EECS 427
Lecture 10: Power/Energy in CMOS Reading: 5.5, 6.3

Outline

- Last time:
  - Adders: Carry select, square-root, log-adders
- Today:
  - Power Dissipation in CMOS
- Sample Logical Effort Questions posted

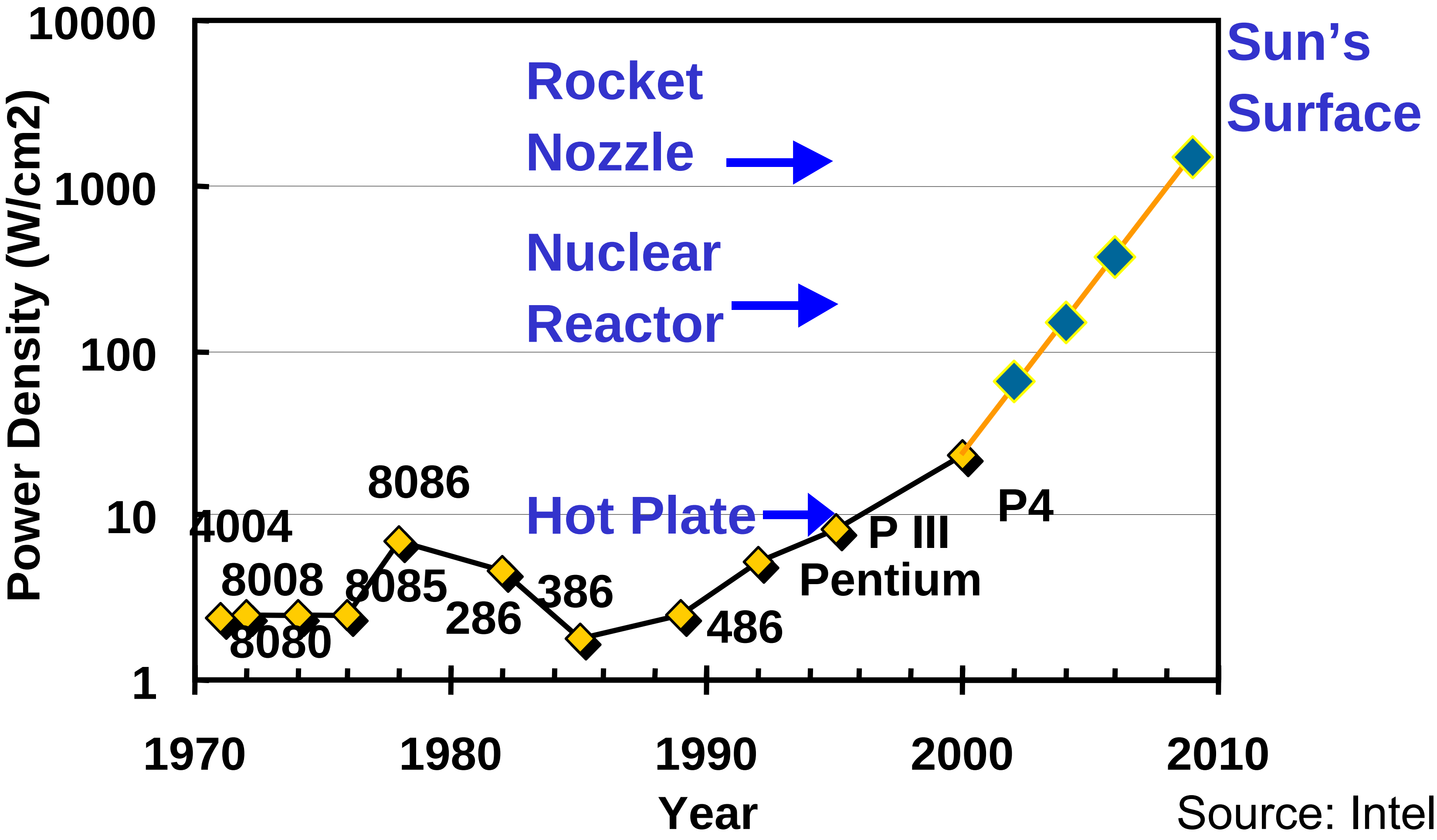
Overview

- Power and energy in CMOS
  - Why is power/energy reduction important
  - What constitutes total power dissipation
  - Arriving at the equations for dissipation
  - Popular approaches to power reduction
- Dynamic logic
  - Alternative to static CMOS
  - Used in industry for very high-speed circuits

