CMOS

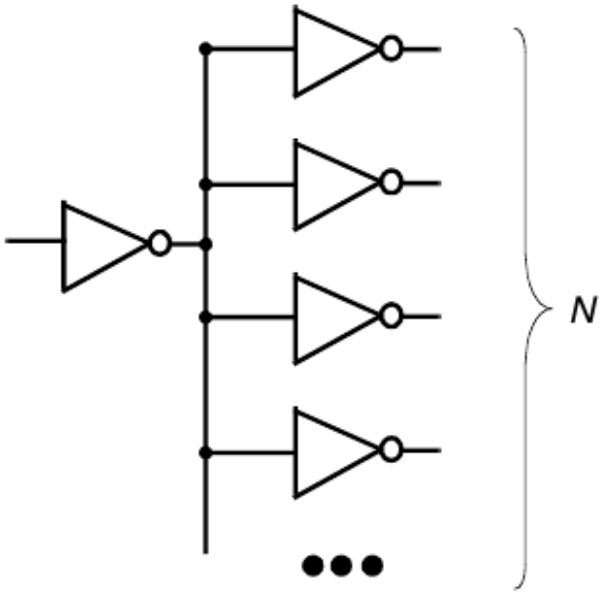
OUTLINE

» Fan-out

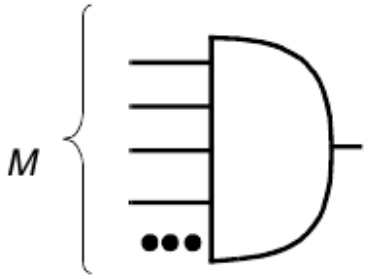
» Propagation delay

» CMOS power consumption

Fan-In and Fan-Out



(a) Fan-out *N*



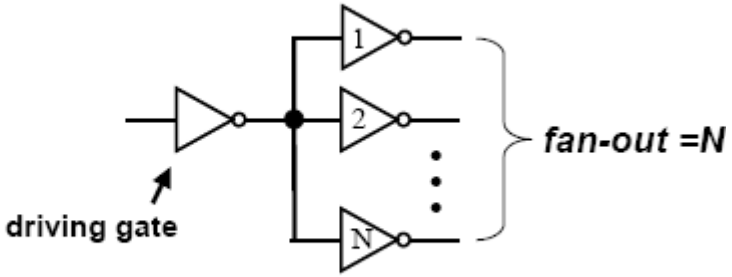
(b) Fan-in *M*

Figure 1.16 Definition of fan-out and fan-in of a digital gate.

- Fan-out คือจำนวนเกตที่เป็นโหลด (Load gate) สูงสุดที่ต่อเป็นเอาท์พุทของเกตที่เป็นตัวขับ (Driving gate) ที่สามารถรับได้
- การเพิ่ม Fan-out ของเกต จะมีผลต่อระดับแรงดันลอจิกที่เอาท์พุท และการเพิ่ม Fan-out ก็เหมือนกับการเพิ่มโหลดให้กับวงจร จึงเป็นการทำให้ประสิทธิภาพของวงจรเสียไป
- Fan-in คือจำนวนอินพุทของเกต
- เกตที่มี Fan-in มาก ๆ จะมีวงจรที่ซับซ้อนขึ้น และเป็นผลต่อคุณสมบัติของวงจร เช่น Delay time เพิ่มขึ้น

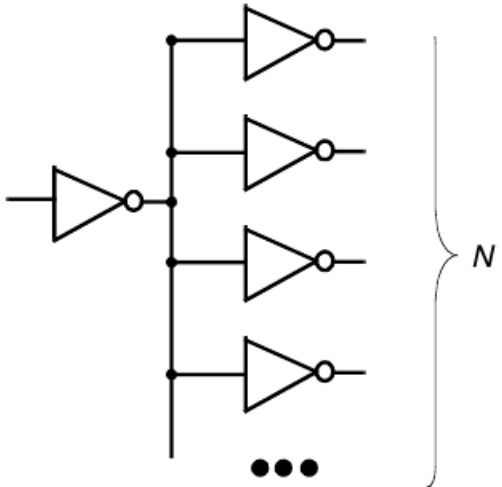
Fan-In and Fan-Out

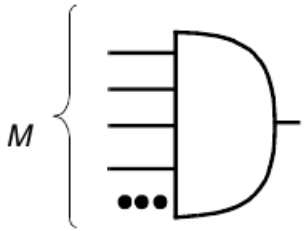
- The ***fan-out*** is the number of gates that are connected to the output of the driving gate:



