1 (a) (i) For inverter A, nMOS, with resistive pull-up, when $V_i = 0$, V_O becomes V_{OH}. M1 does not conduct since no channel formed ($V_{GS} < V_T$). Hence the drain rises to V_{DD}. For inverter B, CMOS, when Vin = 0, VO once again rises to V_{OH}. M1 does not conduct. However, M2, being p-type, is fully conductive. Hence the output rises to V_{DD}.

(ii) For inverter A, when $V_i = V_{DD}$, V_O becomes V_{OL} for that technology. First assume that $V_{DS} = V_{OL}$ is of order 1V or less, so that M1 is in the non-saturated region, $V_{DS} < V_{GS} - V_T$. We can later confirm this assumption. Current in R_L and the drain of M1 are equal, and hence, assuming no current drawn from the output terminal,

$k_n = (W/L) \times k'_n = 4 \times 2 \times 10^{-5}\ AV^{-2} = 8 \times 10^{-5}\ AV^{-2}$

Using the non-saturated form of the equation:

$$\frac{k_1}{2}\left[2(V_{DD} - V_{T1})V_{OL} - V_{OL}^2\right] = \frac{V_{DD} - V_{OL}}{R_L}$$

$4 \times 10^{-5}\ (2 \times (3 - 1)\ V_{OL} - V_{OL}^2) = (3 - V_{OL})/10^5$ Multiply out

$16\ V_{OL} - 4\ V_{OL}^2 = 3 - V_{OL}$

$17\ V_{OL} - 4\ V_{OL}^2 - 3 = 0.$ Solving for V_{OL}

$$V_{OL} = \frac{-17 \pm \sqrt{289 - 48}}{-8} = \frac{-32.52}{-8} \text{ or } \frac{-1.48}{-8}.$$

Only the second root has physical significance; hence $V_{OL} = 0.185$ V

For inverter B, when Vi = V_{DD}, V_O becomes V_{OL}. M2 does not conduct since its $V_{SG2} <$ 1 V. M1 is fully conductive. Hence $V_{OL} = 0$ V.

(b) M2 must have the same ON conductance as M1 in order to equalise the delays arising from the load capacitances being driven. Since the same $|V_{GS}|$ is applied to both devices and they have the same $|V_T|$, it is only necessary to ensure that:

$$k_2 = k_1 \qquad \frac{W_2}{L_2}k'_p = \frac{W_1}{L_1}k'_n \quad \text{so that} \quad \frac{W_2}{L_2} = \frac{W_1}{L_1}\frac{k'_n}{k'_p}$$

Hence for this inverter design, substituting in this equation, the optimal aspect ratio for M2 is 8, as shown.

Where gates have >1 input, the networks of p and n transistors are more complex. In a NAND gate, series-connected transistors are found in the pull-down chain and paralleled transistor channels in the pull-up chain. The converse is so for NOR gates. Gates embodying both kinds of function will have both series and parallel-connected transistors in the pull-up and pull-down chain. As a result, it is no longer possible to identify a single set of device dimensions that will result in equal delays for rising and falling edges on any combination of inputs. Consequently, the procedure normally

followed is to consider the worst-case delays: where series devices are found, their effective resistances must be added; where the devices are in parallel, only one is assumed to be conductive. Typically, this approach will result in selection of device dimensions significantly larger than those needed in inverters.

(c) If the input switches abruptly, and if we may assume that the device itself responds on a timescale short in comparison with the RC delays, then the delay is given by the time taken for C_L to discharge from V_{OH} to $(V_{OH}+V_{OL})/2$. This corresponds to the 50% points customarily taken. Note that the values of V_{OH} and V_{OL} have to be determined for each circuit.

For inverter B, M2 is switched off abruptly by the falling input, so n-channel device M1 alone governs the delay. The initial and final values of V_O as the output rises in response to the input transition are 3.0 V and 1.50 V. In making the transition between these, M1 passes through both modes of MOSFET operation, saturation and non-saturation, according to the precise value of V_{DS} relative to $V_{GS} - V_T$. V_{GS} for M1 is fixed at 3 V throughout this.

