EEC 216 Lecture #1: CMOS Power Dissipation and Trends

Rajeevan Amirtharajah
University of California, Davis

Outline

- **Administrative Details**
- **Why Care About Power?**
- **Trends in CMOS Power Dissipation**
- **Dynamic Power Dissipation**
- **Short Circuit (Overlap) Current**
- **Power-Delay Metric**
- **Energy-Delay Metric**
- **Logic Level Power Estimation**
- **Next Topic: High Level Power Estimation**

Outline

- Administrative Details
- **Why Care About Power?**
- Trends in CMOS Power Dissipation
- Dynamic Power Dissipation
- Short Circuit (Overlap) Current
- Power-Delay Metric
- Energy-Delay Metric
- Logic Level Power Estimation
- Next Topic: High Level Power Estimation

Permissions to Use Conditions & Acknowledgment

- **Permission is granted to copy and distribute this slide set for educational purposes only, provided that the complete bibliographic citation and following credit line is included: "Copyright 2002 J. Rabaey et al."**
  Permission is granted to alter and distribute this material provided that the following credit line is included: "Adapted from (complete bibliographic citation). Copyright 2002 J. Rabaey et al."
  This material may not be copied or distributed for commercial purposes without express written permission of the copyright holders.

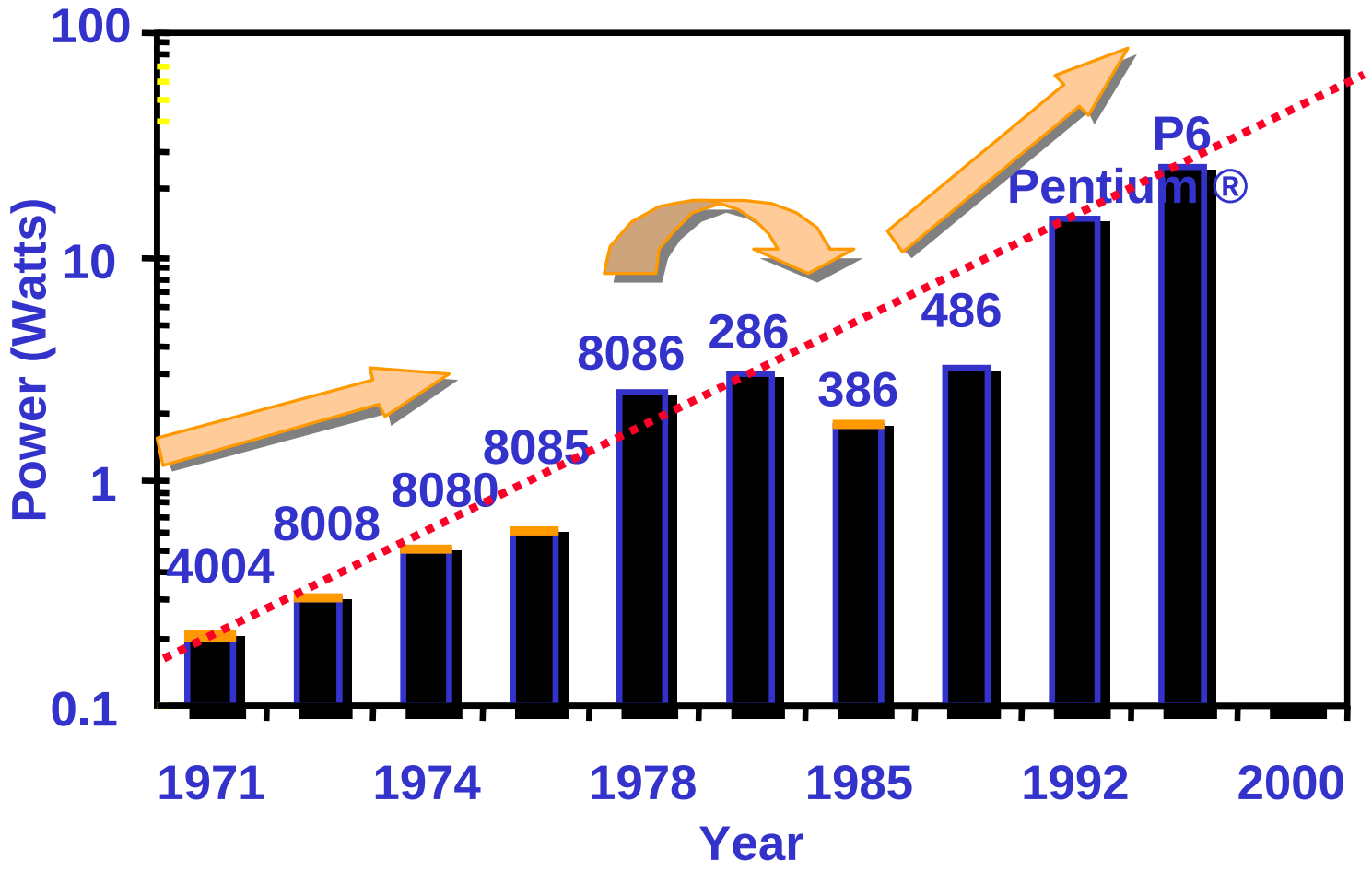
- **Slides 5-10 Adapted from CSE477 VLSI Digital Circuits Lecture Slides by Vijay Narayanan and Mary Jane Irwin, Penn State University**

Why Power Matters

- **Packaging costs**
- **Power supply rail design**
- **Chip and system cooling costs**
- **Noise immunity and system reliability**
- **Battery life (in portable systems)**
- **Environmental concerns**
  - Office equipment accounted for 5% of total US commercial energy usage in 1993
  - *Energy Star* compliant systems for appliances
  - Green data center initiatives by Google, HP, and others

Why worry about power? Power Dissipation

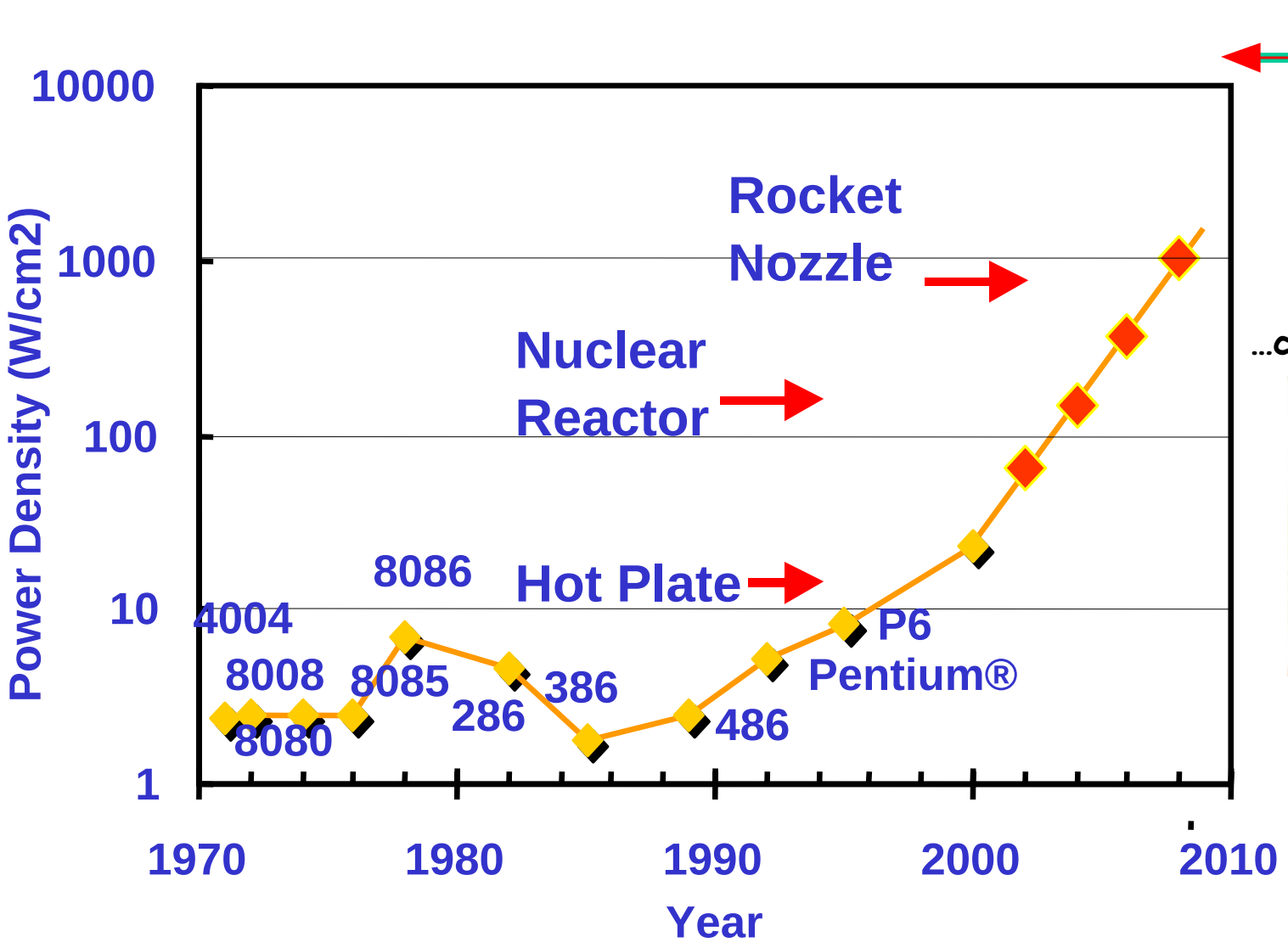
Lead microprocessors power continues to increase



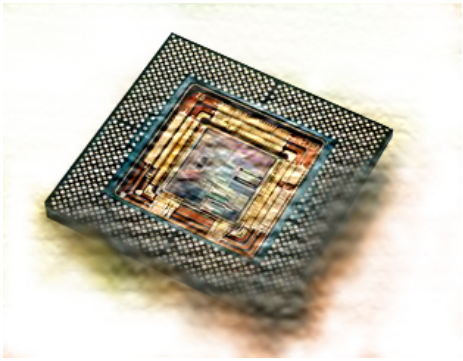
Power delivery and dissipation will be prohibitive