- Fanout leads to increased capacitive load on the driving gate, and therefore longer propagation delay

$$Fanout = \frac{I_{OL(\max)}}{I_{IL(\max)}} \quad \text{หรือ} \quad Fanout = \frac{I_{OH(\max)}}{I_{IH(\max)}}$$

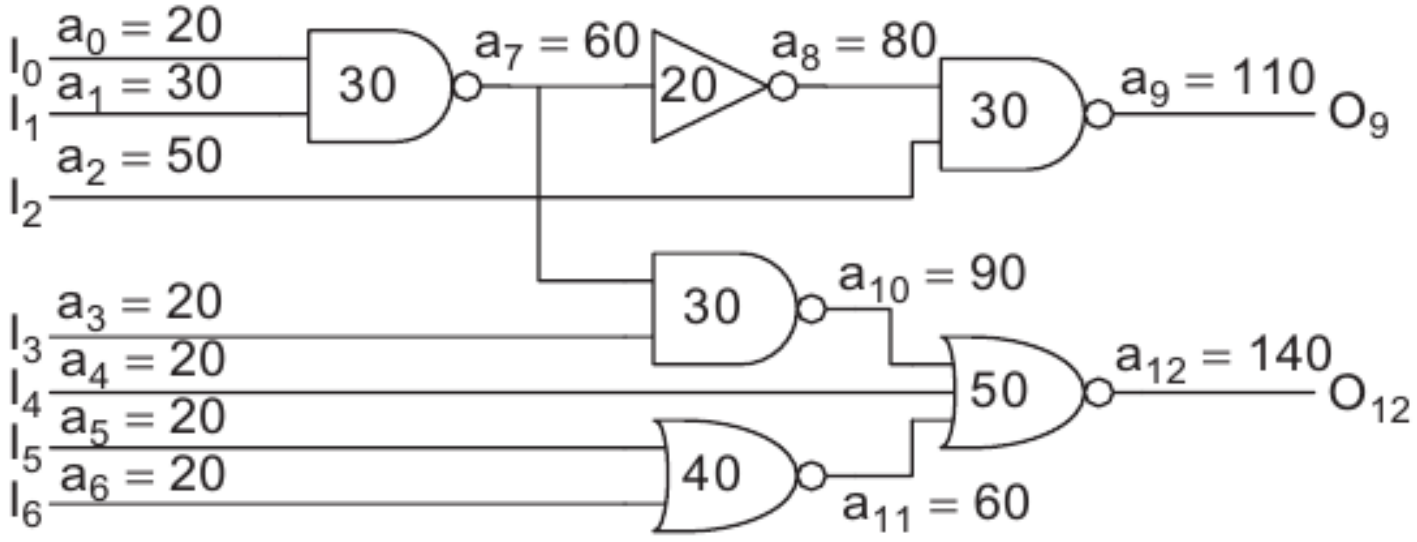




(b) Fan-in *M*

The propagation delay

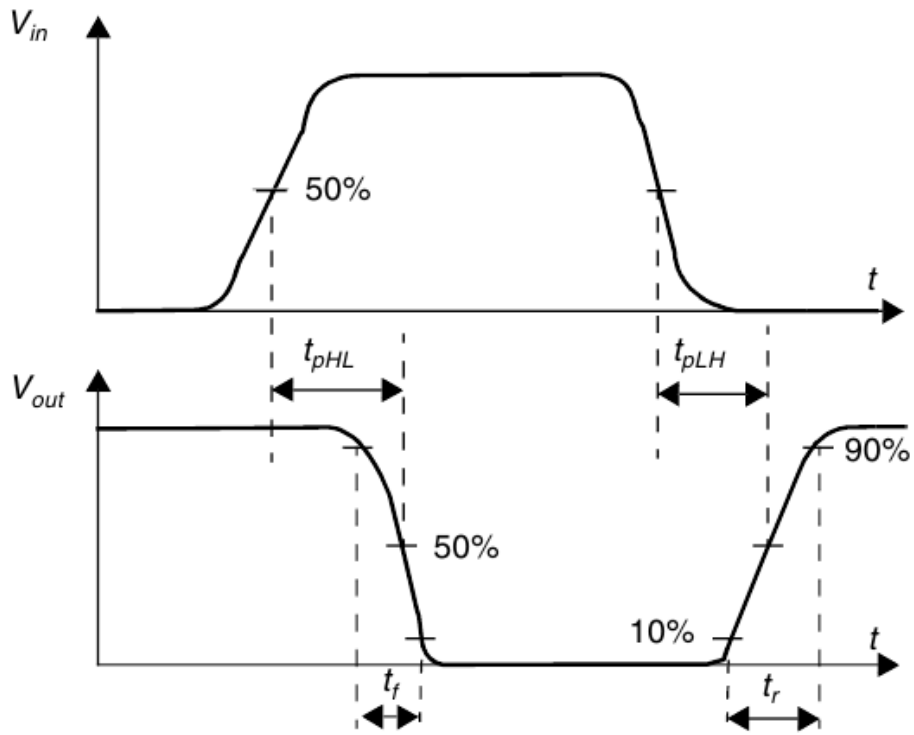
- A very important measure of the performance of a digital system, such as a computer, is the maximum speed at which it is capable of operating.



The propagation delay

The propagation delay *tp* of a gate defines how quickly it responds to a change at its input(s). It expresses the delay experienced by a signal when passing through a gate.

It is measured between the 50% transition points of the input and output waveforms