M1 remains in saturation until $V_{DS} = 2$ V, i.e. until $V_O = 2$ V, and in this interval acts as a current source of magnitude:

$$I_{D1} = \frac{k}{2}(V_{GS} - V_T)^2 \;=\; \frac{1}{2}\times\frac{8}{1}\times 10^{-5}(3-1)^2 \;=\; 16\times 10^{-5}\ \text{A}$$

With this current, the time taken to discharge C_L from 3 V to 2 V, t_{sat}, is:

$$t_{sat} = \frac{C_L}{I_{D2}} \;=\; \frac{5\times 10^{-12}}{16\times 10^{-5}}\,\text{s} \;=\; 31\ \text{ns}$$

After V_O falls to 2 V or below, M2 enters its non-saturated region and its drain current thereafter depends on V_O

$$I_{D1} = \frac{k_1}{2}\left[2(V_{GS} - V_T)V_{DS} - V_{DS}^2\right] \qquad \text{Substituting}$$

$V_{GS} = 3$ V, $V_T = 1$ V, $V_{DS} = V_O$, we get

$$\frac{k_1}{2}\left(2(3-1)V_O - {V_O}^2\right) \;=\; \frac{k_1}{2}\left(4V_O - V_O^2\right)$$

For the falling transition, applying KCL at the output node gives: $\frac{dV_O}{dt} = -\frac{I}{C_L}$, where the – sign refers to a discharge of C_L; gathering terms and integrating we get $t_{non-sat}$:

$$t_{non-sat} \;=\; \int dt \;=\; -\frac{2C_L}{k_1}\int \frac{dV_O}{4V_O - {V_O}^2}$$

Simplifying with partial fractions and putting in the limits for V_O (2 and 1.5 V)

$$t_{non-sat} = \frac{2C_L}{k_1}\frac{1}{4}\int_2^{1.5}\left(\frac{1}{V_O}+\frac{1}{4-V_O}\right)dV_O = -\frac{2C_L}{k_1}\frac{1}{4}\left[\ln(V_O)-\ln(4-V_O)\right]_2^{1.5}$$

$$= \frac{2C_L}{k_1}\frac{1}{4}\left(\ln\frac{2}{2}-\ln\frac{1.5}{2.5}\right) = \frac{2C_L}{k_1}\frac{1}{4}\times 0.5108 = 0.312\times 10^{-7}\times 0.5108\,\text{s}$$

$$= 15.9\text{ ns}$$

The total delay, the sum of t_{sat} and $t_{non-sat}$, or 31 + 15.9 = 46.9 ns.

2.(a) (i) For MOS transistors saturation describes the region of operation where the channel is 'pinched-off' near the drain. This typically corresponds to higher values of V_{DS}. The necessary condition is (for n-channel)

$$V_{DS} > V_{GS} - V_T$$

[diagram showing characteristic I_D vs V_{DS} for various V_{GS}]

If $V_{DS} = V_{GS} - V_T$, pinch-off happens right at the drain. As V_{DS} increases, the pinch-off region advances progressively towards the source. To a first approximation, I_D remains unchanged as V_{DS} increases, and the device acts as a constant current source/sink. A MOSFET biased so that its gate is marginally above V_T and with high V_{DS} will therefore be in the saturation region. As V_{GS} is gradually increased for fixed V_{DS}, the device will first enter its saturation region, with very low I_D. If V_{DS} is now increased, to first order, I_D remains the same, and the pinch-off point simply moves slightly towards the drain; the increasing V_{DS} is reflected as a bigger potential difference across the pinch-off region. Hence the devices in a switching logic gate based on complementary MOS devices inevitably pass through saturation mode during the transition, but this is a transient phenomenon; the devices are in either the non-saturation (or linear) mode, or are cut-off, when the device output is at a regular logic level.

In practice, as V_{DS} increases, the resultant shortening of the channel (channel length modulation) slightly enhances the conductance and there is a gradual increase in I_D.

The near-constant current in this region is a useful characteristic in linear circuits. Also a FET connected with D & G shorted together is by definition in its saturation mode and may be used as a load in place of a resistor – it also occupies less space.

(ii) For bipolar devices saturation is observed when the base current exceeds that required to generate the corresponding collector current, having regard to the load resistance connected in series with the collector circuit, and the device's large-signal current gain. This can be brought about by raising the potential V_{BE} to cause a larger than usual collector current to flow. With a load connected in the collector circuit, V_{CE} falls to a low limiting value V_{CEsat}, typically 0.1 to 0.2 V in a silicon device, and is significantly less that the corresponding V_{BE} (typically about 0.7 V).

In this situation $V_{CB} < 0$ (NPN case) and the CB junction becomes forward-biased; the collector injects electronic charge into the base, accounting for some excess base current.

