

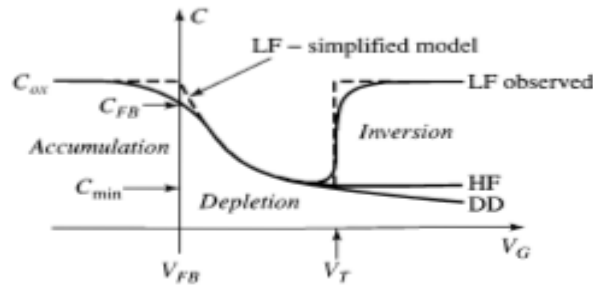
Q.2 a. Explain the C-V characteristics of Ideal MOS System. Explain Flat Band Voltage. (8)

Answer:

Depending on the gate voltage (V_G) ideal MOS system works in three regions of operation

1. Accumulation ($V_G < 0$)
2. Depletion ($V_G > 0$)
3. Inversion ($V_G > V_D$)

In the idealized MOS system at thermal equilibrium, the metal and the semiconductor form two plates of a capacitor. The capacitance C varies according to region of operation and hence is a function of gate voltage V_G .



Flat –Band Voltage: In the idealized MOS system at thermal equilibrium, the metal and the semiconductor form two plates of a capacitor. The capacitor is charged to a voltage corresponding to the difference between the metal and semiconductor work functions. Applying a bias voltage between the metal and silicon causes the system to depart from thermal equilibrium and changes the amount of charge stored on the capacitor. At one particular value the applied voltage exactly compensates the difference in the work functions of the metal and the semiconductor. The stored charge on the MOS capacitor is then reduced to zero, and the fields in the oxide and semiconductor vanish. In this situation, the energy bands in silicon are level or flat in the surface region as well as in the bulk. Because of the effect of the applied voltage on the band diagram, the voltages that produces flat energy bands in the silicon is called the flat-band voltage and usually designated as V_{FB} . The voltage applied to the ideal MOS system to bring it to the flat-band condition equals the difference in the work functions of the metal and silicon.

$$V_{FB} = \phi_M - \phi_S = \phi_{MS}$$

b. Explain the different fabrication steps of n well process for CMOS Inverter. (8)

Answer:

The n-well Process

(4+4)

As indicated earlier, although the p-well process is widely used, n-well fabrication has also gained wide acceptance, initially as a retrofit to nMOS lines. N-well CMOS circuits are also superior to p-well because of the lower substrate bias effects on transistor threshold voltage and inherently lower parasitic capacitances associated with source and drain regions. Typical n-well fabrication steps are illustrated in Figure. The first mask defines the n-well regions. This is followed by a low dose phosphorus implant driven in by a high temperature diffusion step to form the n-wells. The well depth is optimized to

ensure against p-substrate top+ diffusion breakdown without compromising then-well to n+ mask separation. The next steps are to define the devices and diffusion paths, grow field oxide, deposit and pattern the polysilicon, carry out the diffusions, make contact cuts, and finally metalize as before. It will be seen that an n+ mask and its complement may be used to define the n- and p-diffusion regions respectively. These same masks also include the V_{DD} and V_{SS} contacts- (respectively). It should be noted that, alternatively, we could have used a p+ mask and its Complement; since the n + and p + masks are generally complementary.

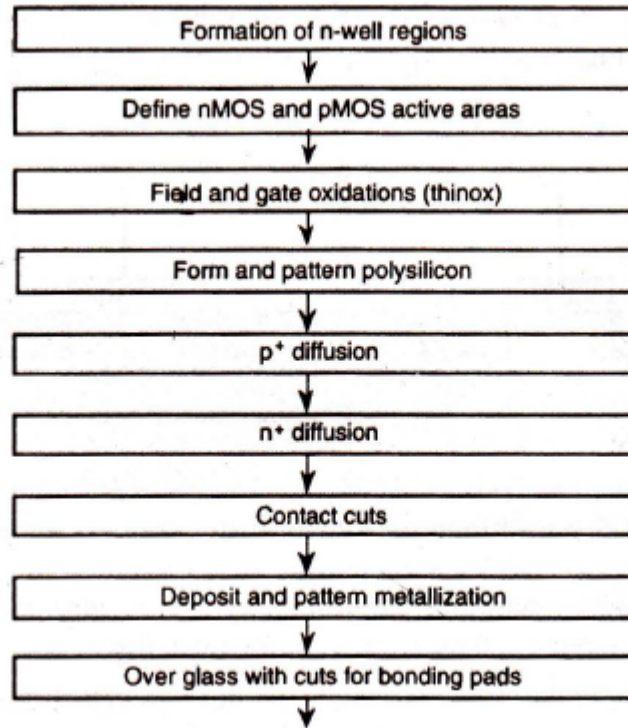


Figure: Main steps in a typical n-well process

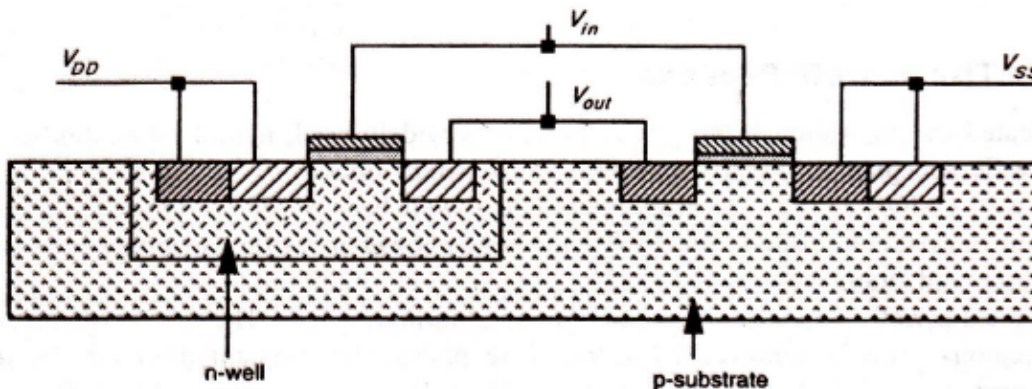


Figure: Cross-sectional view of n-wen CMOS Inverter

Q.3 a. Derive the expression for the threshold voltage of a MOS transistor and explain the significance of different parameters present in the equation. (8)

Answer:

(2 Inif. Expression+2 QSS+Final Parameter)

The threshold voltage V_t may be expressed as:

$$V_t = \phi_{ms} \frac{Q_B - Q_{SS}}{C_0} + 2\phi_{fN}$$

where

Q_B = the charge per unit area in the depletion layer beneath the oxide

Q_{SS} = charge density at Si:SiO₂ interface

C_0 = capacitance per unit gate area

ϕ_{ms} = work function difference between gate and Si

ϕ_{fN} = Fermi level potential between inverted surface and bulk Si.

Now, for polysilicon gate and silicon substrate, the value of ϕ_{ms} is negative but negligible, and the magnitude and sign of V_t are thus determined by the balance between the remaining negative term $\frac{-Q_{SS}}{C_0}$ and the other two terms, both of which are positive. To evaluate V_t , each term is determined as follows:

$$Q_B = \sqrt{2\epsilon_0\epsilon_{Si}qN(2\phi_{fN} + V_{SB})} \text{ coulomb/m}^2$$

$$\phi_{fN} = \frac{kT}{q} \ln \frac{N}{n_i} \text{ volts}$$

$$Q_{SS} = (1.5 \text{ to } 8) \times 10^{-8} \text{ coulomb/m}^2$$

depending on crystal orientation, and where

V_{SB} = substrate bias voltage (negative w.r.t. source for nMOS, positive for pMOS)

q = 1.6×10^{-19} coulomb

N = impurity concentration in the substrate (N_A or N_D as appropriate)

ϵ_{Si} = relative permittivity of silicon ≈ 11.7

n_i = intrinsic electron concentration ($1.6 \times 10^{10}/\text{cm}^3$ at 300°K)

k = Boltzmann's constant = 1.4×10^{-23} joule/°K

The *body effects* may also be taken into account since the substrate may be biased with respect to the source, as shown in Figure 2.3.

