

MODULE 5 - PART B

DELAY MODELS OF CMOS GATES AND INTERCONNECTS

In the last part of Module 5 (Part A), we had derived the expressions for the various types of Delays associated with CMOS inverters. This includes :-

① PROPAGATION DELAY →

a) t_{PLH} (low to high propagation delay)

This is the delay between 50% V_{DD} at the input (V_{IN}) and 50% V_{DD} at the output (V_{OUT}), when the output transitions from low to high.

$$t_{PLH} = R_p C_L \ln(2)$$

$$\text{or } t_{PLH} = 0.69 R_p C_L$$

b) t_{PHL} (high to low propagation delay)

This is the delay between 50% V_{DD} at the input (V_{IN}) and 50% V_{DD} at the output (V_{OUT}), when the output transitions from high to low.

$$t_{PHL} = R_n C_L \ln(2)$$

$$\text{or, } t_{PHL} = 0.69 R_n C_L$$

② RISE TIME →

Rise time or t_r refers to the time when the signal switches from 10% to 90% of its maximum value (V_{DD})

$$t_r = R_p C_L \ln(9) = 2.19 R_p C_L$$

③ FALL TIME →

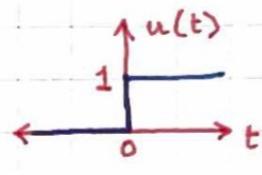
The Fall time or t_f refers to the time when the signal switches from 90% to 10% of its maximum value (V_{DD})

$$t_f = R_n C_L \ln(9) = 2.19 R_n C_L$$

NOTE :-

- ① We assume that V_{IN} is an ideal step function ($u(t)$)

$$u(t) = \begin{cases} 1 & \text{when } t > 0 \\ 0 & \text{when } t < 0 \end{cases}$$



- ② R_p and R_n are lumped resistances.

The expressions for these resistances are:-

- a) LINEAR REGION with LOW V_{DS} →

$$R_p = \frac{1}{\beta_p (V_{DD} - |V_{tp}|)}$$

$$R_n = \frac{1}{\beta_n (V_{DD} - V_{tn})}$$

- b) LINEAR REGION with HIGH V_{DS} →

$$R_p = \frac{1}{\beta_p [(V_{DD} - |V_{tp}|) - |V_{Dsp}|]}$$

$$R_n = \frac{1}{\beta_n [(V_{DD} - V_{tn}) - V_{Dsn}]}$$

- c) SATURATION REGION →

$$R_p = \frac{2 |V_{Dsp}|}{\beta_p (V_{DD} - |V_{tp}|)^2}$$

$$R_n = \frac{2 V_{Dsn}}{\beta_n (V_{DD} - V_{tn})^2}$$

For manual analysis using an RC model for delay, we will use the expressions for R_p and R_n in the LINEAR REGION with low V_{DS} , since these expressions are independent of V_{DS} which is not a constant but changes continuously during transition.

Alternatively, the delay of an inverter can also be written as :-

$$t_{PLH} = \frac{C_L \Delta V_{OUT}}{I_p} = \frac{C_L V_{DD}}{2 I_p}$$

Here, we want this equation of t_{PLH} to be independent of V_{OUT} and hence we will use the saturation region expression for current.

$$I_p = \frac{\beta_p}{2} (V_{DD} - |V_{tp}|)^2$$

($|V_{ds}| = V_{DD}$ when we get maximum current)

$$\therefore t_{PLH} = \frac{C_L V_{DD}}{\beta_p (V_{DD} - |V_{tp}|)^2}$$

Similarly,

$$t_{PHL} = \frac{C_L V_{DD}}{\beta_n (V_{DD} - V_{tn})^2}$$

Using $t_p = \frac{t_{PLH} + t_{PHL}}{2}$ and,

assuming $V_{tn} = |V_{tp}| = V_t$ and,

$V_{DD} \gg V_t$, we have,

$$t_p = \frac{C_L}{2 V_{DD}} \left[\frac{1}{\beta_p} + \frac{1}{\beta_n} \right]$$

NOTE :- Please note that we are using two types of DELAY MODELS.