In bipolar circuits, saturation gives a clearly-defined low V_{CE} with a high V_{BE} and I_B, and a large I_E. The low V_{CE} can be used to define a logic level (as in saturated-mode logic). However, the excess charge stored in the base has to be removed before the circuit can leave saturation and stop conducting. This takes o(50ns) and limits switching speed available with saturated-mode logic.

(b) Applying KCL to the MOS circuit we see that while the device is in saturation:

$$I_D = \frac{k}{2}(V_{GS} - V_T)^2 = \frac{V_{DD} - V_{DS}}{R_L}$$

The device enters saturation as V_{GS} just rises above $V_T = 1$ V, at which point V_{DS} is still at 10V. It remains in saturation while $V_{IN} = V_{GS}$ rises provided $V_{DS} > V_{GS} - V_T$. At the moment where V_{DS} first falls below $V_{GS} - V_T$, the device leaves saturation. Hence we can write $V_{DS} = V_{GS} - V_T$ in the above. Solving for V_{DS}, we find

$$\frac{10 - V_{DS}}{10^4} = 4 \times 10^{-5} \times V_{DS}^2$$

Rearranging: $100 - 10V_{DS} = 4V_{DS}^2$ which has roots: $V_{DS} = -6.4$ V or $V_{DS} = +3.9$ V

Dismiss the first root as it has no physical significance. By inspection, the maximum value of V_{IN} attained before V_{DS} falls below $V_{GS} - V_T$ is 3.9 + 1 = 4.9 V.

Hence the range of V_{GS} values for saturation is +1.0 to +4.9 V

Similarly applying KCL to the bipolar circuit, $I_C = \frac{V_{CC} - V_{CEsat}}{R_C}$. At the onset of saturation, the large-signal current gain still applies, so $I_C = \beta I_B$. The base current is given by $I_B = \frac{V_{IN} - V_{BEsat}}{R_B}$. Note that this value of V_{IN} actually corresponds to V_{IH} for this style of logic. Hence the minimum V_{IN} to satisfy the saturation condition is:

$$V_{IN} = V_{BEsat} + \frac{R_B}{R_C}\frac{V_{CC} - V_{CEsat}}{\beta}$$

Substituting,

$$V_{IN} = 0.7 + \frac{5}{2}\left(\frac{5 - 0.2}{100}\right) = 0.82 \text{ V}$$

Note that any value of V_{IN} greater than or equal to this will produce saturation.

(c) This circuit is similar to the basic diode-transistor logic circuit that evolved to give TTL and its derivatives. However, a key difference is that the output uses a complementary pair, with an active pull-up based on a p-channel transistor.

The circuit is a 2-input NAND, that is, the output will be low if both inputs are high.

The circuit should have the advantage over a more conventional saturating bipolar design of being faster because of the high conductance path provided by T2; also, in the steady-state, the current in the output chain will be lower because one of T1 or T2 will be off while the gate output lies at a regular logic level. However, where good dynamic performance is sought, the relative slowness of switching in pnp bipolar devices will make it difficult to achieve a really fast gate.

3. (a) The terms 0, 2, 16, 18, 24, 26, 28, 30 are put into list 1 in order of how many variable of '1' they contain.

LIST 1

0	00000	√
2	00010	√
16	10000	√
18	10010	√
24	11000	√
26	11010	√
28	11100	√
30	11110	√

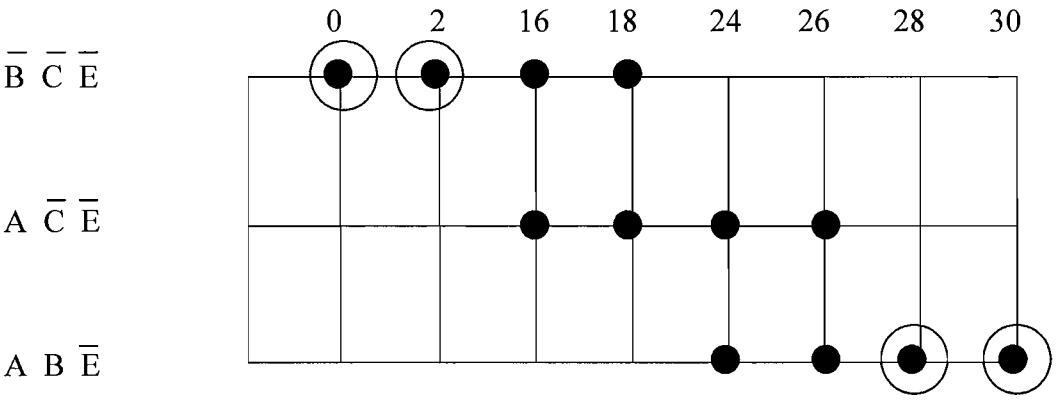
LIST2

0,2	000x0	√
0,16	x0000	√
2,18	x0010	√
16,18	100x0	√
16,24	1x000	√
18,26	1x010	√
24,26	110x0	√
24,28	11x00	√
26,30	11x10	√
28,30	111x0	√