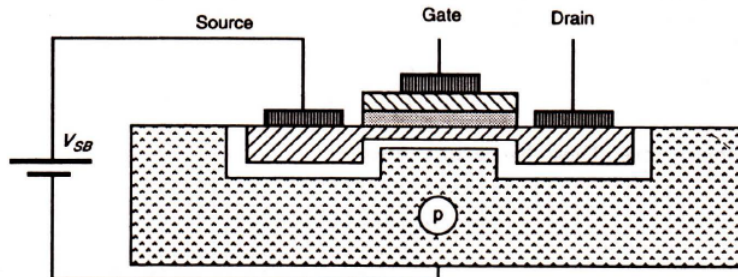


FIGURE 2.3 Body effect (nMOS device shown).

Increasing V_{SB} causes the channel to be depleted of charge carriers and thus the threshold voltage is raised.

Change in V_t is given by $\Delta V_t \approx \gamma(V_{SB})^{1/2}$ where γ is a constant which depends on substrate doping so that the more lightly doped the substrate, the smaller will be the body effect.

Alternatively, we may write

$$V_t = V_t(0) + \left(\frac{D}{\epsilon_{ins} \epsilon_0} \right) \sqrt{2 \epsilon_0 \epsilon_{Si} QN} \cdot (V_{SB})^{1/2}$$

where $V_t(0)$ is the threshold voltage for $V_{SB} = 0$.

- b. Determine the pull up to pull down ratio for an n MOS inverter driven through one or more pass transistors. (8)

Answer: (2 for each step 2x4)

Now consider the arrangement of Figure 2.9 in which the input to inverter 2 comes from the output of inverter 1 but passes through one or more nMOS transistors used as switches in series (called *pass transistors*).

We are concerned that connection of pass transistors in series will degrade the logic 1 level into inverter 2 so that the output will not be a proper logic 0 level. The critical condition is when point A is at 0 volts and B is thus at V_{DD} , but the voltage into inverter 2 at point C is now reduced from V_{DD} by the threshold voltage of the series pass transistor. With all pass transistor gates connected to V_{DD} (as shown in Figure 2.8), there is a loss of

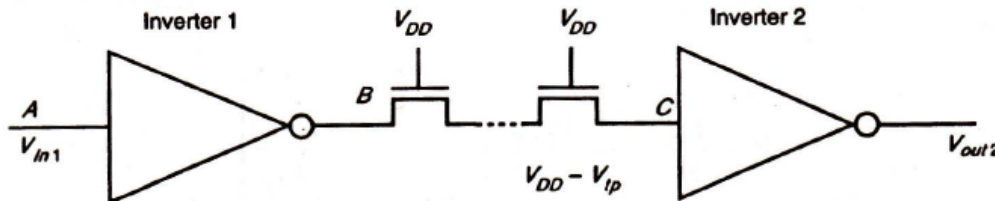


FIGURE 2.9 Pull-up to pull-down ratios for inverting logic coupled by pass transistors.

V_{tp} , however many are connected in series, since no static current flows through them and there can be no voltage drop in the channels. Therefore, the input voltage to inverter 2 is

$$V_{in2} = V_{DD} - V_{tp}$$

where

V_{tp} = threshold voltage for a pass transistor.

We must now ensure that for this input voltage we get out the same voltage as would be the case for inverter 1 driven with input = V_{DD} .

Consider inverter 1 (Figure 2.10(a)) with input = V_{DD} . If the input is at V_{DD} , then the p.d. transistor T_2 is conducting but with a low voltage across it; therefore, it is in its resistive region represented by R_1 in Figure 2.10. Meanwhile, the p.u. transistor T_1 is in saturation and is represented as a current source.

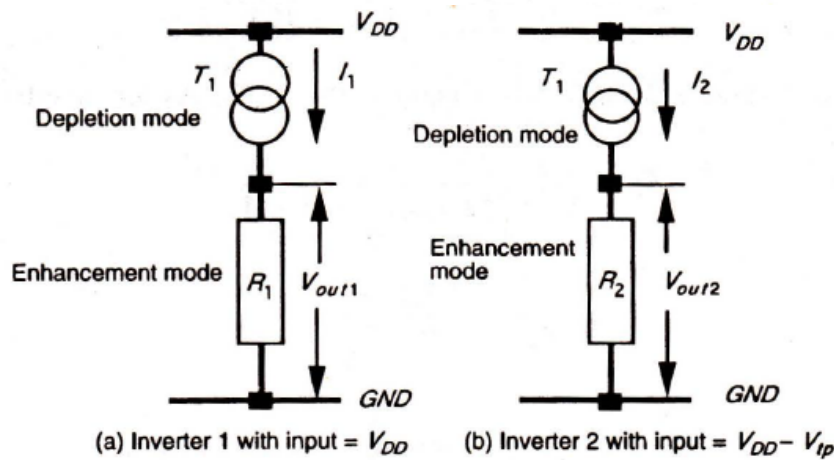


FIGURE 2.10 Equivalent circuits of inverters 1 and 2.

For the p.d. transistor

$$I_{ds} = K \frac{W_{p.d.1}}{L_{p.d.1}} \left((V_{DD} - V_t) V_{ds1} - \frac{V_{ds1}^2}{2} \right) \quad \text{(from 2.4)}$$

Therefore

$$R_1 = \frac{V_{ds1}}{I_{ds}} = \frac{1}{K} \frac{L_{p.d.1}}{W_{p.d.1}} \left(\frac{1}{V_{DD} - V_t - \frac{V_{ds1}}{2}} \right)$$

Note that V_{ds1} is small and $V_{ds1}/2$ may be ignored.

Thus

$$R_1 \approx \frac{1}{K} Z_{p.d.1} \left(\frac{1}{V_{DD} - V_t} \right)$$

Now, for depletion mode p.u. in saturation with $V_{gs} = 0$

$$I_1 = I_{ds} = K \frac{W_{p.u.1}}{L_{p.u.1}} \frac{(-V_{td})^2}{2} \quad (\text{from 2.5})$$

The product

$$I_1 R_1 = V_{out1}$$

Thus

$$V_{out1} = I_1 R_1 = \frac{Z_{p.d.1}}{Z_{p.u.1}} \left(\frac{1}{V_{DD} - V_t} \right) \frac{(V_{td})^2}{2}$$

Consider inverter 2 (Figure 2.10(b)) when input = $V_{DD} - V_{tp}$. As for inverter 1

$$R_2 \approx \frac{1}{K} Z_{p.d.2} \frac{1}{((V_{DD} - V_{tp}) - V_t)}$$

$$I_2 = K \frac{1}{Z_{p.u.2}} \frac{(-V_{td})^2}{2}$$

whence

$$V_{out2} = I_2 R_2 = \frac{Z_{p.d.2}}{Z_{p.u.2}} \left(\frac{1}{V_{DD} - V_{tp} - V_t} \right) \frac{(-V_{td})^2}{2}$$

