EECE 481

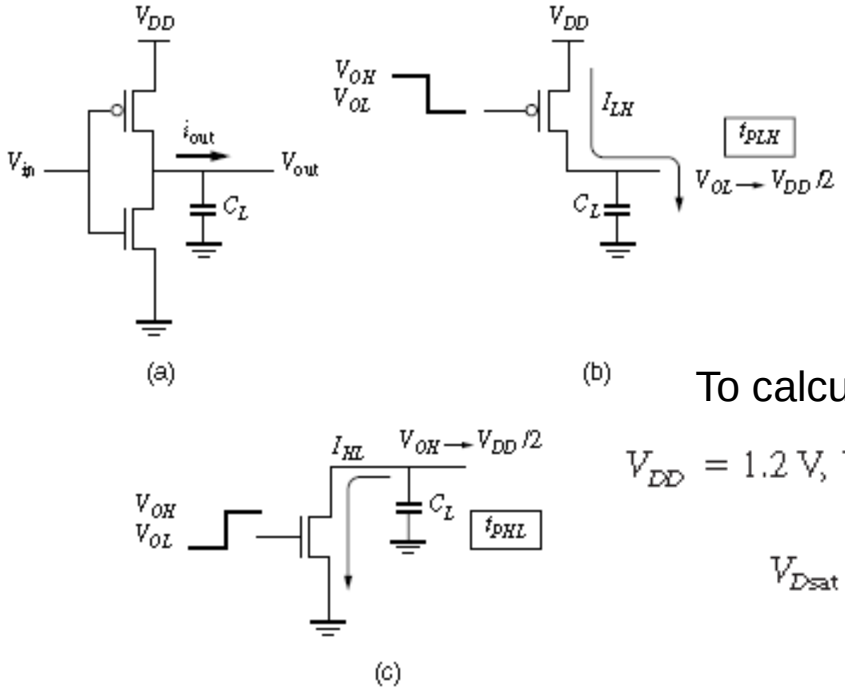
High-Speed CMOS Gate Design

Lecture 9

Reza Molavi
Dept. of ECE
University of British Columbia
reza@ece.ubc.ca

Slides Courtesy : Dr. Res Saleh (UBC), Dr. D. Sengupta (AMD), Dr. B. Razavi (UCLA)

Effective Resistance Calculations



Calculate the delay from the charge/discharge of capacitor

$$I = C_L \frac{dV}{dt} \rightarrow \Delta t \approx \frac{C_L \Delta V}{I_{DS}}$$

$$t_{PHL} = \frac{C_L (V_{DD}/2)}{I_{HL}}$$

To calculate I_{HL}

$$V_{DD} = 1.2\ \text{V},\ V_T = 0.4\ \text{V}, \text{and}\ E_C L = 0.6\ \text{V}$$

$$V_{Dsat} = \frac{(V_{GS} - V_T)E_C L}{(V_{GS} - V_T) + E_C L} = \frac{(1.2 - 0.4)(0.6)}{(1.2 - 0.4) + (0.6)} \cong 0.34\ \text{V}$$

NMOS remains in Saturation throughout T_{PHL}

$$t_{PHL} = \frac{C_L (V_{DD}/2)}{(I_{Dsat})_n}$$

$$R_N = \frac{V_{DD}/2}{0.7(I_{Dsat})_n}$$

Effective Resistance Calculations

Problem:

Using 0.13 μm technology parameters, compute R_{eqn} and R_{eqp} from the equations above for unit-sized devices.

