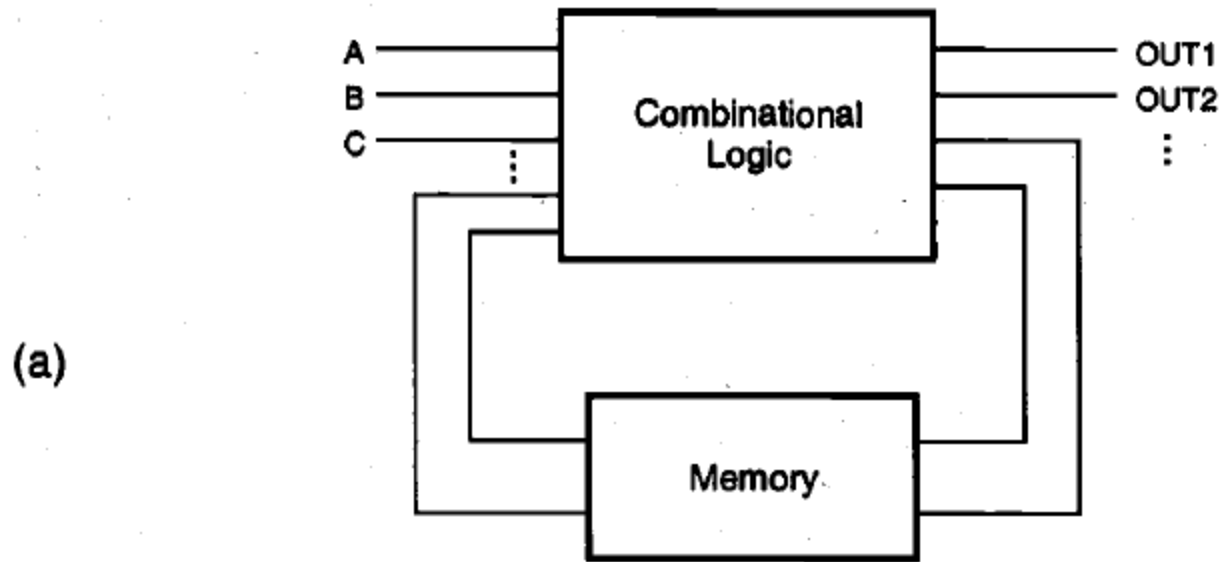


Figure 1. (a) Sequential circuit consisting of a combinational logic block and a memory block in the feedback loop.

Sequential MOS Logic Circuit

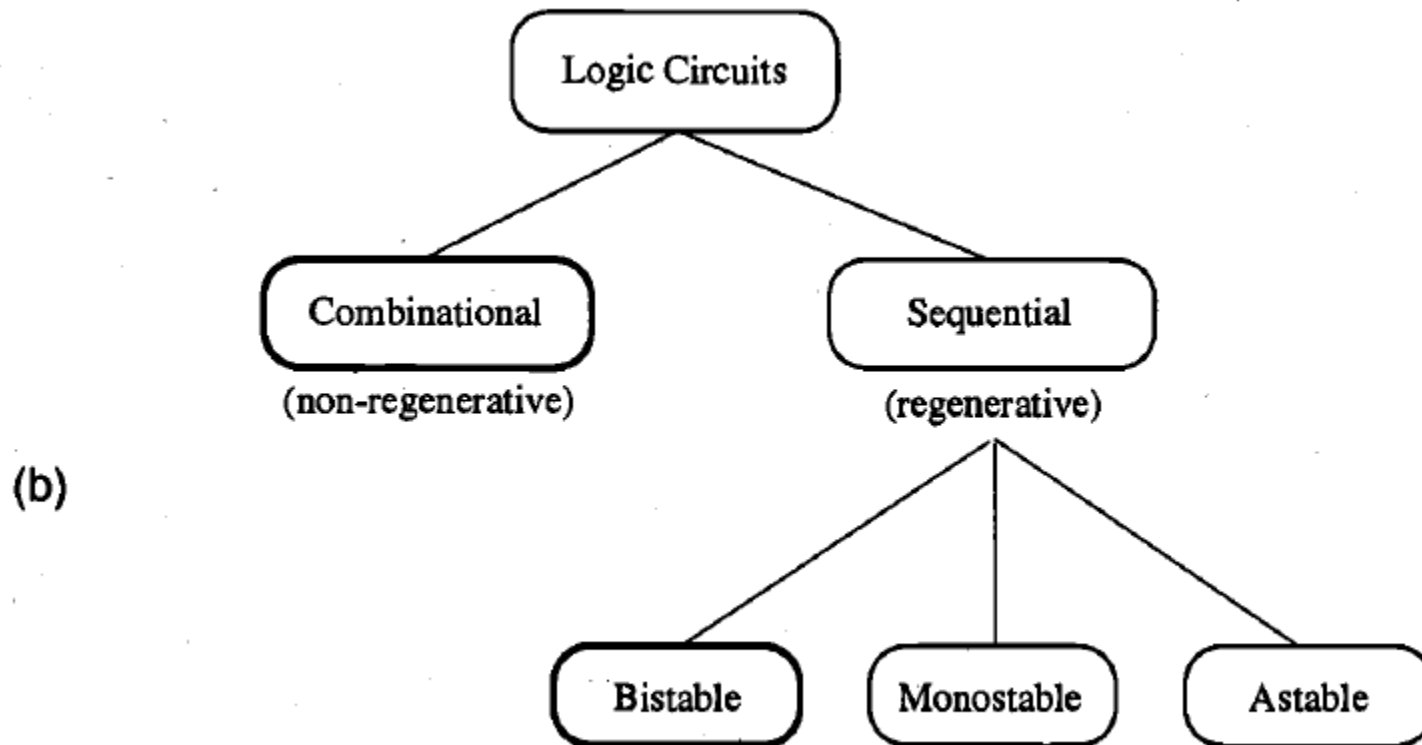


Figure 1. (b) Classification of logic circuits based on their temporal behavior.

Sequential MOS Logic Circuit

- Bistable circuits have, as their name implies, two stable states or operation modes, each of which can be attained under certain input and output conditions.
- Monostable circuits have only one stable operating point (state).
- Even if the circuit experiences an external perturbation, the output eventually returns to the single stable state after a certain time period.

Sequential MOS Logic Circuit

- Finally, in astable circuits, there is no stable operating point or state which the circuit can preserve for a certain time period.
- Consequently, the output of an astable circuit must oscillate without settling into a stable operating mode.
- The ring oscillator circuit is a typical example of an astable circuit.

Sequential MOS Logic Circuit

- Among these three main groups of regenerative circuit types, the bistable circuits are by far the most widely used and the most important class.
- All basic latch and flip-flop circuits, registers, and memory elements used in digital systems fall into this category.

Behavior of Bistable Elements

- The basic bistable element to be examined in this section consists of two identical crosscoupled inverter circuits, as shown in Fig. 2(a).
- Here, the output voltage of inverter (1) is equal to the input voltage of inverter (2), i.e., $v_{o1} = v_{i2}$, and the output voltage of inverter (2) is equal to the input voltage of inverter (1), i.e., $v_{o2} = v_{i1}$.

Behavior of Bistable Elements

- In order to investigate the static input-output behavior of both inverters, we start by plotting the voltage transfer characteristic of inverter (1) with respect to the $v_{o1} - v_{i1}$, axis pair.
- The input and output voltages of inverter (2) correspond to the output and input voltages of inverter (1), respectively.
- Consequently, we can also plot the voltage transfer characteristic of inverter (2) using the same axis pair, as shown in Fig. 2(b).

Behavior of Bistable Elements

- It can be observed that the two voltage transfer characteristics intersect at three points.
- Simple reasoning can help us to conclude that two of these operating points are stable, as indicated in Fig. 2(b).
- If the circuit is initially operating at one of these two stable points, it will preserve this state unless it is forced externally to change its operating point.

Behavior of Bistable Elements

- The gain of each inverter circuit, i.e., the slope of the respective voltage transfer curves, is smaller than unity at the two stable operating points.
- Thus, in order to change the state by moving the operating point from one stable point to the other, a sufficiently large external voltage perturbation must be applied so that the voltage gain of the inverter loop becomes larger than unity.

Behavior of Bistable Elements

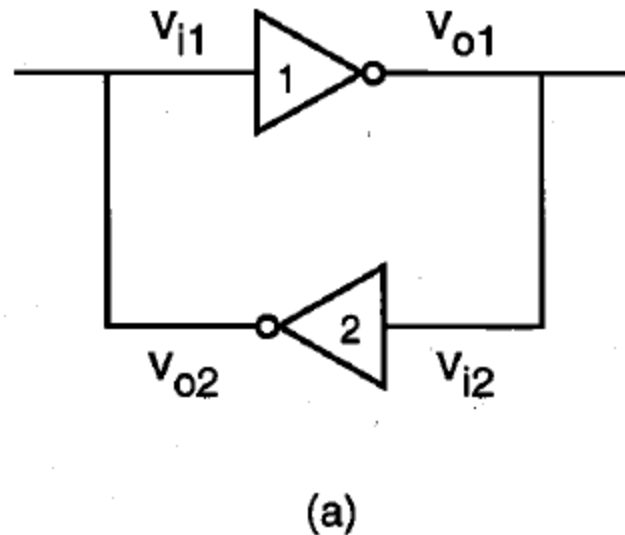


Figure 2. Static behavior of the two-inverter basic bistable element: (a) Circuit schematic. (b) Intersecting voltage transfer curves of the two inverters, showing the three possible operating points. (c) Qualitative view of the potential energy levels corresponding to the three operating points.

Behavior of Bistable Elements

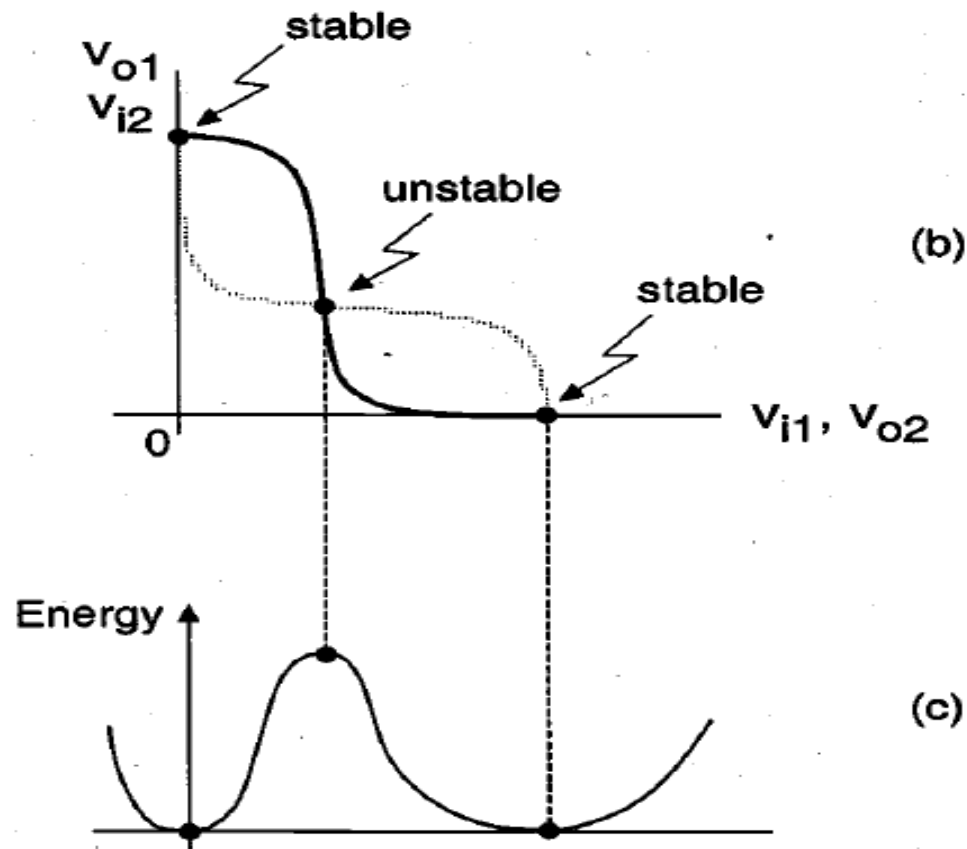


Figure 2. Static behavior of the two-inverter basic bistable element: (a) Circuit schematic. (b) Intersecting voltage transfer curves of the two inverters, showing the three possible operating points. (c) Qualitative view of the potential energy levels corresponding to the three operating points.

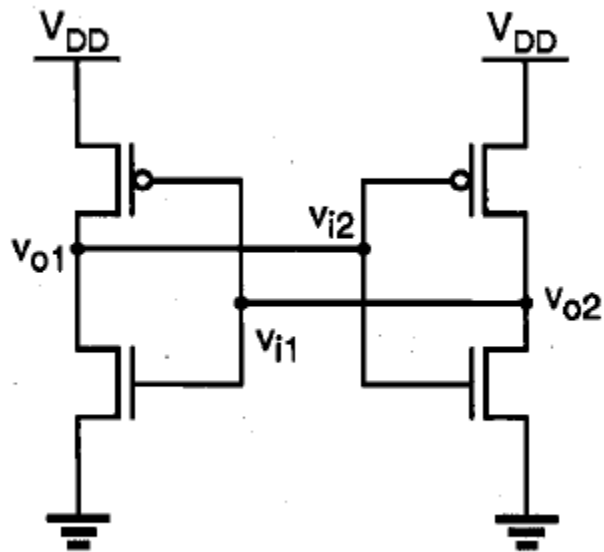
CMOS Bistable Element

- Figure 3(a) shows the circuit diagram of a CMOS two-inverter bistable element.
- At the unstable operating point of this circuit, all four transistors are in saturation, resulting in maximum loop gain for the circuit.
- If the initial operating condition is set at this point, any small voltage perturbation will cause significant changes in the operating modes of the transistors.

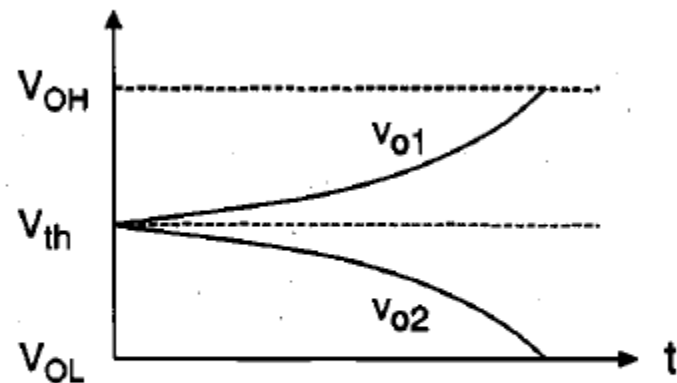
CMOS Bistable Element

- Thus, we expect the output voltages of the two inverters to diverge and eventually settle at V_{OH} and V_{OL} , respectively, as illustrated in Fig. 3(b).

