In summary, this chapter provides a reasonably comprehensive and in-depth of CMOS digital integrated-circuit design, perhaps the most significant area (at least in terms of production volume and societal impact) of electronic circuits.

1. DIGITAL CIRCUIT DESIGN: AN OVERVIEW
We discuss the various technologies and logic-circuit families currently in use, consider the parameters employed to characterize the operation and performance of logic circuits, and finally mention the various styles for digital-system design.

1.1. Digital IC Technologies and Logic-Circuit Families
The chart in Figure 1 shows the major IC technologies and logic-circuit families that are currently in use. Members of each family are made with the same technology, have a similar circuit structure, and exhibit the same basic features. Each logic-circuit family offers a unique set of advantages and disadvantages. In the conventional style of designing systems, one selects an appropriate logic family (e.g., TTL, CMOS, or ECL) and attempts to implement as much of the system as possible using circuit modules (packages) that belong to this family. In this way, interconnection of the various packages is relatively straightforward.

The selection of a logic family is based on such considerations as logic flexibility, speed of operation, availability of complex functions, noise immunity, operating-temperature range, power dissipation, and cost.

CMOS technology is, by a large margin, the most dominant of all the IC technologies available for digital-circuit design. As mentioned earlier, CMOS has replaced NMOS, which was employed in the early days of VLSI (in the 1970s). The most important of which is the much lower power dissipation of CMOS circuits. CMOS has also replaced bipolar as the technology-of-choice in digital-system design, and has made possible levels of integration (or circuit-packing densities), and a range of applications, neither of which would have been possible with bi-polar technology.
1. CMOS logic circuits dissipate much less power than bipolar logic circuits and thus one can pack more CMOS circuits on a chip than is possible with bipolar circuits.

2. The high input impedance of the MOS transistor allows the designer to use charge storage as a means for the temporary storage of information in both logic and memory circuits. This technique cannot be used in bipolar circuits.

3. The feature size (i.e., minimum channel length) of the MOS transistor has decreased dramatically over the years, with some recently reported designs utilizing channel lengths as short as 0.06 µm. This permits very tight circuit packing and, correspondingly, very high levels of integration.

**Bipolar** Two logic-circuit families based on the bipolar junction transistor are in some use at present: TTL and ECL, Transistor-transistor logic (TTL or T2L) was for many years the most widely used logic-circuit family. Its decline was precipitated by the advent of the VLSI era. TTL manufacturers, however, fought back with the introduction of low-power and high-speed versions. In these newer versions, the higher speeds of operation are made possible by preventing the BIT from saturating and thus avoiding the slow turnoff process of a saturated transistor. These nonsaturating versions of TTL utilize the Schottky diode and are called Schottky TTL. Despite all these efforts, TTL is no longer a significant logic-circuit family.

The other bipolar logic-circuit family in present use is emitter-coupled logic (ECL). It is based on the current-switch implementation of the inverter, The basic element of ECL is the differential BIT pair, and, correspondingly, also called current-mode logic (CML), in which saturation is avoided, very high speeds of operation are possible. Indeed, of all the commercially available logic-circuit families, ECL is the fastest. ECL is also used in VLSI circuit design when very high operating speeds are required and the designer is willing to accept higher power dissipation and increased silicon area.

**BiCMOS** BiCMOS combines the high operating speeds possible with BJTs (because of their inherently higher transconductance) with the low power dissipation and other excellent characteristics of CMOS. Like CMOS, BiCMOS allows for the implementation of both analog and digital circuits on the same chip. At present, BiCMOS is used to great advantage in special applications, including memory chips, where it requires a high-speed and lower power dissipation.

**Gallium Arsenide (GaAs)** The high carrier mobility in GaAs results in very high speeds of operation. This has been demonstrated in a number of digital IC chips utilizing GaAs technology. It should be pointed out, however, that GaAs remains an "emerging technology," one that appears to have great potential but has not yet achieved such potential commercially.

1.2. **Logic-Circuit Characterization**

The following parameters are usually used to characterize the operation and performance of a logic-circuit family.
Noise Margins  The static operation of a logic-circuit family is characterized by the voltage transfer characteristic (VTC) of its basic inverter. Figure 2 shows such a VTC and defines its four parameters: $V_{OH}$, $V_{OL}$, $V_{IH}$, and $V_{IL}$. Note that $V_{IH}$ and $V_{IL}$ are defined as the points at which the slope of the VTC is -1. Also indicated is the definition of the threshold voltage $V_M$ or $V_{th}$ as we shall frequently call it, as the point at which $V_O = V_i$.

The robustness of a logic-circuit family is determined by its ability to reject noise, and thus by the noise margins $N_{M_H}$ and $N_{M_L}$,

$$N_{M_H} = V_{OH} - V_{IH}$$  \hspace{1cm} (1)

$$N_{M_L} = V_{IL} - V_{OL}$$  \hspace{1cm} (2)

![Figure 2](image_url)

Figure 2  Typical voltage transfer characteristic (VTC) of a logic inverter, illustrating the definition of the critical points.

An ideal inverter is one for which $N_{M_H}$ = $N_{M_L}$ = $V_{DD}/2$, where $V_{DD}$ is the power-supply voltage. Further, for an ideal inverter, the threshold voltage $V_M$ = $V_{DD}/2$.

Propagation Delay  The dynamic performance of a logic-circuit family is characterized by the propagation delay of its basic inverter. Figure 3 illustrates the definition of the low-to-high propagation delay ($t_{PLH}$) and the high-to-low propagation delay ($t_{PHL}$). The inverter propagation delay ($t_p$) is defined as the average of these two quantities:

$$t_p = \frac{1}{2} (t_{PLH} + t_{PHL})$$  \hspace{1cm} (3)
Power Dissipation The need to minimize the gate power dissipation is motivated by the desire to pack an ever increasing number of gates on a chip, which in turn is motivated by space and economic considerations. Modern digital systems utilize large numbers of gates and memory cells, and thus to keep the total power requirement within reasonable bounds, the power dissipation per gate and per memory cell should be kept as low as possible.

There are two types of power dissipation in a logic gate: static and dynamic. Static power refers to the power that the gate dissipates in the absence of switching action. It results from the presence of a path in the gate circuit between the power supply and ground in one or both of its two states. Dynamic power, on the other hand, occurs only when the gate is switched: An inverter operated from a power supply \( V_{DD} \) and driving a load capacitance \( C \), dissipates dynamic power \( P_D \),

\[
P_D = f CV_{DD}^2
\]  

(A)

Where \( f \) is the frequency at which the inverter is being switched.