Source: Borkar, De Intel®

Why worry about power? Chip Power Density



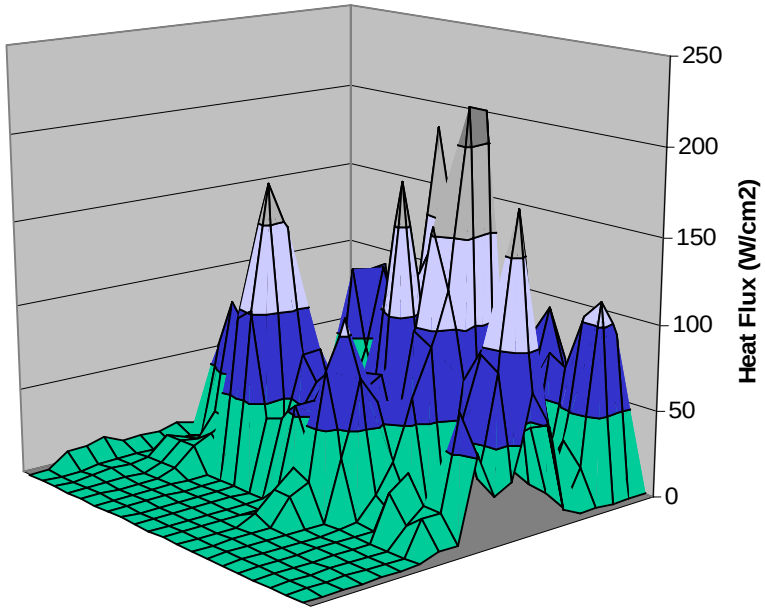
…chips might become hot…



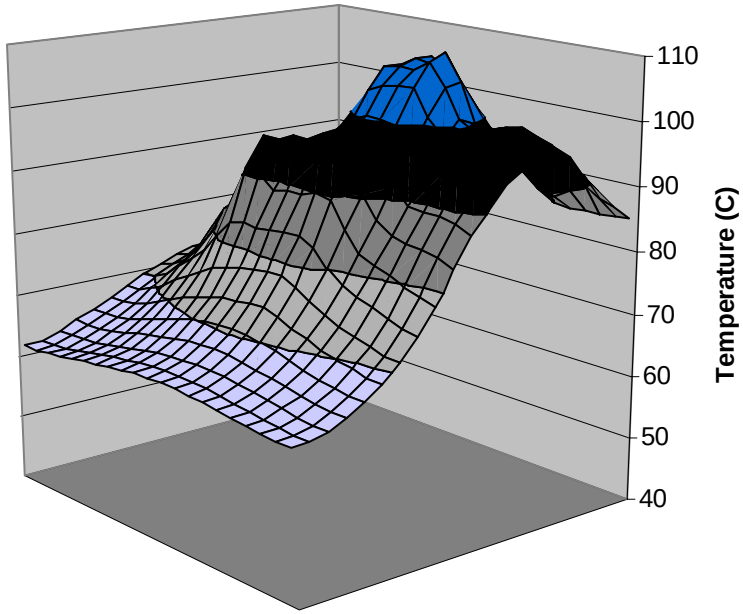
Source: Borkar, De Intel®

Chip Power Density Distribution

Power Map

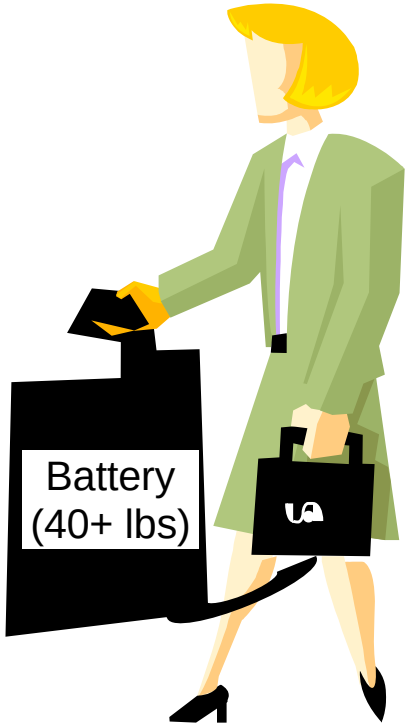


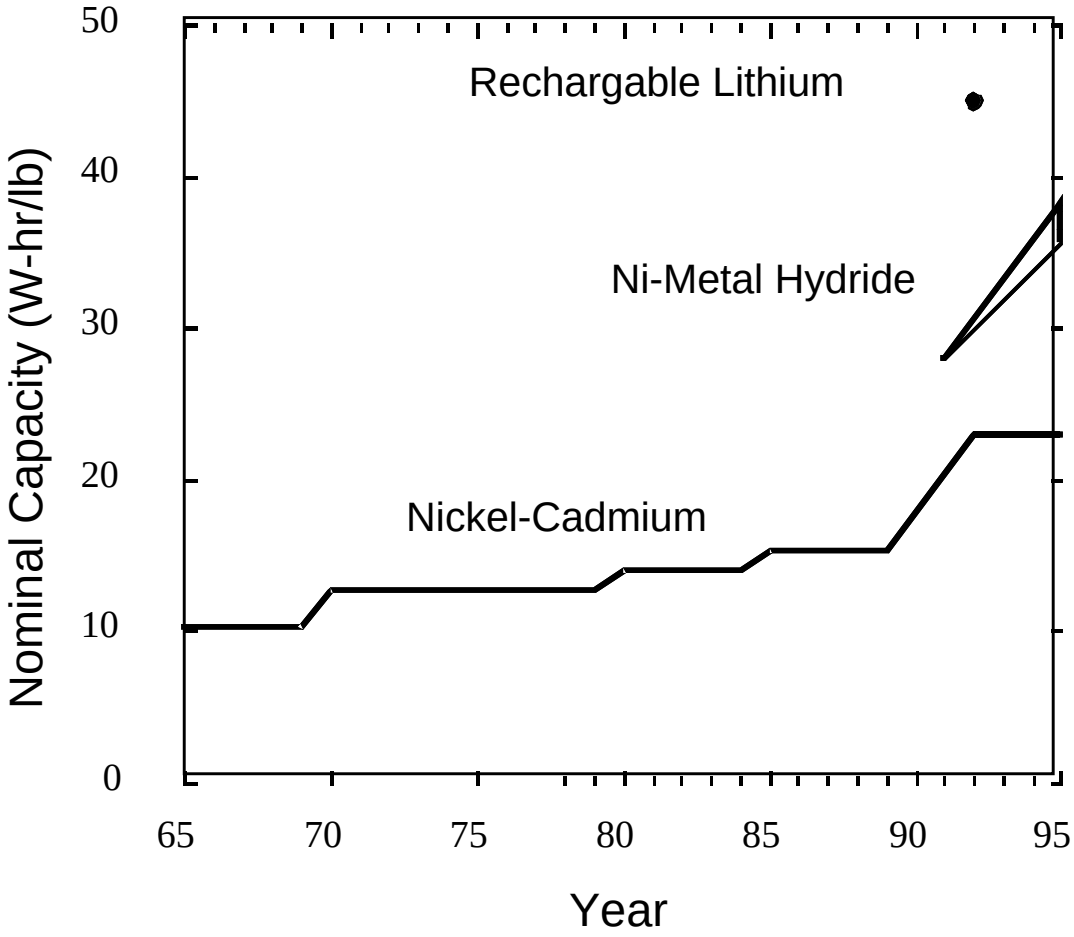
On-Die Temperature



- **Power density is not uniformly distributed across the chip**
- **Silicon not the best thermal conductor (isotopically pure diamond is)**
- **Max junction temperature is determined by hot-spots**
  - Impact on packaging, w.r.t. cooling

