# ELEC 402
# IC Power Consumption

Lecture 15

Reza Molavi
Dept. of ECE
University of British Columbia
reza@ece.ubc.ca

Slides Courtesy : Prof. Sudip Shekhar (UBC)

# Power & Energy

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

- **Instantaneous Power (W):** $P(t) = I(t)V(t)$

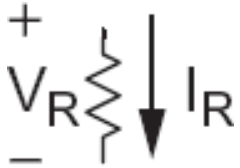
- **Energy (J):** $E = \int_0^T P(t)dt$

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

# Power & Energy in Circuit Elements

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

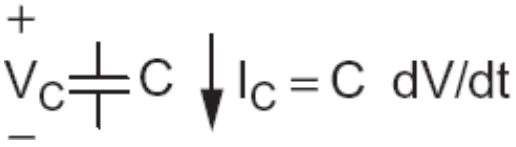

$$P_R(t) = \frac{V_R^2(t)}{R} = I_R^2(t) R$$




**Energy Stored in a Capacitor**

$$E_C = \int_0^{\infty} I(t) V(t) dt = \int_0^{\infty} C \frac{dV}{dt} V(t) dt$$

$$= C \int_0^{V_C} V(t) dV = \tfrac{1}{2} C V_C^2$$




Charging from 0 to $V_C$

# Charging a Cap through an Inverter

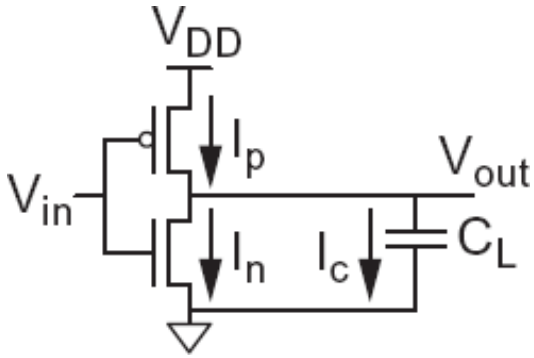
- **When the gate output rises**
  - **Energy stored in capacitor is**

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

  - **But energy drawn from the supply is**

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

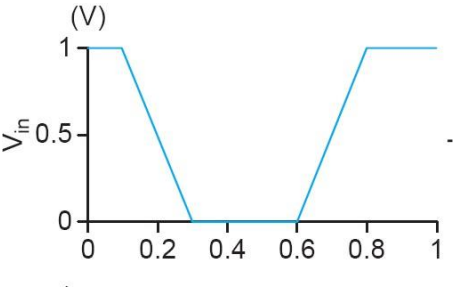
$$= C_L V_{DD} \int_0^{V_{DD}} dV = C_L V_{DD}^2$$

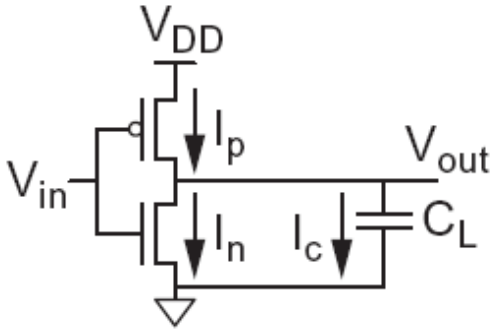



  - **Half the energy from $V_{DD}$ is dissipated in the PMOS transistor as heat, other half stored in capacitor**
- **When the gate output falls**
  - **Energy in capacitor is dumped to GND**
  - **Dissipated as heat in the NMOS transistor**

