Chapter 8
Optimizing Power @ Standby – Circuits and Systems

Slide 8.1

Optimizing Power @ Standby
Circuits and Systems

Jan M. Rabaey
Slide 8.2

Chapter Outline

- Why Sleep Mode Management?
- Dynamic power in standby
  - Clock gating
- Static power in standby
  - Transistor sizing
  - Power gating
  - Body biasing
  - Supply voltage ramping
Arguments for Sleep Mode Management

- Many computational applications operate in burst modes, interchanging active and non-active modes
  - General-purpose computers, cell phones, interfaces, embedded processors, consumer applications, etc.
- Prime concept: Power dissipation in standby should be absolutely minimum, if not zero
- Sleep mode management has gained importance with increasing leakage

<table>
<thead>
<tr>
<th></th>
<th>Fixed Activity</th>
<th>Variable Activity</th>
<th>No Activity - Standby</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>Design Time</td>
<td>Run Time</td>
<td>Clock gating</td>
</tr>
<tr>
<td>Static</td>
<td></td>
<td></td>
<td>Leakage elimination</td>
</tr>
</tbody>
</table>

Slide 8.3
Standby Power Was Not a Concern in Earlier Days

Pentium-1: 15 W (5 V - 66 MHz)
Pentium-2: 8 W (3.3 V - 133 MHz)

Processor in idle mode!
Floating-point unit and cache powered down when not in use

[Source: Intel]
Dynamic Power – Clock Gating

- Turn off clocks to idle modules
  - Ensure that spurious activity is set to zero
- Must ensure that data inputs to the module are in stable mode
  - Primary inputs are from gated latches or registers
  - Or, disconnected from interconnect network
- Can be done at different levels of system hierarchy
Clock Gating

Turning off the clock to non-active components

Disconnecting the inputs

Slide 8.6
Clock gating Efficiently Reduces Power

Without clock gating

With clock gating

30.6 mW

8.5 mW

Power [mW]

0 5 10 15 20 25

90% of FFs clock-gated.

70% power reduction by clock gating alone.

[Ref: M. Ohashi, ISSCC’02] © IEEE 2002

Slide 8.7
Clock Gating

- Challenges to skew management and clock distribution (load on clock network varies dynamically)
- Fortunately state-of-the-art design tools are starting to do a better job
  - For example, physically aware clock gating inserts gaters in clock tree based on timing constraints and physical layout

Power savings

Simpler skew management, less area
Clock Hierarchy and Clock Gating

Example: Clock distribution of dual-core Intel Montecito processor

"Gaters" provided at lower clock-tree levels
Automatic skew compensation

[Ref: T. Fischer, ISSCC’05]
### Trade-Off Between Sleep Modes and Sleep Time

#### Typical operation modes

<table>
<thead>
<tr>
<th>Mode</th>
<th>Active mode</th>
<th>Standby mode</th>
<th>Sleep mode</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>normal processing</td>
<td>fast resume high passive power</td>
<td>slower resume low passive power</td>
</tr>
</tbody>
</table>

Resume-time from clock gating determined by the time it takes to turn on the clock distribution network

**Standby Options:**
- Just gate the clock to the module in question
- Turn off phased-locked loop(s)
- Turn off clock completely

---

*Slide 8.10*
Sleep Modes in μProcessors and μControllers

- PowerPC 603
- Hitachi SH7032
- MIPS 4200

TI MSP430™

- From standby to active in 1 μs using dual clock system

- 0.1-μA power down
- 0.8-μA standby
- 250-μA/MIPS @ 3 V

Slide 8.11
The Standby Design Exploration Space

Trade-off between different operational modes
Should blend smoothly with runtime optimizations

Slide 8.12
Also the Case for Peripheral Devices

### Hard disk

<table>
<thead>
<tr>
<th></th>
<th>P_{sleep} W</th>
<th>P_{active} W</th>
<th>T_{sleep} sec</th>
<th>T_{active} sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBM</td>
<td>0.75</td>
<td>3.48</td>
<td>0.51</td>
<td>6.97</td>
</tr>
<tr>
<td>Fujitsu</td>
<td>0.13</td>
<td>0.95</td>
<td>0.67</td>
<td>1.61</td>
</tr>
</tbody>
</table>

### Wireless LAN Card

<table>
<thead>
<tr>
<th></th>
<th>TX</th>
<th>RX</th>
<th>Doze</th>
<th>Off</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power</td>
<td>1.65 W</td>
<td>1.4 W</td>
<td>0.045 W</td>
<td>0 W</td>
</tr>
<tr>
<td>Transitions</td>
<td>To Off: 62 ms</td>
<td>To Doze: 34 ms</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[Ref: T. Simunic, Kluwer'02]
The Leakage Challenge – Power in Standby

- With clock gating employed in most designs, leakage power has become the dominant standby power source.
- With no activity in module, leakage power should be minimized as well.
  - Remember constant ratio between dynamic and static power …
- Challenge – how to disable unit most effectively given that no ideal switches are available.
Standby Static Power Reduction Approaches