Solution:

For the NMOS device,

$$I_{Dsat} = \frac{W_N v_{sat} C_{ox}(V_{DD} - V_{TN})^2}{(V_{DD} - V_{TN}) + E_{CN}L_N}$$

$$= \frac{(0.1)(10^{-4})8(10^6)1.6(10^{-6})(1.2 - 0.4)^2}{(1.2 - 0.4) + 0.6} \approx 60\ \mu A$$

$$\therefore R_{eqn} = \frac{1.2/2}{0.7(60\ \mu A)} = 14.5\ k\Omega$$

For the PMOS device,

$$I_{Dsat} = \frac{W_P v_{sat} C_{ox}(V_{DD} - |V_{TP}|)^2}{(V_{DD} - |V_{TP}|) + E_{CP}L_P}$$

$$= \frac{(0.1)(10^{-4})8(10^6)1.6(10^{-6})(1.2 - 0.4)^2}{(1.2 - 0.4) + 2.4} \approx 25\ \mu A$$

$$\therefore R_{eqp} = \frac{1.2/2}{0.7(25\ \mu A)} = 33.5\ k\Omega$$

The HSPICE

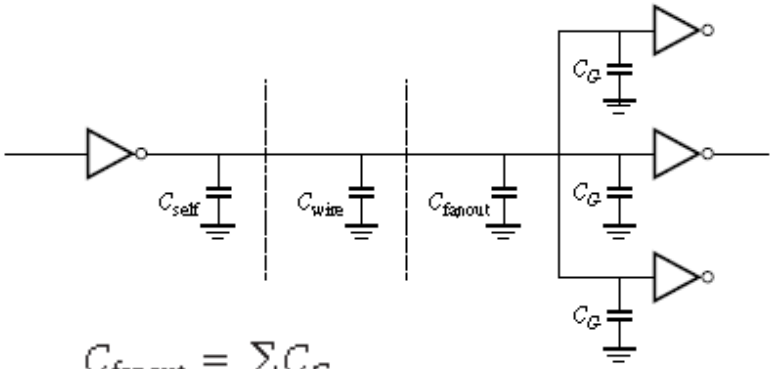
Numbers of 12.5k-ohms

And 30k-ohm

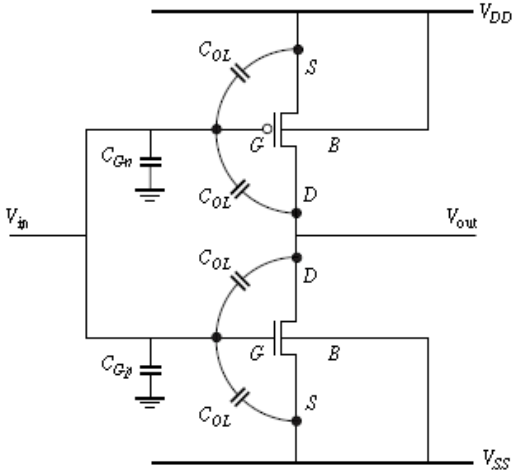
Load Capacitance Calculations

"Fanout Gate Capacitance"

Each gate connected to the output of a CMOS gate (driver) presents a capacitive load , when all summed termed as "*Fanout Gate Capacitance*"



