
MODULE 3

# Switching Characteristics and Interconnect Effects

Dr. Rashmi Seethur
Assistant Professor
Department of Electronics & Communication Engineering

# Contents:

| | |
|---|---|
| R1: Chap 6 (6.1 -6.7) | Introduction, Delay-Time Definitions (6.1,6.2) |
| | Calculation of Delay Times (6.3) |
| | Calculation of Delay Times (6.4) |
| | Estimation of Interconnect Parasitic (6.5) |
| | Calculation of Interconnect Delay (6.6) |
| | Switching Power Dissipation of CMOS Inverters (6.7) |

# CMOS Switching Characteristics

- **CMOS Switching Characteristics**

  - we studied the DC (or Static) characteristics of the CMOS inverter

  - we learned how to calculate: $V_{OH}$, $V_{OL}$, $V_{IL}$, $V_{IH}$, $V_{th}$, $NM_L$, $NM_H$,

  - we learned that we can modify some of these parameters using the W/L ratios of the inverter

  - specifically, we say that the $V_{th}$ is solely dependant on W/L and is usually the most important and most commonly controlled parameter

  - we now turn to the Switching (or AC or Dynamic) behavior of the inverter

  - the switching characteristics give us how *fast* the circuit will run

  - when designing, we must meet both DC and AC specs

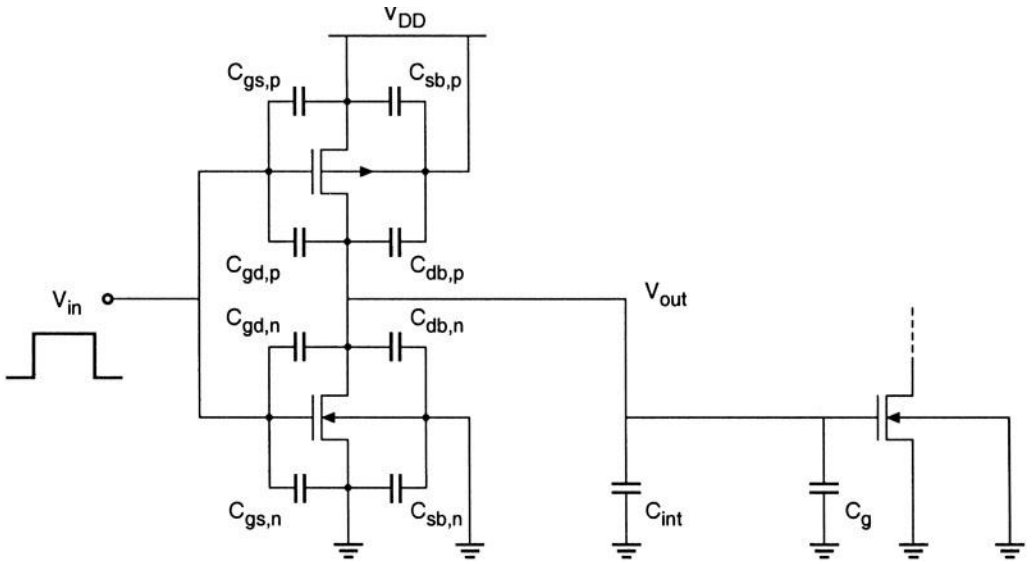
# CMOS Switching Characteristics

- **CMOS Switching Characteristics**

  - in an AC analysis, we need to consider the capacitance in the circuit

  - note that the parasitic inductance tends to be small enough to be ignored (for now!)

  - we consider an inverter that is driving another CMOS device or multiple CMOS devices in parallel

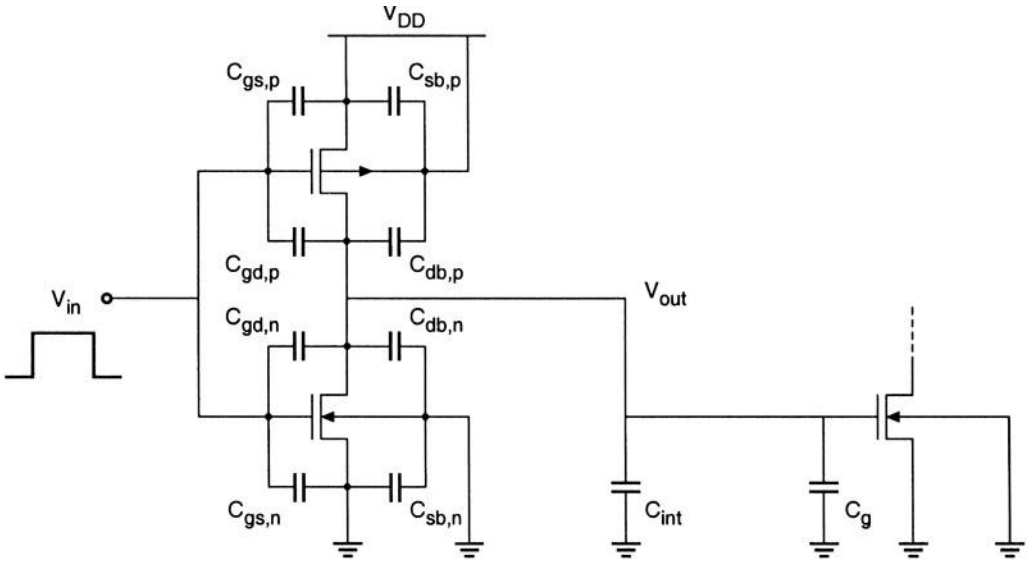


# CMOS Switching Characteristics

- **CMOS Switching Characteristics**

  - there are 4 main groups of capacitance in the circuit

    1) Driver's Oxide Capacitance
    2) Driver's Junction Capacitance
    3) Interconnect Capacitance
    4) Receiver Oxide Capacitance



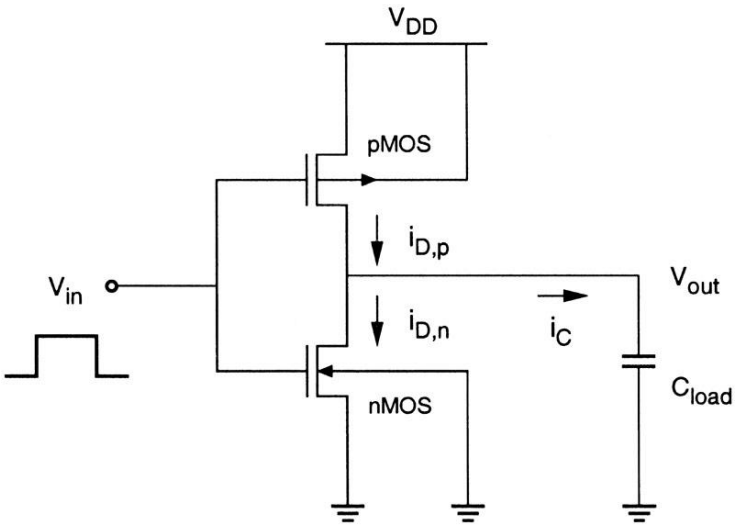
# CMOS Switching Characteristics

- **CMOS Switching Characteristics**

  - we know that all of these capacitances vary as the dimensions of the inverter are altered and for various interconnect configurations

  - in order to get a feel for how the capacitance effects performance, we assume that we can *lump* all of the capacitances into a fixed load capacitance ($C_{load}$)

$$C_{load} = C_{gd,n} + C_{gd,p} + C_{db,n} + C_{db,p} + C_{\text{int}} + C_g$$



# CMOS Switching Characteristics

- **CMOS Switching Characteristics**

  - in this expression we **eliminate** some of the capacitances:

    $C_{sb,n}$, $C_{sb,p}$ : There is no voltage change from $V_{sb,n}$ or $V_{sb,p}$ so there is no net capacitance

    $C_{gs,n}$, $C_{gs,p}$ : Since these are connected between $V_{in}$ and $V_{DD}/V_{SS,}$ the *input* drives these capacitances. It is not part of the capacitance that the device *output* drives.

