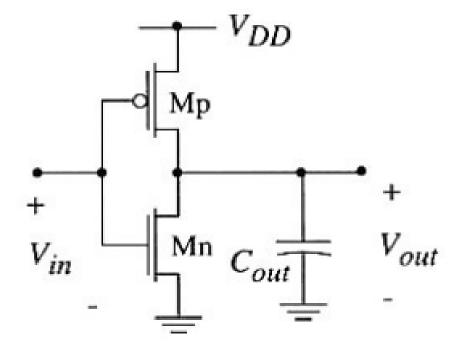
MODULE-2 CMOS Inverter Switching Characteristics

ECT304 VLSI CIRCUIT DESIGN

- Transient switching times are used to calculate data throughput rates and are also important in system timing.
- Switching times are determined by two circuit properties: transistor current flow levels and parasitic capacitances.
- ➤ Both are set by the chip design parameters, and are sensitive to the transistor aspect ratios, layout geometry, and logic routing.
- ➤ To model the basic problem, we introduce the **output capacitance** C_{out} shown in Figure 2.15.

- C_{out} represents the total capacitance at the output node, and consists of contributions from the MOSFETs and the external network.
- For our analytic calculations, C_{out} is assumed to be a linear, time-invariant (LTI) quantity.

Figure 2.15: Output Capacitance



- Switching performance of CMOS digital circuits are characterized by the time intervals required to charge and discharge capacitors at output nodes.
- CMOS inverters use transistors to provide current flow paths between the power supply (M_p) and ground (M_n).
- All switching times are thus set by the current levels and the value of C_{out}.
- ➤ Figure 2.16 shows the inverter input and output voltages as functions of time.
- ➤ The input waveform V_{in}(t) has been taken to have idealized step characteristics to simplifies the calculations and also provides a standard reference.

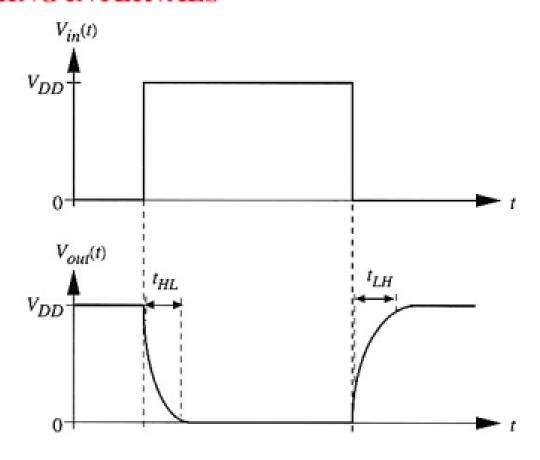
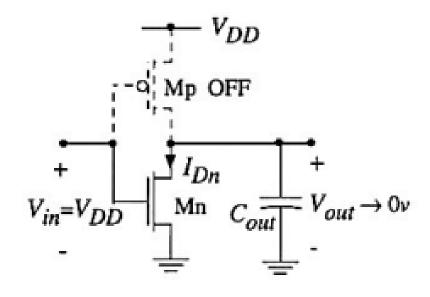


Figure 2.16: Switching Time Definitions

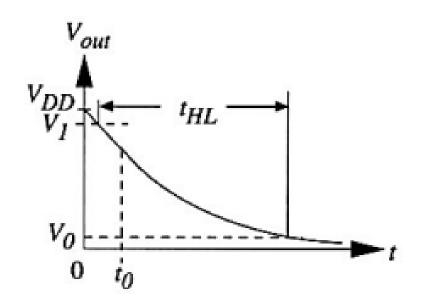
- When the input voltage is low with $V_{in} = 0V$, the output voltage is high at a value of $V_{out} = V_{DD}$
- This corresponds to the case where the nMOS is OFF, while the pMOS is ON and provides the connection to the power supply.
- ightharpoonup Changing the input voltage to high value $V_{in} = V_{DD}$ reverses this.
- Now the nMOS is active while the pMOS is in cut off.
- The capacitor discharges to 0V through M_n , and the output voltage decays to a final value of $V_{out} = 0V$ as shown.
- ➤ The switching time associated with this decay is the output high-to-low time, t_{HL}