CMOS Bistable Element



(a)



(b)

Figure 3. (a) Circuit diagram of a CMOS bistable element. (b) One possibility for the expected time-domain behavior of the output voltages, if the circuit is initially set at its unstable operating point.

Small-signal input and output currents of the inverters

- Consider the bistable circuit shown in Fig. 4, which is initially operating at $v_{o1} = v_{o2} = V_{th}$, i.e., at the unstable operating point.
- For analysis assume that the input (gate) capacitance C_g of each inverter is much larger than its output (drain) capacitance C_d , i.e., $C_g \gg C_d$.

Small-signal input and output currents of the inverters

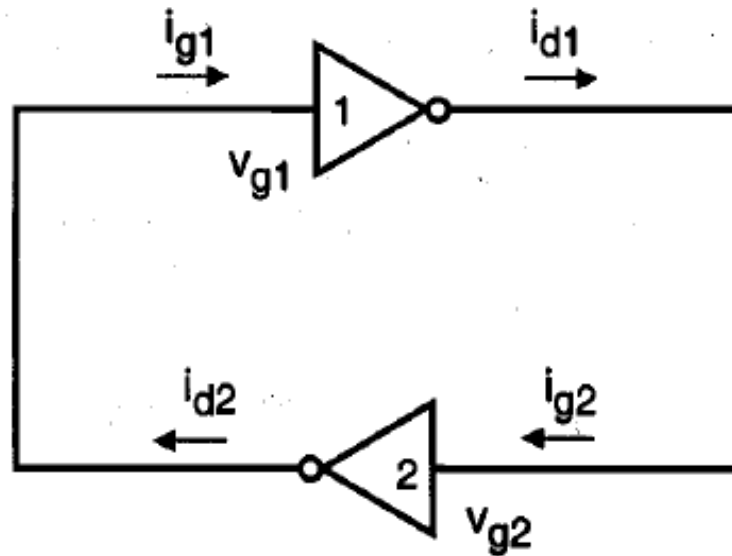


Figure 4. Small-signal input and output currents of the inverters.

Small-signal input and output currents of the inverters

- The small-signal drain current supplied by each inverter (1 and 2) can be expressed, in terms of the small-signal gate voltage of that inverter, as follows.
- Note that the drain current of each inverter is also equal to the gate current of the other inverter.

$$i_{g1} = i_{d2} = g_m v_{g2}$$

$$i_{g2} = i_{d1} = g_m v_{g1}$$

.....(1)

Small-signal input and output currents of the inverters

- Here, g_m represents the small-signal transconductance of the inverter.
- The gate voltages of both inverters can be expressed in terms of the gate charges, q_1 and q_2 .

$$v_{g1} = \frac{q_1}{C_g} \quad v_{g2} = \frac{q_2}{C_g} \quad \dots\dots(2)$$

Small-signal input and output currents of the inverters

- The small-signal gate current of each inverter can be written as a function of the time derivative of its small-signal gate voltage, as follows.

$$i_{g1} = C_g \frac{dv_{g1}}{dt} \quad \text{.....(3)}$$
$$i_{g2} = C_g \frac{dv_{g2}}{dt}$$

Small-signal input and output currents of the inverters

- Combining (Eqn. 1) with (Eqn. 3), we obtain:

$$g_m v_{g2} = C_g \frac{dv_{g1}}{dt} \quad \text{.....(4)}$$

$$g_m v_{g1} = C_g \frac{dv_{g2}}{dt} \quad \text{.....(5)}$$

Small-signal input and output currents of the inverters

- Expressing the gate voltages in terms of the gate charges, these two differential equations can also be written as

$$\frac{g_m}{C_g} q_2 = \frac{dq_1}{dt} \quad \text{.....(6)}$$

$$\frac{g_m}{C_g} q_1 = \frac{dq_2}{dt} \quad \text{.....(7)}$$

Small-signal input and output currents of the inverters

- Both differential equations given in (Eqn. 6) and (Eqn. 7) can now be combined to yield a second-order differential equation describing the time behavior of gate charge q_1 .

$$\frac{g_m}{C_g} q_1 = \frac{C_g}{g_m} \frac{d^2 q_1}{dt^2} \Rightarrow \frac{d^2 q_1}{dt^2} = \left(\frac{g_m}{C_g} \right)^2 q_1 \quad \dots\dots(8)$$

Small-signal input and output currents of the inverters

- This equation can also be expressed in a more simplified form by using τ_o , the transit time constant.

$$\frac{d^2 q_1}{dt^2} = \frac{1}{\tau_0^2} q_1 \quad \text{with } \tau_0 = \frac{C_g}{g_m} \quad \text{.....(9)}$$

Small-signal input and output currents of the inverters

- The time-domain solution of (Eqn. 9) for q_1 yields

$$q_1(t) = \frac{q_1(0) - \tau_0 q_1'(0)}{2} e^{-\frac{t}{\tau_0}} + \frac{q_1(0) + \tau_0 q_1'(0)}{2} e^{+\frac{t}{\tau_0}} \quad \text{.....(10)}$$

- where the initial condition is given as:

$$q_1(0) = C_g \cdot v_{g1}(0) \quad \text{.....(11)}$$

Small-signal input and output currents of the inverters

- Note that $v_{g1} = v_{o2}$ and $v_{o2} = v_{g1}$.
- Replacing the gate charge of both inverters with the corresponding output-voltage variables, we obtain,

$$v_{o2}(t) = \frac{1}{2} \left(v_{o2}(0) - \tau_0 v_{o2}'(0) \right) e^{-\frac{t}{\tau_0}} + \frac{1}{2} \left(v_{o2}(0) + \tau_0 v_{o2}'(0) \right) e^{+\frac{t}{\tau_0}} \quad \text{.....(12)}$$

$$v_{o1}(t) = \frac{1}{2} \left(v_{o1}(0) - \tau_0 v_{o1}'(0) \right) e^{-\frac{t}{\tau_0}} + \frac{1}{2} \left(v_{o1}(0) + \tau_0 v_{o1}'(0) \right) e^{+\frac{t}{\tau_0}} \quad \text{.....(13)}$$

Small-signal input and output currents of the inverters

- For large values of t , the time-domain expressions (Eqn. 12) and (Eqn. 13) can be simplified as follows.

$$\begin{aligned} v_{o1}(t) &\approx \frac{1}{2} \left(v_{o1}(0) + \tau_0 v_{o1}'(0) \right) e^{+\frac{t}{\tau_0}} \\ v_{o2}(t) &\approx \frac{1}{2} \left(v_{o2}(0) + \tau_0 v_{o2}'(0) \right) e^{+\frac{t}{\tau_0}} \end{aligned} \quad \text{.....(14)}$$

The SR Latch Circuit

- Figure 5 shows the circuit structure of the simple CMOS SR latch, which has two such triggering inputs, S (set) and R (reset).
- In the literature, the SR latch is also called an SR flip-flop, since two stable states can be switched back and forth.
- The circuit consists of two CMOS NOR2 gates.
- One of the input terminals of each NOR gate is used to cross-couple to the output of the other NOR gate, while the second input enables triggering of the circuit.

The SR Latch Circuit

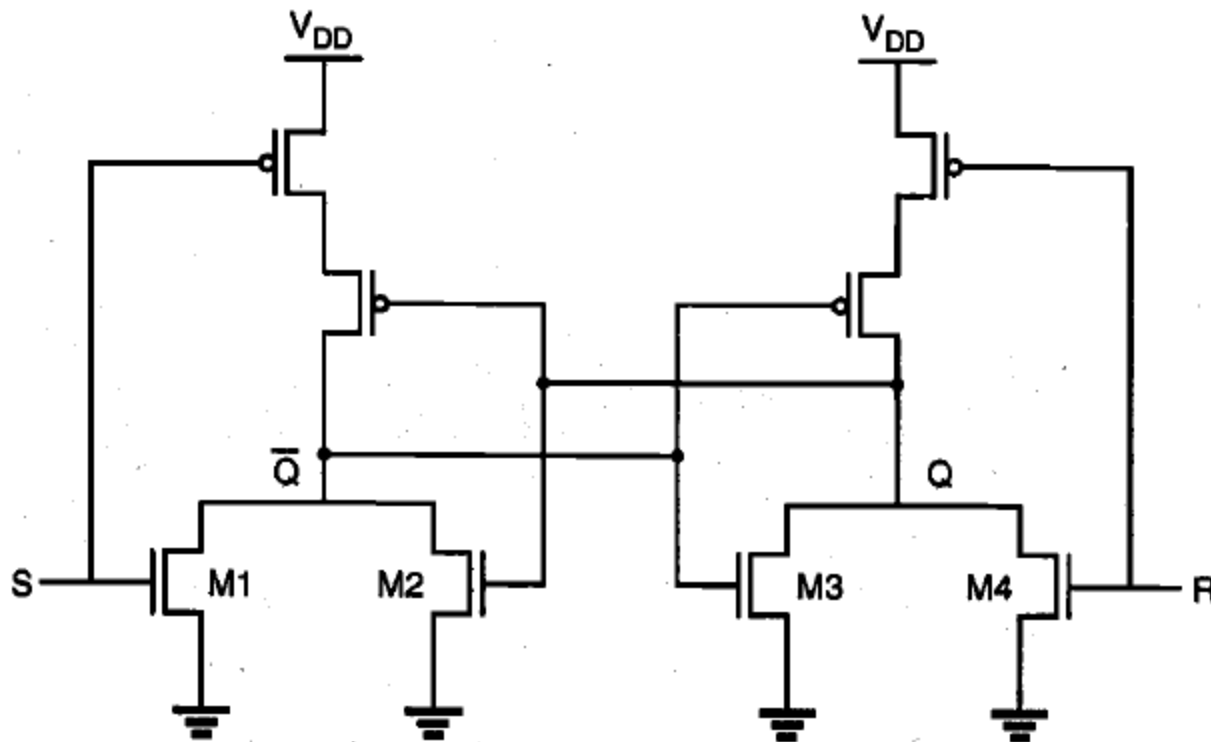


Figure 5 CMOS SR latch circuit based on NOR2 gates.

The SR Latch Circuit

- The SR latch circuit has two complementary outputs, Q and $\bar{Q}$.
- By definition, the latch is said to be in its set state when Q is equal to logic "1" and $\bar{Q}$ is equal to logic "0."
- Conversely, the latch is in its reset state when the output Q is equal to logic "0" and $\bar{Q}$ is equal to "1."
- The gate-level schematic of the SR latch consisting of two NOR2 gates, and the corresponding block diagram representation are shown in Fig. 6.

The SR Latch Circuit

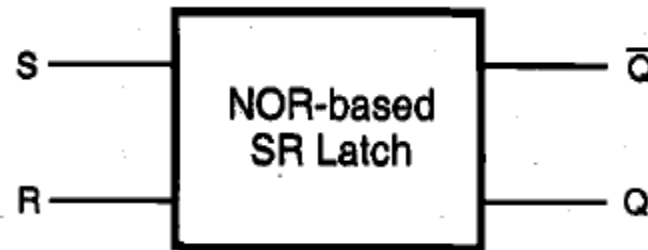
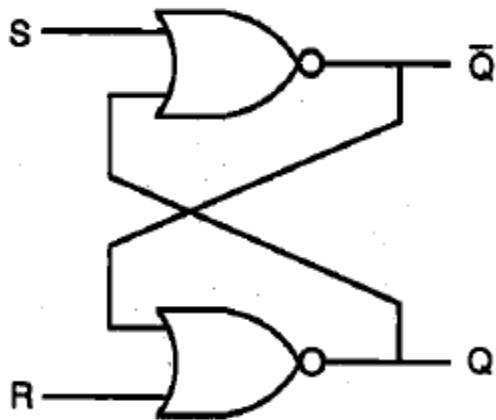


Fig. 6: Gate-level schematic and block diagram of the NOR-based SR latch

The SR Latch Circuit

- It can easily be seen that when both input signals are equal to logic "0," the SR latch will operate exactly like the simple cross-coupled bistable element examined earlier, i.e., it will preserve (hold) either one of its two stable operating points (states) as determined by the previous inputs.

The SR Latch Circuit

- If the set input (S) is equal to logic "1" and the reset input is equal to logic "0," then the output node Q will be forced to logic " 1 " while the output node $\bar{Q}$ is forced to logic "0."
- This means that the SR latch will be set regardless of its previous state.

The SR Latch Circuit

- Similarly, if S is equal to "0" and R is equal to "1" then the output node Q will be forced to "0" while $\bar{Q}$ is forced to "1."
- Thus, with this input combination, the latch is reset, regardless of its previously held state.
- Finally, let us consider the case in which both of the inputs S and R are equal to logic "1".

The SR Latch Circuit

- In this case, both output nodes will be forced to logic "0," which conflicts with the complementarity of Q and $\bar{Q}$.
- Therefore, this input combination is not permitted during normal operation and is considered to be a not allowed condition.
- The truth table of the NOR-based SR latch is summarized in the following:

The SR Latch Circuit

S	R	Q_{n+1}	$\overline{Q_{n+1}}$	Operation
0	0	Q_n	$\overline{Q_n}$	hold
1	0	1	0	set
0	1	0	1	reset
1	1	0	0	not allowed

The SR Latch Circuit

- The operation of the CMOS SR latch circuit shown in Fig. 5 can be examined in more detail by considering the operating modes of the four nMOS transistors, M_1 , M_2 , M_3 , and M_4 .
- If the set input (S) is equal to V_{OH} and the reset input (R) is equal to V_{OL} , both of the parallel-connected transistors M_1 and M_2 will be on.

The SR Latch Circuit

- Consequently, the voltage on node $\bar{Q}$ will assume a logic-low level of $V_{OL} = 0$. At the same time, both M_3 and M_4 are turned off, which results in a logic-high voltage V_{OH} at node Q .
- If the reset input (R) is equal to V_{OH} and the set input (S) is equal to V_{OL} , the situation will be reversed (M_1 and M_2 turned off and M_3 and M_4 turned on).

The SR Latch Circuit

- When both of the input voltages are equal to V_{OL} , there are two possibilities.
- Depending on the previous state of the SR latch, either M_2 or M_3 will be on, while both of the trigger transistors M_1 and M_4 are off.
- This will generate a logic-low level of $V_{OL} = 0$ at one of the output nodes, while the complementary output node is at V_{OH} .