If inverter 2 is to have the same output voltage under these conditions then $V_{out1} = V_{out2}$. That is

$$I_1 R_1 = I_2 R_2$$

Therefore

$$\frac{Z_{p.u.2}}{Z_{p.d.2}} = \frac{Z_{p.u.1}}{Z_{p.d.1}} \frac{(V_{DD} - V_t)}{(V_{DD} - V_{tp} - V_t)}$$

Taking typical values

$$V_t = 0.2V_{DD}$$

$$V_{tp} = 0.3V_{DD}^*$$

$$\frac{Z_{p.u.2}}{Z_{p.d.2}} = \frac{Z_{p.u.1}}{Z_{p.d.1}} \frac{0.8}{0.9}$$

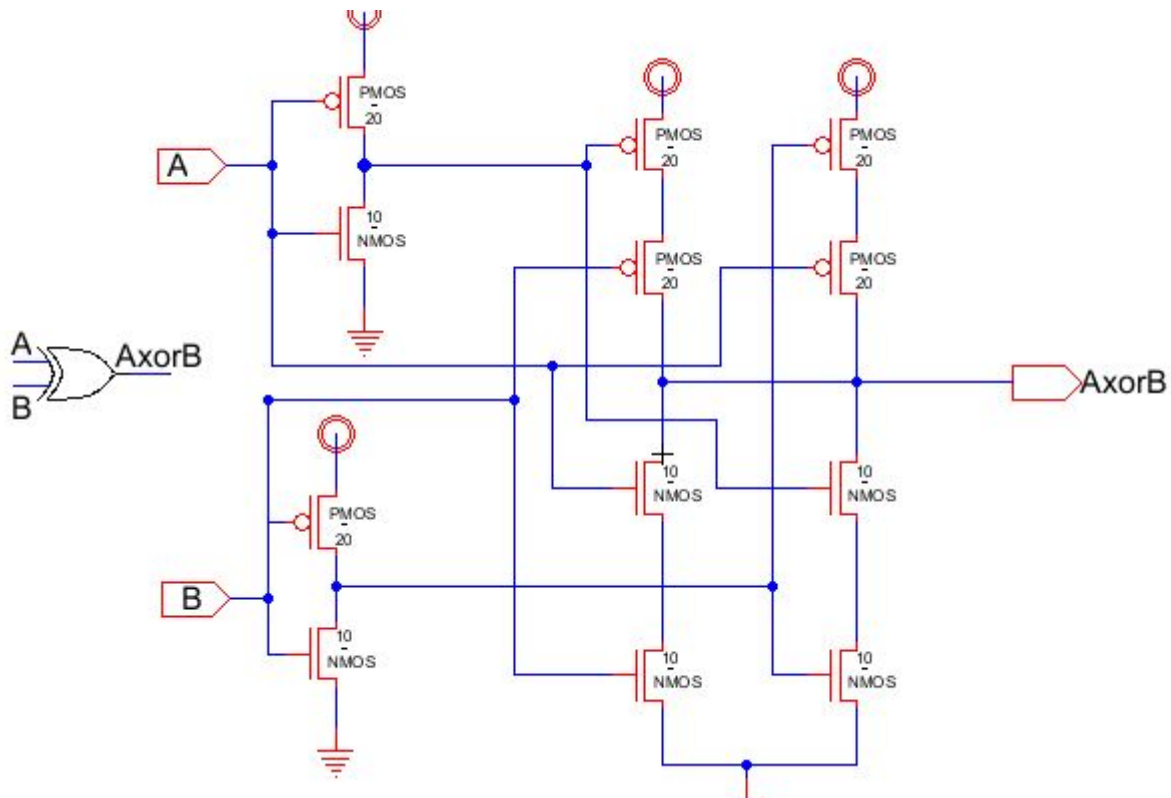
Therefore

$$\frac{Z_{p.u.2}}{Z_{p.d.2}} \doteq 2 \frac{Z_{p.u.1}}{Z_{p.d.1}} = \frac{8}{1}$$

Q.4 a. Draw the schematic for EX-OR gate using CMOS logic. (8)

Answer:

$$(A \text{ xor } B) = \overline{A \cdot B} + \overline{\overline{A} \cdot \overline{B}}$$



b. Explain Lambda (λ) based design rules. (8)

Answer:

(1 M for each rule)

The object of a set of design rules is to allow a ready translation of circuit design concepts, usually in stick diagram or symbolic form, into actual geometry in silicon. The design rules are the effective interface between the circuit/system designer and the fabrication engineer. Clearly, both sides of the interface have a vested interest in making their own particular tasks as easy as possible and design rules usually attempt to provide a workable and reliable compromise that is friendly to both sides.

Lambda-based Design Rules

In general, design rules and layout methodology based on the concept of λ provide a process and feature size-independent way of setting out mask dimensions to scale.

All paths in all layers will be dimensioned in λ units and subsequently λ can be allocated an appropriate value compatible with the feature size of the fabrication process. This concept means that the actual mask layout design takes little account of the value subsequently allocated to the feature size, but the design rules are such that, if correctly obeyed, the mask layouts will produce working circuits for a range of values allocated to λ . For example, λ can be allocated a value of $1.0 \mu\text{m}$ so that minimum feature size on chip will be $2 \mu\text{m}$ (2λ). Design rules, also due to Mead and Conway, specify line widths, separations, and extensions in terms of λ , and are readily committed to memory. Design rules can be conveniently set out in diagrammatic form as in Figure 3.6 for the widths and separation of conducting paths, and in Figure 3.7 for extensions and separations associated with transistor layouts.

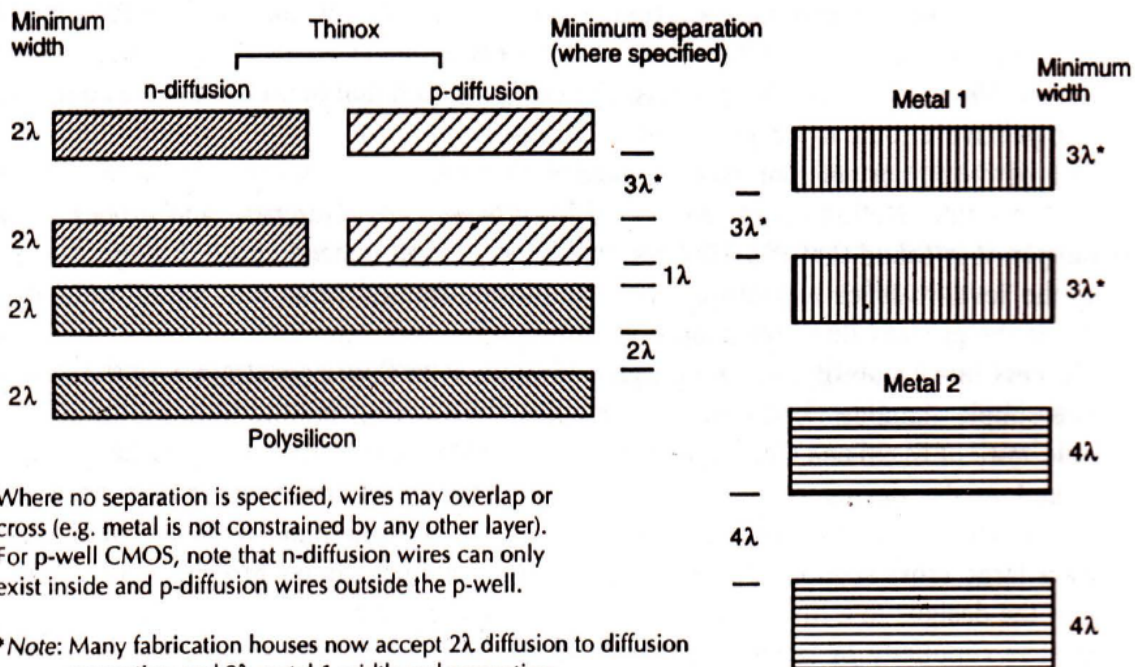


Figure 3-6 Design rules for wires (nMOS and CMOS)

