

1/5 Power Reduction by Global Optimization based on Fine-Grained Body Biasing

Yasumi Nakamura, David Levacq, Limin Xiao, Takuya Minakawa, Taro Niiyama, Makoto Takamiya, and Takayasu Sakurai

The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan

Abstract-A fine-grained body bias control to compensate both the process and design variations is proposed. A test chip was fabricated in 90nm CMOS process. The proposed global optimization scheme reduced power by 23% compared with an as-fabricated chip power and by 11% compared with the power optimized by the conventional local optimization approach. Also, the proposed global optimization scheme reduced power by 19% compared with an as-fabricated chip power within 20 test iterations with simulated annealing algorithm.

I. INTRODUCTION

The adaptive substrate bias control and the post-fabrication tuning of parameters [1-4] are the recent trend in designing power-efficient LSI's to cope with the increasing random device variability. The systematic design variation, which is the error between the simulation results at a design stage and the measured results of a fabricated chip, is also getting an important issue. This is partly because the chip is getting more and more complicated and delay estimation gets inherently more difficult and partly because the recent process introduces systematic delay deviation from designed value by new phenomena such as stress-induced drain current variation and imperfect optical proximity correction.

In the post-fabrication tuning, the parameters can be either locally or globally optimized. [4] shows a local optimization example of the body bias by monitoring critical path replicas in the 21 domains. Local optimization means that the parameter is locally optimized by looking at the local value of the parameter and modifying it to the designed value. This approach, however, cannot compensate the systematic variations and has the mismatch between the critical path replica and the real critical path. This approach may not work

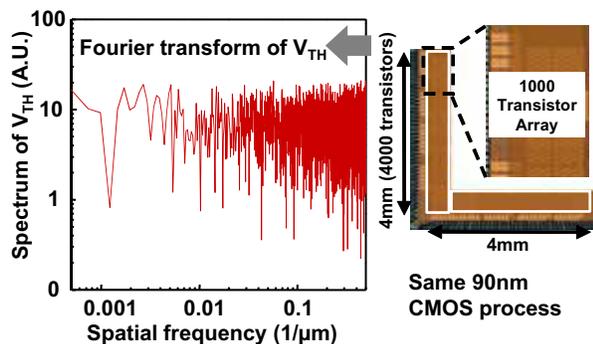


Fig. 1. Spectrum of within-die process variation

for cases where short-range correlated variation is small as is shown in Fig.1 which shows that there is no specific peak in low spatial frequencies.

In contrast, [1-3] show the global optimization of the clock skew in the 52 clock domains [1]. In this approach, however, the required tunable skew circuits consume power and the skew tuning does not reduce the leakage of the circuits.

In this paper, a fine-grained body bias control to compensate both the random (process) and systematic (design) variations is proposed and the globally minimized power at a constant performance is investigated through measurements.

II. LOCAL OPTIMIZATION VS GLOBAL OPTIMIZATION

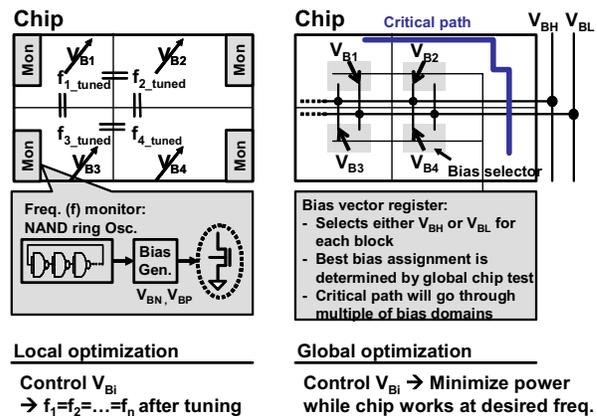


Fig.2. Local optimization (Left) and global optimization (Right).