The SR Latch Circuit

- The static operation modes and voltage levels of the NOR-based CMOS SR latch circuit are summarized in the following table.
- For simplicity, the operating modes of the complementary pMOS transistors are not explicitly listed here.

The SR Latch Circuit

S	R	Q_{n+1}	$\overline{Q}_{n+1}$	Operation
V_{OH}	V_{OL}	V_{OH}	V_{OL}	M1 and M2 on, M3 and M4 off
V_{OL}	V_{OH}	V_{OL}	V_{OH}	M1 and M2 off, M3 and M4 on
V_{OL}	V_{OL}	V_{OH}	V_{OL}	M1 and M4 off, M2 on, or
V_{OL}	V_{OL}	V_{OL}	V_{OH}	M1 and M4 off, M3 on

The SR Latch Circuit

- To calculate the switching times for both output nodes, we first have to find the total parasitic capacitance associated with each node.
- The total lumped capacitance at each output node can be approximated as follows:

$$\begin{aligned} C_Q &= C_{gb,2} + C_{gb,5} + C_{db,3} + C_{db,4} + C_{db,7} + C_{sb,7} + C_{db,8} \\ C_{\bar{Q}} &= C_{gb,3} + C_{gb,7} + C_{db,1} + C_{db,2} + C_{db,5} + C_{sb,5} + C_{db,6} \end{aligned} \quad \text{.....(15)}$$

The SR Latch Circuit

- The circuit diagram of the SR latch is shown in Fig. 7, together with the lumped load capacitances at the nodes Q and $\bar{Q}$.
- Assuming that the latch is initially reset and that a set operation is being performed by applying $S = "1"$ and $R = "0,"$ the rise time associated with node Q can now be estimated as follows.

$$\tau_{rise,Q}(SR-latch) = \tau_{rise,Q}(NOR2) + \tau_{fall,\bar{Q}}(NOR2) \quad \dots\dots(16)$$

The SR Latch Circuit

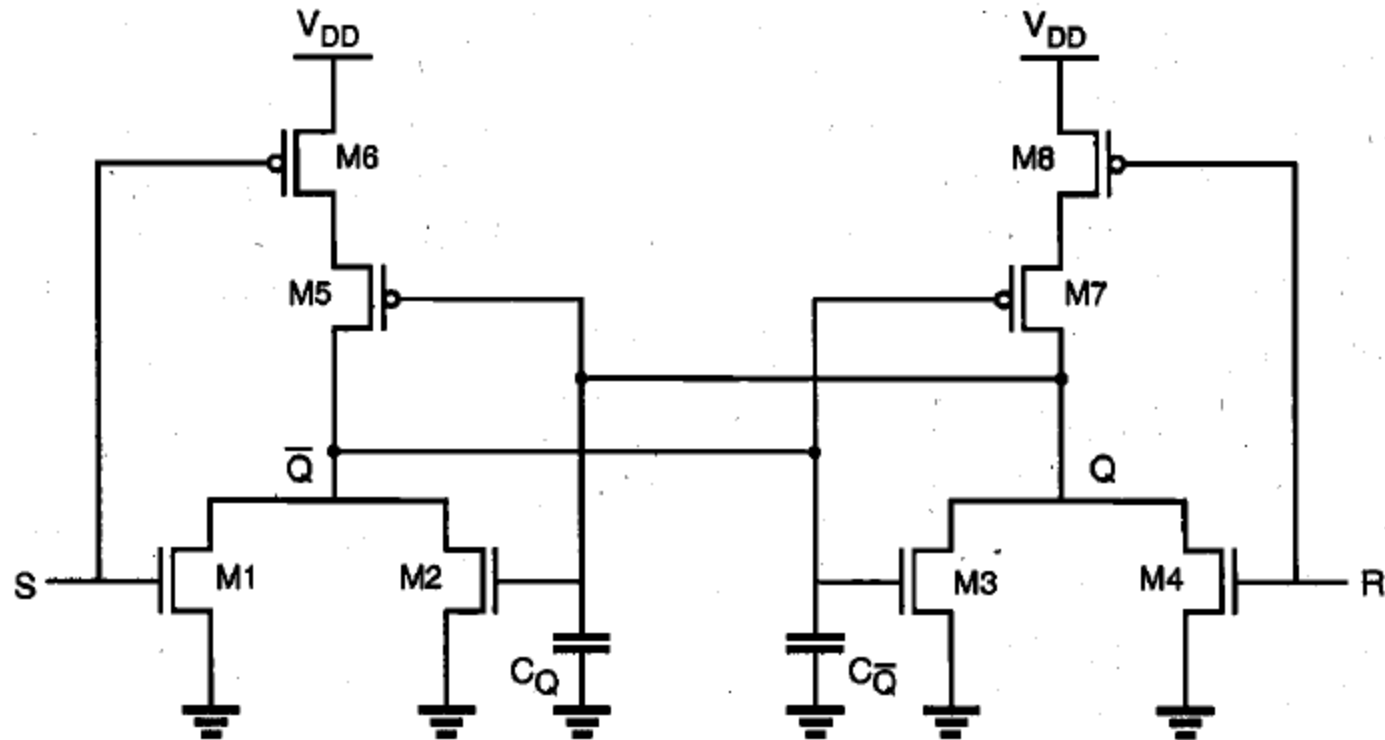


Figure 7. CMOS SR latch circuit based on NOR2 gates.

The SR Latch Circuit

- The calculation of the switching time $\tau_{\text{rise},Q}$ requires two separate calculations for the rise and fall times of the NOR2 gates.
- It is obvious that by considering the two events separately, i.e., first, one of the output node voltages ($\bar{Q}$) falling from high to low due to turn-on of M_1 , followed by the other node voltage (Q) rising from low to high due to turn-off of M_3 , we are bound to overestimate the actual switching time for the SR latch.

The SR Latch Circuit

- Both M_2 and M_4 can be assumed to be off in this process, although M_2 can be turned on as Q rises, thus actually shortening the $\bar{Q}$ node fall time.
- This approach, however, yields a simpler first-order prediction for the time delay, as opposed to the simultaneous solution of two coupled differential equations.

The SR Latch Circuit

- The NOR-based SR latch can also be implemented by using two cross-coupled depletion-load nMOS NOR2 gates, as shown in Fig. 8.
- From the logic point of view, the operation principle of the depletion-load nMOS NOR-based SR latch is identical to that of the CMOS SR latch.

The SR Latch Circuit

- In terms of power dissipation and noise margins, however, the CMOS circuit implementation offers a better alternative, since both of the CMOS NOR2 gates dissipate virtually no static power for preserving a state, and since the output voltages can exhibit a full swing between 0 and V_{DD} .

The SR Latch Circuit

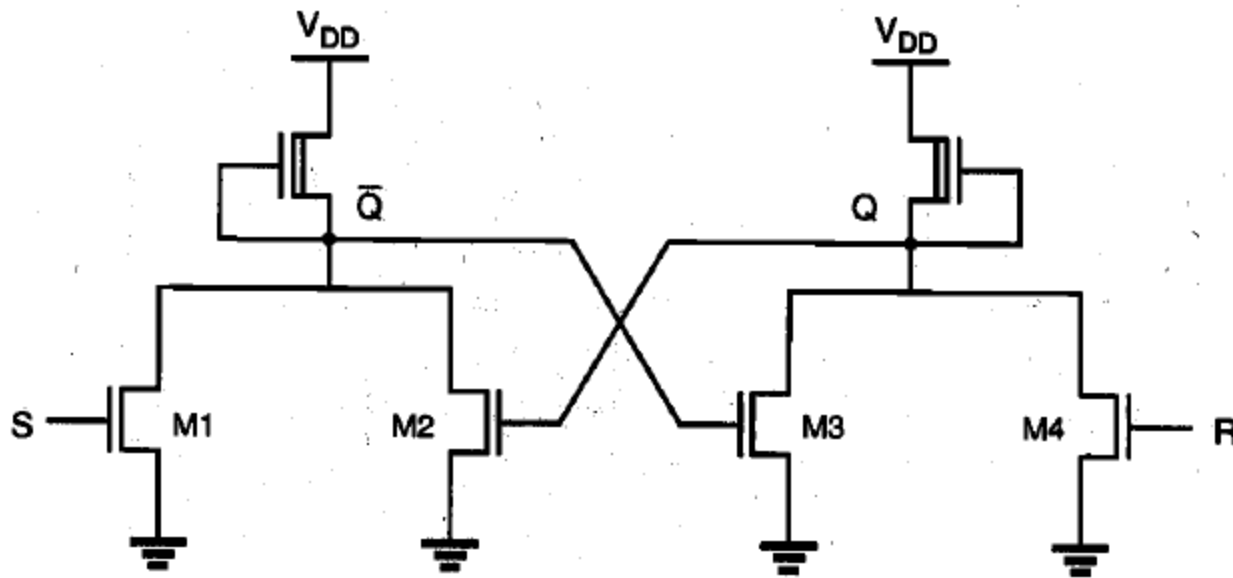


Figure 8. Depletion-load nMOS SR latch circuit based on NOR2 gates.

The SR Latch Circuit

- Instead of using two NOR2 gates, we can use two NAND2 gates, as shown in Fig. 9.
- Here, one input of each NAND gate is used to cross-couple to the output of the other NAND gate, while the second input enables external triggering.

The SR Latch Circuit

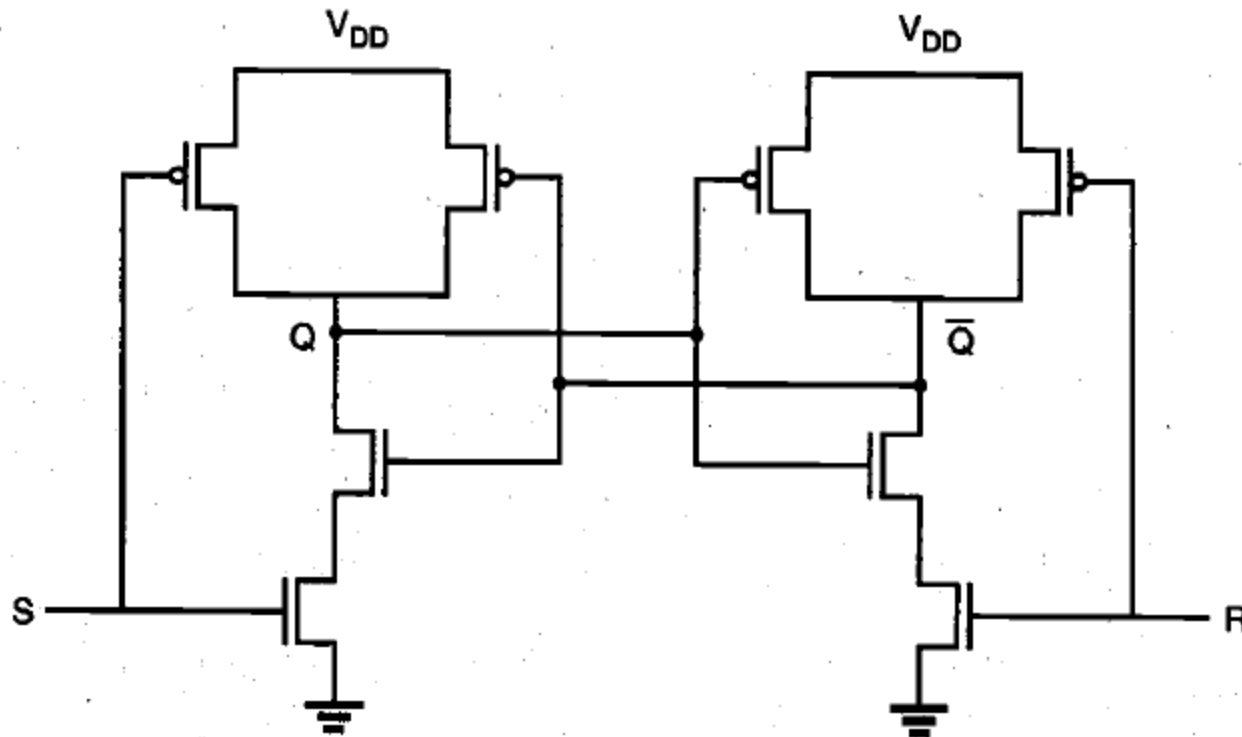


Figure 9. CMOS SR latch circuit based on NAND2 gates.

The SR Latch Circuit

- A close inspection of the NAND-based SR latch circuit reveals that in order to hold (preserve) a state, both of the external trigger inputs must be equal to logic " 1."
- The operating point or the state of the circuit can be changed only by pulling the set input to logic zero or by pulling the reset input to zero.

The SR Latch Circuit

- We can observe that if S is equal to "0" and R is equal to "1," the output Q attains a logic "1" value and the complementary output $\bar{Q}$ becomes logic "0."
- Thus, in order to set the NAND SR latch, a logic "0" must be applied to the set (S) input.
- Similarly, in order to reset the latch, a logic "0" must be applied to the reset (R) input.

The SR Latch Circuit

- The conclusion is that the NAND-based SR latch responds to active low input signals, as opposed to the NOR-based SR latch, which responds to active high inputs.
- Note that if both input signals are equal to logic "0," both output nodes assume a logic-high level, which is not allowed because it violates the complementarity of the two outputs.

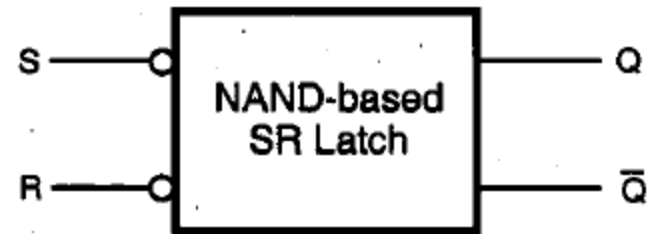
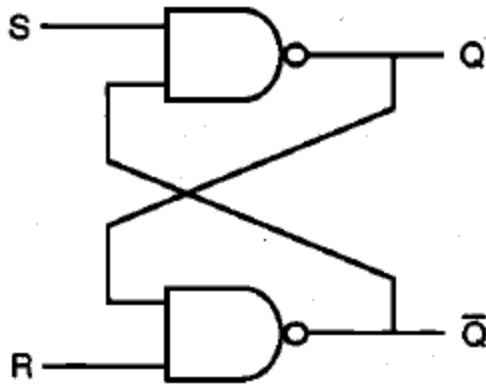
The SR Latch Circuit

- The gate-level schematic and the corresponding block diagram representation of the NAND-based SR latch circuit are shown in Fig. 10.
- The small circles at the S and R input terminals indicate that the circuit responds to active low input signals.