  - this expression does include the interconnect and gate capacitance of the circuits that this inverter is driving

$$C_{load} = C_{gd,n} + C_{gd,p} + C_{db,n} + C_{db,p} + C_{\text{int}} + C_g$$

**Oxides of Driver** — $C_{gd,n} + C_{gd,p}$

**Junctions of Driver** — $C_{db,n} + C_{db,p}$

**Interconnect** — $C_{\text{int}}$

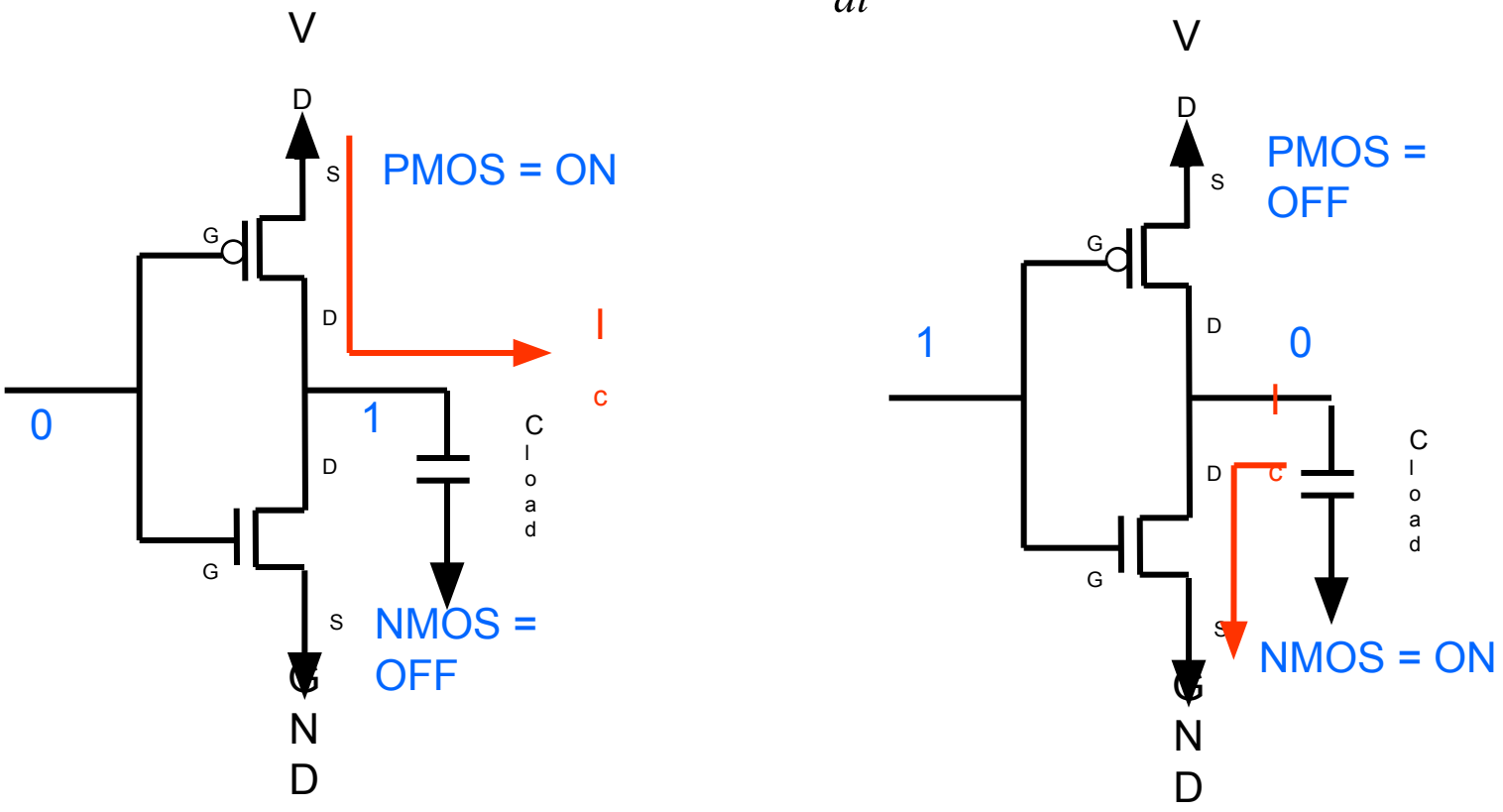
**Oxide of Receiver** — $C_g$

# CMOS Switching Characteristics

- **CMOS Switching Characteristics**

  - the speed of the device describes how fast we can charge or discharge the load capacitor

$$i_C = C\frac{dV}{dt}$$



# CMOS Switching Characteristics

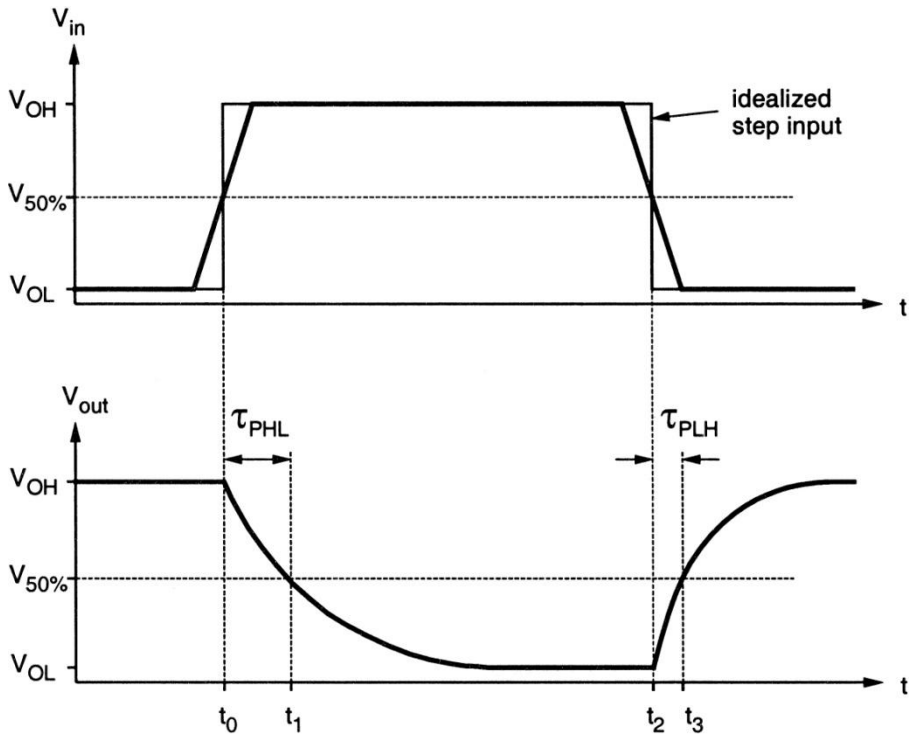
- **Delay Time Definition**

  - the delay is the time it takes to switch from the steady state level to the 50% level

$$V_{50\%} = V_{OL} + \frac{1}{2}\cdot\left(V_{OH} - V_{OL}\right) = \frac{1}{2}\cdot\left(V_{OH} + V_{OL}\right)$$

$$\tau_{PHL} = t_1 - t_0$$

$$\tau_{PLH} = t_3 - t_2$$




  - Note that in CMOS:

    $V_{OH}=V_{DD}$
    $V_{OL}=V_{SS}$

    So $V_{50\%}=V_{DD}/2$

# CMOS Switching Characteristics

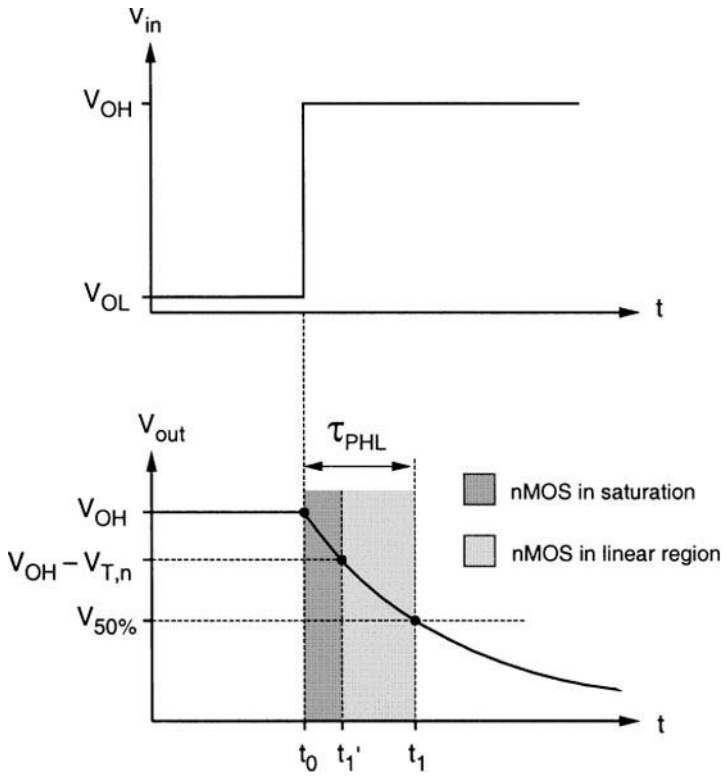
- **Delay Time Derivation ($\tau_{PHL}$)**

  - The current that is used to discharge $C_{load}$ is dictated by the region of operation that the NMOS is in.

  - There are two distinct regions of operation that the NMOS operates in during the transition:

1) $V_{OH}$ to $(V_{OH} - V_{T,n})$ NMOS in Saturation

2) $(V_{OH} - V_{T,n})$ to $V_{50\%}$ NMOS in Linear



# CMOS Switching Characteristics

- **Delay Time Derivation ($\tau_{PHL}$)**

  "Differential Equation Method"

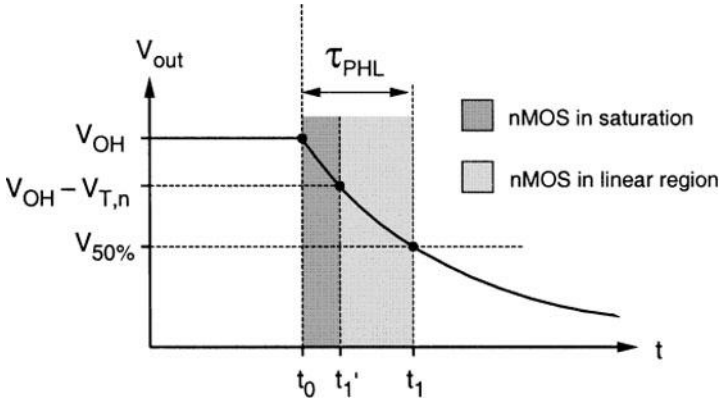
  - we can re-arrange the current expression in the capacitor to be:

$$i_C = -i_{D,n} = C_{load}\frac{dV_{out}}{dt}$$

$$dt = -C_{load}\frac{dV_{out}}{i_{D,n}}$$

  - now we can integrate to solve for *dt*

  - we need to perform two integrals, one for each of the two regions of operation



# CMOS Switching Characteristics

- **Delay Time Derivation ($\tau_{PHL}$)**

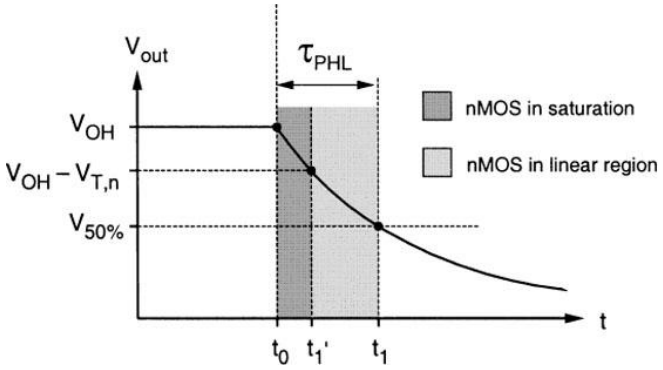
  "Differential Equation Method"

  - For the *saturation* region, our integral is:

$$t_1' - t_0 = \int_{t=t_0}^{t=t_1'} dt = -C_{load} \int_{V_{out}=V_{OH}}^{V_{out}=V_{OH}-V_{T,n}} \left( \frac{1}{i_{D,n-sat}} \right) dV_{out}$$

  - For the *linear* region, our integral is:

$$t_1 - t_1' = \int_{t=t_1'}^{t=t_0} dt = -C_{load} \int_{V_{out}=V_{OH}-V_{T,n}}^{V_{out}=V_{50\%}} \left( \frac{1}{i_{D,n-lin}} \right) dV_{out}$$




  - The delay is simply the sum of these two solutions:

$$\tau_{PHL} = \left(t_1' - t_0\right) + \left(t_1 - t_1'\right)$$

# CMOS Switching Characteristics

- **Delay Time Derivation ($\tau_{PHL}$)**

  "Differential Equation Method"

  - evaluating these integrals and adding the two delays together, we get:

$$\tau_{PHL} = \frac{C_{load}}{k_n\left(V_{OH} - V_{T,n}\right)} \cdot \left[ \frac{2 \cdot V_{T,n}}{V_{OH} - V_{T,n}} + \ln\left( \frac{4\left(V_{DD} - V_{T,n}\right)}{V_{OH} + V_{OL}} - 1 \right) \right]$$

  - we can simplify this further by substituting in $V_{OH}=V_{DD}$ and $V_{OL}=0$

$$\tau_{PHL} = \frac{C_{load}}{k_n\left(V_{DD} - V_{T,n}\right)} \cdot \left[ \frac{2 \cdot V_{T,n}}{V_{DD} - V_{T,n}} + \ln\left( \frac{4\left(V_{DD} - V_{T,n}\right)}{V_{DD}} - 1 \right) \right]$$

# CMOS Switching Characteristics

- **Delay Time Derivation ($\tau_{PLH}$)**

  "Differential Equation Method"

  - we can follow the same process to find $\tau_{PLH}$ using the current equations for the PMOS:

$$\tau_{PLH} = \frac{C_{load}}{k_p\left(V_{DD} - \left|V_{T,p}\right|\right)} \cdot \left[ \frac{2 \cdot \left|V_{T,p}\right|}{V_{DD} - \left|V_{T,p}\right|} + \ln\left( \frac{4\left(V_{DD} - \left|V_{T,p}\right|\right)}{V_{DD}} - 1 \right) \right]$$

  - these solutions are accurate from the standpoint that we use the exact current in the transistors in our derivation of delay.

