Transistor Sizing

So far, we have assumed that to get symmetric rise and fall times:

$$W_p \approx (2 \rightarrow 3) \times W_n$$

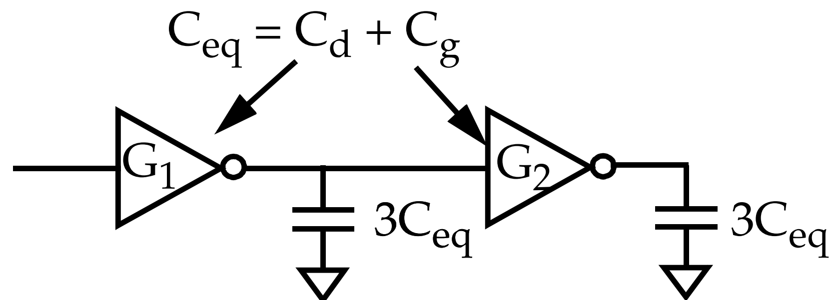
Does this rule reduce overall delay ?

Assume $W_p = 2W_n$

$$t_{inv-pair} = t_{fall} + t_{rise}$$

$$= R(1+2)C_{eq} + 2\left(\frac{R}{2}\right)(1+2)C_{eq}$$

$$= 6RC_{eq}$$

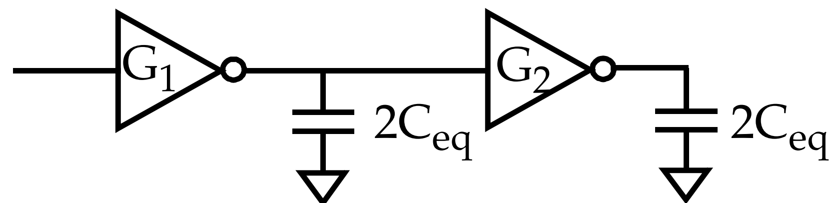


Half the resistance since p is twice as wide but twice the resistance per unit area.

Assume $W_p = W_n$

$$= R(1+1)C_{eq} + 2R(1+1)C_{eq}$$

$$= 6RC_{eq}$$

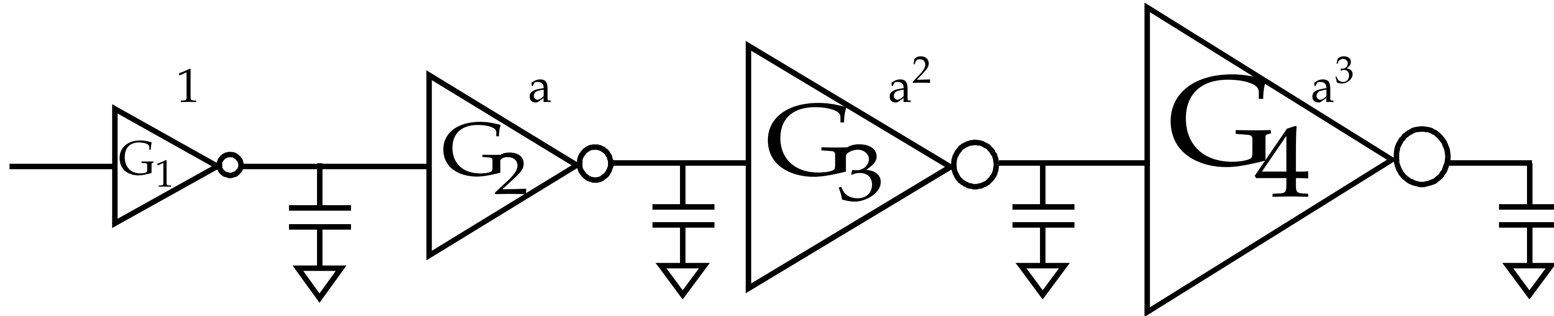


Therefore, in **self-loaded** circuits (circuits without significant routing capacitance and fanouts), *equal sized* devices can be used to reduce power dissipation and area without sacrificing performance (overall delay).

Stage Ratio

How do we drive large load capacitances, e.g. off-chip wires via the I/O pads, long buses, etc. ?

By using a chain of inverters, where each successive inverter is larger than the previous one.



What is the optimal value of *a* (the stage ratio) that both

- Minimizes the delay through the chain.
- Minimizes the area and power.

The magic number *a* is *e* (~2.7) - see analysis in book.
The optimal value may vary depending on process parameters.