$$C_{fanout} = \Sigma C_G$$



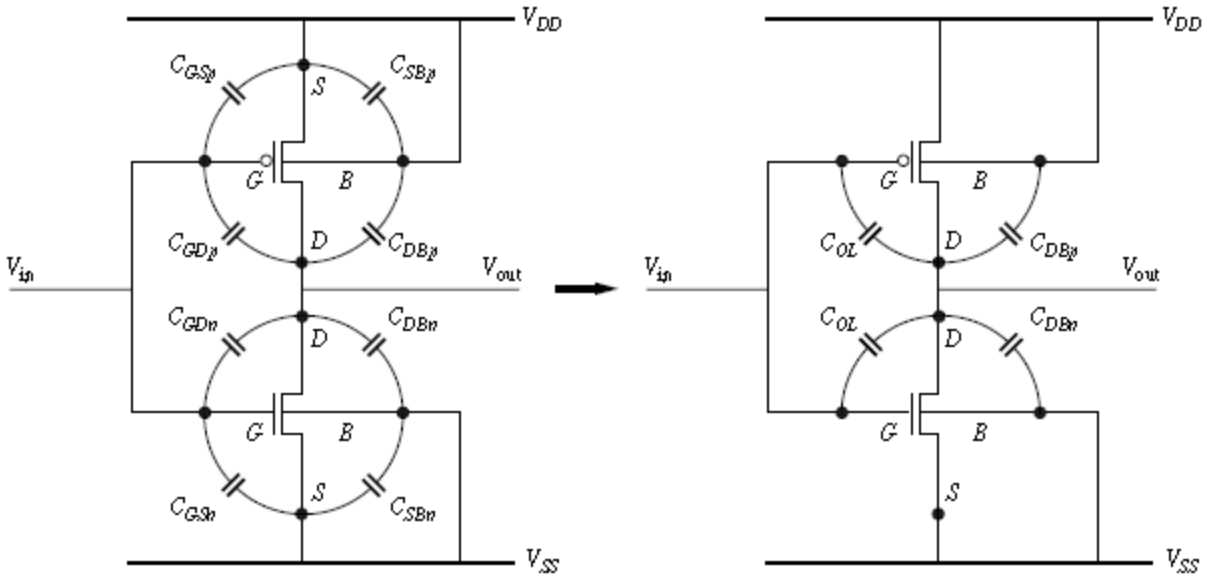
$$C_G = C_{Gn} + 2C_{OL} + C_{Gp} + 2C_{OL}$$
$$= C_{ox}LW_n + 2C_{ol}W_n + C_{ox}LW_p + 2C_{ol}W_p$$
$$= (C_{ox}L + 2C_{ol})(W_n + W_p)$$

$$C_g = C_{ox}L + 2C_{ol} = 1.6\ \text{fF}/\mu\text{m} + 2(0.25\ \text{fF}/\mu\text{m}) \cong 2\ \text{fF}/\mu\text{m}$$

$$\therefore \Sigma C_G = 2\frac{\text{fF}}{\mu\text{m}} \times (\underbrace{W_{p1} + W_{n1}}_{\text{first inverter}} + \underbrace{W_{p2} + W_{n2}}_{\text{second inverter}} + \cdots)$$

Load Capacitance Calculations - II

"Self Capacitance"



$C_{GDp,n} = C_{ol}$ (?)

Another term in load caclulations is the load presented by

The driver itself (all capacitances connected to V_{out})

We have to consider Miller effect! In calculating C_{self}

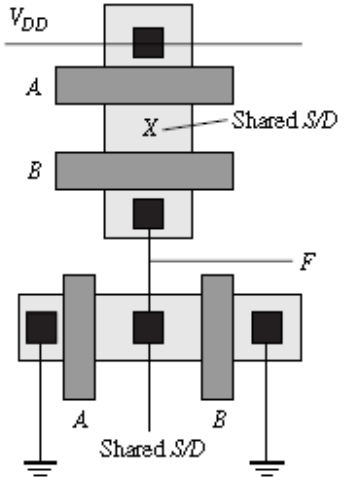
$$C_{self} = C_{DBn} + C_{DBp} + 2C_{OL} + 2C_{OL}$$

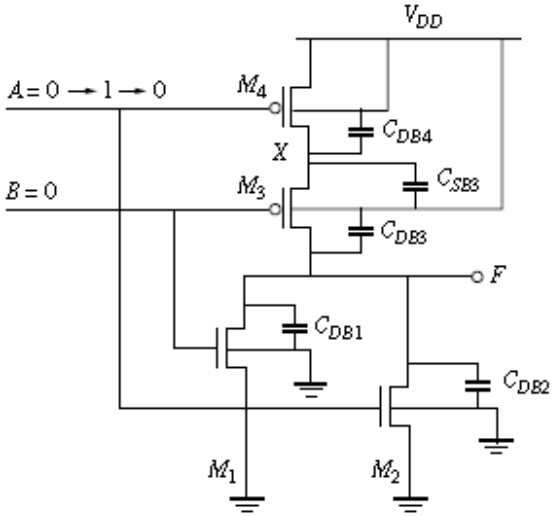
$$= C_{jn}W_n + C_{jp}W_p + 2C_{ol}(W_n + W_p)$$

$$= C_{eff}(W_n + W_p)$$

(Board Notes)

