Title
Energy Savings and Performance Improvement in Subthreshold Using Adaptive Body Bias

Permalink
https://escholarship.org/uc/item/1dq8892w

ISBN
978-1-4503-4972-7

Authors
Sankaranarayanan, Rajsaktish
Guthaus, Matthew R

Publication Date
2017-05-10

DOI
10.1145/3060403.3060421

Peer reviewed
Energy Savings and Performance Improvement in Subthreshold Using Adaptive Body Bias

Rajsaktish Sankaranarayanan
University of California Santa Cruz
rsankara@ucsc.edu

Matthew R. Guthaus
University of California Santa Cruz
mrg@ucsc.edu

ABSTRACT
In subthreshold operation, circuits are more sensitive to the impact of parametric variation due to reduced supply voltages. To meet timing specification and ensure reliable operation, circuits require compensation techniques that mitigate variation. We developed a design methodology to use adaptive forward body bias and reduce worst-case 3σ active energy, delay and standby energy caused by threshold voltage variation. We validated this methodology on the ISCAS85 benchmarks and improved the worst-case metrics in each case, with no loss of performance. Our approach reduces worst-case standby energy and worst-case active energy by up to 21.06% and 18.80%, respectively, on average.

1. INTRODUCTION
Designing circuits for subthreshold operation is challenging as the impact of process variation is more significant at reduced voltages due to the exponential dependence of drain current on gate voltage. Variation in analog circuits like current mirrors and voltage reference circuits is minimized by using large geometry devices. In subthreshold digital design, however, minimum sized devices are optimal for energy efficiency [3]. For a low target frequency, body biasing was more energy efficient than supply voltage scaling in mitigating process and temperature variation [9]. Reverse bias was found to worsen drain current mismatch and forward bias reduced the mismatch [5]. Gate-level clustering with cluster specific body bias was found to improve leakage power [7]. However, these techniques do not adapt to variation, which could limit their effectiveness. Threshold voltage variation could affect transistor delays on critical paths, which in turn affects active energy, circuit performance and thus overall energy efficiency. Critical path replication [4] or approaches like block-based and path-based statistical analyses [10] have been used to measure this. To accurately determine the impact of variation, the exact critical path and its delay need to be estimated.

In this work, we reduce standby energy caused by threshold voltage variation. We developed a design methodology to use forward bias with a body bias regulator [1]. We use a previous regulator, but, develop a digital design methodology to mitigate the impact of variation. This paper presents the methodology of: finding the optimal number of regulators, determining their placement and assigning cells to be biased by these regulators.

The rest of this paper is organized as follows: We present an overview of proposed approach along with a motivational example in Section 2. In Section 3, we describe details of our design methodology and LP formulation. We discuss experimental methods including variation model, static timing analysis and energy measurement in Section 4. We present results in Section 5 and conclusions in Section 6.

2. METHODOLOGY OVERVIEW
We use the regulator circuit shown in Figure 1 which has an NMOS and PMOS transistor in diode-connected mode that are connected at the drains with their wells shorted. This improves device matching in a circuit (an inverter in this case) by producing an output voltage at which the leakage currents match. Even small variations in the doping profile of FETs can lead to exponentially mismatched drain currents. By shorting the wells, a pathway for the minority carriers exists between wells.

2.1 Variation-aware Adaptive Regulation
Consider an inverter with equal rise and fall times, biased by an equal sized regulator as shown in Figure 1. The regulator is designed to output a voltage of \( V_{dd}/2 \) since the two FETs in cutoff act like a voltage divider. In the presence of threshold voltage variation, the actual regulator output voltage will be higher or lower than \( V_{dd}/2 \) depending on the specific FET threshold voltages.

We assume the threshold voltages of NFETs (and PFETs) of both circuits to be normally distributed around the nominal value. According to Pelgrom's model the variance of the parameter mismatch between transistors is directly proportional to the distance between them and inversely proportional to the area of the devices [8]. Given the distance between two transistors, variance of threshold voltage mismatch can be computed using Pelgrom's model. Using variance of mismatch, variance of threshold voltage distribution and uncorrelated random samples from this distribution, we determine the correlation between random variables.

In Figure 2, we compare the worst-case 3σ standby energy of an inverter biased by the regulator, and an inverter biased by a voltage source at \( V_{dd}/2 \). Well-bias and offsets between wells can have many possible combinations to meet target performance and power. Instead, a simple way to mitigate...
impact of variation is biasing both wells by the same amount. For this work, we use $V_{dd}/2$ because, the substrate current is much smaller than drain current and thus scalable to deep subthreshold voltages. In the case of the inverter biased by the regulator, we see the inverter worst-case standby energy decreases over a range of correlation values which changes over distance between inverter and regulator.