The SR Latch Circuit

- The truth table of the NAND SR latch is also shown in the following.
- The same approach used in the timing analysis of NOR-based SR latches can be applied to NAND-based SR latches.

The SR Latch Circuit



S	R	Q_{n+1}	$\bar{Q}_{n+1}$	Operation
0	0	1	1	not allowed
0	1	1	0	set
1	0	0	1	reset
1	1	Q_n	$\bar{Q}_n$	hold

Figure 10. Gate-level schematic and block diagram of the NAND-based SR latch.

The SR Latch Circuit

- The NAND-based SR latch can also be implemented by using two cross-coupled depletion-load NAND2 gates, as shown in Fig. 11.
- While the operation principle is identical to that of the CMOS NAND SR latch (Fig. 9) from the logic point of view, the CMOS circuit implementation again offers a better alternative in terms of static power dissipation and noise margins.

The SR Latch Circuit

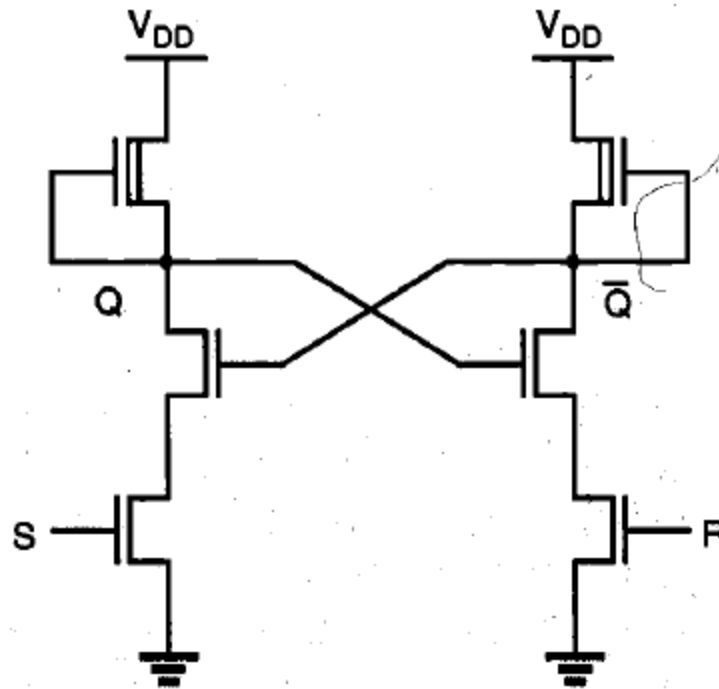


Figure 11. Depletion-load nMOS NAND-based SR latch circuit

Clocked Latch and Flip-Flop Circuits

(Clocked SR Latch)

- All of the SR latch circuits examined in the previous section are asynchronous sequential circuits, which will respond to the changes occurring in input signals at a circuit-delay-dependent time point during their operation.
- To facilitate synchronous operation, the circuit response can be controlled simply by adding a gating clock signal to the circuit, so that the outputs will respond to the input levels only during the active period of a clock pulse.

Clocked Latch and Flip-Flop Circuits

(Clocked SR Latch)

- For simple reference, the clock pulse will be assumed to be a periodic square waveform, which is applied simultaneously to all clocked logic gates in the system.
- The gate-level schematic of a clocked NOR-based SR latch is shown in Fig. 12.
- It can be seen that if the clock (CK) is equal to logic "0," the input signals have no influence upon the circuit response.

Clocked Latch and Flip-Flop Circuits (Clocked SR Latch)

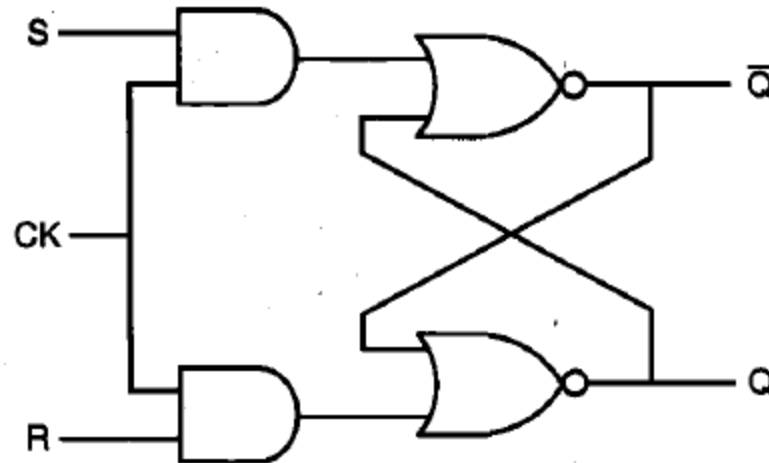


Figure 12. Gate-level schematic of the clocked NOR-based SR latch.

Clocked Latch and Flip-Flop Circuits

(Clocked SR Latch)

- The outputs of the two AND gates will remain at logic "0," which forces the SR latch to hold its current state regardless of the S and R input signals.
- When the clock input goes to logic "1," the logic levels applied to the S and R inputs are permitted to reach the SR latch, and possibly change its state.

Clocked Latch and Flip-Flop Circuits

(Clocked SR Latch)

- In the nonclocked SR latch, the input combination $S = R = "1"$ is not allowed in the clocked SR latch.
- With both inputs S and R at logic " 1," the occurrence of a clock pulse causes both outputs to go momentarily to zero.
- When the clock pulse is removed, i.e., when it becomes "0," the state of the latch is indeterminate.

Clocked Latch and Flip-Flop Circuits

(Clocked SR Latch)

- To illustrate the operation of the clocked SR latch, a sample sequence of CK, S, and R waveforms, and the corresponding output waveform Q are shown in Fig. 13.
- The circuit is strictly level-sensitive during active clock phases, i.e., any changes occurring in the S and R input voltages when the CK level is equal to " 1 " will be reflected onto the circuit outputs.

Clocked Latch and Flip-Flop Circuits (Clocked SR Latch)

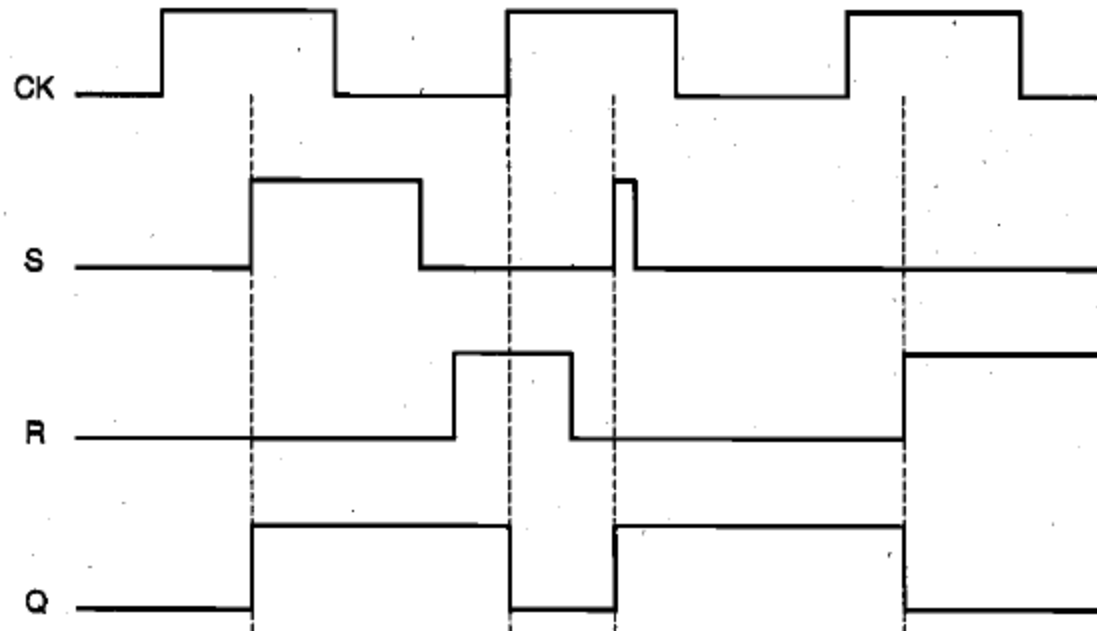


Figure 13. Sample input and output waveforms illustrating the operation of the clocked NOR based SR latch circuit

Clocked Latch and Flip-Flop Circuits

(Clocked SR Latch)

- Consequently, even a narrow spike or glitch occurring during an active clock phase can set or reset the latch, if the loop delay is shorter than the pulse width.

Clocked Latch and Flip-Flop Circuits

(Clocked SR Latch)

- Figure 14 shows a CMOS implementation of the clocked NOR-based SR latch circuit, using two simple AOI gates.
- The AOI-based implementation of the circuit results in a very small transistor count, compared with the alternative circuit realization consisting of two AND2 and two NOR2 gates.

Clocked Latch and Flip-Flop Circuits (Clocked SR Latch)

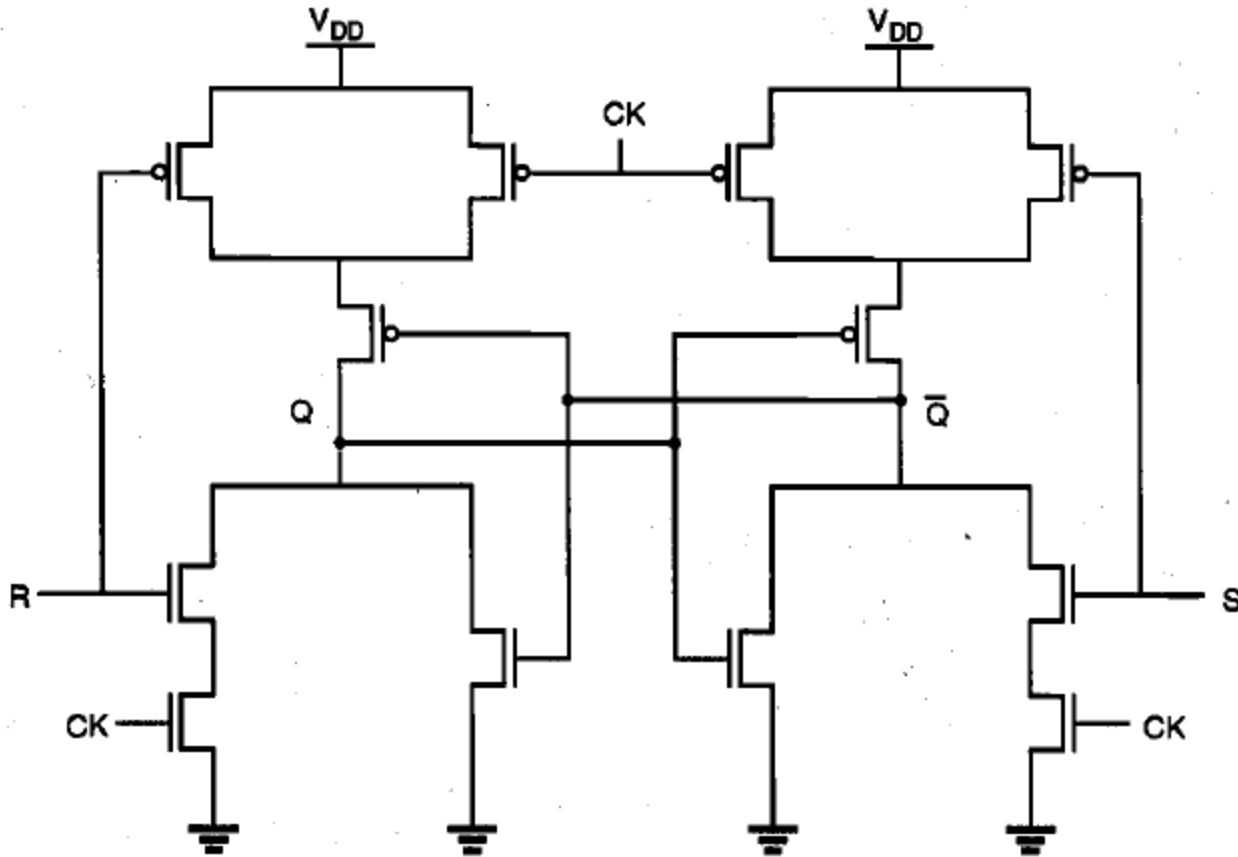


Figure 14. AOI-based implementation of the clocked NOR-based SR latch circuit.

Clocked Latch and Flip-Flop Circuits

(Clocked SR Latch)

- The NAND-based SR latch can also be implemented with gating clock input, as shown in Fig. 15.
- It must be noted, however, that both input signals S and R as well as the clock signal CK are active low in this case.
- This means that changes in the input signal levels will be ignored when the clock is equal to logic "1," and that inputs will influence the outputs only when the clock is active, i.e., CK = "0."

Clocked Latch and Flip-Flop Circuits (Clocked SR Latch)

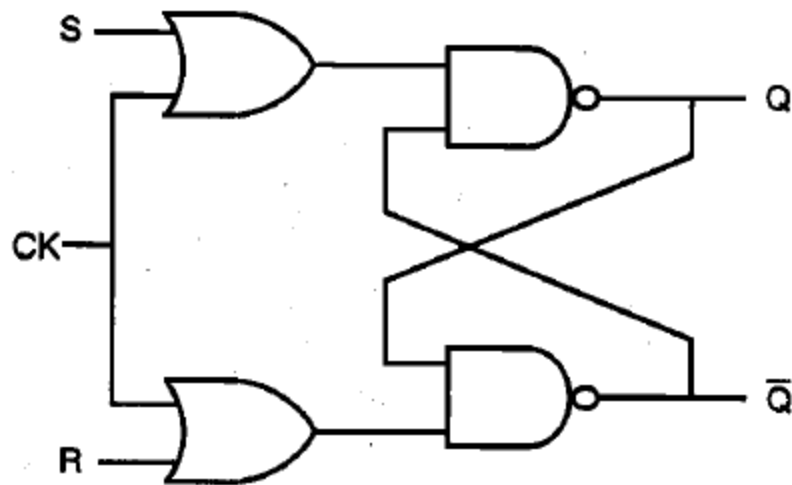


Figure 15. Gate-level schematic of the clocked NAND-based SR latch circuit, with active low inputs.