Delay-Power Product One is usually interested in high-speed performance (low \( t_p \)) combined with low power dissipation. Unfortunately, these two requirements are often in conflict; when designing a gate, if one attempts to reduce power dissipation by decreasing the supply voltage, or the supply current, or both, the current-driving capability of the gate decreases. This in turn results in longer times to charge and discharge the load and parasitic capacitances, and thus the propagation delay increases.

\[
DP = P_D t_p
\]  

(B)

where \( P_D \) is the power dissipation of the gate. Note that \( DP \) has the units of joules. The lower the \( DP \) figure for a logic family, the more effective it is.
**Silicon Area**  
Objective in the design of digital VLSI circuits is the minimization of silicon area per logic gate. Smaller area requirement enables the fabrication of a larger number of gates per chip, which has economic and space advantages from a system design standpoint. Area reduction occurs in three different ways: through advances in processing technology, through advances in circuit-design techniques, and through careful chip layout. In this book, our interest lies in circuit design. As a general rule, the simpler the circuit, the smaller the area required.

The circuit designer has to decide on device sizes. Choosing smaller devices has the obvious advantage of requiring smaller silicon area and at the same time reducing parasitic capacitances and thus increasing speed. Smaller devices, however, have lower current-driving capability, which tends to increase delay. Thus, as in all engineering design problems, there is a trade-off between speed and area.

**Fan-In and Fan-Out**  
The fan-in of a gate is the number of its inputs. Thus, a four-input NOR gate has a fan-in of 4. Fan-out is the maximum number of similar gates that a gate can drive while remaining within guaranteed specifications.

1.3. Styles for Digital System Design
The conventional approach to designing digital systems consists of assembling the system using standard IC packages of various levels of complexity (and hence integration). The advent of VLSI, in addition to providing the system designer with more powerful off-the-shelf components such as microprocessors and memory chips, has made possible alternative design styles. One such alternative is to opt for implementing part or all of the system using one or more custom VLSI chips.

**IC Technology** is the manner in which a digital (gate-level) implementation is mapped on to an IC. IC’s consist of numerous layers (perhaps 10 or more)

The three main IC technologies are:
1) Full-custom/VLSI:
   - All layers are optimized for an embedded system’s particular digital implementation
     - Placing transistors
     - Sizing transistors
     - Routing wires
   - Benefits
     - Excellent performance, small size, low power
   - Drawbacks
     - Long time-to-market
2) Semi-custom ASIC (gate array and standard cell)
   • Lower layers are fully or partially built
     ➢ Designers are left with routing of wires and may be placing some blocks
   • Benefits
     ➢ Good performance, good size.
   • Drawbacks
     ➢ Still require weeks to months to develop
3) PLD (Programmable Logic Device)
   • All layers are already exist
     ➢ Designers can purchase an IC
     ➢ Connection on the IC are either created or destroyed to implement desired functionality. Field-Programmable Gate Array (FPGA) very popular
   • Benefits
     ➢ Almost instant IC availability
   • Drawbacks
     ➢ Bigger, expensive, power hungry, slower

1.4 Design Abstraction and computer Aids
2. DESIGN AND PERFORMANCE ANALYSIS OF THE CMOS INVERTER

2.1 Circuit Structure

The inverter circuit, shown in Fig. 4(a), consists of a pair of complementary MOSFETs switched by the input voltage \( V_I \). Although not shown, the source of each device is connected to its body, thus eliminating the body effect. Usually, the threshold voltages \( V_{t_n} \) and \( V_{t_p} \) are equal in magnitude; that is, \( V_{t_n} = |V_{t_p}| = V_t \) is in the range of 0.2 V to 1 V. The inverter circuit can be represented by a pair of switches operated in a complementary fashion, as shown in Fig. 4(b). As indicated, each switch is modeled by a finite on resistance, which is the source-drain resistance of the respective Transistor, evaluated near \( |V_{DS}| = 0 \).

\[
\begin{align*}
r_{DSN} &= \frac{1}{k' \rho (W/L) \rho (V_{DD} - V_{t})} \\
r_{DSP} &= \frac{1}{k' \rho (W/L) \rho (V_{DD} - V_{t})}
\end{align*}
\] (4) (5)

Figure 4 (a) The CMOS inverter and (b) its representation as a pair of switches operated in a complementary fashion.

2.2 Static Operation

Case 1: With \( V_I = 0 \), \( V_O = V_{OH} = V_{DD} \), and the output node is connected to \( V_{DD} \) through the resistance \( r_{DSP} \) of the pull-up transistor \( Q_P \).

Case 2: with \( V_I = V_{DD} \), \( V_O = V_{OL} = 0 \), and the output node is connected to ground through the resistance \( r_{DSN} \) of the pull-down transistor \( Q_N \).

Thus, in the steady state, no direct-current path exists between \( V_{DD} \) and ground, and the static-current and the static-power dissipation are both zero (leakage effects are usually negligibly small particularly for large-feature-size devices).

The voltage transfer characteristic of the inverter is shown in Fig. 5, from which it is confirmed that the output voltage levels are 0 and \( V_{DD} \), and thus the output voltage swing is the maximum possible. The fact that \( V_{OH} \) and \( V_{OL} \) are independent of device dimensions makes CMOS very different from other forms of MOS logic.
The CMOS inverter can be made to switch at the midpoint of the logic swing, 0 to $V_{DD}$, that is, at $V_{DD}/2$, by appropriately \textit{sizing the transistors}. Specifically, it can be shown that the switching threshold $V_{th}$ (or $V_M$) is given by

$$V_{th} = \frac{V_{DD} - |V_{tp}| + V_{tn}/k_n/k_p}{1 + \sqrt{k_n/k_p}}$$  \hspace{1cm} (6)

where $k_n = k_n'(W/L)_n$ and $k_p = k_p'(W/L)_p$, from which we see that for the typical case where $V_{tn} = |V_{tp}|$, $V_{th} = V_{DD}/2$ for $k_n = k_p$, that is,

$$k_n'(W/L)_n = k_p'(W/L)_p$$  \hspace{1cm} (7)

Thus a symmetrical transfer characteristic is obtained when the devices are designed to have equal transconductance parameters, a condition we refer to as \textit{matching}. Since $\mu_n$ is two to four times larger than $\mu_p$, matching is achieved by making $(W/L)_p$ two to four times (i.e., $\mu_n/\mu_p$ times) $(W/L)_n$,

$$(W/L)_p = \frac{\mu_n}{\mu_p} (W/L)_n$$  \hspace{1cm} (8)

Figure 5 The voltage transfer characteristic (VTC) of the CMOS inverter when $Q_N$ and $Q_P$ are matched. The two devices have the same channel length, $L$, which is set at the minimum allowable for the given process technology. The minimum width of the NMOS transistor is usually one and a half to two times $L$, and the width of the PMOS transistor two to three times that. Since the inverter area can be represented by $W_nL_n + W_pL_p = (W_n + W_p)L_n$, the area of the minimum-size inverter is $(n + p)L_n^2$, and we can use the factor $(n + p)$ as a proxy for area.

