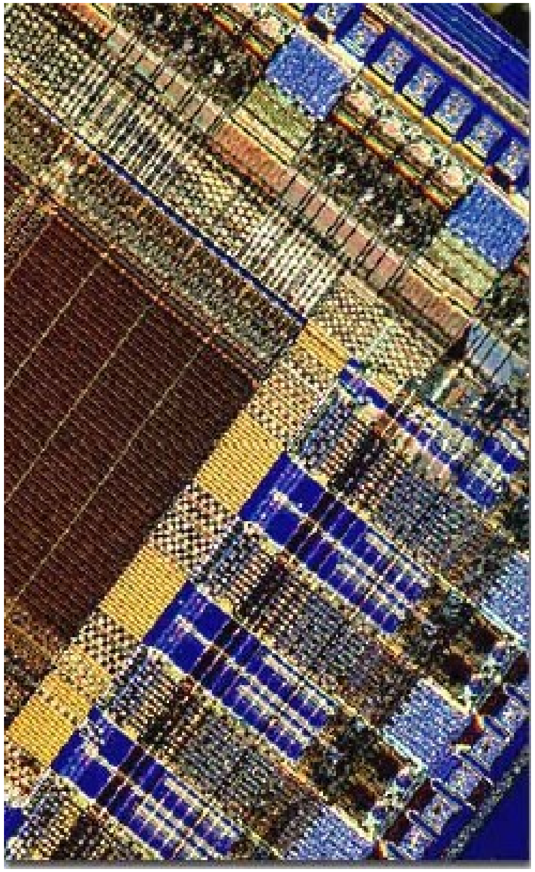
CMOS Digital Integrated Circuits



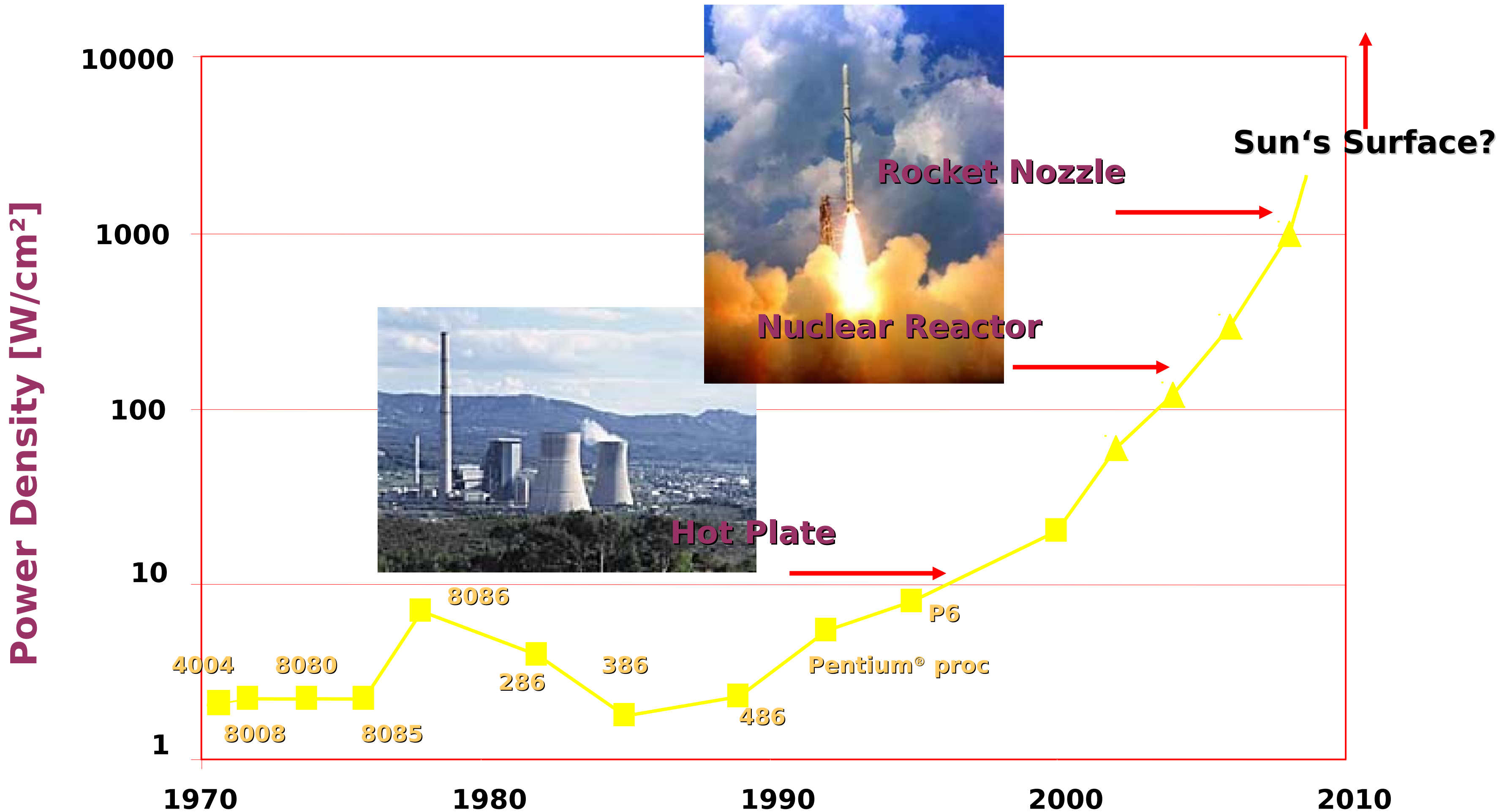
Lec 14

Low-Power CMOS Logic Circuits

Why Worry About Power?

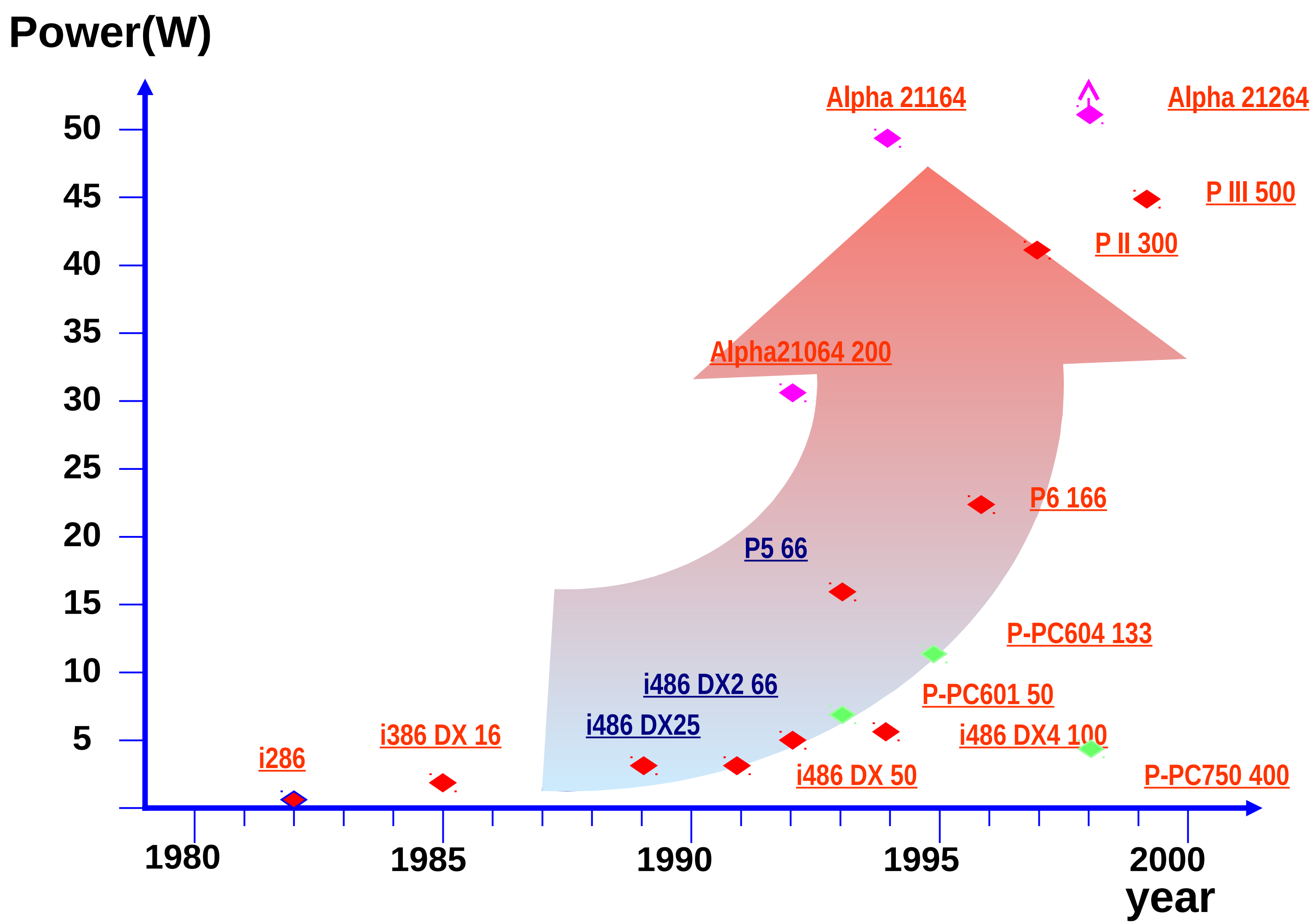
- Battery-powered devices
  - GSM phone, UMTS phone, MP3 player, PDAs
    - » Complexity increases
    - » Energy budget remains the same
- Complex high-speed devices
  - Thermal problems
  - Expensive packaging