Why worry about power ? Battery Size/Weight

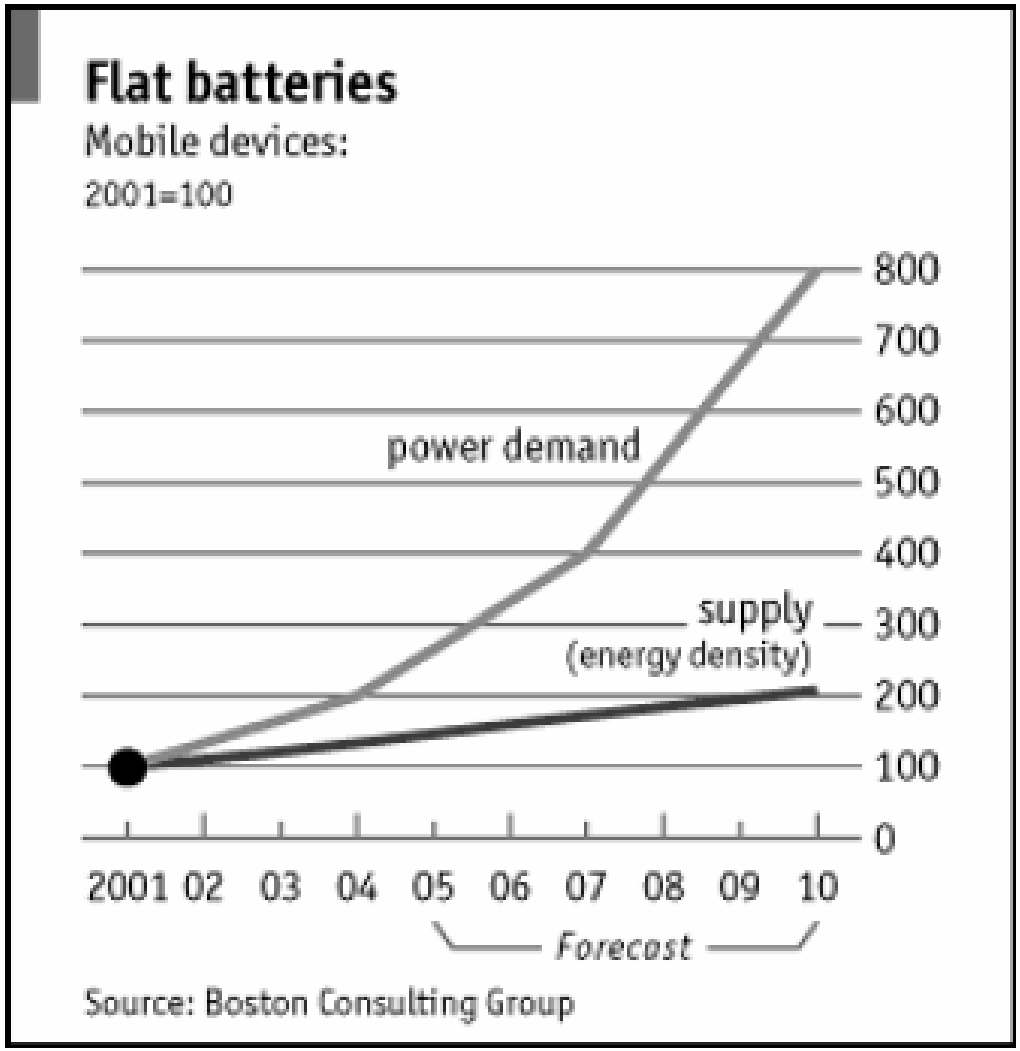




Expected battery lifetime increase over the next 5 years: 30 to 40%

From Rabaey, 1995

Recent Battery Scaling and Future Trends

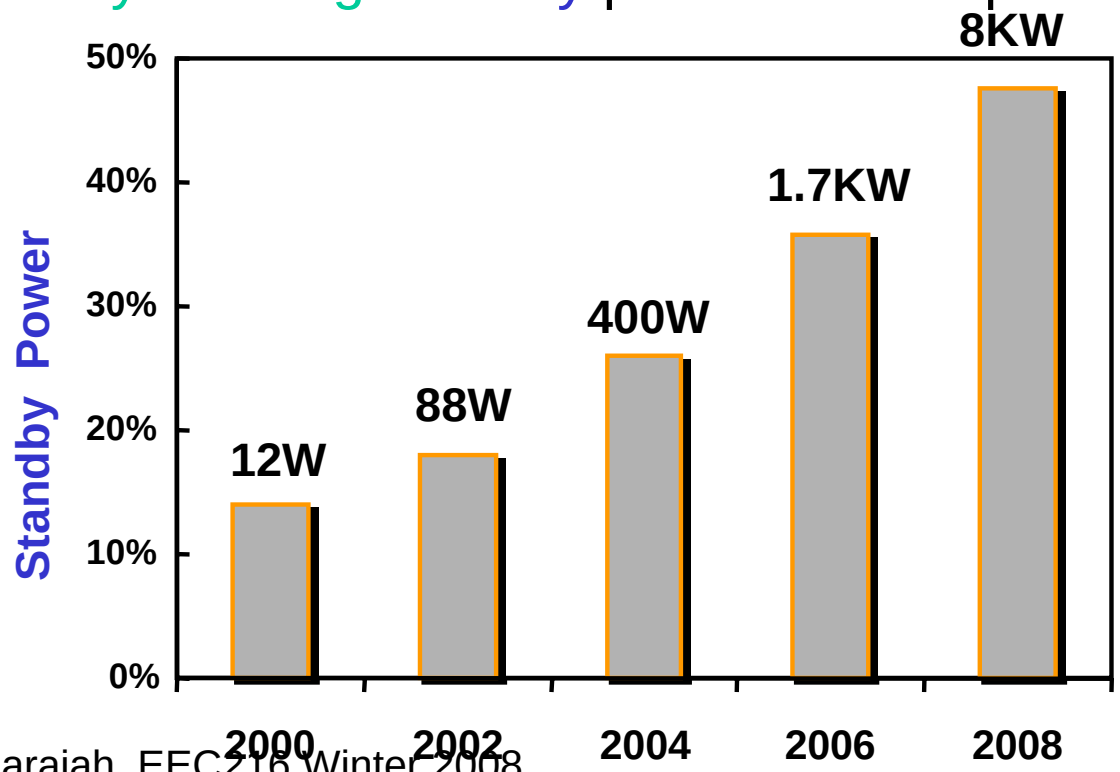


- **Battery energy density increasing 8% per year, demand increasing 24% per year (the Economist, January 6, 2005)**

Why worry about power? Standby Power

Year	2002	2005	2008	2011	2014
Power supply V_{dd} (V)	**1.5**	**1.2**	**0.9**	**0.7**	**0.6**
Threshold V_T (V)	**0.4**	**0.4**	**0.35**	**0.3**	**0.25**

- Drain leakage will increase **if** V_T decreases to maintain noise margins and meet frequency demands, leading to excessive battery draining standby power consumption.



…and phones leaky!

Source: Borkar, De Intel®

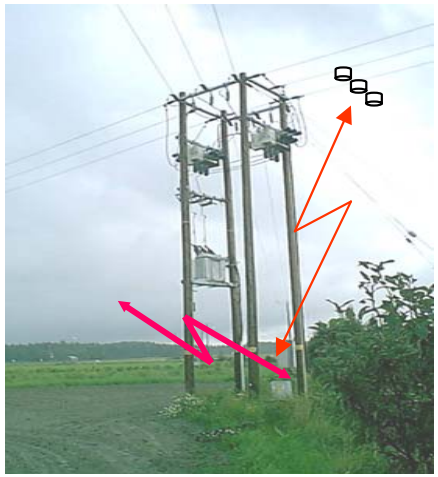
Fundamental Shift: Near Constant V_{DD}

Year	2002	2005	2008	2011	2014
Power supply V_{dd} (V)	**1.5**	**1.2**	**1.0**	**1.0**	**1.0**
Threshold V_T (V)	**0.4**	**0.4**	**0.35**	**0.3**	**0.3**

- Leakage problem has caused voltage scaling with technology to slow down and potentially stop! New devices might restart trend (e.g., FinFET)…
- More emphasis on operating devices below threshold for low power applications

Emerging Microsensor Applications

Industrial Plants and Power Line Monitoring (courtesy ABB)

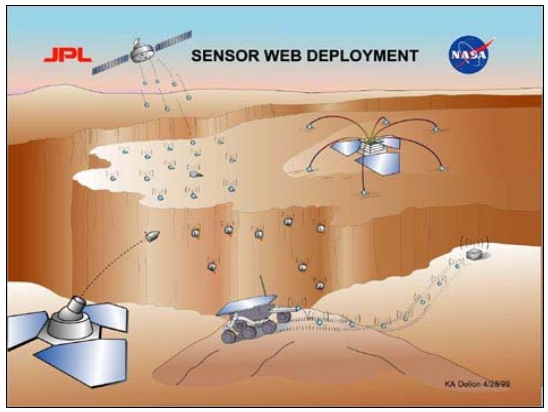


Operating Room of the Future (courtesy John Guttag)

Target Tracking & Detection (Courtesy of ARL)

Location Awareness (Courtesy of Mark Smith, HP)

NASA/JPL *sensorwebs*



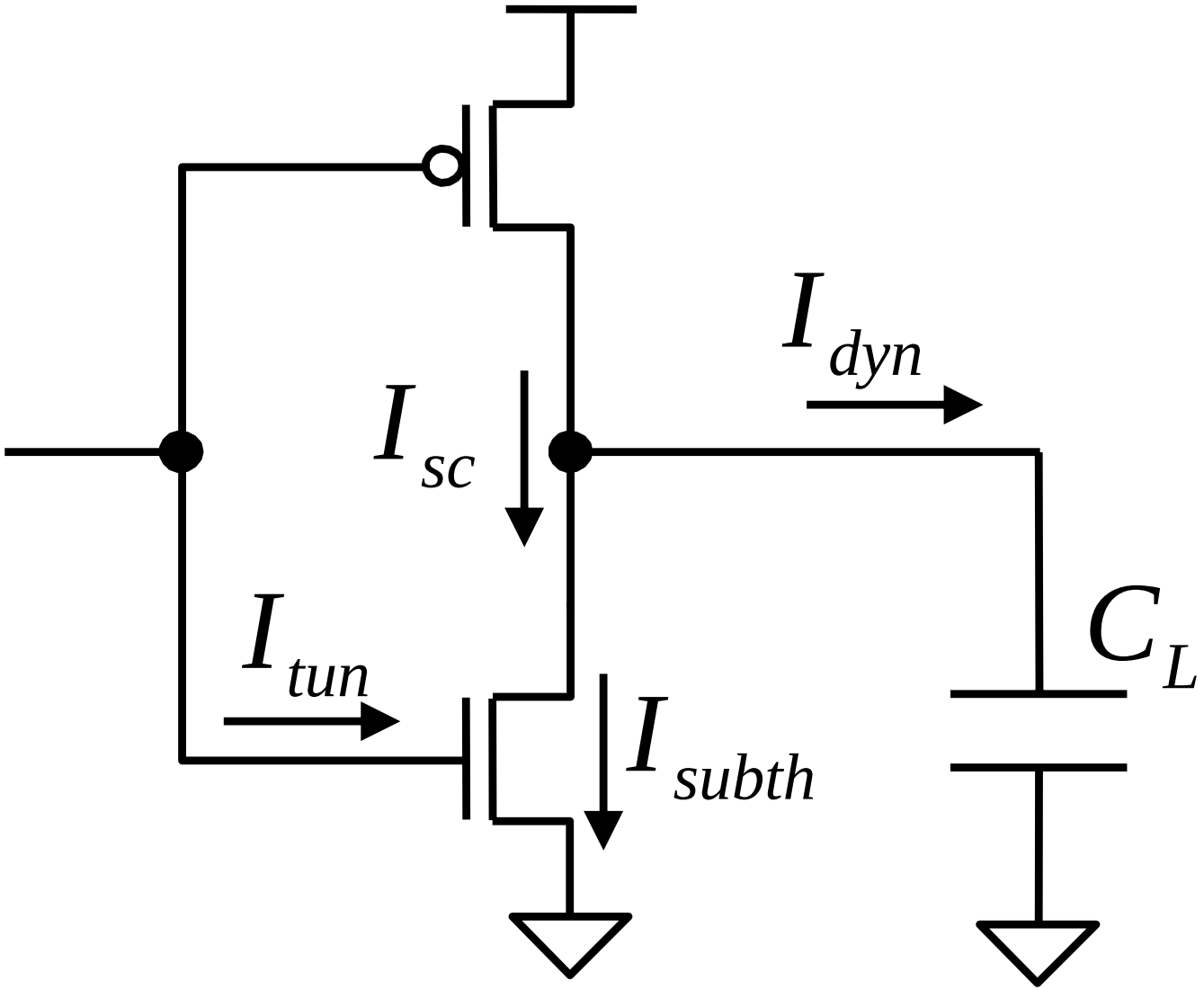
Power as a System Design Consideration

- **Battery operated devices**
  - Reasonable lifetime with limited weight for portability
  - Low to medium performance
  - Examples: PDAs, cell phones, multimedia terminals, laptops, wireless sensors
- **High performance applications**
  - Power delivery and heat removal an issue for system and package cost, reliability
  - On chip hot spots degrade performance
  - Examples: desktops, servers, networking equipment