  - these are still *estimates* and don't include *channel-length-modulation* or small-geometry effects

# CMOS Switching Characteristics

- **Delay Time Derivation ($\tau_{PHL}$)**

"Average Current Method"

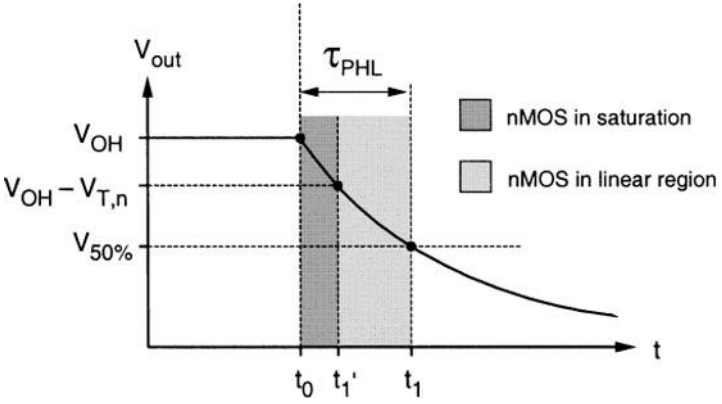
- a simpler technique to estimate the delay is to use the *average current* in the capacitor during the transition.

- this is accomplished by solving for the current at the beginning of the transition and the current at the end of the transition and then averaging the two.

- at the beginning of the High-to-Low transition, the NMOS is in *saturation*

- at the end of the High-to-Low transition, the NMOS is in the *linear region*

$$\tau_{PHL} = \frac{C_{load} \cdot \Delta V_{HL}}{I_{avg,HL}} = \frac{C_{load} \cdot \Delta V_{HL}}{\frac{1}{2} \cdot \left[ i_{D-sat} + i_{D-lin} \right]}$$



# CMOS Switching Characteristics

- **Delay Time Derivation ($\tau_{PHL}$ & $\tau_{PLH}$)**

  "Average Current Method"

  - we can write the expression in terms of the voltages at $V_{in}$ ($V_{gs,n}$) and $V_{out}$ ($V_{ds,n}$):

$$\tau_{PHL} = \frac{C_{load} \cdot \Delta V_{HL}}{\frac{1}{2} \cdot \left[ i_{D-sat}\left(V_{in} = V_{OH}, V_{out} = V_{OH}\right) + i_{D-lin}\left(V_{in} = V_{OH}, V_{out} = V_{50\%}\right) \right]}$$

$$\tau_{PLH} = \frac{C_{load} \cdot \Delta V_{LH}}{\frac{1}{2} \cdot \left[ i_{D-sat}\left(V_{in} = V_{OL}, V_{out} = V_{OL}\right) + i_{D-lin}\left(V_{in} = V_{OL}, V_{out} = V_{50\%}\right) \right]}$$

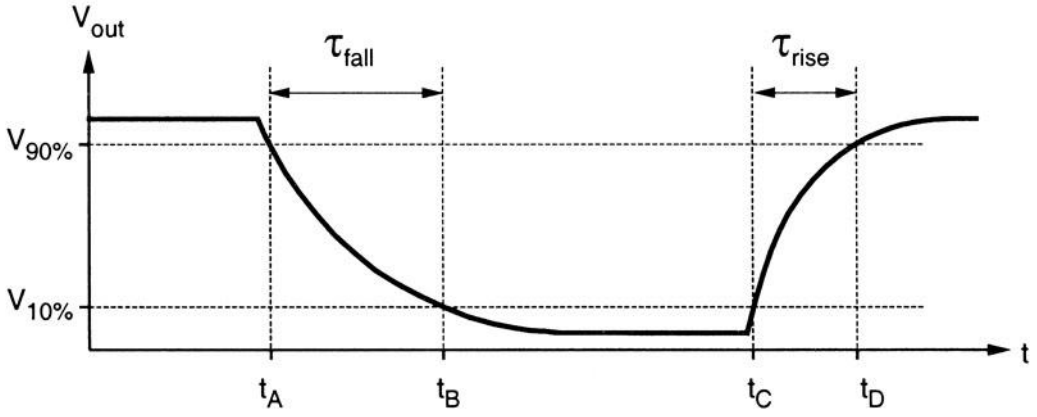
  - this technique tends to be faster and easier to use than the *differential equation* method.

# CMOS Switching Characteristics

- **Rise & Fall Time Definitions ($\tau_{rise}$ & $\tau_{fall}$)**

  - rise time ($\tau_{rise}$) is the time it takes to transition from $V_{10\%}$ to $V_{90\%}$

  - fall time ($\tau_{fall}$) is the time it takes to transition from $V_{90\%}$ to $V_{10\%}$




  - we can use either the (1) *differential equation* or the (2) *average current* technique to solve for these

  - in these transitions, the transistors again operate in both the *saturation* and *linear* regions

  - the only difference is that the limits of the transition are $V_{10\%}$ and $V_{90\%}$

# CMOS Switching Characteristics

- **Non-ideal Inputs**

  - in all of these derivations, we have assumed a perfect step input.

  - if the input is not a perfect step (i.e., it has a finite delay or rise time), it will increase the delay of the gate

  - we can use an RMS estimation to account for the non-ideal input:

$$\tau^2_{PHL(actual)} = \tau^2_{PHL(to\_step\_input)} + \tau^2_{PLH(of\_input)}$$

$$\tau^2_{PLH(actual)} = \tau^2_{PLH(to\_step\_input)} + \tau^2_{PHL(of\_input)}$$

  - we can also estimate the delay of the input if we are only given its rise/fall time by using:

$$\tau_{PLH} = \frac{\tau_{rise}}{2}$$

$$\tau_{PHL} = \frac{\tau_{fall}}{2}$$

# CMOS Switching Characteristics

- **Non-ideal Inputs**

  - we can apply this technique to the rise and fall times also:

$$\tau^2_{rise(actual)} = \tau^2_{rise(to\_step\_input)} + \tau^2_{fall(of\_input)}$$

$$\tau^2_{fall(actual)} = \tau^2_{fall(to\_step\_input)} + \tau^2_{rise(of\_input)}$$

# CMOS Switching Characteristics

- **Designing for Constraints**

  - when we begin a design, we typically start with specification

  - we then size the transistors to achieve the desired performance

  - we saw how the sizes of the transistor effect the DC specs, specifically $V_{th}$

  - we also need to size the transistors so that for a given load capacitance, the gate can achieve a designed delay or rise/fall time.

  - we can use the expressions for delay and rise/fall time that we derived to calculate the necessary transistor sizes.

# CMOS Switching Characteristics

- **Designing for Constraints**

  - the *average current* method is the simplest technique to use:

$$\tau_{PHL} = \frac{C_{load} \cdot \Delta V_{HL}}{\frac{1}{2} \cdot \left[ i_{D-sat}\left(V_{in} = V_{OH}, V_{out} = V_{OH}\right) + i_{D-lin}\left(V_{in} = V_{OH}, V_{out} = V_{50\%}\right) \right]}$$

$$\tau_{PLH} = \frac{C_{load} \cdot \Delta V_{LH}}{\frac{1}{2} \cdot \left[ i_{D-sat}\left(V_{in} = V_{OL}, V_{out} = V_{OL}\right) + i_{D-lin}\left(V_{in} = V_{OL}, V_{out} = V_{50\%}\right) \right]}$$

  - in this expression, we can insert our timing spec in for $\tau_{PHL}$ or $\tau_{PLH}$

  - the RHS of the expression must evaluate to be less than or equal to the timing spec

# CMOS Switching Characteristics

- **Designing for Constraints**

  - in the timing expression, notice that $k_n$ and $k_p$ are parameters under our control

  - these parameters are in the denominator of the timing expression, meaning that as $k_n$ and $k_p$ increase, the delay of the circuit will decrease.