Evolution in Power Density



Source: adapted from Intel

Microprocessor Power Dissipation



Power and Energy

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

- **Instantaneous Power:** $P(t) = i_{DD}(t)V_{DD}$

- **Energy:** $E = \int_0^T P(t)dt = \int_0^T i_{DD}(t)V_{DD}dt$

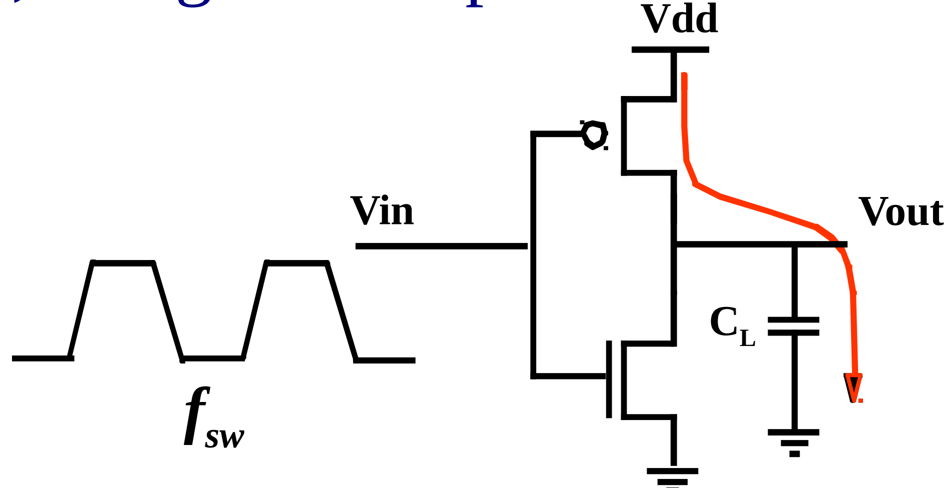
- **Average Power:** $P_{avg} = \frac{E}{T} = \frac{1}{T}\int_0^T i_{DD}(t)V_{DD}dt$

Overview of Power Consumption

- $P_{total} = P_{dynamic} + P_{short\text{-}circuit} + P_{leakage} + P_{static}$
- Dynamic (Switching) Power Consumption ($P_{dynamic}$)
  - Charging and discharging capacitors
- Short Circuit Power Consumption ($P_{short\text{-}circuit}$)
  - Short circuit path between supply rails during switching
- Leakage Power Consumption ($P_{leakage}$)
  - Leaking diodes and transistors
- Static Power Consumption (P_{static})

Dynamic Power (1/3)

- Dynamic power is required to charge and discharge load capacitances when transistors switch.
- One cycle involves a rising and falling output.
- On rising output, charge $Q = CV_{DD}$ is required
- On falling output, charge is dumped to GND



Dynamic Power (2/3)

- **Energy Per Transition**

$$E = \int_0^T i_{DD}(t) V_{DD} dt = V_{DD} \int_0^T i_{DD}(t) dt = C_L V_{DD}^2$$

  - Not a function of frequency!
  - 50% dissipated by Ron
  - 50% stored/delivered in/by CL

- **Dynamic Power**

$$P_{dynamic} = C_L \times V_{DD}^2 \times f$$

- Not a function of transistor sizes!
- Need to reduce C_L, V_{DD}, and f to reduce power.

Dynamic Power (3/3)

- $P_{dynamic}$ = Energy/per-transition $\times$ Transition rate

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

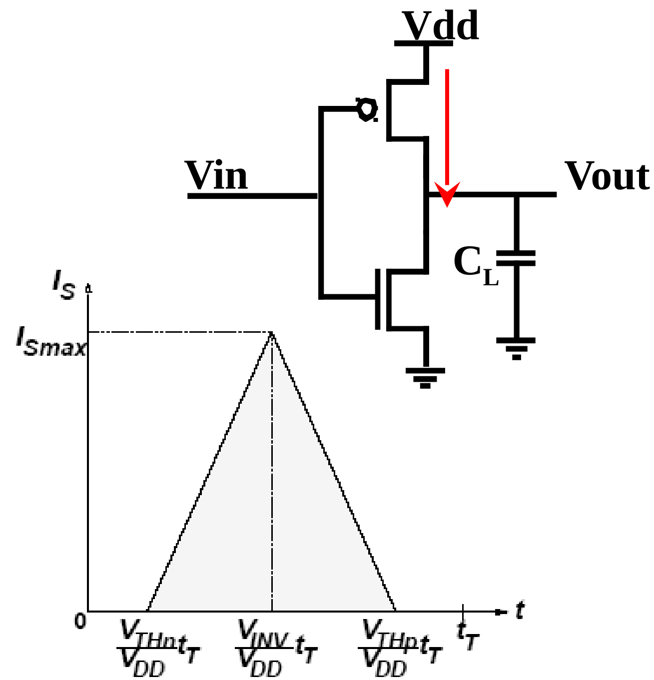
  $= C_L \times V_{DD}^2 \times P_{0\rightarrow 1} \times f$

  $= C_{eff} \times V_{DD}^2 \times f$

- C_{eff} = effective capacitance = $C_L \times P_{0\rightarrow 1}$
- Power dissipation is data dependent
  - ***Function of Switching Activity***
- Activity Factor ($P_{0\rightarrow 1}$)
  - Clock signal: $P_{0\rightarrow 1}$(clk) = 1
  - Data signal: $P_{0\rightarrow 1}$(data) < 0.5

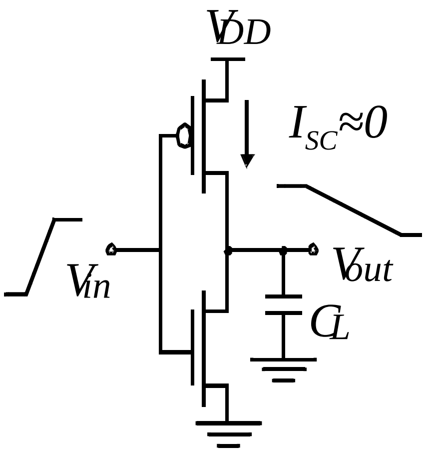
Short Circuit Current (1/2)

- When transistors switch, both nMOS and pMOS networks may be *momentarily ON at once*
- Leads to a blip of “short circuit” current.
- ~ 15% of dynamic power
  - ~85% to charge capacitance C_L
- NMOS and PMOS on
  - Both transistors in saturation
- Long rise / fall times
  - Slow input transition
  - Increase short circuit current

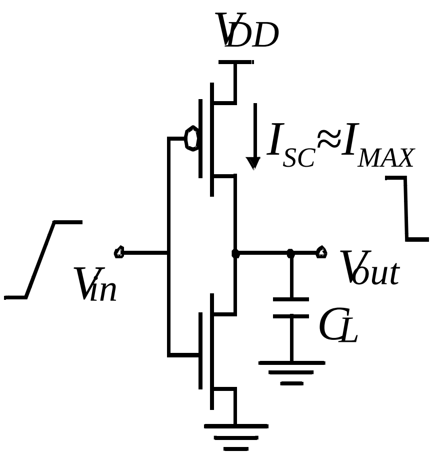


Make input signal transitions fast to save power!