Outline

- Administrative Details
- Why Care About Power?
- **Trends in CMOS Power Dissipation**
- Dynamic Power Dissipation
- Short Circuit (Overlap) Current
- Power-Delay Metric
- Energy-Delay Metric
- Logic Level Power Estimation
- Next Topic: High Level Power Estimation

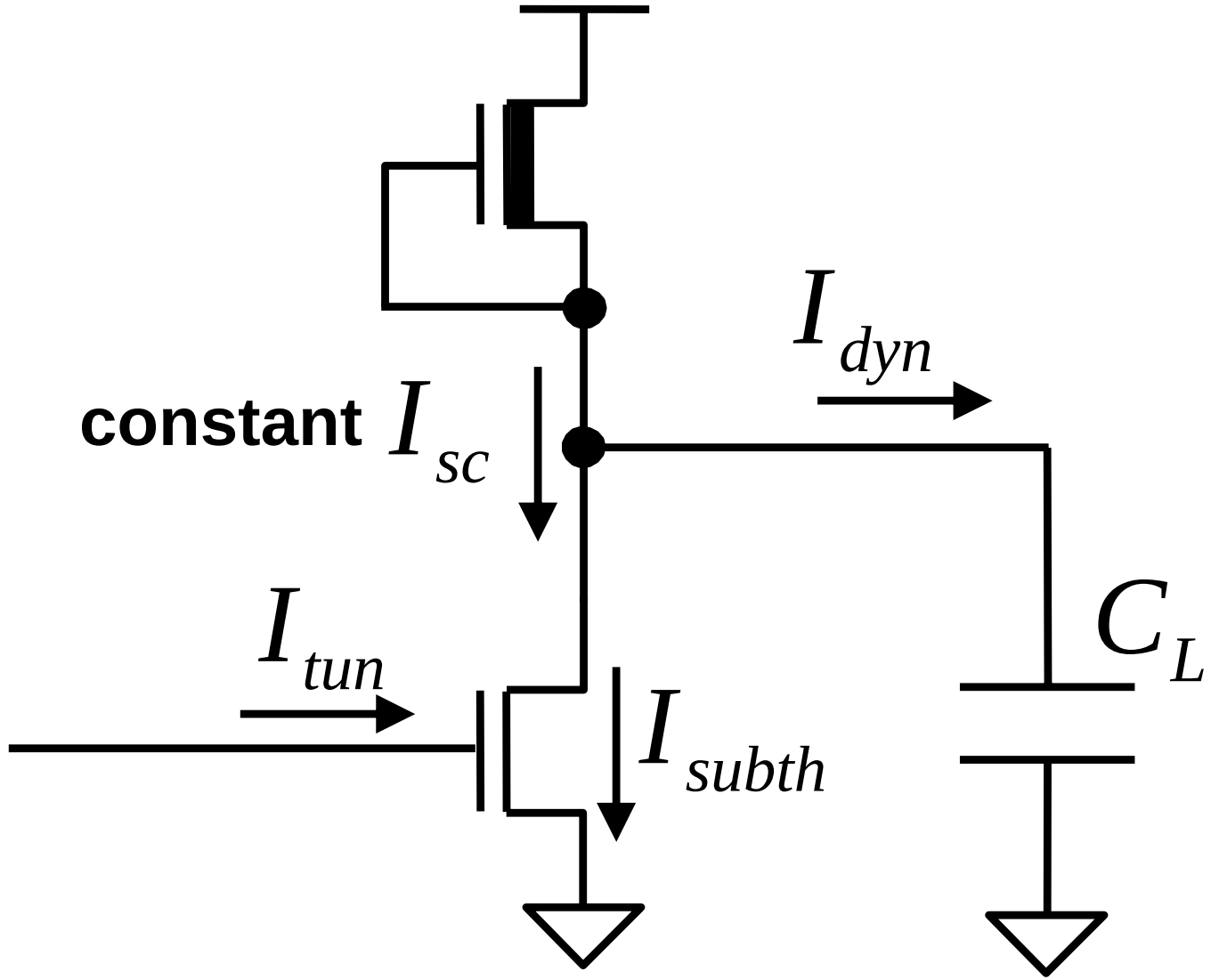
CMOS Inverter Example



Components of CMOS Power Dissipation

- **Dynamic Power**
  - Charging and discharging load capacitances
- **Short Circuit (Overlap) Current**
  - Occurs when PMOS and NMOS devices on simultaneously
- **Static Current**
  - Bias circuitry in analog circuits
- **Leakage Current**
  - Reverse-biased diode leakage
  - Subthreshold leakage
  - Tunneling through gate oxide

Aside: NMOS Inverter Example



- **Depletion NMOS always on, sourcing static current**

Power Dissipation as Technology Scales

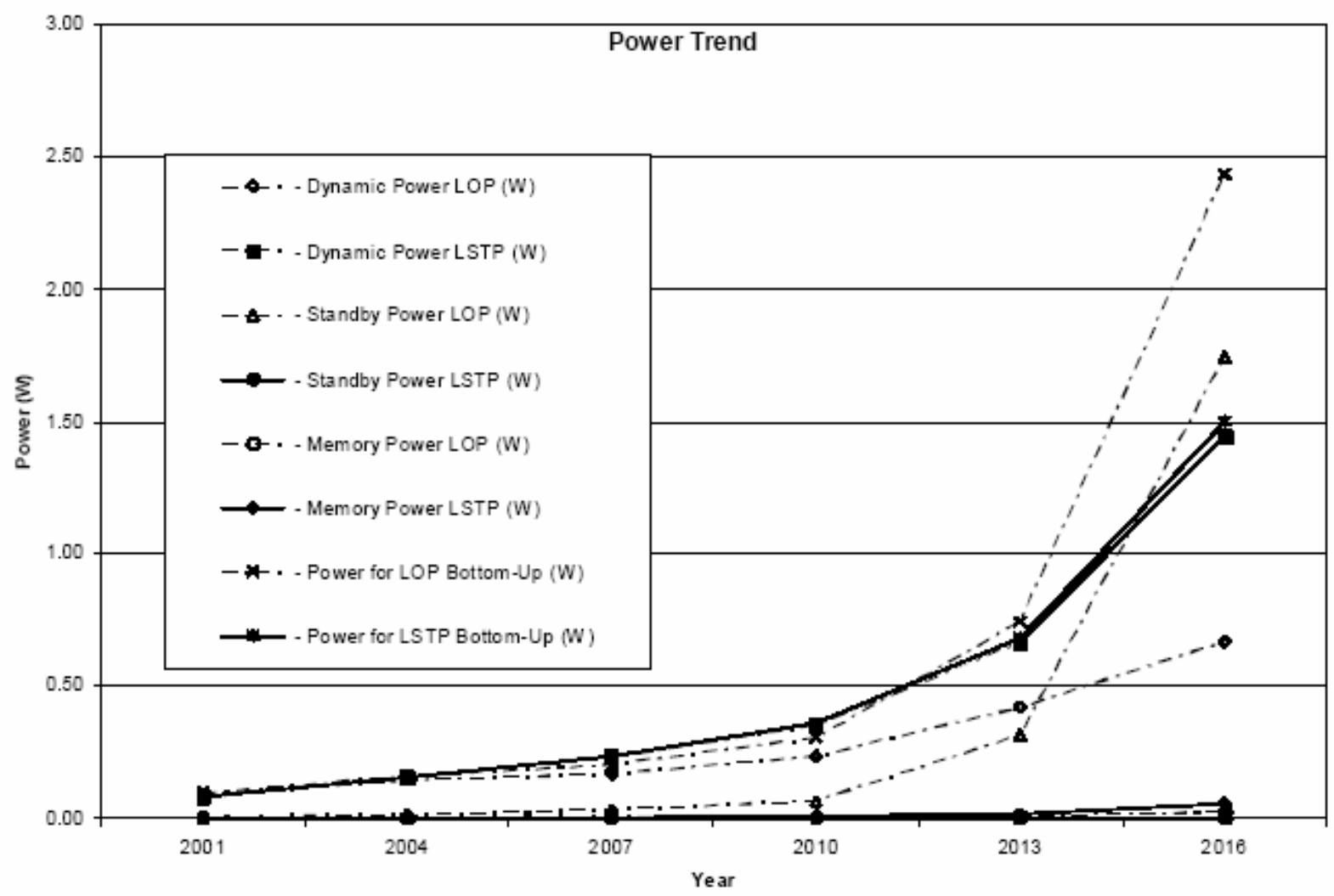


Figure 11 Total Chip Power Trend for PDA Application

- **From 2001 ITRS**

Outline

- Administrative Details
- Why Care About Power?
- Trends in CMOS Power Dissipation
- **Dynamic Power Dissipation**
- Short Circuit (Overlap) Current
- Power-Delay Metric
- Energy-Delay Metric
- Logic Level Power Estimation
- Next Topic: High Level Power Estimation

Extremely Brief MOSFET Review

Saturation: $$I_D = \frac{\mu C_{ox}}{2}\frac{W}{L}\left(V_{GS} - V_T\right)^2\left(1+\lambda V_{DS}\right)$$

Triode: $$I_D = \mu C_{ox}\frac{W}{L}\left(\left(V_{GS} - V_T\right)V_{DS} - \frac{V_{DS}^2}{2}\right)$$

Subthreshold: $$I_D = I_S e^{\frac{V_{GS}}{n^{kT}/_q}}\left(1 - e^{-\frac{V_{DS}}{kT/q}}\right)$$