# Switching Waveforms

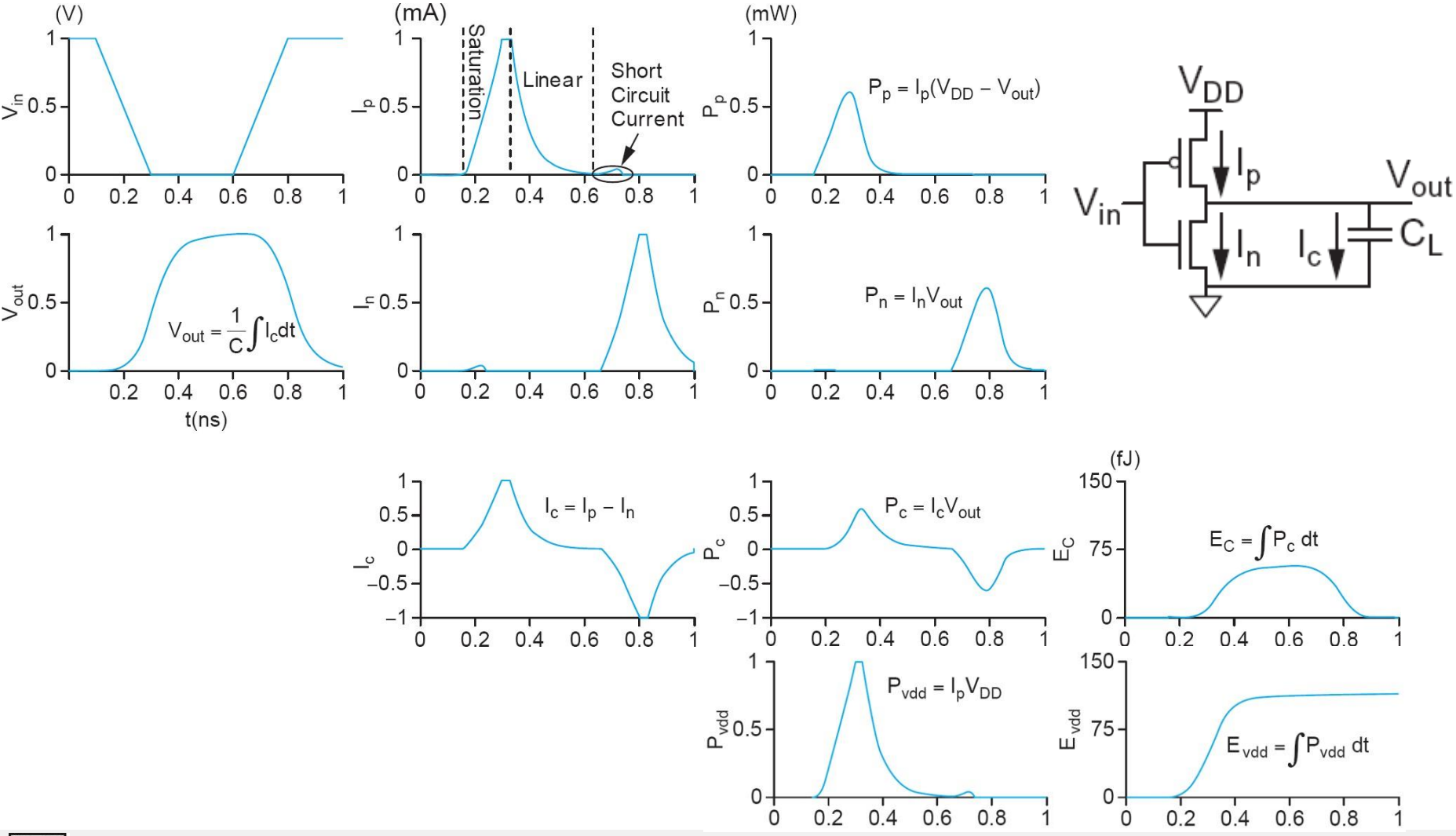
- Example: $V_{DD} = 1.0V$, $C_L = 150fF$, $f = 1GHz$



# Switching Waveforms

- **Example: $V_{DD}$ = 1.0V, $C_L$ = 150fF, f = 1GHz**



# Switching Power

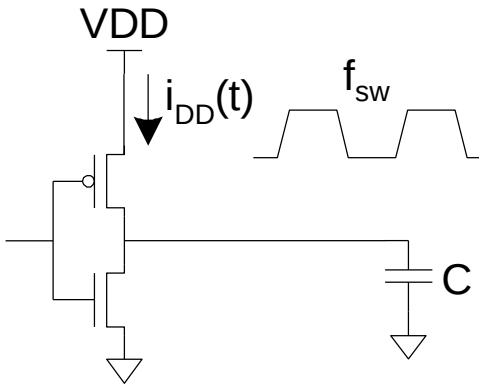
$$P_{switching} = \frac{1}{T_{SW}} \int_0^{T_{SW}} i_{DD}(t) V_{DD} dt$$