Short Circuit Current (2/2)

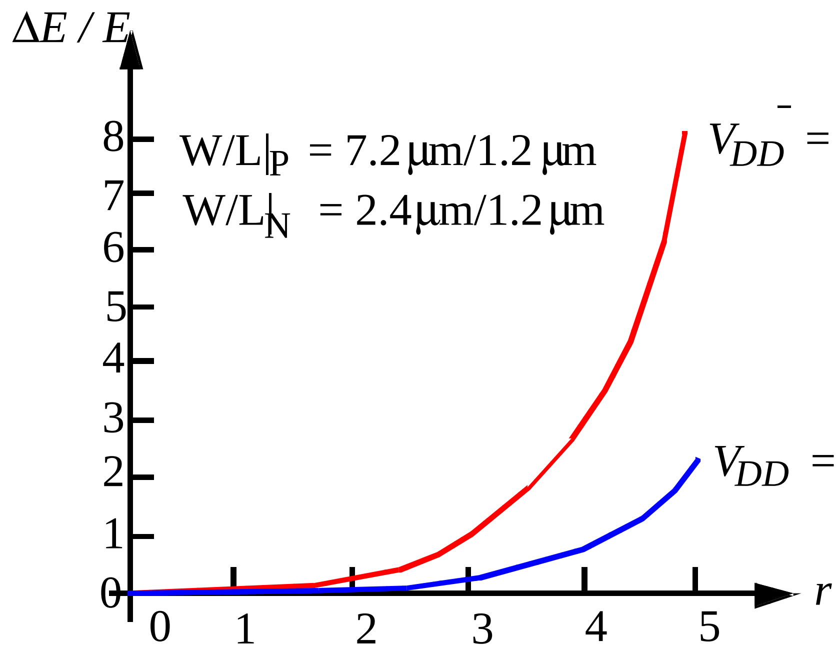


Large capacitive load



Small capacitive load

Because of finite slope of input signal, there is a period when both PMOS and NMOS device are “on” and create a path from supply to ground



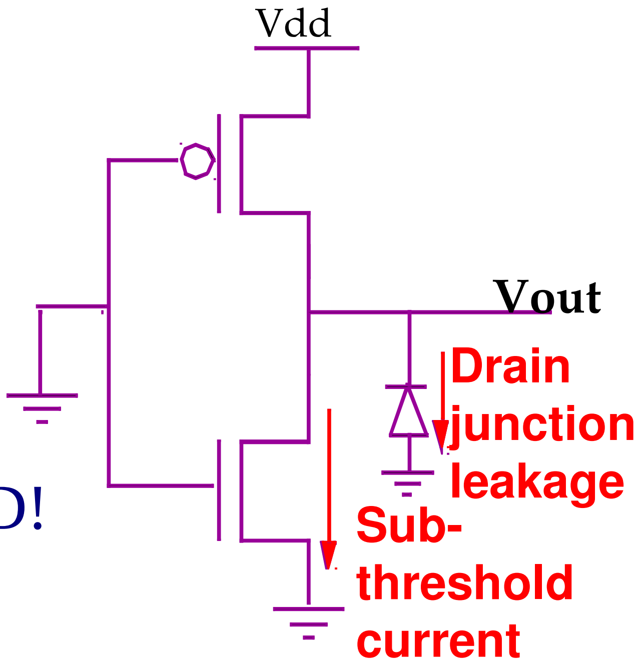
The power dissipation due to short circuit currents is minimized by matching the rise/fall times of the input and output signals.

Leakage

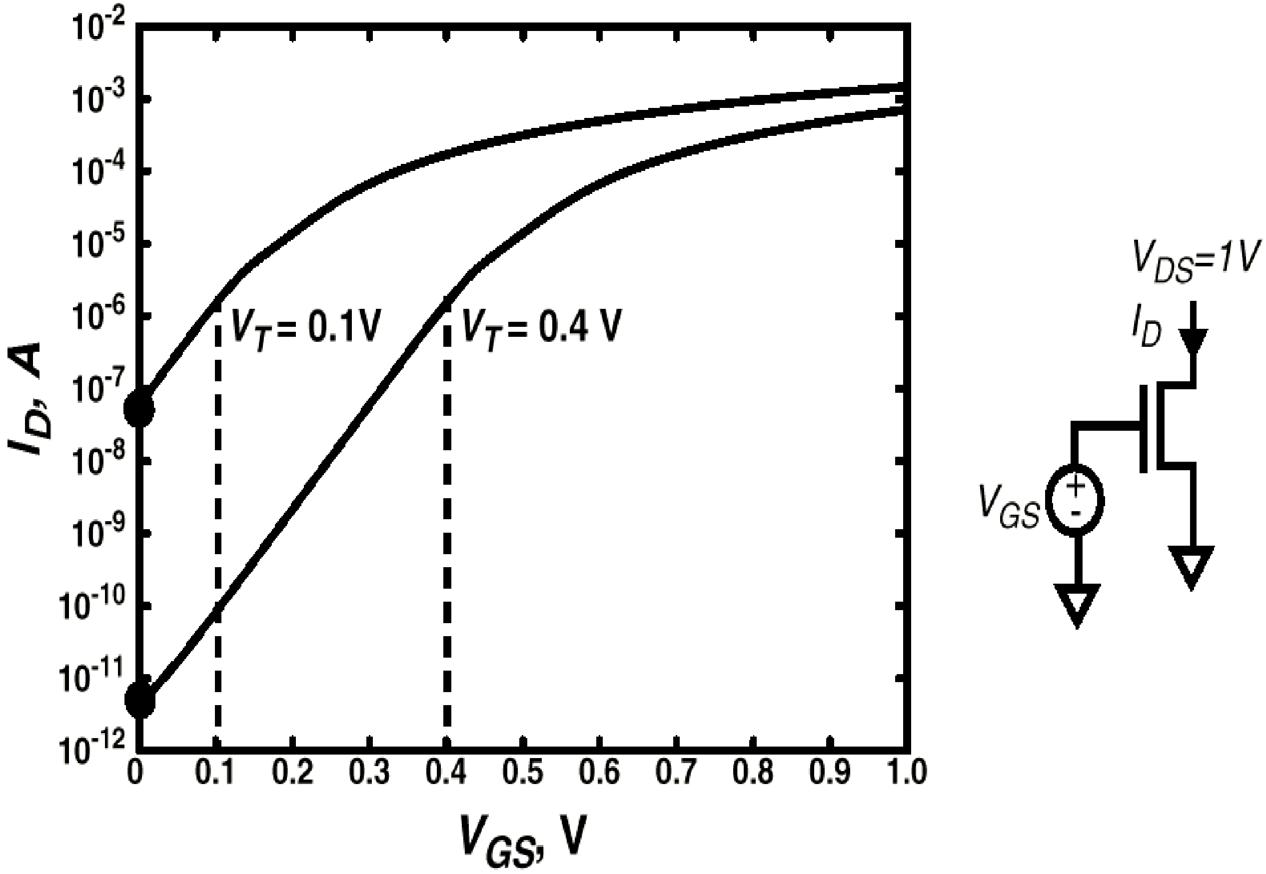
- Sub-threshold current
  - Transistor conducts below Vt
  - For sub-micron relevant
    - » VDD / Vt ratio smaller
    - » Can dominate power consumption!
    - » Especially in idle mode.

  Charge nodes fully to VDD!