Clocked Latch and Flip-Flop Circuits

(Clocked SR Latch)

- For the circuit implementation of this clocked NAND-based SR latch, we can use a simple OAI structure, which is analogous to the AOI-based realization of the clocked NOR SR latch circuit.

Clocked Latch and Flip-Flop Circuits

(Clocked SR Latch)

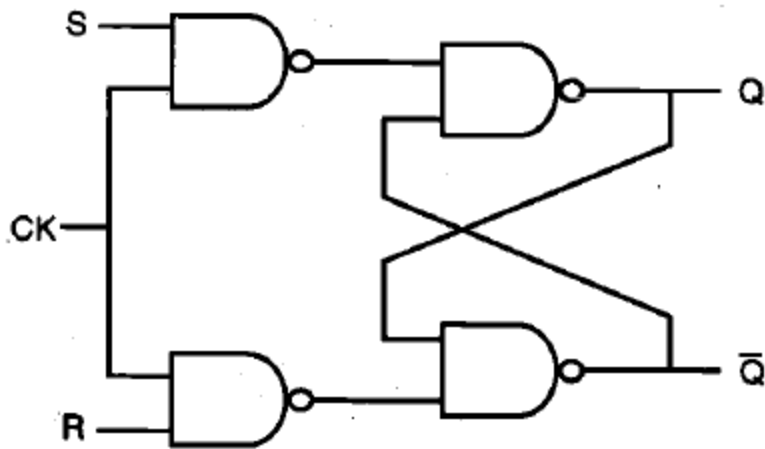
- A different implementation of the clocked NAND-based SR latch is shown in Fig. 16.
- Here, both input signals and the CK signal are active high, i.e., the latch output Q will be set when $CK = 1$, $S = 1$, and $R = 0$.
- Similarly, the latch will be reset when $CK = 1$, $S = 0$, and $R = 1$.
- The latch preserves its state as long as the clock signal is inactive, i.e., when $CK = 0$.

Clocked Latch and Flip-Flop Circuits

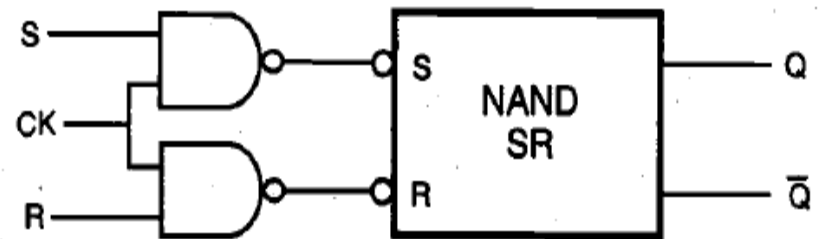
(Clocked SR Latch)

- The drawback of this implementation is that the transistor count is higher than the active low version shown in Fig. 15.

Clocked Latch and Flip-Flop Circuits (Clocked SR Latch)



(a)



(b)

Figure 16. (a) Gate-level schematic of the clocked NAND-based SR latch circuit, with active high inputs. (b) Partial block diagram representation of the same circuit.

Clocked JK Latch

- All simple and clocked SR latch circuits examined to this point suffer from the common problem of having a not-allowed input combination, i.e., their state becomes indeterminate when both inputs S and R are activated at the same time.
- This problem can be overcome by adding two feedback lines from the outputs to the inputs, as shown in Fig. 17.

Clocked JK Latch

- The resulting circuit is called a JK latch.
- Figure 18 shows an all-NAND implementation of the JK latch with active high inputs, and the corresponding block diagram representation.
- The JK latch is commonly called a JK flip-flop.

Clocked JK Latch

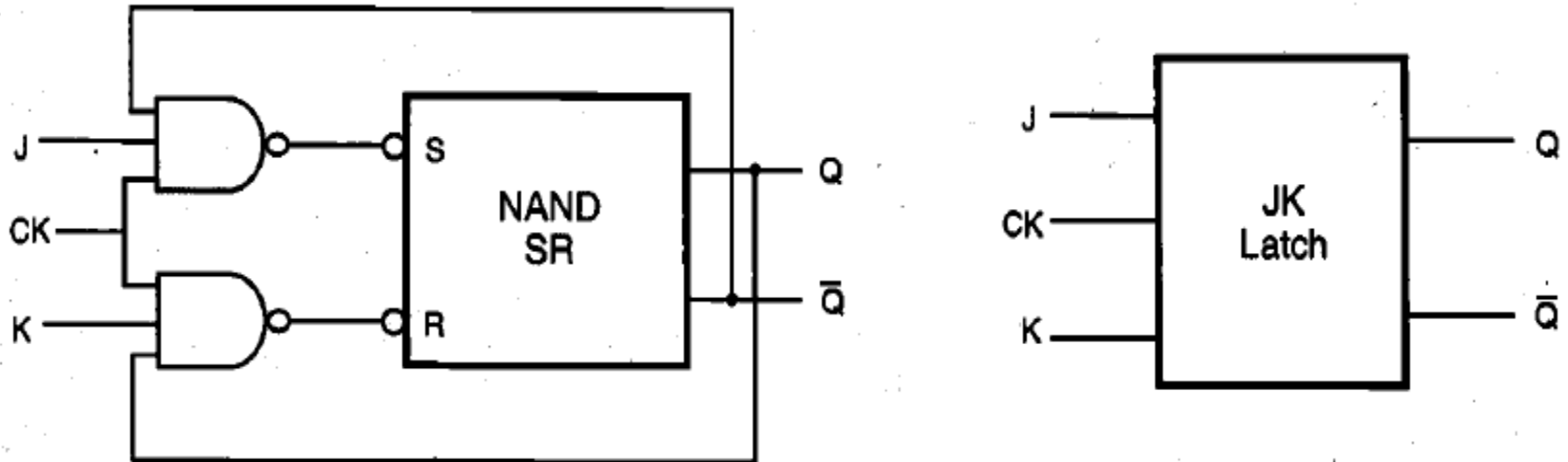


Figure 17. Gate-level schematic of the clocked NAND-based JK latch circuit

Clocked JK Latch

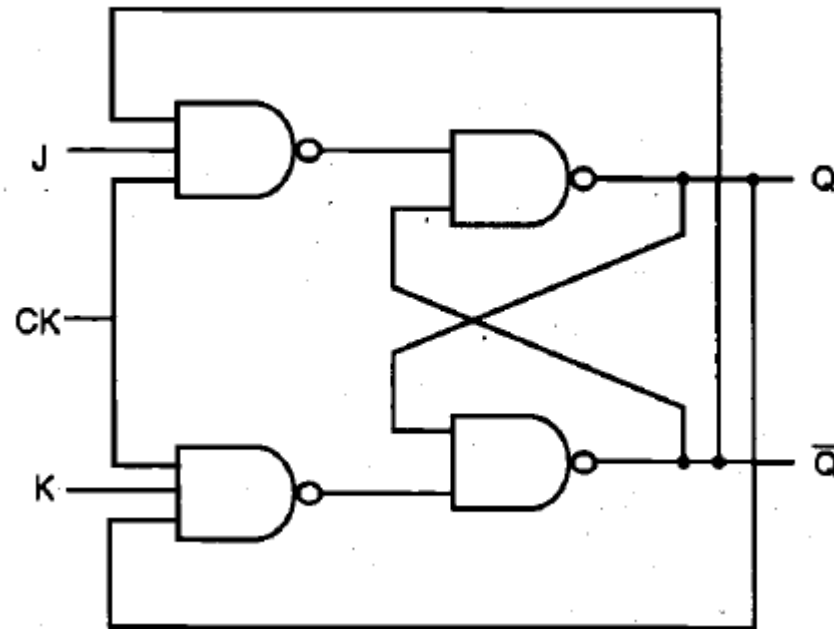


Figure 18. All-NAND implementation of the clocked JK latch circuit.

Clocked JK Latch

- The J and K inputs in this circuit correspond to the set and reset inputs of the basic SR latch.
- When the clock is active, the latch can be set with the input combination ($J = '1', K = '0'$), and it can be reset with the input combination ($J = '0', K = '1'$).
- If both inputs are equal to logic "0," the latch preserves its current state. If, on the other hand, both inputs are equal to " 1 " during the active clock phase, the latch simply switches its state due to feedback.

Clocked JK Latch

- In other words, the JK latch does not have a not-allowed input combination.
- As in the other clocked latch circuits, the JK latch will hold its current state when the clock is inactive ($CK = "0"$).
- The operation of the clocked JK latch is summarized in the truth table as shown below.

Clocked JK Latch

J	K	Q_n	$\overline{Q_n}$	S	R	Q_{n+1}	$\overline{Q_{n+1}}$	Operation
0	0	0	1	1	1	0	1	<i>hold</i>
		1	0	1	1	1	0	
0	1	0	1	1	1	0	1	<i>reset</i>
		1	0	1	0	0	1	
1	0	0	1	0	1	1	0	<i>set</i>
		1	0	1	1	1	0	
1	1	0	1	0	1	1	0	<i>toggle</i>
		1	0	1	0	0	1	

Clocked JK Latch

- Figure 19 shows an alternative, NOR-based implementation of the clocked JK latch, and CMOS realization of this circuit.
- The AOI-based circuit structure results in a relatively low transistor count, and consequently, a more compact circuit compared to the all-NAND realization shown in Fig. 18.

Clocked JK Latch

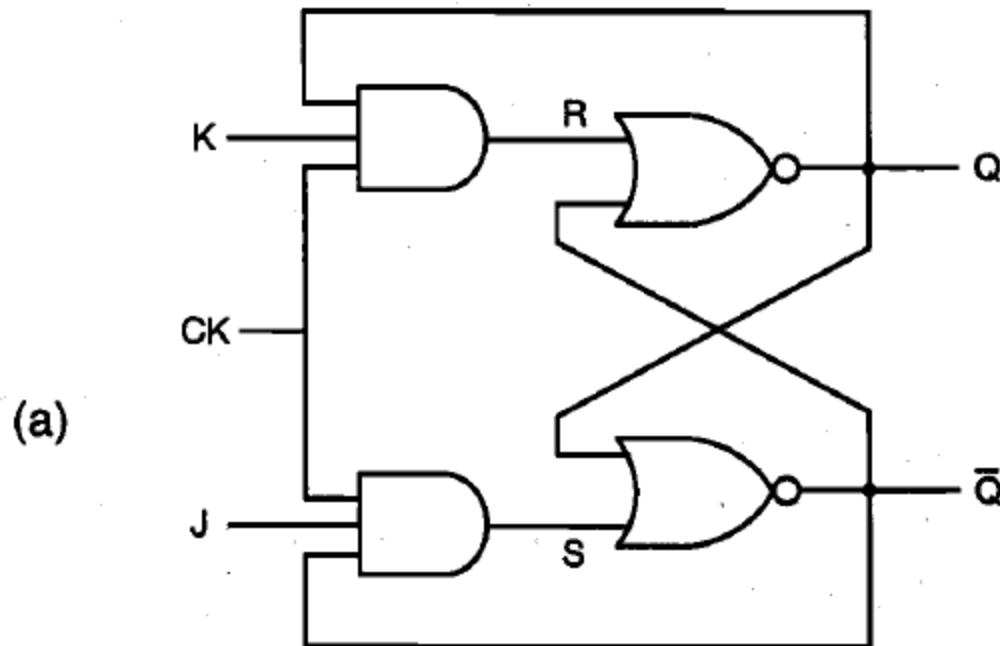


Figure 19 (a) Gate-level schematic of the clocked NOR-based JK latch circuit

Clocked JK Latch

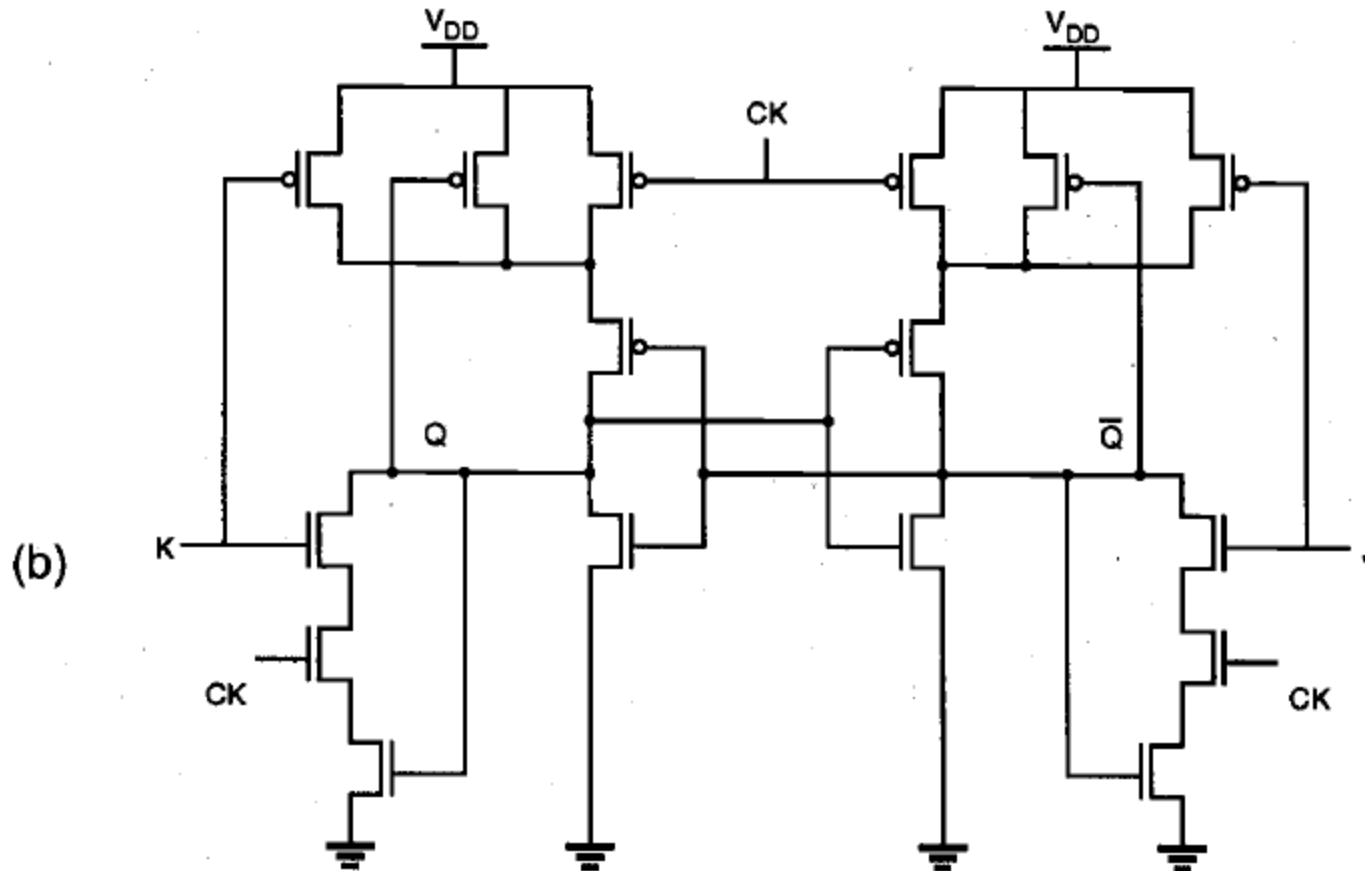


Figure 19 (b) CMOS AOI realization of the JK latch.