Load Capacitance for NAND and NOR





Consider *sharing*

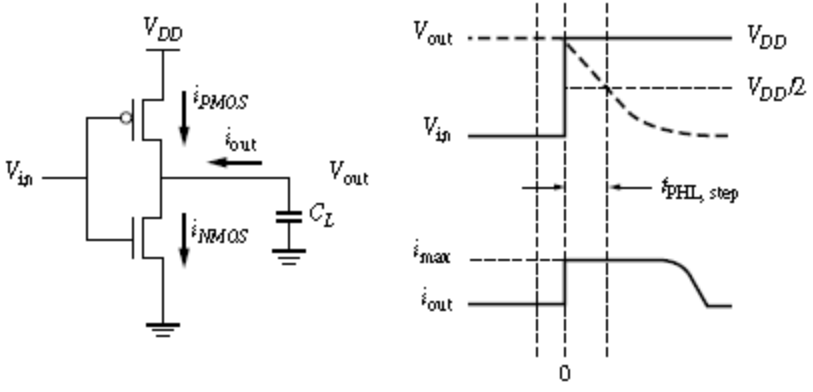
What is worst-case capacitance calculations, why does it matter for speed calculations?

$$C_{self} = \underbrace{C_{DB1} + C_{DB2}}_{n^+ \text{shared } S/D} + C_{DB3} + \underbrace{C_{SB3} + C_{DB4}}_{p^+ \text{shared } S/D}$$

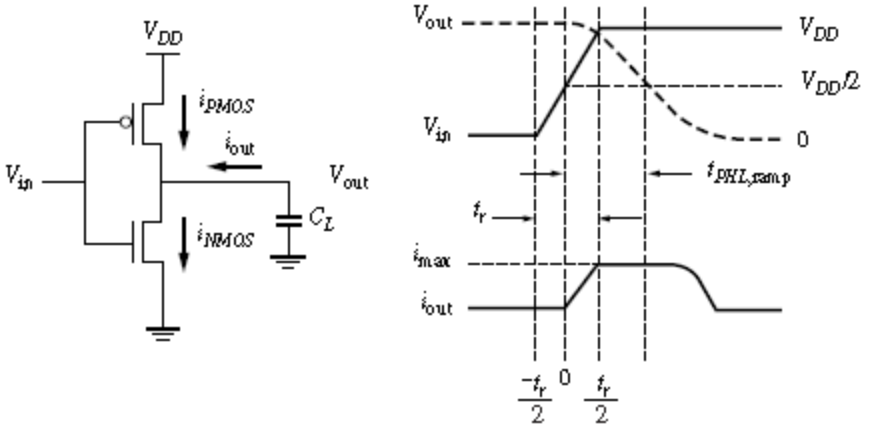
$$= C_{DB12} + C_{DB3} + C_{SDB34}$$

Example: what is the worst case input and output capacitance for a NAND3 CMOS gate (Board Notes)

Response of Inverter to a Ramp voltage change



Response of an inverter to a step V_{in}



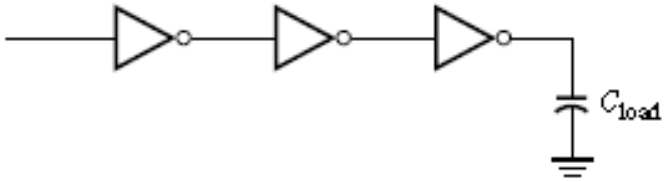
Response of an inverter to a ramp V_{in}

We modeled this propagation delay by *0.7(Ln2)*RC* *(back in propagation delays)*

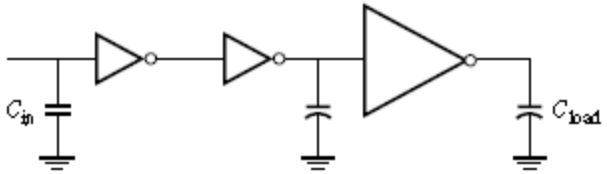
Ramp input makes the propagation delay longer(approximately by 50%) making it ~ *RC*