Why is Power Reduction Important

- Maintaining chip temperature requires more expensive
  - Packaging: Ceramic vs Plastic for example
  - Heat Sinks
- Power Costs
  - Kilowatt hour costs are increasingly important due to improving performance/server and fairly stable performance/watt.
  - Electricity cost cand easily be 40% of total cost. Current trends indicate a further increase in this percentage.
- Power delivery
  - Itanium consumes 130W at 1.3V
  - IR and Ldi/dt drops reduce supply at devices
- Battery Life
  - Low ***energy*** extremely critical for mobile/hand-held applications
  - Li-ion batteries: 100-150Whr/kg. Size/Battery life trade-off.
- Niche applications
  - Many automotive circuits are hermetically sealed. Severe constraints on heat exchange result in desire of inherently lower power circuits.

Power Density in VLSI

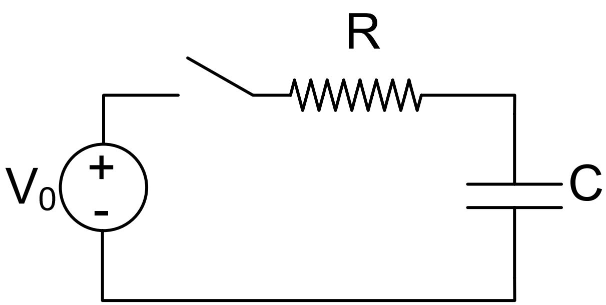


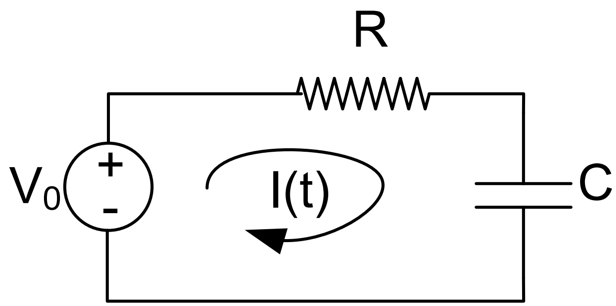
Source: Intel

Where Does the Power Go?

- Dynamic Power
  - Switching Power
    - Charging and discharging of capacitors
  - Short-circuit Power (Crowbar)
    - During switching transients, current flows between Vdd and GND
    - Not dominant – typically assumed to be ~10% of dynamic power
- Static Power
  - Power consumed when there is no switching
  - Leakage: due to non-ideal switches
    - Gradually becoming the dominant component
    - Important during standby

A look at energy dissipation: RC series circuit





- Energy stored equals energy dissipated *in this case.*
- But charge is delivered by non-linear devices in CMOS circuits.
- How does that change the result?
- Is there a device independent way of figuring out the dissipaton??

$$E_C = \frac{1}{2} C V_0^2$$

$$E_{diss}(R) = \int_0^\infty V_R \bullet I \, dt$$

$$E_{diss}(R) = \int_0^\infty I^2 R \, dt$$

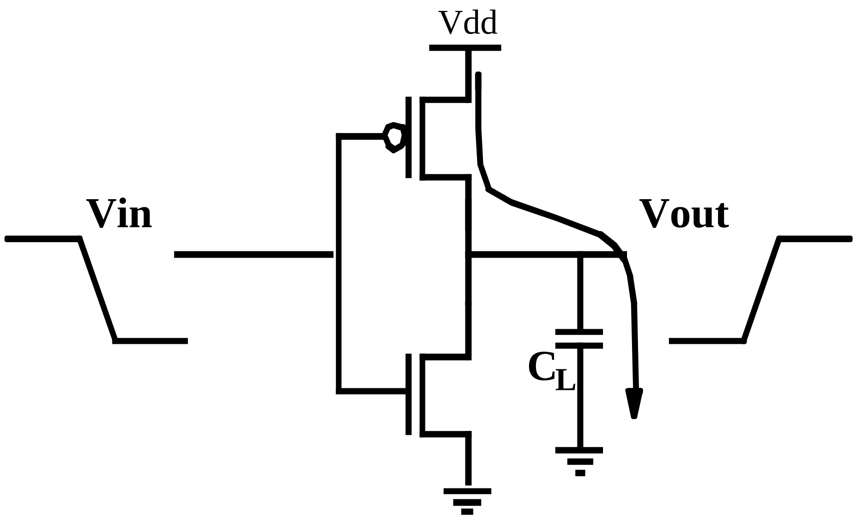
$$E_{diss}(R) = \int_0^\infty \frac{V_0^{\,2}}{R^2} e^{\frac{-2t}{\tau}} R \, dt$$