The propagation delay

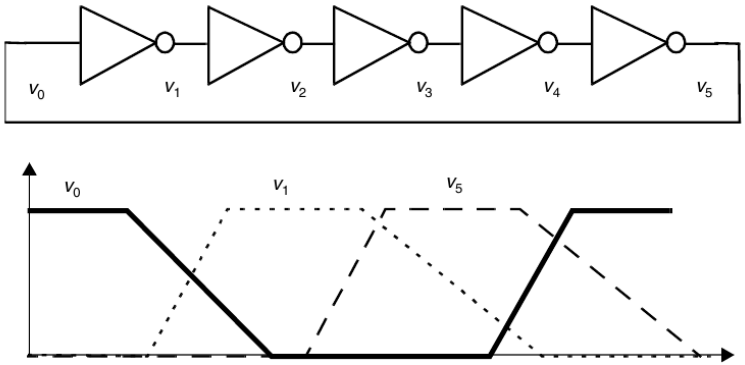
- The t_{pLH} defines the response time of the gate for a low to high (or positive) output transition, while t_{pHL} refers to a high to low (or negative) transition. The propagation delay t_p is defined as the average of the two.

$$t_p = \frac{t_{pLH} + t_{pHL}}{2}$$

The ring oscillator

- A uniform way of measuring the t_p of a gate, so that technologies can be judged on an equal footing, is desirable.
- The de-facto standard circuit for delay measurement is the ring oscillator, which consists of an odd number of inverters connected in a circular chain
- Typically, a ring oscillator needs a least five stages to be operational.