- ➤ If the input voltage is returned to a low voltage V_{in} = 0V, the nMOS is driven into cut off while the pMOS reconnects C_{out} to the power supply.
- This allows to charge to a final voltage of $V_{out} = V_{DD}$ in a characteristic time t_{LH} , the output low-to-high time.
- t_{HL} and t_{LH} represent the times required for the output to stabilize to a final value in response to changes of the input voltage.
- These are limiting factors in the performance of a digital CMOS logic circuit.

\checkmark HIGH-TO-LOW TIME (t_{HI})



(a) Discharge circuit



(b) Output voltage

✓ HIGH-TO-LOW TIME (t_{HL})

- ➤ The output high-to-low time is calculated using the subcircuit in Figure 2.17a.
- ➤ It represents the time interval needed for the output capacitor to discharge through the n-channel MOSFET M_n when M_p is in cut off.
- ➤ t_{HL} is also referred to as the fall time t_f for the circuit since it gives the time needed for the output to decay from a well-defined logic 1 state to a well-defined logic 0 state.
- The discharge is described by the capacitor equation

$$I_{Dn} = -C_{out} \frac{dV_{out}}{dt}$$

EQN 2.26

where we will assume an initial condition of $V_{out}(t=0) = V_{DD}$

- ✓ HIGH-TO-LOW TIME (t_{HL})
- ➤ Minus sign is required because the current is leaving the positive terminal.
- ➤ At the beginning of the discharge, M_n is saturated, so that

$$\frac{\beta_n}{2}(V_{DD} - V_{THn})^2 = -C_{out}\frac{dV_{out}}{dt}$$
 EQN 2.27

- describes the initial discharge.
- ➤ Integrating gives a linear decay in time be means of the function

$$V_{out}(t) = V_{DD} - \frac{\beta_n (V_{DD} - V_{THn})^2 t}{2C_{out}}$$
 EQN 2.28

- This is valid until a time t_{θ} when the output voltage drops to $V_{out} = (V_{DD} V_{THn})$ when the MOSFET enters the non-saturated conduction region.
- This is indicated in Figure 2.17b.

✓ HIGH-TO-LOW TIME (t_{HI})

 \triangleright The value of t_0 is found by setting

$$V_{out}(t_0) = V_{DD} - \frac{\beta_n (V_{DD} - V_{THn})^2 t_0}{2C_{out}} = V_{DD} - V_{THn}$$
EQN 2.29

so that

$$t_0 = \frac{2C_{out}V_{THn}}{\beta_n(V_{DD} - V_{THn})^2}$$
 EQN 2.30

For times $t \ge t_0$, the differential equation describing the discharge is

$$\frac{\beta_n}{2}[2(V_{DD} - V_{THn})V_{out} - (V_{out})^2] = -C_{out}\frac{dV_{out}}{dt} \text{ EQN 2.31}$$

because the nMOS is non-saturated.

- ✓ HIGH-TO-LOW TIME (t_{HL})
- This integrates to

$$V_{out}(t) = (V_{DD} - V_{THn}) \frac{2e^{-(t-t_0)}/\tau_n}{1 + e^{-(t-t_0)}/\tau_n}$$
 EQN 2.32

> where

$$\tau_n = \frac{C_{out}}{\beta_n (V_{DD} - V_{THn})}$$

EQN 2.33

- is the time constant for the discharge circuit.
- ➤ The value of t_{HL} is usually defined between the 10% and 90% voltages and respectively, with

$$V_0 = 0.1 V_{DD}$$
$$V_1 = 0.9 V_{DD}$$

EQN 2.34

for a full-rail output CMOS circuit.