Besides placing the gate threshold at the center of the logic swing, matching the transconductance parameters of $Q_N$ and $Q_P$ provides the inverter with equal current-driving capability in both directions (pull-up and pull-down).
Furthermore, and obviously related, it makes $r_{DSN} = r_{DSP}$. Thus an inverter with matched transistors will have equal propagation delays, $t_{PLH}$ and $t_{PLH}$. When the inverter threshold is at $V_{DD}/2$, the noise margins $NM_H$ and $NM_L$ are equalized, and their values are maximized, such that

$$NM_H = NM_L = 3/8(V_{DD} + 2/3 V_I) \tag{9}$$

A final comment on the inverter VTC, we note that the slope in the transition region, though large, is finite and is given by $-(gm_N + gm_P)(r_{ON}/r_{OP})$.

### 2.3 Dynamic Operation

The propagation delay of the inverter is usually determined under the condition that it is driving an identical inverter. This situation is depicted in Fig. 6. The propagation delay of the inverter comprising $Q_1$ and $Q_2$, which is driven by a low-impedance source $V_I$, and is loaded by the inverter comprising $Q_3$ and $Q_4$.

Indicated in the figure are the various transistor internal capacitances that are connected to the output node of the $(Q_1, Q_2)$ inverter. Replace all the capacitances attached to the inverter output node with a single capacitance $C$ connected between the output node and ground. If we are able to do that, we can utilize the results of the transient analysis performed in Section 4.10. during $t_{PLH}$ or $t_{PLH}$, the output of the first inverter changes from 0 to $V_{DD}/2$ or from $V_{DD}$ to $V_{DD}/2$ respectively. It follows that the second inverter remains in the same state during each of our analysis intervals.

![Figure 6 Circuit for analyzing the propagation delay of the inverter formed by $Q_1$ and $Q_2$, which is driving an identical inverter formed by $Q_3$ and $Q_4$.](image)

1. The gate-drain overlap capacitance of $Q_1$, $C_{gd1}$, can be replaced by an equivalent capacitance between the output node and ground of $2C_{gd1}$. The factor 2 arises because of the Miller effect (Section 6.4.4). Specifically, note that as $V_I$ goes high and $V_O$ goes low by the same amount, the change in voltage across $C_{gd1}$ is twice that amount.
Thus the output node sees in effect twice the value of $C_{gd1}$. The same applies for the gate drain overlap capacitance of $Q_2$, $C_{gd2}$, which can be replaced by a capacitance 2 $C_{gd2}$ between the output node and ground.

2. Each of the drain-body capacitances $C_{db1}$ and $C_{db2}$ has a terminal at a constant voltage. Thus for the purpose of our analysis here, $C_{db1}$ and $C_{db2}$ can be replaced with equal capacitances between the output node and ground.

3. Since the second inverter does not switch states, we will assume that the input capacitances of $Q_3$ and $Q_4$ remain approximately constant and equal to the total gate capacitance $(WLC_{OX} + C_{gsov} + C_{gdo})$. The input capacitance of the load inverter will be

$$C_{g3} + C_{g4} = (WL)C_{OX} + (WL)C_{gsov} + C_{gdo}$$

4. The last component of $C$ is the wiring capacitance $C_W$, which simply adds to the value of $C$. Thus, the total value of $C$ is given by

$$C = 2C_{gd1} + 2C_{gd2} + C_{db1} + C_{db2} + C_{g3} + C_{g4} + C_W$$

Consider the circuit in Fig. 7(a), which applies when $V_I$ goes high and $Q_N$ discharges $C$ from its initial voltage of $V_{DD}$ to the final value of 0. $Q_N$ will be in the saturation mode and then, when $V_O$ falls below $V_{DD} - V_t$, it will go into the triode region of operation.

The approximate expression for $t_{PHL}$ will be

$$t_{PHL} = 1.6C/(k'_n(W/L)nV_{DD})$$

Computing an average value for the discharge current $i_{DN}$ during the interval $t=0$ to $t=t_{PHL}$, $Q_N$ will be saturated, and $i_{DN}(0)$ is given by

$$i_{DN}(0) = \frac{1}{2}k'_n(W/L)n(V_{DD} - V_t)^2$$

At $t=t_{PHL}$, $Q_N$ will be in the triode region, and $i_{DN}(t_{PHL})$ will be

$$i_{DN}(t_{PHL}) = k'_n(W/L)n[(V_{DD} - V_t)V_{DD}/2 - 1/2(V_{DD}/2)^2]$$

The average discharge current can then be found as

$$i_{DN,av} = \frac{1}{2} [i_{DN}(0) + i_{DN}(t_{PHL})]$$

and the discharge interval $t_{PHL}$ computed from

$$t_{PHL} = C\Delta V/i_{DN,av}$$

$$t_{PHL} = (CV_{DD}/2)/i_{DN,av}$$

Utilizing Eqs. (12) through (15) and substituting $V_t \cong 0.2V_{DD}$ gives

$$t_{PHL} \cong 1.7C/(k'_n(W/L)nV_{DD})$$
An expression for the low-to–high inverter delay, $t_{PLH}$, can be written by analogy to the $t_{PHL}$ expression in equation (16),

$$t_{PLH} \approx 1.7C/(k'_{p}(W/L)_{p}V_{DD}) \quad (17)$$

Finally, the propagation delay $t_{p}$ can be found as the average of $t_{PHL}$ and $t_{PLH}$,

$$t_{p} = \frac{1}{2} (t_{PHL} + t_{PLH})$$

Figure 7: Equivalent circuits for determining the propagation delays (a) $t_{PHL}$ and (b) $t_{PLH}$ of the inverter.

Examination of the formulas in Eqs. (16) and (17) enables us to make a number of useful observations:

1. As expected, the two components of $t_{p}$ can be equalized by selecting the (W/L) ratios to equalize $k_{n}$ and $k_{p}$, that is, by matching $Q_{N}$ and $Q_{P}$.

2. Since $t_{p}$ is proportional to $C$, the designer should strive to reduce $C$. This is achieved by using the minimum possible channel length and by minimizing wiring and other parasitic capacitances.

3. Using a process technology with larger transconductance parameter $k'$ can result in shorter propagation delays.
4. Using larger (W/L) ratios can result in a reduction in $t_p$. However, should be exercised here also, since increasing the size of the devices increases the value of $C$, and thus the expected reduction in $t_p$ might not materialize.

