ESE370: Circuit-Level Modeling, Design, and Optimization for Digital Systems

Lec 17: October 16, 2015
Energy Optimization
Ratioed Logic

Previously
- Three components of power
  - \( P_{tot} = P_{static} + P_{dy} + P_{sc} \)
- Restoration and Noise Margins
- CMOS Gates
  - Drive rail-to-rail
  - Only one transistor turned on in steady state
  - Only subthreshold current in steady state

Today
- Energy Tradeoffs
  - Reduce dynamic power
- Ratioed Logic
  - Break all the rules… (lose our nice properties)
  - Not rail-to-rail signals, steady-state-current…
  - Correctness
  - Performance
  - Power
  - Implications

Total Power
- \( P_{tot} = P_{static} + P_{sc} + P_{dy} \)
- \( P_{dy} + P_{sc} = a(\frac{1}{2}C_{load} + C_{sc})V^2f\)
- \( P_{tot} \approx a(\frac{1}{2}C_{load} + C_{sc})V^2f + Vf(W/L)e^{-Vc/(nkT)} \)

Dynamic Power

Switching Power
- \( P_{dy} = (\#_{trans}/\text{clock}) \frac{1}{2}CV^2f \)
- Let \( a = \text{activity factor} \)
  - \( a = \text{average } \#_{\text{tran}}/\text{clock} \)
- \( P_{dy} = a\frac{1}{2}CV^2f \)
- \( P_{sc} = aC_{sc}V^2f \)
Activity Factor

- Let $a =$ activity factor
- $a =$ average $\text{trans}_{\text{tr}}$/clock

$$a = \frac{p(\text{out}_t = 0)p(\text{out}_{t+1} = 1)}{2^n} = \frac{N_0 N_1}{2^n} = \frac{N_0 (2^N - N_0)}{2^{2N}}$$

Reduce Dynamic Power?

- $P_{\text{dyn}} = a^{1/2}CV^2f$
- How do we reduce dynamic power?

Reduce Activity Factor

Tree

- $a = p(\text{out}_t = 0)p(\text{out}_{t+1} = 1)$
- $a = \frac{N_0 N_1}{2^n} = \frac{N_0 (2^N - N_0)}{2^{2N}}$

Tree

- $a = p(\text{out}_t = 0)p(\text{out}_{t+1} = 1)$
- $a = \frac{N_0 N_1}{2^n} = \frac{N_0 (2^N - N_0)}{2^{2N}}$

Chain

- $a = p(\text{out}_t = 0)p(\text{out}_{t+1} = 1)$
- $a = \frac{N_0 N_1}{2^n} = \frac{N_0 (2^N - N_0)}{2^{2N}}$

Chain

- $a = p(\text{out}_t = 0)p(\text{out}_{t+1} = 1)$
- $a = \frac{N_0 N_1}{2^n} = \frac{N_0 (2^N - N_0)}{2^{2N}}$

Slow Down

- $P_{\text{dyn}} = a^{1/2}CV^2f$
- What happens to power contributions as we reduce clock frequency?
Slow Down

- \( P_{dyn} = a^{1/2}CV^2 f \)

- What happens to power contributions as we reduce clock frequency?
  - Example: CMOS circuit consumes equal dynamic and leakage power, \( P \). No short circuit power. The energy consumed in \( T \) seconds is \( 2PT \).

- What suggest about \( V_{th} \)?

Example: CMOS circuit consumes equal dynamic and leakage power, \( P \). No short circuit power. The energy consumed in \( T \) seconds is \( 2PT \).

New power: \( (P/2 + P)2T = 3PT \)
- Increased power!

What does this suggest about \( V_{th} \)?

Reduce \( V_{dd} \)

- What happens as reduce \( V \)?
  - Energy?
  - Static
  - Switching
  - Delay?