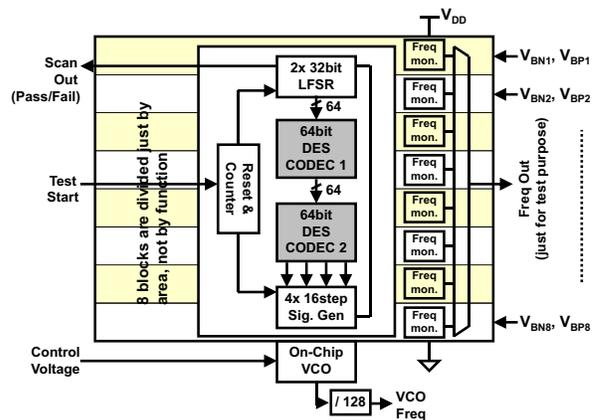


Fig. 3. Block diagram of the fabricated test chip.

There are two approaches to determine the body bias values for each biasing region (Fig.2). One is the local optimization scheme, which aims at making V_{TH} of all the regions is equal to compensate within-die V_{TH} variations [4].

The other is the proposed global optimization scheme, which aims at achieving the lowest power consumption while the real critical path can operate at the desired frequency. Real

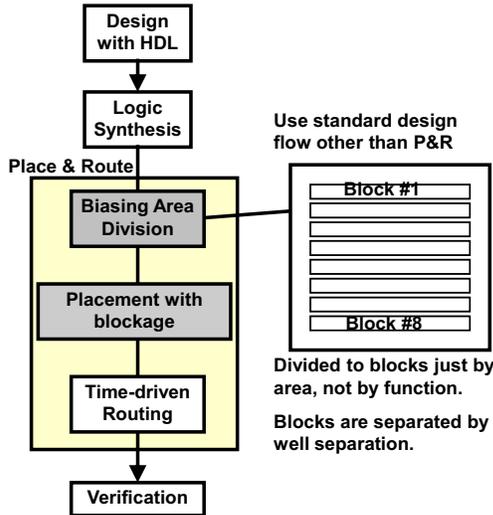


Fig. 4. Design flow of the test chip

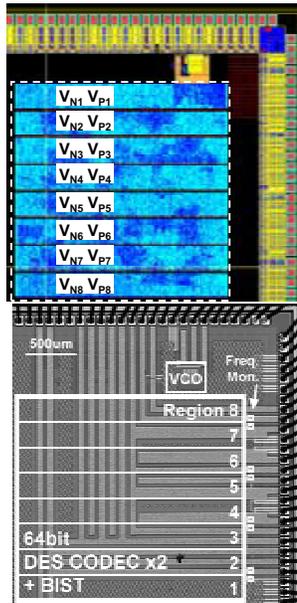


Fig.5. Chip layout (up) and micrograph of fabricated chip (down)

Table 1 Summary of the fabricated chip

Process	1V 90nm CMOS
Area	2400um x 2400um
# of gates	1M gates
Test	BIST (Built-In Self-Test) with 4-bit 16step signature generator

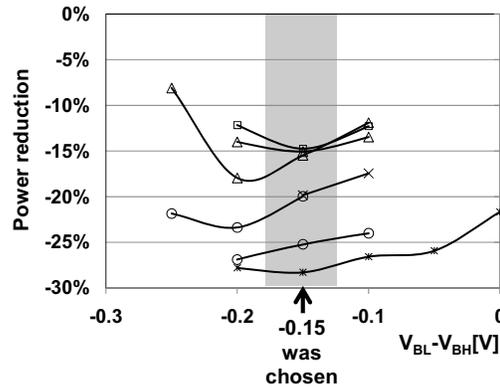


Fig. 6. Measured $V_{BH}-V_{BL}$ – power dependence for 6 chips. This is for PMOS well bias and similar optimization is done for NMOS.

critical paths are tested at a chip level. The global optimization can compensate not only the process variation but also the systematic design variation. This can be used as a method for post-fabrication clock skew tuning (time borrowing) without introducing parametric delay component which is large.