5. A larger supply voltage $V_{DD}$ results in a lower $t_p$. However, $V_{DD}$ is determined by the process technology and thus is often not under the control of the designer. These observations clearly illustrate the conflicting requirements and the trade-offs available in the design of a CMOS digital integrated circuit (and indeed in any engineering design problem).

### 2.4 Dynamic Power Dissipation

The dynamic power dissipated in the CMOS inverter is given by

$$P_D = fCV^2DD$$

(18)

Where $f$ is the frequency at which the gate is switched. It follows that minimizing $C$ is an effective means for reducing dynamic-power dissipation. An even more effective strategy is the use of a lower power-supply voltage.

### 3. CMOS Logic-Gate Circuits

#### 3.1 Basic Structure

A CMOS logic circuit is in effect an extension, or a generalization, of the CMOS inverter:

The CMOS logic gate consists of two networks: the pull-down network (PDN) constructed of NMOS transistors, and the pull-up network (PUN) constructed of PMOS transistors (see Fig. 8). The two networks are operated by the input variables, in a complementary fashion. Thus, for the three-input gate represented in Fig. 8, the PDN will conduct for all input combinations that require a low-output ($Y=0$) and will then pull the output node down to ground, causing a zero voltage to appear at the output, $V_Y=0$. Simultaneously, the PUN will be off, and no direct dc path will exist between $V_{DD}$ and ground. On the other hand, all input combinations that call for a high output ($Y=1$) will cause the PUN to conduct, and the PUN will then pull the output node up to $V_{DD}$, establishing an output voltage $V_Y = V_{DD}$. Simultaneously, the PDN will be cut off, and again, no dc current path between $V_{DD}$ and ground will exist in the circuit.
Since the PDN comprises NMOS transistors, and since an NMOS transistor conducts when the signal at its gate is high, the PDN is activated (i.e., conducts) when the inputs are high. In a dual manner, the PUN comprises PMOS transistors, and a PMOS transistor conducts when the input signal at its gate is low; thus the PUN is activated when the inputs are low.

The PDN and the PUN each utilizes devices in parallel to form an OR function, and devices in series to form an AND function. For the circuit in Fig. 9(a), we observe that $Q_A$ will conduct when $A$ is high ($V_A = V_{DD}$) and will then pull the output node down to ground ($V_Y = 0V, Y = 0$). Similarly, $Q_B$ conducts and pulls $Y$ down when $B$ is high. Thus $Y$ will be low when $A$ is high or $B$ is high, which can be expressed as

$$\bar{Y} = A + B$$

or equivalently

$$Y = \overline{A + B}$$

Figure 8 Representation of a three-input CMOS logic gate. The PUN comprises PMOS transistors, and the PDN comprises NMOS transistors.

Figure 9 Examples of pull-down networks.
The PDN in Fig. 9(b) will conduct only when \( A \) and \( B \) are both high simultaneously.
Thus \( Y \) will be low when \( A \) is high and \( B \) is high,

\[
\bar{Y} = AB
\]

or equivalently

\[
Y = AB
\]

As a final example, the PDN in Fig. 9(c) will conduct and cause \( Y \) to be 0 when \( A \) is high or when \( B \) and \( C \) are both high, thus

\[
\bar{Y} = A + BC
\]

or equivalently

\[
Y = A + BC
\]

Figure 10 Examples of pull-up networks.

Next consider the PUN examples shown in Fig. 10. The PUN in Fig. 10(a) will conduct and pull \( Y \) up to \( V_{DD} (Y=1) \) when \( A \) is low or \( B \) is low, thus

\[
Y = \overline{A + \overline{B}}
\]

The PUN in Fig. 10(b) will conduct and produce a high output (\( V_Y = V_{DD}, Y=1 \)) only when \( A \) and \( B \) are both low, thus

\[
Y = \overline{AB}
\]

Finally, the PUN in Fig. 10(c) will conduct and cause \( Y \) to be high (logic 1) if \( A \) is low or if \( B \) and \( C \) are both low, thus

\[
Y = \overline{A} + \overline{BC}
\]
Figure 11 Usual and alternative circuit symbols for MOSFETs

Fig 11 shows our usual symbols (left) and the corresponding "digital" symbols (right). Observe that the symbol for the PMOS transistor with a circle at the gate terminal is intended to indicate that the gate terminal of the PMOS transistor is an active low input.

Remember that for an NMOS transistor, the drain is the terminal that is at the higher voltage (current flows from drain to source), and for a PMOS transistor the source is the terminal that is at the higher voltage (current flows from source to drain).

3.2 The Two-Input NOR Gate

Consider the CMOS gate that realizes the two-input NOR function

\[ y = \overline{A + B} = \overline{A \overline{B}} \]  \hspace{1cm} (19)

\( Y \) is to be low (PDN conducting) when \( A \) is high or \( B \) is high. Thus the PDN consists of two parallel NMOS devices with \( A \) and \( B \) as inputs (i.e., the circuit in Fig.9(a)). For the PUN, we note from the second expression in Eq. (19) that \( Y \) is to be high when \( A \) and \( B \) are both low. Thus the PUN consists of two series PMOS devices with \( A \) and \( B \) as the inputs (i.e., the circuit in Fig. 10(b)). Putting the PDN and the PUN together gives the CMOS NOR gate shown in Fig. 12.
3.3 The Two-Input NAND Gate

The two-input NAND function is described by the Boolean expression

\[ Y = \overline{A \cdot B} = \overline{A} + \overline{B} \]  

(20)

To synthesize the PDN, we consider the input combinations that require \( Y \) to be low: There is only one such combination, namely, \( A \) and \( B \) both high. Thus, the PDN simply comprises two NMOS transistors in series (such as the circuit in Fig. 9(b)). To synthesize the PUN, we consider the input combinations that result in \( Y \) being high. These are found from the second expression in Eq. (20) as \( A \) low or \( B \) low. Thus, the PUN consists of two parallel PMOS transistors with \( A \) and \( B \) applied to their gates (such as the circuit in Fig. 10(a)). Putting the PDN and PUN together results in the CMOS NAND gate implementation shown in Fig. 13.
3.4 A Complex Gates
These are AND-OR-INVERT(AOI) and OR-AND-INVERT(OAI) gates. Both the complex gates have a propagation delay equivalent to that of a single NAND or NOR gate.

Let us implement the function

\[ F = \overline{AB + CD} \]

Here, \( AB \) and \( CD \) are two AND functions and their sum is the OR function, which is finally Inverted. Thus \( F \) can be implemented as an AOI gate. Figure 14(c) shows the CMOS realization of an AOI gate. The truth table for this gate and its logic Equivalent circuit for the AOI gate are shown in figs.14(b) and 14(a), respectively.