Clocked JK Latch

- While there is no not-allowed input combination for the JK latch, there is still a potential problem.
- If both inputs are equal to logic " 1 " during the active phase of the clock pulse, the output of the circuit will oscillate (toggle) continuously until either the clock becomes inactive (goes to zero), or one of the input signals goes to zero.

Clocked JK Latch

- To prevent this undesirable timing problem, the clock pulse width must be made smaller than the input-to-output propagation delay of the JK latch circuit.
- This restriction dictates that the clock signal must go low before prevents uncontrolled oscillation difficult to implement for most practical applications.

Clocked JK Latch

- Assuming that the clock timing constraint described is the output of JK latch will toggle (change its state) only once for each clock pulse, if both inputs are equal to logic " 1 " (Fig. 20).
- A circuit which is operated exclusively in this mode is called a toggle switch.

Clocked JK Latch

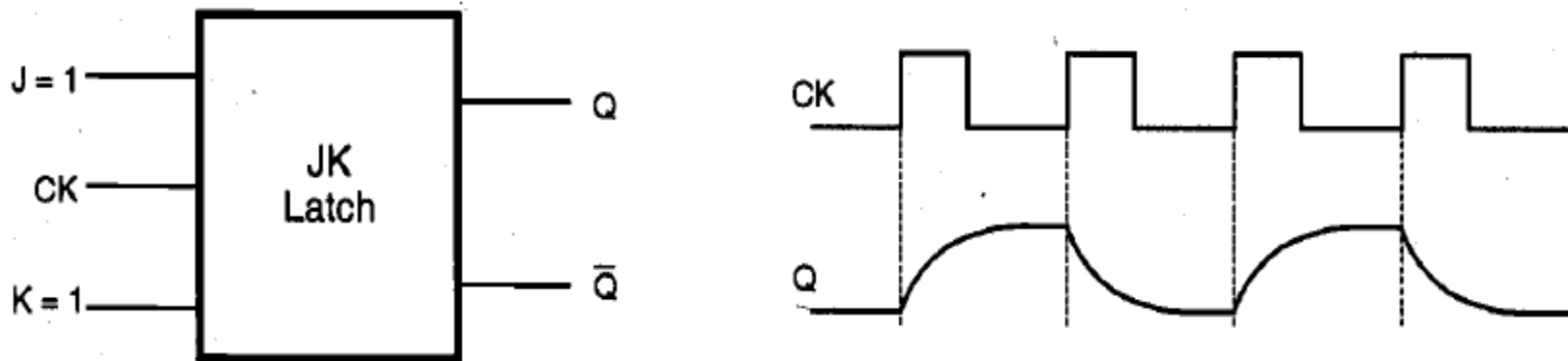


Figure 20. Operation of the JK latch as a toggle switch.

Master-Slave Flip-Flop

- Most of the timing limitations encountered in the previously examined clocked latch circuits can be prevented by using two latch stages in a cascaded configuration.
- The key operation principle is that the two cascaded stages are activated with opposite clock phases.
- This configuration is called the master-slave flip-flop.

Master-Slave Flip-Flop

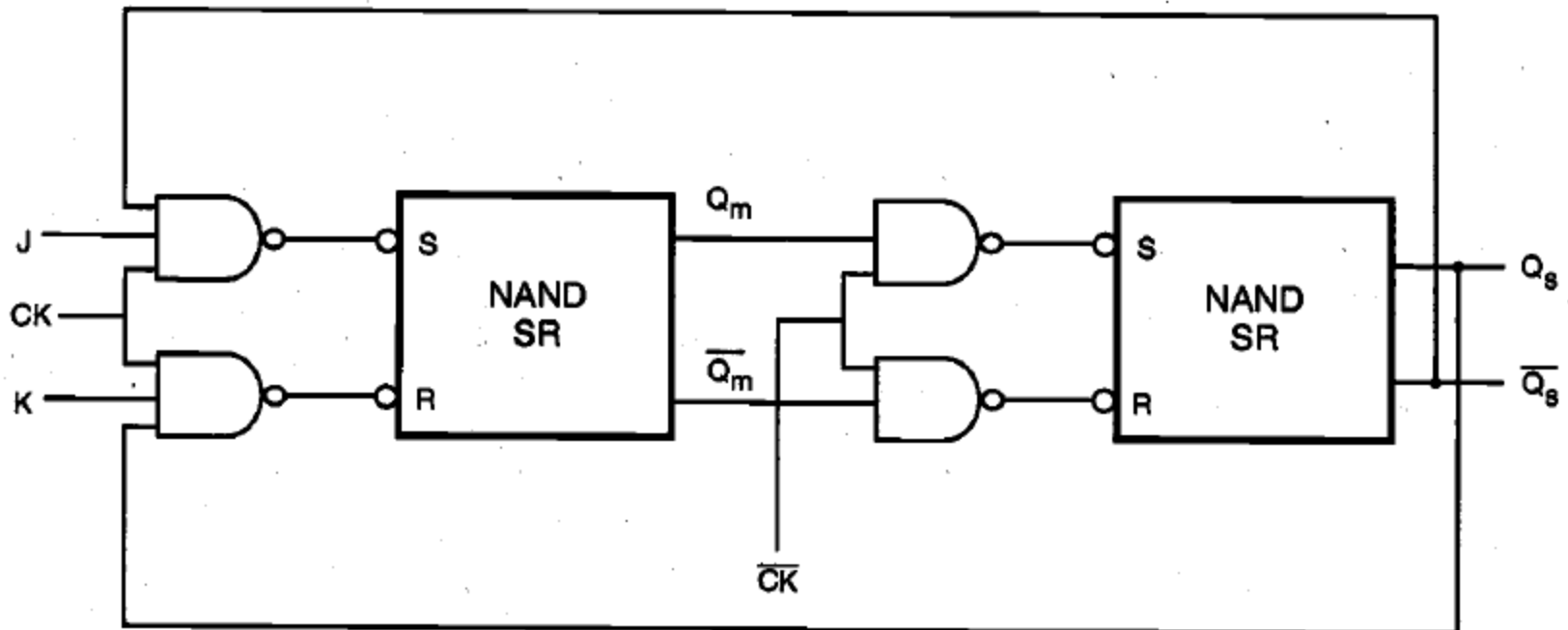


Figure 21. Master-slave flip-flop consisting of NAND-based JK latches.

Master-Slave Flip-Flop

- The input latch in Fig. 21, called the "master," is activated when the clock pulse is high.
- During this phase, the inputs J and K allow data to be entered into the flip-flop, and the first-stage outputs are set according to the primary inputs.
- When the clock pulse goes to zero, the master latch becomes inactive and the second-stage latch, called the "slave," becomes active.

Master-Slave Flip-Flop

- The output levels of the flip-flop circuit are determined during this second phase, based on the master-stage outputs set in the previous phase.

Master-Slave Flip-Flop

- Since the master and the slave stages are effectively decoupled from each other with the opposite clocking scheme, the circuit is never transparent, i.e., a change occurring in the primary inputs is never reflected directly to the outputs.
- This very important property clearly separates the master-slave flip-flop from all of the latch circuits examined earlier.

Master-Slave Flip-Flop

- Figure 22 shows a sample set of input and output waveforms associated with the JK master-slave flip-flop to study the basic operation principles.

Master-Slave Flip-Flop

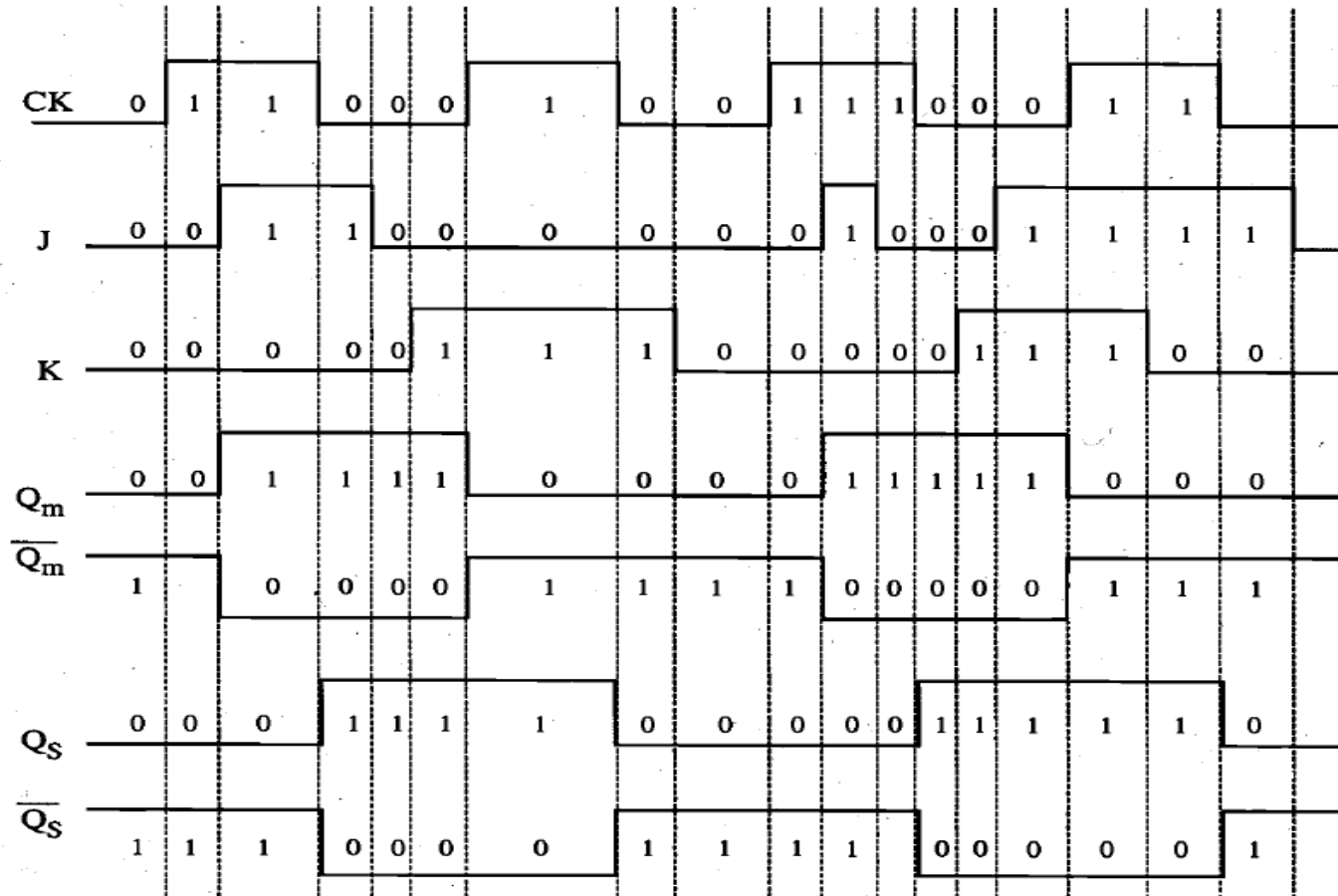


Figure 22. Sample input and output waveforms of the master-slave flip-flop circuit

Master-Slave Flip-Flop

- Because the master and the slave stages are decoupled from each other, the circuit allows for toggling when $J = K = "1,"$ but it eliminates the possibility of uncontrolled oscillations since only one stage is active at any given time.
- A NOR-based alternative realization for the master-slave flip-flop circuit is shown in Fig. 23.

Master-Slave Flip-Flop

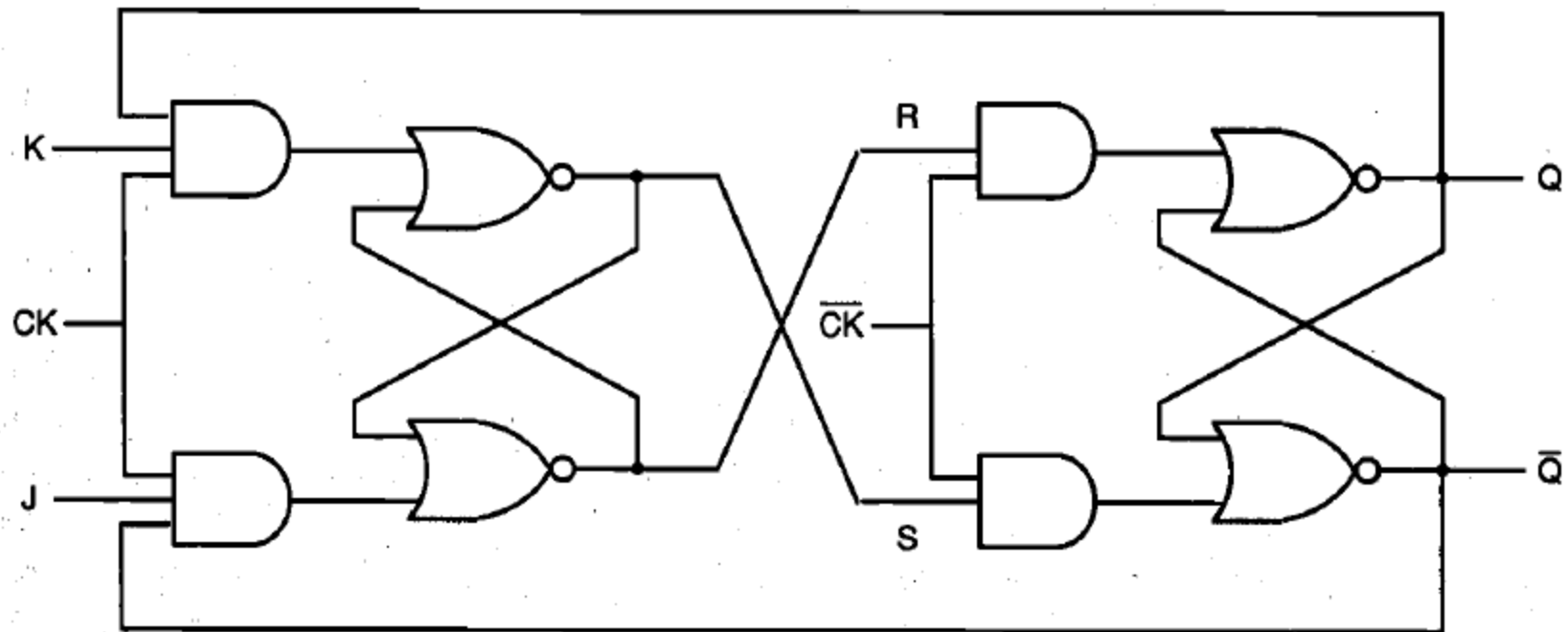


Figure 23. NOR-based realization of the JK master-slave flip-flop.