Power Dissipation

Two components of power dissipation in CMOS circuits:

- Static power
- Dynamic power

Static power dissipation:

- Reverse-bias leakage current through parasitic diodes formed by source/drain diffusion and n-well diffusion.
- Through-current of pseudo-nMOS devices.
- Subthreshold conduction (current that flows when $V_{in} < V_{tn}$).

  Becoming more important as power supply is scaled down.

Static Dissipation

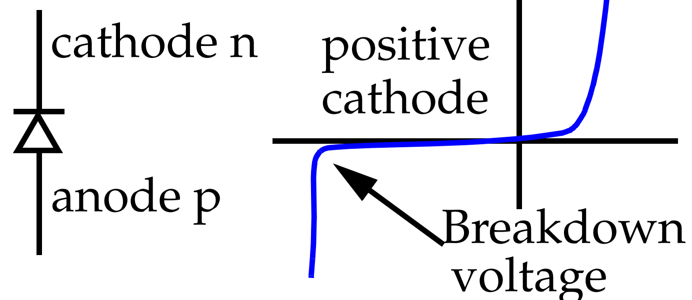
Reverse-bias leakage current

Diode equation: $I = A_d I_s \left(e^{\frac{qV}{kT}} - 1 \right)$ where I_s = the saturation current.
V = diode voltage.
kT/q = 25.8mV at 300 degrees K.
A_d = area of the diode

when V is negative, exponential term becomes small.

$I_{leakage} = 0.1nA \rightarrow 0.5nA/device$
at room temperature

$P_{leakage}$ of an inverter at 5V = 1 or 2 nW



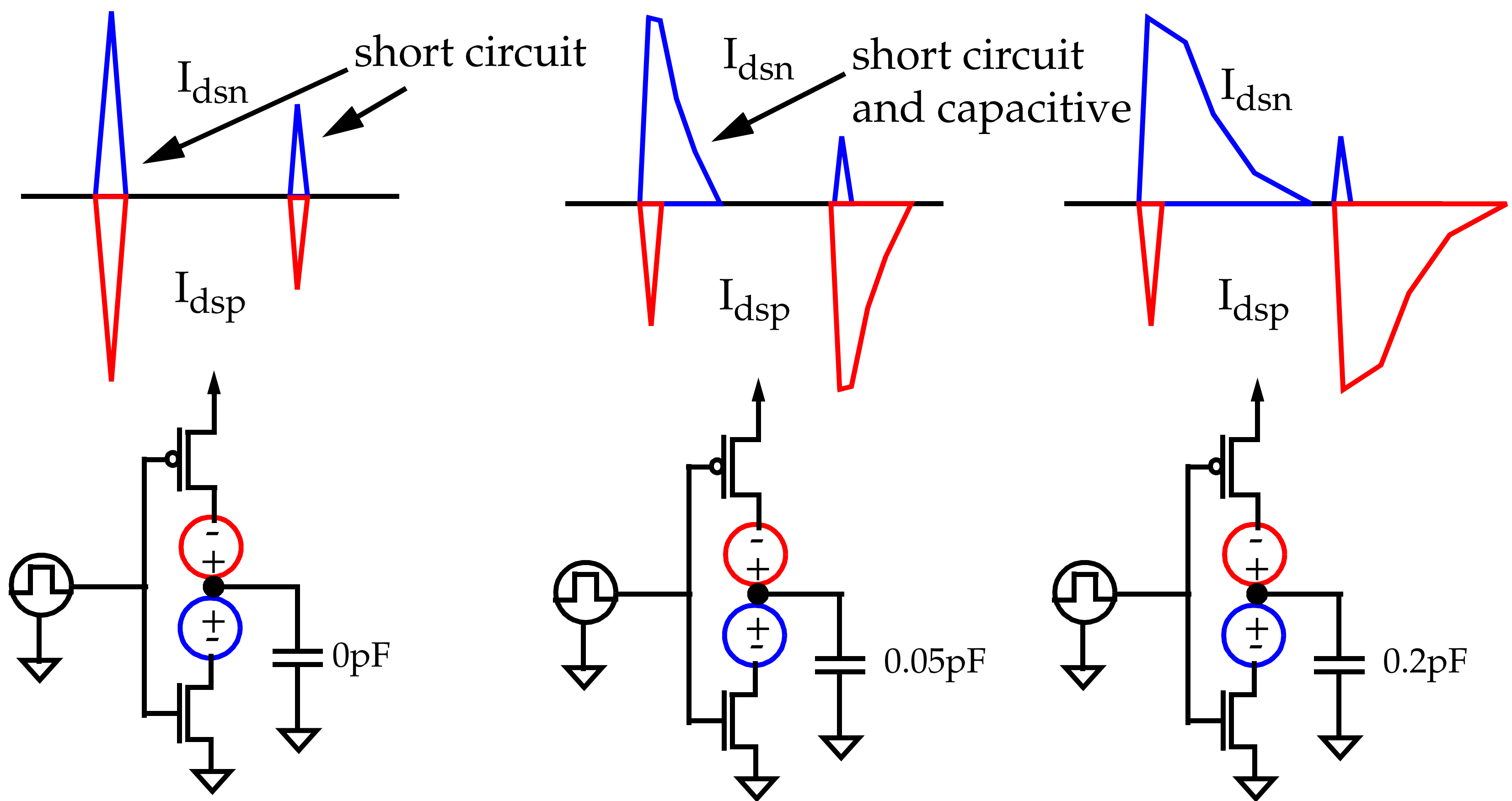
Total Static Power dissipation:

$$P_s = \sum_{1}^{n} (\text{leakage current} \times \text{supply voltage}) + \sum_{1}^{m} \text{through-current}$$

Dynamic Dissipation

The current required to charge/discharge capacitive load usually dominates the crowbar (short circuit) current.

For example:



Capacitive Dissipation

However, slow rise and fall times will increase crowbar current of driven gates.

Assuming a step input and a repetition frequency of f_p, the average dynamic power, P_d, is expressed as:

$$P_d = C_L V_{DD}^2 f_p$$

Therefore power is proportional to

- The switching frequency
- The capacitive load.

But goes up as V_{DD}^2.

Also, power is independent of device parameters, such as V_t or beta.

Total Power Dissipation

$P_{total} = P_s + P_d + P_{short\text{-}circuit}$ (see text P. 236)

Detailed analysis of power is often impractical.

Consider the following simplifications:

- Calculate total capacitance driven by the gates in the circuit.
- Estimate the percentage of the devices operating at the max clock frequency (e.g. 50%).
- Use the dynamic power dissipation expression:

$$P_d = \frac{\text{percent-activity} \times C_{total} \times V_{DD}^2}{t_p} \qquad \text{where} \quad f_p = \frac{1}{t_p}$$

Power minimization:

- Use complementary logic gates to reduce through current (static)
- Use minimum-size devices to reduce diffusion leakage (static).
- Reduce V_{DD}, the frequency and the switched capacitance (dynamic).

Sizing Routing Conductors

The size of metal conductors is important because:

- Metal migration.
- Power supply noise and integrity.
- RC delay (considered previously).

Electro-migration is the transport of metal ions through a conductor induced by direct current.

A 'safe' value of current density, J, is:

$$0.4mA/\mu m \quad \text{to} \quad 1.0mA/\mu m$$

For example, consider a clock buffer that drives a 100 pF load at 50 MHz:

$$P_d = C_L V_{DD}^2 f_p = 100\times10^{-12} \times 25 \times 50\times10^{6} = 125mW$$

$$I = \frac{P}{V} = \frac{125mW}{5V} = 25mA$$

Using $0.5mA/\mu m$ as the limit, the wire width should be at least $50\mu m$
A safe value would be $100\mu m$

Sizing Routing Conductors

Power supply noise and integrity:

IR drops on V_{DD} and V_{SS} (voltage drops due to current spikes and the resistance of the metal) can occur causing gates to fail.

What is the voltage drop (ground bounce) in the power and ground wires if the buffer is 500 microns from the power and ground pads ?

