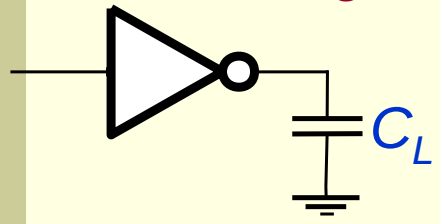
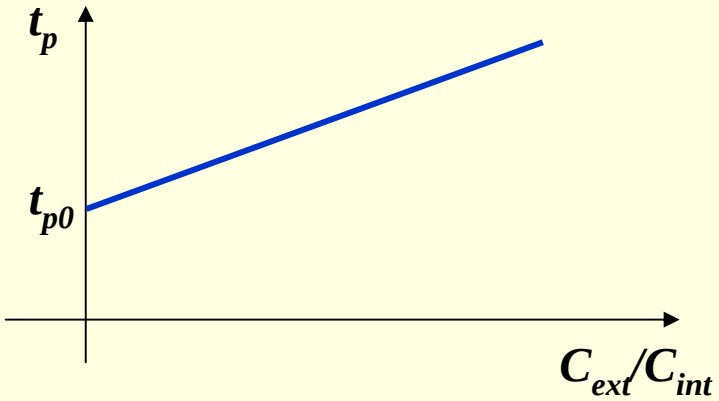
Dynamic operation – 20

A simple model for the propagation delay

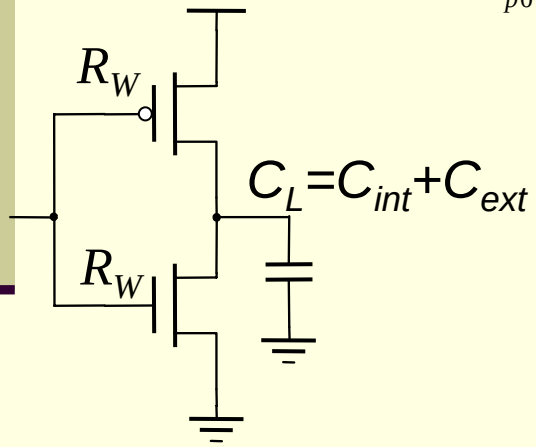
- Symmetric inverter (rise and fall delays are identical)
- Total capacitance is linear
- Minimum length devices



$$t_p = 0.69 R_W C_L =$$
$$0.69 R_W C_{int}(1 + C_{ext}/C_{int}) =$$
$$t_{p0}(1 + C_{ext}/C_{int})$$



Propagation delay vs. C_{ext}/C_{int} ratio



1. t_{p0} (for minimum-L devices) is independent of the sizing (W's) of the gate;
2. Making W infinitely large eliminates the impact of any external load

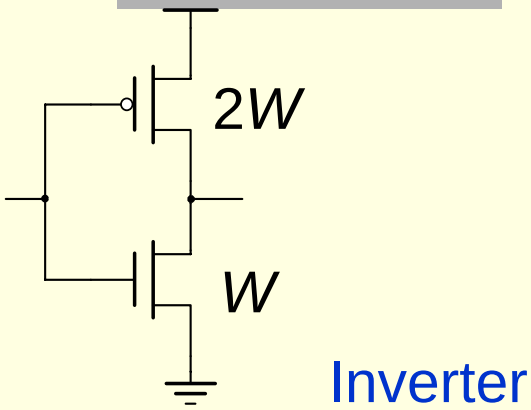
C_{int} – intrinsic (self-loading) output capacitance
C_{ext} – extrinsic load capacitance (fan-out + wiring)

Source: Rabaey

Dynamic operation – 21

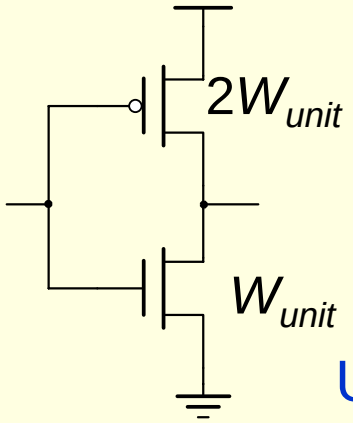
Inverter Delay

- Minimum length devices
- 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



Inverter

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



Unit inverter

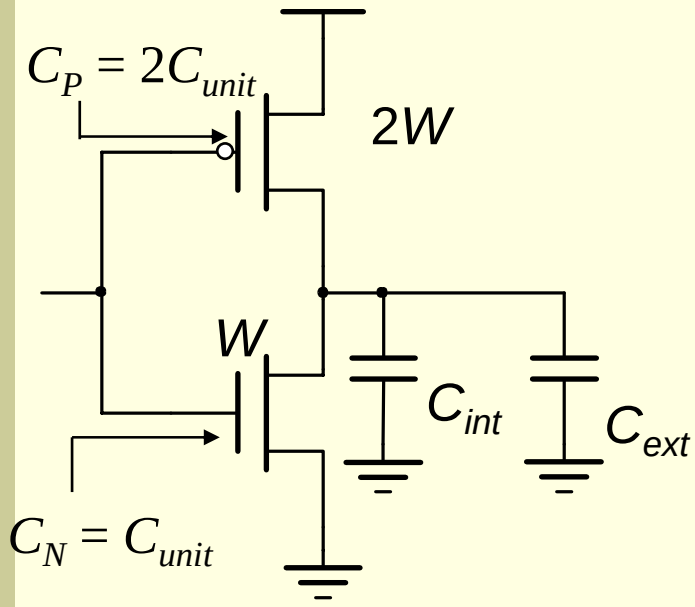
Delay (D): $t_{pHL} = (\ln 2)\, R_N C_L \quad t_{pLH} = (\ln 2)\, R_P C_L$

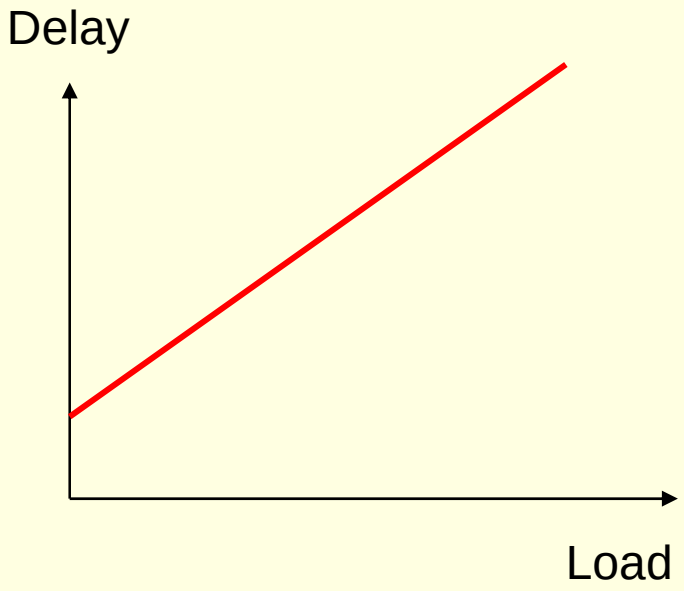
Load for previous stage: $C_{gin} = 3\frac{W}{W_{unit}} C_{unit}$

