Announcements

- Homework 4 due next week
- Quiz #4 next Monday
Outline

› Last lecture
  › DVS
  › Clock gating
  › Managing leakage
› This lecture
  › Power gating
  › Body bias

5. Low Power Design
J. Power Gating
## Power /Energy Optimization Space

<table>
<thead>
<tr>
<th>Energy</th>
<th>Constant Throughput/Latency</th>
<th>Variable Throughput/Latency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Active</strong></td>
<td>Design Time</td>
<td>Sleep Mode</td>
</tr>
<tr>
<td>Logic design</td>
<td></td>
<td>Clock gating</td>
</tr>
<tr>
<td>Scaled $V_{DD}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trans. sizing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multi-$V_{DD}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Leakage</strong></td>
<td>Stack effects</td>
<td>Sleep T’s</td>
</tr>
<tr>
<td>Trans sizing</td>
<td>Scaling $V_{DD}$</td>
<td>Multi-$V_{DD}$</td>
</tr>
<tr>
<td>$+ Multi-V_{Th}$</td>
<td></td>
<td>Variable $V_{Th}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ Input control</td>
</tr>
</tbody>
</table>

---

## Dynamic Sleep Transistor

**Active mode**

- PMOS forward body bias
- ON: gate overdrive
- Virtual $V_{cc}$
- Noise on virtual supply
- Virtual $V_{ss}$

*Courtesy of J. Tschanz, Intel (ISSCC'03)*
**Dynamic Sleep Transistor**

**Idle mode**

- OFF: gate underdrive
- PMOS reverse body bias
- Virtual \(V_{CC}\)
- Virtual \(V_{SS}\)
- Virtual supply collapse

Courtesy of J. Tschanz, Intel (ISSCC’03)

---

**How to Size the Sleep Transistor?**

- Don’t need both header and footer
- 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 is not for free – it will degrade the performance in active mode
- Charging and discharging the virtual rails costs energy
- Need to sequentially wake up
Sleep Transistor

High-VTH transistor has to be very large for low resistance in linear region. Low-VTH transistor needs much less area for the 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</td>
<td>1450X</td>
<td>3130X</td>
<td>11.5X</td>
</tr>
<tr>
<td>reduction Virtual supply bounce</td>
<td>60 mV</td>
<td>59 mV</td>
<td>58 mV</td>
</tr>
</tbody>
</table>

Courtesy: R. Krishnamurthy, Intel

Sleep Transistor Layout

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

Tschanz, ISSCC’03
Sleep in Standard Cells

<table>
<thead>
<tr>
<th>Schematics</th>
<th>All HVT (hvt_ND2)</th>
<th>All LVT (lvt_ND2)</th>
<th>Footswitch (fs_ND2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perf.</td>
<td>1X</td>
<td>1.5X - 2X</td>
<td>1.4X - 1.8X</td>
</tr>
<tr>
<td>Leakage</td>
<td>1X</td>
<td>70X - 100X</td>
<td>&gt; 1X</td>
</tr>
<tr>
<td>Area</td>
<td>1X</td>
<td>1X</td>
<td>1.25X</td>
</tr>
</tbody>
</table>

Uvieghara, ISSCC’04

Sleep Transistor Grid

No sleep transistor

PMOS & NMOS sleep transistors

Virtual $V_{CC}$

Virtual $V_{SS}$

Tschanz, ISSCC’03
Preserving State

- Virtual supply collapse in sleep mode will cause the loss of state in registers
- Putting the registers at nominal VDD would preserve the state
  - These registers leak
  - The second supply needs to be routed as well
- Can lower VDD in sleep
  - Some impact on robustness, noise and SEU immunity
- State preservation and recovery

Register Design

[Mutoh95]
## 5. Low Power Design

### K. Voltage Scaling in Sleep

### Power /Energy Optimization Space

<table>
<thead>
<tr>
<th></th>
<th>Constant Throughput/Latency</th>
<th>Variable Throughput/Latency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Energy</strong></td>
<td>Design Time</td>
<td>Sleep Mode</td>
</tr>
<tr>
<td><strong>Active</strong></td>
<td>Logic design, Scaled $V_{DD}$</td>
<td>Clock gating</td>
</tr>
<tr>
<td></td>
<td>Trans. sizing, Multi-$V_{DD}$</td>
<td></td>
</tr>
<tr>
<td><strong>Leakage</strong></td>
<td>Stack effects, Trans sizing</td>
<td>Sleep T’s, Multi-$V_{DD}$</td>
</tr>
<tr>
<td></td>
<td>Scaling $V_{DD}$, + Multi-$V_{Th}$</td>
<td>Variable $V_{Th}$, + Input control</td>
</tr>
</tbody>
</table>
Leakage vs. Supply