3. ON-CHIP REGULATOR METHODOLOGY

Standard cell design flows use a set of pre-designed libraries of cells. From Figure 2, we see savings in worst-case 3σ energy when the regulator is nearby and thus correlates better with the cell. The regulator circuit consumes standby energy and incurs area. Our methodology makes a trade-off in the number of regulators and the performance of the circuit. We target filler cell locations for regulator insertion.

3.1 Regulator Design

We use the Nangate Open Cell Library implemented using 45nm FreePDK process technology. The library cells were redesigned for equal rise and fall times in the worst case switching condition. The impact of variation can vary between a cell and its associated regulator, which in turn could affect the leakage current match and hence energy savings. For improved matching, the transistor dimensions of the regulator need be similar to that of the library cells. So, we custom designed a set of regulators based on the functionality and transistor dimensions of the library cells.

3.2 Cell Characterization

The worst-case standby energy of the inverter follows a linear trend over distance between inverter and regulator. So, to use this methodology in a standard cell based flow targeting worst-case energy savings, we characterized all cells in Nangate Open Cell Library for power and performance using cell specific regulators and with the variation model that was used to characterize the inverter circuit described earlier. Using the characterization data, we fit linear models for the worst-case standby energy savings of each cell in the library, as a function of distance.

3.3 Design Implementation

We developed a CAD flow with industry standard tools for synthesis and placement of standard cells and a separate regulator placement and assignment tool. The regulator assignment is formulated as a Linear Program (LP), which is described in the next section. In Figure 3, we show the CAD flow of our proposed design and verification methodology. Starting with Verilog description of the benchmarks, we synthesized them using Synopsys Design Compiler and Nangate Open Cell Library to obtain gate level netlists. These netlists were placed using Synopsys IC Compiler. The placement contains filler cells in the unused areas. Our LP optimizer reads in the placed netlists containing geometric coordinates of the cell and filler instances. Because the regulators consume standby power, we do not want to replace all of the filler cells, but they are candidate locations for regulator instances. Our optimizer solves for the optimal clustering of cells biased by regulators, using LP Solve [2].

3.4 Regulator Placement and Assignment

Insertion of each regulator incurs a standby energy cost. To achieve worst-case energy savings at the circuit level we determine the optimal number of regulators, the placement of these regulators and the cell(s) assigned to a regulator.

We model the problem of assigning regulators to one or more groups of standard cells as an LP. We demonstrate two algorithms to solve this LP problem. The first is an exhaustive approach yielding the optimal solution, while the second is a faster, heuristic solution. The two approaches differ in the number of LP constraints formulated and how they generate the constraints.

In the optimal solution, we consider all filler cells as candidate locations for regulators, each of which can be paired with any cell on chip.

In the heuristic approach, we derive candidate subsets of filler locations from the set of all filler locations, for regulator insertion. We then proceed to determine the regulators to make clusters with cell instances. For these subsets, we consider only rows immediately adjacent to a cell, for each cell in the design. Using this row adjacency constraint, we build LP constraints and attempt to solve the LP. If the LP does not converge, we increase the set of rows from which we derive the subset of filler cells, for each cell in the design. Again we attempt to solve the LP. The goal of the heuristic is to determine the smallest number of rows from which the subsets can be derived leading to a solvable LP. The goal of the LP formulation is to identify a subset of regulators from the available regulators and determine the grouping of standard cells to be assigned to those selected regulators.

The LP formulations of both algorithms have identical cost functions and differ only in the number of constraints. So, for brevity we present the equations of the optimal formulation and indicate the areas where they differ.

Both algorithms have similar inputs, namely, placed file containing fillers and cells and the cell-specific characterization data. The output from both algorithms is the gate clustering and regulator assignment. For brevity, we present the heuristic and indicate where it differs from the optimal solution. Lines 2.5 and 6 constitute the iterative part of the heuristic solution only and are not applied in the optimal solution. All other aspects are common to both algorithms.

To physically implement our methodology we leverage features available in place and route tools. We first create relative placement groups for each cell cluster connected to a regulator and associate a power domain with that group. We then designate the output net of each regulator as a supply net and connect those nets to cell bias pins.