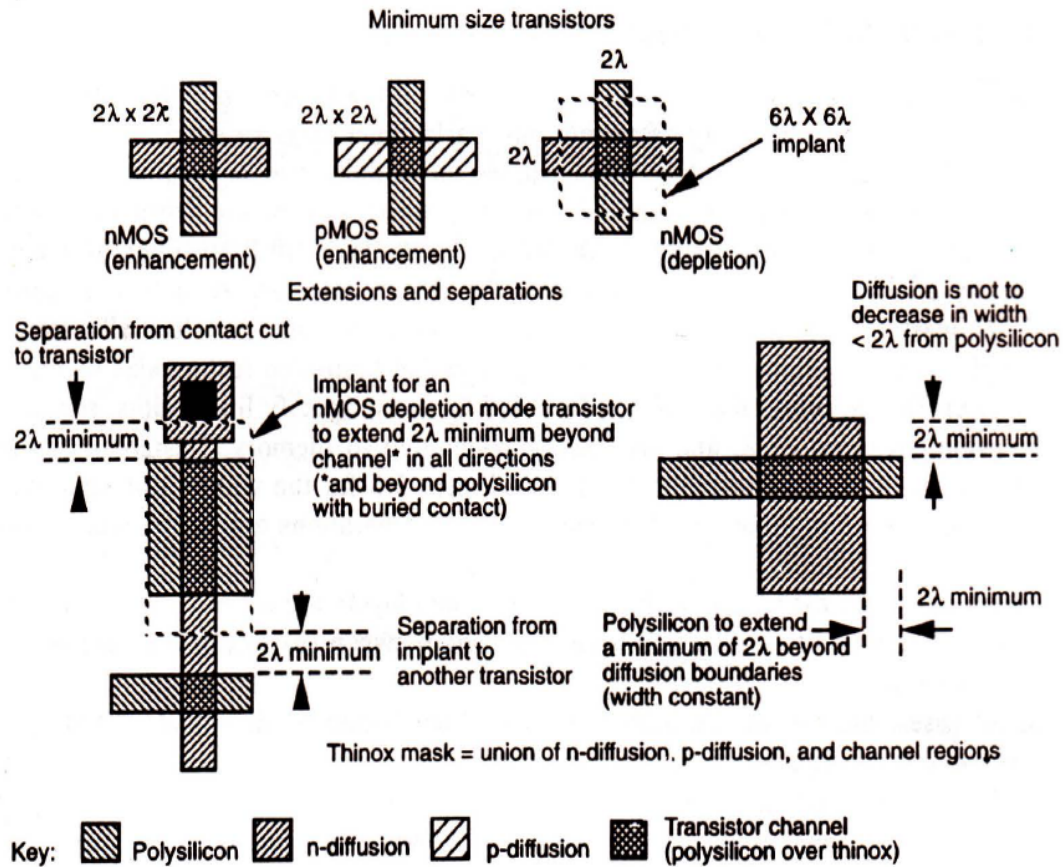


FIGURE 3.7 Transistor design rules (nMOS, pMOS and CMOS).

Q.5 a. Explain how to estimate wiring capacitances? (8)

Answer:

Three sources of wiring capacitances are

2 for each cap (2x4)

1. Fringing Fields:

Capacitance due to fringing field effects can be a major component of the overall capacitance of interconnect wires. For fine line metallization, the value of fringing field capacitance (C_{ff}) can be of the same order as that of the area capacitance. Thus, C_{ff} should be taken into account if accurate prediction of performance is needed.

$$C_{ff} = \epsilon_{SiO_2} \epsilon_0 l \left[\frac{\pi}{1n \left\{ 1 + \frac{2d}{t} (1 + \sqrt{1 + \frac{t}{d}}) \right\}} - \frac{t}{4d} \right]$$

where

l = wire length

t = thickness of wire

d = wire to substrate separation

Then, total wire capacitance

$$C_w = C_{area} + C_{ff}$$

2. Interlayer Capacitances:

Quite obviously the parallel plate effects are present between one layer and another. For example, some thought on the matter will confirm the fact that, for a given area, metal to polysilicon capacitance must be higher than metal to substrate. The reason for not taking such effects into account for simple calculations is that the effects occur only where layers cross or when one layer underlies another, and in consequence interlayer capacitance is highly dependent on layout. However, for regular structures it is readily calculated and contributes significantly to the accuracy of circuit modeling and delay calculation.

3. Peripheral Capacitance :

The source and drain n-diffusion regions (n-active regions for Orbit processes) form junctions with the p-substrate or p-well at well-defined and uniform depths; similarly for p-diffusion (p-active) regions in n-substrates or n-wells. For diffusion regions, each diode thus formed has associated with it a peripheral (side-wall) capacitance in picofarads per unit length which, in total, can be considerably greater than the area capacitance of the diffusion region to substrate; the smaller the source or drain area, the greater becomes the relative value of the peripheral capacitance.

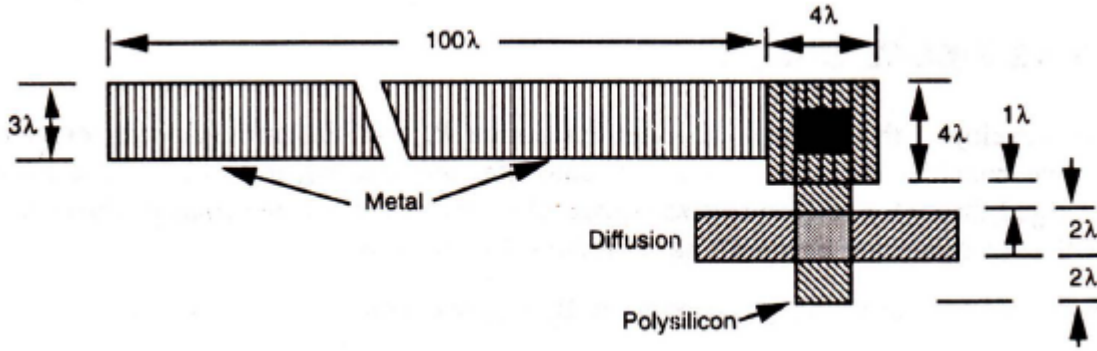
For Orbit processes, the n-active and p-active regions are formed by impurity implant at the surface of the silicon and thus, having negligible depth, they have negligible peripheral capacitance.

However, for n- and p-regions formed by a diffusion process, the peripheral capacitance is important and becomes particularly so as we shrink the device dimensions.

In order to calculate the total diffusion capacitance we must add the contributions of area and peripheral components

$$C_{total} = C_{area} + C_{periph}$$

- b. Calculate area capacitance values associated with structures occupying more than one layer as shown in figure given below (8)**



Typical area capacitance values for MOS circuits

Capacitance	Value in $pF \times 10^{-4}/\mu m^2$ (Relative values in brackets)		
	5 μm	2 μm	1.2 μm
Gate to channel	4 (1.0)	8 (1.0)	16 (1.0)
Diffusion (active)	1 (0.25)	1.75 (0.22)	3.75 (0.23)
Polysilicon* to substrate	0.4 (0.1)	0.6 (0.075)	0.6 (0.038)
Metal 1 to substrate	0.3 (0.075)	0.33 (0.04)	0.33 (0.02)
Metal 2 to substrate	0.2 (0.05)	0.17 (0.02)	0.17 (0.01)
Metal 2 to metal 1	0.4 (0.1)	0.5 (0.06)	0.5 (0.03)
Metal 2 to polysilicon	0.3 (0.075)	0.3 (0.038)	0.3 (0.018)

Notes: Relative value = specified value/gate to channel value for that technology.

