Announcements

- Homework 3 due today.
- Homework 4 will be posted later today.
- Special office hours from 1:30-3pm at BWRC (in lieu of Tuesday)
Today’s lecture

- Power consumption
- Inverter chain sizing

Power Dissipation
Where Does Power Go in CMOS?

- **Dynamic Power Consumption**
  - Charging and Discharging Capacitors

- **Short Circuit Currents**
  - Short Circuit Path between Supply Rails during Switching

- **Leakage**
  - Leaking diodes and transistors

Dynamic Power Dissipation

\[
\text{Energy/transition} = C_L \cdot V_{dd}^2
\]

\[
\text{Power} = \text{Energy/transition} \cdot f = C_L \cdot V_{dd}^2 \cdot f
\]

- Not a function of transistor sizes!
- Need to reduce \( C_L \), \( V_{dd} \), and \( f \) to reduce power.
**Node Transition Activity and Power**

- Consider switching a CMOS gate for \( N \) clock cycles

\[
E_N = C_L \cdot V_{dd}^2 \cdot n(N)
\]

\( E_N \): the energy consumed for \( N \) clock cycles
\( n(N) \): the number of 0->1 transition in \( N \) clock cycles

\[
P_{\text{avg}} = \lim_{N \to \infty} \frac{E_N}{N} \cdot f_{\text{clk}} = \left( \lim_{N \to \infty} \frac{n(N)}{N} \right) \cdot C_L \cdot V_{dd}^2 \cdot f_{\text{clk}}
\]

\[
\alpha_{0 \to 1} = \lim_{N \to \infty} \frac{n(N)}{N}
\]

\[
P_{\text{avg}} = \alpha_{0 \to 1} \cdot C_L \cdot V_{dd}^2 \cdot f_{\text{clk}}
\]

---

**Short Circuit Currents**

![Short Circuit Currents Diagram](image)

<table>
<thead>
<tr>
<th>( V_{in} )</th>
<th>( V_{out} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( V_{in} ) (V)</th>
<th>( I_{SS} ) (mA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.0</td>
<td>0.01</td>
</tr>
<tr>
<td>4.0</td>
<td>0.01</td>
</tr>
<tr>
<td>3.0</td>
<td>0.01</td>
</tr>
<tr>
<td>2.0</td>
<td>0.01</td>
</tr>
<tr>
<td>1.0</td>
<td>0.01</td>
</tr>
<tr>
<td>0.0</td>
<td>0.01</td>
</tr>
</tbody>
</table>

![Current vs Voltage Graph](image)
How to keep Short-Circuit Currents Down?

Short circuit current goes to zero if $t_{\text{fall}} >> t_{\text{rise}}$, but can’t do this for cascade logic, so ...

Minimizing Short-Circuit Power

- Keep the input and output rise/fall times the same ($< 10\%$ of Total Consumption) 
  (IEEE Journal of Solid-State Circuits, August 1984)
- If $V_{dd} < V_{tn} + |V_{tp}|$ then short-circuit power can be eliminated

EE141
Sub-threshold current one of most compelling issues in low-energy circuit design!

Reverse-Biased Diode Leakage

\[ I_{DL} = J_S \times A \]

\[ J_S = 10-100 \text{ pA}/\mu\text{m}^2 \text{ at } 25 \text{ deg C for 0.25\mu m CMOS} \]

\[ J_S \text{ doubles for every 9 deg C}! \]
Subthreshold Leakage Component

- Leakage control is critical for low-voltage operation

\[ P_{\text{stat}} = P_{(I_{\text{in}}=1)} \cdot V_{\text{dd}} \cdot I_{\text{stat}} \]

Wasted energy ...
Should be avoided in most cases,
but could help reducing energy in others (e.g. sense amps)
Principles for Power Reduction

- Prime choice: Reduce voltage!
  - Recent years have seen an acceleration in supply voltage reduction
  - Design at very low voltages still open question (0.6 … 0.9 V by 2010!)

- Reduce switching activity
- Reduce physical capacitance

Sizing of an Inverter Chain
Inverter Chain

If $C_L$ is given:
- How many stages are needed to minimize the delay?
- How to size the inverters?

May need some additional constraints.

Inverter Delay

- Minimum length devices, $L=0.25\mu m$
- Assume that for $W_P = 2W_N = 2W$
  - same pull-up and pull-down currents
  - approx. equal resistances $R_N = R_P$
  - approx. equal rise $t_{PLH}$ and fall $t_{P HL}$ delays
- Analyze as an RC network

$$R_P = R_{\text{unit}} \left( \frac{W_P}{W_{\text{unit}}} \right)^{-1} \approx R_{\text{unit}} \left( \frac{W_N}{W_{\text{unit}}} \right)^{-1} = R_N = R_W$$

Delay ($D$): $t_{PLH} = (\ln 2) R_N C_L$, $t_{P HL} = (\ln 2) R_P C_L$

Load for the next stage: $C_{gin} = \frac{3}{W_{\text{unit}}} C_{\text{unit}}$
Assumptions: no load -> zero delay

