

# A Yield and Speed Enhancement Scheme under Within-die Variations on 90nm LUT Array

Kazuya Katsuki, Manabu Kotani, Kazutoshi Kobayashi and Hidetoshi Onodera  
 Graduate School of Informatics, Kyoto University, Kyoto, Japan.  
 {katsuki, kotani, kobayasi, onodera}@vlsi.kuee.kyoto-u.ac.jp

**Abstract**—In this paper, we propose a yield and speed enhancement scheme using a reconfigurable device. An LUT array LSI is fabricated on a 90nm process to measure process variations of LUTs. D2D and WID variations are clearly observed. Reconfiguration using the measurement process variations boosts yield and also increases the average operating speed by 4.1%. In addition, it is proved that expansion of WID variations make the proposed method more effective.

## I. INTRODUCTION

Process scaling makes it possible to integrate billions of transistors on a single die. It is quite difficult to manufacture such small transistors with similar characteristics. Scaling down increases variations of transistor performance. In addition, high- $k$  materials are said to affect variations seriously[1]. Transistor performances are different die-to-die (D2D) and also within-die (WID). [2] reveals that WID variations are apparently observed in a 90nm process, which become dominant according to the process scaling[3], [4].

Degradations of transistor performance by variations impacts gate delay. If the performance of transistors along some critical path becomes worse, a fabricated chip does not work correctly at the required speed. Finally it causes yield loss of ASICs. To avoid the loss we design implemented circuits with large amount of timing margins, but increase of the margin results in increase of power consumption and expansion of the circuit area.

In this paper, we propose a yield and speed enhancement scheme against WID variations using the reconfigurable architecture. On reconfigurable devices, functions are allocated after manufacturing. In conventional reconfigurable hardware, each reconfigurable functional block is considered to have the same performance. In the proposal method we measure WID variations after manufacturing, and allocate the functions suitably considering the measurement results. It enhances not only yield and speed, but also narrows timing margins.

Section II explains the principle of the proposed method in details. An LUT array to measure process variations is fabricated in a 90nm process. Its structure and measurement results are shown in section III. Section IV shows experimental results of yield and speed enhancement by the proposed method considering measurement results. Finally conclusion are written in section V.

## II. PRINCIPLE OF THE SPEED AND YIELD ENHANCEMENT

There are many candidates of critical paths in a chip. If even one of these paths become slower by within-die (WID) variations, the fabricated chip become slower. Therefore WID variations uniformly degrade the performance of LSIs. That possibility increases as the transistor count. As the result, almost all fabricated chip may be slower and the yield drops significantly if WID variations become dominant.

The reconfigurable structure mitigates these performance degradation. Figure 1 shows the fundamental idea of the proposed speed and yield enhancement scheme compared with conventional cell-based fixed-structured ASICs. First of all, we fabricate a reconfigurable devices. WID variations are measured on every chip. Forecasting the variations is very difficult, but measuring them after manufacturing is much easier. When reconfiguration, functional blocks are placed according to the measured variations of each chip and their lengths of critical paths. That is to say, a functional block with longer critical path is allocated in the area of faster transistors and a block with shorter critical path is in the area of slower transistors. This method enhances not only yield and the maximum operating clock frequency of fabricated chips, but also slashes the timing margin by leveraging WID variations rather than compensating them.

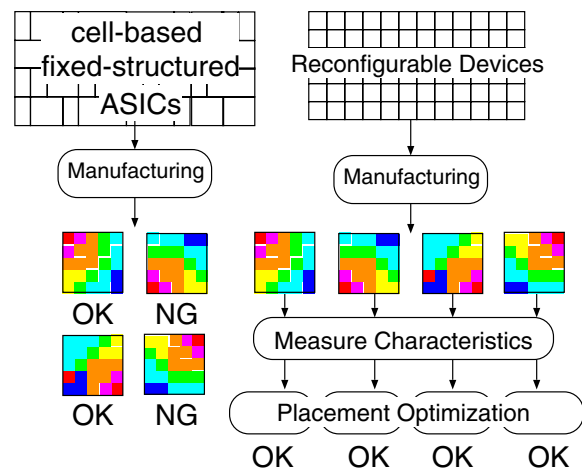


Fig. 1. Fundamental idea of the proposed yield enhancement scheme. Left: conventional cell-based fixed-structured ASICs. Right: proposed reconfigurable devices configurations of which are optimized as measured within-die variations.

