Chapter 17: Low-Power Design

Keshab K. Parhi and Viktor Owall
IC Design Space
VLSI Digital Signal Processing Systems

• Technology trends:
  – 200-300M chips by 2010 (0.07 micron CMOS)

• Challenges:
  – Low-power DSP algorithms and architectures
  – Low-power dedicated / programmable systems
  – Multimedia & wireless system-driven architectures
  – Convergence of Voice, Video and Data
  – LAN, MAN, WAN, PAN
  – Telephone Lines, Cables, Fiber, Wireless
  – Standards and Interoperability
Power Consumption in DSP

- **Low performance portable applications:**
  - Cellular phones, personal digital assistants
  - Reasonable battery lifetime, low weight
- **High performance portable systems:**
  - Laptops, notebook computers
- **Non-portable systems:**
  - Workstations, communication systems
  - DEC alpha: 1 GHz, 120 Watts
  - Packaging costs, system reliability
Power Dissipation

Two measures are important

• **Peak power (Sets dimensions)**

$$P_{\text{peak}} = V_{\text{DD}} \times i_{\text{DDmax}}$$

• **Average power (Battery and cooling)**

$$P_{\text{av}} = \frac{V_{\text{DD}}}{T} \int_{0}^{T} i_{\text{DD}}(t) \, dt$$
CMOS Power Consumption

\[ P_{\text{tot}} = P_{\text{dyn}} + P_{\text{sc}} + P_{\text{leakage}} = \]

\[ = \alpha f C_L V_{DD}^2 + V_{DD} I_{\text{sc}} + I_{\text{leakage}} V_{DD} \]

\[ \alpha = \text{probability for switching} \]
Dynamic Power Consumption

Energy charged in a capacitor
\[ E_C = CV^2/2 = C_L V_{DD}^2/2 \]

Energy \( E_C \) is also discharged, i.e.
\[ E_{tot} = C_L V_{DD}^2 \]

Power consumption
\[ P = C_L V_{DD}^2 f \]
Off-Chip Connections have High Capacitive Load

↓

Reduced off Chip Data Transfers by System Integration

Ideally a Single Chip Solution

↓

Reduced Power Consumption
Switching Activity ($\alpha$):

**Example**

Due to correlation
Increased Switching Activity due to Glitching

Extra transition due to race
Dissipates energy
Clock Gating and Power Down

Only active modules should be clocked!

Control circuitry is needed for clock gating and power down and needs wake-up.
Carry Ripple

Transitions due to carry propagation
Balancing Operations

Example: Addition
Delay as function of Supply

Fig. 17.3 Circuit delay as a function of supply voltage.
Delay as function of Threshold

Fig. 17.4 Circuit delay as a function of supply voltage for varying threshold voltages.
Dual $V_T$ Technology

Reduced $V_{DD}$ $\alpha$ Increased delay
Low $V_T$ $\alpha$ Faster but Increased Leakage

Low $V_T$ in critical path
High $V_T$ stand-by

- **High $V_T$ α low leakage**
- **Low leakage in stand by when high $V_T$ transistors turned off**

Diagram notes:
- $V_{DD}$
- CL
- Fast
- High leakage
- Standby

Chapter 17
Low Power Gate Resizing

- Systematic capture and elimination of slack using fictitious entities called *Unit Delay Fictitious Buffers*.
- Replace unnecessary fast gates by slower lower power gates from an underlying gate library.
- Use a simple relation between a gate’s speed and power and the UDF’s in its fanout nets. Model the problem as an *efficiently solvable ILP* similar to retiming.
- In *Proceedings of ARVLSI’99* Georgia Tech.

Critical Path = 8, UDF’s in Boxes
Dual Supply Voltages for Low Power

- Components on the Critical Path exhibit no slack but components off the critical path exhibit excessive slack.
- A high supply voltage VDDH for critical path components and a low supply voltage VDDL for non critical path components.
- Throughput is maintained and power consumption is lowered.

Dual Supply Voltages for Low Power

- Systematic capture and elimination of slack using fictitious entities called *Unit Delay Fictitious Buffers*.
- Switch unnecessarily fast gates to lower supply voltage VDDL thereby saving power; critical path gates have a high supply voltage of VDDH.
- Use a simple relation between a gate’s speed/power and supply voltage with the UDF’s in its fanout nets. Model the problem as an *approximately solvable ILP*.

![Diagram showing critical path and UDF's in boxes](image-url)

Chapter 17
Dual Threshold CMOS VLSI for Low Power

- Systematic capture and elimination of slack using fictitious entities called *Unit Delay Fictitious Buffers*.
- Gates on the critical path have a low threshold voltage VTL and unnecessarily fast gates are switched to a high threshold voltage VTH.
- Use a simple relation between a gate’s speed/power and threshold voltage with the UDF’s in its fanout nets. Model the problem as an *efficiently approximable 0-1 ILP*.