  - this means that *larger* = *faster*

  - we typically leave the lengths of the NMOS and PMOS transistors equal to each other

  - we also typically set the lengths to the smallest possible dimension for a given process.

  - this gives us the highest transconductance for a given *Length* and also minimizes the area.

  - a given design process consists of the following steps:

    1) set $L_p=L_n=L_{min}$
    2) find the $W_p/W_n$ ratio that will yield the desired $V_{th}$
    3) find the minimum values for $W_p$ and $W_n$ to achieve timing
    4) combine the minimum sizes and the $W_p/W_n$ ratio to select final sizes
    5) round up the dimensions to give additional margin and standard sizes
       (i.e., 4.927um rounds up to 5um)

# CMOS Switching Characteristics

- **Area vs. Delay**

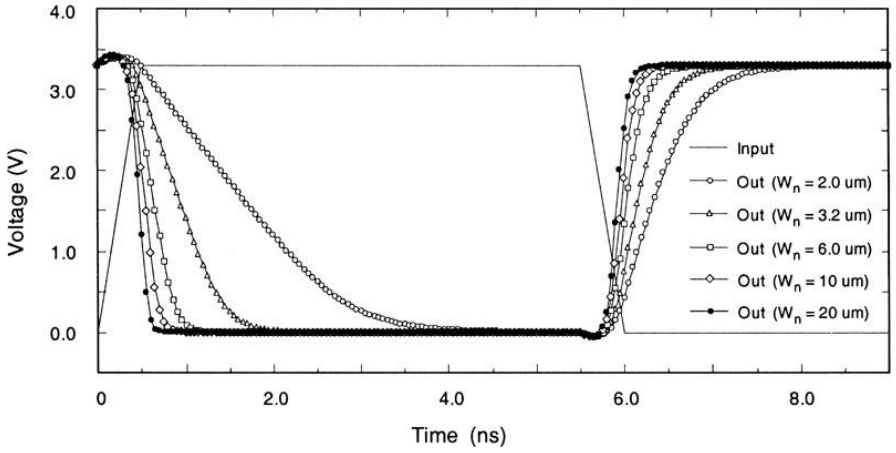
  - we've seen that *larger* = *faster* for a given inverter

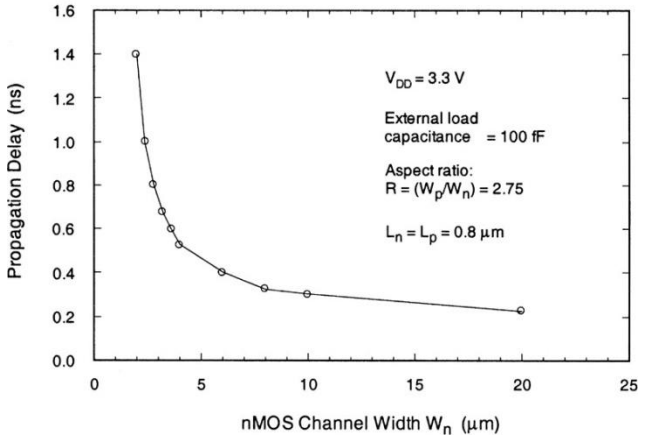
  - however, we have made an assumption that the load capacitance is independent of transistor size

  - we know what a portion of the load capacitance comes from the driver oxide and driver junctions

  - this means that as the inverter gets larger, so does the capacitance

  - this leads to a point of *diminishing returns* with regards to reducing delay



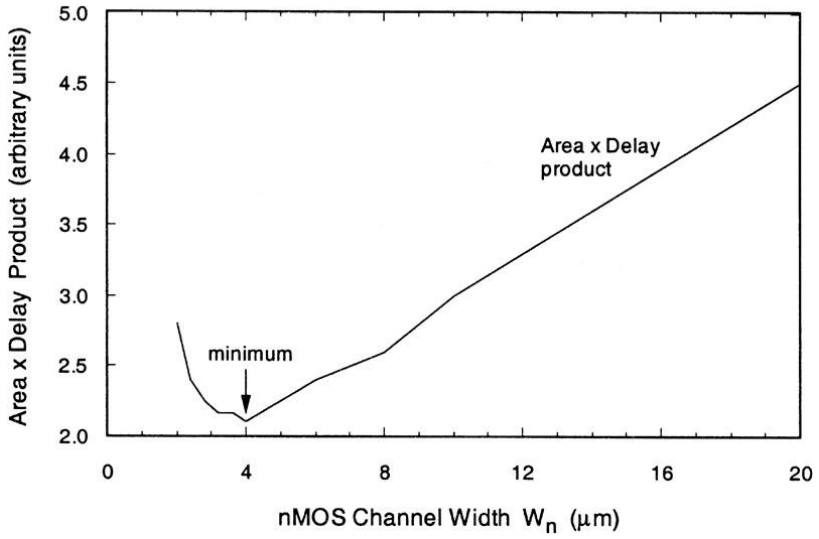
# CMOS Switching Characteristics

- **Area vs. Delay**

  - we can look at the **Area X Delay Product** to gauge the *quality* of a design with regards to efficient area usage.

  - typically we will see an inflection point which indicates the point at which increasing the size of the transistors to decrease delay is out-weighed by the negative impact of increasing the area used on the silicon.

  - if a timing specifications requires an excessively large sized gate, it typically means that the process is not sufficient to meet timing.



# CMOS Switching Characteristics

- **Interconnect**

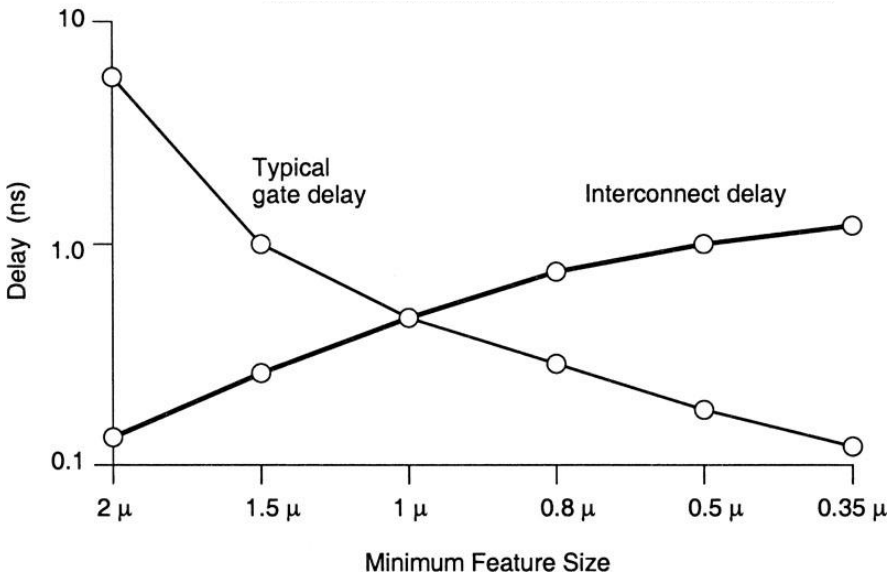
  - one of the components in the load capacitance is the interconnect.

  - the interconnect refers to the polysilicon and metal layers that are used to connect the gates together.

  - as sizes on-chip shrink, we've seen that the scaling of interconnect is a big problem because the delay actually increases as you get smaller.

  - in addition, the delay scales quadradically with length meaning that intra-module traces and global interconnect can create significant timing challenges.

  - in modern processes, the delay of the interconnect is actually more than the switching delay of the transistors.



# CMOS Switching Characteristics

- **Interconnect Modeling**

  - modeling of the interconnect describes the equivalent circuits we use to describe the electrical behavior of the materials.