$$E_{diss}(R) = \frac{V_0^{\,2}}{R} \bullet \frac{\tau}{2} \left[-e^{\frac{-2t}{\tau}} \right]_0^\infty$$

$$E_{diss}(R) = \frac{1}{2} C V_0^2 = E_C$$

Switching Power

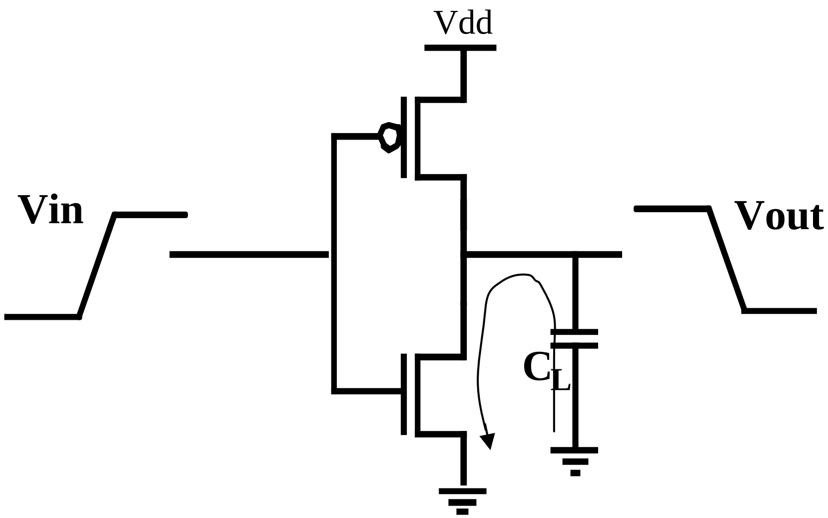
- Output 0-1 transition



Energy drawn from supply =
Energy stored in C_L =
Energy dissipated in pulldown device =