A ramp voltage change is a more realistic scenario when we have a full chain of inverters driving in cascade

Gate Sizing for Optimal Path delay



In order to drive a large load we can NOT use an arbitrary large gate to minimize the delay (it just shifts the problem to the previous stage)



A proper question is drive a load (C_{load}) with an input capacitance of C_{in} (and of course optimizing the delay)

Optimal Path Delay Design

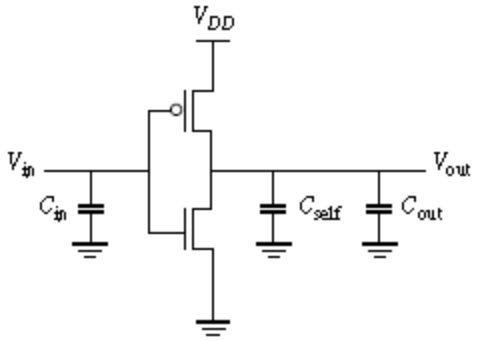
There are two unknowns, the number of gates and the size of each (*two degrees of freedom*)

$$C_{in} = C_g(W_n + W_p) = C_g(W_n + 2W_n) = C_g(3W_n)$$

$$R_{eff} = R_{eqn}\left(\frac{L_n}{W_n}\right)$$

$$\tau_{inv} = R_{eff}C_{in} = R_{eqn}\left(\frac{L_n}{W_n}\right)C_g(3W_n) = 3R_{eqn}C_gL_n$$

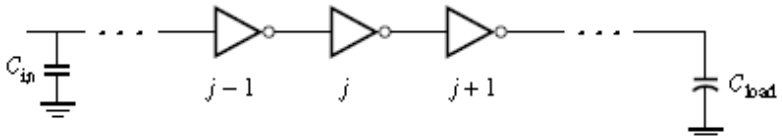
This is the *intrinsic delay* of a gate (specific tag) used many places



$$t_{delay} = R_{eff}C_{in}\left[\frac{C_{out}}{C_{in}} + \frac{C_{self}}{C_{in}}\right] = \tau_{inv}\left[\frac{C_{out}}{C_{in}} + \gamma_{inv}\right]$$

Gate Sizing for Optimal Path delay

$$\text{total\_delay} = \sum_{j=1}^{N} \tau_{inv}\left(\frac{C_{j+1}}{C_j} + \gamma_{inv}\right)$$



$$\text{total\_delay} = \sum_j \tau_{inv}\left(\frac{C_g W_{j+1}}{C_g W_j} + \gamma_{inv}\right) = \sum_j \tau_{inv}\left(\frac{W_{j+1}}{W_j} + \gamma_{inv}\right)$$

Let's consider two consecutive ones

$$D_j = \tau_{inv}\left(\frac{W_j}{W_{j-1}} + \gamma_{inv}\right) + \tau_{inv}\left(\frac{W_{j+1}}{W_j} + \gamma_{inv}\right)$$

Minimize the delay (find a W that makes $dD_j/dW = 0$)

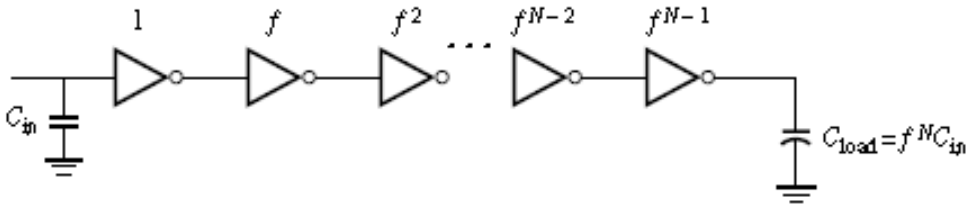
$$\frac{\partial D_j}{\partial W_j} = \tau_{inv}\frac{1}{W_{j-1}} - \tau_{inv}\frac{W_{j+1}}{W_j^2} = 0$$

$$\therefore \frac{W_j}{W_{j-1}} = \frac{W_{j+1}}{W_j}$$

$$\therefore W_j = \sqrt{W_{j+1}W_{j-1}}$$

If the size of each gate is the geometric mean of the two gates (previous and after) the delay Is minimum!