"Classical" MOSFET model, will discuss deep submicron modifications as necessary

CMOS Delay and Power Dissipation

Delay: $\Delta t = \frac{C\Delta V}{I} = \frac{C_L V_{dd}}{I_D}$

Total Power:

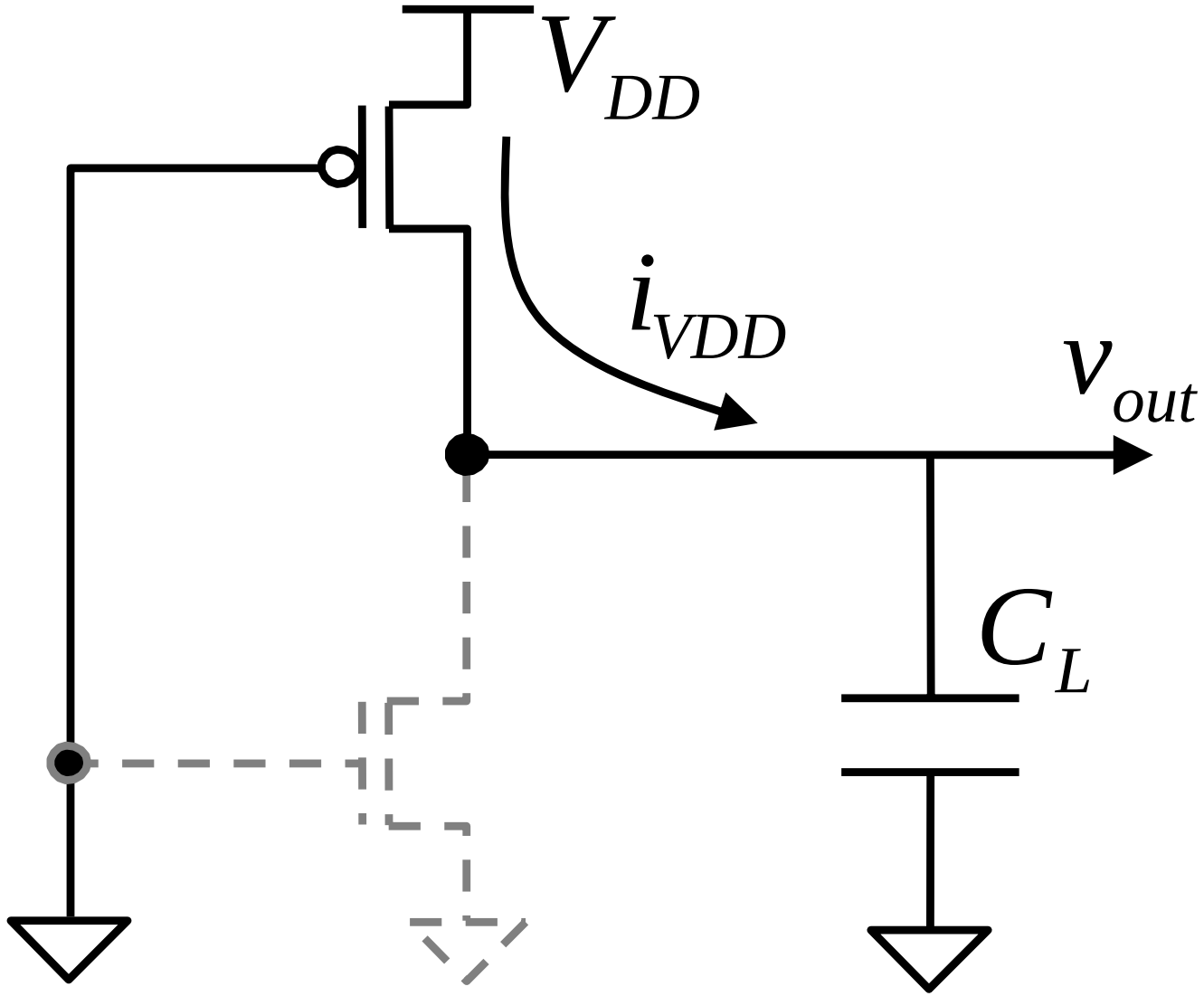
$$P_{TOT} = P_{dyn} + P_{sc} + P_{stat} + P_{leak}$$

$$= \alpha C_L V_{dd}^2 f + V_{dd} I_{peak} \left(\frac{t_r + t_f}{2} \right) f$$

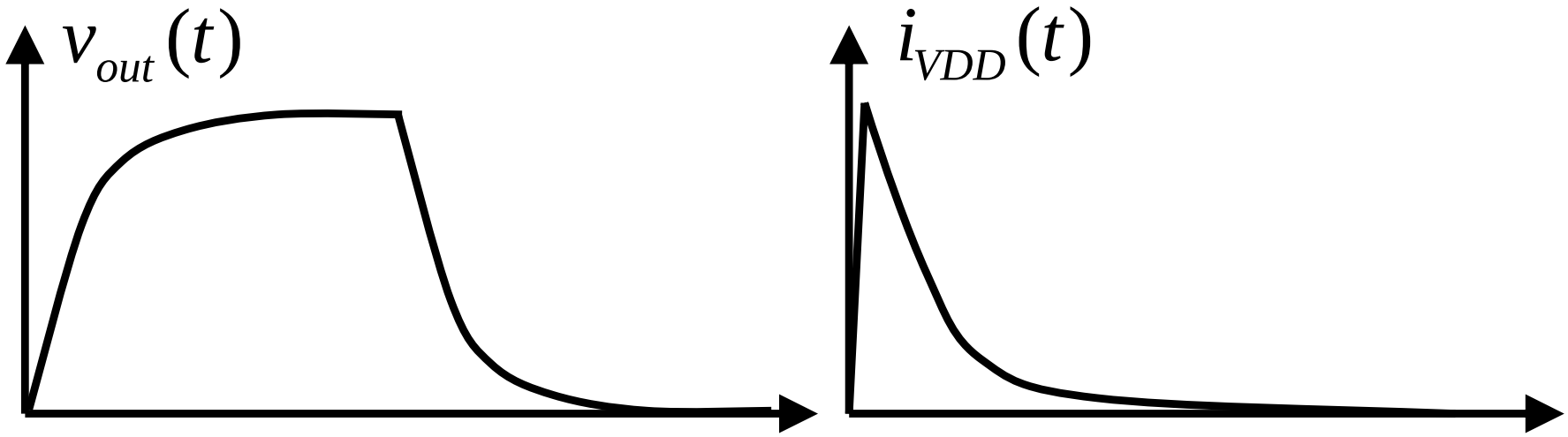
$$+ V_{dd} I_{static} + V_{dd} I_{leak}$$

To reduce power, minimize each term – starting with the biggest! Historically, biggest has been dynamic power…

Low-to-High Transition Equivalent Circuit



Energy Drawn From Power Supply



$$E_{VDD} = \int_0^\infty P_{VDD}(t)dt = \int_0^\infty i_{VDD}(t)V_{DD}dt$$

$$= V_{DD}\int_0^\infty C_L \frac{dv_{out}}{dt}dt = C_L V_{DD}\int_0^{V_{DD}} dv_{out}$$

$$= C_L V_{DD}^2$$

Energy Stored on Load Capacitor

$$E_C = \int_0^\infty i_{VDD}(t) v_{out} dt$$

$$= \int_0^\infty C_L \frac{dv_{out}}{dt} v_{out} dt = C_L \int_0^{V_{DD}} v_{out} dv_{out}$$

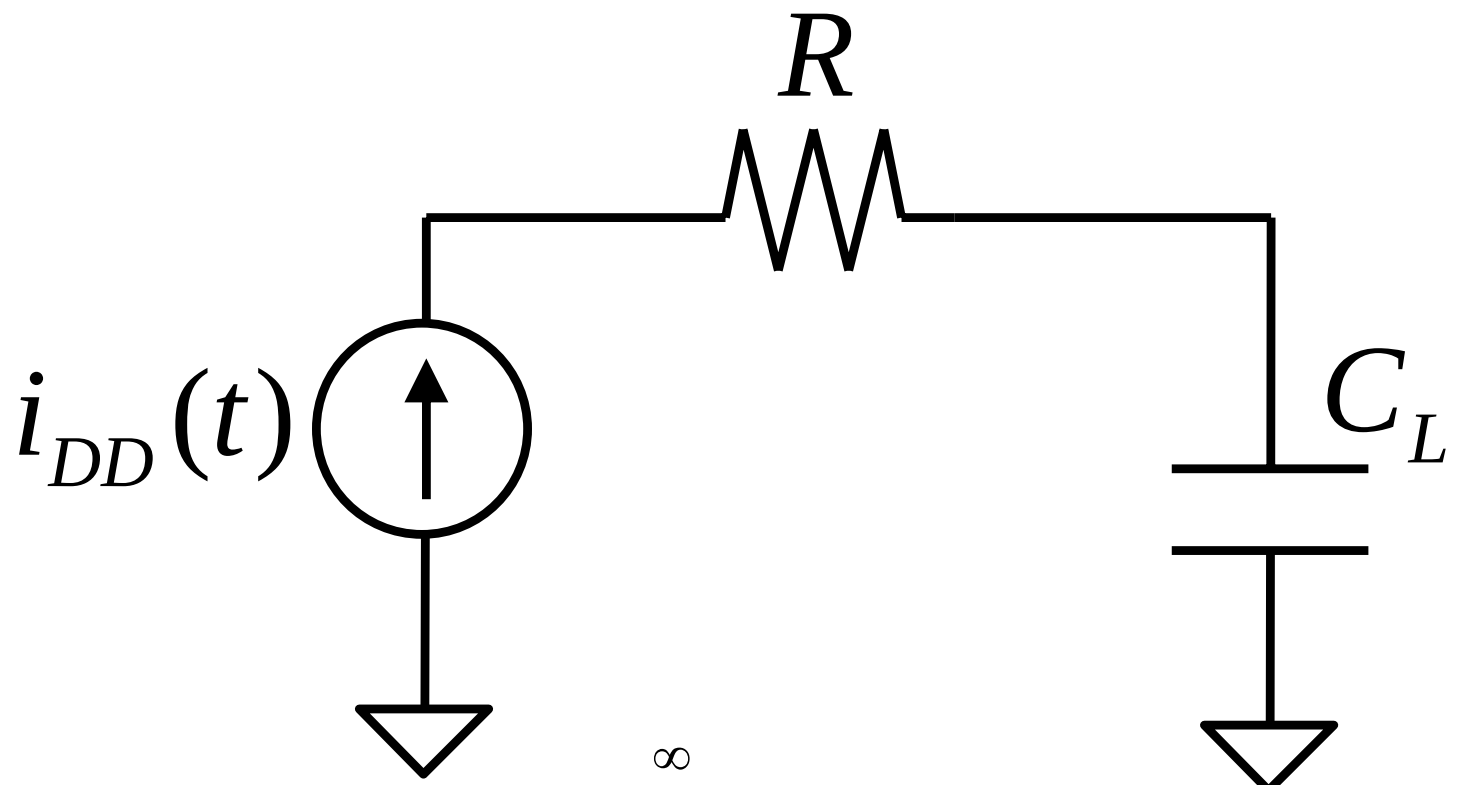
$$= \frac{1}{2} C_L V_{DD}^2$$

- **Compared to E_{VDD}, we see that $\frac{C_L V_{DD}^2}{2}$ dissipated**
- **Same amount dissipated when capacitor discharged**
- **Independent of MOSFET resistance**

Aside: Can We Do Better on Energy?