CMOS D-Latch and Edge-Triggered Flip-Flop

- With the widespread use of CMOS circuit techniques in digital integrated circuit design, a large selection of CMOS-based sequential circuits have also gained popularity and prominence, especially in VLSI design.
- Throughout this chapter, we have seen examples showing that virtually all of the latch and flip-flop circuits can be implemented with CMOS gates, and that their design is quite straightforward.

CMOS D-Latch and Edge-Triggered Flip-Flop

- However, direct CMOS implementations of conventional circuits such as the clocked JK latch or the JK master slave flip-flop tend to require a large number of transistors.

CMOS D-Latch and Edge-Triggered Flip-Flop

- In this section, we will see that specific versions of sequential circuits built primarily with CMOS transmission gates are generally simpler and require fewer transistors than the circuits designed with conventional structuring.
- As an introduction to the issue, let us first consider the simple D-latch circuit shown in Fig. 24.

CMOS D-Latch and Edge-Triggered Flip-Flop

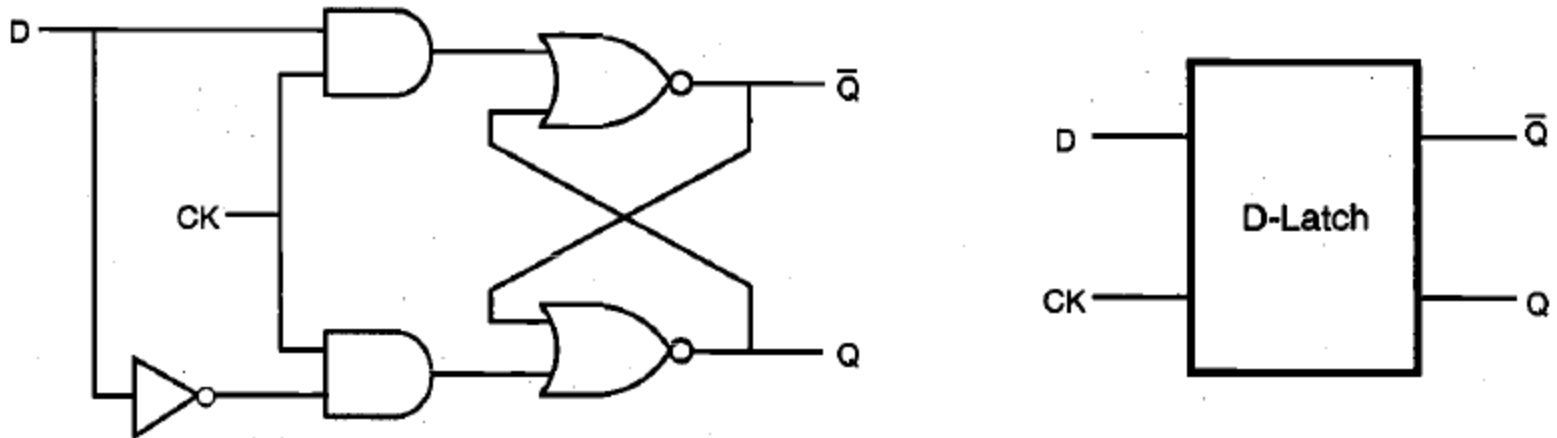


Figure 24. Gate-level schematic and the block diagram view of the D-latch.

CMOS D-Latch and Edge-Triggered Flip-Flop

- The gate-level representation of the D-latch is simply obtained by modifying the clocked NOR-based SR latch circuit.
- Here, the circuit has a single input D, which is directly connected to the S input of the latch.

CMOS D-Latch and Edge-Triggered Flip-Flop

- The input variable D is also inverted and connected to the R input of the latch.
- It can be seen from the gate-level schematic that the output Q assumes the value of the input D when the clock is active, i.e., for $CK = "1."$
- When the clock signal goes to zero, the output will simply preserve its state. Thus, the CK input acts as an enable signal which allows data to be accepted into the D-latch.

CMOS D-Latch and Edge-Triggered Flip-Flop

- The D-latch finds many applications in digital circuit design, primarily for temporary storage of data or as a delay element.
- Consider the circuit diagram given in Fig. 25, which shows a basic two-inverter loop and two CMOS transmission gate (TG) switches.

CMOS D-Latch and Edge-Triggered Flip-Flop

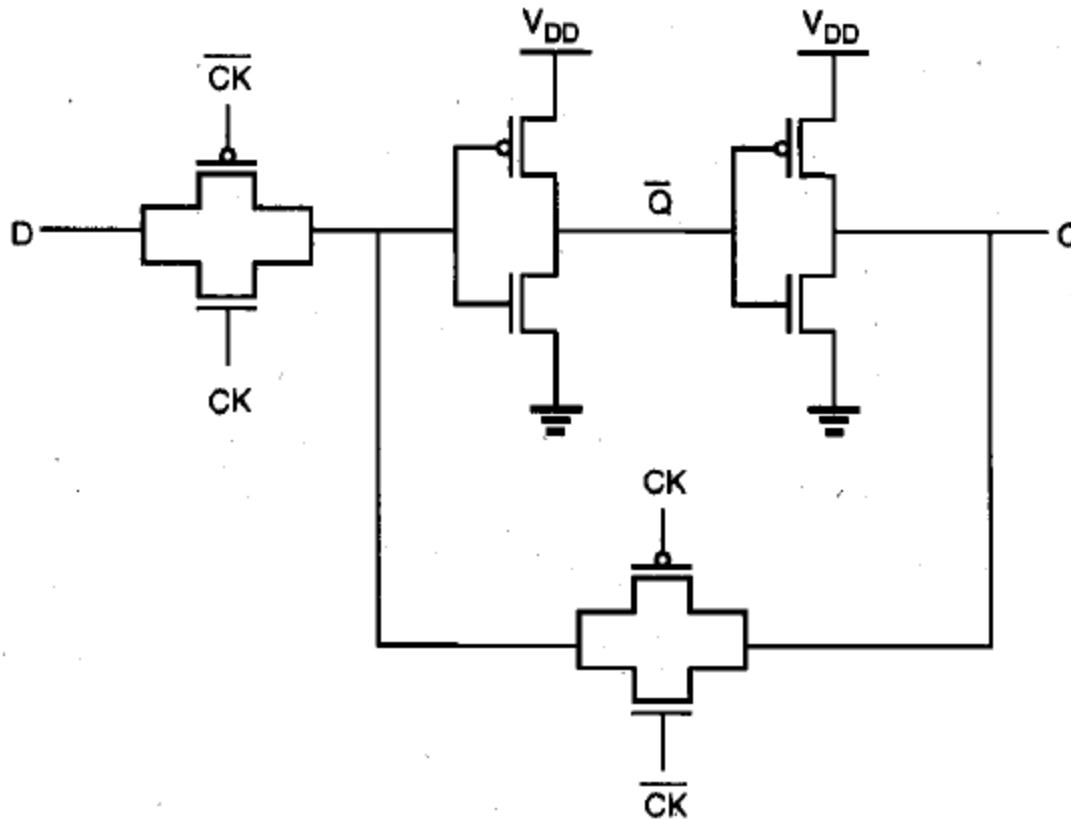


Figure 25. CMOS implementation of the D-latch(version 1)

CMOS D-Latch and Edge-Triggered Flip-Flop

- The TG at the input is activated by the CK signal, whereas the TG in the inverter loop is activated by the inverse of the CK signal, $\overline{\text{CK}}$.
- Thus, the input signal is accepted (latched) into the circuit when the clock is high, and this information is preserved as the state of the inverter loop when the clock is low.

CMOS D-Latch and Edge-Triggered Flip-Flop

- The operation of the CMOS D-latch circuit can be better visualized by replacing the CMOS transmission gates with simple switches, as shown in Fig. 26.
- A timing diagram accompanying this figure shows the time intervals during which the input and the output signals should be valid (unshaded).

CMOS D-Latch and Edge-Triggered Flip-Flop

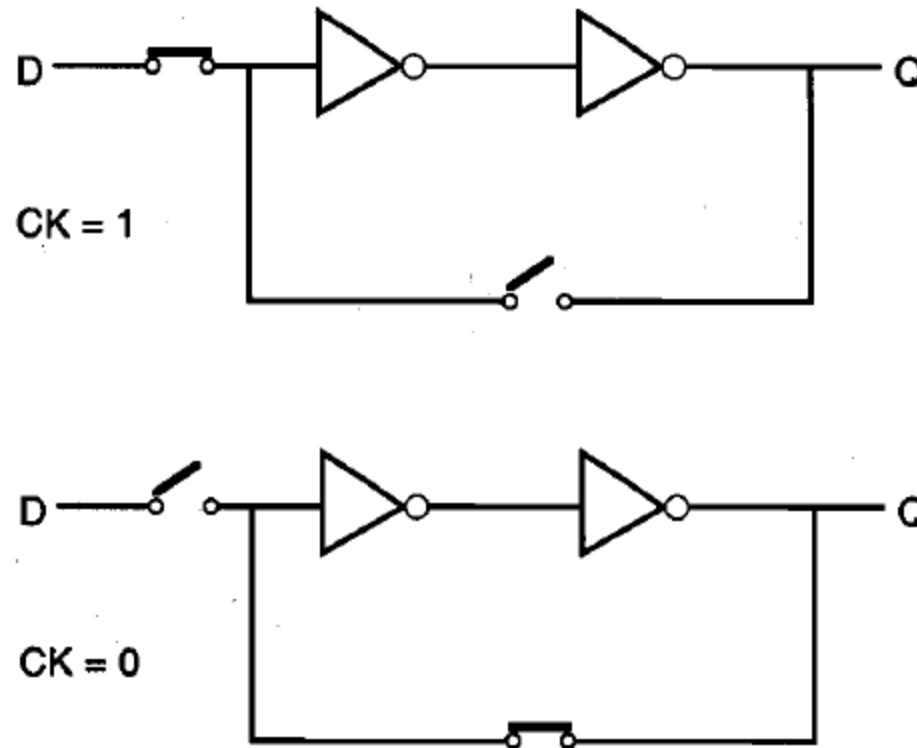


Figure 26. Simplified schematic view and the corresponding timing diagram of the CMOS D latch circuit, showing the setup time and the hold time.

CMOS D-Latch and Edge-Triggered Flip-Flop

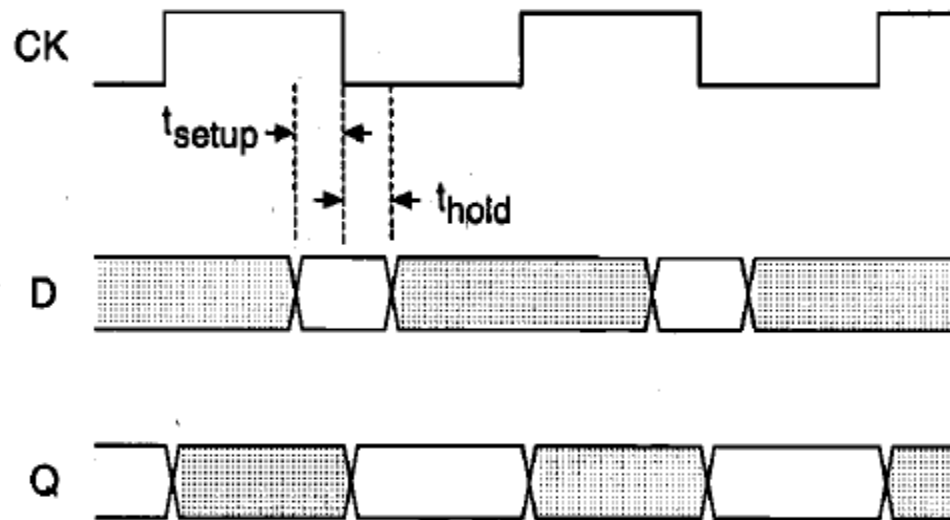


Figure 26. Simplified schematic view and the corresponding timing diagram of the CMOS D latch circuit, showing the setup time and the hold time.

CMOS D-Latch and Edge-Triggered Flip-Flop

- The D-latch shown in Fig. 25 is not an edge-triggered storage element because the output changes according to the input, i.e., the latch is transparent, while the clock is high.
- The transparency property makes the application of this D-latch unsuitable for counters and some data storage implementations.

CMOS D-Latch and Edge-Triggered Flip-Flop

- Figure 27 shows a different version of the CMOS D-latch.
- The circuit contains two tristate inverters, driven by the clock signal and its inverse.
- Although the circuit appears to be quite different from that shown in Fig. 25, the basic operation principle of the circuit is the same as that shown in Fig. 26.
- The first tri-state inverter acts as the input switch, accepting the input signal when the clock is high.

