

# **Lecture 10: Power**

# What is due 10/3

- ☐ Reading 5.1 – 5.3
- ☐ Midterm will be on 10/10
- ☐ Quiz #5 will be on thursday

# Outline

---

- ☐ Power and Energy
- ☐ Dynamic Power
- ☐ Static Power

# Power and Energy

- ❑ Power is drawn from a voltage source attached to the  $V_{DD}$  pin(s) of a chip.
- ❑ Instantaneous Power:  $P(t) =$
- ❑ Energy:  $E =$
- ❑ Average Power:  $P_{avg} =$

# Power in Circuit Elements

$$P_{VDD}(t) = I_{DD}(t)V_{DD}$$



$$P_R(t) = \frac{V_R^2(t)}{R} = I_R^2(t)R$$



$$\begin{aligned} E_C &= \int_0^{\infty} I(t)V(t)dt = \int_0^{\infty} C \frac{dV}{dt} V(t)dt \\ &= C \int_0^{V_C} V(t)dV = \frac{1}{2} CV_C^2 \end{aligned}$$



# Charging a Capacitor

- When the gate output rises
  - Energy stored in capacitor is
- But energy drawn from the supply is

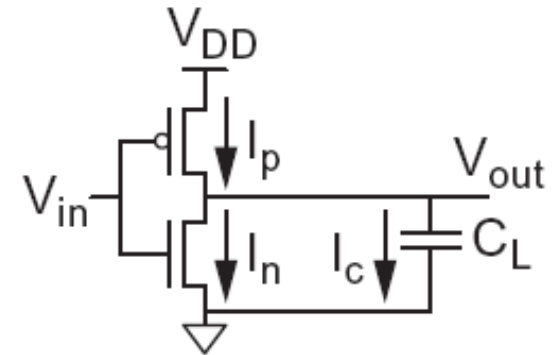
$$E_C = \frac{1}{2} C_L V_{DD}^2$$

$$E_{VDD} = \int_0^{\infty} I(t) V_{DD} dt = \int_0^{\infty} C_L \frac{dV}{dt} V_{DD} dt$$

$$= C_L V_{DD} \int_0^{V_{DD}} dV = C_L V_{DD}^2$$

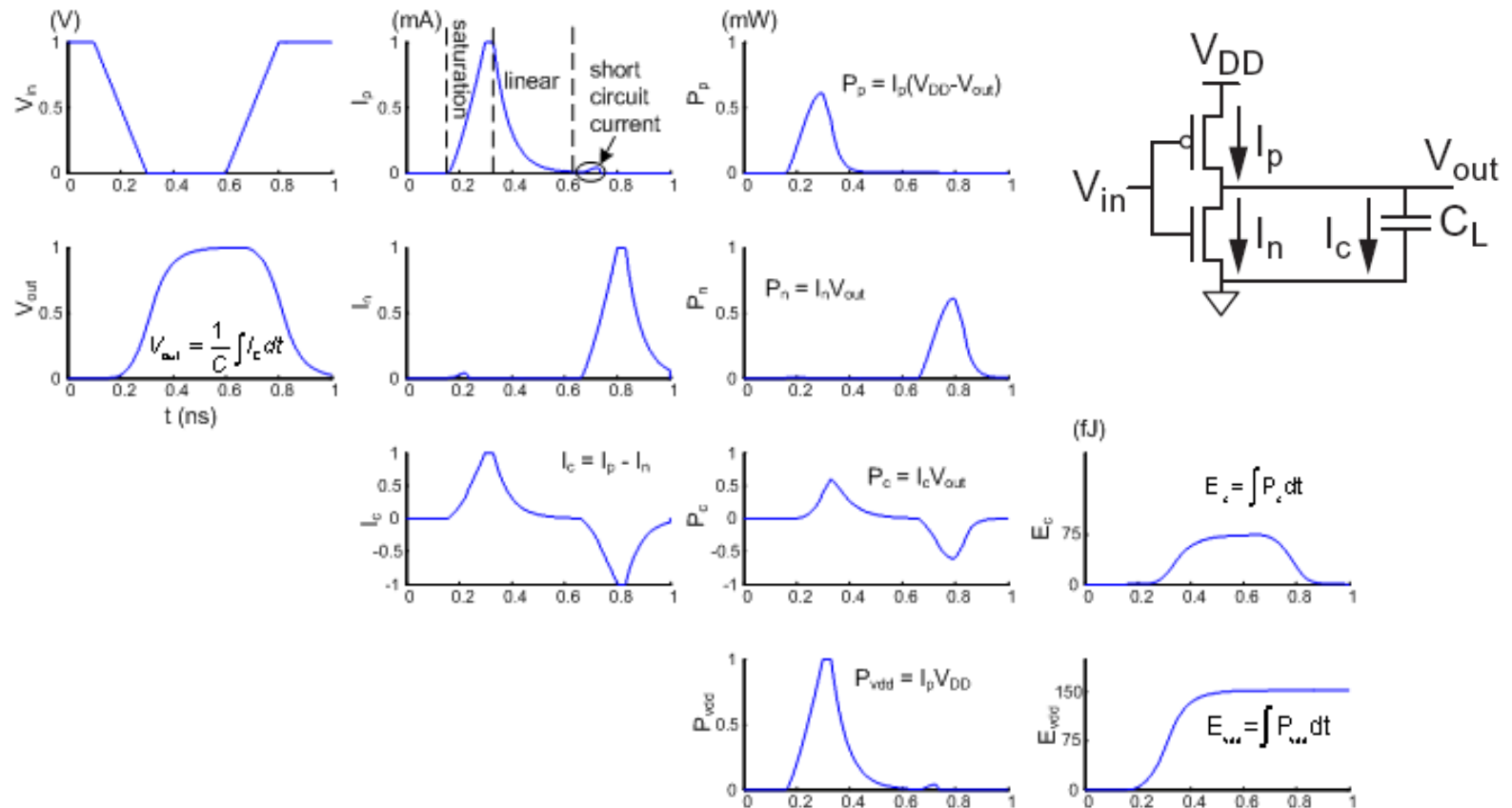
- Half the energy from  $V_{DD}$  is dissipated in the pMOS transistor as heat, other half stored in capacitor