To evaluate the effect of the fine-grain global body bias optimization, a test chip was fabricated in 1V 90nm CMOS and measured. Fig. 3 shows the block diagram of the test chip. The circuit under test is two series-connected 64bit DES CODECs, which is driven by 2 32bit LFSR input vector generator, and 4 out of 64bit output are compacted with 16step signature generator. The chip has 8 body bias domains and each region has a 31-stage FO3 NAND ring oscillator divided by 64 as a frequency monitor. Test is carried out with a PC with 16ch D/A board which generate the body bias voltages.

The design flow is shown in Fig.3. The only difference between a normal digital circuit and the proposed circuit is to divide the chip into multiple body biasing domains. In the test chip, this division is done just by area and not by function. This means that any division can be made without considering the functional borders and can be applied to any chip.

The micrograph and layout of the fabricated chip is shown in Fig.5. As shown in Table 1, it occupies 2400um x 2400um in area.

III. METHODS FOR DETERMINING V_B VECTOR

Since each body bias voltage (V_{BP1-8} for PMOS and V_{BN1-8} for NMOS) can take any analog value, finding the best value to obtain the lowest power out of millions of patterns is difficult. In this paper, each body bias voltage is limited to “high” or “low”, namely V_{BH} and V_{BL} . V_{BH} is set to the lowest value where the chip can operate when $V_{BN1}=V_{BN2}=\dots=V_{BN8}=V_{BP1}=\dots=V_{BP8}=V_{BH}$.

V_{BL} is then set to $V_{BH}-0.15V$. Fig. 5 shows the reason, the $V_{BL}-V_{BH}$ – Power reduction ratio dependence for 6 chips. Here, power reduction ratio = 0 when the power consumption is same as with the worst case V_{BH} or the V_{BH} where all the chips can operate. Power reduction ratio shows a gentle minimum between -0.2V and -0.1V for various chips.

Three methods are tried to determine the V_B vector, exhaustive search, best vector LUT and simulated annealing.

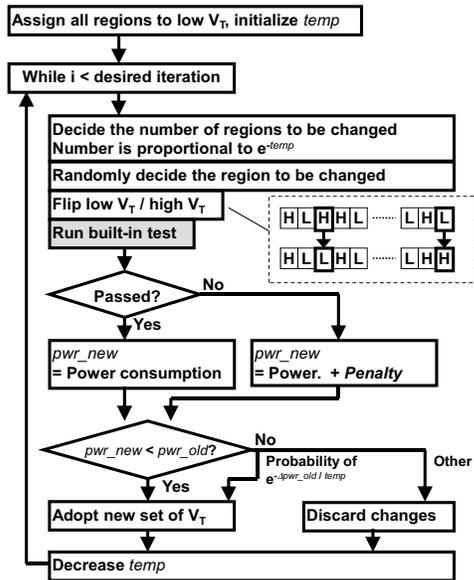


Fig. 6. Simulated annealing algorithm applied to the test chip

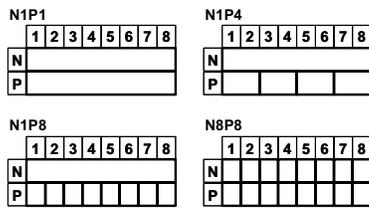


Fig. 7. Bias domain division and its name

A. Exhaustive Enumeration

The simplest way to find the best vector to reduce power is to test all possible vectors. It is possible for small number of regions, though, even for 8 regions or 16 parameters (V_{BN} , V_{BP} x 8), 65536 tests are required, which is not practical if the number of domains are large.

B. Best Vector LUT

An alternative method is to apply the exhaustive test (or long-series simulated annealing) in the development stage and find out the best vector look-up table (LUT) for each V_{TN} , V_{TP} . This will take time in the development stage but once the LUT is established, there is not test time overhead in each die test.