5. Low Power Design
L. Multiple Thresholds
## Power / Energy Optimization Space

<table>
<thead>
<tr>
<th>Energy</th>
<th>Constant Throughput/Latency</th>
<th>Variable Throughput/Latency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>Design Time</td>
<td>Sleep Mode</td>
</tr>
<tr>
<td></td>
<td>Logic design</td>
<td>Scaled V&lt;sub&gt;DD&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td>Trans. sizing</td>
<td>Clock gating</td>
</tr>
<tr>
<td></td>
<td>Multi-V&lt;sub&gt;DD&lt;/sub&gt;</td>
<td></td>
</tr>
<tr>
<td>Leakage</td>
<td>Stack effects</td>
<td>Sleep T's</td>
</tr>
<tr>
<td></td>
<td>Trans sizing</td>
<td>Multi-V&lt;sub&gt;DD&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td>Scaling V&lt;sub&gt;DD&lt;/sub&gt;</td>
<td>Variable V&lt;sub&gt;Th&lt;/sub&gt;</td>
</tr>
<tr>
<td></td>
<td>+ Multi-V&lt;sub&gt;Th&lt;/sub&gt;</td>
<td>+ Input control</td>
</tr>
</tbody>
</table>

## Technology Options

![Normalized currents](image)

- **HP**
- **LP (LOP)**

- Normalized currents: I<sub>IL, HS, LP</sub>, I<sub>IL, LP</sub>, I<sub>IL, HP</sub>, I<sub>IL</sub>
Using Multiple Thresholds

- Cell-by-cell $V_T$ assignment (not block level)
- Allows us to minimize leakage
- Achieves all-low-$V$ performance

5. Low Power Design
L. Transistor Sizing
<table>
<thead>
<tr>
<th></th>
<th>Constant Throughput/Latency</th>
<th>Variable Throughput/Latency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Energy</strong></td>
<td>Design Time</td>
<td>Sleep Mode</td>
</tr>
<tr>
<td><strong>Active</strong></td>
<td>Logic design</td>
<td>Sleep Mode</td>
</tr>
<tr>
<td></td>
<td>Scaled $V_{DD}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trans. sizing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Multi-$V_{DD}$</td>
<td></td>
</tr>
<tr>
<td><strong>Leakage</strong></td>
<td>Stack effects</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trans sizing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scaling $V_{DD}$</td>
<td>Sleep T’s</td>
</tr>
<tr>
<td></td>
<td>+ Multi-$V_{Th}$</td>
<td>Multi-$V_{DD}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Variable $V_{Th}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+ Input control</td>
</tr>
</tbody>
</table>

**Power /Energy Optimization Space**

- **Active**
  - Logic design
  - Scaled $V_{DD}$
  - Trans. sizing
  - Multi-$V_{DD}$

- **Leakage**
  - Stack effects
  - Trans sizing
  - Scaling $V_{DD}$
  - Multi-$V_{DD}$
  - Variable $V_{Th}$
  - Input control

**Longer Channels**

- **35%** longer gates reduce leakage by 35%
- Increases switching energy by 21% with $W/L = \text{const.}$

- Attractive when don't have to increase $W$ (memory)
- Doubling $L$ reduces leakage by 3x (in 0.13um)
- Much stronger effect in 28nm!
- Effect improves with more aggressive devices
## 5. Low Power Design
### M. Multiple Supplies

### Power/Energy Optimization Space

<table>
<thead>
<tr>
<th>Energy</th>
<th>Design Time</th>
<th>Sleep Mode</th>
<th>Run Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>Logic design</td>
<td>Clock gating</td>
<td>DFS, DVS</td>
</tr>
<tr>
<td></td>
<td>Scaled $V_{DD}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trans. sizing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Multi-$V_{DD}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leakage</td>
<td>Stack effects</td>
<td>Sleep T's</td>
<td>+ Variable $V_{th}$</td>
</tr>
<tr>
<td></td>
<td>Trans sizing</td>
<td>Multi-$V_{DD}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scaling $V_{DD}$</td>
<td>Variable $V_{th}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+ Multi-$V_{th}$</td>
<td>+ Input control</td>
<td></td>
</tr>
</tbody>
</table>
5. Low Power Design