① RC MODEL → For manual analysis using the RC MODEL, we need R to be independent of V_{OUT} since, V_{OUT} keeps changing and we cannot assign a constant value to V_{OUT} . We have seen in MODULE 3 that, for an nMOS, $V_{OUT} = V_{DS}$ and for a pMOS, $V_{OUT} = V_{DD} - |V_{DS}|$. Therefore, we need a resistance model that is independent of the V_{DS} of the transistor. (C however, is already a constant) We get this V_{DS} independent model in the LINEAR REGION where, $R_{\text{Linear}} = 1/\beta(V_{DD} - V_t)$ and V_{DS} is small.

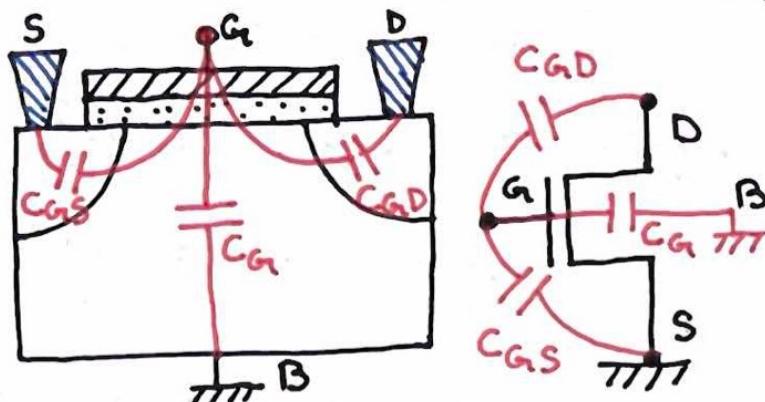
② CV/I MODEL → The second delay model that we use is where the delay is given by $\frac{CV_{DD}}{2I}$. Here, for manual analysis, the current 'I' needs to be independent of V_{OUT} or V_{DS} . This happens when the Transistor is in saturation since we know that the equation of the current here is given by $I_{\text{sat}} = \frac{\beta}{2}(V_{DD} - V_t)^2$.

Thus, we see that depending on the delay model that we are using, we will use that particular operating region of the transistor which will allow us to have an analytical model of delay where all the parameters are constant and hence can be computed easily.

We will now try to MODEL this load capacitance C_L .

CAPACITANCE MODEL of a TRANSISTOR

Let us look a little closely at the capacitances associated with an nMOS.



First let us look at the various components of capacitance looking into the gate of a transistor. These are :-

① GATE - to - BODY Capacitance (C_{GB}) →

This is the capacitance formed by the gate on one side and the channel or Body on the other side and the thin oxide in between.

② GATE - to - SOURCE Capacitance (C_{GS}) →

Since the edge of the gate and the edge of the source are not perfectly aligned, there is an overlap region between the gate and the source. This capacitance formed by the gate on one side, source on the other, and the thin oxide in between is called the Gate-to-Source Capacitance C_{GS} .

③ GATE - to - DRAIN Capacitance (C_{GD}) →

Since the edge of the gate and the edge of the drain are not perfectly aligned, there is an overlap region between the gate and the drain. This capacitance formed by the gate on one side, drain on the other and the thin oxide in between is called the Gate - to - Drain Capacitance C_{GD} .

NOTE: ① C_{GS} and C_{GD} are both much smaller than C_G . C_{GS} , C_{GD} are approximately 10% to 15% of C_G .

② Here, C_G is responsible for the transistor action (generating high ON current). This is because C_G is the same as C_{ox} . C_G is connected to GND on one side before the channel forms and is connected to the channel after the channel forms.
 $\therefore C_G = C_{ox}$ (when channel forms)

③ C_{GD} and C_{GS} are not desired as they do not contribute to the current but are inevitable in the manufacturing process. These are hence called parasitic capacitances.

Before we proceed with our analysis, we will need some background knowledge from ECE 2040.

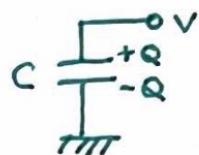
We have so far studied capacitances as DC elements and defined $C = \frac{Q}{V}$

However, when we are studying switching circuits, we are interested in the ac component of capacitance which represents the change in CHARGE in response to a change in VOLTAGE across it.

i.e.

$$C_{ac} = \frac{\Delta Q}{\Delta V}$$

Let us assume that we have a circuit as shown below :-



When V voltage is applied across the capacitor C , the charge across the capacitance is given by, $Q = CV$.