Source: Rabaey

Dynamic operation – 22

Inverter Delay





Delay = $0.69R_W(C_{int} + C_{ext}) = 0.69R_WC_{int} + 0.69R_WC_{ext} = 0.69R_W C_{int}(1+ C_{ext} /C_{int})$
= Delay (Internal) + Delay (Load)

Note:
$R_W \propto L/W$
$C_{int} \propto WL$

Source: Rabaey

Dynamic operation – 23

Delay Formula

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

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

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

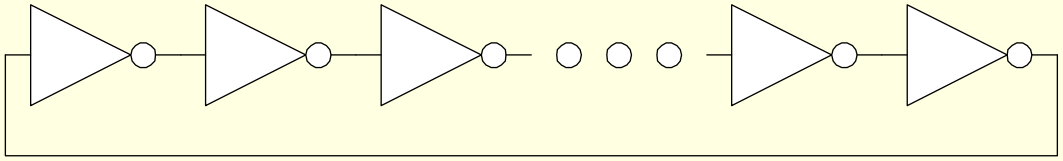
Note:
$R_W \propto L/W$
$C_{int} \propto WL$
$t_{p0} \propto L^2 \rightarrow$ minimum L for minimum delay

Source: Rabaey

Dynamic operation – 24

Ring oscillators - 1

N: (odd) number of inverters (usually >5)



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

$C_{int} = \gamma C_{gin}$ with $\gamma \approx 1$
$f = C_{ext}/C_{gin} = 1$

$$f_{osc} = \frac{1}{2Nt_p} \rightarrow t_p = \frac{1}{2Nf_{osc}}$$

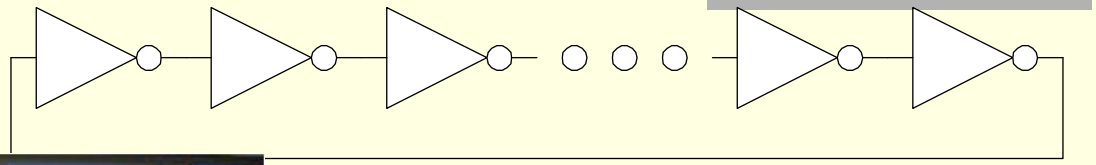
Ring oscillators are used as process monitors to verify if a chip is faster or slower than nominally expected.
Ex: 31-stage ring oscillator in a 180 nm process oscillates at 540 MHz.

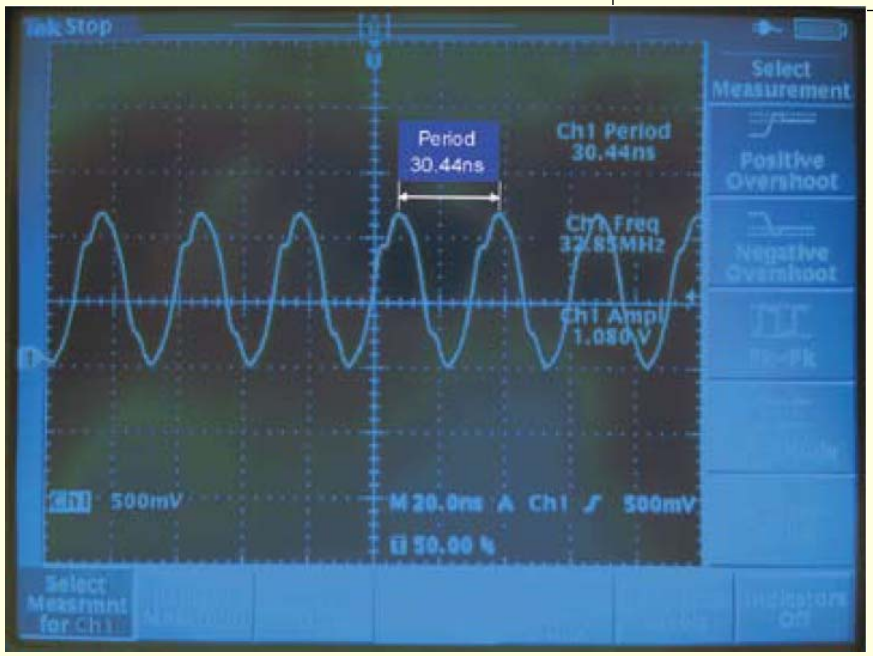
Sources: Rabaey
Weste

Dynamic operation – 25

Ring oscillators - 2

N: (odd) number of inverters (usually >5)





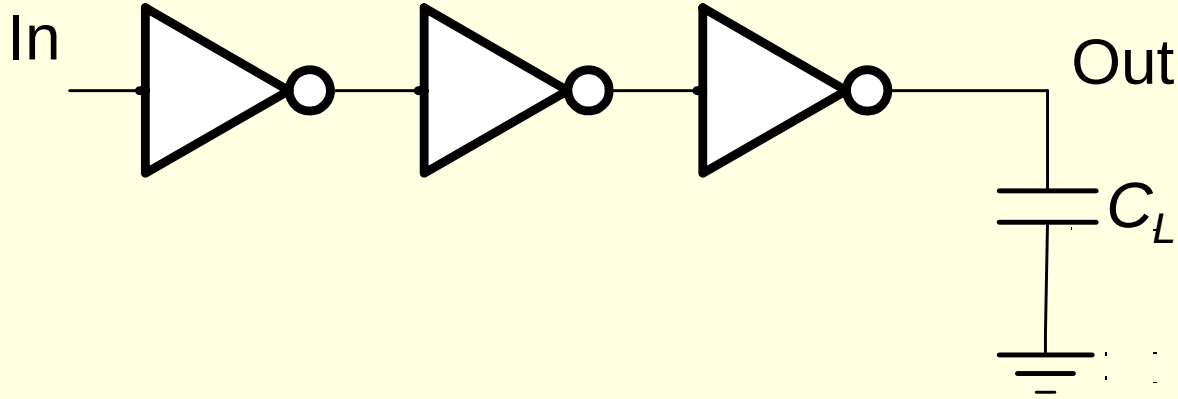
V_{DD}	0.8V	1.0V	1.2V
Frequency	17.2MHz	26.3MHz	33.4MHz

Ring oscillator output as viewed with an oscilloscope. The time base is set to 20ns per division and the voltage scale is set to 500mV per division. The measured wave period is 30.44ns and the measured amplitude is 1.08V when the supply voltage is 1.2V.

Source:
www.keithley.com.cn/data?asset=51070

Dynamic operation – 26

Inverter chain - 1



If C_L is given:

- How many stages are needed to minimize the delay?
- How to size the inverters?