Gate Sizing for Optimal Path delay



Therefore, we can consider the size of gates a geometric sequence with factor *f*

$$f^N C_{\text{in}} = C_{\text{load}}$$

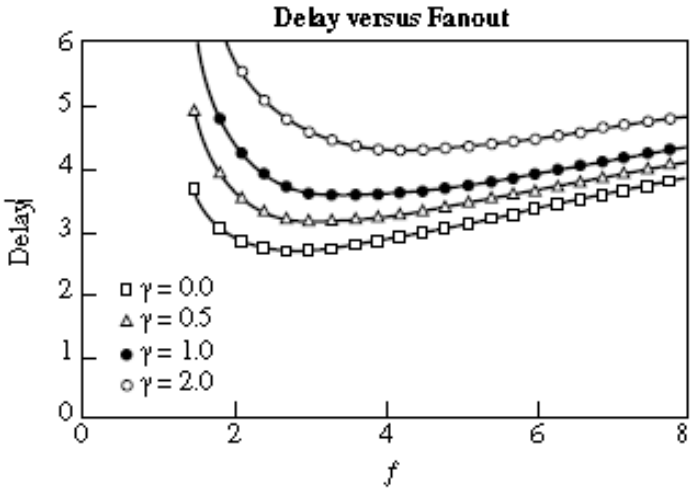
$$\therefore N = \frac{\ln(C_{\text{load}}/C_{\text{in}})}{\ln f}$$

$$\text{total\_delay} = N \times \tau_{\text{inv}}\left(\frac{C_j}{C_{j-1}} + \gamma_{\text{inv}}\right)$$

$$\text{total\_delay} = \frac{\ln(C_{\text{load}}/C_{\text{in}})}{\ln f} \times \tau_{\text{inv}}(f + \gamma_{\text{inv}})$$

What *f* makes the total delay a minimum?

Optimum Fan-out for CMOS Gates



The optimum value of f depends on γ (for $\gamma = 0$, it is e).

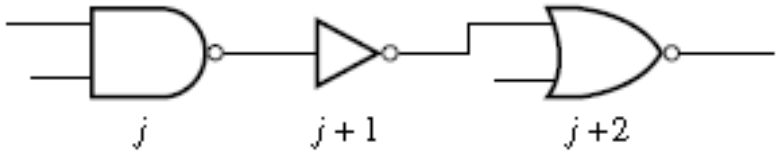
It is similar calculation if we have chain of NAND and NORs:

$$\tau_{\text{nand}} = R_{\text{eff}} C_{\text{in}} = R_{\text{eqn}} \left(\frac{L_n}{W_n} \right) 4 W_n C_g = 4 R_{\text{eqn}} C_g L_n$$

$$\tau_{\text{nor}} = R_{\text{eff}} C_{\text{in}} = R_{\text{eqn}} \left(\frac{L_n}{W_n} \right) 5 W_n C_g = 5 R_{\text{eqn}} C_g L_n$$

$$\text{total\_delay} = \sum_j \tau_{\text{nand}} \left(\frac{C_{j+1}}{C_j} + \gamma_{\text{nand}} \right)$$

Optimum Fan-out for CMOS Gates



$$\text{total\_delay} = \tau_{\text{nand}}\left(\frac{C_{j+1}}{C_j} + \gamma_{\text{nand}}\right) + \tau_{\text{inv}}\left(\frac{C_{j+2}}{C_{j+1}} + \gamma_{\text{inv}}\right) + \tau_{\text{nor}}\left(\frac{C_{j+3}}{C_{j+2}} + \gamma_{\text{nor}}\right)$$