- **Seems that $C_L V_{DD}^2$ is pretty fundamental**
  - Independent of resistance, circuit delay
- **Static CMOS logic basically configures FETs as switches connected to voltage sources**
  - Transient determined by capacitor dynamics, RC_L
- **But, suppose we use a current source to charge the capacitor instead…**

Preview: Adiabatic Charging



$$E_{IDD} = \int_0^\infty i_{DD}(t) v_{IDD}(t) dt$$

- **By controlling the current, we can control the voltage developed across the resistor, and reduce power consumption by charging slowly**

Deriving Dynamic Power

- **Each charge/discharge cycle dissipates total energy** E_{VDD}
- **To compute power, account for switching the circuit at frequency** f
- **Typically, output does not switch every cycle, so we scale the power by the probability of a transition** α
- **Putting it all together, we derive the dynamic power component of CMOS power dissipation:**

$$P_{dyn} = \alpha C_L V_{DD}^2 f$$

Why Is Power Getting Worse?

- **Same reason performance is getting better: scaling!**
  - Finer lithography and thinner gate oxides imply faster devices…
  - …but we have to scale the supply voltage to maintain reliability…
  - …so in turn we have to decrease the threshold voltages to maintain performance…
  - …while we pack more devices onto bigger die
- **In the bad old days, supply voltage was fixed while devices got smaller (fixed voltage scaling)**
  - Did not reap the full benefits of scaling MOS technology

Competing Factors in Power Trend

- **Recently, we scaled supply voltages more closely with lithography generations**
- **Suppose full scaling (constant electric field) with factor 1/S**
  - Intrinsic gate delay RC falls as 1/S: Good!
  - Intrinsic gate energy CV^2 falls as $1/S^3$: Very Good!
  - Power (energy/delay) falls as $1/S^2$: Not Quite as Good…
  - Gate power density (power/gate area) fixed at 1: No worse than previous generations?
- **Real power drivers: bigger die, more gates, more leakage – secondary effects of scaling**

Minimizing Dynamic Power I

$$P_{dyn} = \alpha C_L V_{DD}^2 f$$

- **Voltage and frequency scaling (lower V_{DD}, f)**
  - Run high speed circuits at lower voltage to meet performance constraint
  - Reduce clock frequency with parallel functional units
  - Relax critical path constraints by pipelining
- **For a fixed circuit, fundamental tradeoff between speed and voltage implies speed/power tradeoff**

Minimizing Dynamic Power II

$$P_{dyn} = \alpha C_L V_{DD}^2 f$$

- **Reduce capacitive load (lower C_L)**
  - Minimum device sizes if performance allows
  - Compact and custom layout
  - Shorten or eliminate long wires, including clock net
- **Choose appropriate logic style: Complementary Pass-Gate Logic (CPL)?**

Minimizing Dynamic Power III

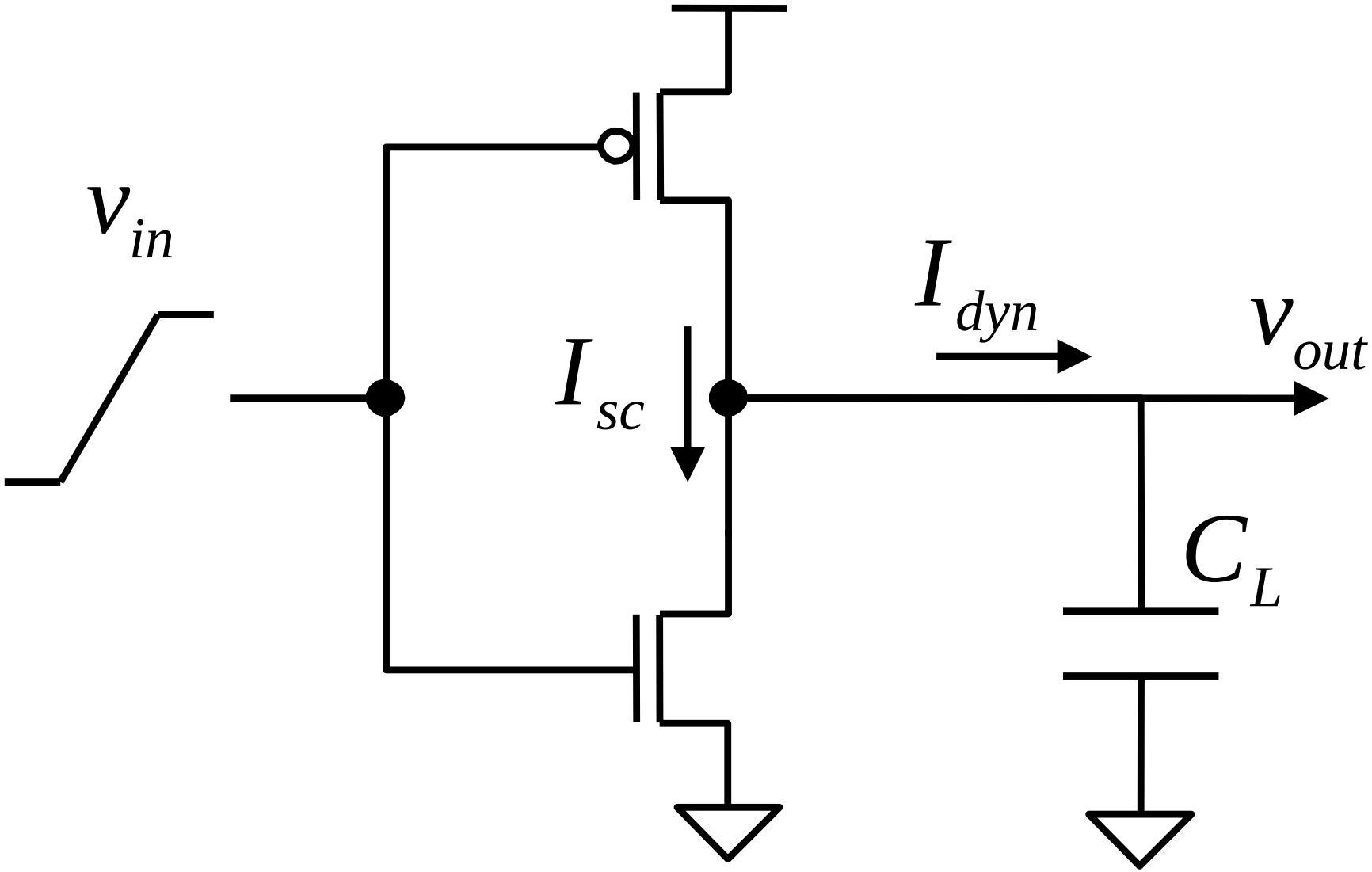
$$P_{dyn} = \alpha C_L V_{DD}^2 f$$

- **Reduce activity factor (lower α)**
  - Clock gating
  - Latch module inputs to eliminate glitches
- **Recode data to change statistics, for example use Gray coding for counters, state machines**
- **Sometimes literature refers to Effective Capacitance or Switched Capacitance:** $C_{EFF} = C_{SW} = \alpha C_L$