$$R = \frac{500\mu m}{100\mu m}(0.05\Omega/\mu m) = 0.25\Omega$$

$$IR = 0.025A \times 0.25 = 6.25mV$$

Design Margining

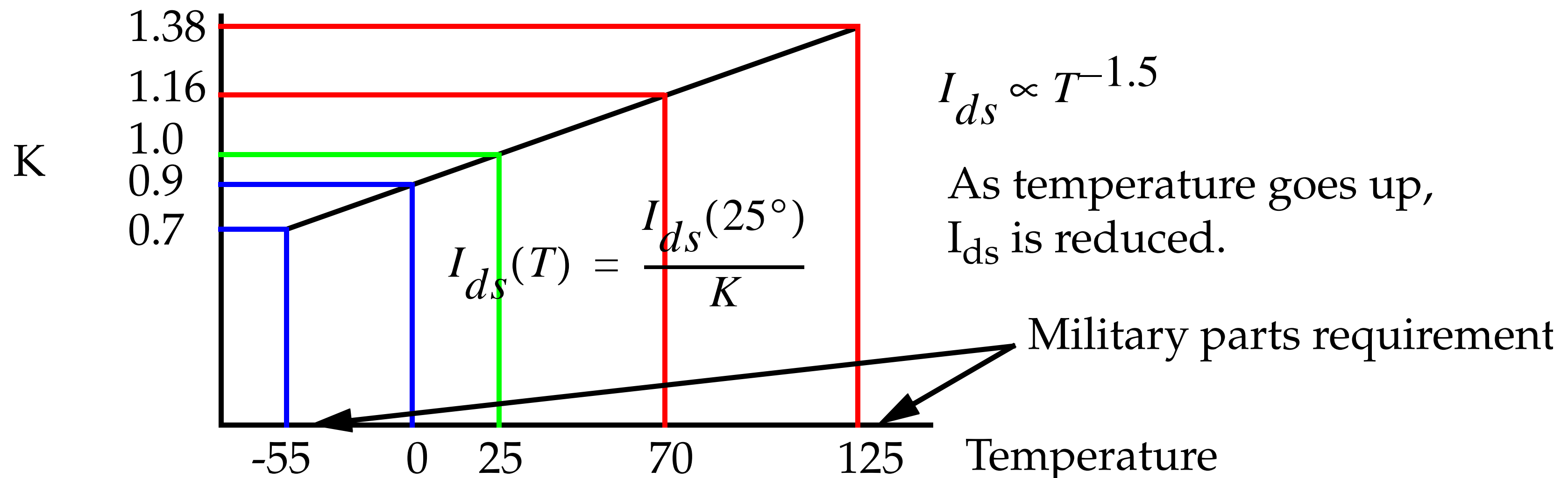
Sources of variations that effect nominal circuit behavior (2 environmental, 1 manufacturing):

- Operating Temperature
- Supply Voltage: Data sheets give +/- 10%, e.g., 3.0 to 3.6 for 3.3V.
- Process Variation: Normal to keep parts within 2 or 3 sigma.

We must design the circuit to operate over all extremes of these variables.

Temperature:

What happens to I_{ds} with temperature ?



$I_{ds} \propto T^{-1.5}$

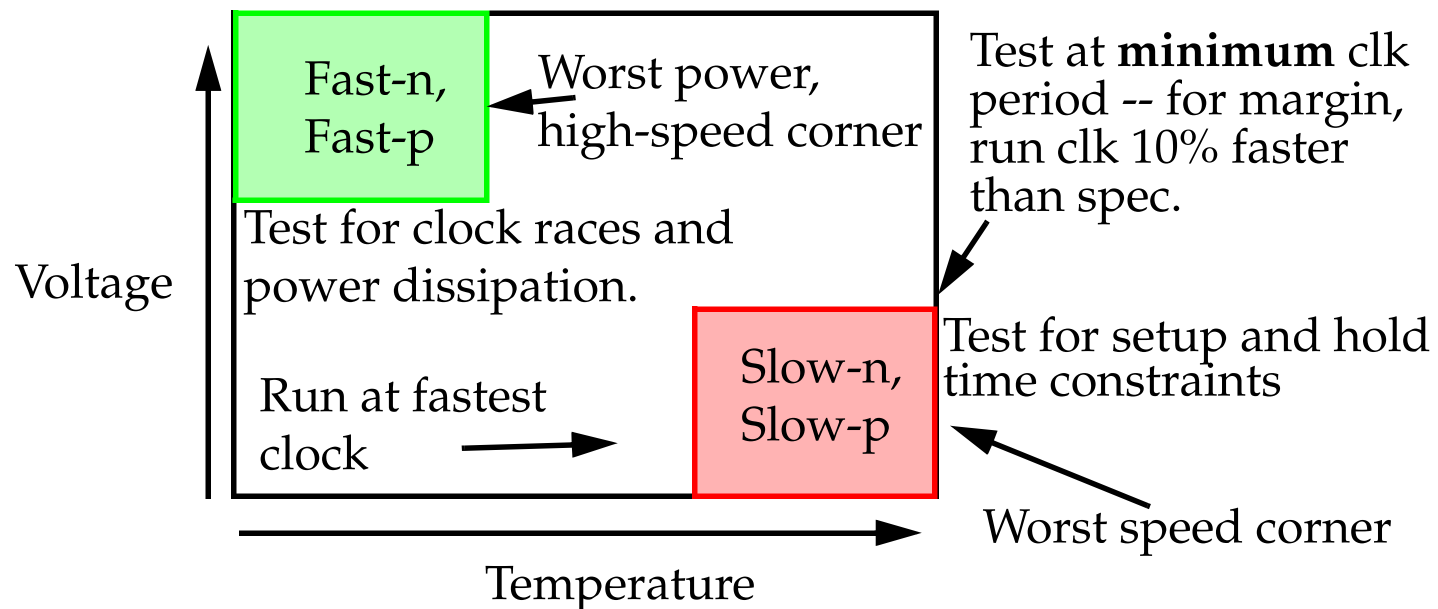
As temperature goes up, I_{ds} is reduced.