N. Dynamic Threshold Scaling
### Power /Energy Optimization Space

<table>
<thead>
<tr>
<th>Energy</th>
<th>Constant Throughput/Latency</th>
<th>Variable Throughput/Latency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Design Time</td>
<td>Sleep Mode</td>
</tr>
<tr>
<td>Active</td>
<td>Logic design</td>
<td>Clock gating</td>
</tr>
<tr>
<td></td>
<td>Scaled ( V_{DD} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trans. sizing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Multi-( V_{DD} )</td>
<td></td>
</tr>
<tr>
<td>Leakage</td>
<td>Stack effects</td>
<td>Sleep T’s</td>
</tr>
<tr>
<td></td>
<td>Trans sizing</td>
<td>Multi-( V_{DD} )</td>
</tr>
<tr>
<td></td>
<td>Scaling ( V_{DD} )</td>
<td>Variable ( V_{Th} )</td>
</tr>
<tr>
<td></td>
<td>+ Multi-( V_{Th} )</td>
<td>+ Input control</td>
</tr>
</tbody>
</table>

### Dynamic Body Bias

- Similar concept to dynamic voltage scaling
- Control loop adjusts the substrate bias to meet the timing
  - Can be used just as runtime/sleep
- Limited range of threshold adjustments (<100mV)
- Limited leakage reduction (<10x)
- No delay penalty
  - Can increase speed by forward bias
- Energy cost of charging/discharging the substrate capacitance
  - (but doesn’t need a regulator)
Dynamic Body Bias

**Active mode**
- Forward body bias (FBB)
- Local $V_{CC}$ tracking

**Idle mode**
- Reverse body bias (RBB)
- Triple well needed

Tschanz, ISSCC’03
**Body Bias Layout**

- **Sleep transistor LBGs**
- **ALU core LBGs**

<table>
<thead>
<tr>
<th>Component</th>
<th>Value</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>

**Total Active Power Savings**

(Fixed activity: $\alpha = 0.05$)

- PMOS sleep transistor (1.32V): Max 18%
- Body bias (1.28V): active: FBB, idle: ZBB Max 8%

Power savings for $T_{OFF} > \sim 100$ idle cycles

Reference: 450mV FBB to core with clock gating, 1.28V, 4.05GHz, 75°C
Body Biasing and Variations

- Body biasing with a local control loop can be used to lower the impact of process variations
- Used to limit die-to-die and within-die variations

Normalized Delay vs $V_{DD}$ & $V_{TH}$

[Sakurai, Kuroda]
Self-Adjusting Threshold-Voltage Scheme (SATS)

low $V_{th}$ → large leakage → SSB ON → deep $V_{BB}$ → high $V_{th}$

high $V_{th}$ → little leakage → SSB OFF → shallow $V_{BB}$ → low $V_{th}$

- control $V_{th}$ to adjust leakage current
- compensate $V_{th}$ fluctuation

Substrate Biasing

150nm CMOS technology

Frequency too low
Leakage too high

$P_{max} = 20 \text{ W/cm}^2$

$P_{leak,\text{max}} = P_{max} - \alpha CV^2 f$

$110^\circ C$
$V_{CC}: 1.1V$
$\alpha: 0.05$

Tschanz, JSSC 11/02
Effectiveness of Substrate Bias

**Die-to-die variations**

- Normalized leakage vs. normalized frequency for Accepted dies: NBB and ABB.
- Frequency Variation:
  - NBB: 4.1%
  - ABB: 0.69%

**Within-die variations**

- Normalized leakage vs. normalized frequency for Accepted dies: ABB and WID-ABB.
- Frequency Variation:
  - ABB: 0.69%
  - WID-ABB: 0.21%

97% in highest bin
Techniques Summary

- Sleep transistor - up to ~25x leakage reduction
- Standby supply reduction - ~3-4x leakage reduction
- Reverse bias - ~3x leakage reduction
- Standby supply + reverse bias - ~10x leakage reduction

Next Lecture

- Optimal supplies