Reduction of Leakage Currents with Mixed Gates in Deep Submicron Technologies

Frank Sill, Claas Cornelius, Dirk Timmermann
University of Rostock
Institute of Applied Microelectronics and Computer Engineering
Outline

1. Motivation
2. Basics
3. Mixed Gates
4. Algorithm
5. ISCAS – Benchmarks
6. Conclusion
1. Motivation
Why Thinking about Leakage?
Trend: Power

Power Density (W/cm²)

- SiO2 Lkg
- SD Lkg
- Active

90nm, 65nm, 45nm, 32nm, 22nm, 16nm

S. Borkar, ‘05
Focus of this Work

1. Modification of established Leakage Reduction Approach (DTCMOS)

2. Allocation algorithm
2. Basics
Power Dissipation in CMOS

- $I_{sub}$ occurs if $V_g < V_{th}$
- carriers move by diffusion along surface
- $I_{gate}$ caused by direct tunneling through gate oxide
Power & Delay Dependence

\[ P = p_t \cdot f_{CLK} \cdot C_L \cdot V_{DD}^2 + I_0 \frac{W_T}{W_0} \cdot 10^{-\frac{V_{TH}}{S}} \cdot V_{DD} \]

\[ t_d = \frac{k \cdot Q}{I} = \frac{k' \cdot C_L \cdot V_{DD}}{(W / L) \cdot (V_{DD} - V_{TH})^{\alpha_k}} \]

w.o. gate leakage

Sakurai, ‘01
Power & Delay Dependence

\[ P = p_t \cdot f_{\text{CLK}} \cdot C_L \cdot V_{DD}^2 + I_0 \cdot \frac{W_T}{W} \cdot 10^{\frac{-V_{TH}}{S}} \cdot V_{DD} \]

\[ t_d = \frac{k \cdot Q}{I} = \frac{k' \cdot C_L \cdot V_{DD}}{(W / I) \cdot (V - V_{TH})^{\alpha_k}} \]

**Problem:**

fast transistors with high power dissipation (low \( V_{th} \))

or

slow transistors with low power dissipation (high \( V_{th} \))

Sakurai, '01
$T_{ox}$ vs. Delay and Leakage

Inverter (BPTM 65nm)

Leakage vs. Delay and Leakage

Thickness of gate oxide (Tox) [m]

Leakage

Delay
Problem:

fast transistors with high power dissipation (low $T_{ox}$)

or

slow transistors with low power dissipation (high $T_{ox}$)

Thickness of gate oxide (Tox) [m]

1.40E-09  1.60E-09  1.70E-09  1.80E-09  2.00E-09  2.20E-09
Dual $V_{th}$ (DTCMOS) / Dual $T_{ox}$ (DTtoCMOS)

- Use different $V_{th}$’s / $T_{ox}$’s
  - use lower $V_{th}$ / $T_{ox}$ for devices on the critical paths
  - use higher $V_{th}$ / $T_{ox}$ for devices off the critical paths
- Decrease power without performance penalty
- Approaches at:
  - Gate level (V.Sundararajan et al., LPED’99)
  - Transistor level (L.Wei et al., DAC’99)
3. Mixed Gates
Mixed Pull-Down/Up-Paths

Goal (for fast gates): keep the delay while leakage decreases

if all transistors in gates are dimensioned regularly
( = PMOS and NMOS have same resistance)
→ different output slopes
up to now: sizing of transistors

idea: use different $V_{th}$ / $T_{ox}$ within a gate to adapt the slopes
Mixed Stacks

**Goal:** additional gate types at constant mask count

only two gate types in DTCMOS / DToCMOS at gate level (HVTO, LVTO)

→ problem: more LVTO gates after optimization as needed to keep the delay

**idea:** mixed $V_{th} / T_{ox}$ within a gate (stack & Pull-up/Pull-down path)
Up to now: *two kinds of cells* (DTCMOS/DToCMOS)
4. Algorithm
Slack

all Inputs of G1 arrived

G1 ready with evaluation

delay of G1

Slack for G1

all inputs of G2 arrived

University of Rostock
Institute of Applied Microelectronics and Computer Engineering
**Weight $\psi_n$**

\[ \psi_n = I_{\text{leak\_diff}} \cdot \left(1 + t_{\text{slack}} - t_{\text{diff}}\right) \cdot \left(\text{pos}_{\text{weight}}\right)^{-0.5} \]

- $I_{\text{leak\_diff}}$ = difference of leakage of weighted gate type compared to leakage of current gate type
- $t_{\text{diff}}$ = difference of delay of weighted gate type compared to delay current gate type
- $\text{pos}_{\text{weight}} = \sum_{\text{gate\_inputs}} + \sum_{\text{connected\_gates\_on\_output}}$
Allocation algorithm

**MLVTO netlist**
- Estimate signal probabilities

**For each gate:**
- Estimate slack / leakage
- Estimate $\psi_n$

**Insert HVTO / MVTO gates**

**Pin Reordering**
- no further optimizations

**MG netlist**
5. ISCAS - Benchmarks
LVT vs. MG

Reduction of Leakage Currents

<table>
<thead>
<tr>
<th>Component</th>
<th>Leakage Reduction (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>c432</td>
<td>50%</td>
</tr>
<tr>
<td>c499</td>
<td>60%</td>
</tr>
<tr>
<td>c880</td>
<td>70%</td>
</tr>
<tr>
<td>c1335</td>
<td>80%</td>
</tr>
<tr>
<td>c1908</td>
<td>90%</td>
</tr>
<tr>
<td>c2610</td>
<td>100%</td>
</tr>
<tr>
<td>c3540</td>
<td>110%</td>
</tr>
<tr>
<td>c5315</td>
<td>120%</td>
</tr>
<tr>
<td>c6288</td>
<td>130%</td>
</tr>
<tr>
<td>c7552</td>
<td>140%</td>
</tr>
</tbody>
</table>
DTCMOS vs. MG

Reduction of Leakage Currents

c432  c499  c880  c1335  c1908  c2610  c3540  c5315  c6288  c7552
5. Conclusion

- Subthreshold current and gate oxide leakage dominate leakage power
- Mixed Gates (MG) combine advantages of DTCMOS and DToCMOS at transistor and gate level
- Average 65% (vs. LVT) and 18% (vs. DTCMOS) leakage reduction at constant delay
Thank you!

Contact:
email: frank.sill@uni-rostock.de
Tel.: +49 381 4987278