- Transistor stacking
- Power gating
- Body biasing
- Supply voltage ramping
Transistor Stacking

- Off-current reduced in complex gates (see leakage power reduction @ design time)
- Some input patterns more effective than others in reducing leakage
- Effective standby power reduction strategy:
  - Select input pattern that minimizes leakage current of combinational logic module
  - Force inputs of module to correspond to that pattern during standby
- Pros: Little overhead, fast transition
- Con: Limited effectiveness
Transistor Stacking

Combinational Module

Latches

Clk

Standby

Latches

[Ref: S. Narendra, ISLPED’01]
Forced Transistor Stacking

Useful for reducing leakage in non-critical shallow gates (in addition to high $V_{TH}$)

[Ref: S. Narendra, ISLPED'01]

Slide 8.18
Power Gating

- Disconnect module from supply rail(s) during standby
- Footer or header transistor, or both
- Most effective when high-$V_{th}$ transistors are available
- Easily introduced in standard design flows
- But ... Impact on performance

Very often called “MTCMOS” (when using high- and low-threshold devices)

[Ref: T. Sakata, VLSI’93; S. Mutoh, ASIC’93]
Power Gating – Concept

Leakage current reduces because
- Increased resistance in leakage path
- Stacking effect introduces source biasing

\[ V_{DD} \]
\[ V_S = I_{leak} R_S \]

(Similar effect at PMOS side)

Slide 8.20
Power Gating Options

- NMOS sleeper transistor more area-efficient than PMOS
- Leakage reduction more effective (under all input patterns) when both footer and header transistors are present

Slide 8.21
Other option: Boosted-Gate MOS (BGMOS)

CMOS logic
- low $V_{TH}$
- thin $T_{OX}$

Leak cut-off Switch (LS)
- high $V_{TH}$
- thick $T_{OX}$
  (eliminates tunneling)

[T. Inukai, CICC’00]
Other Option: Boosted-Sleep MOS
(also called Super-Cutoff CMOS or SCCMOS)

- CMOS logic
  - low $V_{TH}$
  - thin $T_{OX}$

- Leak cut-off Switch (LS)
  - normal (or high) $V_{TH}$
  - normal $T_{OX}$

Area-efficient

[Ref: T. Inukai, CICC’00]
Virtual Supplies

Active Mode

Standby Mode

Noise on virtual supplies

Virtual supply collapse

[Ref: J. Tschanz, JSSC’03]
Decoupling Capacitor Placement

- Decap on supply rails
- Decap on virtual rails

Performance:
- Convergence time: ✓
- Oxide leakage savings: ✓

Oxide leakage savings:
- Longer time constant:
- Reduced leakage:

[Ref: J. Tschanz, JSSC’03]
Leakage Power Savings versus Decap

- Low-leakage 133 nF decap on virtual $V_{CC}$
- No decap on virtual $V_{CC}$

Normalized leakage power in idle mode

1.32 V
75°C

[Ref: J. Tschanz, JSSC'03]
How to Size the Sleep Transistor?

- Sleep transistor is not free – it will degrade the performance in active mode
- Circuits in active mode see the sleep transistor as extra power-line resistance
  - The wider the sleep transistor, the better
- Wide sleep transistors cost area
  - Minimize the size of the sleep transistor for given ripple (e.g., 5%)
  - Need to find the worst-case vector
Sleep Transistor Sizing

- High-$V_{TH}$ transistor must be very large for low resistance in linear region
- Low-$V_{TH}$ transistor needs less area for same resistance.

<table>
<thead>
<tr>
<th></th>
<th>MTCMOS</th>
<th>Boosted Sleep</th>
<th>Non-Boosted Sleep</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sleep TR size</td>
<td>5.1%</td>
<td>2.3%</td>
<td>3.2%</td>
</tr>
<tr>
<td>Leakage power reduction</td>
<td>1450x</td>
<td>3130x</td>
<td>11.5x</td>
</tr>
<tr>
<td>Virtual supply bounce</td>
<td>60 mV</td>
<td>59 mV</td>
<td>58 mV</td>
</tr>
</tbody>
</table>

[Ref: R. Krishnamurthy, ESSCIRC'02]
Preserving State

- Virtual supply collapse in sleep mode causes the loss of state in registers
- Keeping the registers at nominal $V_{DD}$ preserves the state
  - These registers leak ...
- Can lower the $V_{DD}$ in sleep
  - Some impact on robustness, noise, and soft-error immunity
Latch-Retaining State During Sleep

Black-shaded devices use low- $V_{TH}$ transistors
All others are high- $V_{TH}$

[Ref: S. Mutoh, JSSC’95]
MTCMOS Derivatives Preventing State Loss

Clamping

Retention

Reduce voltage and retain state
Sleep Transistor Placement

No sleep transistors

With headers and footers

Slide 8.32
Sleep Transistor Layout

Slide 8.33

<table>
<thead>
<tr>
<th>Area overhead</th>
<th>PMOS</th>
<th>6%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NMOS</td>
<td>3%</td>
</tr>
</tbody>
</table>