The CMOS realization of the OR-AND-INVERT(OAI) gate is the dual of that for the AND-OR-INVERT(AOI) gate and is easily obtained by flipping that latter end-for-end while interchanging all NMOS circuits with PMOS circuits and vice versa, as shown in fig. 15(c). The truth table for this gate and its logic Equivalent circuit for the OAI gate are shown in figs.15(b) and 15(a), respectively.
The output expression F is given as

\[ F = (A+B)(C+D) \]

Consider next the more complex logic function

\[ y = \overline{A(B + CD)} \]  

(21)

Since \( \overline{Y} = A(B + CD) \), we see that \( Y \) should be low for \( A \) high and simultaneously either \( B \) high or \( C \) and \( D \) both high, from which the PDN is directly obtained. To obtain the PUN, we need to express \( Y \) in terms of the complemented variables. We do this through repeated application of DeMorgan's law, as follows:

\[ Y = \overline{A(B + CD)} \]
\[ = A + B + CD \]
\[ = \bar{A} + \overline{B} \quad \overline{CD} \]
\[ = \bar{A} + \bar{B} \quad (\bar{C} + \bar{D}) \quad (22) \]

Thus, \( Y \) is high for \( A \) low or \( B \) low and either \( C \) or \( D \) low. The corresponding complete CMOS circuit will be as shown in Fig. 16.

![Figure 16 CMOS realization of a complex gate.](image)

### 3.5 Obtaining the PUN from the PDN and Vice Versa

From the CMOS gate circuits considered thus far (e.g., that in Fig. 16), we observe that the PDN and the PUN are dual networks: Where a series branch exists in one, a parallel branch exists in the other. Thus, we can obtain one from the other, a process that can be simpler than having to synthesize each separately from the Boolean expression of the function.

### 3.6 The Exclusive-OR Function

An important function that often arises in logic design is the exclusive-OR (XOR) function,

\[ y = A \bar{B} + \bar{A} \; B \quad (23) \]

We observe that since \( \bar{Y} \) (rather than \( Y \)) is given, it is easier to synthesize the PUN. We note, however, that unfortunately \( Y \) is not a function of the complemented variables only (as we would like it to be). Thus, we will need additional inverters. The PUN obtained directly from Eq. (23) is shown in Fig. 17(a). Note that the \( Q_1, Q_2 \) branch
realizes the first term \((A \, B)\), whereas the \(Q_3, \, Q_4\) branch realizes the second term \((A \, B)\).

As for synthesizing the PDN, we can obtain it as the dual network of the PUN in Fig. 17(a). Alternatively, we can develop an expression for \(Y\) and use it to synthesize the PDN. DeMorgan’s law can be applied to the expression in Eq. (23) to obtain \(\overline{Y}\) as

\[
\overline{Y} = AB + \overline{A} \overline{B}
\]

The corresponding PDN will be as in Fig. 17(b), which shows the CMOS realization of the exclusive-OR function except for the two additional inverters.

![Figure 17 Realization of the exclusive-OR (XOR) function:](image)

3.7 Summary of the Synthesis Method
1. The PDN can be most directly synthesized by expressing \(Y\) as a function of the uncomplemented variables. If complemented variables appear in this expression, additional inverters will be required to generate them.
2. The PUN can be most directly synthesized by expressing \(y\) as a function of the complemented variables and then applying the uncomplemented variables to the gates of the PMOS transistors. If uncomplemented variables appear in the expression, additional inverters will be needed.
3. The PDN can be obtained from the PUN (and vice versa) using the duality property.

3.8 Transistor Sizing

Once a CMOS gate circuit has been generated, the only significant step remaining in the design is to decide on \(W/L\) ratios for all devices. These ratios usually are selected to provide the gate with current-driving capability in both directions equal to that of the basic inverter, The reader will recall from Section 4 that for the basic inverter
design, we denoted \((W/L)_n = n\) and \((W/L)_p = p\), where \(n\) is usually 1.5 to 2 and, for a matched design, \(p = (\mu_n/\mu_p)n\). Thus, we wish to select individual \(W/L\) ratios for all transistors in a logic gate so that the PDN should be able to provide a capacitor discharge current at least equal to that of an NMOS transistor with \(W/L = n\), and the PUN should be able to provide a charging current at least equal to that of a PMOS transistor with \(W/L = p\). This will guarantee a worst-case gate delay equal to that of the basic inverter.

"worst case" means that in deciding on device sizing, we should find the input combinations that result in the lowest output current and then choose sizes that will make this current equal to that of the basic inverter. The issue of determining the current-driving capability of a circuit consisting of a number of MOS devices. In other words, we need to find the equivalent \(W/L\) ratio of a network of MOS transistors.

Consider the parallel and series connection of MOSFETs and find the equivalent \(W/L\) ratios. The derivation of the equivalent \(W/L\) ratio is based on the fact that the on resistance of a MOSFET is inversely proportional to \(W/L\). Thus, if a number of MOSFETs having ratios of \((W/L)_1, (W/L)_2, \ldots\) are connected in series, the equivalent series resistance obtained by adding the on-resistances will be

\[
R_{\text{series}} = r_{DS1} + r_{DS2} + \ldots = (\text{constant}/(W/L)_1) + (\text{constant}/(W/L)_2) + \ldots
\]

\[
= \text{constant}[(1/(W/L)_1) + (1/(W/L)_2) + \ldots]
\]

\[
= (\text{constant}/(W/L)_{eq})
\]

resulting in the following expression for \((W/L)_{eq}\) for transistors connected in series;

\[
(W/L)_{eq} = [1/(1/(W/L)_1) + (1/(W/L)_2) + \ldots] \quad (25)
\]

Similarly, we can show that the parallel connection of transistors with \(W/L\) ratios of \((W/L)_1, (W/L)_2, \ldots\), results in an equivalent \(W/L\) of

\[
(W/L)_{eq} = (W/L)_1 + (W/L)_2 + \ldots \quad (26)
\]

As an example, two identical MOS transistors with individual \(W/L\) ratios of 4 result in an equivalent \(W/L\) of 2 when connected in series and of 8 when connected in parallel.
As an example of proper sizing, consider the four-input NOR in Fig. 18. Here, the worst case (the lowest current) for the PDN is obtained when only one of the NMOS transistors is conducting. We therefore select the \( W/L \) of each NMOS transistor to be equal to that of the NMOS transistor of the basic inverter, namely, \( n \). For the PUN, however, the worst-case situation (and indeed the only case) is when all inputs are low and the four series PMOS transistors are conducting. Since the equivalent \( W/L \) will be one-quarter of that of each PMOS device, we should select the \( W/L \) ratio of each PMOS transistor to be four times that of \( Q_p \) of the basic inverter, that is, \( 4p \).