$W_{unit} = 1$

$k$ is a constant, equal to 0.69

Delay = $kR_W(C_{int} + C_L) = kR_W C_{int} + kR_W C_L = kR_W C_{int}(1 + C_L / C_{int})$

= Delay (Internal) + Delay (Load)
Delay Formula

\[ \text{Delay} \sim R_W (C_{int} + C_L) \]

\[ t_p = kR_W C_{int} \left( 1 + C_L / C_{int} \right) = t_{p0} \left( 1 + f / \gamma \right) \]

- \( C_{int} = \gamma C_{gin} \) with \( \gamma = 1 \)
- \( f = C_L / C_{gin} \) - effective fanout
- \( R = R_{unit} / W ; C_{int} = WC_{unit} \)
- \( t_{p0} = 0.69R_{unit}C_{unit} \)

Apply to Inverter Chain

\[ t_p = t_{p_1} + t_{p_2} + \ldots + t_{p_N} \]

\[ t_{p_j} \sim R_{unit}C_{unit} \left( 1 + \frac{C_{gin,j+1}}{\gamma C_{gin,j}} \right) \]

\[ t_p = \sum_{j=1}^{N} t_{p_j} = t_{p0} \sum_{i=1}^{N} \left( 1 + \frac{C_{gin,j+1}}{\gamma C_{gin,j}} \right) \cdot C_{gin,N+1} = C_L \]
Optimal Tapering for Given $N$

Delay equation has $N - 1$ unknowns, $C_{gin,2} - C_{gin,N}$

Minimize the delay, find $N - 1$ partial derivatives

Result: $C_{gin,j+1}/C_{gin,j} = C_{gin,j}/C_{gin,j-1}$

Size of each stage is the geometric mean of two neighbors

$$C_{gin,j} = \sqrt{C_{gin,j-1}C_{gin,j+1}}$$

- each stage has the same effective fanout ($C_{out}/C_{in}$)
- each stage has the same delay

---

Optimum Delay and Number of Stages

When each stage is sized by $f$ and has same eff. fanout $f$:

$$f^N = F = C_L / C_{gin,1}$$

Effective fanout of each stage:

$$f = \sqrt[N]{F}$$

Minimum path delay

$$t_p = Nt_{p0} \left(1 + \sqrt[N]{F} / \gamma \right)$$
Example

$C_L/C_1$ has to be evenly distributed across $N = 3$ stages:

$$f = \sqrt[3]{8} = 2$$

Optimum Number of Stages

For a given load, $C_L$ and given input capacitance $C_{in}$, find optimal sizing $f$.

$$C_L = F \cdot C_{in} = f^N C_{in} \quad \text{with} \quad N = \frac{\ln F}{\ln f}$$

$$t_p = N t_{p0} \left( \frac{F^{1/N}}{f} - 1 \right) = t_{p0} \frac{\ln F}{\gamma} \left( \frac{f}{\ln f} + \frac{\gamma}{\ln f} \right)$$

$$\frac{\partial t_p}{\partial f} = t_{p0} \frac{\ln F \cdot \ln f - 1 - \gamma}{\gamma \ln^2 f} = 0$$

For $\gamma = 0$, $f = e$, $N = \ln F$:

$$f = \exp \left( 1 + \frac{\gamma}{f} \right)$$
Optimum Effective Fanout $f$

Optimum $f$ for given process defined by $\gamma$

$$f = \exp\left(1 + \frac{\gamma}{f}\right)$$

$f_{opt} = 3.6$
for $\gamma=1$

---

Impact of Loading on $tp$

With Self-Loading $\gamma=1$
Normalized delay function of $F$

$$t_p = N t_{p0} \left( 1 + \sqrt[3]{F / \gamma} \right)$$

<table>
<thead>
<tr>
<th>$F$</th>
<th>Unbuffered</th>
<th>Two Stage</th>
<th>Inverter Chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>11</td>
<td>8.3</td>
<td>8.3</td>
</tr>
<tr>
<td>100</td>
<td>101</td>
<td>22</td>
<td>16.5</td>
</tr>
<tr>
<td>1000</td>
<td>1001</td>
<td>65</td>
<td>24.8</td>
</tr>
<tr>
<td>10,000</td>
<td>10,001</td>
<td>202</td>
<td>33.1</td>
</tr>
</tbody>
</table>

Buffer Design

<table>
<thead>
<tr>
<th>N</th>
<th>f</th>
<th>$t_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>64</td>
<td>65</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>2.8</td>
<td>15.3</td>
</tr>
</tbody>
</table>
What about power consumption (and area)?

\[ C_{\text{tot}} = C_i + eC_i + ... + e^N C_i \]
\[ = C_i \left(1 + e + ... + e^N\right) \]
\[ = C_i + C_i e^N + C_i e \left(1 + e + ... + e^{N-2}\right) \]

Overhead = \( (E^n - 1)/(E - 1) \)

e.g. \( C_L = 20\text{pF} \), \( C_i = 50\text{fF} \) \( \rightarrow N = 6 \)
Fixed: 20pF
overhead: 11.66pF !!!

---

Delay versus Area and Power

![Diagram showing normalized area and delay against tapering factor, f]