Design Corners

Sources of process variation include changes in doping densities, oxide thickness and line width variations.

The following boundary combinations may result

- Fast-n, Fast-p
- Fast-n, Slow-p
- Slow-n, Fast-p
- Slow-n, Slow-p

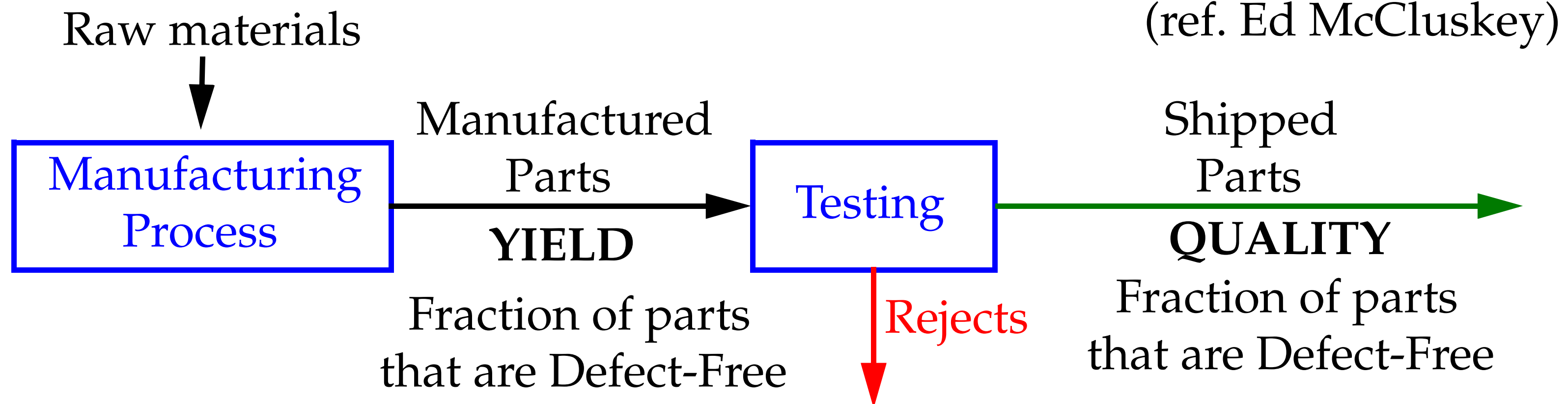


Yield

Defined as:

$$Y = \frac{\text{No. of Good Chips of Wafer}}{\text{Total No. of chips}} 100\%$$

Yield verses Testing



Yield is influenced by:

- Technology
- Chip Area
- Layout

Yield

A simple model for yield (Seed's model):

$$Y = e^{-\sqrt{AD}}$$

where A = chip area
D = defect density (lethal defects/cm^2).

Clearly, yield decreases dramatically as the area of the chip increases.

Yield and testing are related by

$$DL = 1 - Y^{(1 - T)}$$

where DL = Fraction of devices shipped that are defective (measured in defective parts/million or DPM).
Y = Yield.
T = Test coverage percentage.

Another exponential function that states that if yield is low, we better have high test coverage (+99%) otherwise we ship lots of bad parts.

Device testing is a course of its own - stayed tuned.

First Order Approximations of Scaling

Constant field scaling: 1/alpha scaling applied to all dimensions, device voltages and concentration densities.

- I_{ds} per transistor scales by 1/alpha.
- # of transistors per unit area scales by alpha^2.
- Current density scales by alpha, power density remains constant (VI/A), e.g., (1/alpha*1/alpha)*alpha^2

Constant voltage scaling: V_{DD} is held constant while process is scaled.

- I_{ds} per transistor scales by alpha.
- # of transistors per unit area scales up by alpha^2.
- Current density scales by alpha^3, power density scales by alpha^3.

Lateral scaling: Only the gate length is scaled (*gate-shrink*).

- I_{ds} per transistor scales by alpha.
- # of transistors per unit area scales by alpha.
- Current density scales by alpha^2, power density scales by alpha^2.