  Discharge nodes completely to GND!
- Drain leakage current
  - Reverse biased junction diodes



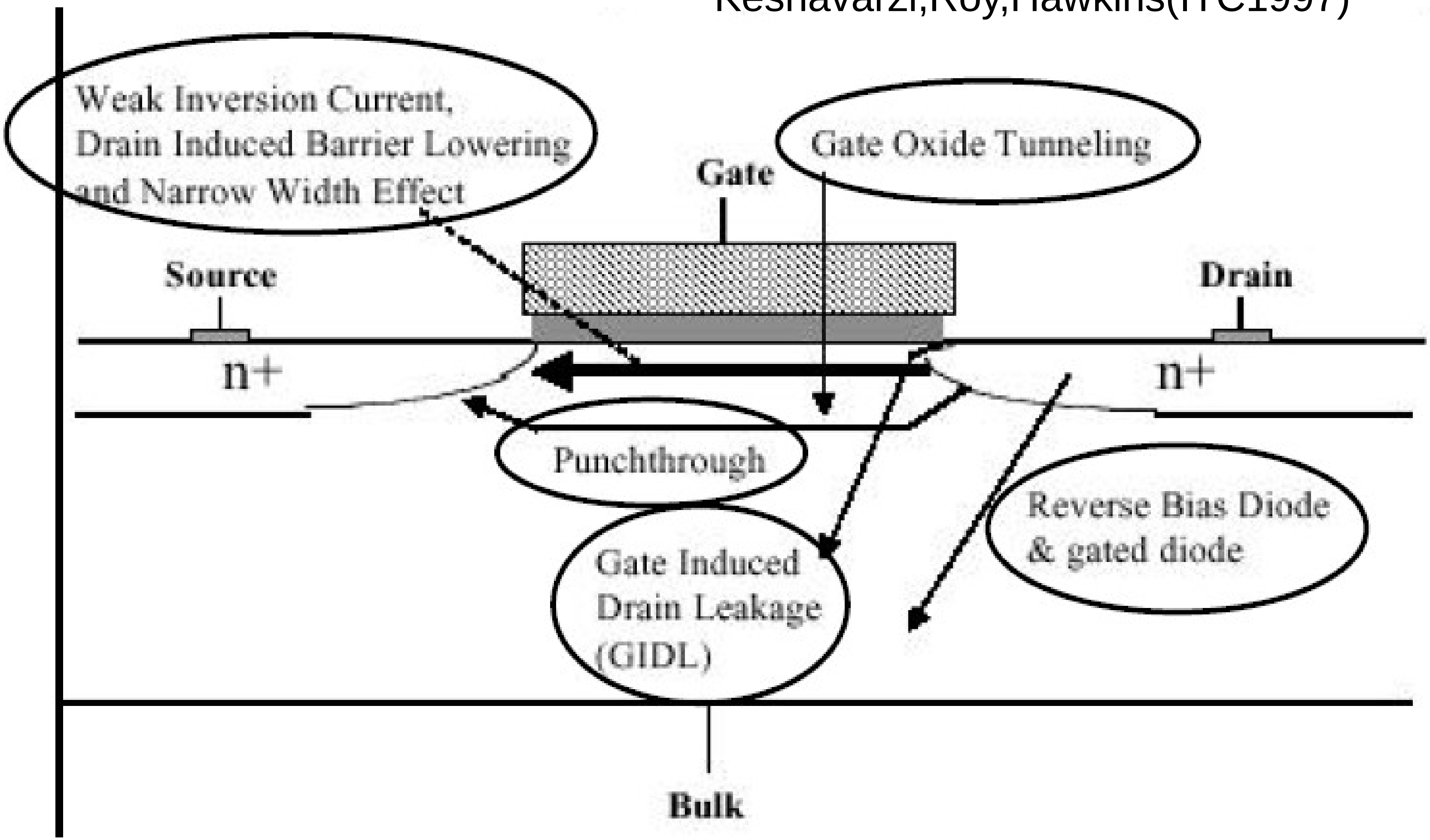
Sub-threshold Leakage Component



- Leakage control is critical for low-voltage operation

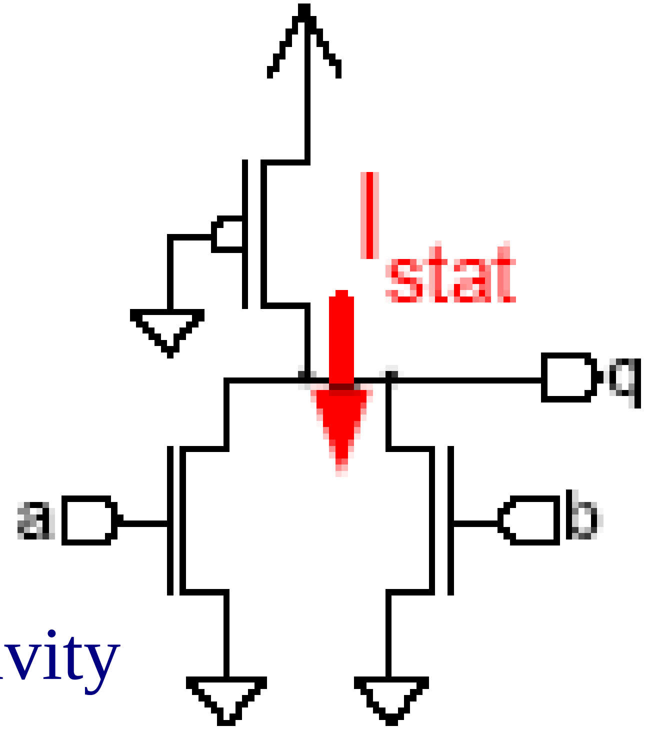
Source of Leakage Current

Keshavarzi,Roy,Hawkins(ITC1997)



Static Power Consumption

- Pseudo-NMOS logic style
  - PMOS as resistor
  - PDN as static CMOS logic
- Static current
  - When output low
- Power consumption
  - Even without switching activity



Power Dissipation for Various CMOS Circuits

Chip	Intel 386	DEC Alpha 21064	Cell based ASIC
Minimum feature size	1.5μm	0.75μm	0.5μm
Number of gates	36,808	263,666	10,000
f_{CLK}	16MHz	200MHz	110MHz
V_{DD}	5V	3.3V	3V
P_{total}	1.41W	32w	0.8w
Logic gates	32%	14%	9%
Clock Distribution	9%	32%	30%
Interconnect	28%	14%	15%
I/O drivers	26%	37%	43%

Design for Low Power (1/4)

- **Good Ideas**
  - On all levels
    - » Software
    - » Algorithm
    - » Architecture
    - » Gate
    - » Transistor
    - » Process technology
- **Bad Ideas**
  - Apply one method
  - Do it as late as possible

Consider low power design from the beginning!

Design for Low Power (2/4)

- **System Level**
  - Power management
    - » Power-down mode
    - » Global clock gating
    - » Dynamic voltage scaling
  - Hardware/software co-design
    - » Early (simplified) power estimation
    - » Partitioning of functionality
    - » Minimum instructions for execution not code size

Design for Low Power (3/4)

- **Algorithm**
  - Arithmetic
    - » Choice of number representation
    - » Pre-computation
  - Concurrency
    - » Parallelism - Trade area for power
    - » To reduce frequency

Design for Low Power (4/4)

- **Architecture**
  - Pipelining
    - » Allows voltage scaling: Increased throughput because frequency could be increased => lower supply voltage instead
  - Redundancy
    - » Minimize shared resources to lower signal activity (buses)
  - Data encoding
    - » Energy efficient state encoding
    - » Example: Gray code, One hot encoding
  - Clocking
    - » Gated clocks
    - » Self-timed circuits

Voltage Scaling (1/3)

- $P_{dynamic} = f \times C \times V_{DD}^2$
  - Quadratic influence
- Delay
  - Increased
- Power delay product
  - Improved