As another example, we show in Fig. 19 the proper sizing for a four-input NAND gate. Comparison of the NAND and NOR gates in Figs. 18 and 19 indicates that because \( p \) is usually two to three times \( n \), the NOR gate will require much greater area than the NAND gate. For this reason, NAND gates are generally preferred for implementing combinational logic functions in CMOS.
3.9 Effects of Fan-In and Fan-Out on Propagation Delay

Each additional input to a CMOS gate requires two additional transistors, One NMOS and one PMOS. This is in contrast to other forms of MOS logic, where each additional input requires only one additional transistor. The additional transistor in CMOS not only increases the chip area but also increases the total effective capacitance per gate and in turn increases the propagation delay. The size-scaling method described earlier compensates for some (but not all) of the increase in $t_p$. By increasing device size, we are able to preserve the current-driving capability. The capacitance $C$ increases because of both the increased number of inputs and increase in device size. Thus $t_p$ will still increase with fan-in, a fact that imposes a practical limit on the fan-in of, say, the NAND gate to about 4.

An increase in a gate’s fan-out adds directly to its load capacitance and, thus, increases its propagation delay.

Thus although CMOS has many advantages, it does suffer from increased circuit complexity when the fan-in and fan-out are increased, and from the corresponding effects of this complexity on both chip area and propagation delay.
4 PASS-TRANSISTOR LOGIC CIRCUITS

Simple approach for implementing logic functions utilizes series and parallel combinations of switches that are controlled by Input logic variables to connect the input and output nodes (see Fig. 20). Each of the switches can be implemented either by a single NMOS transistor (Fig. 21(a)) or by a pair of complementary MOS transistors connected in what is known as the CMOS transmission-gate configuration (Fig. 21(b)).

This form of logic utilizes MOS transistors in the series path from input to output, to pass or block signal transmission, it is known as pass-transistor logic (PTL). As mentioned earlier, CMOS transmission gates are frequently employed to implement the switches, giving this logic-circuit form the alternative name, transmission-gate logic.

Figure 20 Conceptual pass-transistor logic gates. (a) Two switches, controlled by the input variables $B$ and $C$, when connected in series in the path between the input node to which an input variable $A$ is applied and the output node (with an implied load to ground) realize the function $Y = ABC$. (b) When the two switches are connected in parallel, the function realized is $Y = A(B + C)$.

Figure 21 Two possible implementations of a voltage-controlled switch connecting nodes $A$ and $Y$: (a) single NMOS transistor and (b) CMOS transmission gate.
4.1 An Essential Design Requirement

An essential requirement in the design of PTL circuits is ensuring that every-circuit node has at all times a low-resistance path to $V_{DD}$ or ground. Consider the situation depicted in Fig. 22(a): A switch $S_1$ (usually part of a larger PTL network, not shown) is used to form the AND function of its controlling variable $B$ and the variable $A$ available at the output of a CMOS inverter. The output $Y$ of the PTL circuit is shown connected to the input of another inverter. Obviously, if $B$ is high, $S_1$ closes and $Y = A$. Node $Y$ will then be connected either to $V_{DD}$ (if $A$ is high) through $Q_2$ or to ground (if $A$ is low) through $Q_1$.

But, what happens when $B$ goes low and $S_1$ opens? Node $Y$ will now become a high-impedance node. If initially, $v_Y$ was zero, it will remain so. However, if initially, $v_Y$ was high at $V_{DD}$, this voltage will be maintained by the charge on the parasitic capacitance $C$, but for only a time: The inevitable leakage currents will slowly discharge $C$, and $v_Y$ will diminish correspondingly. In any case, the circuit can no longer be considered a static combinational logic circuit.

The problem can be easily solved by establishing for node $Y$ a low-resistance path that is activated when $B$ goes low, as shown in Fig. 22(b). Here, another switch $S_2$ controlled by $B$ is connected between $Y$ and ground. When $B$ goes low, $S_2$ closes and establishes a low-resistance path between $Y$ and ground,
Figure 22 A basic design requirement of PTL circuits is that every node have, at all times, a low-resistance path to either ground or \( V_{DD} \). Such a path does not exist in (a) when \( B \) is low and \( S_1 \) is open. It is provided in (b) through switch \( S_2 \).

### 4.2 Operation with NMOS Transistors as Switches

PTL circuit with single NMOS transistors results in a simple circuit with small area and small node capacitances. Consider the circuit shown in Fig. 23, where an NMOS transistor \( Q \) is used to implement a switch connecting an input node with voltage \( v_j \) and an output node. The total capacitance between the output node and ground is represented by capacitor \( C \). The switch is shown in the closed state with the control signal applied to its gate being high at \( V_{DD} \). The operation of the circuit as the input voltage \( V_i \) goes high (to \( V_{DD} \)) at time \( t = 0 \). We assume that initially the output voltage \( v_o \) is zero and capacitor \( C \) is fully discharged.

When \( v_j \) goes high, the transistor operates in the saturation mode and delivers a current \( i_D \) to charge the capacitor,

\[
i_D = \frac{1}{2} k_n (V_{DD} - v_o - V_t)^2
\]  

(27)

where \( k_n = k'_n (W/L) \), and \( V_t \), is determined by the body effect since the source is at a voltage \( v_o \) relative to the body,

\[
V_t = V_{to} + \gamma (\sqrt{V_o + 2 \varphi_f} - \sqrt{2 \varphi_f})
\]  

(28)

Thus, initially (at \( t = 0 \)), \( V_t = V_{to} \) and the current \( i_D \) is relatively large. However, as \( C \) charges up and \( v_o \) rises, \( V_t \) increases (Eq.28) and \( i_D \) decreases. Observe from Eq. (27) that \( i_D \) reduces to zero when \( v_o \) reaches \( V_{DD} - V_t \). Thus the high output voltage (\( V_{OH} \)) will not be equal to \( V_{DD} \); rather, it will be lower by \( V_t \), the value of \( V_t \) can be as high as 1.5 to 2 times \( V_{to} \)!

![Figure 23](image)

Figure 23 Operation of the NMOS transistor as a switch in the implementation of PTL circuits. This analysis is for the case with the switch closed (\( v_C \) is high) and the input going high (\( v_I = V_{DD} \)).

The propagation delay \( t_{PLH} \) of the PTL gate of Fig. 23 can be determined as the time for \( v_o \) to reach \( V_{DD}/2 \).
Figure 24 shows the NMOS switch circuit when $v_j$ is brought down to 0 V, assume that initially $v_O = V_{DD}$. Thus at $t = 0^+$, the transistor conducts and operates in the saturation region,

$$i_D = \frac{1}{2} k_n (V_{DD} - V_t)^2$$  \hspace{1cm} (29)

Since the source is now at 0 V (note that the drain and source have interchanged roles), there will be no body effect, and $V_t$ remains constant at $V_t0$. As C discharges, $v_O$ decreases and the transistor enters the triode region at $v_O = (V_{DD} - V_t)$. The capacitor discharge continues until $C$ is fully discharged and $v_O = 0$. Thus, the NMOS transistor provides $V_{OL} = 0$, or a "good 0." Again, the propagation delay $t_{PHL}$ can be determined.