LIST 3

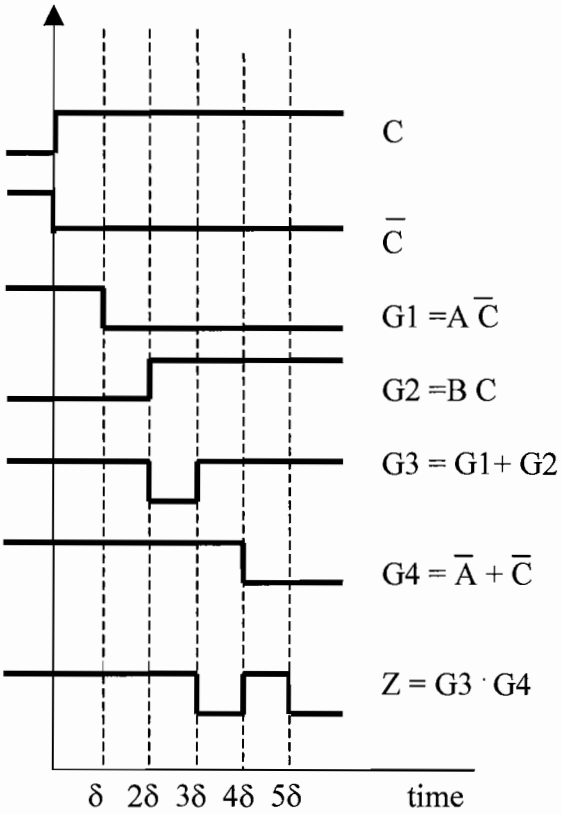
0,2,16,18 0,16,2,18	x00x0	$\overline{B}\ \overline{C}\ \overline{E}$
16,18,24,26 16,24,18,26	1x0x0	$A\ \overline{C}\ \overline{E}$
24,26,28,30 24,28,26,30	11xx0	$A\ B\ \overline{E}$

All the terms from lists 1 and 2 have combined according to $PQ + P\overline{Q} = P$. The list 3 contains all the Principal implicants (PIs) that can be put in a PI table.



Top and bottom PIs are essential. They also recognise all the original terms. So the simplest expression is $\overline{B}\ \overline{C}\ \overline{E} + A\ B\ \overline{E}$ [40%]

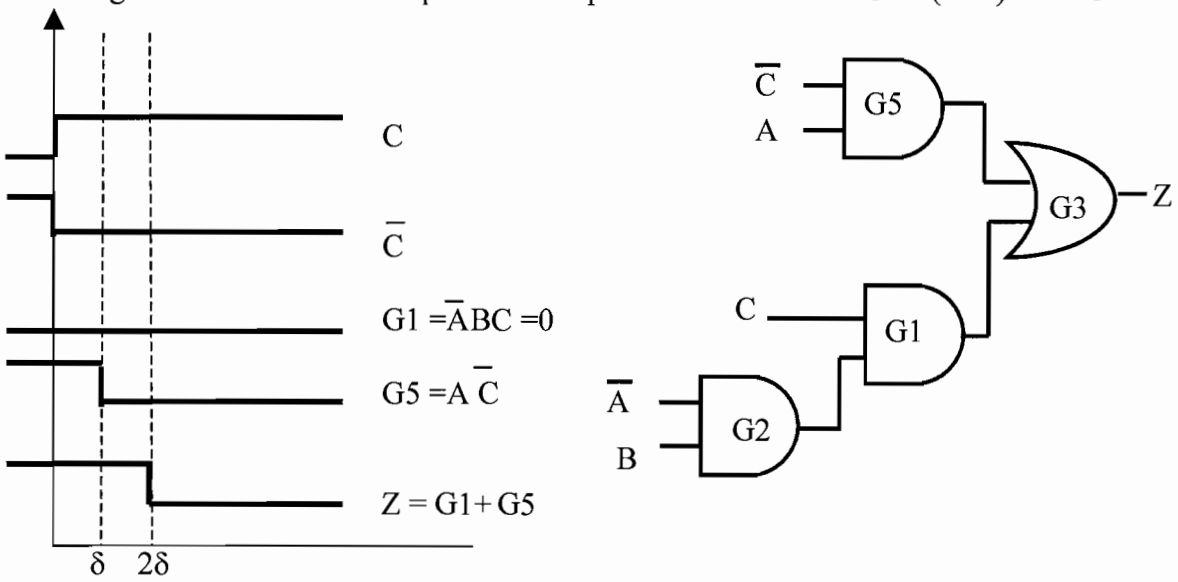
(b) The input C is used in three different paths. A dynamic hazard is possible.