*Poly. 1 and Poly. 2 are similar (also silicides where used).

Answer:

Consider the metal area (less the contact region where the metal is connected to polysilicon and shielded from the substrate)

$$\text{Ratio} = \frac{\text{Metal area}}{\text{Standard gate area}} = \frac{100\lambda \times 3\lambda}{4\lambda^2} = 75$$

$$\text{Metal capacitance } C_m = 75 \times 0.075 = 5.625 \square C_g$$

Consider the polysilicon area (excluding the gate region)

$$\text{Polysilicon area} = 4\lambda \times 4\lambda + 3\lambda \times 2\lambda = 22\lambda^2$$

Therefore

$$\text{Polysilicon capacitance } C_p = \frac{22}{4} \times 0.1 = .55 \square C_g$$

For the transistor,

$$\text{Gate capacitance } C_g = 1 \square C_g$$

Therefore

$$\text{Total capacitance } C_T = C_m + C_p + C_g \doteq 7.20 \square C_g$$

Q.6 a. What is short channel effect? Explain

(8)

Answer:

(2+6)

Short-Channel Devices

A MOSFET device is considered to be short when the channel length is the same order of magnitude as the depletion-layer widths (X_{dD} , X_{dS}) of the source and drain junction. As the channel length L is reduced to increase both the operation speed and the number of components per chip, the so-called short-channel effects arise.

Short-Channel Effects

The short-channel effects are attributed to two physical phenomena:

1. the limitation imposed on electron drift characteristics in the channel,
2. the modification of the threshold voltage due to the shortening channel length.

In particular five different short-channel effects can be distinguished:

1. drain-induced barrier lowering and punchthrough
2. surface scattering
3. velocity saturation
4. impact ionization
5. hot electrons

1. Drain-induced barrier lowering and punchthrough

When the depletion regions surrounding the drain extends to the source, so that the two depletion layer merge (i.e., when $x_{dS} + x_{dD} = L$), punchthrough occurs. Punchthrough can be minimized with thinner oxides, larger substrate doping, shallower junctions, and obviously with longer channels. The current flow in the channel depends on creating and sustaining an inversion layer on the surface. If the gate bias voltage is not sufficient to invert the surface ($V_{GS} < V_{T0}$), the carriers (electrons) in the channel face a potential barrier that blocks the flow. Increasing the gate voltage reduces this potential barrier and, eventually, allows the flow of carriers under the influence of the channel electric field. In small-geometry MOSFETs, the potential barrier is controlled by both the gate-to-source voltage V_{GS} and the drain-to-source voltage V_{DS} . If the drain voltage is increased, the potential barrier in the channel decreases, leading to drain-induced barrier lowering (DIBL). The reduction of the potential barrier eventually allows electron flow between the source and the drain, even if the gate-to-source voltage is lower than the threshold voltage. The channel current that flows under this conditions ($V_{GS} < V_{T0}$) is called the sub-threshold current.

2. **Surface scattering:** As the channel length becomes smaller due to the lateral extension of the depletion layer into the channel region, the longitudinal electric field component ϵ_y increases, and the surface mobility becomes field-dependent. Since the carrier transport in a MOSFET is confined within the narrow inversion layer, and the surface scattering (that is the collisions suffered by the electrons that are accelerated toward the interface by ϵ_x) causes reduction of the mobility, the electrons move with great difficulty parallel to the interface, so that the average surface mobility, even for small values of ϵ_y , is about half as much as that of the bulk mobility.
3. **Velocity saturation:** The performance short-channel devices is also affected by velocity saturation, which reduces the transconductance in the saturation mode. At low ϵ_y , the electron drift velocity v_{de} in the channel varies linearly with the

electric field intensity. However, as ϵ_y increases above 10^4 V/cm, the drift velocity tends to increase more slowly, and approaches a saturation value of $v_{de(sat)}=10^7$ cm/s around $\epsilon_y =10^5$ V/cm at 300 K. Note that the drain current is limited by velocity saturation instead of pinchoff. This occurs in short channel devices when the dimensions are scaled without lowering the bias voltages

4. **Impact ionization:** Another undesirable short-channel effect, especially in NMOS, occurs due to the high velocity of electrons in presence of high longitudinal fields that can generate electron-hole (e-h) pairs by impact ionization, that is, by impacting on silicon atoms and ionizing them. It happens as follow: normally, most of the electrons are attracted by the drain, while the holes enter the substrate to form part of the parasitic substrate current. Moreover, the region between the source and the drain can act like the base of an npn transistor, with the source playing the role of the emitter and the drain that of the collector. If the aforementioned holes are collected by the source, and the corresponding hole current creates a voltage drop in the substrate material of the order of .6V, the normally reversed-biased substrate-source pn junction will conduct appreciably. Then electrons can be injected from the source to the substrate, similar to the injection of electrons from the emitter to the base. They can gain enough energy as they travel toward the drain to create new eh pairs. The situation can worsen if some electrons generated due to high fields escape the drain field to travel into the substrate, thereby affecting other devices on a chip.
5. **Hot electrons:** Another problem, related to high electric fields, is caused by so-called hot electrons. This high energy electrons can enter the oxide, where they can be trapped, giving rise to oxide charging that can accumulate with time and degrade the device performance by increasing V_T and affect adversely the gate's control on the drain current.

b. Explain the difference between static Logic circuits and Dynamic Logic Circuits. (8)

Answer:

(1 M each diff)

Dynamic logic is temporary (transient) in that output levels will remain valid only for a certain period of time

- Static logic retains its output level as long as power is applied

Dynamic logic is normally done with charging and selectively discharging capacitance (i.e. capacitive circuit nodes) – Precharge clock to charge the capacitance – Evaluate clock to discharge the capacitance depending on condition of logic inputs

Advantages over static logic:

- Avoids duplicating logic twice as both N-tree and P-tree, as in standard CMOS
- Typically can be used in very high performance applications
- Very simple sequential memory circuits; amenable to synchronous logic
- High density achievable
- Consumes less power (in some cases)

Disadvantages compared to static logic:

- Problems with clock synchronization and timing

– Design is more difficult

Q.7 a. Design and explain 4 X 4 crossbar switch.

(8)

Answer:

(4+4 Fig.)

Any general purpose n-bit shifter should be able to shift incoming data by up to n - 1 places in a right-shift or left-shift direction. If we now further specify that all shifts should be on an 'end-around' basis, so that any bit shifted out at one end of a data word will be shifted in at the other end of the word, then the problem of right shift or left shift is greatly eased. In fact, a moment's consideration will reveal, for a 4-bit word, that a 1-bit shift right is equivalent to a 3-bit shift left and a 2-bit shift right is equivalent to a 2-bit shift left, etc. Thus we can achieve a capability to shift left or right by zero, one, two, or three places by designing a circuit which will shift right only (say) by zero, one, two, or three places. Data could be loaded from the output of the ALU and shifting effected; then the outputs of each stage of the shift register would provide the required parallel output to be returned to the register array (or elsewhere in the general case). However, there is danger in accepting the obvious without question. Many designers, used to the constraints of TTL, MSI, and SSI logic, would be conditioned to think in terms of such standard arrangements. When designing VLSI systems, it pays to set out exactly what is required to assess the best approach. In this case, the shifter must have:

- input from a four-line parallel data bus;
- four output lines for the shifted data;
- means of transferring input data to output lines with any shift from zero to three bits inclusive.