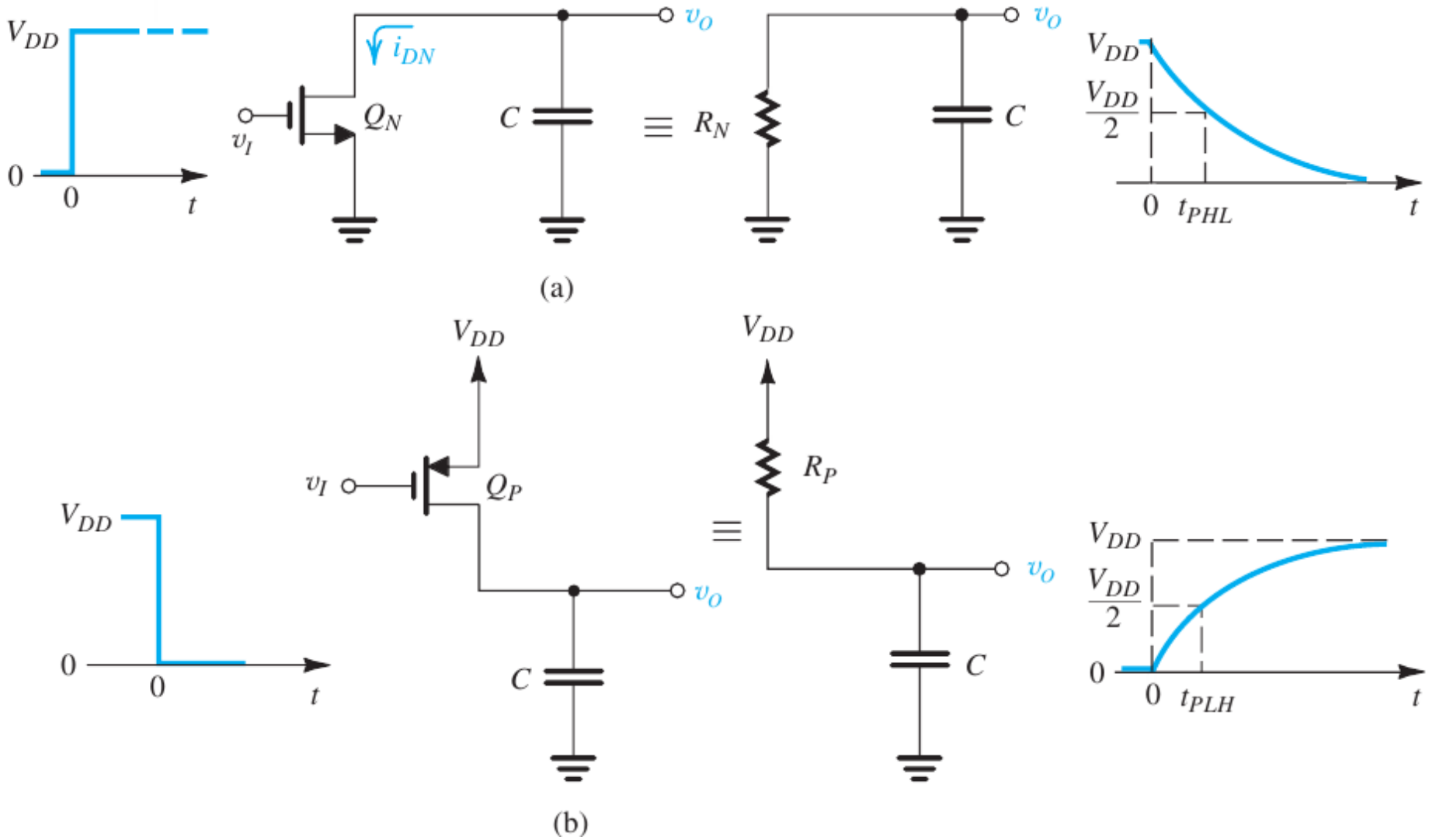
The ring oscillator



Determining the Propagation Delay

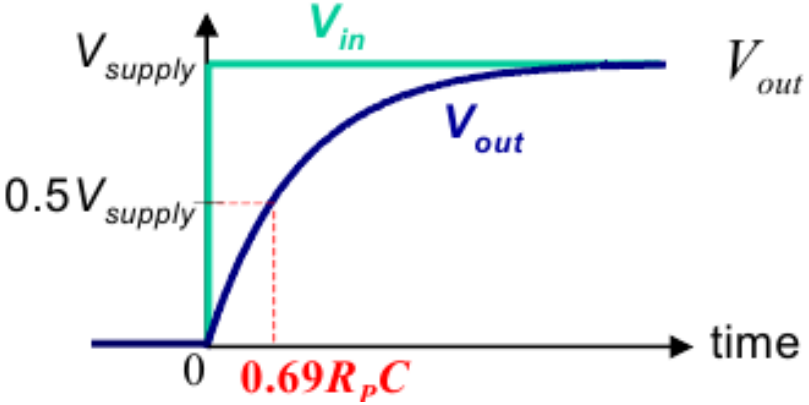
- The most fundamental way to compute delay is to develop a physical model of the circuit of interest, write a differential equation describing the output voltage as a function of input voltage and time, and solve the equation.
- The solution of the differential equation is called the transient response, and the delay is the time when the output reaches VDD /2.