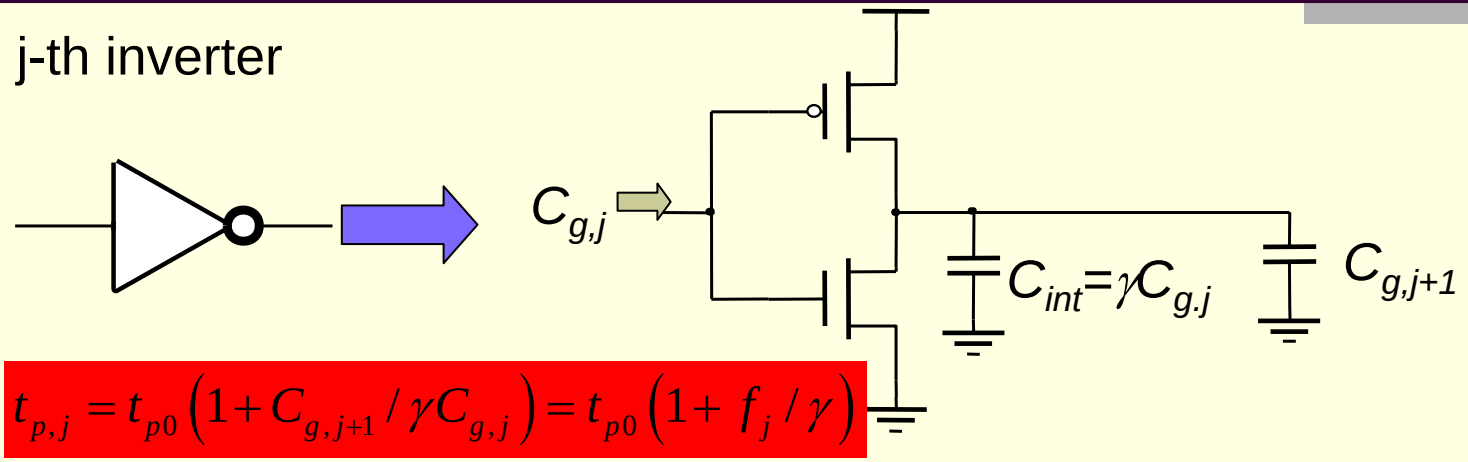
May need some additional constraints.

Source: Rabaey

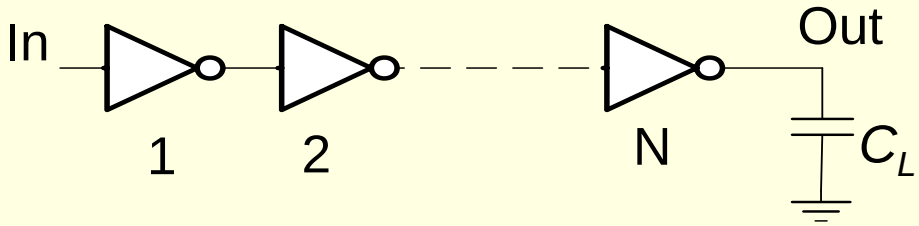
Dynamic operation – 27

Inverter chain - 2

j-th inverter



$$t_{p,j} = t_{p0}\left(1 + C_{g,j+1}/\gamma C_{g,j}\right) = t_{p0}\left(1 + f_j/\gamma\right)$$



$$t_p = \sum_{j=1}^{N} t_{p,j} = t_{p0}\sum_{j=1}^{N}\left(1 + C_{g,j+1}/\gamma C_{g,j}\right); \quad C_{g,N+1} = C_L$$

N-1 unknowns: $C_{g,2}$, $C_{g,3}$, ….$C_{g,N-1}$, $C_{g,N}$

First inverter is minimally sized

Source: Rabaey

Dynamic operation – 28

Inverter chain - 3

Let's minimize $t_p = \sum_{j=1}^{N} t_{p,j} = t_{p0} \sum_{j=1}^{N} \left(1 + C_{g,j+1}/\gamma C_{g,j}\right); \quad C_{g,N+1} = C_L$

Taking the N-1 derivatives partial derivatives and equating them to 0 we find that $C_{g,j} = \sqrt{C_{g,j-1} C_{g,j+1}}$

Thus, each inverter is sized up by the same factor f wrt the preceding gate

$$f = C_{g,j+1}/C_{g,j} = \sqrt[N]{C_L/C_{g,1}} = \sqrt[N]{F}$$

The minimum delay is

$t_p = t_{p0} \sum_{j=1}^{N} \left(1 + f/\gamma\right) = N t_{p0} \left(1 + \sqrt[N]{F}/\gamma\right)$ ⇒ *What´s N that minimizes delay?*

Source: Rabaey

Dynamic operation – 29

Inverter chain - 4

$$t_p = t_{p0}\sum_{j=1}^{N}\left(1+f/\gamma\right) = Nt_{p0}\left(1+\sqrt[N]{F}/\gamma\right)$$ → $dt_p/dN=0$