C. Simulated Annealing

Simulated annealing is an algorithm to find a global minimum value in large search space by stochastic approach. To apply simulated annealing to this bias vector optimization, the algorithm shown in Fig. 6 is adopted. The key point here is to add a “penalty constant” to the consumed power if the circuit fails since if the circuit fails to operate at the desired frequency, the bias vector is “very bad” even if the power is low. With the introduction of the penalty, the search becomes a simple bound-free minimum search.

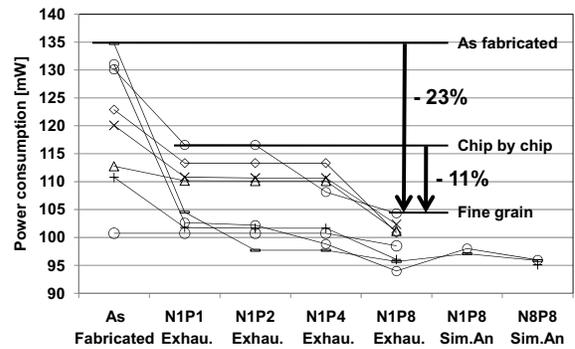


Fig. 8. Measured relationship between grain size and power consumption

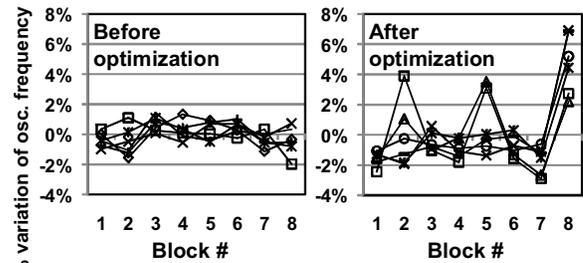


Fig. 9. NAND ring monitor frequency for each bias block. Each line corresponds to a chip

The difference among these three methods are shown in the following section.

IV. MEASUREMENT RESULTS

In this section, the word “NiPj” denotes that there are i bias domains for NMOS and j bias domains for PMOS as shown in Fig. 7.

First, the relationship between grain size and power consumption is shown in Fig. 8(a). Since the within-die process variation is small and random (see Fig. 9), the same bias is happened to be set to all the bias domains. From the figure, it is clear that the finer the grain size is, the lower the power consumption becomes. N1P8 bias control is shown to reduce 23% of the power consumption compared with as-fabricated chip without post-tuning.

NAND ring oscillation frequency at the lowest power consumption in N1P8 exhaustive search is shown in Fig. 9. From this, it is seen that within-die variations are below 3% while design variations goes up to 10% which is seen from the figure at right hand side.

The measured relationship between number of test iterations and the power reduction in the proposed simulated annealing method for one chip is shown in Fig. 10. Within 20 iterations, more than 19% power reduction is achieved.

Fig. 11 shows the comparison among the three methods for N1P8 parameter set. The power reduction efficiency of one certain chip is compared the methods themselves. Exhaustive search of course shows the best result of 29% power reduction but it may not be practical in cases because the test time increase is considerably increased. On the other hand, by using

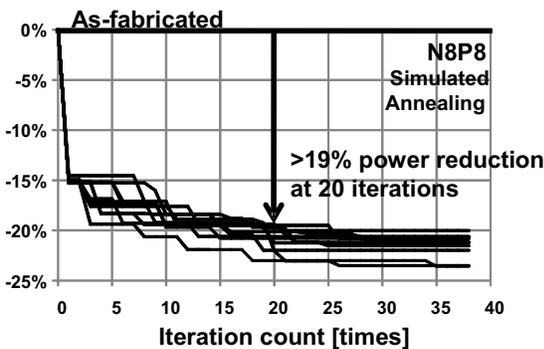


Fig. 10. Measured iteration - power characteristics in simulated annealing

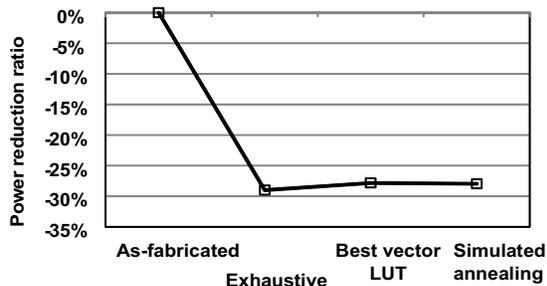


Fig. 11. Comparison among exhaustive search, LUT, and simulated annealing (NIP8)