The propagation delays of the inverter.



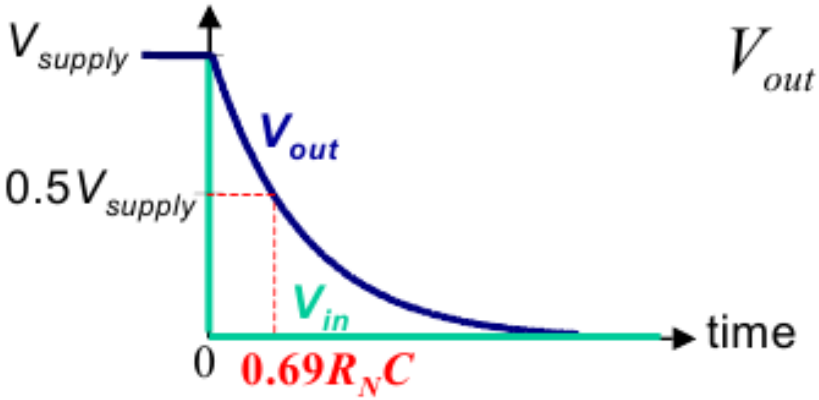
The propagation delays of the inverter.

Example: Output voltage changing from "low" to "high"



$$V_{out}(t) = V_{supply}\left(1 - e^{-t/R_PC}\right)$$

Example: Output voltage changing from "high" to "low"



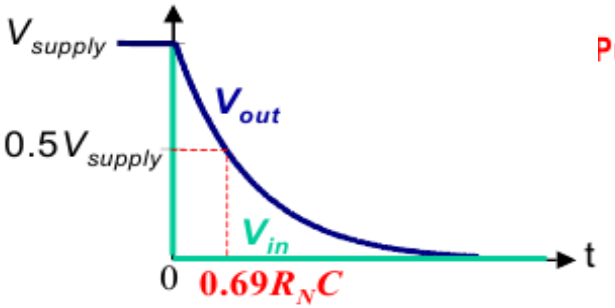
$$V_{out}(t) = V_{high}e^{-t/R_NC}$$

Model the MOSFET in the ON state as a resistive switch:

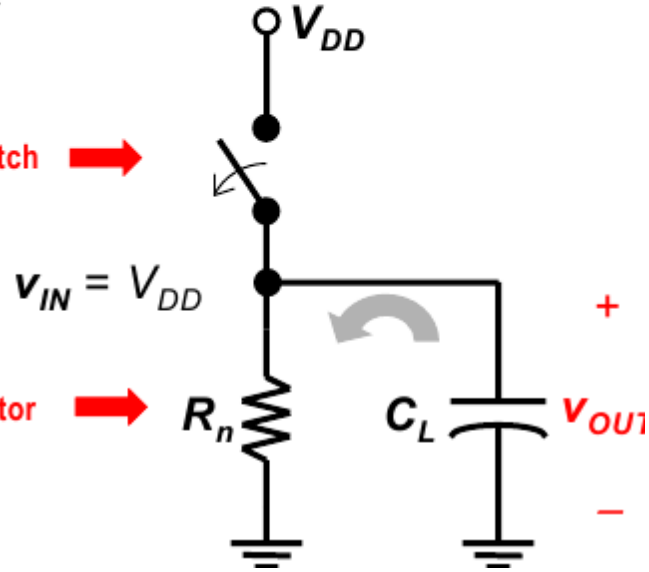
Case 1: V_{out} changing from High to Low
(input signal changed from Low to High)