- ✓ HIGH-TO-LOW TIME (t_{HL})
- This can be computed by using the integrals

$$t_{HL} = C_{out} \int_{(V_{DD} - V_{THn})}^{V_1} \frac{dV_{out}}{I_{Dn(sat)}} + C_{out} \int_{V_0}^{(V_{DD} - V_{THn})} \frac{dV_{out}}{I_{Dn(nonsat)}}$$

EQN 2.35

- > or by determining the required time intervals from the equations above.
- Either approach gives the result

$$t_{HL} = s_n \tau_n$$
 EQN 2.36

where

$$s_n = \frac{2(V_{THn} - V_0)}{(V_{DD} - V_{THn})} + \ln\left(\frac{2(V_{DD} - V_{THn})}{V_0} - 1\right)$$
 EQN 2.37

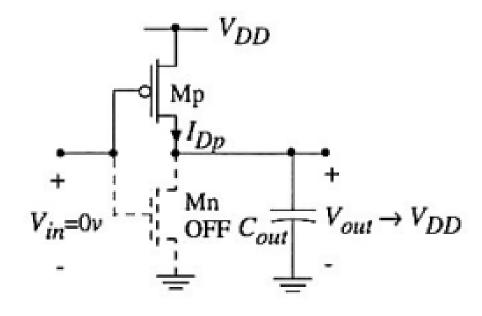
- is a voltage-dependent scaling multiplier.
- ➤ The first term in represents the time when M_n is saturated, while the second term is due to non-saturated conduction.

- ✓ HIGH-TO-LOW TIME (t_{HL})
- The definition of the time constant τ_n allows us to write that $\tau_n = R_n C_{out}$ where

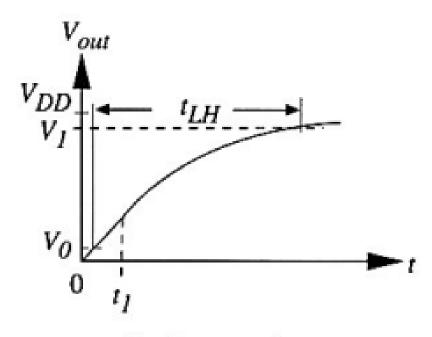
$$R_n = \frac{1}{\beta_n (V_{DD} - V_{THn})}$$
 EQN 2.38

- represents an equivalent LTI value for the drain-to-source resistance.
- The concept of the MOSFET resistance must be used with care, since the MOSFET is inherently a non-linear device while resistances are usually assumed to be linear.

✓ LOW-TO-HIGH TIME (t_{LH})



(a) Charge circuit



(b) Output voltage

✓ LOW-TO-HIGH TIME (t_{LH})

- The low-to-high time t_{LH} also known as the rise time t_r is found in the same manner as t_{HL} .
- During this time interval, M_n is in cut off while M_p is conducting from the power supply.
- ➤ As shown in Figure 2.18a, t_{LH} is the time required to charge C_{out} through M_p.
- \triangleright It is also referred to as the charge time t_{ch} in the literature.
- Charging is described by the general equation

$$I_{Dp} = C_{out} \frac{dV_{out}}{dt}$$

EQN 2.39

 \triangleright with the initial condition $V_{out}(t=0) = 0V$

✓ LOW-TO-HIGH TIME (t_{LH})

➤ When the charging starts, M_p is saturated and the integration gives

$$V_{out}(t) = \frac{\beta_p (V_{DD} - |V_{THp}|)^2 t}{2C_{out}}$$
 EQN 2.40

This is valid until a time t₁

$$t_1 = \frac{2C_{out}|V_{THp}|}{\beta_p(V_{DD} - |V_{THp}|)^2}$$
 EQN 2.41

- \triangleright where $V_{out}(t_1) = |V_{THp}|$
- ➤ This point is shown in the graph of Figure 2.18b.