Switching Power

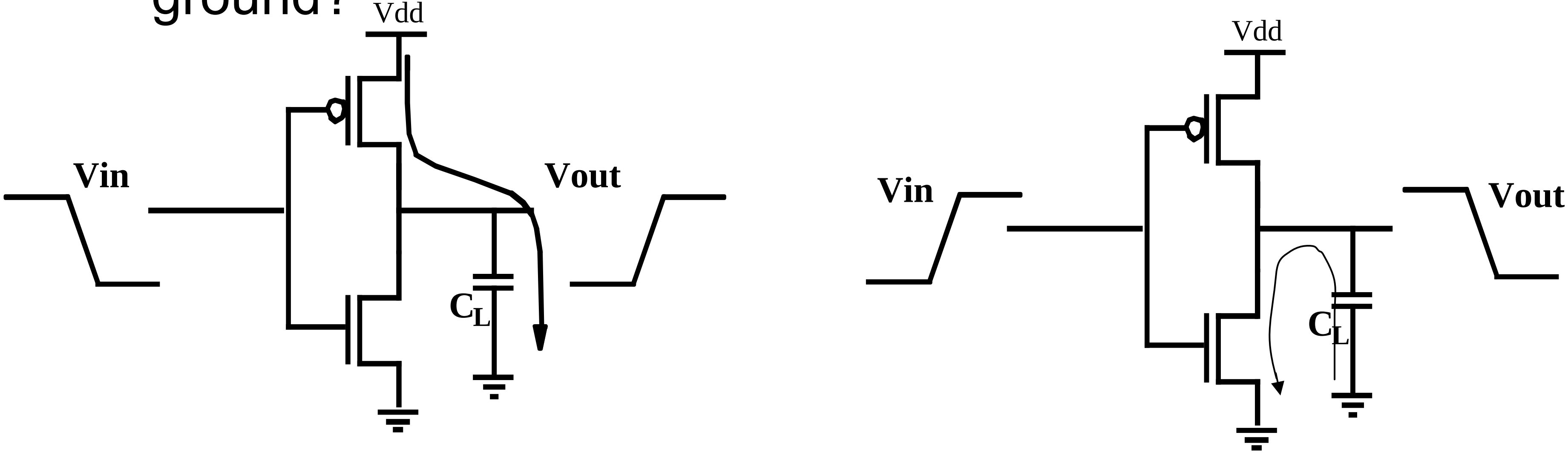
- Output 0-1 transition



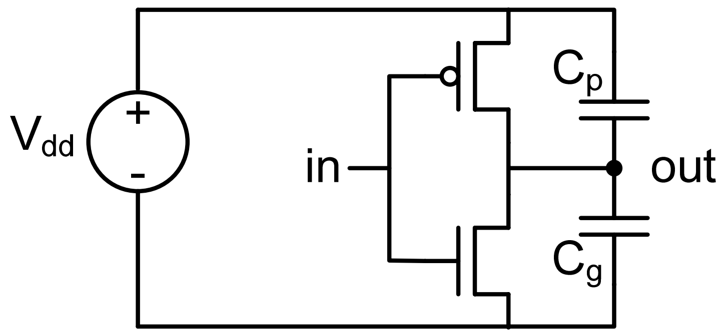
Energy drawn from supply =
Energy stored in C_L =
Energy dissipated in pulldown device =

Switching Power

- Over an entire charge-discharge sequence then …
  - Energy drawn from the supply =
  - Energy dissipated in the inverter =
- Energy provided by the supply is a straightforward way of tracking energy dissipation.
- But what if all load capacitance is not coupled to ground?



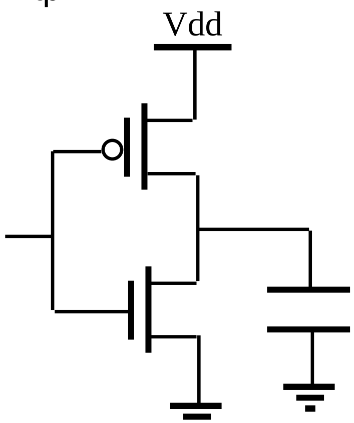
Switching Power

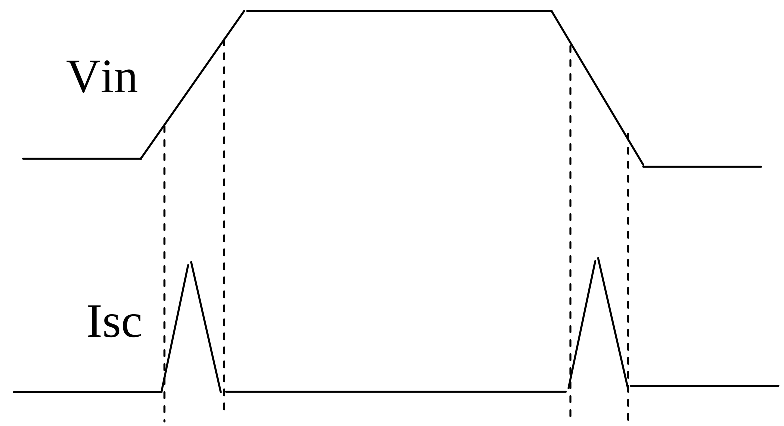


	Energy (supply)	Energy (PMOS)	Energy (NMOS)
Input 1→0			
Input 0→1			
Total			

Short Circuit Power (Crowbar)

- Rise and fall times are not ideal
  - Both nMOS and pMOS conducting between V_{tn} and (Vdd – $|V_{tp}|$)