Now, if we increase V to $V + \Delta V$,

the charge across C , increases to

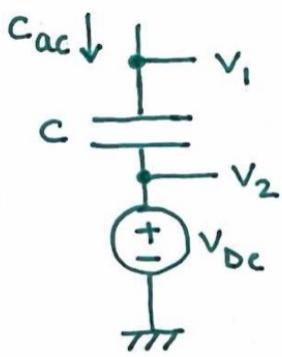
$$Q + \Delta Q = C(V + \Delta V) = CV + C\Delta V$$

$$\therefore \Delta Q = C\Delta V$$

\therefore The EFFECTIVE CAPACITANCE looking in, say,

$$C_{ac} = \frac{\Delta Q}{\Delta V} = \frac{C\Delta V}{\Delta V} = C$$

Now, let us assume a modification of this circuit where one terminal is connected to a DC voltage source (V_{DC}) and we want to find C_{ac}



Here, V_2 is always fixed at V_{DC} . Let us assume that V_1 is initially at V voltage and then it increases to

$$(V + \Delta V) \quad \text{i.e. } V_1(\text{initial}) = V$$

$$V_1(\text{final}) = V + \Delta V$$

$$\therefore Q_{(\text{initial})} = C(V_1 - V_2) = C(V - V_{DC})$$

$$Q_{(\text{final})} = C(V + \Delta V - V_{DC})$$

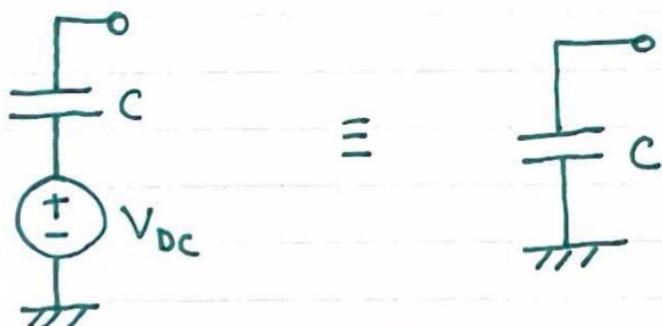
$$\therefore \Delta Q = Q_{(\text{final})} - Q_{(\text{initial})}$$

$$\therefore \Delta Q = C(V + \Delta V - V_{DC}) - C(V - V_{DC})$$

$$\text{or, } \Delta Q = C \Delta V$$

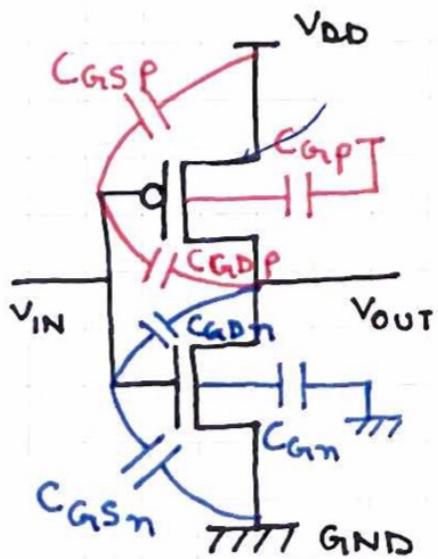
$$\therefore C_{ac} = \frac{\Delta Q}{\Delta V} = C$$

Thus, we see that the DC source has no effect on the ac component of the capacitance. In other words, we can treat a DC voltage source in series with a capacitor as a SHORT when we are interested in the ac component.



NOTE: This is true for LINEAR CAPACITANCE only i.e. when the value of the capacitance is independent of the voltage across it and is a CONSTANT value

INPUT CAPACITANCE of an INVERTER



As we look into the input node (V_{IN}) of an inverter, we observe multiple capacitors contributed by the nMOS and pMOS transistors as we have seen before.