- When the gate output falls
  - Energy in capacitor is dumped to GND
  - Dissipated as heat in the nMOS transistor



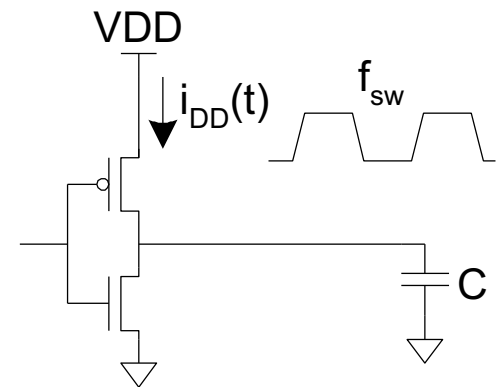
# Switching Waveforms

□ Example:  $V_{DD} = 1.0 \text{ V}$ ,  $C_L = 150 \text{ fF}$ ,  $f = 1 \text{ GHz}$



# Switching Power

$$\begin{aligned} P_{\text{switching}} &= \frac{1}{T} \int_0^T i_{DD}(t) V_{DD} dt \\ &= \frac{V_{DD}}{T} \int_0^T i_{DD}(t) dt \\ &= \frac{V_{DD}}{T} [T f_{\text{sw}} C V_{DD}] \\ &= C V_{DD}^2 f_{\text{sw}} \end{aligned}$$





# Activity Factor

- ❑ Suppose the system clock frequency =  $f$
- ❑ Let  $f_{sw} = \alpha f$ , where  $\alpha$  = activity factor
  - If the signal is a clock,  $\alpha = 1$
  - If the signal switches once per cycle,  $\alpha = 1/2$

- ❑ Dynamic power:

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

# Short Circuit Current

- ❑ When transistors switch, both nMOS and pMOS networks may be momentarily ON at once
- ❑ Leads to a blip of “short circuit” current.
- ❑  $< 10\%$  of dynamic power if rise/fall times are comparable for input and output
- ❑ We will generally ignore this component