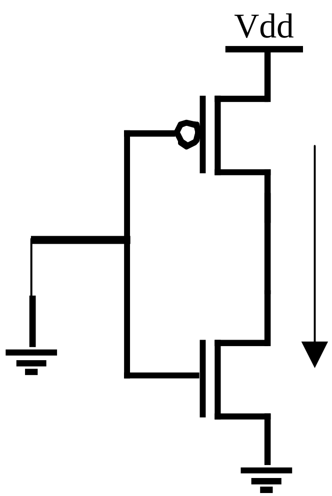




Energy dissipated =

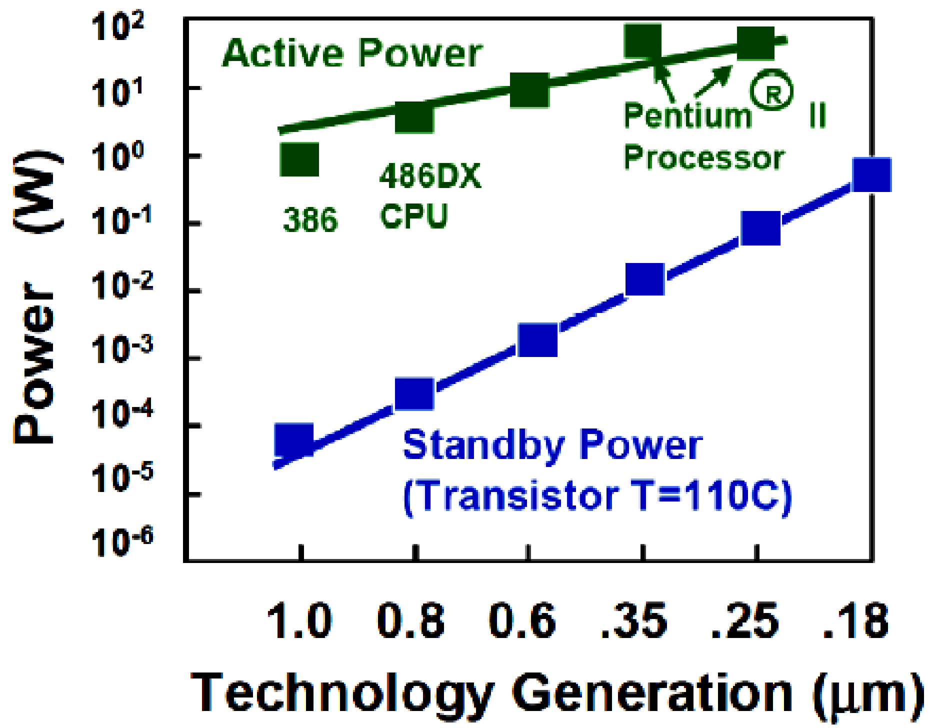
Short circuit power =

Static Power Consumption



Wasted energy …
Should be avoided as much as possible

$$I_{sub} \approx k_0 \frac{W}{L} e^{\frac{V_{gs}^{'} - \zeta V_{ds} - \gamma V_{sb}}{V_T}} \left(1 - e^{\frac{V_{ds}}{V_T}}\right)$$



Total power dissipation

$$P \approx V_{dd} \bullet k_0 \frac{W}{L} e^{\frac{nV'_{gs} - \zeta V_{ds} - \gamma V_{sb} - V_{th}}{V_T}} \left(1 - e^{\frac{V_{ds}}{V_T}} \right) + sCV_{dd}^2 f + V_{dd} I_{sc} t_{sc} f$$

- Total power dissipation is the combination of all three main sources of power dissipation.
  - Dynamic switching power
  - Dynamic short-circuit current (crowbar)
  - Static leakage (sub-threshold, junction)

Common methods for power minimization

$$P \approx V_{dd} \bullet k_0 \frac{W}{L} e^{\frac{nV_{gs}^{'} - \zeta V_{ds} - \gamma V_{sb} - V_{th}}{V_T}} \left(1 - e^{\frac{V_{ds}}{V_T}} \right) + sCV_{dd}^2 f + sV_{dd} I_{sc} t_{sc} f$$

Static Dissipation — Dynamic Dissipation

- Power minimization
  - Switching power usually dominates dynamic dissipation (Crowbar power accounts for about 10-15% of total dynamic dissipation)
  - Static leakage can dominate systems with low switching activity.
    - Typical value of s in datapaths is 10-20%
    - Important in systems with significant stand-by time