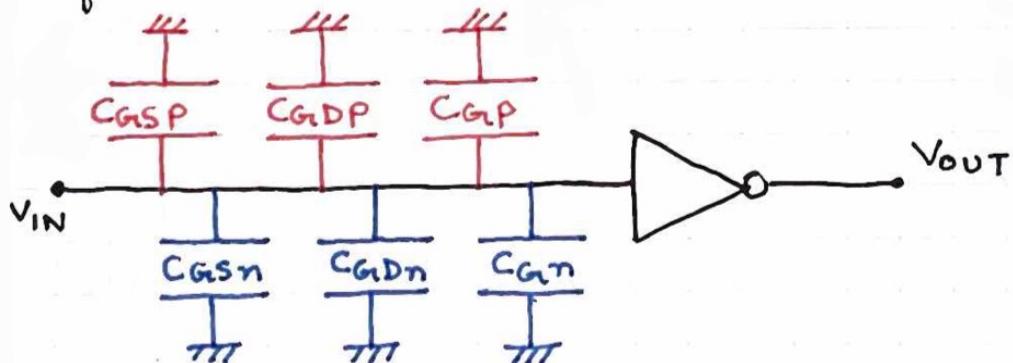
C_{GSN} and C_{GSN} appear between V_{IN} and GND.

C_{GSp} and C_{Gp} appear between V_{IN} and V_{DD} . However, we have seen that in ac analysis, we can consider these capacitors between V_{IN} and GND since the DC source V_{DD} can be considered as a short.

C_{GDp} and C_{GDn} appear between V_{IN} and V_{OUT} and since these are both switching nodes, the analysis is very complex and is beyond the scope of this course. These are called MILLER capacitances and you will learn more about these in advanced courses.

For now, we will assume that C_{GDp} and C_{GDn} also appear between V_{IN} and GND.

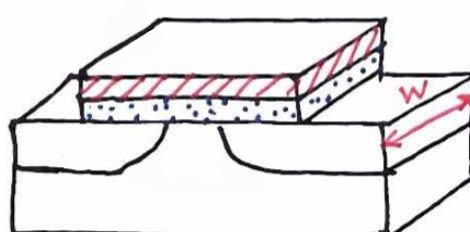
Hence, the total capacitance at the INPUT of the inverter is :-



Since, these capacitances are in parallel, we can write,

$$C_{IN} = C_{GSp} + C_{GDp} + C_{Gp} + \\ C_{GSn} + C_{GDn} + C_{Gn}$$

NOTE :-



All of these capacitors are proportional to the widths (W) of the corresponding nMOS and pMOS transistors.

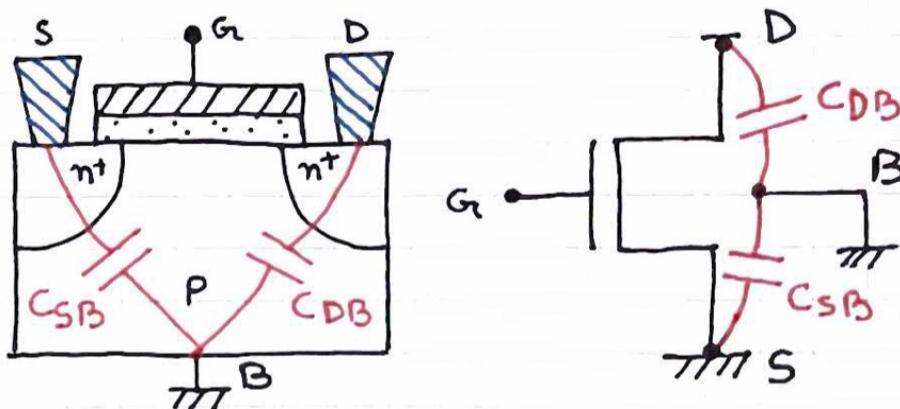
OUTPUT CAPACITANCE of a MOSFET

Previously, we had looked at the input or gate capacitance of a MOSFET. Now we will look at the output capacitance.

The Source and the Drain of the MOSFET also have capacitances to the Body terminal.

This happens, because there are p-n junctions (reverse biased) between the S and D and the Body.

We have seen this in Module 2
These are also called JUNCTION CAPACITORS.



Here the capacitances are :-

- ① DRAIN - to - BODY JUNCTION CAPACITANCE (C_{DB}) →

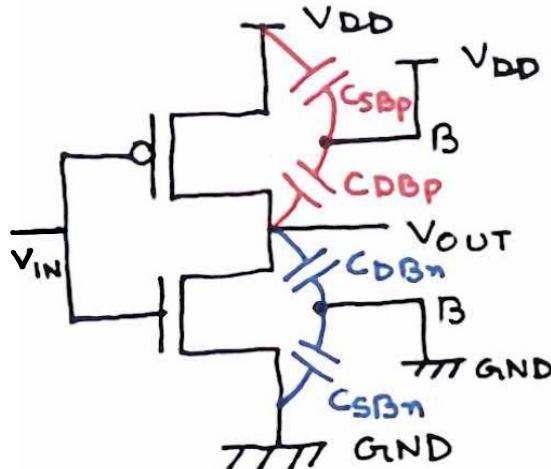
This is the capacitance formed by the p-n junction between the Drain and the Body.