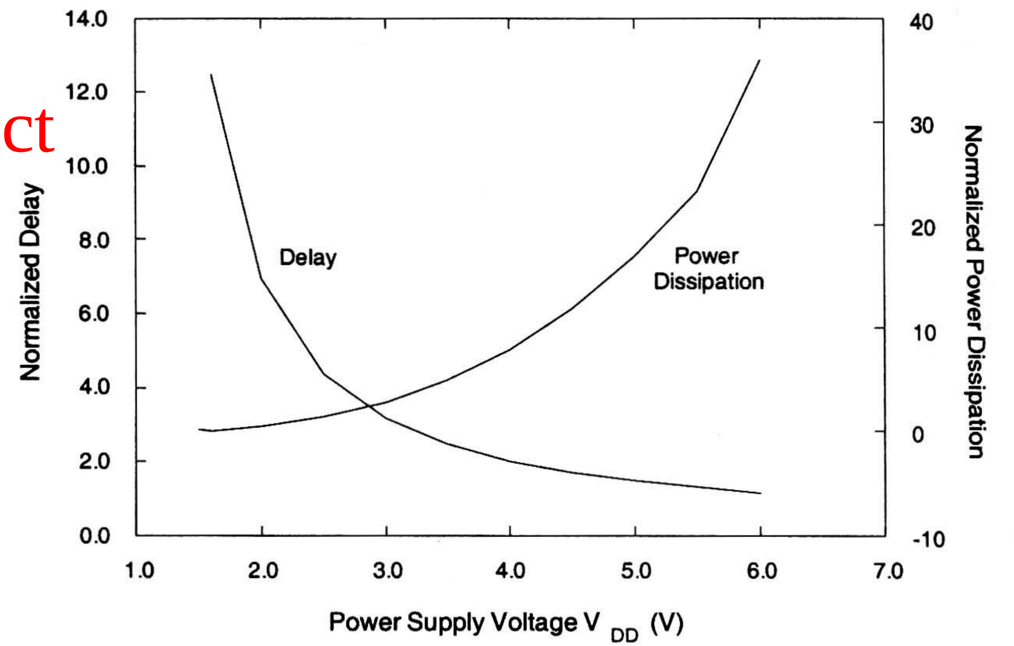


Figure 11.9. Normalized propagation delay and average switching power dissipation of a CMOS inverter, as a function of the power supply voltage V_{DD}.

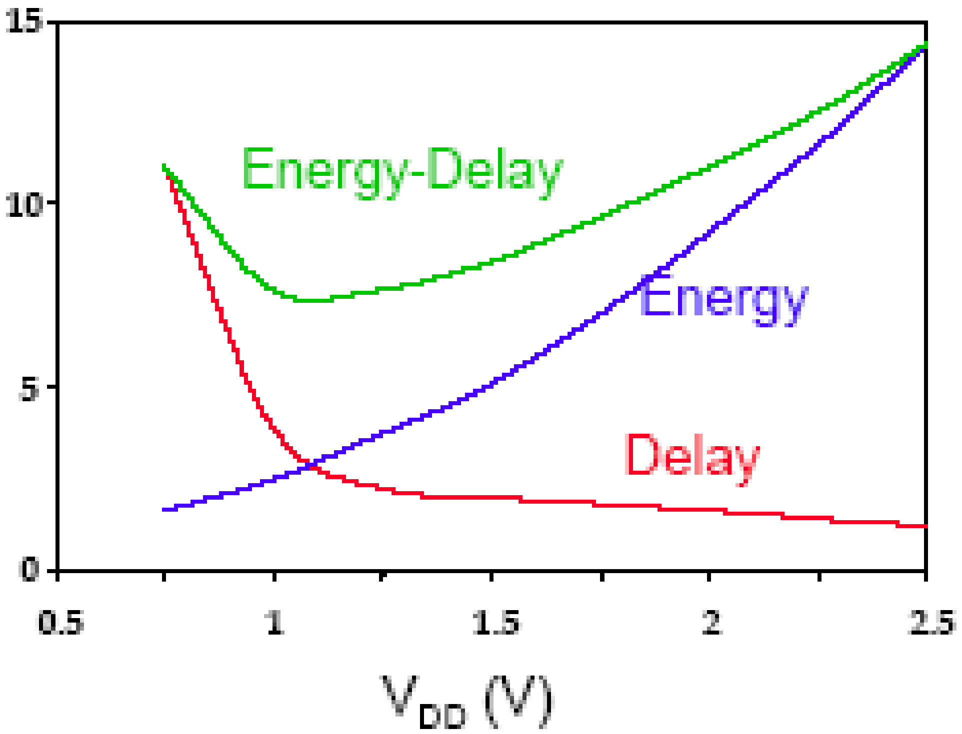
Voltage Scaling (2/3)

EDP (Energy Delay Product)

- Measure for energy efficiency

■ Lower supply voltage

- Less energy
- Increased delay

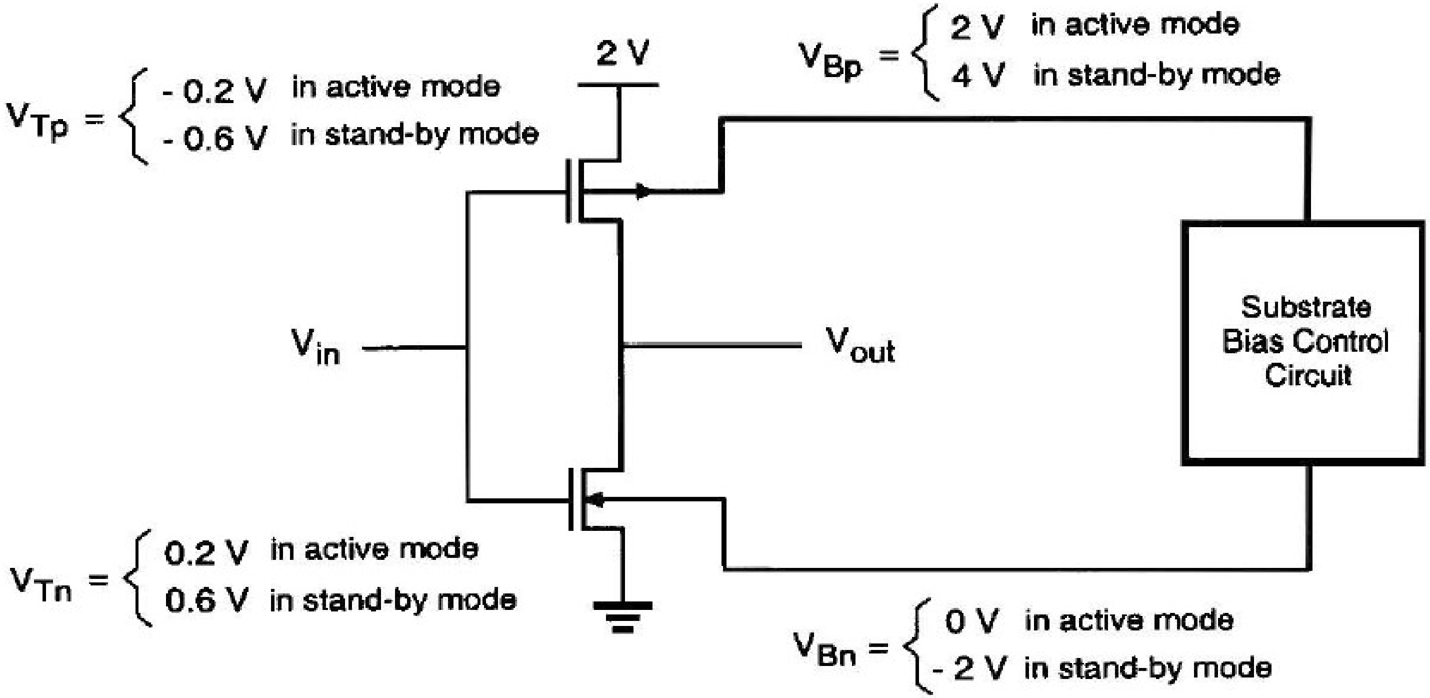


Voltage Scaling (3/3)

- Dual voltage supply
- Internal voltage
  - Reduced internal voltage 1.2V
    - » For low power operation
- External voltage
  - Compatible IO voltage 3.3V
    - » To interface other ICs