- NMOSFET(s) connect V_{out} to *GND*

$$t_{pHL} = 0.69 \cdot R_n C_L$$

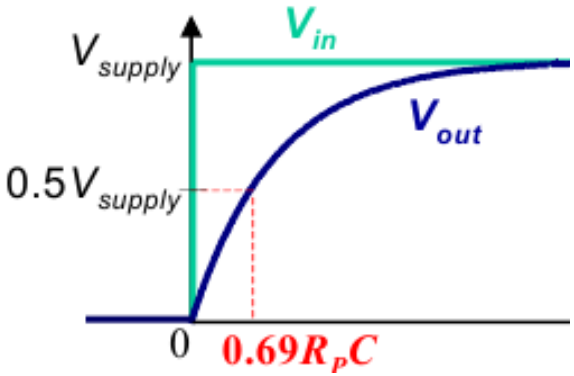


Pull-up network is modeled as an open switch



Pull-down network is modeled as a resistor

The time to reach the 50% point is easily computed as $t = \ln(2)\tau = 0.69\tau$



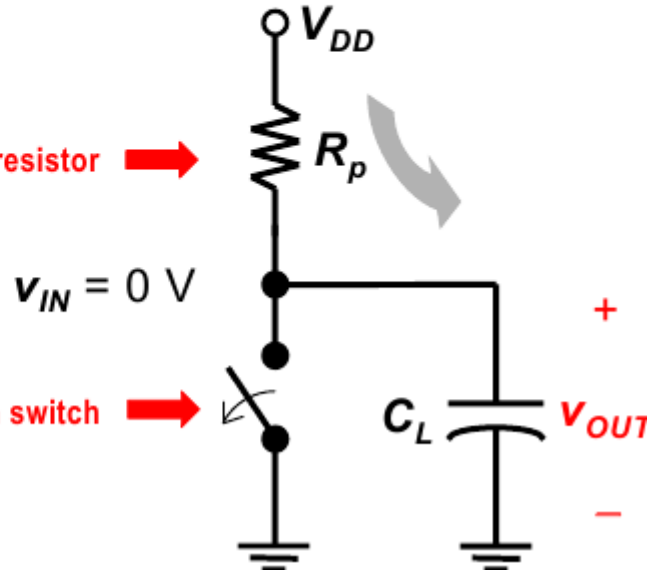
Case 2: V_{out} changing from Low to High
(input signal changed from High to Low)

- PMOSFET(s) connect V_{out} to V_{DD}

$$t_{pLH} = 0.69 \cdot R_p C_L$$

Pull-up network is modeled as a resistor

Pull-down network is modeled as an open switch



Power and Energy Consumption

- The power consumption of a design determines how much energy is consumed per operation, and much heat the circuit dissipates

$$P_{peak} = i_{peak} V_{supply} = max[p(t)]$$

$$P_{av} = \frac{1}{T}\int_0^T p(t)dt = \frac{V_{supply}}{T}\int_0^T i_{supply}(t)dt$$

where *p(t)* is the instantaneous power, *isupply* is the current being drawn from the supply voltage *Vsupply* over the interval $t \in [0,T]$, and *ipeak* is the maximum value of *isupply* over that interval.

- The dissipation can further be decomposed into static and dynamic components.