CMOS D-Latch and Edge-Triggered Flip-Flop

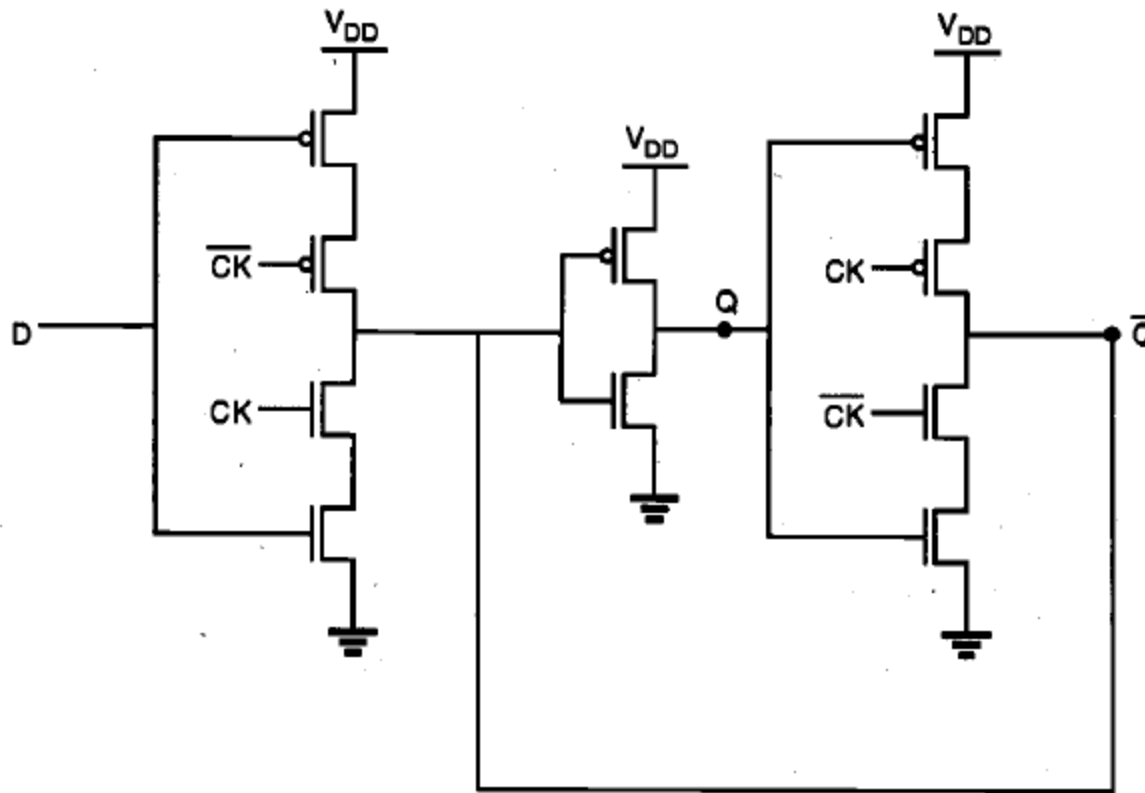


Figure 27. CMOS implementation of the D-latch (version 2)

CMOS D-Latch and Edge-Triggered Flip-Flop

- At this time, the second tristate inverter is at its high-impedance state, and the output Q is following the input signal.
- When the clock goes low, the input buffer becomes inactive, and the second tristate inverter completes the two-inverter loop, which preserves its state until the next clock pulse.

CMOS D-Latch and Edge-Triggered Flip-Flop

- Finally, consider the two-stage master-slave flip-flop circuit shown in Fig. 28, which is constructed by simply cascading two, D-latch circuits.
- The first stage (master) is driven by the clock signal, while the second stage (slave) is driven by the inverted clock signal.
- Thus, the master stage is positive level sensitive, while the slave stage is negative level-sensitive.

CMOS D-Latch and Edge-Triggered Flip-Flop

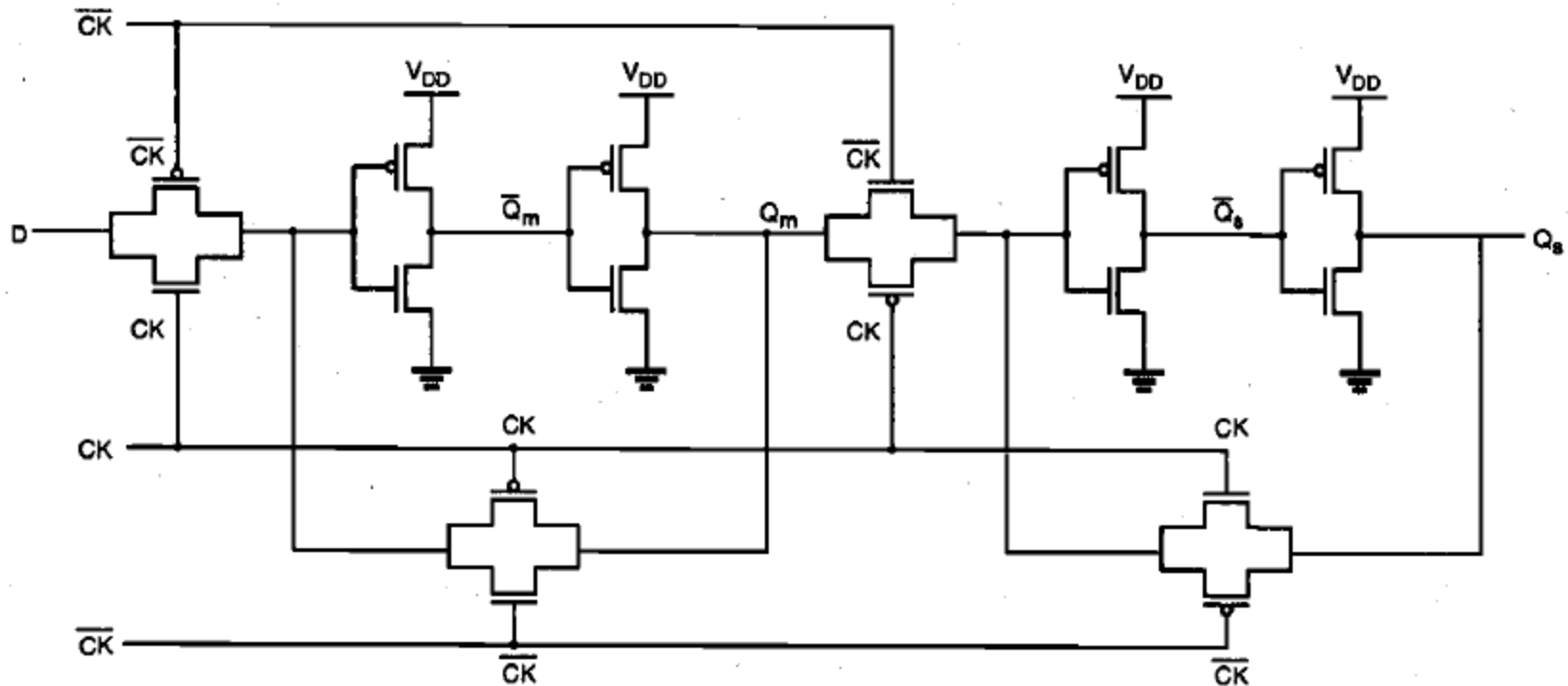


Figure 28. CMOS negative (falling) edge-triggered master-slave D flip-flop (DFF).

CMOS D-Latch and Edge-Triggered Flip-Flop

- When the clock is, high, the master stage follows the D input while the slave stage holds the previous value.
- When the clock changes from logic "1" to logic "0" the master latch ceases to sample the input and stores the D value at the time of the clock transition.
- At the same time, the slave latch becomes transparent, passing the stored master value Q_m to the output of the slave stage, Q_s .

CMOS D-Latch and Edge-Triggered Flip-Flop

- The input cannot affect the output because the master stage is disconnected from the D input.
- When the clock changes again from logic 0" to 1," the slave latch locks in the master latch output and the master stage starts sampling the input again.
- Thus, this circuit is a negative edge-triggered D flip-flop by virtue of the fact that it samples the input at the falling edge of the clock pulse.

CMOS D-Latch and Edge-Triggered Flip-Flop

- Another implementation of edge-triggered D flip-flop is shown in Fig. 29, which consists of six NAND3 gates.
- This D flip-flop is positive edge-triggered as illustrated in the waveform chart in Fig. 30.
- Initially, all the signal values except for S are 0, i.e., $(S, R, CK, D) = (1, 0, 0, 0)$, and $Q = 0$.
- In the second phase, both D and R switch to 1, i.e., $(S, R, CK, D) = (1, 0, 1, 1)$, but no change in Q occurs and the Q value remains at 0.

CMOS D-Latch and Edge-Triggered Flip-Flop

- However, in the third phase, if CK goes to high, i.e., $(S, R, CK, D) = (1, 1, 1, 1)$, the output of gate 2 switches to 0, which in turn sets the output of the last stage SR latch to 1.
- Thus, the output of this D flip-flop switches to 1 at the positive-going edge of the clock signal, CK.
- However, as can be observed in the ninth phase of the waveform diagram chart, the Q output is not affected by the negative-going edge of CK, nor by other signal changes.

CMOS D-Latch and Edge-Triggered Flip-Flop

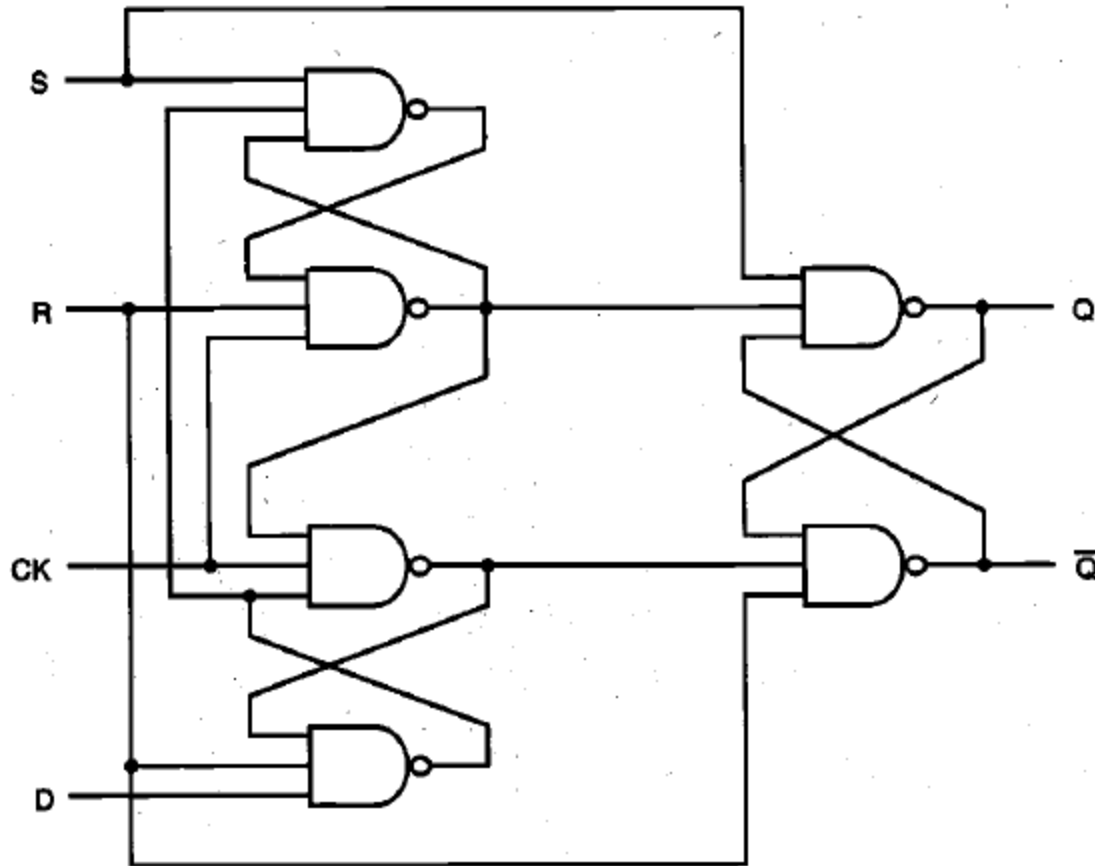


Figure 29. NAND3-based positive edge-triggered D flip-flop circuit

CMOS D-Latch and Edge-Triggered Flip-Flop

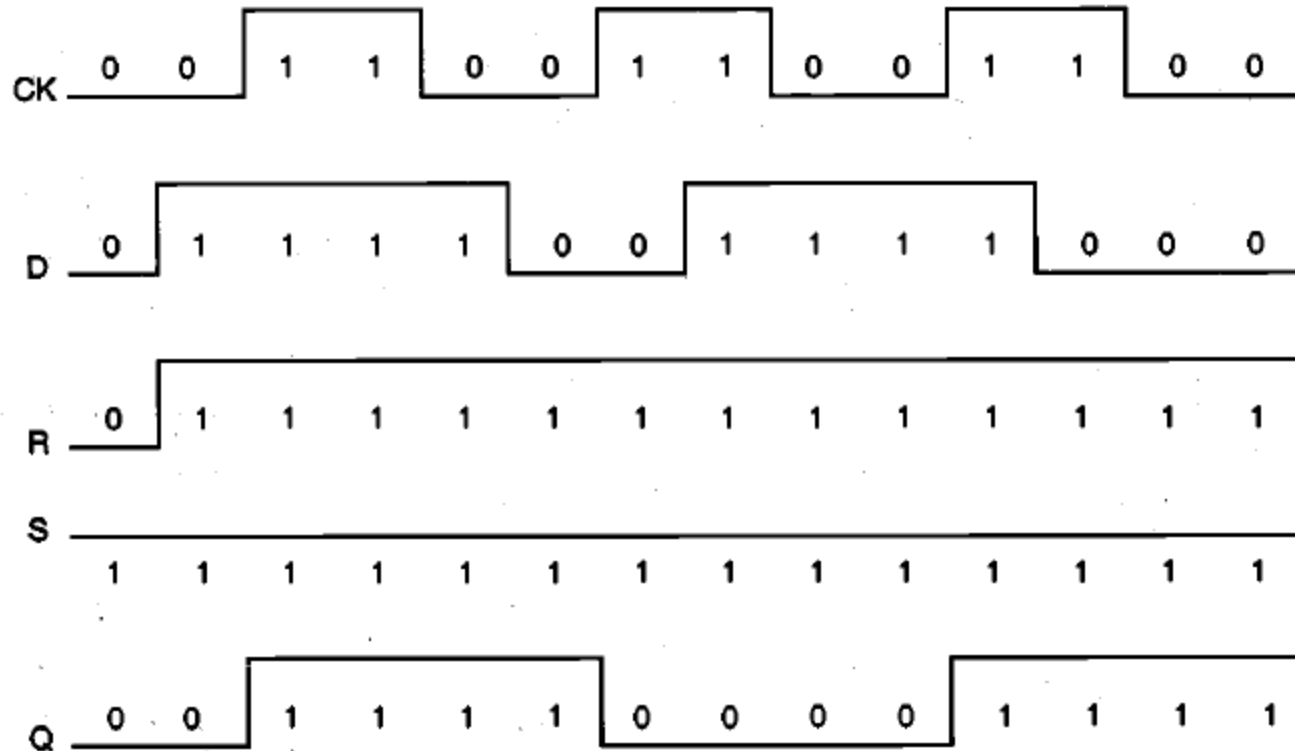


Figure 30. Timing diagram of the positive edge-triggered D flip-flop.