### 4.3 The Use of CMOS Transmission Gates as Switches

Great improvements in static and dynamic performance are obtained when the switches are implemented with CMOS transmission gates. The transmission gate utilizes a pair of complementary transistors connected in parallel. It acts as an excellent switch, providing bidirectional current flow, and it exhibits an on-resistance that remains almost constant for wide ranges of input voltage. These characteristics make the transmission gate not only an excellent switch in digital applications but also an excellent analog switch in such applications as data converters and switched-capacitor filters.

Figure 25(a) shows the transmission-gate switch in the "on" position with the input, $V_i$, rising to $V_{DD}$ at $t = 0$. Assuming, as before, that initially the output voltage is zero, we see that $Q_N$ will be operating in saturation and providing a charging current of

$$i_{DN} = \frac{1}{2} k_n (V_{DD} - v_O - V_m)^2$$  \hspace{1cm} (30)

where, as in the case of the single NMOS switch, $V_m$ is determined by the body effect,

$$V_m = V_{t0} + \gamma (\sqrt{V_O + 2\varphi_f} - \sqrt{2\varphi_f})$$  \hspace{1cm} (31)
Transistor $Q_N$ will conduct a diminishing current that reduces to zero at $v_O = V_{DD} - V_t$. Observe, however, that $Q_P$ operates with $V_{SG} = V_{DD}$ and is initially in saturation,

$$i_{DP} = \frac{1}{2} k_p (V_{DD} - |V_{tp}|)^2$$  \hspace{1cm} (32)$$

where, since the body of $Q_P$ is connected to $V_{DD} - |V_{tp}|$ remains constant at the value $V_{to}$, assumed to be the same value as for the n-channel device. The total capacitor-charging current is the sum of $i_{DN}$ and $i_{DP}$. Now, $Q_p$ will enter the triode region at $v_O = |V_{tp}|$, but will continue to conduct until $C$ is fully charged and $v_O = V_{OH} = V_{DD}$, Thus, the P-channel device will provide the gate with a "good 1." The value of $t_{PLH}$ can be calculated using usual techniques. Note, however, that adding the PMOS transistor increases the value of $C$.

When $v_i$ goes low, as shown in Fig.25(b)

![Figure 25](image)

**Figure 25** Operation of the transmission gate as a switch in PTL circuits with (a) $v_i$ high and (b) $v_i$ low.

$Q_N$ and $Q_P$ interchange roles. Analysis of the circuit in Fig.25(b) will indicate that $Q_P$ will cease conduction when $v_O$ falls to $|V_{tp}|$, where $|V_{tp}|$ is given by
\[ |V_{tp}| = V_{to} + \gamma \left( \sqrt{V_{DD} - v_o + 2\varphi_f} - \sqrt{2\varphi_f} \right) \]  \hspace{1cm} (33)

Transistor Q_N, however, continues to conduct until C is fully discharged and \( v_O = V_{OL} = 0 \) V, a "good 0."

We conclude that transmission gates provide far superior performance, both static and dynamic, than is possible with single NMOS switches. The Disadvantage is increased circuit complexity, area, and capacitance.

### 4.4 Pass-Transistor Logic Circuit Examples

Figure 26 shows a PTL realization of a two-to-one multiplexer: Depending on the logic value of C, either A or B is connected to the output Y. The circuit realizes the Boolean function

\[ Y = CA + \overline{CB} \]

![Figure 26](image)

Our second example is an efficient realization of the exclusive-OR (XOR) function. The circuit shown in Fig. 27, utilizes four transistors in the transmission gates and another four for the two inverters needed to generate the complements A and B, for a total of eight transistors. Note that 12 transistors are needed in the realization with complementary CMOS.
Our final PTL example is the circuit shown in Fig. 28. It uses NMOS switches with low or zero threshold. Observe that both the input variables and their complements are employed and that the circuit generates both the Boolean function and its complement. Thus this form of circuit is known as **complementary pass-transistor logic** (CPL). The circuit consists of two identical networks of pass transistors with the corresponding transistor gates controlled by the same signal (B and $\overline{B}$). The inputs to the PTL, however, are complemented: $A$ and $B$ for the first network, and $A$ and $B$ for the second. The circuit shown realizes both the AND and NAND functions.
5 DYNAMIC LOGIC CIRCUITS

To place dynamic-logic-circuit techniques into perspective, let's take stock of the various logic-circuit styles we have studied.

Complementary CMOS excels in nearly every performance category: It is easy to design, has the maximum possible logic swing, is robust from a noise-immunity standpoint, dissipates no static power, and can be designed to provide equal low-to-high and high-to-low propagation delays. Its main disadvantage is the requirement of two transistors for each additional gate input, which for high fan-in gates can make the chip area large and increase the total capacitance and, correspondingly, the propagation delay and the dynamic power dissipation.

Pseudo-NMOS reduces the number of required transistors at the expense of static power dissipation. Pass-transistor logic can result in simple small-area circuits but is limited to special applications and requires the use of complementary inverters to restore signal levels, especially when the switches are simple NMOS transistors.

The dynamic logic techniques is to maintain the low device count of pseudo-NMOS while reducing the static power dissipation to zero. This is achieved at the expense of more complex, and less robust, design.

5.1 Basic Principle

Figure 29(a) shows the basic dynamic-logic gate. It consists of a pull-down network (PDN) that realizes the logic function in exactly the same way as the PDN of a complementary CMOS gate or a pseudo-NMOS gate. Here, however, we have two switches in series that are periodically operated by the clock signal $\phi$ whose waveform is shown in Fig. 29(b). When $\phi$ is low, $Q_p$ is turned on, and the circuit is said to be in the setup or precharge phase. When $\phi$ is high, $Q_p$ is off and $Q_e$ turns on, and the circuit is in the evaluation phase. Finally, note that $C_L$, denotes the total capacitance between the output node and ground.

During precharge, $Q_p$ conducts and charges capacitance $C_L$ so that, at the end of the precharge interval, the voltage at Y is equal to $V_{DD}$. Also during precharge, the inputs $A$, $B$, and $C$ are allowed to change and settle to their proper values. Observe that because $Q_e$ is off, no path to ground exists.