In looking for a way of meeting these requirements, we should also attempt to take best advantage of the technology; for example, the availability of the switch-like MOS pass transistor and transmission gate. We must also observe the strategy decided on earlier for the direction of data and control signal flow, and the approach adopted should make this feasible. Remember that the overall strategy in this case is for data to flow horizontally and control signals vertically. A solution which meets these requirements emerges from the days of switch and relay contact based switching networks-the crossbar switch. Consider a direct MOS switch implementation of a 4 x 4 crossbar switch, as in Figure

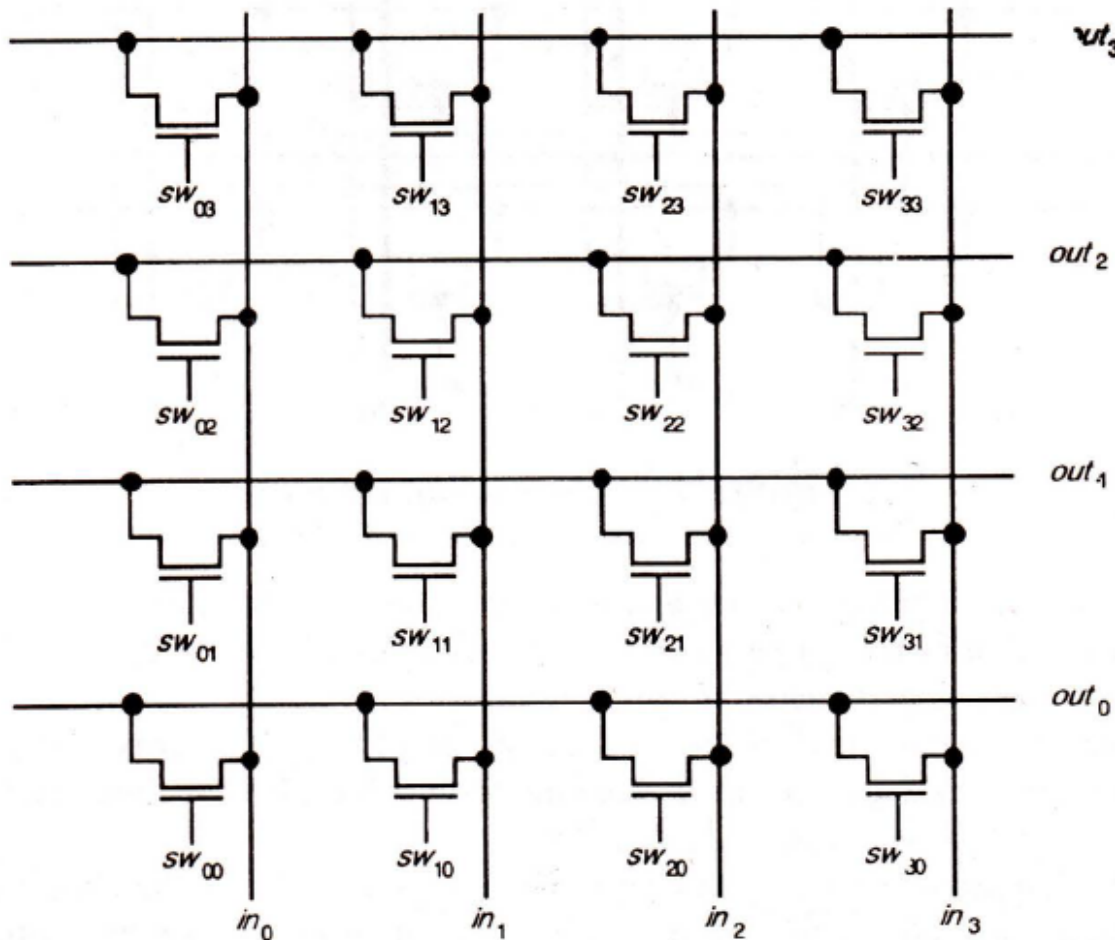


Figure: 4x4 crossbar switch

- b. Draw and explain a possible arrangement of the adder elements for both arithmetic and logical functions. (8)

Answer:

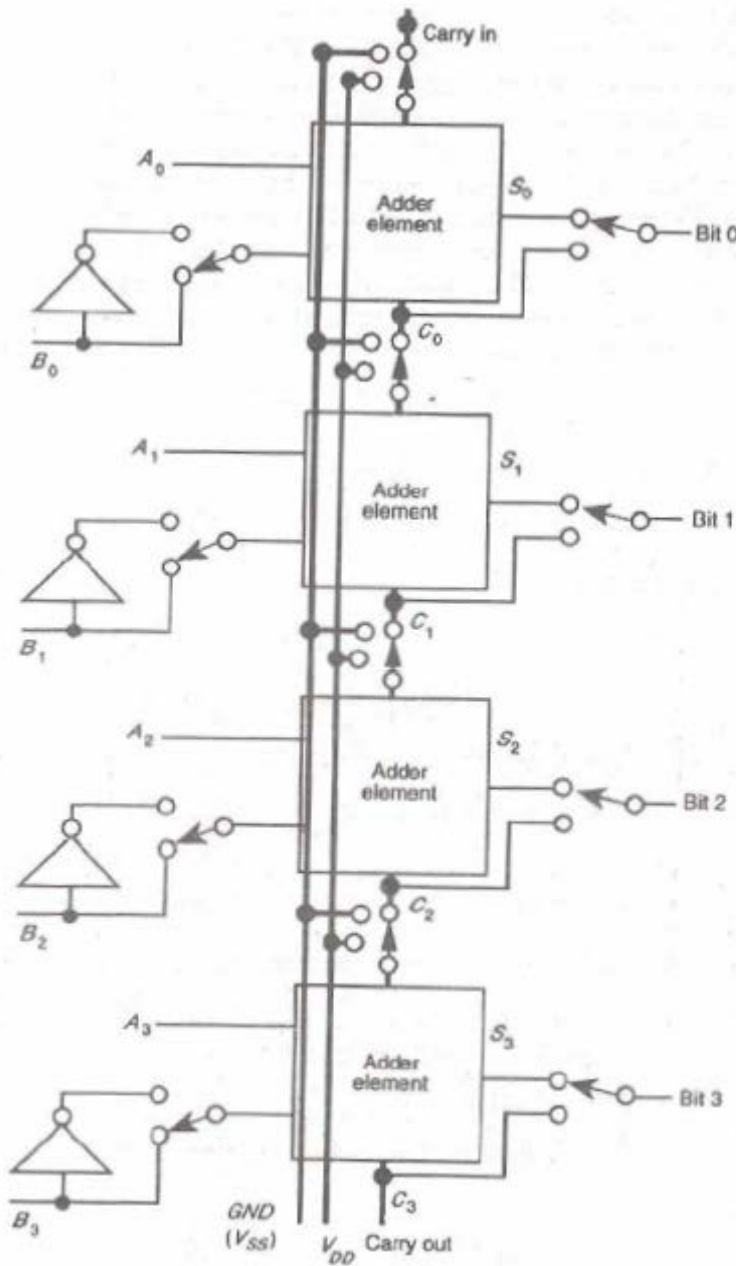


FIGURE 8.12 4-bit ALU.