There is a **dynamic** hazard as the output changes from 1 to 0 to 1 to 0 before it settles down after 5 δ.

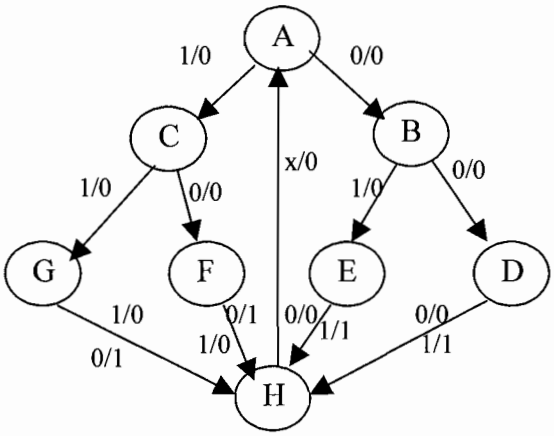
[30%]

The expression is Z = (AC' + BC) (A' +C') = AA'C' + A'BC + AC' + BCC'. The last term gives the hazard. The expression is equivalent to A'BC+ AC' = (A'B)C + AC'



The minimal delay in the output when C changes from 0 to 1 for A=B=1 can be obtained as 2δ. This is because G2 does not switch. For better overall speed (when A and B also change, one can swap G2 and G5 given a total delay of 3δ. From the time diagram one can see that the hazard is removed.

4. This is a sequential Mealy circuit (the output depends both on the present state and the input). We start from the reset state A. After each 3 bits, the circuit is to wait a clock pulse. This means that the third bit should send the circuit to a waiting state H.



PRESENT STATE & coding y1y2y3	NEXT STATE/OUTPUT for input x=0 and x=1	
	x=0	x=1
A (000)	B/0	C/0
B (001)	D/0	E/0
C (101)	F/0	G/0
D (110)	H/0	H/1
E (111)	H/0	H/1
F (010)	H/1	H/0
G (011)	H/1	H/0
H (100)	A/0	A/0

Decimal (ROM word line)	INPUT & PRESENT STATE & x y1y2y3		NEXT STATE $y1^+y2^+y3^+$ (= D1D2D3 inputs of the bistables)	OUTPUT Z
0	0	000	001	0
1	0	001	110	0
2	0	010	100	1
3	0	011	100	1
4	0	100	000	0
5	0	101	010	0
6	0	110	100	0
7	0	111	100	0
8	1	000	101	0
9	1	001	111	0
10	1	010	100	0
11	1	011	100	0
12	1	100	000	0
13	1	101	011	0
14	1	110	100	1
15	1	111	100	1

One can notice that D=E and F=G but that does not help the implementation of the Rom as we still need 3 bits for coding.