✓ LOW-TO-HIGH TIME (t_{l.H})

For times $t \ge t_1$, M_p is non-saturated and the output voltage is described by

$$V_{out}(t) = V_{DD} - (V_{DD} - |V_{THp}|) \frac{2e^{-(t-t_1)/\tau_p}}{1 + e^{-(t-t_1)/\tau_p}}$$
 EQN 2.42

where the charging time constant is

$$\tau_p = \frac{C_{out}}{\beta_p (V_{DD} - |V_{THp}|)}$$
 EQN 2.43

▶ Defining \mathbf{t}_{LH} as the time to charge \mathbf{C}_{out} from \mathbf{V}_{0} (the 10% point) to \mathbf{V}_{1} (the 90% point) gives

$$t_{LH} = s_p \tau_p$$
 EQN 2.44

- ✓ LOW-TO-HIGH TIME (t_{LH})
- ➤ with

$$s_p = \frac{2(|V_{THp}| - V_0)}{(V_{DD} - |V_{THp}|)} + \ln\left(\frac{2(V_{DD} - |V_{THp}|)}{V_0} - 1\right) \quad EQN \ 2.45$$

- as the multiplier for this time interval.
- Note that t_{LH} has the same form as the fall time except that pMOS parameters appear instead of the nMOS quantities.
- A pMOS resistance may be approximated by

$$R_p = \frac{1}{\beta_p (V_{DD} - |V_{THp}|)}$$
 EQN 2.46

- such that gives the charging time constant.
- \triangleright Both R_p and R_n are inversely proportional to (W/L); increasing the aspect ratio thus decreases the equivalent resistance.

✓ MAXIMUM SWITCHING FREQUENCY

- The sum of the transient times $(t_{HL} + t_{LH})$ represents the minimum time needed for a gate to undergo a complete switching cycle, i.e., for the output to change from a logic 1 to a logic 0 voltage, and then back up to a logic 1 value.
- > The maximum switching frequency is given by

$$f_{max} = \frac{1}{t_{HL} + t_{LH}} = \frac{1}{s_n \tau_n + s_p \tau_p}$$

EQN 2.47

- This represents the maximum rate of data transfer for the gate.
- \triangleright In system design, the working value of f_{max} is set by the slowest gate or data path element in the network.

✓ MAXIMUM SWITCHING FREQUENCY

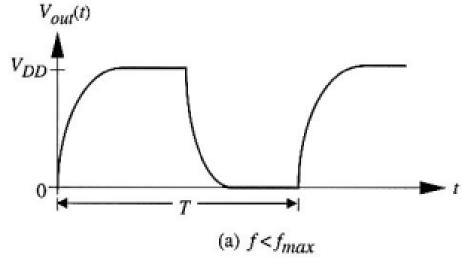
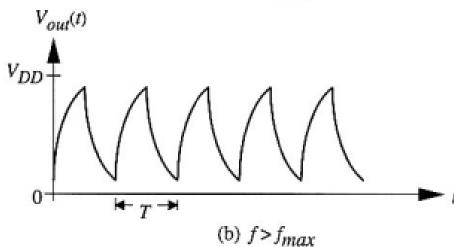


Figure 2.19: Output Voltage Waveforms



✓ MAXIMUM SWITCHING FREQUENCY

- Figure 2.19 illustrates the importance of f_{max} for the inverter.
- For signal frequencies $f < f_{max}$ as in Figure 2.19a, the output has sufficient time to react to changes in the inputs and exhibits proper form.
- If we increase the signal frequency to $f > f_{max}$ the circuit does not have enough time to complete the charge or discharge event.
- This gives an output signal that has a limited amplitude that may cause a logic error.