[Ref: J. Tschanz, JSSC'03]
Dynamic Body Biasing

- Increase thresholds of transistors during sleep using reverse body biasing
  - Can be combined with forward body biasing in active mode
- No delay penalty

**But**

- Requires triple-well technology
- Limited range of threshold adjustments (<100 mV)
  - Not improving with technology scaling
- Limited leakage reduction (<10x)
- Energy cost of charging/discharging the substrate capacitance

*Slide 8.34*
Dynamic Body Biasing

Active mode: Forward Body Bias

- Low threshold, high performance
- High threshold, low leakage

Standby mode: Reverse Body Bias

Can also be used to compensate for threshold variations

[Ref's: T Kuroda ISSCC'96; J. Tschanz, JSSC'03]
The Dynamics of Dynamic Body Bias

Needs level-shifting and voltage-switch circuitry

[Ref: K. Seta, ISSCC'95]
Body Bias Layout

- Sleep transistor LBGs
- ALU core LBGs

```
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of ALU core LBGs</td>
<td>30</td>
</tr>
<tr>
<td>Number of sleep transistor LBGs</td>
<td>10</td>
</tr>
<tr>
<td>PMOS device width</td>
<td>13 mm</td>
</tr>
<tr>
<td>Area overhead</td>
<td>8%</td>
</tr>
</tbody>
</table>
```

[Ref: J. Tschanz, JSSC'03]

Slide 8.37
Application-specific processor (SH-mobile)
- 250 nm technology
- core at 1.8 V
- I/O at 3.3 V
- 3.3M transistors

[Ref: M. Miyazaki, Springer/06]
Effectiveness of Dynamic Body Biasing

Practical $V_{TH}$ tuning range less than 150 mV in 90 nm technology

Slide 8.39
Supply Voltage Ramping (SVR)

- Reduce supply voltage of modules in sleep mode
  - Can go to 0 V if no state-retention is necessary
  - Down to state retention voltage otherwise, (see Memory in next chapter), or move state to persistent memory before power-down
- Most effective leakage reduction technique
  - Reduces current and voltage

But
- Needs controllable voltage regulator
  - Becoming present more often in modern integrated system designs
- Longer reactivation time

Simplified version switches between $V_{DD}$ and GND (or $V_{DDL}$)

[Ref: M. Sheets, VLSI'06]
Supply Ramping

- Standby power = $V_{DD}(\text{standby}) \times I_{\text{leak}}(\text{standby})$
- Modules must be isolated from neighbors
- Creating “voltage islands”

Slide 8.41
Supply Ramping – Impact

Leakage power as a function of the supply voltage (90 nm)

Because of DIBL, dropping supply voltage causes dramatic reduction in leakage – Can go as low as 300 mV before data retention is lost.
Integration in Standard-Cell Layout Methodology

- Power switch cell easily incorporated into standard design flow
  - Cell has same pitch as existing components
  - No changes required to cell library from foundry
- Switch design can be independent of block size

Slide 8.43
## Standby Leakage Management – A Comparison

<table>
<thead>
<tr>
<th></th>
<th>Transistor Stacking</th>
<th>Power Gating</th>
<th>Dynamic Body Biasing</th>
<th>Supply Voltage Ramping</th>
</tr>
</thead>
</table>
| **Pros**               | Conventional technology  
No performance impact | Conventional technology  
Conceptually simple  
Most effective | Re-use of standard designs  
No performance impact | Most effective  
Also available in switched version |
| **Cons**               | Limited impact  
Special registers | Performance impact of serial transistor  
Changes in design flow | Triple well  
Slow activation  
Does not fare well with technology scaling | Needs voltage regulator or extra rails  
Slow activation |
| **Potential Savings**  | 5–10                 | 2–40         | 2–1000               | Huge                   |
Some Long-Term Musings

- Ideal power-off switch should have zero leakage current ($S = 0$ mV/decade)
- Hard to accomplish with traditional electronic devices
- Maybe possible using MEMS – mechanical switches have a long standing reputation for good isolation

[Ref: N. Abele, IEDM'05]
Today’s designs are not leaky enough to be truly power–performance optimal! Yet, when not switching, circuits should not leak!

Clock gating effectively eliminates dynamic power in standby

Effective standby power management techniques are essential in sub-100 nm design
- Power gating the most popular and effective technique
- Can be supplemented with body biasing and transistor stacking
- Voltage ramping probably the most effective technique in the long range (if gate leakage becomes a bigger factor)

Emergence of “voltage or power” domains
References

Books and Book Chapters

Articles
- T. Kuroda et al., “A 0.9 V 150 MHz 10 mW 4 mm2 2-D discrete cosine transform core processor with variable-threshold-voltage scheme,” JSSC, 31(11), pp. 1770–1779, Nov. 1996.
References (cont.)

- TI MSP430 Microcontroller family, http://focus.ti.com/lit/Slab034n/Slab034n.pdf

Slide 8.48