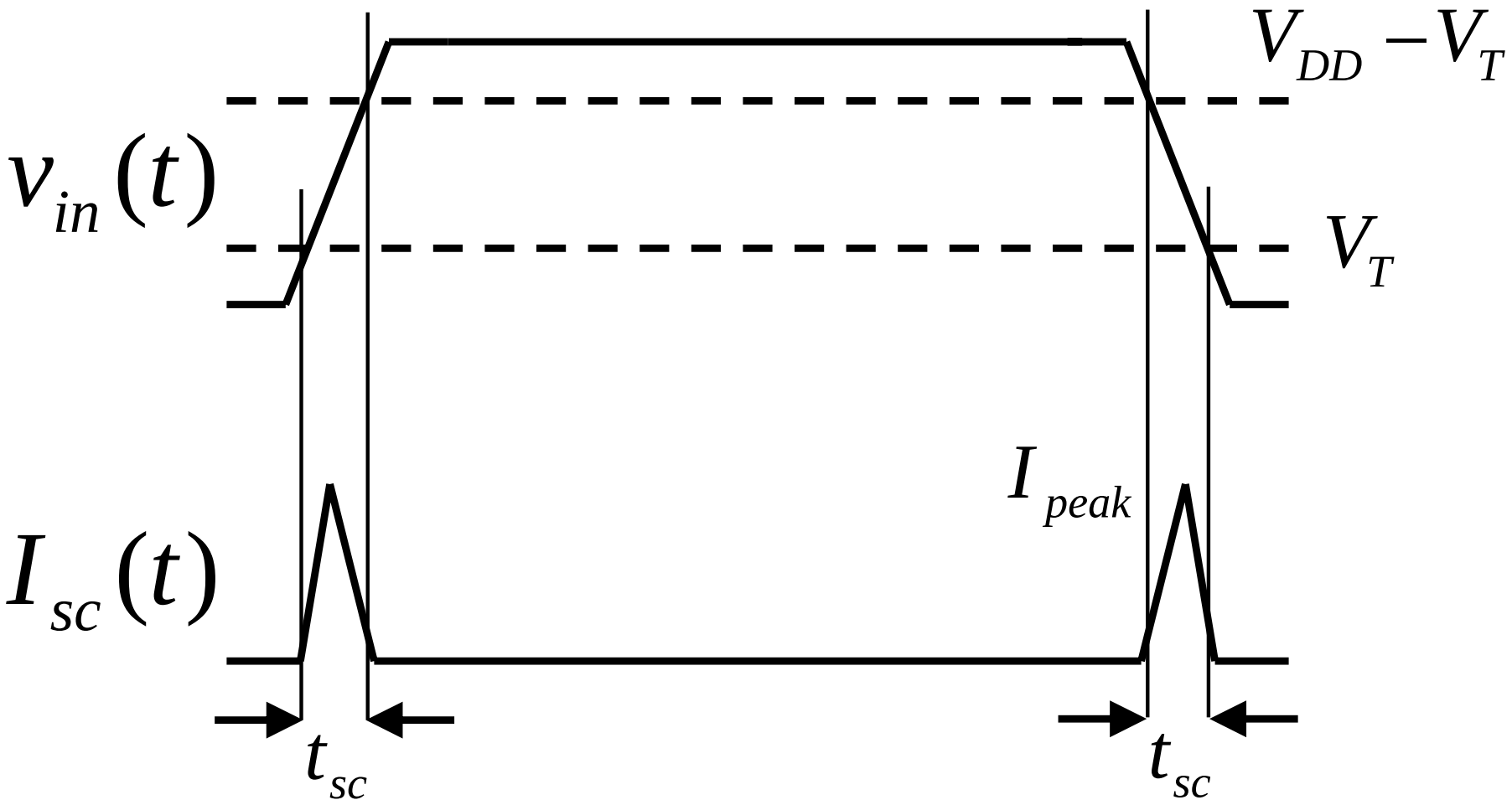
Outline

- Administrative Details
- Why Care About Power?
- Trends in CMOS Power Dissipation
- Dynamic Power Dissipation
- **Short Circuit (Overlap) Current**
- Power-Delay Metric
- Energy-Delay Metric
- Logic Level Power Estimation
- Next Topic: High Level Power Estimation

CMOS Inverter Short Circuit Current



Short Circuit Current Triangle Approx.



Short Circuit Power Derivation

$$E_{sc} = \int_0^{\infty} I_{SC}(t) V_{DD} dt$$

$$= V_{DD} \frac{I_{peak} t_{sc}}{2} + V_{DD} \frac{I_{peak} t_{sc}}{2} = t_{sc} V_{DD} I_{peak}$$

- **Similar to deriving the dynamic power, we must account for the switching frequency and probability of a charge/discharge cycle to estimate short circuit power:**

$$P_{sc} = t_{sc} V_{DD} I_{peak} \alpha f = C_{sc} V_{DD}^2 f$$

- **Note that we can assign a "short circuit" effective capacitance C_{sc} since $\alpha t_{sc} I_{peak}$ has units of charge**

Duration of Both Devices Conducting

- **Assume input voltage is linear**
- **Devices start conducting when input gets above a threshold voltage, stop conducting when input gets within a threshold drop of the rail:**

$$t_{sc} = \frac{V_{DD} - 2V_T}{V_{DD}} t_{0-100\%} \approx \frac{V_{DD} - 2V_T}{V_{DD}} \times \frac{t_{r(f)}}{0.8}$$

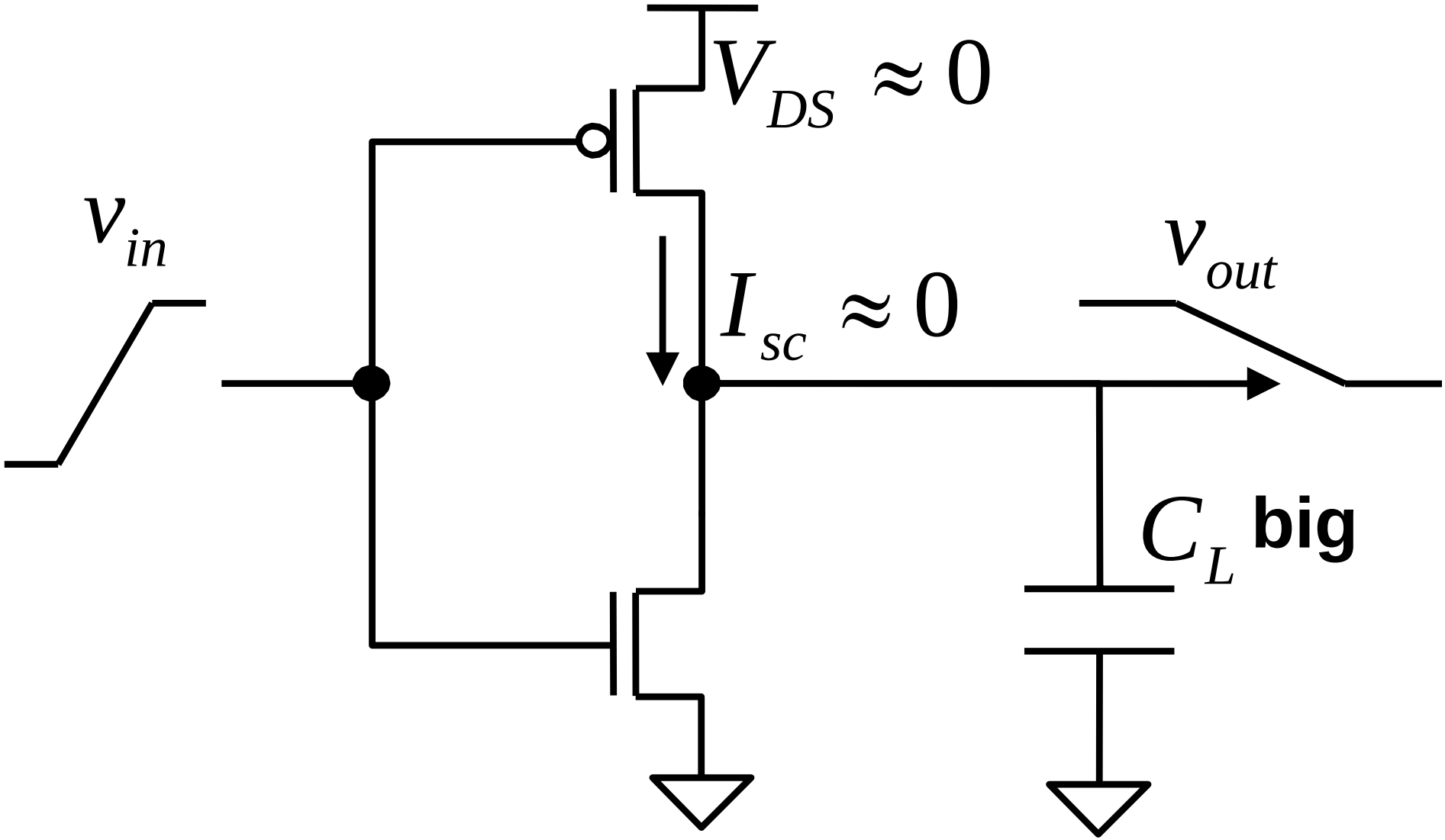
- **If** $V_T = 0.1V_{DD}$ **, then** $t_{sc} \approx t_{r(f)}$ **, i.e. the short circuit current flows during the rise/fall time of the circuit input**

Short Circuit (Overlap) Current Power

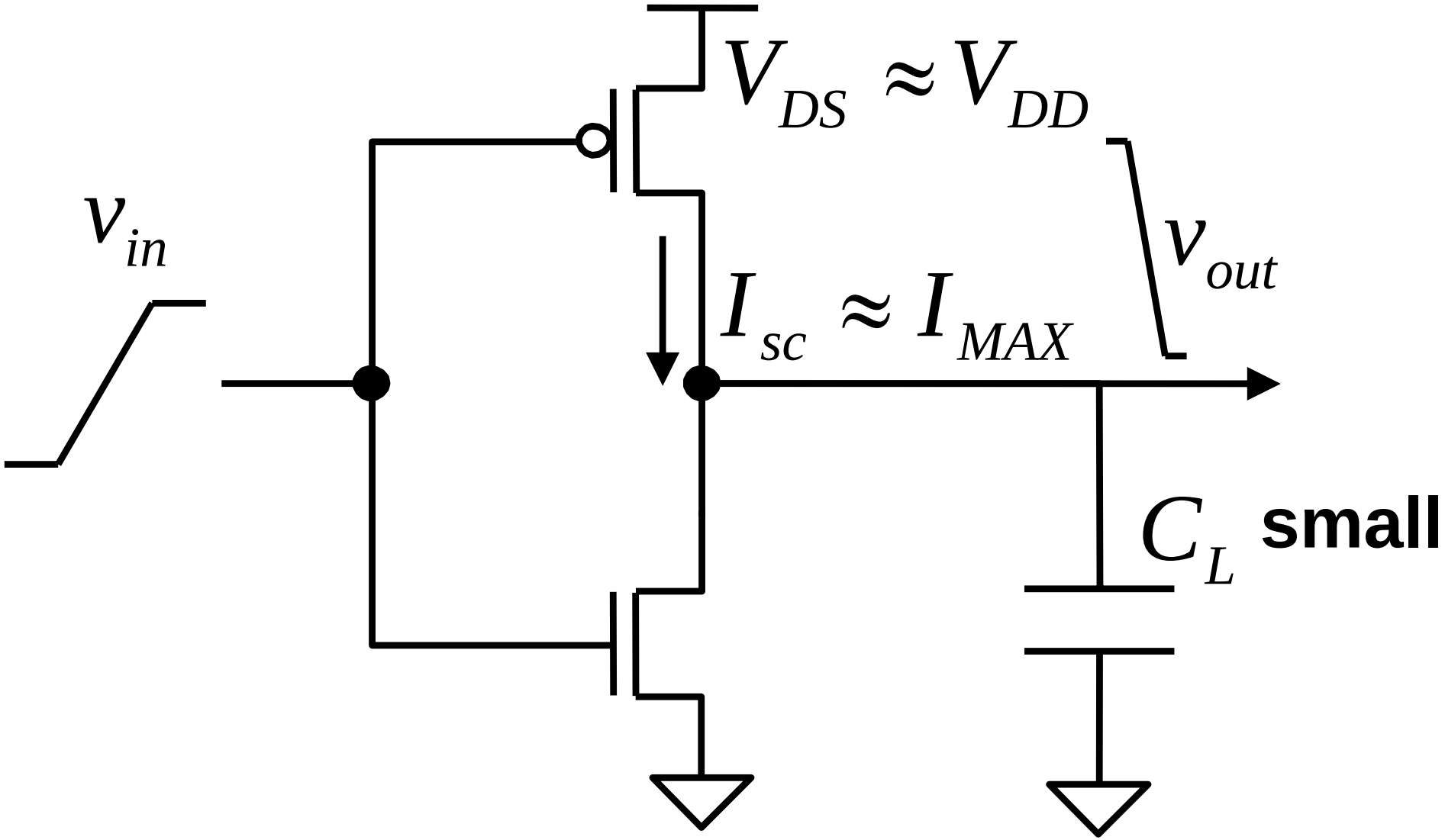
$$P_{sc} = V_{DD} I_{peak} \left(\frac{t_r + t_f}{2} \right) f$$