Common methods for power minimization

$$P \approx k_0 V_{dd} \frac{W}{L} e^{\frac{nV'_{gs} - \zeta V_{ds} - \gamma V_{sb} - V_{th}}{V_T}} + sCV_{dd}^2 f$$

Leakage Power Dissipation ⟷ Switching power

- Power minimization techniques usually involve controlling one of s, C, V_{dd}, f, V_{th} as per equation
- Such changes often involve tradeoffs (eg. area, power, complexity)

Altering *switching capacitance* for low power dissipation

$$P \approx k_0 V_{dd} \frac{W}{L} e^{\frac{nV_{gs}^{'} - \zeta V_{ds} - \gamma V_{sb} - V_{th}}{V_T}} + sCV_{dd}^2 f$$

Leakage Power Dissipation — Switching power

- Extremely important term in switching power.
- Clock gating (s = 2 for clock nets)
- Gate sizing: Size down gates in non-critical paths (delay "wall")
- Glitch avoidance
- Bus encoding / Data encoding
- Avoiding pre-charging/dynamic mechanisms or use conditional discharge (eg. in flip-flops)

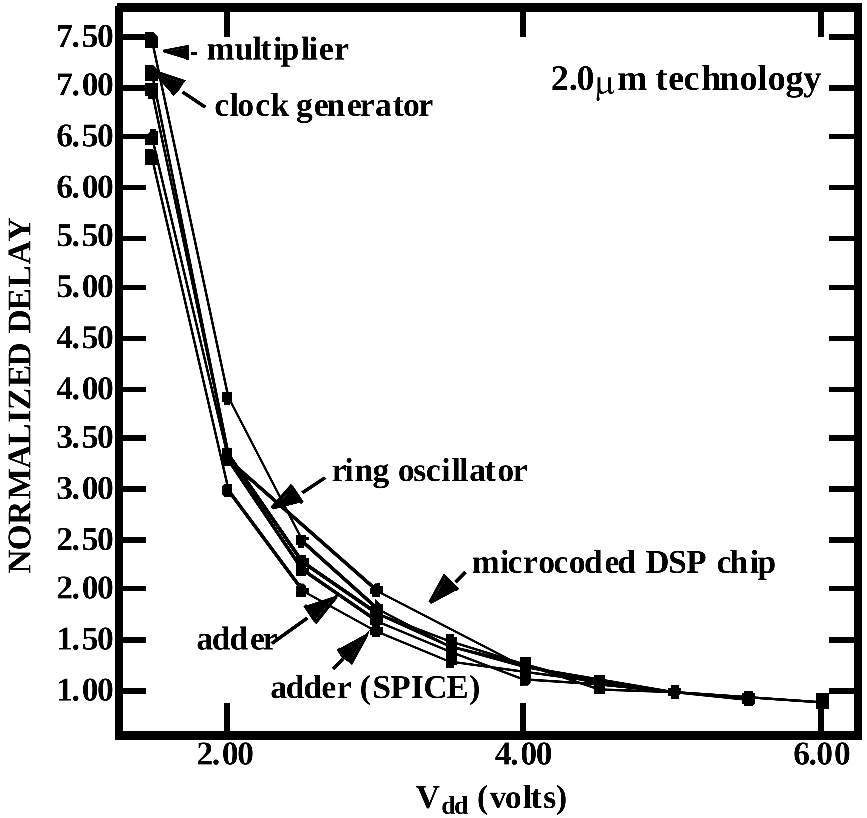
Altering V_{dd} for low power dissipation

$$P \approx k_0 V_{dd} \frac{W}{L} e^{\frac{nV'_{gs} + \zeta V_{ds} - \gamma V_{sb} - V_{th}}{V_T}} + sCV_{dd}^2 f$$

Leakage Power Dissipation — Switching power

- Supply Voltage provides quadratic reduction in switching ***energy.***
- Voltage scaling results in increased delay.
- Voltage-scaling (through sizing or pipelining)
- Dynamic voltage-scaling (at runtime)
- Low-swing logic/signaling (harms noise margins)
- CVS (Clustered Voltage Scaling)