The minimum delay is for N obtained from →

$$\gamma + \sqrt[N]{F}\left(1-\ln F/N\right)=0$$

or, equivalently →

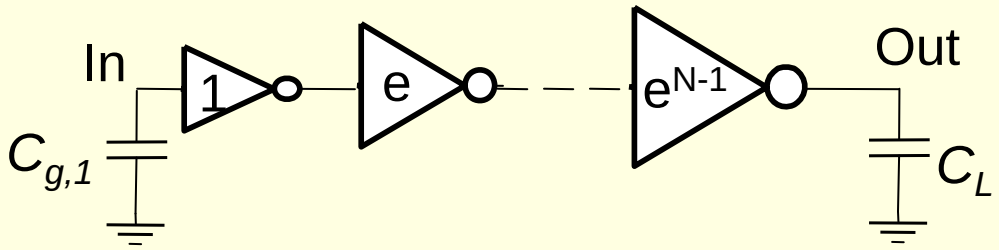
$$f = e^{(1+\gamma/f)}$$

Canonical case: $\gamma=0$ →

$$f = e, \qquad N = \ln F = \ln\left(C_L/C_{g,1}\right)$$

$$t_p = et_{pi}\ln\left(C_L/C_{g,1}\right)$$

t_{pi}: propagation delay of unit inverter loaded with another unit inverter

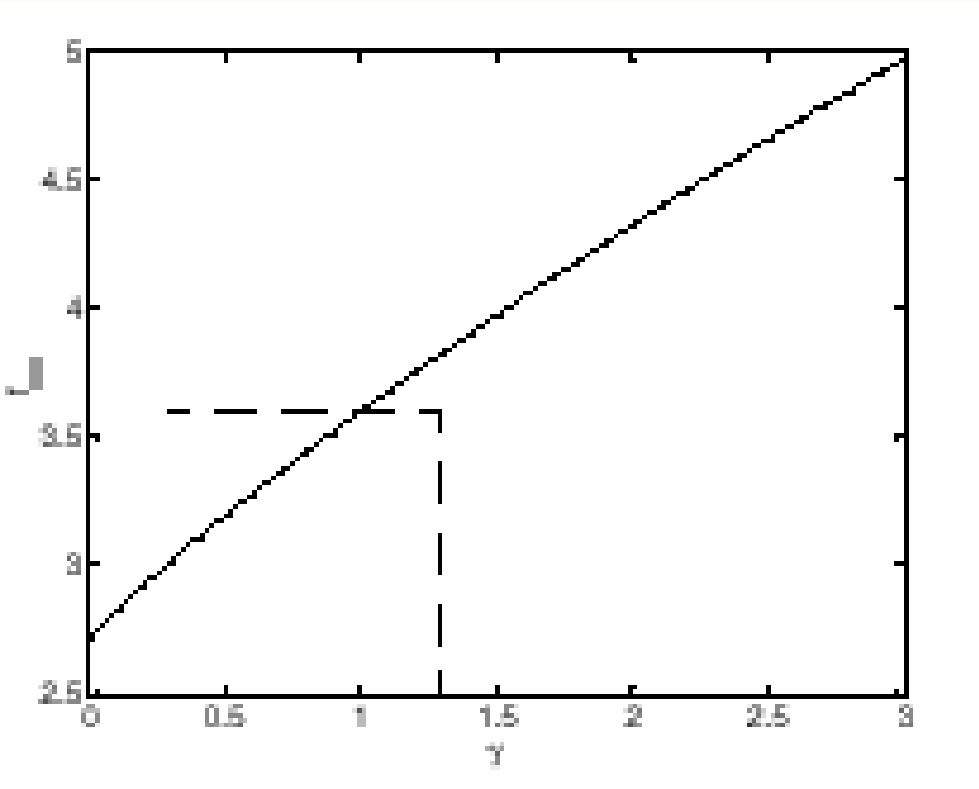


Source: Rabaey

Dynamic operation – 30

Inverter chain - 5

Optimum effective fan-out f $\quad f = \exp(1+\gamma/f)$



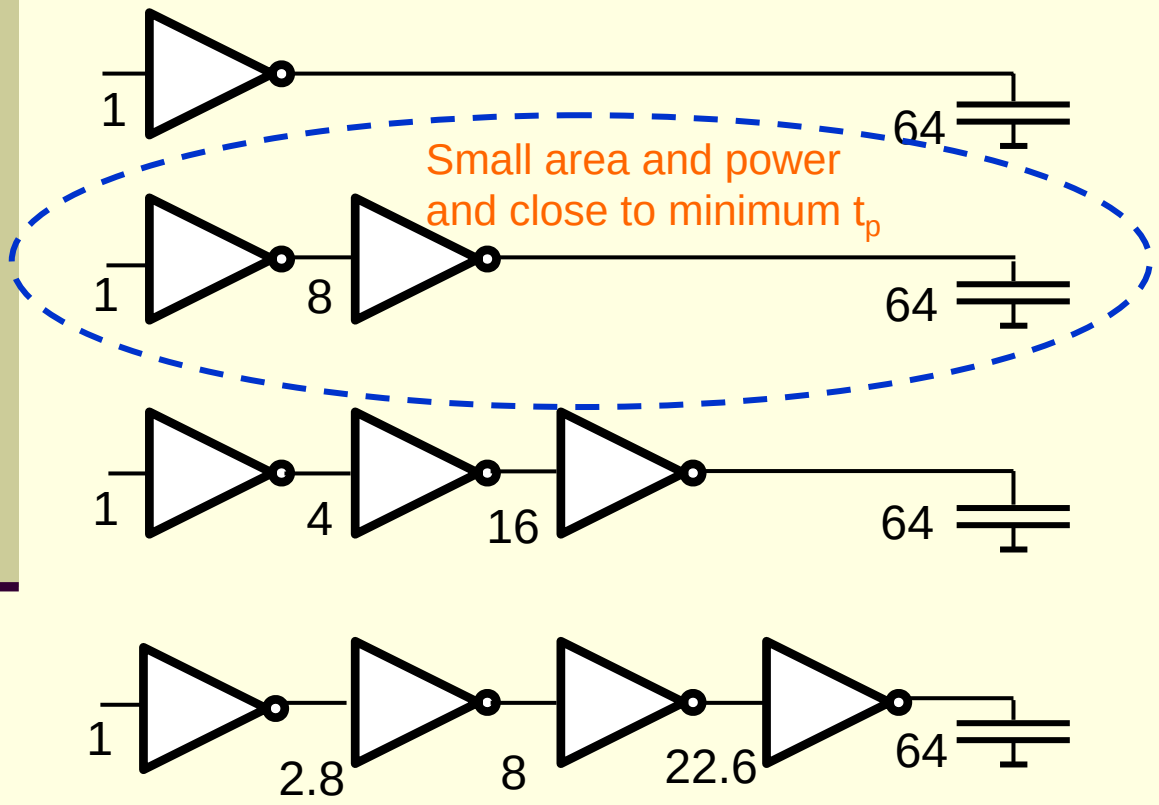
f_{opt} = 3.6
for γ=1

Source: Rabaey

Dynamic operation – 31

Inverter chain - 6

Buffer Design



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

*** Values normalized to t_{po}**

Sources: Weste and Rabaey

Power, energy, and energy delay – 1

- Power is drawn from a voltage source attached to the V_{DD} pin(s) of a chip.

- Instantaneous Power: $p(t) = i(t)v(t)$

  Delivered by the power source $p(t) = i_{DD}(t)V_{DD}$

- Energy: $E = \int_0^T p(t)dt = \int_0^T i(t)v(t)dt$

- Average Power: $P_{\text{avg}} = \frac{E}{T} = \frac{1}{T}\int_0^T i(t)v(t)dt$

Source: Weste

Power, energy, and energy delay – 2

Where Does Power Go in CMOS?

- **Dynamic Power Consumption**

  Charging and Discharging Capacitors

- **Short Circuit Currents**

  Short Circuit Path between Supply Rails during Switching

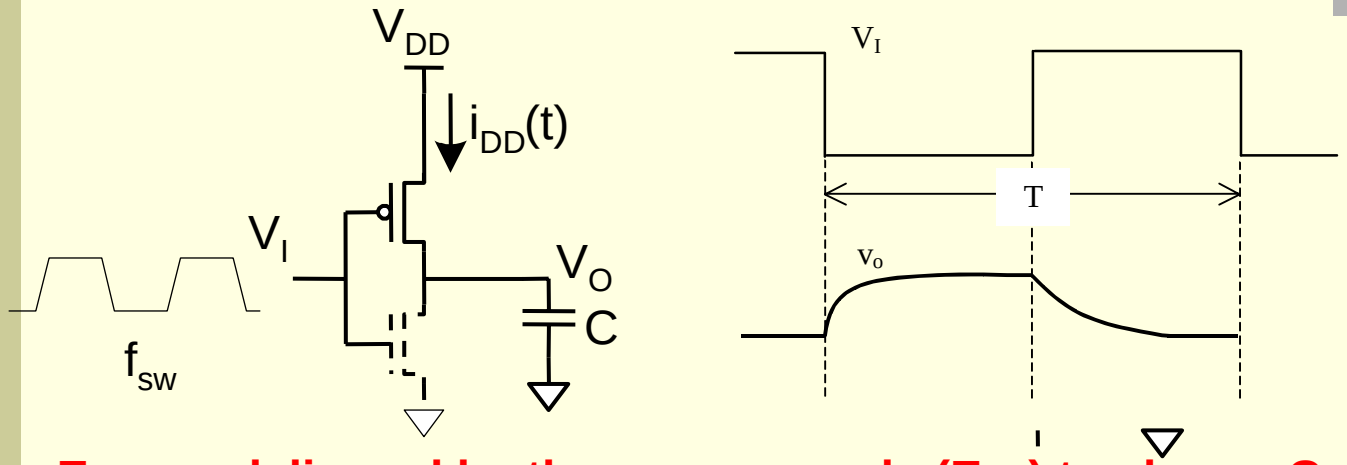
- **Leakage**

  Leaking diodes and transistors

Source: Rabaey

Power, energy, and energy delay – 3

Dynamic Power Dissipation



Energy delivered by the power supply (E_{DD}) to charge C

$$E_{\mathrm{DD}} = \int_0^T i_{DD}(t) V_{DD}\, dt = V_{DD} \int_0^T i_{DD}(t)\, dt = V_{DD} \int_0^{V_{DD}} C\, dv_o = C V_{DD}^{\ 2}$$