The delay through stages j and j + 1 is given by

$$D_{j+1} = \tau_{\text{nand}}\left(\frac{C_{j+1}}{C_j} + \gamma_{\text{nand}}\right) + \tau_{\text{inv}}\left(\frac{C_{j+2}}{C_{j+1}} + \gamma_{\text{inv}}\right)$$

For minimum delay, we take the derivative with respect to C_{j+1}:

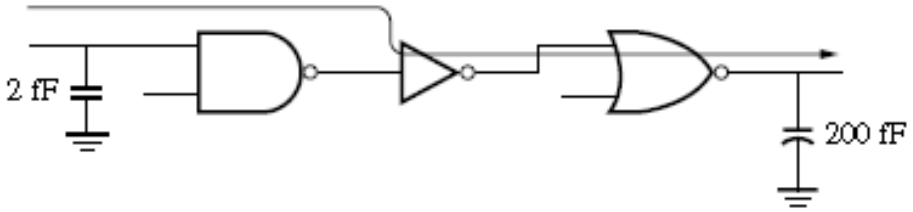
$$\frac{\partial D_{j+1}}{\partial C_{j+1}} = \tau_{\text{nand}}\left(\frac{1}{C_j}\right) - \tau_{\text{inv}}\left(\frac{C_{j+2}}{C_{j+1}^2}\right) = 0$$

$$\therefore \tau_{\text{nand}}\left(\frac{C_{j+1}}{C_j}\right) = \tau_{\text{inv}}\left(\frac{C_{j+2}}{C_{j+1}}\right)$$

$$\therefore \tau_{\text{nand}} FO_j = \tau_{\text{inv}} FO_{j+1}$$

Optimum Fan-out for CMOS Gates

Example: Find the device sizes that optimize the delay through the indicated path for the circuit below.



Optimum Fan-out for CMOS Gates

We must equalize the fanout portion of the delay. Therefore,

$$\therefore \tau_{nand}\left(\frac{C_{j+1}}{C_{in}}\right) = \tau_{inv}\left(\frac{C_{j+2}}{C_{j+1}}\right) = \tau_{nor}\left(\frac{C_{load}}{C_{j+2}}\right)$$

We take the product of these three components and then obtain the geometric mean:

$$\text{Fanout\_delay} = \sqrt[3]{\tau_{nand}\left(\frac{C_{j+1}}{C_{in}}\right) \times \tau_{inv}\left(\frac{C_{j+2}}{C_{j+1}}\right) \times \tau_{nor}\left(\frac{C_{load}}{C_{j+2}}\right)}$$

$$= \sqrt[3]{\tau_{nand} \times \tau_{inv} \times \tau_{nor}\left(\frac{C_{load}}{C_{in}}\right)} = \sqrt[3]{4 \times 3 \times 5\left(\frac{200}{2}\right) \times R_{eqn}C_gL_n}$$

$$= 18.2\, R_{eqn}C_gL_n$$

Therefore, the input capacitance for each gate can be computed by setting the fanout delay to the above result:

$$\tau_{nor}\left(\frac{C_{load}}{C_{j+2}}\right) = 5R_{eqn}C_gL_n\left(\frac{200\text{ fF}}{C_{j+2}}\right) = 18.2\, R_{eqn}C_gL_n$$

$$\therefore C_{j+2} = 55\text{ fF}$$

$$\tau_{inv} = \left(\frac{C_{j+2}}{C_{j+1}}\right) = 3R_{eqn}C_gL_n\left(\frac{55\text{ fF}}{C_{j+1}}\right) = 18.2\, R_{eqn}C_gL_n$$

$$\therefore C_{j+1} = 9.1\text{ fF}$$

$$\tau_{nand}\left(\frac{C_{j+1}}{C_{in}}\right) = 4\, R_{eqn}C_gL_n\left(\frac{9.1\text{ fF}}{C_{in}}\right) = 18.2\, R_{eqn}C_gL_n$$

$$\therefore C_{in} = 2\text{ fF}$$