During the evaluation phase, $Q_p$ is off and $Q_e$ is turned on. Now, if the input combination is one that corresponds to a high output, the PDN does not conduct (just as in a complementary CMOS gate) and the output remains high at $V_{DD}$ thus $V_{OH} = V_{DD}$. Observe that no low-to-high propagation delay is required, thus $t_{PLH}=0$. On the other hand, if the combination of inputs is one that corresponds to a low output, the appropriate NMOS transistors in the PDN will conduct and establish a path between the output node and ground through the on-transistor $Q_e$. Thus $C_L$ will be discharged through the PDN, and the voltage at the output node will reduce to $V_{OL} = 0$ V. The high-to-low propagation delay $t_{PHL}$ can be calculated in exactly the same way as for a complementary CMOS circuit except that here we have an additional transistor, $Q_e$, in the series path to ground.
5.2 Nonideal Effects

We now briefly consider various sources of non ideal operation of dynamic logic circuits.

**Noise Margins** Since, during the evaluation phase, the NMOS transistors begin to conduct for $V_i = V_{th}$, and thus the noise margins will be

$$NM_L = V_{th}$$
$$NM_H = V_{DD} - V_{th}$$

**Output Voltage Decay Due to Leakage Effects**

In the absence of a path to ground through the PDN, the output voltage will ideally remain high at $V_{DD}$. This, however, is based on the assumption that the charge on $C_L$ will remain intact. In practice, there will be leakage current that will cause $C_L$ to slowly discharge and $v_\gamma$ to decay. The principal source of leakage is the reverse current of the reverse-biased junction between the drain diffusion of transistors connected to the output node and the substrate. Such currents can be in the range of $10^{-12}$ A to $10^{-15}$ A, and they increase rapidly with temperature (approximately doubling for every 10°C rise in temperature). Thus the circuit can malfunction if the clock is operating at a very low frequency and the output node is not "refreshed" periodically.

**Charge Sharing**

There is another and often more serious way for $C_L$ to lose some of its charge and thus cause $v_\gamma$ to fall significantly below $V_{DD}$. To see how this can happen, refer to Fig.30,
which shows only $Q_1$ and $Q_2$, the two top transistors of the PDN, together with the precharge transistor $Q_P$. Here, $C_1$ is the capacitance between the common node of $Q_1$ and $Q_2$ and ground. At the beginning of the evaluation phase, after $Q_P$ has turned off and with $C_L$ charged to $V_{DD}$ (Fig. 30), we assume that $C_1$ is initially discharged and that the inputs are such that at the gate of $Q_1$ we have a high signal, whereas at the gate of $Q_2$ the signal is low. We can easily see that $Q_1$ will turn on, and its drain current $i_{D1}$ will flow as indicated.

Thus $i_{D1}$ will discharge $C_L$ and charge $C_1$. Although eventually $i_{D1}$ will reduce to zero, $C_L$ will have lost some of its charge, which will have been transferred to $C_1$. This phenomenon is known as charge sharing.

A serious problem arises if one attempts to cascade dynamic logic gates. Consider the situation depicted in Fig. 31, where two single-input dynamic gates are connected in cascade. During the precharge phase, $C_{L1}$ and $C_{L2}$ will be charged through $Q_{P1}$ and $Q_{P2}$ respectively. Thus, at the end of the precharge interval, $V_{\gamma1} = V_{DD}$ and $V_{\gamma2} = V_{DD}$. Now consider what happens in the evaluation phase for the case of high input $A$. Obviously, the correct result will be $y_1$ low ($V_{\gamma1} = 0$ V) and $y_2$ high ($V_{\gamma2} = V_{DD}$). As the evaluation phase begins, $Q_1$ turns on and $C_{L1}$ begins to discharge. However, simultaneously, $Q_2$ turns on and $C_{L2}$ also begins to discharge. Only when $V_{\gamma1}$ drops below $V_{in}$ will $Q_2$ turn off. Unfortunately, however, by that time, $C_{L2}$ will have lost a significant amount of its charge, and $V_{\gamma2}$ will be less than the expected value of $V_{DD}$. (It is important to note that in dynamic logic, once charge has been lost, it cannot be recovered.)
Figure 31  Two single-input dynamic logic gates connected in cascade. With the input \( A \) high, during the evaluation phase \( C_{L2} \) will partially discharge and the output at \( Y_2 \) will fall lower than \( V_{DD} \), which can cause logic malfunction.

5.3 Domino CMOS Logic

Domino CMOS logic is a form of dynamic logic that results in cascadable gates. Figure 32 shows the structure of the Domino CMOS logic gate. We observe that it is simply the basic dynamic-logic gate of Fig.29(a) with a static CMOS inverter connected to its output. Operation of the gate is straightforward. During precharge, \( X \) will be raised to \( V_{DD} \), and the gate output \( Y \) will be at 0V. During evaluation, depending on the combination of the input variables, either \( X \) will remain high and thus the output \( Y \) will remain low (\( t_{PHL} = 0 \)) or \( X \) will be brought down to 0V and the output \( Y \) will rise to \( V_{DD} \) (\( t_{PLH} \) finite). Thus, during evaluation, the output either remains low or makes only one low-to-high transition.

Consider the situation in Fig.33(a), where we show two Domino gates connected in cascade. For simplicity, we show single input gates. At the end of precharge, \( X_1 \) will be at \( V_{DD} \), \( Y_1 \) will be at 0V, \( X_2 \) will be at \( V_{DD} \), and \( Y_2 \) will be at 0V. As the preceding case, assume \( A \) is high at the beginning of evaluation. Thus, as \( \phi \) goes up, capacitor \( C_{L1} \) will begin discharging, pulling \( X_1 \) down. Meanwhile, the low input at the gate of \( Q_2 \) keeps \( Q_2 \) off, and \( C_{L2} \) remains fully charged. When \( v_{X1} \) falls below the threshold voltage of inverter \( I_1 \), \( Y_1 \) will go up turning \( Q_2 \) on, which in turn begins to discharge \( C_{L2} \) and pulls \( X_2 \) low. Eventually, \( Y_2 \) rises to \( V_{DD} \).
Figure 32  The Domino CMOS logic gate. The circuit consists of a dynamic-MOS logic gate with a static-CMOS inverter connected to the output. During evaluation, $Y$ either will remain low (at 0 V) or will make one 0-to-1 transition (to $V_{DD}$).

From this description, we see that because the output of the Domino gate is low at the beginning of evaluation, no premature capacitor discharge will occur in the subsequent gate in the cascade. As indicated in Fig.33(b), output $Y_1$ will make a 0-to-1 transition $t_{PLH}$ seconds after the rising edge of the clock. Subsequently, output $Y_2$ makes a 0-to-1 transition after another $t_{PLH}$ interval.

Figure 10.33(a) Two single-input domino CMOS logic gates connected in cascade. (b) Waveforms during the evaluation phase.

Domino CMOS logic finds application in the design of address decoders in memory chips.

ALL THE BEST