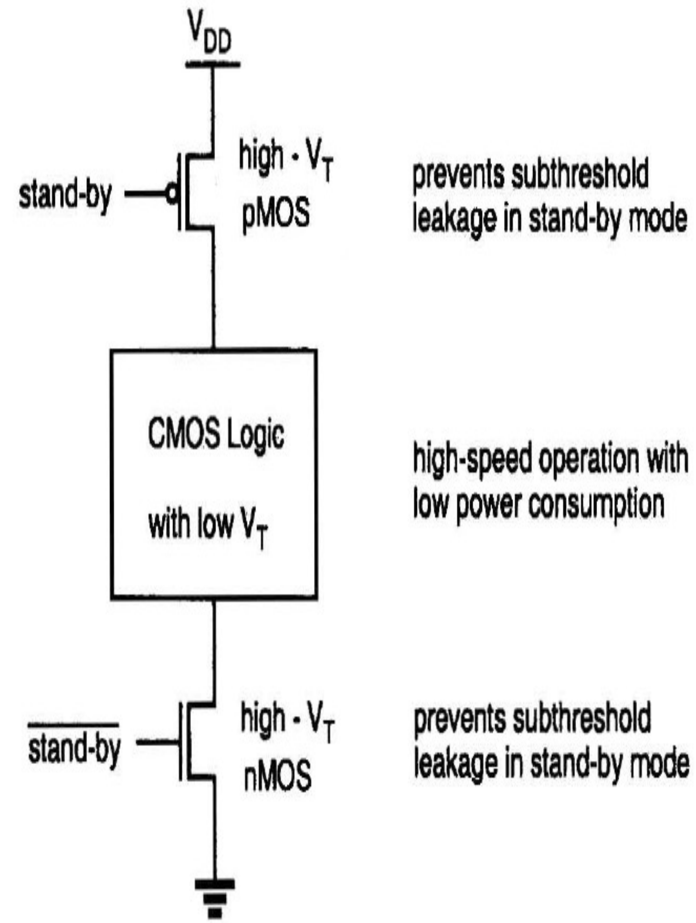
Variable-Threshold CMOS (VTCMOS) Circuits

- An efficient way to reduce ***subthreshold leakage currents***
  - Require twin-well or triple-well CMOS technology to apply different substrate bias voltages.
  - Separate power pins may be required if the substrate bias voltages level are not generated on-chip.



Multiple-Threshold CMOS (MTCMOS) Circuits

- **Active Mode**
  - High-V_T transistors are turned on.
  - Logic gates consisting of low-V_T transistors can operate with low switching power dissipation and small propagation delay.
- **Standby Mode**
  - High-V_T transistors are turned off, and the conduction paths can be effectively cut off.
- The series-connected standby transistors increase the overall circuit area and add extra parasitic capacitance and delay.



Pipelining Approach (1/3)

- $P_{reference} = f_{clock} \times C_{total} \times V_{DD}^2$
  - C_{total}
    - » Capacitance switched in the input and output register array
    - » Capacitance switched to implement the logic function

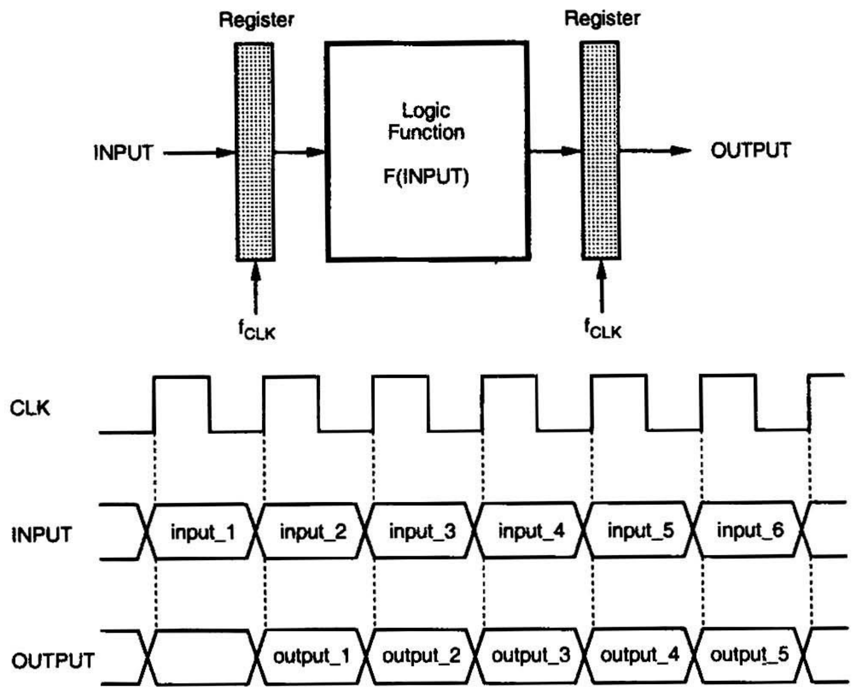


Figure 11.15. Single-stage implementation of a logic function and its simplified timing diagram.

Pipelining Approach (2/3)

- ***Speed Improvement***
  - **By inserting N pipeline registers, clock frequency (throughput) my be increased about N times**
- ***Power Saving***
  - **One may keep pipelined clock frequency constant by reducing supply voltage to save power**

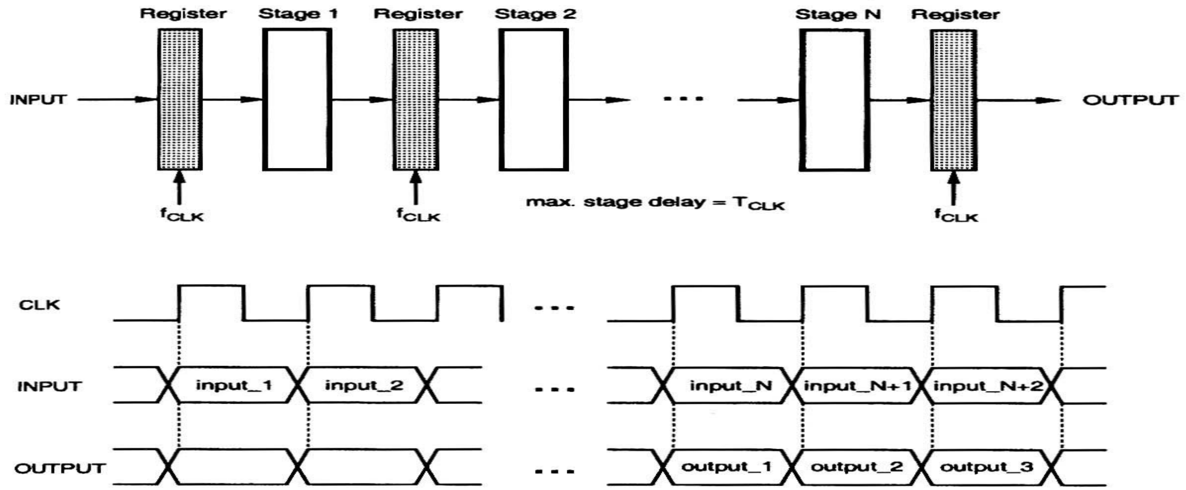


Figure *11.16.* N-stage pipeline structure realizing the same logic function as in Fig. 11.15. The maximum pipeline stage delay is equal to the clock period, and the latency is N clock cycles.

Pipelining Approach (3/3)

Power Saving

τ_p(pipeline_stage) = $\tau_{P,max}$(input-to-output)/N = T_{CLK}

- The logic blocks between two successive register *can operate N-times slower* while maintaining the same functional throughout.
- V_{DD} can be reduced to a new value $V_{DD,new}$

$P_{pipeline} = [C_{total} \times + (N-1)\, C_{reg}] \times V_{DD,New}^2 \times f_{clock}$

$$\frac{P_{pipeline}}{P_{reference}} = \frac{\left[C_{total} + (N-1)C_{reg}\right] V_{DD,New}^2 f_{CLK}}{C_{total} V_{DD}^2 f_{CLK}}$$

$$= \left[1 + \frac{C_{reg}}{C_{total}}(N-1)\right] \frac{V_{DD,New}^2}{V_{DD}^2}$$

- Example: V_{DD}=5V, f_{CLK}=20MHz, four stage pipeline. V_{th}=0.8V, C_{reg}/C_{total}=0.1 $\Rightarrow$ Each stage can four times slower than the original system. $V_{DD,New}$=2V(See Fig11.10).
- The overall power reduction factor is 0.2. The power saving is about 80%.