- Let $m$ be the number of cell instances in the design.
- Let $i$ be an index variable such that $i = 0, 1, \ldots, (m-1)$.
- Let $C$ represent the set of all cells indicated by $C_i$, where $i = 0, 1, \ldots, (m-1)$.
- Let $n$ represent the number of regulators in the design.
- Let $j$ be an index variable such that $j = 0, 1, \ldots, (n-1)$.
Our formulation is described below:

\[ \text{Minimize } \sum_{i=1}^{n} \sum_{j=1}^{m} Y_{ij} + m \cdot a \quad \text{such that} \]
\[ \sum_{j=1}^{m} c_{i} X_{i,j} \leq Y_{i} \quad \forall i = 0, 1, 2 \ldots (m-1) \]
\[ \sum_{j=1}^{m} X_{i,j} = 1 \quad \forall i = 0, 1, 2 \ldots (m-1) \]
\[ X_{p} = \{0, 1\} \quad \forall p = 1, 2 \ldots (m \cdot n) \]

The LP solver solves for decision variables \( X_{ij} \) which determines the optimal grouping of cells biased by regulators. Our formulation is described below:

\[ \text{Minimize } \sum_{i=1}^{n} \sum_{j=1}^{m} Y_{ij} + m \cdot a \quad \text{such that} \]
\[ \sum_{j=1}^{m} c_{i} X_{i,j} \leq Y_{i} \quad \forall i = 0, 1, 2 \ldots (m-1) \]
\[ \sum_{j=1}^{m} X_{i,j} = 1 \quad \forall i = 0, 1, 2 \ldots (m-1) \]
\[ X_{p} = \{0, 1\} \quad \forall p = 1, 2 \ldots (m \cdot n) \]

The cost function in Eq.1 describes the goal of this formulation, which is to minimize the number of clusters into which all standard cells can be grouped, such that each cluster is connected to a regulator. Here \( Y_{i} \) denotes the linearized worst-case 3\( \sigma \) standby energy cost of a cell \( i \) from amongst the set of all cell instances \( C \), when biased by a regulator. Summed over the set of all cell instances, our goal is to minimize the cost of clustering all the cell instances in the design. The second term indicates the standby energy cost of the regulators needed to cluster all the cell instances. The set of constraints denoted by Eq.2 determine which of the cells get clustered together to be biased by a common regulator and get assigned a regulator. The coefficient term is the energy cost of driving a particular cell with a particular regulator. This term is obtained by precharacterizing the cells of the standard cell library and the distance between the regulator and the cell. The LHS of constraint Eq.2 represents the energy cost of biasing a given standard cell by each of the available regulators. The RHS of constraint Eq.2 ensures this cell is biased by one of the available regulators only. The constraint Eq.3 ensures each cell has to be driven by a regulator and enables grouping of cells to a common regulator. The constraint Eq.4 indicates \( X_{ij} \) is a binary decision variable taking values \( \{0,1\} \).

The above described constraints apply to the Algorithm-1. For Algorithm-2, the indices of the variables and the limits of the summation are not constant \( n \) and instead take variable values based on the number of regulator candidates available in the rows adjacent to each cell.
critical path delay, we then simulate the circuit at $f_{\text{max}}$ using random input stimuli and measure active and standby energy per cycle using Synopsys Hspice. We perform this process of applying variation, determining critical path, finding $f_{\text{max}}$ and using it to compute energy using 1000 iterations of Monte Carlo simulation. We then compare the benchmark performance with an unbiased circuit and a circuit biased using a voltage source at $V_{\text{dd}}/2$, described earlier in Section 2, comparable to an off-chip bias source. Since the regulators are connected to on-chip supply voltage, our results include the energy overhead of the regulators.

5. RESULTS

In this section, we present the results of our optimization process and the impact of optimization driven body bias on energy and performance of ISCAS85 benchmarks.

The run time of our optimization process was measured for several benchmarks using the optimal and heuristic approaches. Across the benchmarks, the improvement in run-time of the heuristic solution, was 68% on average. This is due to the reduced subset of candidate regulators considered for each cell by our optimizer.

We found the active energy, standby energy and delay following log-normal distributions. So, we evaluate the distributions using typical parameters namely mean ($\mu$) and worst-case $3\sigma$. The regulator bias method provides improvement by delivering a forward bias which is adapted to the cell variation in that spatial vicinity. This locally relevant forward bias offers a better matching of the FET off-currents and lowers the cell threshold voltage, thus improving its performance. This increased performance enables scaling the circuit to lowered supply voltages, thus saving worst-case active energy and standby energy.