Critical Path = 8, UDF’s in Boxes

![Diagram](image-url)
## Experimental Results

- **Table:** ISCAS’85 Benchmark Ckts Resizing (20 Sizes) Dual VDD Dual

<table>
<thead>
<tr>
<th>Ckt</th>
<th>#Gates</th>
<th>Power Savings</th>
<th>CPU(s)</th>
<th>Power Savings</th>
<th>CPU(s)</th>
<th>Power Savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1908</td>
<td>880</td>
<td>15.27%</td>
<td>87.5</td>
<td>49.5%</td>
<td>739.05</td>
<td>84.92%</td>
</tr>
<tr>
<td>c2670</td>
<td>1211</td>
<td>28.91%</td>
<td>164.38</td>
<td>57.6%</td>
<td>1229.37</td>
<td>90.25%</td>
</tr>
<tr>
<td>c3540</td>
<td>1705</td>
<td>37.11%</td>
<td>312.51</td>
<td>57.7%</td>
<td>1743.75</td>
<td>83.36%</td>
</tr>
<tr>
<td>c5315</td>
<td>2351</td>
<td>41.91%</td>
<td>660.56</td>
<td>62.4%</td>
<td>4243.63</td>
<td>91.56%</td>
</tr>
<tr>
<td>c6288</td>
<td>2416</td>
<td>5.57%</td>
<td>69.58</td>
<td>62.7%</td>
<td>7736.05</td>
<td>61.75%</td>
</tr>
<tr>
<td>c7552</td>
<td>3624</td>
<td>54.05%</td>
<td>1256.76</td>
<td>59.6%</td>
<td>9475.1</td>
<td>90.90%</td>
</tr>
</tbody>
</table>

HEAT: Hierarchical Energy Analysis Tool

- **Salient features:**
  - Based on stochastic techniques
  - Transistor-level analysis
  - Effectively models glitching activity
  - Reasonably fast due to its hierarchical nature
Theoretical Background

- Signal probability:
  - \( S = \frac{T_{clk}}{T_{gd}} \), where
    \( T_{clk} \): clock period
    \( T_{gd} \): smallest gate delay

- Transition probability:

- Conditional probability:
State Transition Diagram Modeling

\[ \text{node}_2(n+1) = (1 - x_1(n)) + x_1(n) \cdot x_2(n) \cdot \text{node}_2(n) \]

\[ \text{node}_3(n+1) = (1 - x_1(n)) + (1 - x_2(n)) \]
The HEAT algorithm

- Partitioning of systems unit into smaller sub-units
- State transition diagram modeling
- Edge energy computation (HSPICE)
- Computation of steady-state probabilities (MATLAB)
- Edge activity computation
- Computation of average energy

\[ \text{Energy} = W_j \cdot EA_j \]
Performance Comparison

Finite field arithmetic -- Addition and Multiplication

\[ A = a_{m-1} \alpha^{m-1} + \ldots + a_1 \alpha + a_0 \]

\[ B = b_{m-1} \alpha^{m-1} + \ldots + b_1 \alpha + b_0 \]

\[ A + B = (a_{m-1} + b_{m-1}) \alpha^{m-1} + \ldots + (a_1 + b_1) \alpha + (a_0 + b_0) \]

\[ A \cdot B = (a_{m-1} \alpha^{m-1} + \ldots + a_1 \alpha + a_0)(b_{m-1} \alpha^{m-1} + \ldots + b_1 \alpha + b_0) \mod(p(x)) \]

Polynomial addition over GF(2)

one’s complement operation --&gt; XOR gates

Polynomial multiplication and modulo operation

(modulo primitive polynomial p(x))
Programmable finite field multiplier

Array-type

Parallel

Digit-serial

MAC2

+ DEGRED2

Four Instr.

MAC2

MAC2

DEGRED2

DEGRED2
Finite field arithmetic--programmable finite field multipliers

Programmability:-primitive polynomial $p(x)$
 -field order $m$

How to achieve programmability:-control circuitry
 -zero, pre & post padding

Polynomial multiplication
Polynomial modulo operation

Array-type multiplication
Fully parallel multiplication
Digit-serial/parallel multiplication

Data-path architectures for low energy RS codecs

• Advantages of having two separate sub-arrays
  – Example: Vector-vector multiplication over GF(2^m)

\[
\begin{bmatrix}
A_0 & A_1 & \ldots & A_{n-1}
\end{bmatrix}
\begin{bmatrix}
B_0 \\
B_1 \\
\vdots \\
B_{n-1}
\end{bmatrix} = (A_0B_0 + \ldots + A_{n-1}B_{n-1}) \mod (p(x))
\]

– Assume energy(parallel multiplier)=Eng

Energy(MAC8x8)=0.25 Eng  \quad \text{Total Energy(parallel)=Eng}*n
Energy(DEGRED7)=0.75 Eng  \quad \text{Total Energy(MAC-D7)=0.25Eng}*n+0.75Eng

\[
s = \frac{\text{Eng} \cdot (n - (0.25n + 0.75))}{\text{Eng} \cdot n} \approx 75\%
\]
Data-path architectures for low-power RS encoder

- Data-paths
  - One parallel finite field multiplier
  - Digit-serial multiplication: MACx and DEGREDy
Data-path architectures for low energy RS codecs

• Data-path:
  – one parallel finite field multiplier
  – Digit-serial multiplication: MACx and DEGREDy

MAC8 + DEGRED2
MAC8 + DEGRED1
MAC4 + DEGRED2
MAC4 + DEGRED1
MAC8 + DEGRED4
MAC8 + DEGRED2

Low power design challenges

- System Integration
- Application Specific architectures for Wireless/ADSL/Security
- Programmable DSPs to handle new application requirements
- Low-Power Architectures driven by Interconnect, Crosstalk in DSM technology
- How Far are we away from PDAs/Cell Phones for wireless video, internet access and e-commerce?