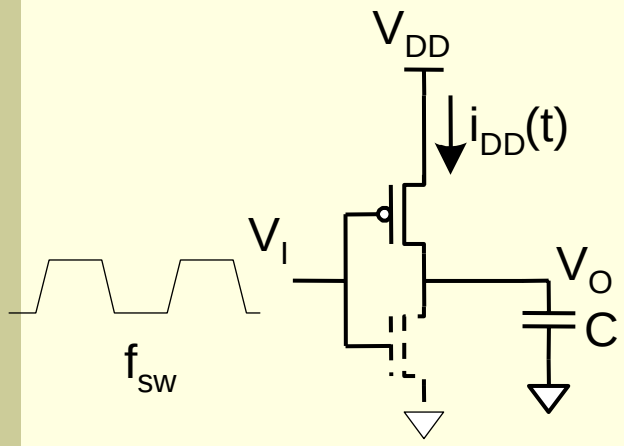
The energy stored in the fully charged capacitor is

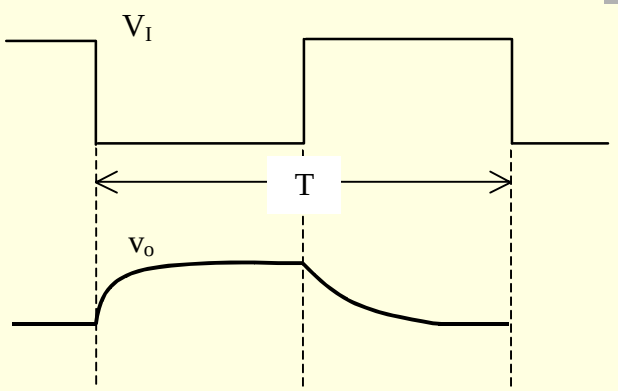
$$E_C = \int_0^T v_o\ i_C\, dt = \int_0^{V_{DD}} v_o\ C\, dv_o = C\frac{V_{DD}^2}{2}$$

Where´s the other half of the energy delivered by V_{DD}?

Power, energy, and energy delay – 4

Dynamic Power Dissipation





Where´s the energy delivered by V_{DD}?

One half of the energy is stored in C whereas the other half is converted into heat in the pull-up network

$$E\left(M_p\right) = CV_{DD}^{\ 2}/2$$

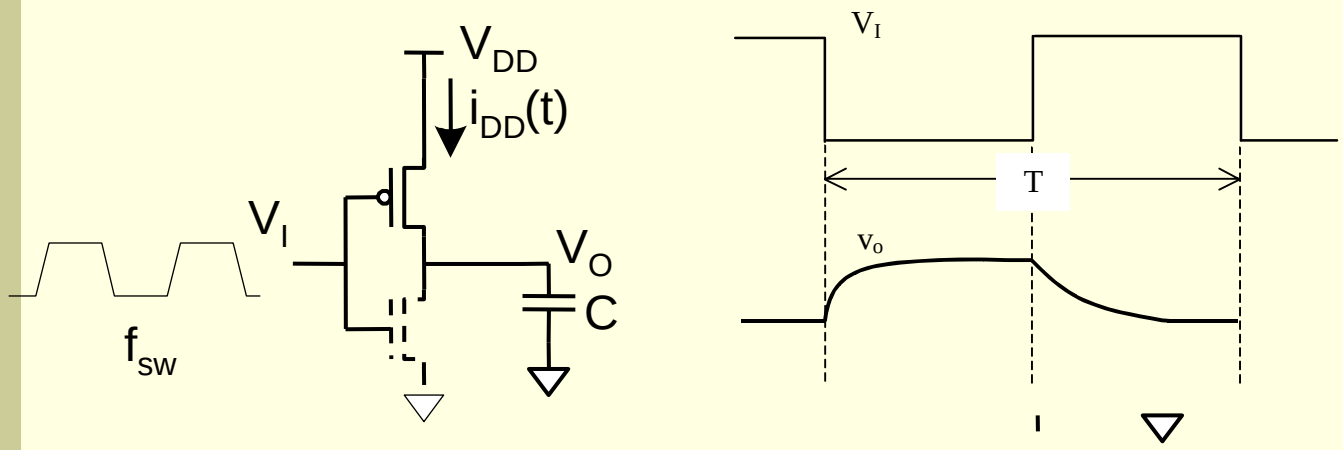
$$E\left(M_N\right) = CV_{DD}^{\ 2}/2$$

HEAT

During the 1 →0 transition of the output, the energy stored on C is dissipated into the n-channel transistor

Power, energy, and energy delay – 5

Dynamic Power Dissipation



$$P_{\text{dynamic}} = \frac{1}{T}\int_0^T i_{DD}(t)V_{DD}\,dt = \frac{V_{DD}}{T}\int_0^T i_{DD}(t)\,dt$$

$$= \frac{V_{DD}}{T}\left[Tf_{\text{sw}}CV_{DD}\right] = CV_{DD}{}^2 f_{\text{sw}}$$

clock frequency = f_{ck}

$f_{sw} = \alpha f_{ck}$, $\alpha \rightarrow$ activity factor

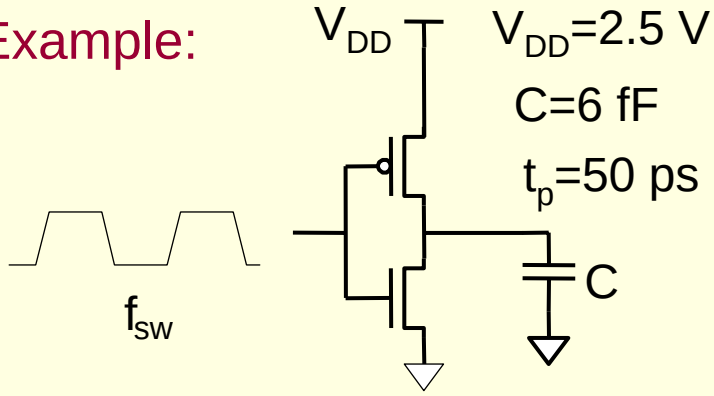
For low power reduce C, V_{DD}, and f_{sw}