Lower V_{dd} Increases Delay



$$T_d = \frac{C_L * V_{dd}}{I}$$

$$I \sim (V_{dd} - V_t)^2$$

$$\frac{T_{d(Vdd=2)}}{T_{d(Vdd=5)}} = \frac{(2) * (5 - 0.7)^2}{(5) * (2 - 0.7)^2}$$

$$\approx 4$$

The exponent will change next time we see this…

- **Relatively independent of logic function and style.**

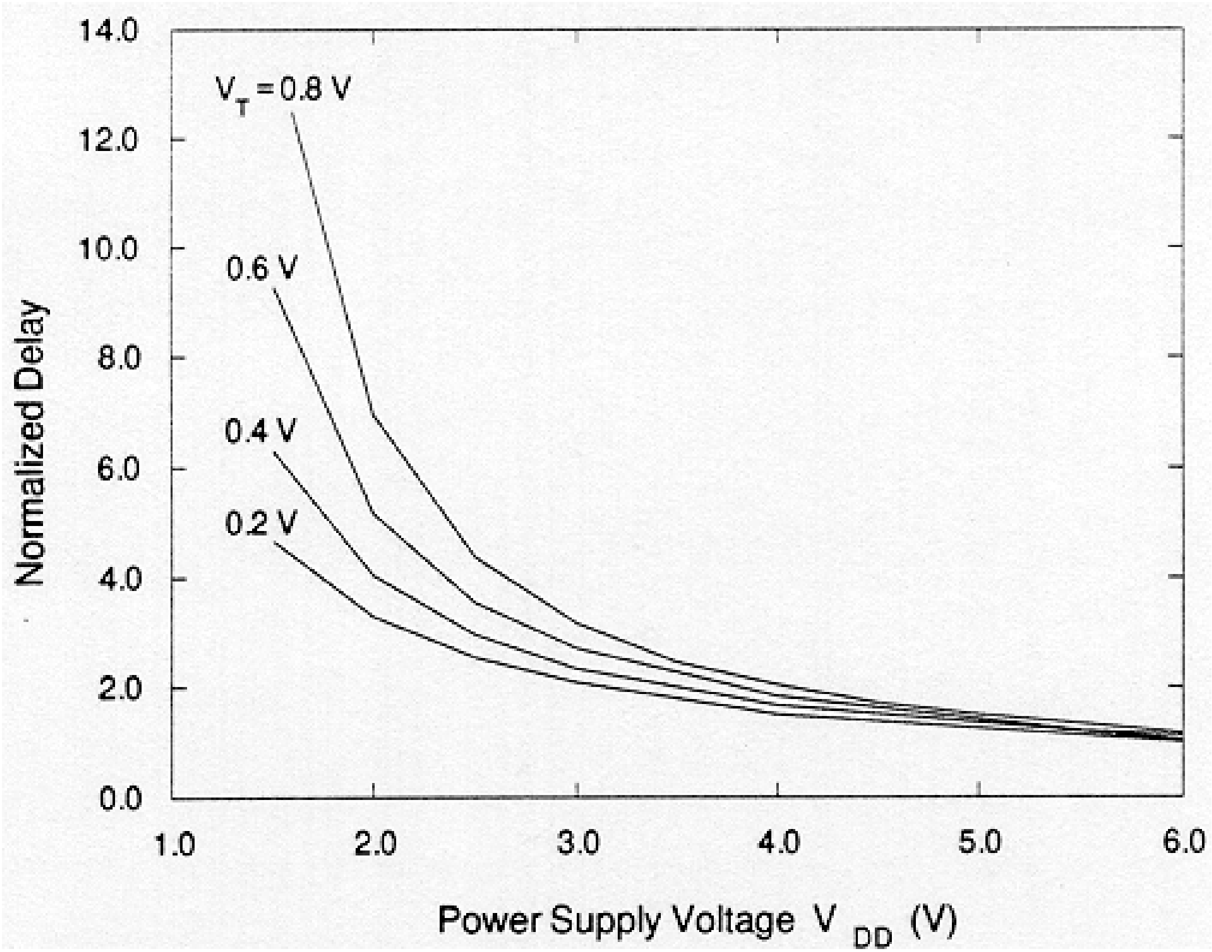
Altering *f* for low power dissipation

$$P \approx k_0 V_{dd} \frac{W}{L} e^{\frac{nV'_{gs} - \zeta V_{ds} - \gamma V_{sb} - V_{th}}{V_T}} + sCV_{dd}^2 f$$