# Power Dissipation Sources

- ❑  $P_{\text{total}} = P_{\text{dynamic}} + P_{\text{static}}$
- ❑ Dynamic power:  $P_{\text{dynamic}} = P_{\text{switching}} + P_{\text{shortcircuit}}$ 
  - Switching load capacitances
  - Short-circuit current
- ❑ Static power:  $P_{\text{static}} = (I_{\text{sub}} + I_{\text{gate}} + I_{\text{junct}} + I_{\text{contention}})V_{\text{DD}}$ 
  - Subthreshold leakage
  - Gate leakage
  - Junction leakage
  - Contention current

# Dynamic Power Example

- ❑ 1 billion transistor chip
  - 50M logic transistors
    - Average width:  $12 \lambda$
    - Activity factor = 0.1
  - 950M memory transistors
    - Average width:  $4 \lambda$
    - Activity factor = 0.02
  - 1.0 V 65 nm process
  - $C = 1 \text{ fF}/\mu\text{m}$  (gate) +  $0.8 \text{ fF}/\mu\text{m}$  (diffusion)
- ❑ Estimate dynamic power consumption @ 1 GHz.  
Neglect wire capacitance and short-circuit current.

# Solution

$$C_{\text{logic}} = (50 \times 10^6)(12\lambda)(0.025 \mu m / \lambda)(1.8 fF / \mu m) = 27 \text{ nF}$$

$$C_{\text{mem}} = (950 \times 10^6)(4\lambda)(0.025 \mu m / \lambda)(1.8 fF / \mu m) = 171 \text{ nF}$$

$$P_{\text{dynamic}} = [0.1C_{\text{logic}} + 0.02C_{\text{mem}}](1.0)^2(1.0 \text{ GHz}) = 6.1 \text{ W}$$

# Dynamic Power Reduction

□  $P_{\text{switching}} = \alpha C V_{DD}^2 f$

- Try to minimize:
- Activity factor
  - Capacitance
  - Supply voltage
  - Frequency

# Activity Factor Estimation

- ❑ Let  $P_i = \text{Prob}(\text{node } i = 1)$ 
  - $\bar{P}_i = 1 - P_i$
- ❑  $\alpha_i = \bar{P}_i * P_i$
- ❑ Completely random data has  $P = 0.5$  and  $\alpha = 0.25$
- ❑ Data is often not completely random
  - e.g. upper bits of 64-bit words representing bank account balances are usually 0
- ❑ Data propagating through ANDs and ORs has lower activity factor
  - Depends on design, but typically  $\alpha \approx 0.1$

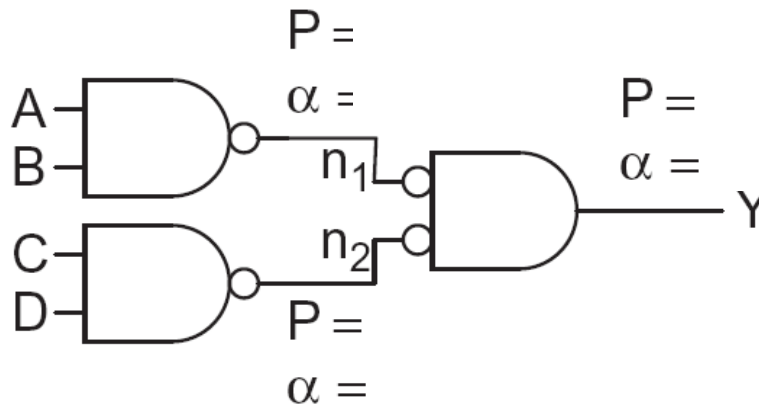
# Switching Probability

Gate	$P_Y$
AND2	$P_A P_B$
AND3	$P_A P_B P_C$
OR2	$1 - \bar{P}_A \bar{P}_B$
NAND2	$1 - P_A P_B$
NOR2	$\bar{P}_A \bar{P}_B$
XOR2	$P_A \bar{P}_B + \bar{P}_A P_B$