Parallelism

- Preserve original throughput by lowering supply voltage to save power dissipation

$$P_{parallel} = NC_{total} \cdot V_{DD,New}^{2} \cdot f_{CLK}/N + C_{reg} \cdot V_{DD,New}^{2} \cdot f_{CLK}$$

$$= (1 + C_{reg}/C_{total})\, C_{total} \cdot V_{DD,New}^{2} \cdot f_{CLK}$$

$$\frac{P_{parallel}}{P_{reference}} = \frac{V_{DD,New}^{2}}{V_{DD}^{2}} \left(1 + \frac{C_{reg}}{C_{total}} \right)$$

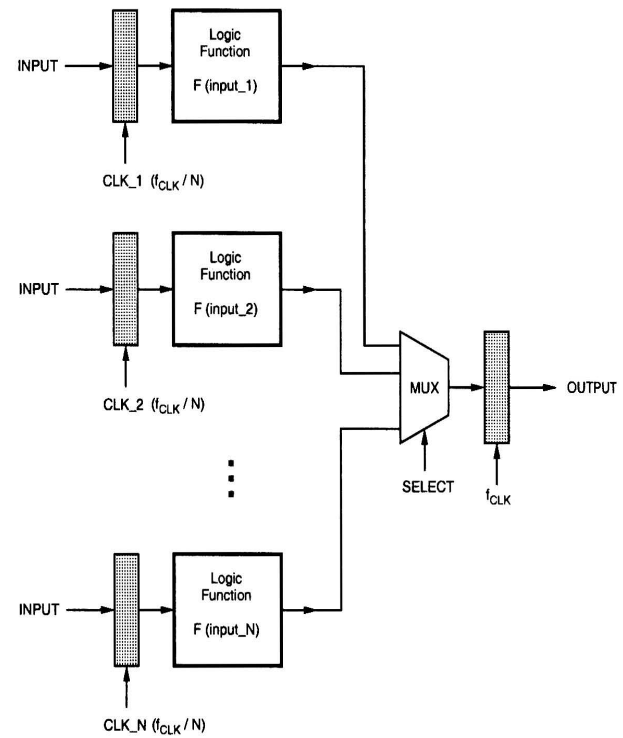


Figure 11.17. N-block parallel structure realizing the same logic function as in Fig. 11.15. Notice that the input registers are clocked at a lower frequency of (f_{CLK} / N).

- **What do you pay?**
  - Area is increased.
  - Latency is increased.
- Can combine pipelining with parallelism to further improve the speed and power.

Switching Activity Reduction (1/5)

Power Consumption is Data Dependent

- Static Circuit
  - Example 1: 2 input static NOR gate

    Assume P(A=1)=1/2, P(B=1)=1/2.

    P(out=1)=1/4

    $P_{0\rightarrow 1}=P(out=0)P(out=1)=3/4\times 1/4=3/16$

    $C_{eff}=3/16\times$

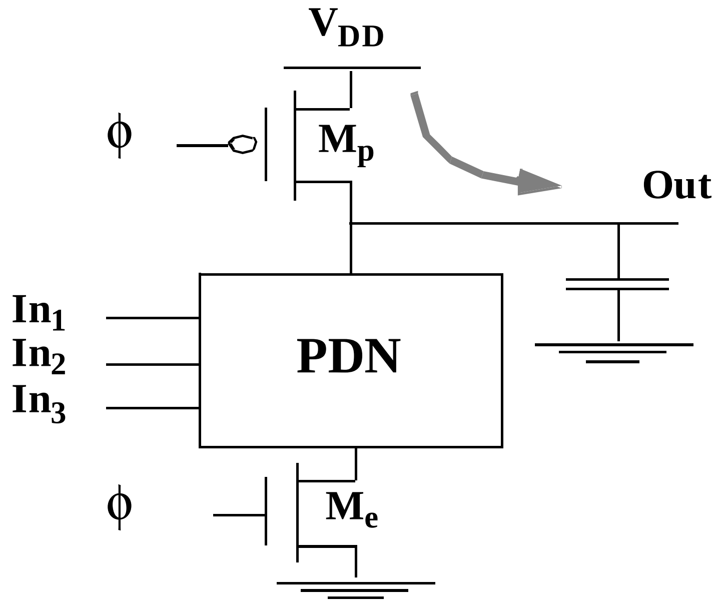
A	B	Out
0	0	1
0	1	0
1	0	0
1	1	0

Truth Table of 2 input NOR gate

Switching Activity Reduction (2/5)

Power Consumption is Data Dependent

- Dynamic Circuit



Power is Only Dissipated when Out=0!

$$C_{EFF} = P(Out=0).C_L$$

Switching Activity Reduction (3/5)

Power Consumption is Data Dependent

- Dynamic Circuit
  - Example 2: 2 input dynamic NOR gate

    Assume P(A=1)=1/2, P(B=1)=1/2.

    P(out=0)=3/4

    $C_{eff}=3/4 \times C_L$

Switching activity is always higher in dynamic circuits

Switching Activity Reduction (4/5)
Glitch Reduction

- **Dynamic hazards**
  - Caused by unbalanced delays
  - Usually 8% - 25% of dynamic power
- **Suspicious for glitches**
  - Deep logic depth
  - Ripple of carry in adder

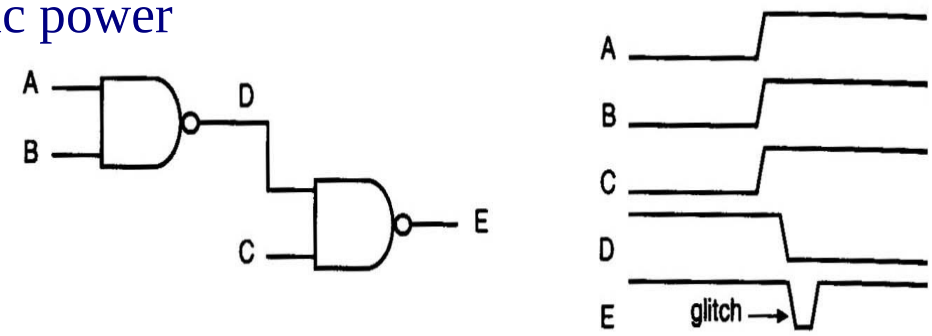


Figure 11.22. Signal glitching in multi-level static CMOS circuits.

- **Relief**
  - Equalize lengths of timing paths through design.
  - Reduce logic depth: Pipelining