Leakage Power Dissipation ⟷ Switching power

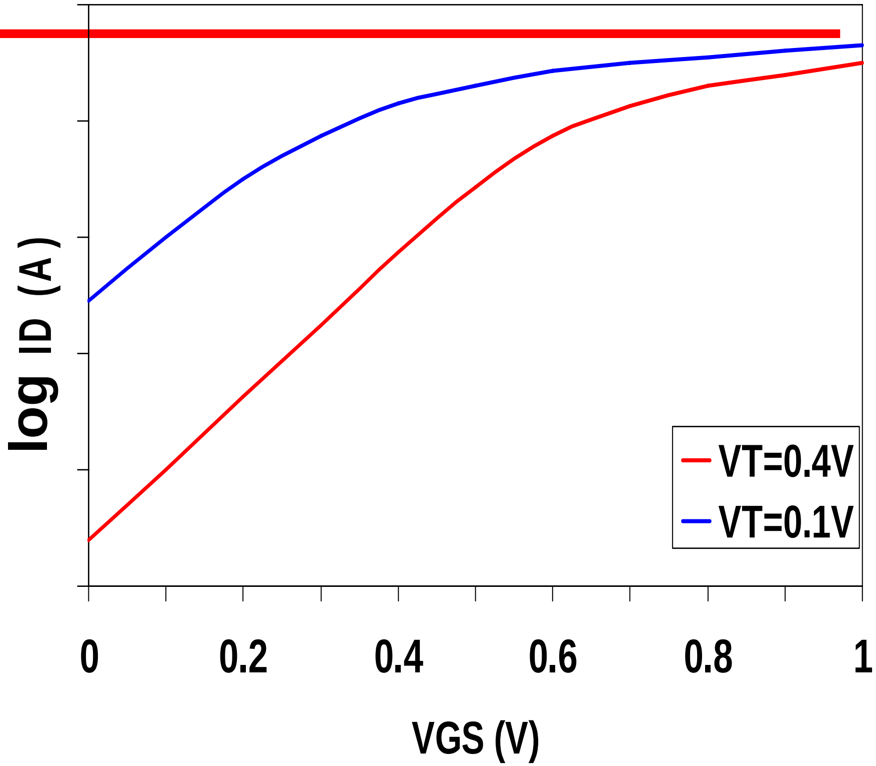
- Scaling the frequency reduces switching power
- Switching **energy** remains unchanged.
- However, leakage energy dissipation is increased
- Scaling only frequency does not achieve energy reduction
  - For the same application, energy dissipation actually increases
- A possible option for high performance cores which cannot transfer heat away from the chip fast enough.

Reducing V_{th} to offset delay penalty



Leakage as a Function of V_T

- Reducing the V_T **increases** the sub-threshold leakage current (exponentially)
  - ~90mV reduction in V_T increases leakage by 10X
- But, reducing V_T **decreases** gate delay (increases performance)



Altering V_{th} for low power dissipation

$$P \approx k_0 V_{dd} \frac{W}{L} e^{\frac{nV'_{gs} - \zeta V_{ds} - \gamma V_{sb} - V_{th}}{V_T}} + sCV_{dd}^2 f$$

Leakage Power Dissipation — Switching power

- Dual V_{th} design
  - Determine the critical paths during the design phase, use low V_T devices on those paths for speed (requires dual-Vt, variation issues)
  - Use a high V_T in rest of logic to control leakage
  - Can provide total leakage reduction of up to 80%
- Body biasing
  - Change the substrate potential to reduce leakage current
- MTCMOS
  - Use the “stack effect” and high V_{th} devices to achieve leakage reduction
  - Performance degradation (frequency, start-up time)

Summary

- Power reduction is as important as increasing speed in IC design today.
- Three major components of power in CMOS
  - Dynamic: charging capacitors → dominant
  - Short-circuit: small, typically ignore
  - Static: subthreshold leakage, growing fast already dominant in certain structures (eg. memories)