Q.8 a. Briefly explain the operation of 6T SRAM memory structure with read and write circuitry. Explain their timing diagrams. (8)

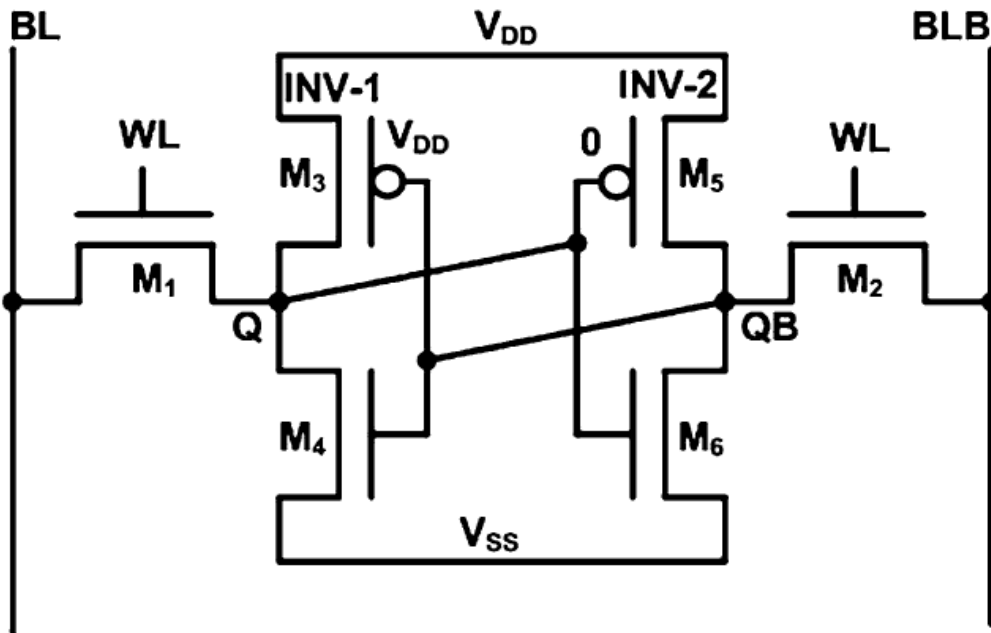
Answer:

(4 Fig. + 4 T. Diag)

Read Operation: It is assumed that the internal data storage nodes Q and QB are at '0' and '1'. To read the bitcell contents, the following sequence of steps are performed

- Conventionally to read a bitcell, the bitlines (BL and BLB) are precharged to the supply voltage (VDD). In some SRAM designs these bitlines are precharged to intermediate level of 0 and VDD.
- The wordline (WL) is asserted to high.

- Rise the WL from ‘0’ to ‘1’, result, one of the bitcell sides (node) stores the logical ‘0’; that side of the bitline is discharged through the pass-gate and pulldown transistors. In standard 6T, as shown in Fig, devices M1 and M4 discharges the precharged bitlines BL.
 - If BLB goes to low (or discharges), then the bitcell holds a logic ‘1’ value.
 - If BL goes to low (or discharges), then the bitcell holds a logic ‘0’ value.
- Depending upon whether the bitline BL or BLB is discharged, the bitcell is read as a logical ‘1’ or ‘0’. A sense amplifier converts the differential signal exists on BL and BLB to a logic-level output.
- De-assert the wordline (WL) back to 0.



Write operation: The write operation or flipping the bitcell contents when initially assuming that the internal data storage nodes Q and QB are at ‘0’ and ‘1’, respectively, as shown in Fig., consists of the following sequence of steps:

- Initially, wordline (WL) = 0.
- Precharge the bitlines (BL and BLB) to the supply voltage (VDD).
- After precharge, both the bitlines (BL and BLB) are disconnected from the supply voltage (VDD).
- Wordline (WL) is activated to high (data enters the bitcell during this step).
- Place the data value on the BL and the complementary data value on BLB.
- The bitline BLB connected to the data storage node QB via M2, is driven to the ground potential by a write driver through the M2 pass-gate transistor, while the BL is remained held at VDD to pull node Q to high via M1 pass-gate transistor.
- As node Q and QB flip their states de-assert the wordline (WL) back to 0.

b. Discuss the different ground rules for successful design. (8)

Answer: (1 M for each point 1x8)

GROUND RULES FOR SUCCESSFUL DESIGN

This section is intended to provide a convenient focus for design information. From our considerations of system design up to this point a number of ground rules, aspects of philosophy, and some basic data have emerged which help to ease the design process and ensure success.

1. The ratio rules

(a) for nMOS inverters and inverter-like stages

Zp.u.:Zpd. ratio = 4:1 when driven from another inverter

Zp.u.:Zpd. ratio= 8:1 when driven through one or more pass transistor(s)

where

$$Z = L/W \text{ for the channel in question}$$

(b) for CMOS, a 1: 1 ratio is normally used to minimize area, but for pseudo-nMOS inverters etc., a ratio Zp.u.:Zp.d. = 3:1 is required.

2. Design rules. Never bend the rules.

3. Inverter pair delay In general terms, the delay through a pair of similar nMOS inverters is

$$T_d = (1 + Z_{p.u./Z_{p.d.}})\tau$$

and for a minimum size CMOS complementary inverter pair

$$T_d = 7\tau$$

4. Cascaded inverters for driving capacitive load (C_L)

The approach is to use N cascaded inverters, each one of which is larger than the preceding stage by a width factor f.

It has been shown that the number 'N' of stages required is given by

$$N = \ln(y) / \ln(f)$$

Where

$$y = \frac{C_L}{\square C_g}$$

It can also be shown that total delay is minimized if f assumes the value e (base of natural logarithms); that is, each stage should be approximately 2.7* times wider than its predecessor. This applies to CMOS as well as nMOS inverters.

5. Propagation delay through cascaded pass transistors or transmission gates

$$T_d = n^2rc(\tau)$$

Where

n = number in series

r = relative series resistance per transistor or per transmission gate in terms of R_s

c = relative capacitance gate to channel per transistor or per transmission gate in terms of $\square C_g$.

Normally, no more than four pass transistors or transmission gates should be connected in series without buffering.

6. Subsystem/leaf-cell design guidelines

(a) Define the requirements properly and carefully.

(b) Consider communication paths most carefully in order to develop sensible placing of subsystems and leaf-cells.

(c) Draw a floor plan (alternating with (b) as necessary).

(d) Aim for regular structures so that design is largely a matter of replication.

- (e) Draw stick diagrams for basic cells, leaf-cells, and/or subsystems or enter the design in symbolic form.
 - (f) Convert to a mask level layout.
 - (g) Carefully and thoroughly check each mask layout for design rule errors and simulate circuit or logical operation. Correct as necessary, rechecking as corrections are made.
7. Restrictions associated with MOS pass transistors and transmission gates
 - (a) No more than four in series without buffering.
 - (b) No pass transistor gate must be driven from the output of one or more pass transistors, since logic 1 levels are degraded by threshold voltage V_{tp} (where V_{tp} can be as high as 0.3 VDD).
 - (c) When designing switch logic networks of pass transistors or transmission gates, care must be taken to deliberately implement both the logic 1 and logic 0 output conditions.
 8. Storage of logic levels on the gate capacitance of transistors
 - (a) Gate/channel capacitance is suitable for storing a bit, but care must be taken to allow for the finite decay time (about 0.25 msec at room temperature).
 - (b) It is quite allowable to construct pass transistors, etc. under metal layers to save space. This is often convenient and is used, for example, in some multiplexer layouts, but care must be taken with overlying metal wires where gate/channel capacitance is used for bit storage.
 - (c) Restrictions also apply to logic level storage on the input capacitance of a Nand gate except for the input nearest the GND or Vss rail.
 9. Enhanced clocking. One of the basic limitations on the use of simple MOS pass transistors (see Point 9 above) is the degradation of logic 1 levels by V_{tp} and the consequent inability of one pass transistor to drive the gate of a second (or more) pass transistor. This is particularly bothersome in clocking networks and a solution to this problem is to run all clock lines at a voltage level above VDD.
 10. The maximum allowable current density in aluminum wires is $1 \text{ mA}/\mu\text{m}^2$. Otherwise, metal migration may occur. Current density must be particularly carefully considered if the circuit is to be scaled down.
 11. Scaling effects.
 12. Set out rules of -system timing at an early stage in design.
 13. Avoid bus contentions by setting out bus utilization diagrams or tables, particularly in complex systems and/or where bidirectional buses are used.
 14. Do not take liberties with the design rules but do take account of the ground rules and guidelines.
 15. Remember, IC designers should expect their systems to function first time around* and this will happen if the design concepts are correct and if the rules are obeyed.