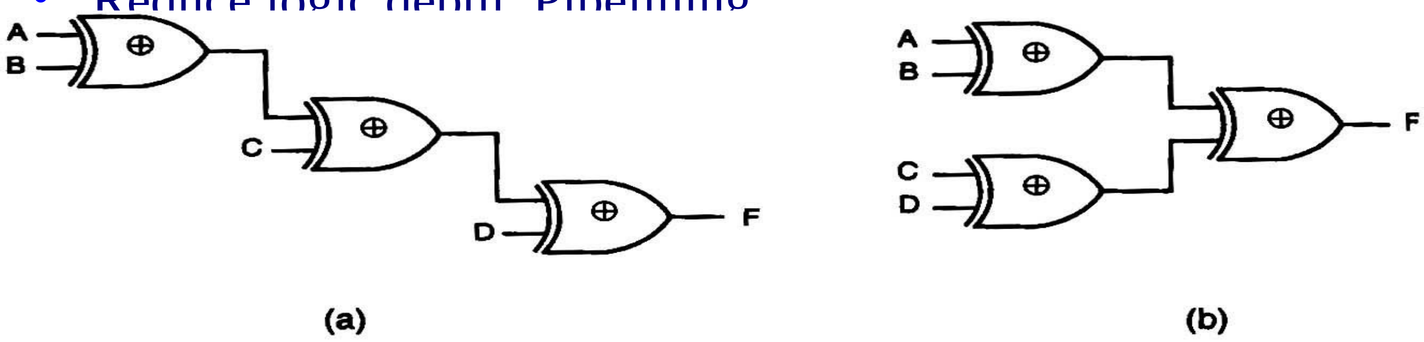
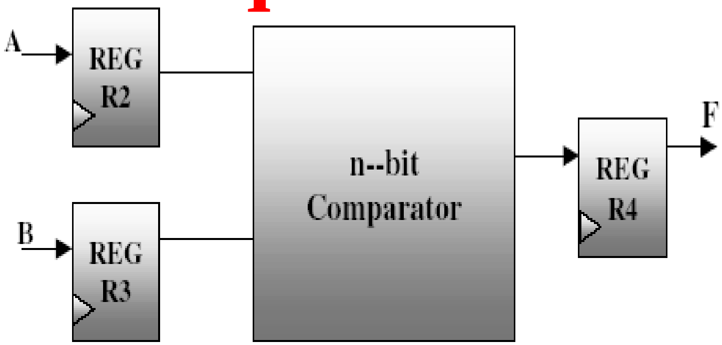
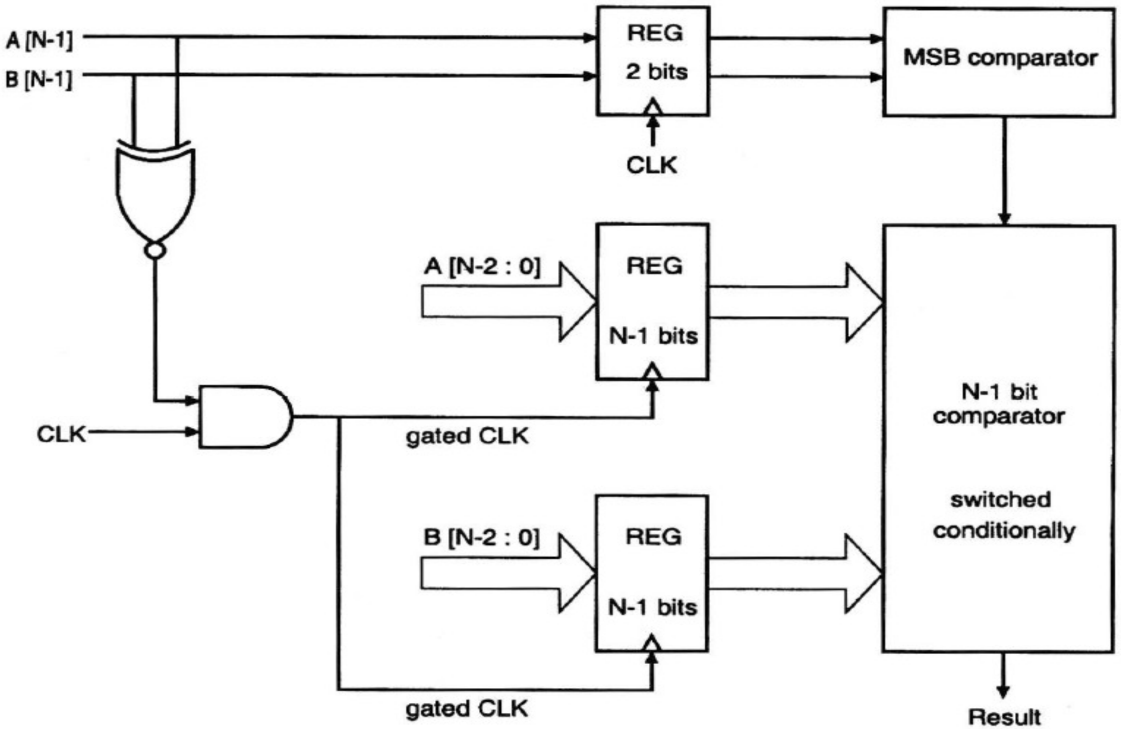


Figure 11.23. **(a) Implementation of a four-input parity (XOR) function using a chain structure. (b) Implementation of the same function using a tree structure which will reduce glitching transitions.**

Switching Activity Reduction (5/5)
Pre-Computation Technique

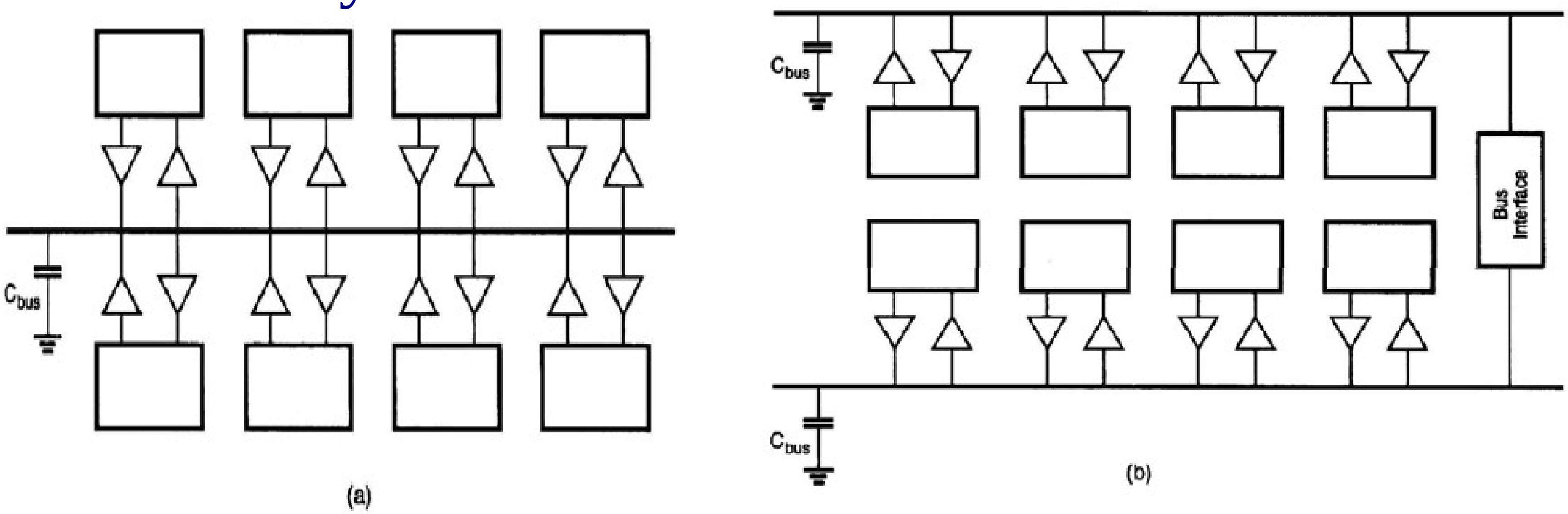
- Saves power by not enabling registers R2 and R3 in half (50%) of cases





Reduction of Switched Capacitance

- Resource Sharing
  - Causes switching overhead
  - Increases effective capacitance
- Global buses vs. Local interconnect
- Locality: Shorter wires



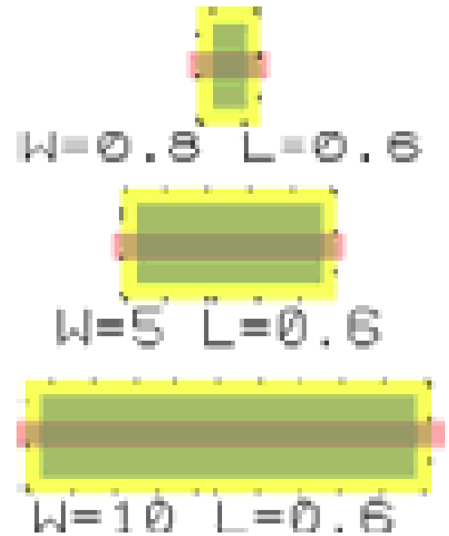
Transistor Sizing

Use Minimal Transistor Where Possible

- Transistor width W
  - Current driving capability

    $I_D = K \times (W/L) \times ......$
  - Capacitance

    $C = C_{OX} \times W \times L$
  - Large W
    - For dominating interconnect
- Minimum transistors
  - Lowest capacitance
  - Optimal for low power



Design for Low Power (Cont.)

- **Process Technology**
  - V_{DD} reduction
  - Threshold voltage
    - » High threshold voltage
    - » Double-threshold devices
      - Low threshold for high speed
      - High threshold for low power
  - Silicon on insulator (SOI)

Conclusion

- Power consumption
  - Dynamic, Short circuit, Leakage, Static
- Design for low power
  - Motivation for VLSI innovation
  - On all levels! System level … process tech.
  - Lowest possible
    - » Supply voltage V_{DD}
    - » Effective capacitance C_{eff}
    - » Clock frequency f_{CLK}