Reminder:

- \( V_{dd}=1V, V_{thn}=|V_{thp}|=300mV \)

<table>
<thead>
<tr>
<th>( V_{in} )</th>
<th>( I_{static} )</th>
<th>( I_{dynamic} )</th>
<th>( I_{sc} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0V</td>
<td>180pA</td>
<td>125nA</td>
<td></td>
</tr>
<tr>
<td>140mV</td>
<td>6nA</td>
<td>100nA</td>
<td></td>
</tr>
<tr>
<td>400mV</td>
<td>36nA</td>
<td>18nA</td>
<td></td>
</tr>
<tr>
<td>500mV</td>
<td>36nA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>600mV</td>
<td>36nA</td>
<td>18nA</td>
<td></td>
</tr>
<tr>
<td>860mV</td>
<td>6nA</td>
<td>100nA</td>
<td></td>
</tr>
<tr>
<td>1V</td>
<td>180pA</td>
<td>125nA</td>
<td></td>
</tr>
</tbody>
</table>

Reduce \( V_{dd} \):

- \( V_{dd}=520mV, V_{thn}=|V_{thp}|=300mV \)

<table>
<thead>
<tr>
<th>( V_{in} )</th>
<th>( I_{static} )</th>
<th>( I_{dynamic} )</th>
<th>( I_{sc} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0V</td>
<td>180pA</td>
<td>39.6nA</td>
<td></td>
</tr>
<tr>
<td>140mV</td>
<td>6nA</td>
<td>14.4nA</td>
<td></td>
</tr>
<tr>
<td>260mV</td>
<td>111nA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>360mV</td>
<td>6nA</td>
<td>10.8nA</td>
<td></td>
</tr>
<tr>
<td>500mV</td>
<td>180pA</td>
<td>36nA</td>
<td></td>
</tr>
</tbody>
</table>
Reduce V (no velocity saturation)

- $\tau_{gd} = \frac{Q}{I} = \frac{(CV)}{I}$
- $I_d = \frac{\mu C_{ox}}{2}(W/L)(V_g - V_{TH})^2$
- $\tau_{gd}$ impact?
- $\tau_{gd} \propto \frac{1}{V}$

Ignoring leakage:

$E \propto V^2$

$E \tau^2 = Const$

Reduce V (velocity saturation)

- $\tau_{gd} = \frac{Q}{I} = \frac{(CV)}{I}$
- $I_d = \frac{\nu_{sat} C_{ox}}{2}(W/L)(V_{gs} - V_{TH} - V_{DSAT}/2)$

Reduce $V_{dd}$ (velocity saturation)

- $V_{thn} = |V_{thp}| = 300mV, V_{in} = V_{dd}$, estimate $E_T$

<table>
<thead>
<tr>
<th>$V_{dd}$</th>
<th>$I_{th}$</th>
<th>$t_{th}$</th>
<th>$E_{switch}$</th>
<th>$E_T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1V</td>
<td>126uA</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>700mV</td>
<td>72uA</td>
<td>1.225</td>
<td>0.49</td>
<td>0.6</td>
</tr>
<tr>
<td>500mV</td>
<td>36uA</td>
<td>1.75</td>
<td>0.25</td>
<td>.437</td>
</tr>
<tr>
<td>350mV</td>
<td>9uA</td>
<td>4.9</td>
<td>0.12</td>
<td>.588</td>
</tr>
<tr>
<td>250mV</td>
<td>111nA</td>
<td>295</td>
<td>0.07</td>
<td>20.6</td>
</tr>
</tbody>
</table>

Reduce Short-Circuit Power?

- $P_{sc} = \frac{1}{2}C_{sc}V^2 f$

$E = V_{dd} \times I_{th} \times t_{th} \times \frac{1}{2}$

Increase $V_{th}$?

- What is impact of increasing threshold on
  - Delay?
  - Leakage?
Increase $V_{th}$

- $\tau_{gd} = \frac{Q}{I} = \frac{(CV)}{I}$
- $I_{ds} = (\nu_{sat} C_{ox})(W)(V_{gs}-V_{T_H}-V_{DSAT}/2)$

Increase $V_{th}$?

- What is impact of increasing threshold on
  - Delay?
  - Leakage?