  - the type of model we use is a trade-off between accuracy and simulation time

  - we typciall use 1 of the 3 following models:

| | Typical Uses |
|---|---|
| 1) Lumped Capacitance | inter-module |
| 2) RC network | intra-module and global |
| 3) Transmission Line | global and off-chip |

# CMOS Switching Characteristics

- **Interconnect Modeling**

  - we choose the appropriate model based on the rise/fall time of the driver relative to the *prop delay* of the interconnect

  - the *prop delay* ($t_{prop}$) is the time it takes for the wave to travel down the length of the interconnect:

  - the velocity of a wave in a dielectric is given by:

$$v = \frac{c}{\varepsilon_r}$$

  - the *prop delay* can then be given by:

$$t_{prop} = \frac{length}{v}$$

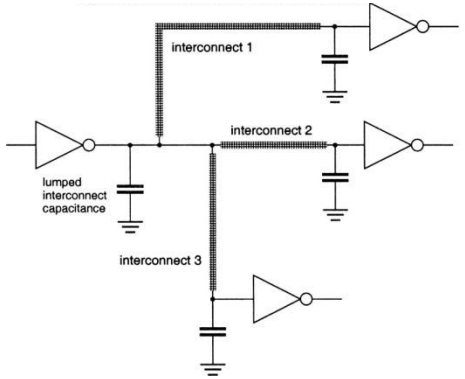
# CMOS Switching Characteristics

- **Interconnect Modeling**

  - we move between a *lumped* (C or RC) and a *distributed (*transmission line) model as follows:

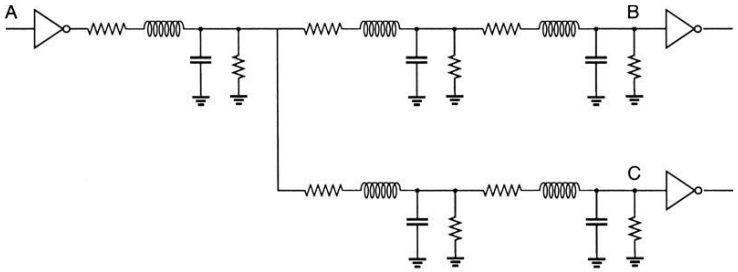
"Lumped"

$$\tau_{rise} < 2.5 \cdot \left( \frac{l}{v} \right)$$




"Distributed"

$$\tau_{rise} \geq 2.5 \cdot \left( \frac{l}{v} \right)$$

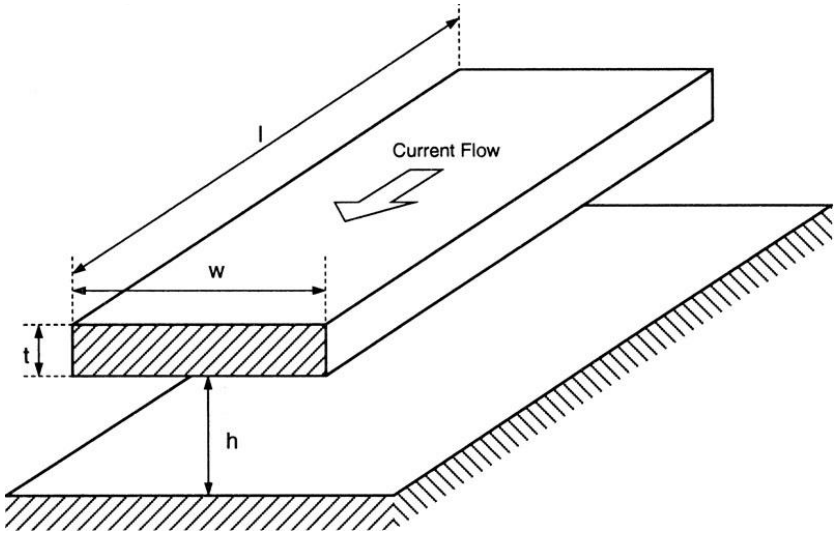


# CMOS Switching Characteristics

- **Interconnect Resistance**

  - resistance is based on the geometry and materials of the interconnect

$$R = \frac{\rho \cdot l}{A} = R_S \cdot (\#\_of\_squares)$$

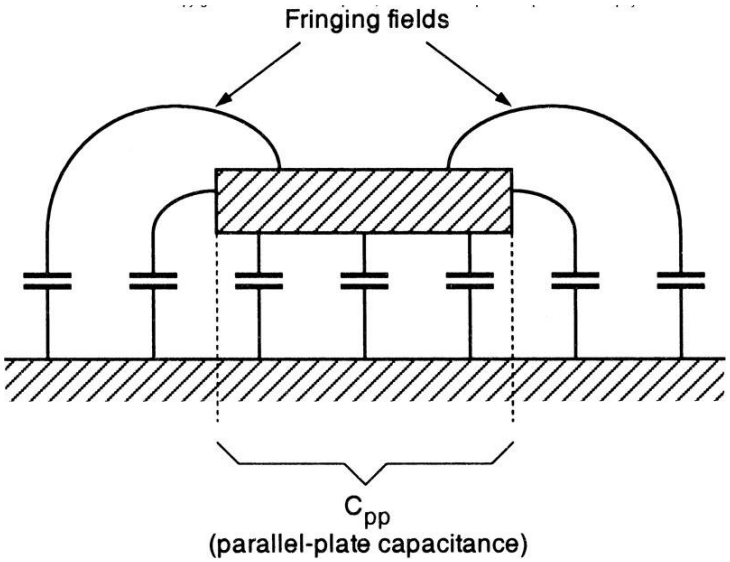


# CMOS Switching Characteristics

- **Interconnect Capacitance**

  - capacitance depends on the surface area of the conductor, the insulating materials between the conductors, and the distance between the conductors.

$$C = \frac{\varepsilon \cdot A}{t}$$



# CMOS Switching Characteristics

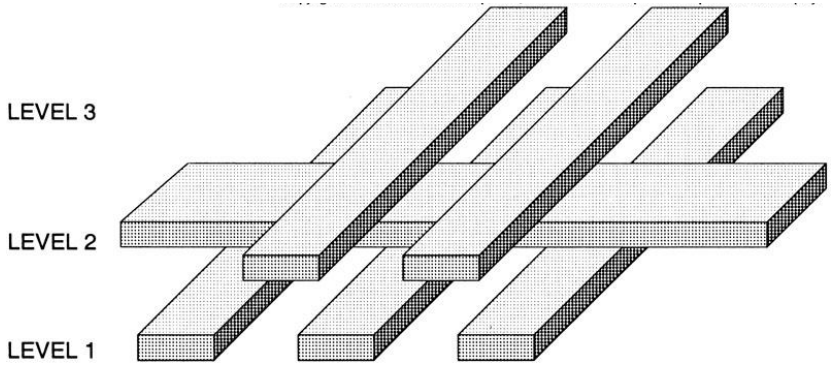
- **Interconnect Capacitance**

  - interconnect modeling becomes a complex problem due to the 3D geometries present on-chip

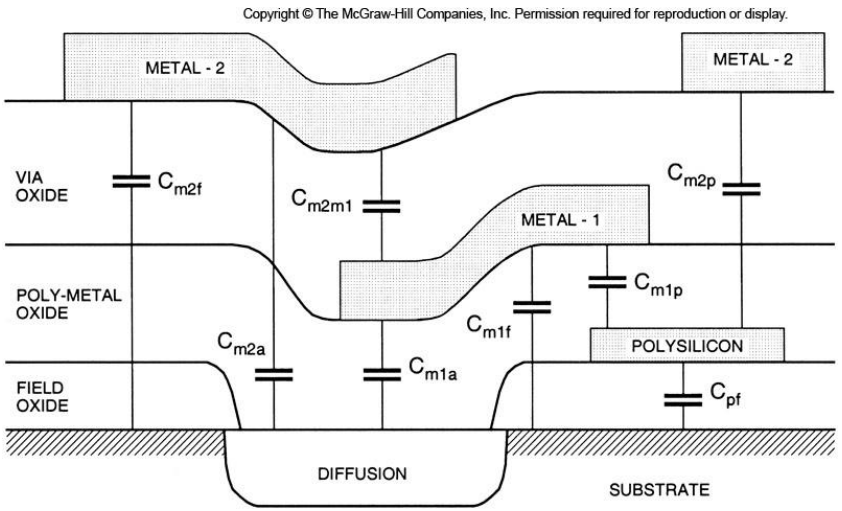
  - we typically take a *guess* at the capacitance of the interconnect for initial simulations.