Q.9 a. Explain the aspects of design tools.

(8)

Answer:

ASPECTS OF DESIGN TOOLS

1. **Graphical Entry Layout:** Textual entry of layouts was at one time quite widely used and special textual entry editors are in existence and may well be used for small subsystem layout. However, such tools have been virtually swept aside by a much more convenient and highly interactive method of producing layouts for which monochrome or color graphics terminals are used, and on which the layout is built up and displayed during the

design process. Such systems are mostly 'menu driven', in that menus of possible actions at various stages of the design are displayed on the screen beside the display of the current layout detail. Some form of cursor allows selection and/or placement of geometric features, etc., and the cursor may also allow selection of menu items or, alternatively, these may also be selected from a keyboard. Positioning of the cursor may be effected from the keyboard in simple systems and/or cursor position may be controlled from a bitpad digitizer or from a 'mouse', etc.

2. **Design Verification Prior to Fabrication :** It is not enough to have good design tools for-producing mask and system layout detail. It is essential that such tools be complemented by equally effective verification software capable of handling large systems and with reasonable computing power requirements. The nature of the tools required will depend on the way in which an integrated circuit design is represented in the computer. Two basic approaches are:
 1. Mask level layout languages, such as CIF, which are well suited to physical layout description but not for capturing the design intent.
 2. Circuit description languages where the primitives are circuit elements such as transistors, wires, and nodes. In general, such languages capture the design intent but do not directly describe the physical layout associated with the design.
 By and large, therefore, the designer's needs may include the following.

3. Design Rule Checkers (DRC):

The cost in time and facilities in mask-making and in fabricating a chip from those masks is such that all possible errors must be eliminated before mask-making proceeds. Once a design has been turned h. to silicon there is little that can be done if it doesn't work. The wise designer will check for errors at all stages of the design, namely:

1. at the pencil and paper stage of the design of leaf-cells;
2. at the leaf-cell level once the layout is complete (e.g. when the CIF code for that leaf-cell has been generated);
3. at the su\.)system level to check that butting together and wiring up of leaf-cells is correctly done;
4. once the entire system layout has been completed.

The nature of physical layout verification 'design rule checking (DRC)' software may depend on whether the design rules are absolute or lambda-based, or on whether or not the layout is on a fixed or virtual grid.

4. Circuit Extractors

If design information exists in the form of physical layout data (as in CIF code form), then a circuit extractor program which will interpret the physical layout in circuit terms is required. Although the designer could use the extracted data to check against his or her design intent, it is normally fed directly into a simulator so that the computer may be used to interpret the findings of the extractor. (An example of a circuit extractor program is NET from Integrated Silicon Design Pty Ltd.)

5. Simulators

From mask layout detail it is possible to extract a circuit description in a form suitable for input to a simulator. Programs that do this are referred to as circuit extractors. The circuit description contains information about circuit components and their

interconnections. This information is subsequently transformed by the simulator into a set of equations from which the predictions of behaviour are made.

b. Explain different VLSI design styles. (8)

Answer:

When wishing to implement a system design in silicon, various approaches are possible and, of course, a wide range of technologies is available to choose from. The designer must choose an appropriate design style, but at *this* point it must be stressed that in no case will the choice of style hide the lack of a competent and systematic approach by the designer. However, we may summarize the possibilities into three broad categories:

1. Full custom design of the complete system for implementation in the chosen technology. In this case, the designer designs all the circuitry and all interconnection/communication paths.
2. Semi-custom design using a library of standard leaf-cells together with specially designed circuits and subsystems which are placed appropriately in the floor plan and interconnected to achieve the desired functional performance. In this case, the designer designs a limited amount of circuitry and the majority of interconnections/communications.
3. Gate array (uncommitted logic array) design in which standard logic elements are presented for the designer to interconnect to achieve the desired functional performance. In this case, the design is that of the interconnections and communications only.

Once again the boundaries between these categories may be blurred. For example, full custom design seldom involves the complete design of the entire chip; input/output pad circuits are more or less accepted as standard components and are generally available to the custom designer.

In all cases it is desirable to take a hierarchical approach to the system design in which the principles of iteration or replication (regularity) can be used to reduce the complexity of the design task.

The designer is usually concerned with a number of key design parameters. These will include:

1. performance, in terms of the function to be performed, the required speed of operation and the power dissipation of the system;

2. time' —for the design/development cycle;
3. testability;
4. the size of the die, which is determined by the area occupied by the circuitry and in turn has a marked impact on the likely yield in production and on the cost of bonding and packaging and testing. Large die sizes are generally associated with poor yields and high costs.

Full custom design tends to achieve the best results, but *only* if the designer is fully conversant with the fundamental aspects of design in silicon so that parameters can be optimized. However, full custom design parameter optimization is usually at the expense of parameter 2, the time taken to design.

Semi-custom and gate array designs both have penalties in area and often in speed and this is contributed to by the fact that not all the available logic will be used. This is due to the need for generality in gate array and standard cell geometries. However, it may often be the case that gate arrays will be faster than a prototype full custom design in, say, MPC form and the final custom designs must often be carefully optimized.

Once the approach is chosen, there remains the design philosophy which ranges through the following general possibilities.

1. *Hand-crafted design* in which, for example, the mask layouts are drawn on squared paper with layer encoding and are then digitized to give a machine-readable form of the mask detail. Digitization can be done 'by hand', with entry of coordinates through keyboard or by more direct digitization of the drawn layout using a digitizer pad and cursor.
2. *Computer-assisted textual entry* of mask detail through a keyboard using some specially developed language employing a text editing program. Such programs may have relatively low-level capabilities, allowing the entry of rectangular boxes, and 'wires', etc. only, or may be at a higher level and allow symbolic entry of circuit elements such as transistors and structures.
3. *Computer-assisted graphical* —of mask — through either a monochrome or color graphics terminal, again with the aid of the appropriate entry, display, and editing software.

In cases 2 and 3 the software usually aids the processes of hierarchical system design in that leaf-cells (or symbols) can be instanced many times, each instance being placed as appropriate in the floorplan. Subsystems thus created may themselves be repeatedly instanced and placed as required to build up the system hierarchy.

Such tools obviously encourage *regularity* and are generally used with a *generate then verify* design philosophy.

4. *Silicon compiler-based design* in which a high level approach is taken to design, and special languages, analogous to high level programming language compilers, are developed to allow the designer to specify the system requirements in a manner which is convenient and compact. The silicon compiler program then translates this input code into a mask design which will generate a circuit in silicon to meet the specified system requirements. Such programs are the subject of much research and

development work at this particular time. Indeed, the work has reached a stage at which silicon compilers have been in use for some time and there are textbooks on the subject (e.g. Ayres, 1983).

TEXT BOOK

- I. Basic VLSI Design, Douglas A. Pucknell and Kamran Eshraghian, PHI, 3rd Edition, 2007