<table>
<thead>
<tr>
<th>$V_{dd}$</th>
<th>$V_{in}$</th>
<th>$I_{ds}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>300mV</td>
<td>326uA</td>
<td>1</td>
</tr>
<tr>
<td>460mV</td>
<td>97uA</td>
<td>1.3</td>
</tr>
<tr>
<td>600mV</td>
<td>72uA</td>
<td>1.75</td>
</tr>
</tbody>
</table>

Idea

- Tradeoff
  - Speed
  - Switching energy
  - Leakage energy
- Energy-Delay tradeoff: $E\tau^2$, $Et$

Ratioed Logic

Previously

- Restoration and Noise Margins
  - Allows for gate abstraction
- CMOS Gates
  - Drive outputs rail-to-rail
  - Only one transistor turned on in steady state
  - Only subthreshold leakage current in steady state

Today

- Ratioed Gates
  - Break all the rules… (nice properties)
    - No rail-to-rail outputs, steady-state-current is not subthreshold…
  - Logic correctness
  - Performance
  - Power
  - Implications
Idea

- Building both pull-up and pull-down can be expensive – many gates
- Seems wasteful to build logic function twice
  - Once in pullup, once in pulldown
  - Large gate capacitance

Maybe only need to build one

- Build NFET pulldown
  - Exploit high N mobility
  - traditional

Ratioed Inverter

- Does this work?
  - What is $V_{out}$ for $V_{in} = Gnd$?
  - What is $V_{out}$ for $V_{in} = V_{dd}$?

Ratioed Inverter in 22nm

Ratioed Inverter

- How do we need to size N to make it work?

Ratioed Inverter in 22nm
How do we need to size P to make it work?

P vs. N

Conclude: still prefer N to P for ratioed logic

Noise Margin Tradeoff

What is impact of increasing noise margin?
- On size
- On input capacitance

Worst-Case Output Drive Strength?

- $R_{\text{drive}}$
- $R_{\text{drive \ when \ input \ low}}$
- $R_{\text{drive \ when \ input \ high}}$
**Ratioed Inverter Sizing**

- What causes knee in curve at high end?

![DC Transfer Curve for Ratioed Inverter](image)

**Non-VSAT Equations**

- P looks roughly like current source
- Must Match current with N
  - Determine $V_{ds}$
  - How does $W_n$ affect $V_{ds}$?

$$I_{DS} = \frac{\mu_n C_{ox}}{2} \left( \frac{W}{L} \right) \left[ (V_{GS} - V_T) \right]^2$$

$$I_{DS} = \mu_n C_{ox} \left( \frac{W}{L} \right) \left[ (V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right]$$

**Size for $R_o/2$ drive?**

- How do we size for $R_o/2$ drive?

**Power?**

- $I_{static}$?
- Input low-Output high?
  - $I_{I_{leak}}$
- Input high-Output low?
  - $I_{pmos_{on}}$
  - $V_{dd}/(R_o/2)$ -- for our sample case

**Compare Static CMOS**

- For $R_{drive}=R_o/2$ inverter
  - Total Transistor Width?
  - Input capacitance load?

![Static CMOS Circuit](image)
Power

\[ P_{\text{tot}} \approx a(\frac{1}{2}C_{\text{load}} + C_{\text{sc}})V^2f \]
\[ + p(V_{\text{out}}=\text{low})V^2/R_{\text{pun}} \]
\[ + (1-p(V_{\text{out}}=\text{low}))V_{\text{in}}(W/L)e^{V_{\text{t}}/(nkT/q)} \]

\( p(V_{\text{out}}=\text{low}) \) – probability the output is low

How size for \( R_{\text{pun}}/2 \) drive?

How size for \( R_{\text{pun}}/2 \) drive?

How size for \( R_{\text{pun}}/2 \) drive?

How size \( K \)-input nor?

When better than CMOS nor-\( k \)?

Better = smaller, lower input capacitance

Ideas

- There are other logic disciplines
- We have the tools to analyze
- Ratioed Logic
  - Tradeoff noise margin for
    - Reduced area
    - Capacitive load
  - Dissipates static power in one mode
Admin

- Project – Milestone due Wednesday
  - Should have read it already
  - Build and simulate baseline design over weekend
    - Ch 11 in textbook
  - Start list of optimizations to try
    - If don’t have list come to office hours ready to talk about it (Khanna: Monday, Giesen: Tuesday)

- Extra office hours next week on Thursday and Friday
  - Giesen: Thursday 6-8pm
  - Khanna: Friday 4-6pm