$$= V_{DD} f_{SW} \int_0^{T_{SW}} i_{DD}(t) dt$$

$$= V_{DD} f_{SW} \int_0^{T_{SW}} C \frac{dv_C}{dt} dt$$

$$= V_{DD} f_{SW} C \int_0^{T_{SW}} \frac{dv_C}{dt} dt = V_{DD} f_{SW} C \int_0^{V_{DD}} dv_C$$

$$= C V_{DD}^2 f_{SW}$$



# Activity Factor

- Suppose the system clock frequency = f
- Let $f_{sw} = \alpha f$, where $\alpha$ = activity factor = probability that the circuit node transitions from 0 to 1.
  - If the signal is a clock, $\alpha = 1$
  - If the signal switches once per cycle, $\alpha = ½$
- 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

  (as clock period goes further down, i.e. increasing frequency, this assumption becomes less accurate).

# Power Dissipation Sources

- $P_{total} = P_{dynamic} + P_{static}$
- Dynamic power: $P_{dynamic} = P_{switching} + P_{shortcircuit}$
  - Switching load capacitances
  - Short-circuit current
- Static power: $P_{static} = (I_{sub} + I_{gate} + I_{junct})V_{DD}$
  - 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**
  - **1.0V 65nm process ($\lambda$ ~ 30 nm)**
  - **C = 1 fF/$\mu$m (gate) + 1 fF/$\mu$m (diffusion)**
- **Estimate dynamic power consumption @ 1 GHz. Neglect wire capacitance (!) and short-circuit current.**

$$C_{\text{logic}} = \left(50\times10^{6}\right)\left(12\lambda\right)\left(0.030\mu m/\lambda\right)\left(2\,fF/\mu m\right) = 27\ \text{nF}$$

$$C_{\text{mem}} = \left(950\times10^{6}\right)\left(4\lambda\right)\left(0.030\mu m/\lambda\right)\left(2\,fF/\mu m\right) = 171\ \text{nF}$$

$$P_{\text{dynamic}} = \left\lceil \left\lfloor 0.1C_{\text{logic}} + 0.02C_{\text{mem}} \right\rfloor \right\rceil (1.0)^2\,(1.0\ \text{GHz}) = 6.1\ \text{W}$$

# Dynamic Power Reduction

- $P_{\text{switching}} = \alpha C V_{DD}{}^2 f$
- **Try to minimize:**
  - **Activity factor**
  - **Capacitance**
  - **Supply voltage**
  - **Frequency**

# Activity Factor Estimation

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

- Let $P_i$ = Prob(node i = 1)
  - $\overline{P}_i = 1 - P_i$
- $\alpha_i = \overline{P}_i * P_i$ (node is 0 on one cycle and 1 on the next)
- 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 have lower activity factor
  - Depends on design, but typically $\alpha \approx 0.1$

# Switching Probability

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

- Assuming the inputs are uncorrelated

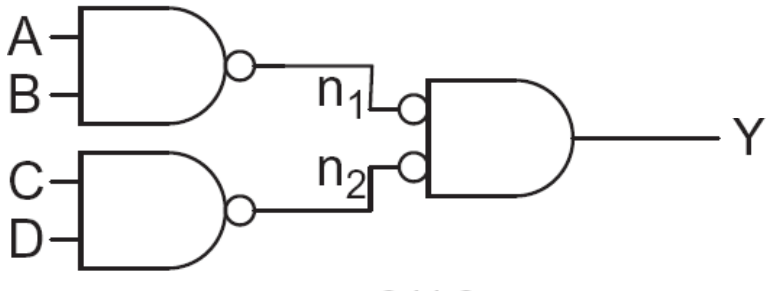
| Gate | $P_Y$ |
|---|---|
| AND2 | $P_A P_B$ |
| AND3 | $P_A P_B P_C$ |
| OR2 | $1 - \overline{P}_A \overline{P}_B$ |
| NAND2 | $1 - P_A P_B$ |
| NOR2 | $\overline{P}_A \overline{P}_B$ |
| XOR2 | $P_A \overline{P}_B + \overline{P}_A P_B$ |