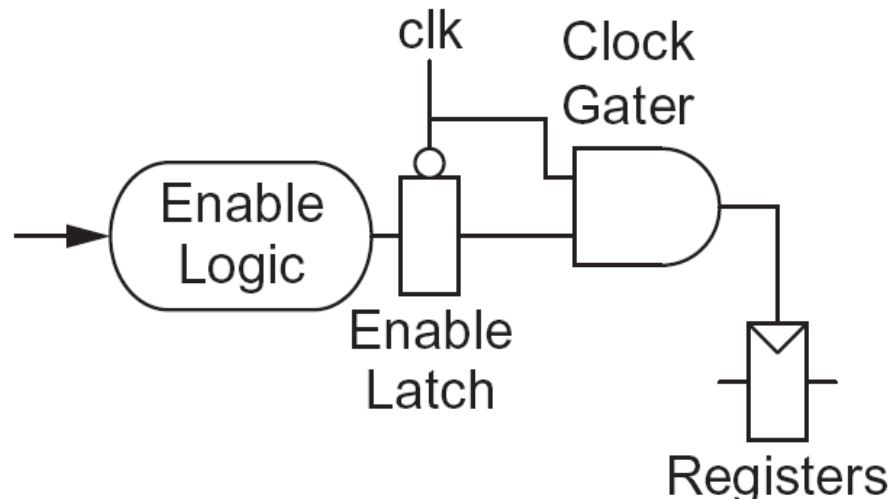
# Example

- ❑ A 4-input AND is built out of two levels of gates
- ❑ Estimate the activity factor at each node if the inputs have  $P = 0.5$



# Clock Gating

- ❑ The best way to reduce the activity is to turn off the clock to registers in unused blocks
  - Saves clock activity ( $\alpha = 1$ )
  - Eliminates all switching activity in the block
  - Requires determining if block will be used



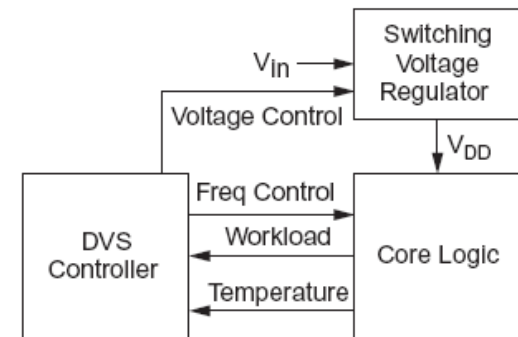
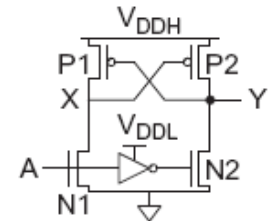
# Capacitance

---

- ❑ Gate capacitance
  - Fewer stages of logic
  - Small gate sizes
- ❑ Wire capacitance
  - Good floorplanning to keep communicating blocks close to each other
  - Drive long wires with inverters or buffers rather than complex gates

# Voltage / Frequency

- ❑ Run each block at the lowest possible voltage and frequency that meets performance requirements
- ❑ Voltage Domains
  - Provide separate supplies to different blocks
  - Level converters required when crossing from low to high  $V_{DD}$  domains
- ❑ Dynamic Voltage Scaling
  - Adjust  $V_{DD}$  and  $f$  according to workload



# Static Power

---

- ❑ Static power is consumed even when chip is quiescent.
  - Leakage draws power from nominally OFF devices
  - Ratioed circuits burn power in fight between ON transistors

# Static Power Example

- ❑ Revisit power estimation for 1 billion transistor chip
- ❑ Estimate static power consumption
  - Subthreshold leakage
    - Normal  $V_t$ : 100 nA/ $\mu\text{m}$
    - High  $V_t$ : 10 nA/ $\mu\text{m}$
    - High  $V_t$  used in all memories and in 95% of logic gates
  - Gate leakage 5 nA/ $\mu\text{m}$
  - Junction leakage negligible

# Solution

$$W_{\text{normal-}V_t} = (50 \times 10^6)(12\lambda)(0.025 \mu\text{m} / \lambda)(0.05) = 0.75 \times 10^6 \mu\text{m}$$

$$W_{\text{high-}V_t} = \left[ (50 \times 10^6)(12\lambda)(0.95) + (950 \times 10^6)(4\lambda) \right] (0.025 \mu\text{m} / \lambda) = 109.25 \times 10^6 \mu\text{m}$$

$$I_{\text{sub}} = \left[ W_{\text{normal-}V_t} \times 100 \text{ nA}/\mu\text{m} + W_{\text{high-}V_t} \times 10 \text{ nA}/\mu\text{m} \right] / 2 = 584 \text{ mA}$$

$$I_{\text{gate}} = \left[ (W_{\text{normal-}V_t} + W_{\text{high-}V_t}) \times 5 \text{ nA}/\mu\text{m} \right] / 2 = 275 \text{ mA}$$

$$P_{\text{static}} = (584 \text{ mA} + 275 \text{ mA})(1.0 \text{ V}) = 859 \text{ mW}$$

