EE141- Spring 2003
Lecture 19

Arithmetic
Power

Announcements

• Homework 7 due next Thursday
• Hardware Lab 1 next week
• Project 1 results by early next week
Today’s lecture

- Multipliers
- Power in CMOS

Multipliers
The Binary Multiplication

\[
101010 \\
\times 1011 \\
\text{AND operation} \\
101010 \\
000000 \\
+ 101010 \\
\hline 111001110
\]

The Array Multiplier
Carry-Save Multiplier

\[ t_{\text{mult}} = (N-1)t_{\text{carry}} + (N-1)t_{\text{and}} + t_{\text{merge}} \]

Multiplier Floorplan

X and Y signals are broadcasted through the complete array.
Wallace-Tree Multiplier

Multipliers —Summary

- Optimization Goals Different Vs Binary Adder
- Once Again: Identify Critical Path
- Other possible techniques
  - Logarithmic versus Linear (Wallace Tree Mult)
  - Data encoding (Booth)
  - Pipelining

FIRST GLIMPSE AT SYSTEM LEVEL OPTIMIZATION
The Binary Shifter

The Barrel Shifter

Area Dominated by Wiring
4x4 barrel shifter

\[ \text{Width}_{\text{barrel}} \sim 2 \, p_m \, M \]

Logarithmic Shifter
0-7 bit Logarithmic Shifter

\[ \text{width}_{\log} = p_m \left( 2^K + \left( 1 + 2 + \ldots + 2^{K-1} \right) \right) = p_m \left( 2^K + 2^K - 1 \right) \]

Power
The Importance of Power Awareness

- Crucial for Portable Applications
  » Determines battery lifetime
- Crucial for High-Performance Applications
  » Determines cooling and energy costs
  » Many designs today are power limited

The Power Challenge

- 400 million computers in the world
  » 0.16 PW (PetaWatt = $10^{15}$ W) of power dissipation
  » Equivalent to 26 nuclear plants!
- Data centers represent the absolute challenge
  » 1 single server rack is between 5 and 20 kW
  » 100’s of those racks in a single room!
Power-and-Energy Challenges

Power and energy management and minimization have emerged as some of the most dominant roadblocks. The best opportunity lies in a very aggressive scaling and adaptation of supply and threshold values in concert with a careful orchestration of the system activity.

Portability:
Battery storage the limiting factor

- Little change in basic technology
  - store energy using a chemical reaction
- Battery capacity doubles every 10 years
- Energy density/size, safe handling are limiting factor

<table>
<thead>
<tr>
<th>Energy density of material</th>
<th>KWH/kg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gasoline</td>
<td>14</td>
</tr>
<tr>
<td>Lead-Acid</td>
<td>0.04</td>
</tr>
<tr>
<td>Li polymer</td>
<td>0.15</td>
</tr>
</tbody>
</table>
Battery Progress

Facture 4 over the last 10 years!

Power Dissipation in CMOS

- Dynamic power
  » Charging capacitances
  » Dominant today
- Leakage power
  » Leaky transistors
  » Concern in low-activity, portable devices
- Short circuit power
- Static power
  » E.g. pseudo-NMOS
Dynamic Power of a CMOS Gate

\[ E_{0 \to 1} = C_L V_{dd}^2 \]

\[ E_{\text{cap}} = \int_0^T P_{\text{cap}}(t) dt = \int_0^T V_{dd} i_{\text{supply}}(t) dt = V_{dd} \int_0^T C_L dV_{out} = C_L \cdot V_{dd}^2 \]

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 \( 0 \to 1 \) transition in \( N \) clock cycles

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

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

\[ P_{\text{avg}} = \alpha_{0 \to 1} \cdot C_L \cdot V_{dd}^2 \cdot f_{\text{clk}} \]
Factors Affecting Transition Activity, $\alpha_{0\rightarrow1}$

- "Static" component (does not account for timing)
  - Type of Logic Function (NOR vs. XOR)
  - Type of Logic Style (Static vs. Dynamic)
  - Signal Statistics
  - Inter-signal Correlations

- "Dynamic" or timing dependent component
  - Circuit Topology
  - Signal Statistics and Correlations

Type of Logic Function: NOR vs. XOR

Example: Static 2 Input NOR Gate

Assume:
- $p(A=1) = 1/2$
- $p(B=1) = 1/2$

Then:
- $p(\text{Out}=1) = 1/4$
- $p(0 \rightarrow 1) = p(\text{Out}=0) \cdot p(\text{Out}=1) = 3/4 \times 1/4 = 3/16$

$\alpha_{0\rightarrow1} = 3/16$
Type of Logic Function: NOR vs. XOR

Example: Static 2 Input XOR Gate

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>Out</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Truth Table of a 2 input XOR gate

Assume:
\[ p(A=1) = 1/2 \]
\[ p(B=1) = 1/2 \]

Then:
\[ p(\text{Out}=1) = 1/2 \]
\[ p(0 \rightarrow 1) = p(\text{Out}=0) \cdot p(\text{Out}=1) = 1/2 \times 1/2 = 1/4 \]