- Activity Factor of the output = $\overline{P_Y} * P_Y$

# Example

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

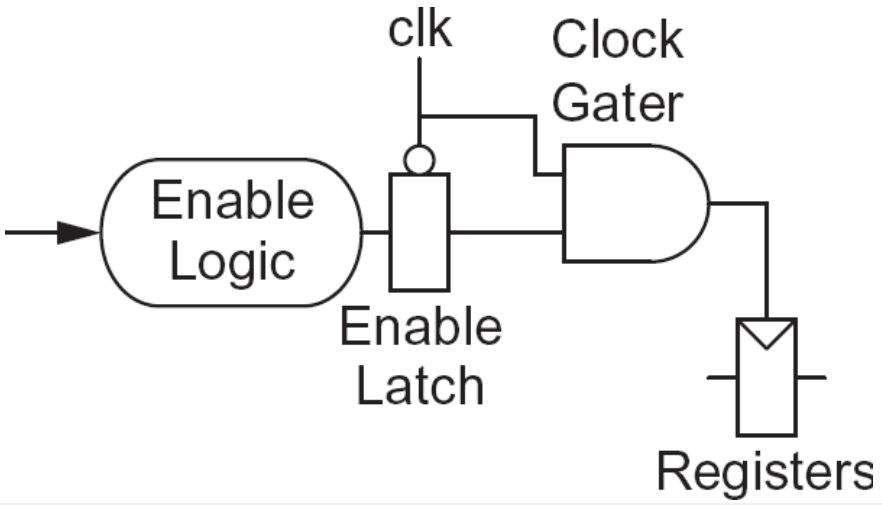
- An AND4 is built out of two levels of gates
- Estimate the activity factor at each node if the inputs have P = 0.5



# Clock Gating

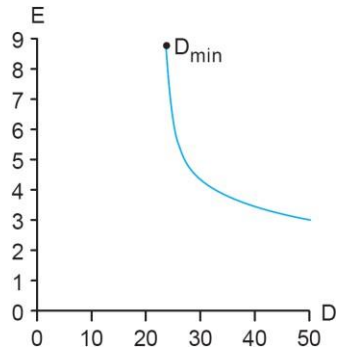
$$P_{\text{switching}} = \alpha C V_{DD}^{\ 2} f$$

- **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**
  - **Disable clock early in the clock distribution network ($\alpha$ = 1) and high cap blocks.**
  - **Use an enable latch to ensure that enable does not change before the clock falls**



# Capacitance

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

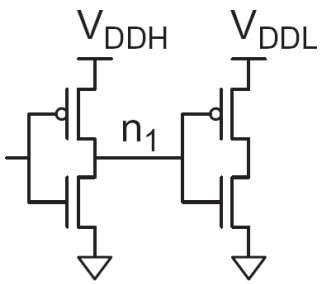


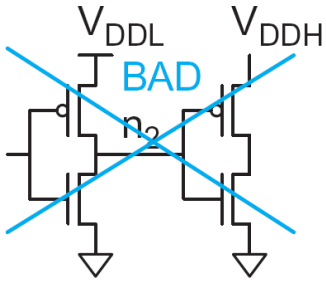

- **Gate capacitance**
  - **Fewer stages of logic**
  - **Small gate sizes**
  - **FO8-12 if needed to reduce energy at slight overhead of delay**
- **Wire capacitance**
  - **Good floorplanning to keep communicating blocks close to each other**
  - **Drive long wires with inverters or buffers rather than complex gates**

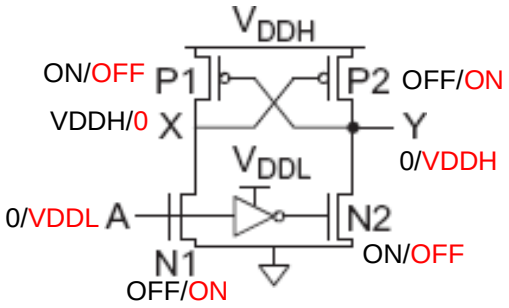
# Voltage

$$P_{switching} = \alpha C V^2 f$$

- **Run each block at the lowest possible voltage that meets performance requirements**
- **Voltage Domains**
  - **Provide separate supplies to different blocks**
  - **Level converters required when crossing from low to high $V_{DD}$ domains**



# Voltage & Frequency

$$P_{\text{switching}} = \alpha C V$$

- **Run each block at the lowest possible voltage and frequency that meets performance requirements**

- **Dynamic Voltage Frequency Scaling (DVFS)**
  - **Adjust $V_{DD}$ and f according to workload**
  - **Cubic reduction in power may be achieved**