Source: Weste

Power, energy, and energy delay – 6

Dynamic Power Dissipation

Example:



Energy delivered by the power supply (E_{DD}) to charge C

$$E_{DD} = CV_{DD}^{\ 2} = 6 \cdot 2.5^2 = 37.5 \text{ fJ}$$

Assume that $f_{ck}=1/4t_p=5$ GHz

For $f_{sw}=f_{ck}=5$ GHz, the average dynamic power dissipation is

$$P_{\text{dynamic}} = CV_{DD}^{\ 2} f_{sw} = 37.5 \cdot 5 \text{ fJ} \cdot \text{GHz} = 187.5 \ \mu\text{W}$$

For an activity factor of 0.1, the average dynamic power dissipation is ~ 18 μW

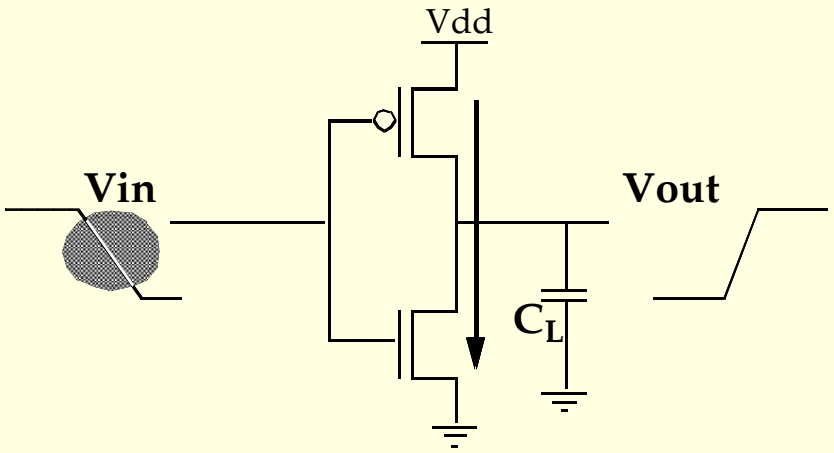
One million identical inverters with the same activity factor of 0.1 would give a total power dissipation of ~ 18 W

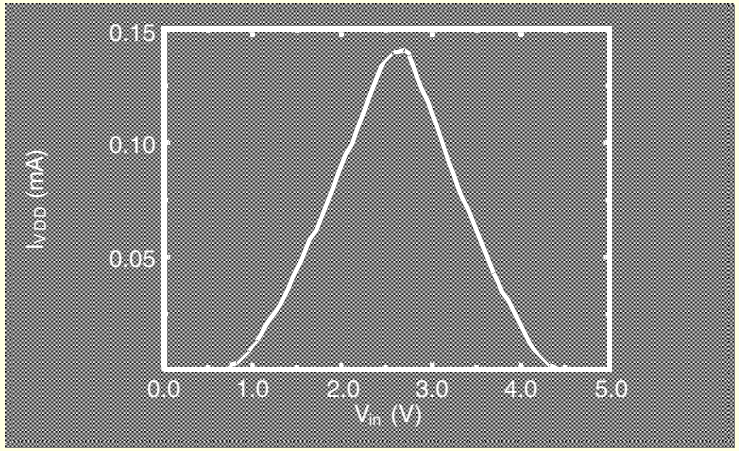
Source: Weste

Power, energy, and energy delay – 7

Short Circuit Currents

- When transistors switch, both nMOS and pMOS transistors may be momentarily ON at once
- Typically < 10% of dynamic power if rise/fall times are comparable for input and output