  - once we start physically laying out our design, we can use the CAD tool to *extract* the actual capacitance and back annotate it into our simulation.

  - we then run a new simulation with accurate capacitance models to verify timing is still met post-layout.





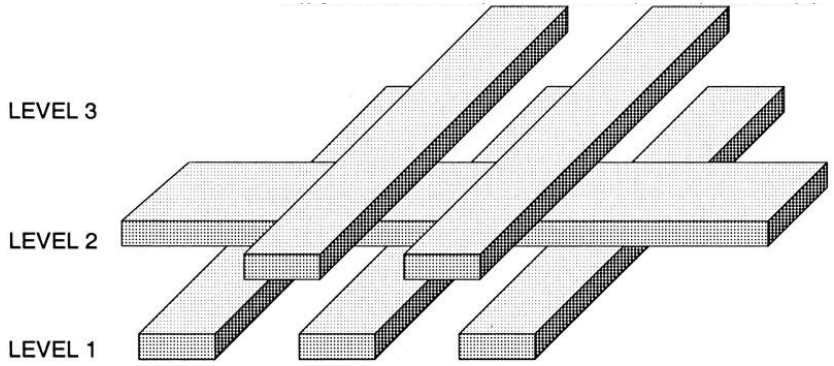
# CMOS Switching Characteristics

- **Interconnect Capacitance**

  - *Cross-talk* refers to the noise that is generated on a line due to capacitive coupling from neighboring lines that are switching.

  - as geometries get smaller, lines are closer together so capacitance goes up.

  - we can reduce cross-talk by separating the traces or inserting ground lines between the signals, but this takes area.



# CMOS Switching Characteristics

- **Elmore Delay**

  - when we model interconnect using RC networks, it doesn't take many branches in the net before the KVL/KCL solution for the delay gets complex.

  - *Elmore Delay* is a technique to estimate the overall delay between two nodes of an RC network tree.

  - in Elmore Delay, we find the equivalent RC network of the path between two nodes by:

    - summing the delay of each segment in our path-of-interest

    - we construct a set of RC networks as seen by our path-of-interest and then sum them together

    - we walk through the series resistance in our path-of-interest.

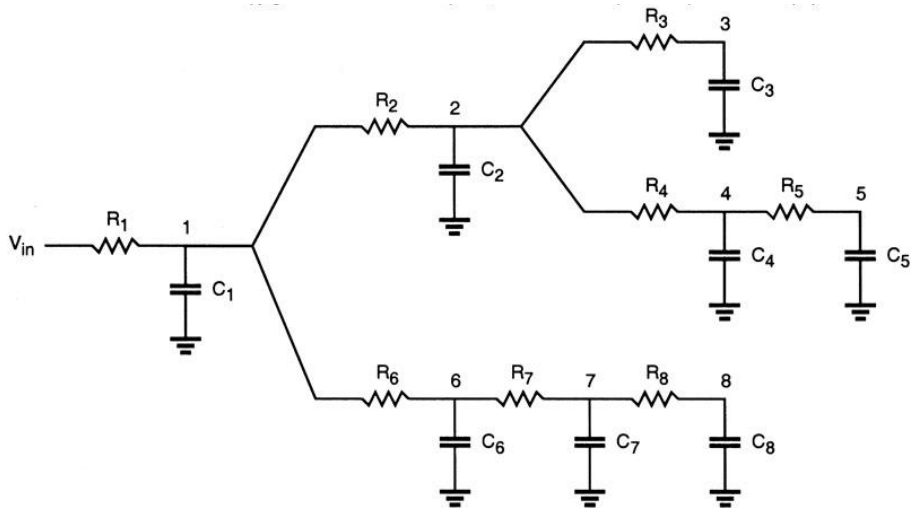
    - for each resistor node in our path-of interest, we include RC's in our expression as follows:

      - C's NOT in our path-of-interest are included as ($R_{seg} \cdot C_x$)
      - C's that ARE in our path-of-interest can't be seen if they are on the far side of a resistor in our path-of-interest
      - as we get to the end of our path-of-interest, we can see all of the downstream Capacitances past our end-node.

# CMOS Switching Characteristics

- **Elmore Delay**

  - example: Find the expression for the equivalent RC from $V_{in}$ to node 7:




$$\tau_{D7} = (R_1)\cdot(C_1) + (R_1)\cdot(C_2) + (R_1)\cdot(C_3) + (R_1)\cdot(C_4) + (R_1)\cdot(C_5) + (R_1 + R_6)\cdot(C_6) + (R_1 + R_6 + R_7)\cdot(C_7) + (R_1 + R_6 + R_7)\cdot(C_8)$$

# CMOS Switching Characteristics

- **Dynamic Power Consumption**

  - in theory, a CMOS gate does not consume any Static Power because the NMOS and PMOS transistors are in the cut-off regions when driving $V_{OH}$ or $V_{OL}$

  - we know what there is leakage current in cut-off, however to the first order we neglect it.

  - the majority of the power is due to the charging and discharging of $C_{load}$

  - this is called *Dynamic Power* because it is AC in nature and only occurs when the gate switches

  - this current is described as:

$$i_C = C_{load} \frac{dV_{out}}{dt}$$

  - since the current consumed is proportional to the number of times that the gate switches, we need to make an assumption to the number of times per second that $V_{out}$ switches

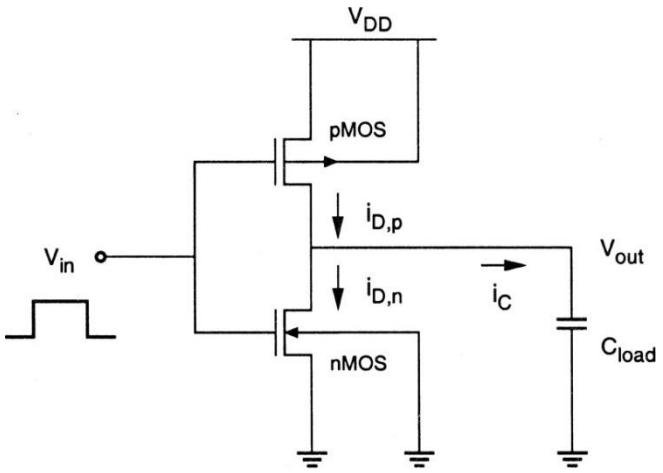
  - since we have a binary system, we can assume that the output will be a '0' 50% of the time and a '1' 50% of the time.

