BSC~Microsoft Research Centre

Tassadaq Hussain

VLSI
Design

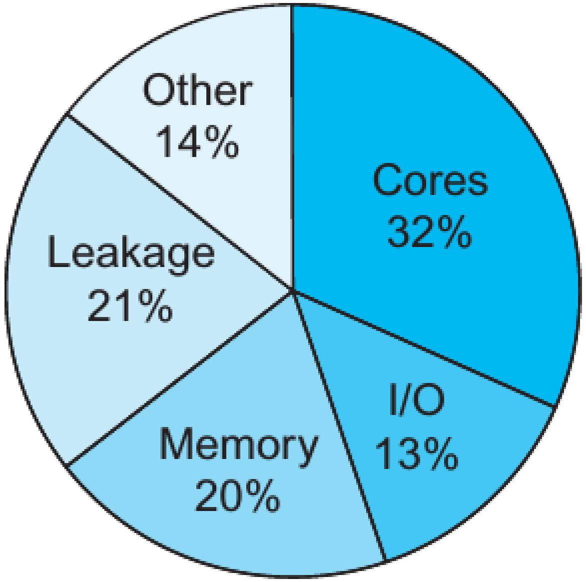
Power

Outline

Power and Energy

Dynamic Power

Static Power



RIPHAH INTERNATIONAL UNIVERSITY

Power and Energy

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

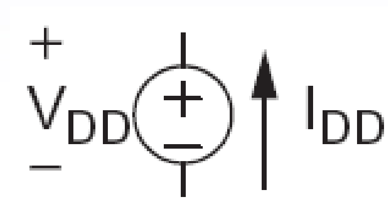
Instantaneous Power:

Energy:

Average Power:

Power in Circuit Elements

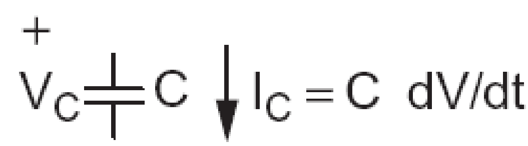
$$P_{VDD}(t) = I_{DD}(t)\,V_{DD}$$



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

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

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



Charging a Capacitor

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

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

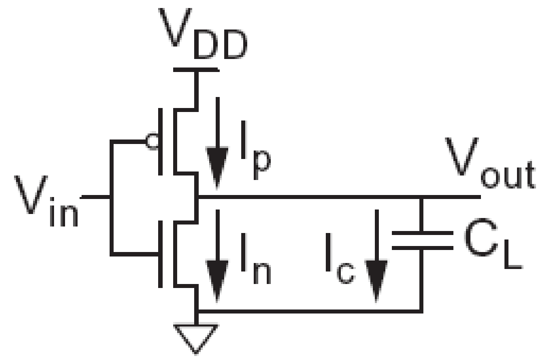
$$= C_L V_{DD} \int_0^{V_{DD}} dV = C_L V_{DD}^2$$

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



RIPHAH INTERNATIONAL UNIVERSITY

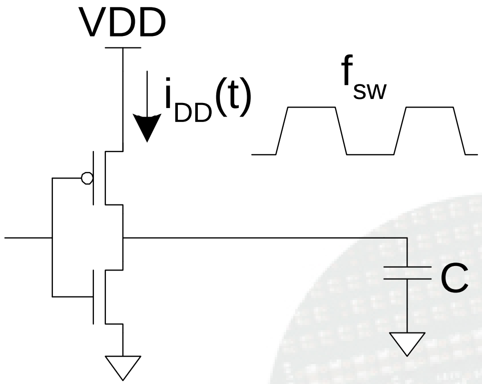
Switching Power

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

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

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

$$= CV_{DD}{}^{2} f_{\text{sw}}$$



Activity Factor

Suppose the system clock frequency = f

Let $f_{sw} = \alpha f$, where α = activity factor

If the signal is a clock, $\alpha = 1$

If the signal switches once per cycle, $\alpha = ½$

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

RIPHAH INTERNATIONAL UNIVERSITY

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

$P_{total} = P_{dynamic} + P_{static}$

Dynamic power: $P_{dynamic} = P_{switching} + P_{shortcircuit}$

Switching load capacitances

Short-circuit current

Static power: $P_{static} = (I_{sub} + I_{gate} + I_{junct} + I_{contention})V_{DD}$

Subthreshold leakage

Gate leakage

Junction leakage

Contention current

A digital system-on-chip in a 1 V 65 nm process (with 50 nm drawn channel lengths and $\lambda = 25$ nm) has 1 billion transistors, of which 50 million are in logic gates and the remainder in memory arrays. The average logic transistor width is 12 λ and the average memory transistor width is 4 . The memory arrays are divided into banks and only the necessary bank is activated so the memory activity factor is 0.02. The static CMOS logic gates have an average activity factor of 0.1. Assume each transistor contributes 1 fF/um of gate capacitance and 0.8 fF/um of diffusion capacitance. Neglect wire capacitance for now (though it could account for a large fraction of total power). Estimate the switching power when operating at 1 GHz.

