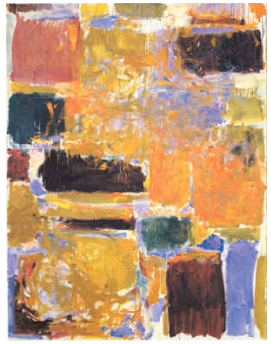
CMOS Inverter: Power Dissipation and Sizing

Professor Chris H. Kim

University of Minnesota
Dept. of ECE

chriskim@umn.edu

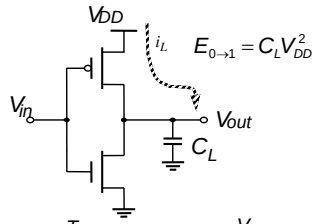
CMOS Inverter Power Dissipation



Where Does Power Go in CMOS?

- **Switching power**
  - **Charging capacitors**
- **Leakage power**
  - **Transistors are imperfect switches**
- **Short-circuit power**
  - **Both pull-up and pull-down on during transition**
- **Static currents**
  - **Biasing currents, in e.g. memory**

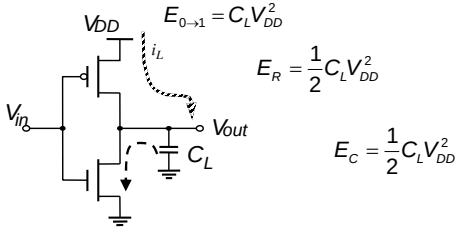
Dynamic Power Consumption



$$E_{0\to1} = \int_0^T P_{DD}(t)dt = V_{DD}\int_0^T i_{DD}(t)dt = V_{DD}\int_0^{V_{DD}} C_L dv_{out} = C_L V_{DD}^2$$

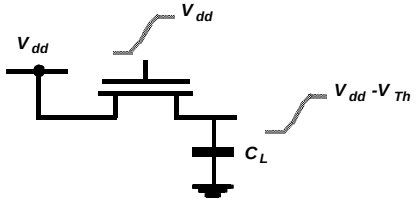
$$E_C = \int_0^T P_C(t)dt = \int_0^T v_{out} i_L(t)dt = \int_0^{V_{DD}} C_L v_{out} dv_{out} = \frac{1}{2} C_L V_{DD}^2$$

Dynamic Power Consumption



- One half of the energy from the supply is consumed in the pull-up network and one half is stored on C_L
- Energy from C_L is dumped during the 1→0 transition

Circuits with Reduced Swing



$$E_{0\to1} = C_L V_{DD} \left(V_{DD} - V_{Th} \right)$$

Dynamic Power Consumption

Power = Energy/transition • Transition rate

$$= C_L V_{DD}^2 \bullet f_{0\to1}$$

$$= C_L V_{DD}^2 \bullet f \bullet P_{0\to1}$$

$$= C_{switched} V_{DD}^2 \bullet f$$

- Power dissipation is data dependent – depends on the switching probability
- Switched capacitance $C_{switched} = C_L \bullet P_{0\to1}$

Transition Activity and Power

- Energy consumed in N cycles, E_N:

$$E_N = C_L \bullet V_{DD}^2 \bullet n_{0\to1}$$

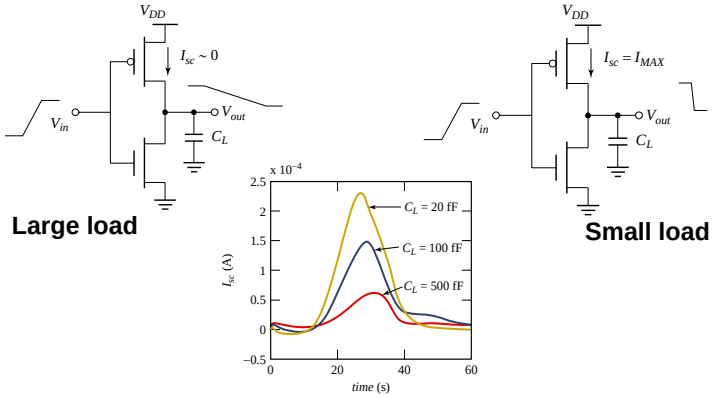
$n_{0\to1}$ – number of 0→1 transitions in N cycles

$$P_{avg} = \lim_{N\to\infty} \frac{E_N}{N} \cdot f = \left(\lim_{N\to\infty} \frac{n_{0\to1}}{N} \right) \cdot C_L \cdot V_{DD}^2 \cdot f$$

$$\alpha_{0\to1} = \lim_{N\to\infty} \frac{n_{0\to1}}{N} \cdot f$$

$$P_{avg} = \alpha_{0\to1} \cdot C_L \cdot V_{DD}^2 \cdot f$$

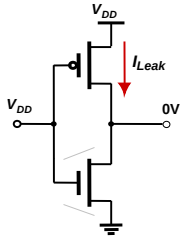
Short Circuit Current



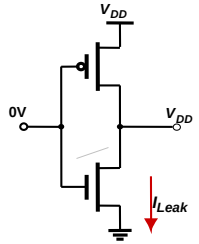
- **Short circuit current is usually well controlled**