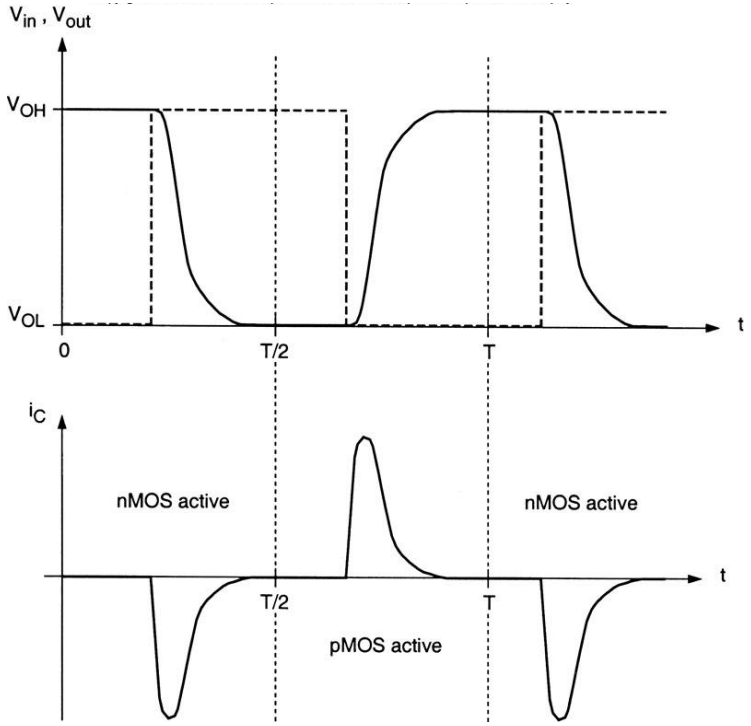
  - we can model the voltage on $V_{out}$ as a periodic square wave

# CMOS Switching Characteristics

- **Dynamic Power Consumption**

  - current will be drawn from $V_{DD}$ and sunk into $V_{SS}$ during a transition



# CMOS Switching Characteristics

- **Dynamic Power Consumption**

  - assuming a periodic input and output waveform, the average power dissipated by a device over one period is given as:

$$P_{avg} = \frac{1}{T}\int_0^T v(t)\cdot i(t)dt$$

  - we split up the period into two sections:

  0 →T/2  $V_{in}$ transitions from a 0 to a 1, the NMOS discharges $C_{load}$

  T/2 →T  $V_{in}$ transitions from a 1 to a 0, the PMOS charges $C_{load}$

# CMOS Switching Characteristics

- **Dynamic Power Consumption**

  - we can now re-write our average power expression as:

$$P_{avg} = \frac{1}{T}\left[\int_0^{T/2} V_{out} \cdot \left(-C_{load}\frac{dV_{out}}{dt}\right)dt + \int_{T/2}^{T} (V_{DD} - V_{out}) \cdot \left(C_{load}\frac{dV_{out}}{dt}\right)dt\right]$$

$$P_{avg} = \frac{1}{T}\left[\left(-C_{load}\frac{V_{out}^2}{2}\right)\Bigg|_0^{T/2} + \left(V_{DD} \cdot V_{out} \cdot C_{load} - C_{load}\frac{V_{out}^2}{2}\right)\Bigg|_{T/2}^{T}\right]$$

$$P_{avg} = \frac{1}{T} \cdot C_{load} \cdot V_{DD}^2$$

$$P_{avg} = f \cdot C_{load} \cdot V_{DD}^2$$

# CMOS Switching Characteristics

- **Dynamic Power Consumption**

  - a more qualitative view of this power consumption is as follows:

    Capacitance is defined as:

$$C = \frac{Q}{V}$$

    Each cycle, the average current in the capacitor is:

$$I_{AVG} = \frac{Q}{T} = \frac{C \cdot V}{T}$$

    Power is I·V, which gives:

$$P_{AVG} = V_{AVG} \cdot I_{AVG} = V \cdot \frac{C \cdot V}{T} = \frac{1}{T} \cdot C_{load} \cdot V_{out}^2 = f \cdot C_{load} \cdot V_{out}^2$$

# CMOS Switching Characteristics

- **Power Delay Product (PDP)**

  - another *quality* measure of a design is the PDP

  - this is a measure of the energy required to switch logic levels in a given period.

  - qualitatively, Power x Time is:

$$P_{avg} = \frac{1}{time} \cdot C_{load} \cdot V_{DD}^2$$

$$P_{avg} \cdot \tau = C_{load} \cdot V_{DD}^2$$

# CMOS Switching Characteristics

- **Power Delay Product (PDP)**

  - as the delay goes down, the power goes up.

  - the power going up is due to the increase in intrinsic junction capacitance of the driver.

  - the delay reaches an ~asymptotic limit as the size is increased.

  - the power increases as the size is increased.

  - looking at the PDP can give an estimate of when you are optimally sized to deliver energy in the most effective manner.

PDP=(C * V*V)

# Ring Oscillator

- **Ring Oscillator**

  - if we connect a chain of inverters in a loop and have an ODD number of inverters, the circuit is inherently unstable.

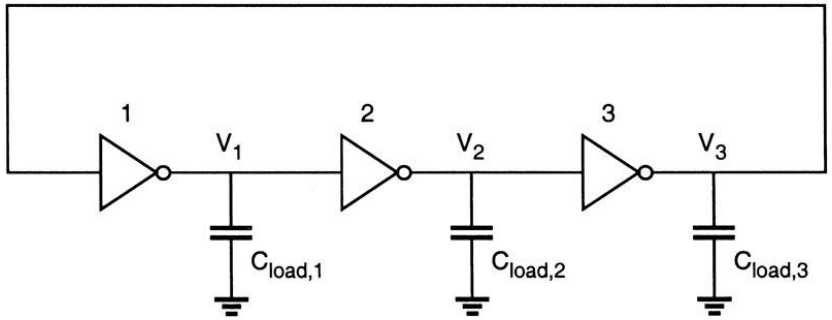
  - the circuit will *oscillate* between a 0 and 1 indefinitely.

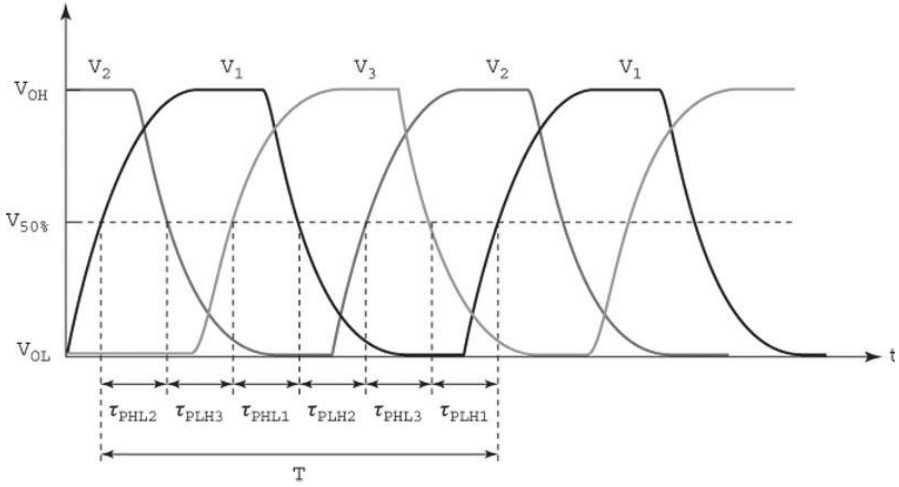
  - the frequency of the oscillation depends on the gate delay of the inverter.

  - this type of circuit is commonly used to test the device delay of a given process.

  - this can also be used to create a clock.

  - the clock frequency of the ring oscillator is not typically controlled tight enough to be used as the system clock.




$$f = \frac{1}{2 \cdot n_{inv} \cdot \tau_{inv}}$$