- ② SOURCE - To - BODY JUNCTION CAPACITANCE (C_{SB}) →

This is the capacitance formed by the p-n junction between the Source and the Body.

NOTE:- Both of these capacitors are parasitic capacitors.

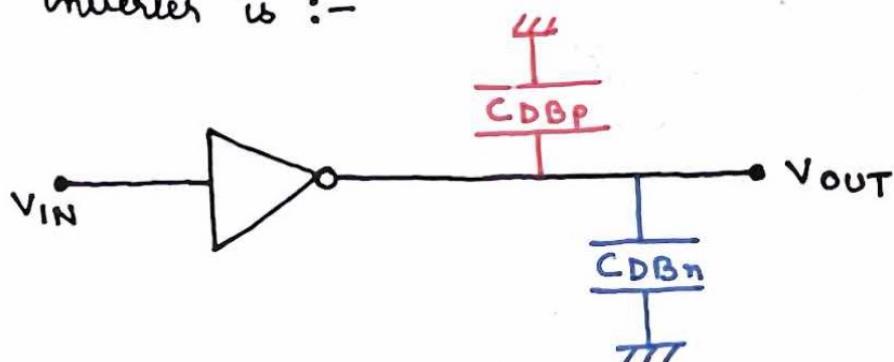
OUTPUT CAPACITANCE of an INVERTER



Now we will look at the total capacitance at the output (V_{OUT}) of the INVERTER.

C_{DBn} appears between V_{OUT} and ground (GND). C_{DBp} appears between V_{OUT} and V_{DD} . Since, V_{DD} is a DC source, we have seen that C_{DBp} can be considered to be between V_{OUT} and GND.

C_{SBn} and C_{SPB} are not connected to V_{OUT} and therefore do not contribute to the output capacitance. Hence, the total capacitance at the OUTPUT of the inverter is :-



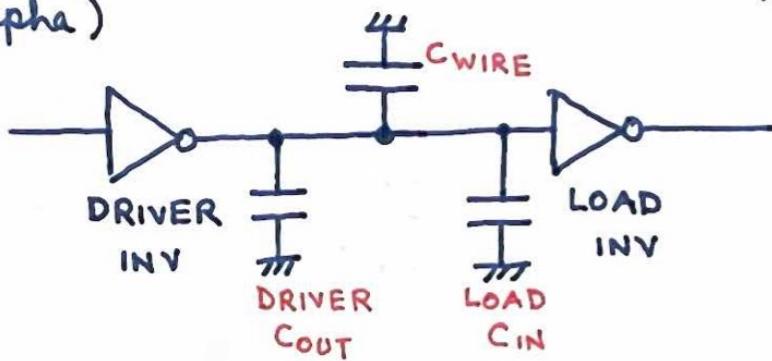
Since, these capacitances are in parallel, we can write,

$$C_{OUT} = C_{DBp} + C_{DBn}$$

NOTE : These capacitances too, like the input capacitances are proportional to the corresponding nMOS and pMOS widths.

EFFECT OF TRANSISTOR SIZING ON DELAY

Let us consider an inverter driving another inverter and let us assume that the inverters have been sized as $\frac{W_P}{W_n} = \alpha$ (alpha)



The total capacitance driven by the DRIVER INVERTER, is :-

$$C_L = C_{OUT} + C_{WIRE} + C_{IN}$$

where, C_{OUT} → the total output capacitance of the DRIVER INV

C_{WIRE} → parasitic capacitance from the interconnect wire

C_{IN} → the total input capacitance of the LOAD INV.