Transistor Leakage

- **Transistors that are supposed to be off - leak**

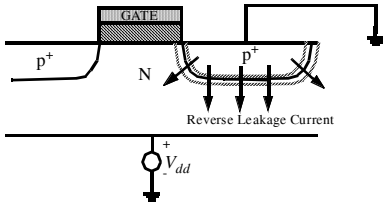


Input at V_{DD}



Input at 0

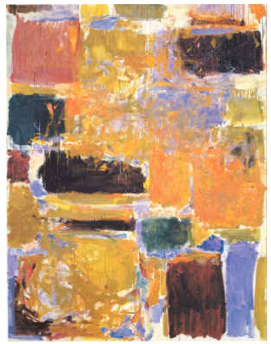
Diode Leakage



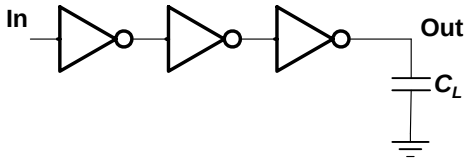
$$I_{DL} = J_S \times A$$

***JS* = 10-100 pA/mm2 at 25 deg C for 0.25um CMOS**
***JS* doubles for every 9 deg C!**
Much smaller than transistor leakage in deep submicron

Sizing of an Inverter Chain



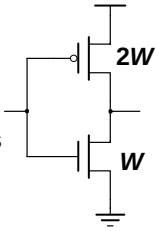
Inverter Chain



If C_L is given:
- How many stages are needed to minimize the delay?
- How to size the inverters?
May need some additional constraints.

Inverter Delay

- **Minimum length devices, L=0.25um**
- **Assume that for $W_P = 2W_N = 2W$**
  - **same pull-up and pull-down currents**
  - **approx. equal resistances $R_N = R_P$**
  - **approx. equal rise t_{pLH} and fall t_{pHL} delays**
- **Analyze as an RC network**

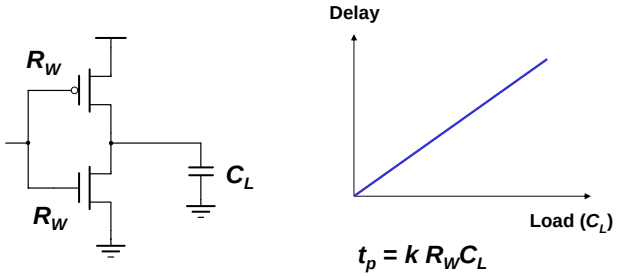


$$R_P = R_{unit}\left(\frac{W_P}{W_{unit}}\right)^{-1} \approx R_{unit}\left(\frac{W_N}{W_{unit}}\right)^{-1} = R_N = R_W$$

Delay (D): t_{pHL} = (ln 2) $R_N C_L$ **t_{pLH} = (ln 2) $R_P C_L$**

Load for the next stage: $C_{gin} = 3\frac{W}{W_{unit}} C_{unit}$

Inverter with Load

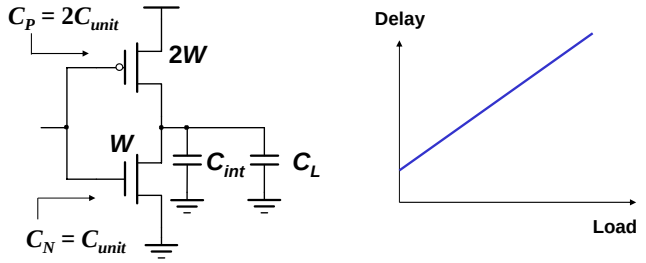


$t_p = k\, R_W C_L$

k is a constant, equal to 0.69 for step input
Assumptions: no load -> zero delay

$W_{unit} = 1$

Inverter with Load



Delay = $kR_W(C_{int} + C_L) = kR_W C_{int} + kR_W C_L = kR_W C_{int}(1 + C_L / C_{int})$
= Delay (Intrinsic) + Delay (Load)

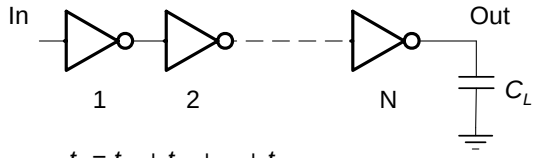
Delay Formula

$$Delay \sim R_W (C_{int} + C_L)$$

$$t_p = kR_W C_{int}(1 + C_L / C_{int}) = t_{p0}(1 + f/\gamma)$$

$C_{int} = \gamma C_{gin}$ with $\gamma \approx 1$
$f = C_L/C_{gin}$ - effective fanout
$R = R_{unit}/W$; $C_{int} = WC_{unit}$
$t_{p0} = 0.69 R_{unit} C_{unit}$