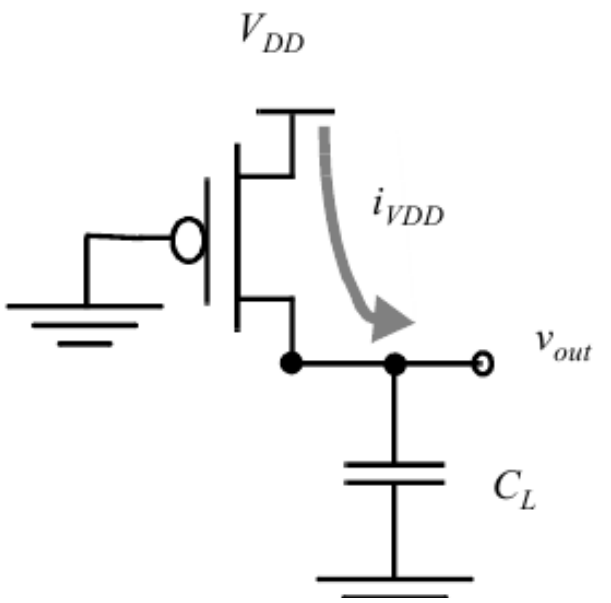
$$P_{stat} = I_{stat} V_{DD}$$

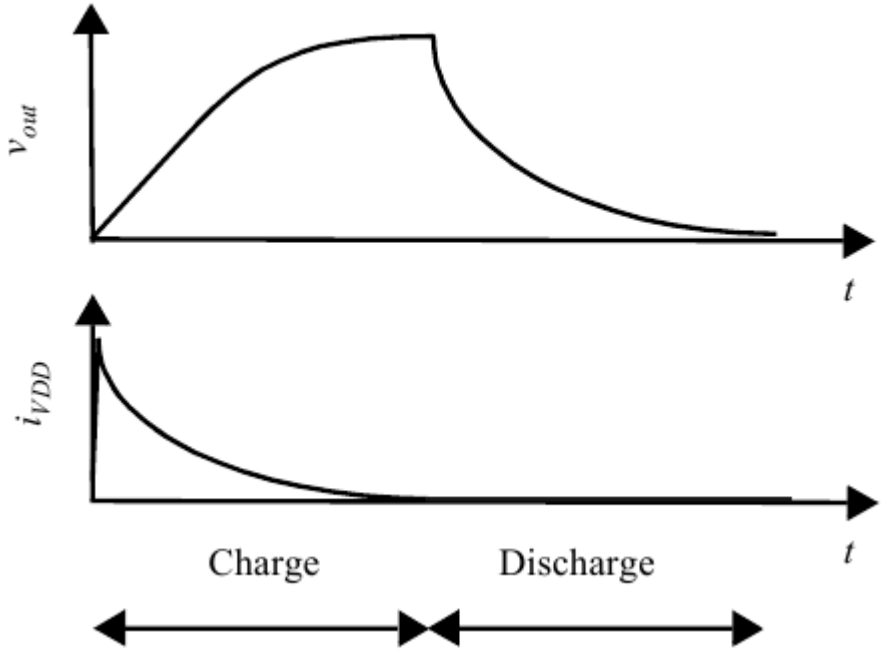
$$P_{\text{dyn}} = f C V_{DD}^2$$

Where *Istat* is the current that flows between the supply rails in the absence of switching activity

where *f* is the frequency at which the gate is switched. It follows that minimizing *C* is an effective means for reducing dynamic power dissipation

Dynamic Dissipation due to Charging and Discharging Capacitances





$$E_{VDD} = \int_0^\infty i_{VDD}(t) V_{DD} dt = V_{DD} \int_0^\infty C_L \frac{dv_{out}}{dt} dt = C_L V_{DD} \int_0^{V_{DD}} dv_{out} = C_L V_{DD}^2$$

$$E_C = \int_0^\infty i_{VDD}(t) v_{out} dt = \int_0^\infty C_L \frac{dv_{out}}{dt} v_{out} dt = C_L \int_0^{V_{DD}} v_{out} dv_{out} = \frac{C_L V_{DD}^2}{2}$$

- If the gate is switched on and off f0→1 times
- per second, the power consumption equals

$$P_{dyn} = C_L V_{DD}^2 f_{0 \rightarrow 1}$$

f0→1 represents the frequency of energy-consuming transitions, this is 0→1 transitions for static CMOS.

If the input signals remain unchanged, no switching happens, and the dynamic power consumption is zero!

$$P_{dyn} = C_L V_{DD}^2 f_{0\to 1} = C_L V_{DD}^2 P_{0\to 1} f = C_{EFF} V_{DD}^2 f$$

$f_{0\to 1}$ = frequency of 0→1 transitions ("switching activity")

f = clock rate (maximum possible event rate)

Effective capacitance C_{EFF} = average capacitance charged every clock cycle

Reference books

- CMOS VLSI Design A Circuits and Systems Perspective Fourth Edition, Neil H. E. Weste ,David Money Harris.
- VLSI Design and Tools : ดร. ธีรยศ เวียงทอง
- Sedra/Smith,*Microelectronic Circuits*, 6^{th} edition
- Digital Integrated Circuits : A Design Perspective 2 Edition : Jan M. Rabaey, Anantha Chandrakasan, and Borivoje Nikolic