> TRANSIENT EFFECTS ON THE VTC

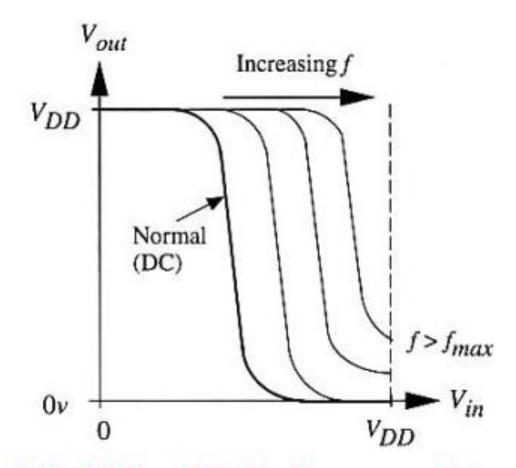


Figure 2.20: VTC modified for frequency effects

✓ TRANSIENT EFFECTS ON THE VTC

- ➤ Although the VTC is defined to be the DC transfer curve, it is useful for illustrating the transient switching effects by generating the family of curves as shown in Figure 2.20.
- This corresponds to merging the transient behaviour manifest in $V_{out}(t)$ and $V_{in}(t)$ by eliminating the time t as a variable.
- ➤ For low switching frequencies we obtain the usual plot that gives the DC behaviour.
- As the signal frequency is increased, the behaviour of V_{out} as a function of V_{in} shows that there is a change in the response of the network.
- When the switching frequency exceeds f_{max} , then the circuit cannot respond to the quickly changing input.
- For this case, the output voltage never reaches a value of **0V**.

✓ PROPAGATION DELAY

- Logic delay through a gate is described by the propagation delay time
 t_p
- > t_p is the average time needed for the output to respond to a change in the input logic state.

$$t_P = \frac{t_{PHL} + t_{PLH}}{2}$$
 EQN 2.48

- Where t_{PHL} and t_{PLH} represent the propagation delays for a high-to-low, and a low-to-high transition, respectively.
- \triangleright Let us define the 50% voltage points as $V_{1/2} = 0.5V_{DD}$
- Then, t_{PHL} and t_{PLH} are defined by the time intervals between the input and output voltages as shown in Figure 2.21.

✓ PROPAGATION DELAY

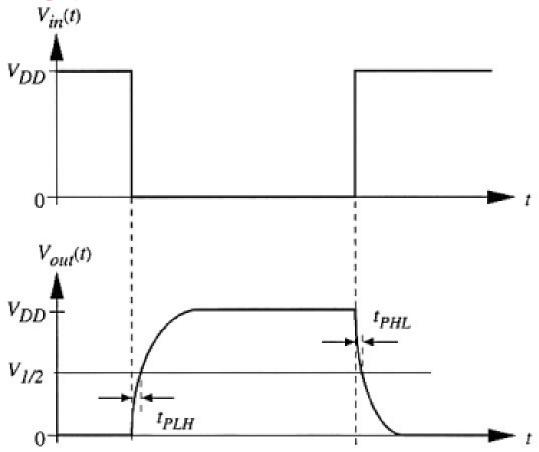


Figure 2.21: Propagation Delay with a step input voltage

✓ PROPAGATION DELAY

- The high-to-low propagation delay represents the time needed for the output to fall from V_{DD} to V_{I} to simplify the calculations, we usually approximate $V_{I} = V_{DD}/2$
- This yields the general expression

$$t_{PHL} = C_{out} \int_{(V_{DD} - V_{THn})}^{V_{DD}} \frac{dV_{out}}{I_{Dn(sat)}} + C_{out} \int_{V_{DD}/2}^{(V_{DD} - V_{THn})} \frac{dV_{out}}{I_{Dn(nonsat)}}$$
EQN 2.49

> which defines the basic integrals. Evaluating yields

$$t_{PHL} = s_n' \tau_n$$
 EQN 2.50

ightharpoonup where $\tau_n = R_n C_{out}$ is the time constant, and the new scaling factor is

$$s_{n}' = \frac{2V_{THn}}{(V_{DD} - V_{THn})} + \ln\left(\frac{4(V_{DD} - V_{THn})}{V_{DD}} - 1\right)$$
 EQN 2.51

✓ PROPAGATION DELAY

The value of is computed in the same manner with

$$t_{PLH} = s_p' \tau_p$$

EQN 2.52

where

$$s_{p}' = \frac{2|V_{THp}|}{(V_{DD} - |V_{THp}|)} + \ln\left(\frac{4(V_{DD} - |V_{THp}|)}{V_{DD}} - 1\right) \text{ EQN 2.53}$$

provides the scaling factor. Combining terms thus gives

$$t_P = \frac{1}{2}(s_n'\tau_n + s_p'\tau_p)$$
 EQN 2.54

as the total propagation delay.