Since each regulator could output a bias voltage different from another regulator, we consider inter-well spacing for wells at different potential to measure the impact on circuit area. This area overhead spans a range of 17.5% at the minimum to 24.5% at the maximum across all benchmarks with an average of 19.4%.

From Figure 4, we see the regulator method offers improvement in worst-case delay of the circuit for all benchmarks compared to an unbiased circuit. The improvements decrease as the supply voltage is scaled down. This is because, at lower operating voltages, the impact of variation is much higher than the applied bias compensation. The improvement in delay spans a range of 10.89% at the minimum to 50.16% at the maximum. Considering all benchmarks, the average improvement in worst-case delay is 36.93%, 27.85% and 12.74% at 350mV, 300mV and 250mV respectively using Algorithm-2, compared with an unbiased circuit.

From Figure 5, we see the regulator method offers savings in worst-case active energy of the circuit for all benchmarks when compared against an unbiased circuit and an offchip biased circuit. We have verified this using both our algorithms. The active energy savings on average compared to an unbiased circuit across all benchmarks include 14.52% and 4.50% for Algorithm-1 and Algorithm-2 respectively. Compared to an offchip biased circuit, the savings are 18.84% and 9.20% using Algorithm-1 and Algorithm-2 respectively. We can see Algorithm-1 slightly outperforms Algorithm-2 in both cases of comparison, namely against an unbiased circuit and offchip biased circuit. This is because, the optimal solution algorithm offers a wider choice of regulators to choose from, resulting in improved savings.

In Table 1 we list the savings in worst-case $3\sigma$ standby energy using our methodology. We compare this against unbiased circuits and with offchip biased circuits using both algorithms. Algorithm-1 offers savings across all benchmarks, while Algorithm-2 offers savings in some cases. In other cases there is a slight increase in the worst-case standby energy. This corresponds to the cases where the active energy savings are also at the lowest, due to the reduced

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>c432</td>
<td>0.92</td>
<td>0.69</td>
<td>41.90</td>
<td>0.87</td>
<td>5.70</td>
<td>1.04</td>
<td>0.60</td>
<td>42.30</td>
</tr>
<tr>
<td>c499</td>
<td>1.15</td>
<td>0.98</td>
<td>1.64</td>
<td>1.24</td>
<td>-8.13</td>
<td>1.32</td>
<td>0.98</td>
<td>25.75</td>
</tr>
<tr>
<td>c353</td>
<td>1.16</td>
<td>0.90</td>
<td>3.30</td>
<td>1.26</td>
<td>-8.61</td>
<td>1.29</td>
<td>0.90</td>
<td>30.24</td>
</tr>
<tr>
<td>c1908</td>
<td>1.67</td>
<td>1.35</td>
<td>12.52</td>
<td>1.65</td>
<td>1.41</td>
<td>1.74</td>
<td>1.35</td>
<td>22.41</td>
</tr>
<tr>
<td>c2670</td>
<td>2.00</td>
<td>1.96</td>
<td>4.07</td>
<td>1.98</td>
<td>1.15</td>
<td>2.04</td>
<td>1.96</td>
<td>3.92</td>
</tr>
<tr>
<td>c354</td>
<td>1.94</td>
<td>1.80</td>
<td>5.21</td>
<td>1.95</td>
<td>-0.82</td>
<td>2.04</td>
<td>1.80</td>
<td>11.76</td>
</tr>
<tr>
<td>c5315</td>
<td>2.34</td>
<td>2.16</td>
<td>3.15</td>
<td>2.36</td>
<td>-0.57</td>
<td>2.43</td>
<td>2.16</td>
<td>21.06</td>
</tr>
<tr>
<td>Avg</td>
<td>10.26</td>
<td>9.41</td>
<td>-8.13</td>
<td>19.4%</td>
<td>-8.61</td>
<td>2.16</td>
<td>1.24</td>
<td>36.93%</td>
</tr>
</tbody>
</table>

6. CONCLUSION

In this work we developed a design methodology to use forward bias and reduced the impact of threshold voltage variation. We formulated optimal assignment of forward bias as a LP optimization in this methodology. We reduce worst-case standby energy and active energy by up to 21.06% and 18.80% on average respectively and reduce worst-case delay by up to 37.8% on average. The improvements are available over a range of deep subthreshold voltages.

7. REFERENCES