Transition Probabilities

\[ P_{0 \rightarrow 1} (\text{NOR, NAND}) = \frac{2^N - 1}{2^N} \]
\[ P_{0 \rightarrow 1} (\text{XOR}) = 1/4 \]
Power Consumption of Dynamic Gate

Power only dissipated when previous Out = 0

Dynamic Power Consumption is Data Dependent

Dynamic 2-input NOR Gate

Assume signal probabilities
\[ P_{A=1} = 1/2 \]
\[ P_{B=1} = 1/2 \]

Then transition probability
\[ P_{0\rightarrow1} = P_{\text{out}=0} \times P_{\text{out}=1} \]
\[ = 3/4 \times 1 = 3/4 \]

Switching activity always higher in dynamic gates!
\[ P_{0\rightarrow1} = P_{\text{out}=0} \]
**Another Logic Style: Dynamic DCVSL**

Guaranteed transition for every operation!

\[ \alpha_{0\rightarrow1} = 1 \]

---

**Inter-signal Correlations**

(a) Logic circuit without reconvergent fanout

\[ p_{0\rightarrow1} = (1 - p_a p_b) p_a p_b = 3/16 \]

(b) Logic circuit with reconvergent fanout

\[ p_Z = p(C=1|B=1) \cdot p(B=1) \]

\[ p_{0\rightarrow1} = 0 \]

- Need to use conditional probabilities to model inter-signal correlations!
- CAD tools required for such analysis
Taking Delay into Account: Glitching or “Dynamic Hazards”

May cause $a_{0,1} > 1$

Example: Adder Circuit
Solution: Balanced delay paths

![Diagram of balanced delay paths]

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!)
  - Reducing thresholds to improve performance increases leakage
- Reduce switching activity
- Reduce physical capacitance
Review: EDP Plot

![Graph showing Energy-Delay (norm) versus V_DD (V)]

Architecture Trade-off for Fixed-rate Processing
Reference Data Path

- Critical path delay \( T_{\text{adder}} + T_{\text{comparator}} = 25\text{ns} \) 
  \( \Rightarrow f_{\text{ref}} = 40\text{Mhz} \)
- Total capacitance being switched = \( C_{\text{ref}} \)
- \( V_{\text{dd}} = V_{\text{ref}} = 5V \)
- Power for reference datapath = \( P_{\text{ref}} = C_{\text{ref}} V_{\text{ref}}^2 f_{\text{ref}} \)

(From [Chamtrakavanij92] (IEEE JSSC))
**Parallel Data Path**

- The clock rate can be reduced by half with the same throughput ⇒ $f_{\text{par}} = f_{\text{ref}} / 2$
- $V_{\text{par}} = V_{\text{ref}} / 1.7$, $C_{\text{par}} = 2.15C_{\text{ref}}$
- $P_{\text{par}} = (2.15C_{\text{ref}})(V_{\text{ref}}/1.7)^2(f_{\text{ref}}/2) = 0.36P_{\text{ref}}$

**Pipelined Data Path**

- Critical path delay is less ⇒ $\max [T_{\text{adder}}, T_{\text{comparator}}]$
- Keeping clock rate constant: $f_{\text{pipe}} = f_{\text{ref}}$
  Voltage can be dropped ⇒ $V_{\text{pipe}} = V_{\text{ref}} / 1.7$
- Capacitance slightly higher: $C_{\text{pipe}} = 1.15C_{\text{ref}}$
- $P_{\text{pipe}} = (1.15C_{\text{ref}})(V_{\text{ref}}/1.7)^2f_{\text{ref}} = 0.39P_{\text{ref}}$
A Simple Data Path: Summary

<table>
<thead>
<tr>
<th>Architecture type</th>
<th>Voltage</th>
<th>Area</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple datapath (no pipelining or parallelism)</td>
<td>5V</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Pipelined datapath</td>
<td>2.9V</td>
<td>1.3</td>
<td>0.39</td>
</tr>
<tr>
<td>Parallel datapath</td>
<td>2.9V</td>
<td>3.4</td>
<td>0.36</td>
</tr>
<tr>
<td>Pipeline-Parallel</td>
<td>2.0V</td>
<td>3.7</td>
<td>0.2</td>
</tr>
</tbody>
</table>

How Low a Voltage can be Used?

* Capacitance overhead starts to dominate at “high” levels of parallelism and results in an optimum voltage
Activity Reduction

- More easily accomplished at higher levels of abstraction
- Some options at circuit and logic levels:
  » Choice of circuit style
  » Technology mapping
  » Logic restructuring (see later)
  » Data encoding
  » Data-dependent enabling of computational modules
  » Gated clocks

Circuit-level Activity Reduction

from [Aldina94] (1994 International Workshop on Low-power Design)
Circuit-Level Activity Encoding

Conditional Inversion Coding for Interconnect

Transistor Sizing for Low-Power

- Larger sized devices are useful only when interconnect dominated
- Minimum sized devices are usually optimal for low-power
Transistor Sizing for Fixed Throughput

Given: performance requirement
Determine: size and voltage that minimizes power

---

EE141