# Leakage Control

- ❑ Leakage and delay trade off
  - Aim for low leakage in sleep and low delay in active mode
- ❑ To reduce leakage:
  - Increase  $V_t$ : *multiple  $V_t$* 
    - Use low  $V_t$  only in critical circuits
  - Increase  $V_s$ : *stack effect*
    - *Input vector control* in sleep
  - Decrease  $V_b$ 
    - *Reverse body bias* in sleep
    - Or forward body bias in active mode



# Gate Leakage

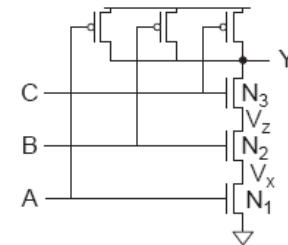
- ❑ Extremely strong function of  $t_{ox}$  and  $V_{gs}$ 
  - Negligible for older processes
  - Approaches subthreshold leakage at 65 nm and below in some processes
- ❑ An order of magnitude less for pMOS than nMOS
- ❑ Control leakage in the process using  $t_{ox} > 10.5 \text{ \AA}$ 
  - High-k gate dielectrics help
  - Some processes provide multiple  $t_{ox}$ 
    - e.g. thicker oxide for 3.3 V I/O transistors
- ❑ Control leakage in circuits by limiting  $V_{DD}$

# NAND3 Leakage Example

□ 100 nm process

$$I_{gn} = 6.3 \text{ nA} \quad I_{gp} = 0$$

$$I_{offn} = 5.63 \text{ nA} \quad I_{offp} = 9.3 \text{ nA}$$



Input State (ABC)	$I_{sub}$	$I_{gate}$	$I_{total}$	$V_x$	$V_z$
000	0.4	0	0.4	stack effect	stack effect
001	0.7	0	0.7	stack effect	$V_{DD} - V_t$
010	0.7	1.3	2.0	intermediate	intermediate
011	3.8	0	3.8	$V_{DD} - V_t$	$V_{DD} - V_t$
100	0.7	6.3	7.0	0	stack effect
101	3.8	6.3	10.1	0	$V_{DD} - V_t$
110	5.6	12.6	18.2	0	0
111	28	18.9	46.9	0	0

Data from [Lee03]

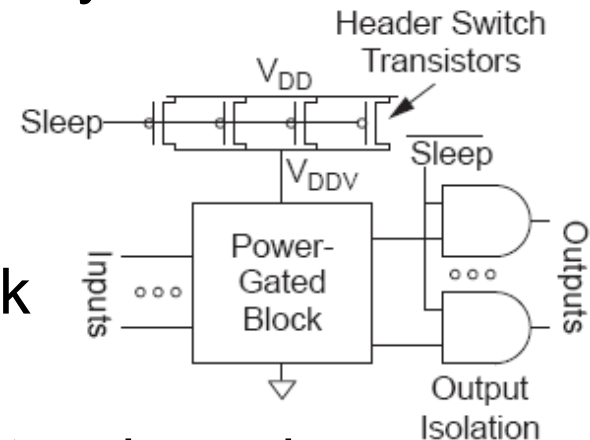
# Junction Leakage

- ❑ From reverse-biased p-n junctions
  - Between diffusion and substrate or well
- ❑ Ordinary diode leakage is negligible
- ❑ Band-to-band tunneling (BTBT) can be significant
  - Especially in high- $V_t$  transistors where other leakage is small
  - Worst at  $V_{db} = V_{DD}$
- ❑ Gate-induced drain leakage (GIDL) exacerbates
  - Worst for  $V_{gd} = -V_{DD}$  (or more negative)

# Power Gating

- ❑ Turn OFF power to blocks when they are idle to save leakage

- Use virtual  $V_{DD}$  ( $V_{DDV}$ )
- Gate outputs to prevent invalid logic levels to next block



- ❑ Voltage drop across sleep transistor degrades performance during normal operation
  - Size the transistor wide enough to minimize impact
- ❑ Switching wide sleep transistor costs dynamic power
  - Only justified when circuit sleeps long enough