Now, let us assume :-

① WIDTH of the nMOS in the DRIVER INV
→ W_d

② WIDTH of the pMOS in the DRIVER INV
→ αW_d

③ WIDTH of the nMOS in the LOAD INV
→ W_l

④ WIDTH of the pMOS in the LOAD INV
 $\rightarrow \alpha W_L$

$$\therefore C_{OUT} = C_{DBp} + C_{DBn}$$

$$\text{or, } C_{OUT} = \alpha W_d \tilde{C}_{DBp} + W_d \tilde{C}_{DBn}$$

where, \tilde{C}_{DBp} and \tilde{C}_{DBn} are the DRAIN-TO-BODY Capacitance per unit width of the DRIVER pMOS and nMOS respectively.

$$\text{or, } C_{OUT} = W_d (\tilde{C}_{OUT})$$

$$\text{where, } \tilde{C}_{OUT} = \alpha \tilde{C}_{DBp} + \tilde{C}_{DBn}$$

similarly, we can write,

$$C_{IN} = C_{GSp} + C_{GDp} + C_{Gp} + \\ C_{GSn} + C_{GDN} + C_{Gn}$$

$$= \alpha W_L (\tilde{C}_{GSp} + \tilde{C}_{GDp} + \tilde{C}_{Gp}) + \\ W_L (\tilde{C}_{GSn} + \tilde{C}_{GDN} + \tilde{C}_{Gn})$$

where, \tilde{C}_{GSp} , \tilde{C}_{GDp} , \tilde{C}_{Gp} , \tilde{C}_{GSn} , \tilde{C}_{GDN} and \tilde{C}_{Gn} are capacitances per unit width of the LOAD pMOS and nMOS.

We have seen in Module 5 - Part A that a simplified form of the DELAY equation is given by,

$$t_p = \frac{C_L}{2V_{DD}} \left[\frac{1}{\beta_p} + \frac{1}{\beta_n} \right]$$

This is the PROPAGATION DELAY for the DRIVER INVERTER.

Here,

$$\beta_p = \frac{\alpha W_d}{L} \mu_p C_{ox} \quad \text{and,}$$

$$\beta_n = \frac{W_d}{L} \mu_n C_{ox}$$

(we assume that C_{ox} and L are the same for all the transistors in the DRIVER and LOAD inverters)

$$\therefore \frac{1}{\beta_p} + \frac{1}{\beta_n} = \frac{1}{W_d} \left[\frac{L}{\alpha \mu_p C_{ox}} + \frac{L}{\mu_n C_{ox}} \right]$$

$$\text{or, } \frac{1}{\beta_p} + \frac{1}{\beta_n} = \frac{1}{W_d} \left[\frac{1}{\tilde{\beta}} \right]$$

where,

$$\frac{1}{\tilde{\beta}} = \frac{L}{\alpha \mu_p C_{ox}} + \frac{L}{\mu_n C_{ox}}$$

We can also write,

$$C_L = C_{OUT} + C_{WIRE} + C_{IN}$$

$$\text{or, } C_L = W_d \tilde{C}_{OUT} + C_{WIRE} + \\ W_L \left[\alpha (\tilde{C}_{GSP} + \tilde{C}_{GDP} + \tilde{C}_{GP}) + \right. \\ \left. (\tilde{C}_{GSn} + \tilde{C}_{GDN} + \tilde{C}_{GN}) \right]$$

$$\text{or, } C_L = W_d \tilde{C}_{OUT} + C_{WIRE} + W_L \tilde{C}_{IN}$$

$$\text{where, } \tilde{C}_{IN} = \alpha (\tilde{C}_{GSP} + \tilde{C}_{GDP} + \tilde{C}_{GP}) + \\ (\tilde{C}_{GSn} + \tilde{C}_{GDN} + \tilde{C}_{GN})$$

Hence, we can write the propagation delay, t_p , as :-

$$t_p = \frac{C_L}{2V_{DD}} \left[\frac{1}{\beta_p} + \frac{1}{\beta_n} \right]$$

$$= \frac{W_d \tilde{C}_{out} + C_{WIRE} + W_L \tilde{C}_{in}}{2V_{DD}} \left[\frac{1}{W_d \tilde{\beta}} \right]$$

$$= \frac{\tilde{C}_{out}}{2V_{DD}\tilde{\beta}} + \frac{C_{WIRE} + W_L \tilde{C}_{in}}{2V_{DD}W_d \tilde{\beta}}$$

$$t_p = \frac{\tilde{C}_{out}}{2V_{DD}\tilde{\beta}} + \frac{C_{WIRE} + W_L \tilde{C}_{in}}{2V_{DD}W_d \tilde{\beta}} \quad \dots(i)$$