Apply to Inverter Chain



$$t_p = t_{p1} + t_{p2} + \ldots + t_{pN}$$

$$t_{pj} \sim R_{unit} C_{unit}\left(1 + \frac{C_{gin,j+1}}{C_{gin,j}}\right)$$

$\gamma = 1$

$$t_p = \sum_{j=1}^{N} t_{p,j} = t_{p0} \sum_{i=1}^{N}\left(1 + \frac{C_{gin,j+1}}{C_{gin,j}}\right), \quad C_{gin,N+1} = C_L$$

Optimal Tapering for Given *N*

Delay equation has *N* - 1 unknowns, $C_{gin,2}$ – $C_{gin,N}$

Minimize the delay, find *N* - 1 partial derivatives

Result: $C_{gin,j+1}/C_{gin,j} = C_{gin,j}/C_{gin,j-1}$

Size of each stage is the geometric mean of two neighbors

$$C_{gin,j} = \sqrt{C_{gin,j-1} C_{gin,j+1}}$$

- **each stage has the same effective fanout (C_{out}/C_{in})**
- **each stage has the same delay**

Optimum Delay and Number of Stages

When each stage is sized by *f* and has same eff. fanout *f*:

$$f^N = F = C_L / C_{gin,1}$$

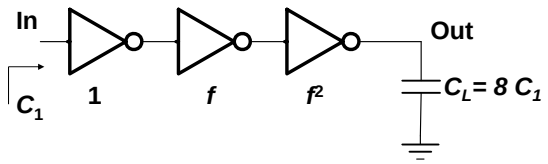
Effective fanout of each stage:

$$f = \sqrt[N]{F}$$

Minimum path delay

$$t_p = N t_{p0}\left(1 + \sqrt[N]{F}\right)$$

Example



C_L/C_1 has to be evenly distributed across N = 3 stages:

$$f = \sqrt[3]{8} = 2$$

Optimum Number of Stages

For a given load, C_L and given input capacitance C_{in}
Find optimal number of stages, N, and optimal sizing, f

$$C_L = F \cdot C_{in} = f^N C_{in} \quad \text{with} \quad N = \frac{\ln F}{\ln f}$$

$$t_p = N t_{p0}\left(F^{1/N}/\gamma + 1\right) = \frac{t_{p0} \ln F}{\gamma}\left(\frac{f}{\ln f} + \frac{\gamma}{\ln f}\right)$$

$$\frac{\partial t_p}{\partial f} = \frac{t_{p0} \ln F}{\gamma} \cdot \frac{\ln f - 1 - \gamma/f}{\ln^2 f} = 0$$

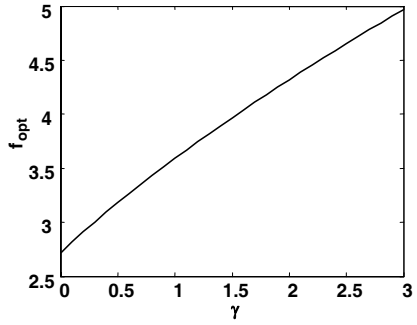
For $\gamma = 0$, $f = e$, $N = \ln F$

$$f = \exp\left(1 + \gamma/f\right)$$

Optimum Effective Fanout *f*

Optimum f for given process defined by γ

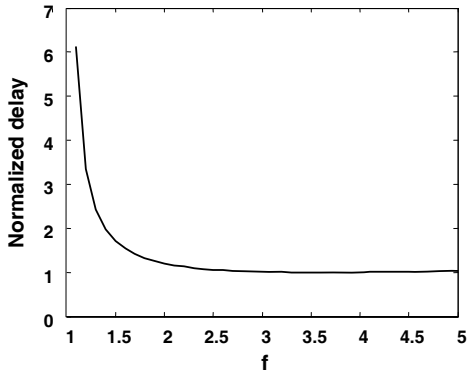
$$f = \exp\left(1 + \gamma/f\right)$$



f_{opt} = 3.6 for γ=1

Impact of Loading on *tp*

With self-loading γ=1



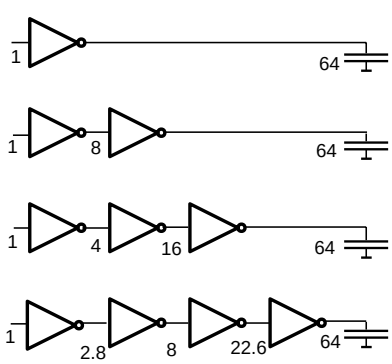
Normalized Delay Function of *F*

$$t_p = Nt_{p0}\left(1+\sqrt[N]{F}/\gamma\right)$$

$$t_p = Nt_{p0}\left(1+\sqrt[N]{F}\right), \text{with } \gamma = 1$$

F	Unbuffered	Two Stage	Inverter Chain
10	11	8.3	8.3
100	101	22	16.5
1000	1001	65	24.8
10,000	10,001	202	33.1

Buffer Design



N	f	t_p
1	64	65
2	8	18
3	4	15
4	2.8	15.3