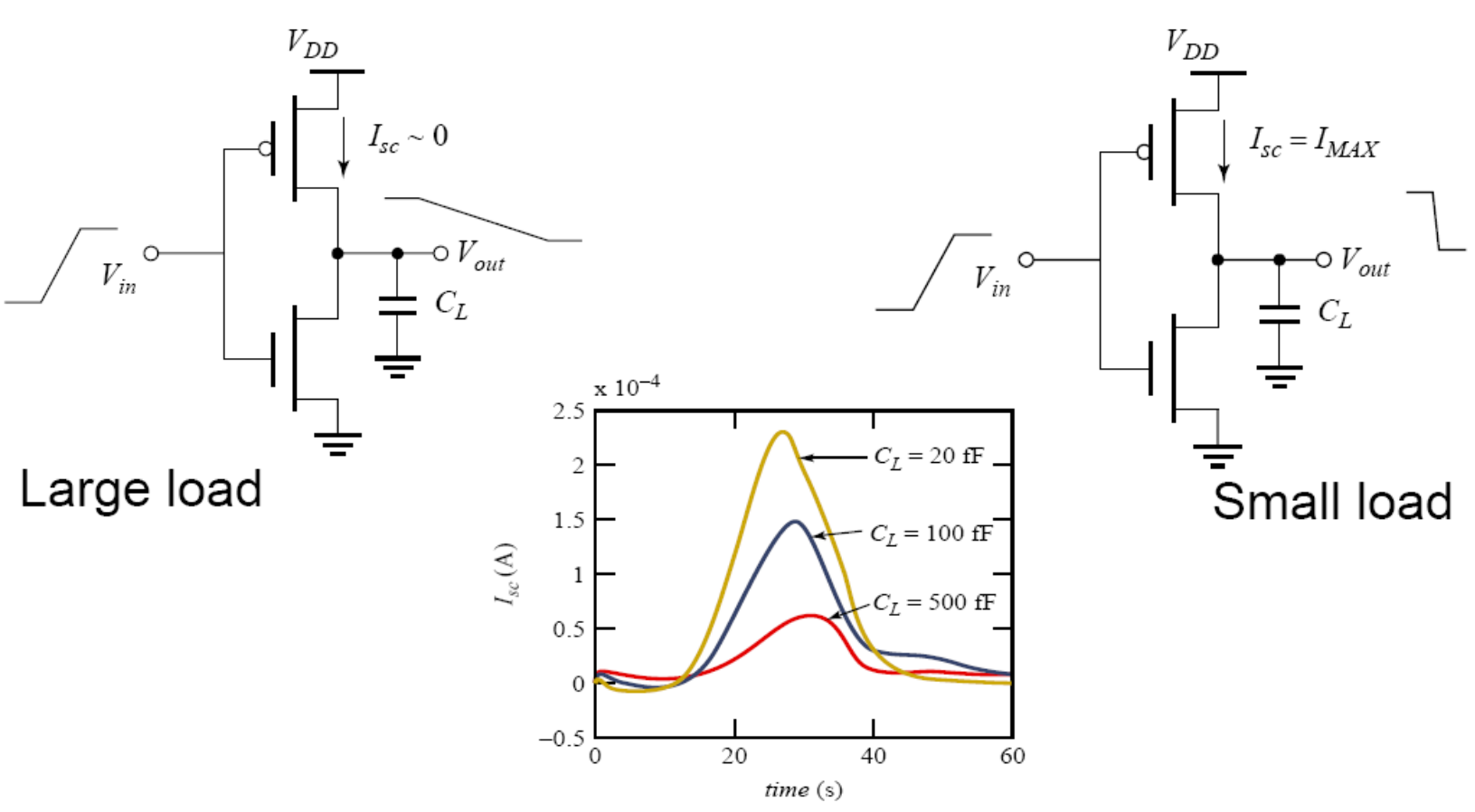




Source: Rabaey

Power, energy, and energy delay – 8

Short Circuit Currents

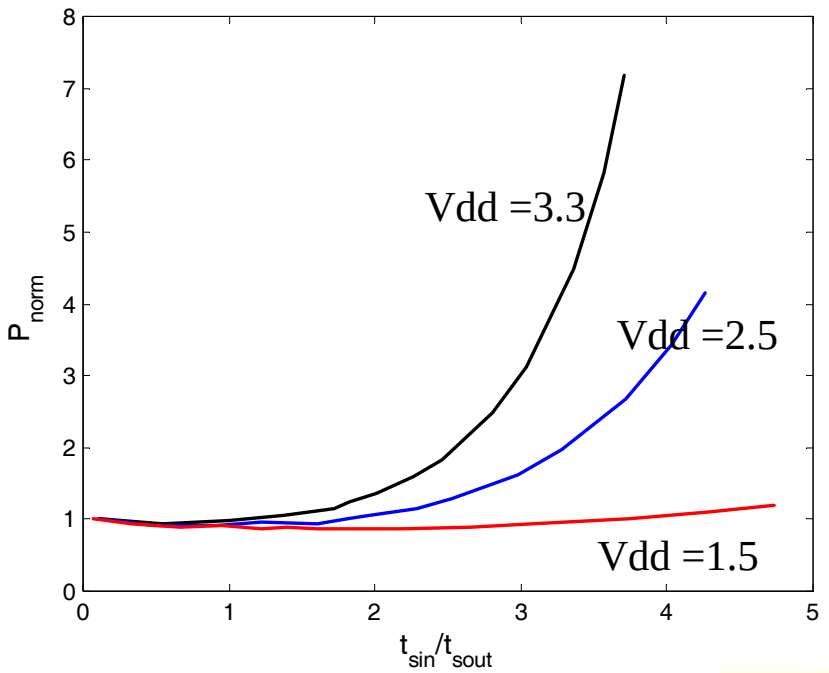


Source: Rabaey

Power, energy, and energy delay – 9

Short Circuit Currents

Minimizing Short-Circuit Power

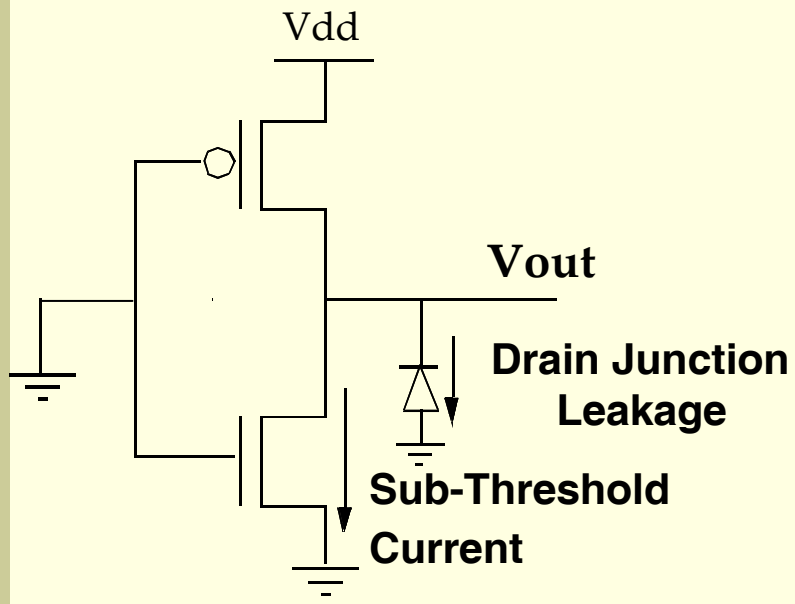


- Keep the input and output rise/fall times the same
  (< 10% of Total Consumption)
  from [Veendrick84]
  (*IEEE Journal of Solid-State Circuits*, August 1984)
- If $V_{dd} < V_{tn} + |V_{tp}|$ then short-circuit power can be *eliminated*!

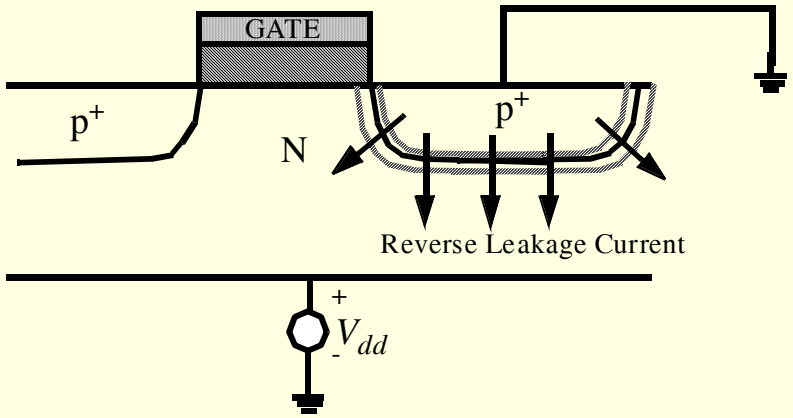
Source: Rabaey

Power, energy, and energy delay – 10

Leakage → static dissipation



Reverse-Biased Diode Leakage



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

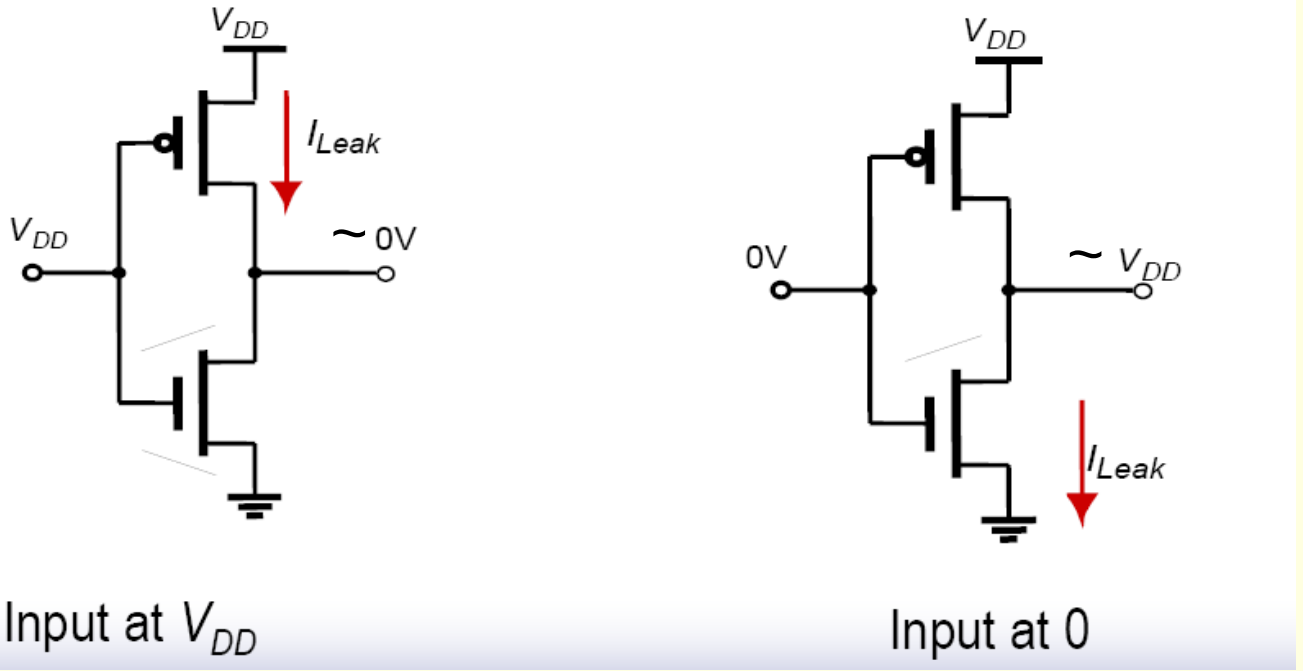
Sub-threshold current one of most compelling issues in low-energy circuit design!