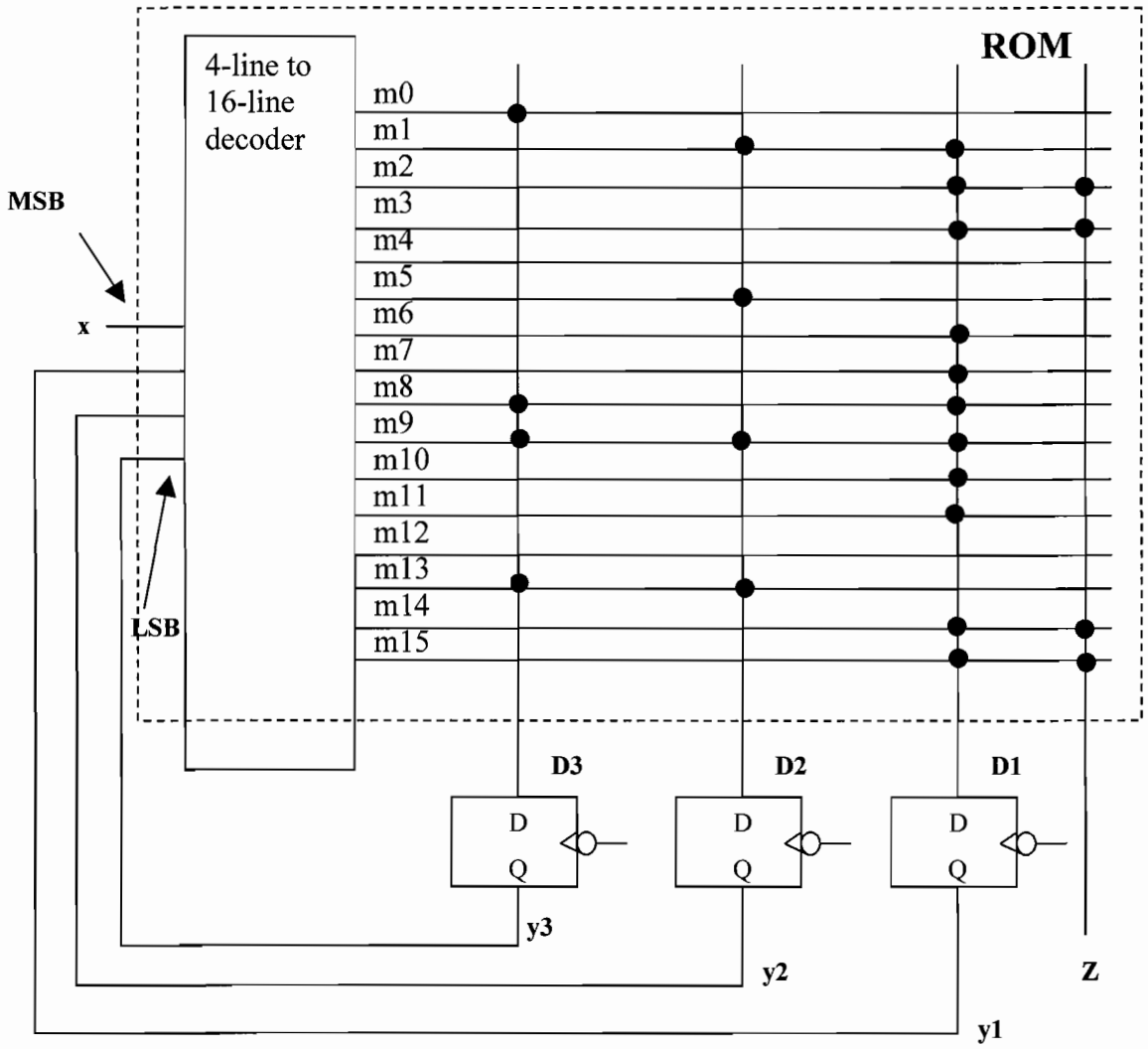
$D1 = y1^+ = \Sigma (1, 2, 3, 6,7,8,9,10,11,14,15)$

$D2 = y2^+ = \Sigma (1, 5, 9, 13)$

$D3 = y3^+ = \Sigma (0, 8, 9, 13)$

$Z = \Sigma (2, 3, 14, 15)$

[60%]



The size of the memory needed is 16 words X 4 output lines = 84 bits

[40%]

Numerical answers 3B2 –2009

1 (a) V_{OL} = 0.185 V for inverter A, V_{OL} = 0 V for inverter B

(c) $I_{D1} = 16\times10^{-5}$ A $\quad t_{sat} = \frac{C_L}{I_{D2}}$ = s = 31 ns $t_{non-sat}$ = 15.9 ns]

The total delay, the sum of t_{sat} and $t_{non-sat}$, or 31 + 15.9 = 46.9 ns

2. (b) V_{DS} = +3.9 V; the range of V_{GS} values for saturation is +1.0 to +4.9 V

V_{IN} = 0.82 V

3. (a) Simplified function $T = \overline{B}\,\overline{C}\,\overline{E} + A\,B\,\overline{E}$

(b) Z = A'BC+ AC'. The minimal delay in the output when C changes from 0 to 1 for A=B=1 can be obtained as 2δ. The total delay = 3δ

4. D1= $y1^{+}$= Σ (1, 2, 3, 6,7,8,9,10,11,14,15)

D2= $y2^{+}$= Σ (1, 5, 9, 13)

D3= $y3^{+}$= Σ (0, 8, 9, 13)

Z= Σ (2, 3, 14, 15)