- **Proportional to switching activity, same as dynamic power**
- **Voltage and frequency scaling (lower V_{DD}, f)**
  - **Run high speed circuits at lower voltage to meet performance**
  - **Reduce clock frequency with parallel functional units**
  - **Relax critical path constraints by pipelining**
- **Decrease input rise and fall times?**

Short Circuit Current With Large Load



Short Circuit Current With Small Load



Minimizing Short Circuit Power

- **Peak current determined by MOSFET saturation current, so directly proportional to device sizes**
- **Peak current also strong function of ratio between input and output slopes as shown in previous 2 slides**
- **For individual gate, minimize short circuit current by making output rise/fall time much bigger than input rise/fall time**
  - Slows down circuit
  - Increases short circuit current in fanout gates
- **Compromise: match input and output rise/fall times**

Some Final Words on Short Circuit Power

- **When input and output rise/fall times are equalized, most power is associated with dynamic power**
  - <10% devoted to short circuit currents
- **Can eliminate short circuit dissipation entirely by very aggressive voltage scaling**
  - Need $V_{DD} < V_{Tn} + |V_{Tp}|$
  - Both devices can't be on simultaneously
- **Short circuit power becoming less important in deep submicron**
  - Threshold voltages not scaling as fast as supply voltages

Preview: Lowering Static Power

- **Reduce static current**
  - Minimize bias circuitry
  - Use switched current sources
- **Reduce subthreshold leakage**
  - Use only high threshold devices
  - Use high threshold devices to interrupt leakage paths (e.g., at bottom of NMOS logic tree)
  - Adjust threshold using body bias
- **Reduce gate tunneling leakage**
  - Use high k gate dielectric and thicker gate oxide

Power vs. Performance

- **Performance optimization and low power can go hand in hand…**
  - Optimize circuits for maximum speed, then decrease voltage until performance requirement met
  - Pipelining, parallelization, other performance oriented architectural features can also help reduce voltage, clock frequency
- **…but can also oppose each other**
  - Architectural and circuit techniques can increase capacitance
  - High performance (low threshold) devices result in large leakage currents

Outline

- Administrative Details
- Why Care About Power?
- Trends in CMOS Power Dissipation
- Dynamic Power Dissipation
- Short Circuit (Overlap) Current
- **Power-Delay Metric**
- Energy-Delay Metric
- Logic Level Power Estimation
- Next Topic: High Level Power Estimation

Review: Dynamic Power

- **Each charge/discharge cycle dissipates total energy** $E_{VDD} = C_L V_{DD}^2$
- **To compute power, account for switching the circuit at frequency** f
- **Typically, output does not switch every cycle, so we scale the power by the probability of a transition** α
- **Putting it all together, we derive the dynamic power component of CMOS power dissipation:**

$$P_{dyn} = \alpha C_L V_{DD}^2 f$$

How Can We Quantify Quality of Design?

- **Want to compare designs quantitatively on a level playing field**
  - Two characteristics we care most about are delay and power consumption
  - Delay and power are related
- **CMOS logic operates by moving charge (storing energy) on and off capacitors**
  - Faster energy transfer…
  - …implies higher power consumption…
  - means faster logic gates!

Power-Delay Product

$$PDP = P_{av} t_{pd}$$

- **Product of average power and propagation delay t_{pd} is generally a constant**
  - For given technology and fixed gate topology
- **Define propagation delay as average between low-to-high transition delay and high-to-low transition delay:**

$$t_{pd} = \frac{t_{pLH} + t_{pHL}}{2}$$

- **PDP = Energy consumed by gate per switching event (Watts x seconds = Joules)**

Power-Delay Product Example

- **Assume that dynamic power dominates and that gates switch at highest possible frequency:**

$$f_{fast} = 1 \Big/ 2t_{pd}$$

$$PDP = C_L V_{DD}^2 f_{fast} t_{pd} = \frac{C_L V_{DD}^2}{2}$$

- **Recovery capacitor energy from before**

How Good Is Power-Delay Product Metric?

- **Measures energy needed for switching gate which is important**
  - Optimum voltage would be lowest possible to meet minimum performance
  - Reasonable for many applications, but not all
- **Consider metric that accounts for both energy and performance**

Outline

- Administrative Details
- Why Care About Power?
- Trends in CMOS Power Dissipation
- Dynamic Power Dissipation
- Short Circuit (Overlap) Current
- Power-Delay Metric
- **Energy-Delay Metric**
- Logic Level Power Estimation
- Next Topic: High Level Power Estimation

Energy-Delay Product

$$EDP = PDP \times t_{pd} = P_{av} t_{pd}^2$$

$$= \frac{C_L V_{DD}^2}{2} t_{pd}$$

- **Weight performance more heavily than PDP**
  - Enables more flexible power-performance tradeoff
- **Higher voltages decrease delay but increase energy**
- **Lower voltages decrease energy but increase delay**
- **Therefore there exists an optimum supply voltage**

Energy-Delay Product Example

- **Find optimal supply voltage using EDP**

  Delay Equation: $\Delta t = \frac{C\Delta V}{I} = \frac{C_L V_{dd}}{I_D}$

  Drain Current with Velocity Saturation (Empirical):

$$I_D = v_{sat} C_{ox} W \left(V_{GS} - V_T - \frac{V_{DSAT}}{2} \right) \qquad V_{DSAT} \approx \frac{L v_{sat}}{\mu}$$

- **Velocity saturation common in today's short channel devices**
  - MOSFETS rely on drift current
  - Carrier velocity depends linearly on electric field until it maxes out

Energy-Delay Product Example (cont.)

- **Plugging in equations for delay and EDP:**

$$t_{pd} \approx \frac{\alpha C_L V_{DD}}{V_{DD} - V_T - \frac{V_{DSAT}}{2}} \qquad EDP = \frac{\alpha C_L^2 V_{DD}^3}{2\left(V_{DD} - V_T - \frac{V_{DSAT}}{2}\right)}$$

- **Find the optimum by setting** $\frac{dEDP}{dV_{DD}} = 0$

- **Optimal supply voltage:** $V_{DDopt} = \frac{3}{2}\left(V_T + \frac{V_{DSAT}}{2}\right)$

- **For typical submicron devices with thresholds ~ 0.5 V, the optimum power supply is ~ 1V**

Outline

- Administrative Details
- Why Care About Power?
- Trends in CMOS Power Dissipation
- Dynamic Power Dissipation
- Short Circuit (Overlap) Current
- Power-Delay Metric
- Energy-Delay Metric
- **Logic Level Power Estimation**
- Next Topic: High Level Power Estimation

CMOS Complex Gate Dynamic Power

- **Dynamic power formula for more complex gates identical to the inverter from before**
- **Main difference is in switching activity** α
- **Switching activity is a strong function of logic operation being implemented**
- **Energy drawn from the power supply when output transitions low to high**
- **For static CMOS with independent inputs, probability of low to high transition is probability output low in cycle N times probability output high in cycle N+1**

Complex Gate Activity Factor

Fundamental Activity Factor Equation:

$$\alpha_{0\to1} = p_0 p_1 = p_0\left(1 - p_0\right)$$

- **Assuming inputs are independent, uniformly distributed, then any N input static gate has transition probability:**

$$\alpha_{0\to1} = \frac{N_0}{2^N}\frac{N_1}{2^N} = \frac{N_0\left(2^N - N_0\right)}{2^{2N}}$$

- **N_0 is the number of 0s in truth table output column**
- **N_1 is the number of 1s in truth table output**

Simple Activity Factor Examples

INV ($N_0 = 1$): $\alpha_{0\to1} = \frac{1}{4}$ **XOR ($N_0 = 2$):** $\alpha_{0\to1} = \frac{1}{4}$

NOR2 ($N_0 = 3$): $\alpha_{0\to1} = \frac{3}{16}$ **NAND2 ($N_0 = 1$):** $\alpha_{0\to1} = \frac{3}{16}$

NOR3 ($N_0 = 7$): $\alpha_{0\to1} = \frac{7}{64}$ **NAND3 ($N_0 = 1$):** $\alpha_{0\to1} = \frac{7}{64}$

- **Note that for NORs and NANDs converges to** $\frac{1}{2^N}$
- **Could expand list to include Full Adders, AOIs, etc.**

Signal Statistics

- **Assuming inputs are uniformly distributed not really valid**
- **Passing through logic gates can significantly modify signal statistics**
- **Assumption of independence also has problems**
  - Even if inputs independent, signals become correlated as they pass through logic (statistics are “colored”)

Next Topic: High Level Power Estimation

- **Examine signal statistics in more detail**
  - Look at impact of circuit topology
- **Impact of glitches on activity factor**
- **High level estimation flows and techniques**