simulated annealing, the shorter test time is expected and the power reduction ratio stays the same as 28%. As to the best vector LUT approach, the power reduction ratio is 28% and in this look-up table approach, no test time overhead is needed since the time-consuming optimization efforts are made only once in the development stage. In measuring this value, the best vector is not the best vector for the specific chip but it is the best vector for the chip set.

IV. DISCUSSIONS

In order to divide a logic circuit into multiple bias blocks, extra area for well separation is required. Fig. 12 shows the relationship between the number of division and the area overhead caused by this well separation. In the fabricated process, this overhead does not exceed 5% until 12 divisions for 5mm x 5mm chip.

V. CONCLUSION

The fine-grained body bias control to compensate both the random (process) and systematic (design) variations was proposed and the effectiveness was demonstrated with the 90 nm CMOS test chips. The proposed scheme compensates die-to-die V_{TH} variation and the systematic design variations. Undesired inequality of critical path delay among pipeline stages are compensated in this scheme. Compared with as-fabricated chip, proposed global optimization approach reduces the power by 23% and by 11% compared with the

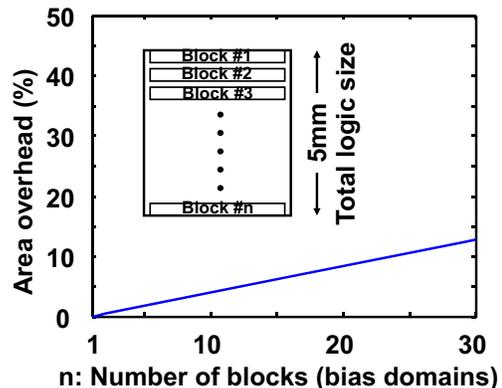


Fig. 12. Number of blocks vs area overhead (calculated)

power optimized by the conventional local optimization approach. Also, the proposed global optimization scheme reduced power by 19% compared with as-fabricated chip power within 20 test iterations with simulated annealing algorithm. The best vector LUT approach is also practical. The proposed schemes are considered to be promising for achieving the power-efficient LSI's in scaled devices with reasonable area and test overhead.

ACKNOWLEDGMENTS

This work is partially supported by MEXT and Hitachi, Ltd. The VLSI chips were fabricated through the chip fabrication program of VLSI Design and Education Center (VDEC), the University of Tokyo, with the collaboration by STARC, Fujitsu Limited, Matsushita Electric Industrial Company Limited., NEC Electronics Corporation, Renesas Technology Corporation, and Toshiba Corporation.

REFERENCES

- [1] D. Deleagnes, J. Douglas, B. Kommandur, and M. Patyra, " Designing a 3GHz, 130nm, Pentium® 4 processor," IEEE Symp. on VLSI Circuits, pp. 130-133, June 2002.
- [2] E. Takahashi, Y. Kasai, M. Murakawa, and T. Higuchi, "A post-Silicon clock timing adjustment using genetic algorithms," IEEE Symp. on VLSI Circuits, pp. 13-16, June 2003.
- [3] J. Friedrich, B. McCredie, N. James, B. Huott, B. Curran, E. Fluhr, G. Mittal, E. Chan, Y. Chan, D. Plass, S. Chu, H. Le, L. Clark, J. Ripley, S. Taylor, J. Dilullo, and M. Lanzerotti, "Design of the Power6 microprocessor," ISSCC, pp. 96-97, Feb. 2007.
- [4] J. Tschanz, J. Kao, S. Narendra, R. Nair, D. Antoniadis, A. Chandrakasan, and V. De, "Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage," ISSCC, pp. 422-423, Feb. 2002.