Source: Rabaey

Power, energy, and energy delay – 11

Leakage → static dissipation

Transistors that are supposed to be off - leak

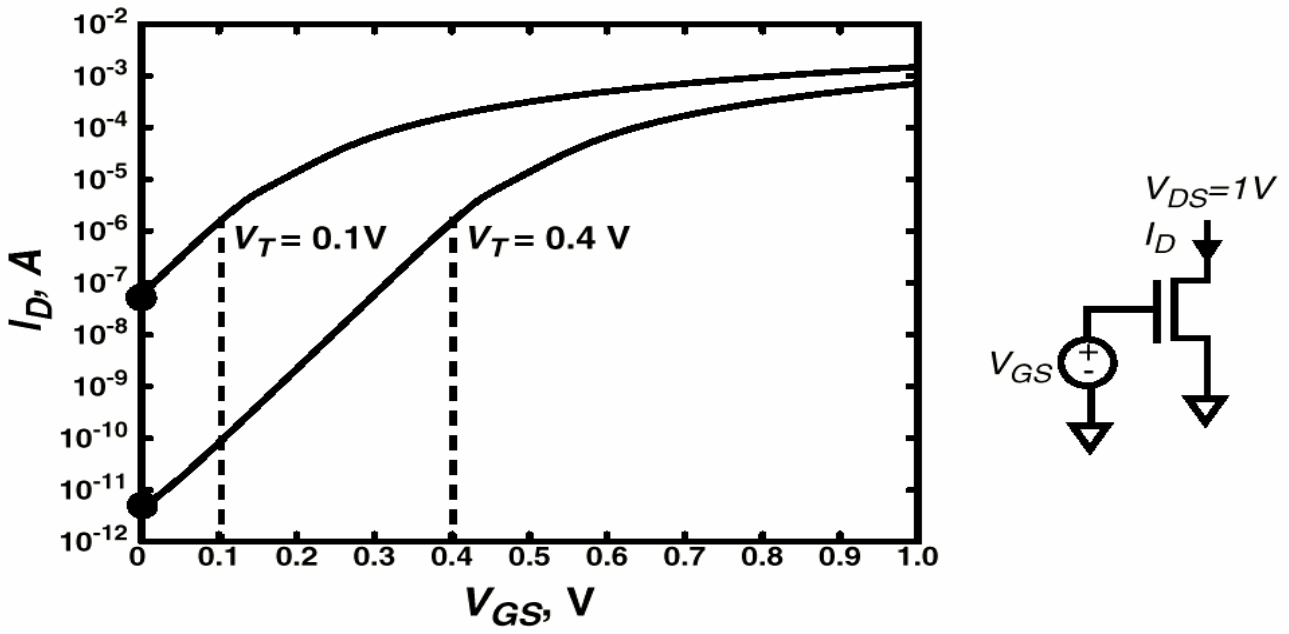


Source: Rabaey

Power, energy, and energy delay – 12

Leakage → static dissipation

Subthreshold Leakage Component



- Leakage control is critical for low-voltage operation

In our simplified model, currents for V_{GS} below V_T were assumed to be zero. However, subthreshold current is very important, especially for advanced technologies (low V_T's) and can be the major component of power dissipation.

Source: Rabaey

Power, energy, and energy delay – 13

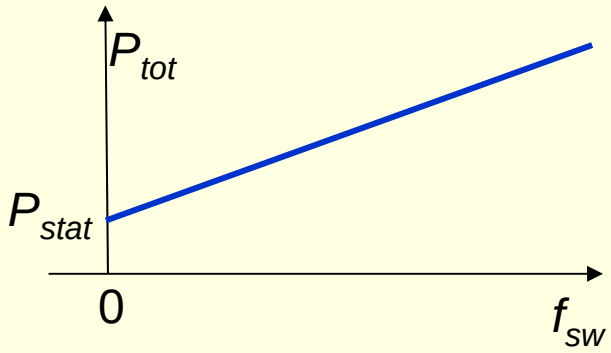
Leakage → static dissipation

Static power is due to the current that flows between supply rails in the absence of switching activity $P_{\text{stat}} = I_{stat} V_{DD}$

$$P_{\text{tot}} = P_{stat} + P_{dyn}$$

Diode leakage + subthreshold current

Charge/discharge Cs + short-circuit current



Source: Rabaey

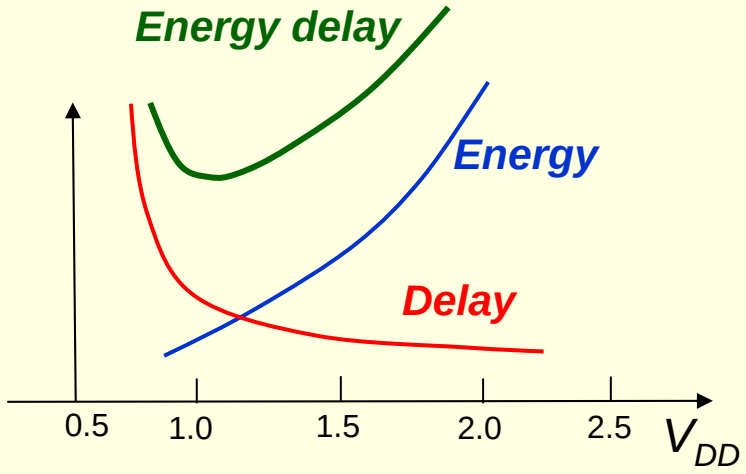
Power, energy, and energy delay – 14

Power-Delay Product (PDP) =Average energy consumed per switching event (0→1 or 1→0)

$$PDP = \frac{CV_{DD}^2}{2}$$

Energy-Delay Product (EDP) = quality metric of gate = $E \times t_p$

$$EDP = PDP \cdot t_p = \frac{CV_{DD}^2}{2} t_p$$



Source: Rabaey