Dynamic Power Example

1 billion transistor chip

50M logic transistors

- Average width: 12 λ
- Activity factor = 0.1

950M memory transistors

- Average width: 4 λ
- Activity factor = 0.02

1.0 V 65 nm process

C = 1 fF/μm (gate) + 0.8 fF/μm (diffusion)

Estimate dynamic power consumption @ 1 GHz. Neglect wire capacitance and short-circuit current.

RIPHAH INTERNATIONAL UNIVERSITY

Solution

$$C_{\text{logic}} = \left(50\times10^{6}\right)\left(12\lambda\right)\left(0.025\mu m/\lambda\right)\left(1.8\, fF/\mu m\right) = 27\text{ nF}$$

$$C_{\text{mem}} = \left(950\times10^{6}\right)\left(4\lambda\right)\left(0.025\mu m/\lambda\right)\left(1.8\, fF/\mu m\right) = 171\text{ nF}$$

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

Dynamic Power Reduction

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

Try to minimize:

Activity factor

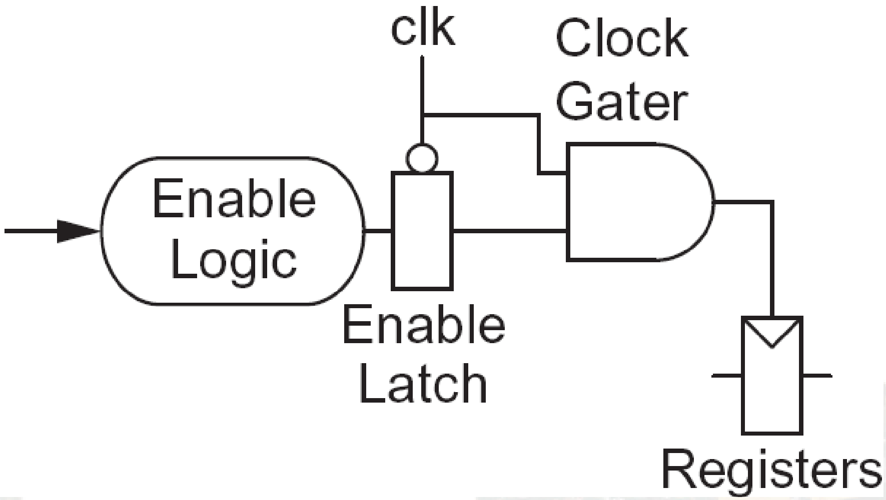
Capacitance

Supply voltage

Frequency

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

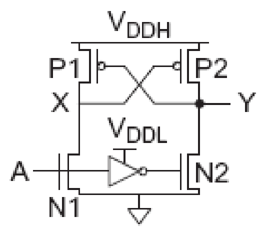
Run each block at the lowest possible voltage and frequency that meets performance requirements

Voltage Domains

Provide separate supplies to different blocks

Level converters required when crossing

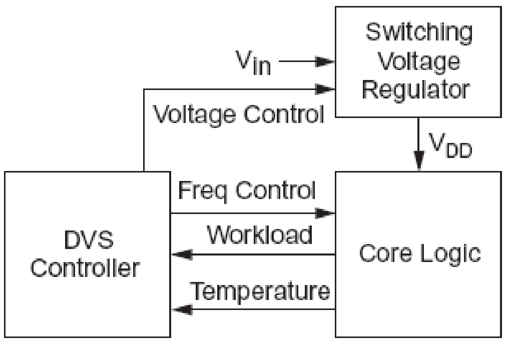
from low to high V_{DD} domains



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

Ratio circuits burn power in fight between ON transistors

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

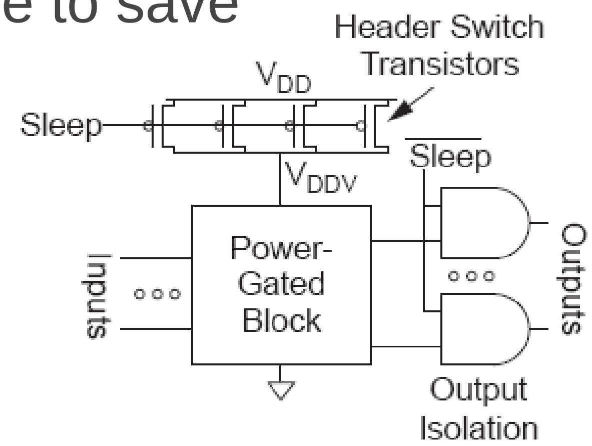
Power Gating

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