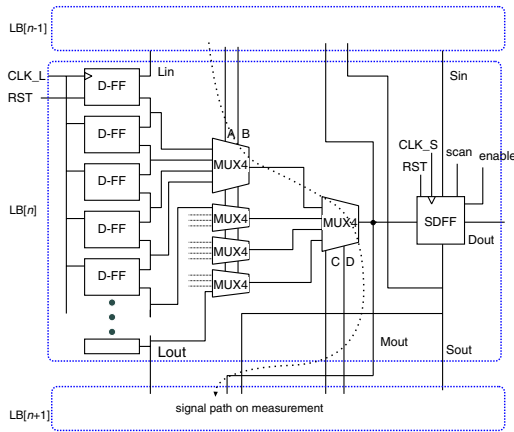


Fig. 2. Structure of a logic block. A signal is transmitted along the dashed arrow through two MUX4s per LB at the measurement.

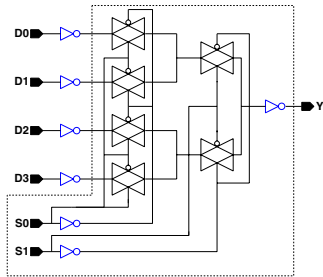


Fig. 3. Schematic diagram of an MUX4. Within-die variations of the region inside the dashed line can be detected by measurement.

### III. AN LUT ARRAY FABRICATED ON A 90NM PROCESS

We have fabricated an LUT array on a 90nm process to measure process variations of a reconfigure structure. Generally speaking, regular structures such as FPGAs or SRAMs have small process variations than irregular structures like ASICs or microprocessors. But in the upcoming nanometer technologies, such regular structures will have distinguished WID variations.

#### A. Structure of the LUT array

Fig. 2 shows the structure of a logic block (LB) which contains a 4-bit LUT and a scan flip-flop (SDFF). An LUT consists of 16 flip-flops to store an LUT configuration and five MUX4s (4-input multiplexers). Fig. 3 shows the schematic diagram of a MUX4. We can measure process variations of MUX4s in the region inside the dashed line. The output signal **Mout** from the MUX4 is sent to the adjacent LUT. Fig. 4 shows the array structure of logic blocks in the fabricated chip. They are laid out in a fractal structure to observe scalable process variations. If they are laid out in a line, WID variations may be canceled. The fractal structure makes it possible to measure WID variations in scalable square regions.

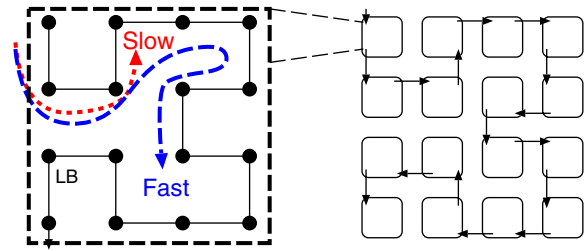


Fig. 4. Structure of the LUT array. LBs are connected in a fractal structure to observe scalable process variations.

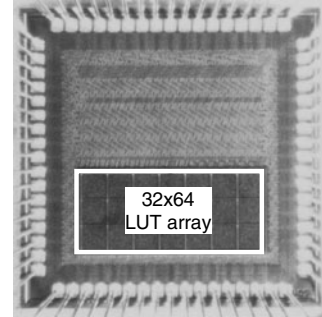


Fig. 5. Chip micrograph of a 90nm LUT array LSI including 2,048 logic blocks located at the bottom.

#### B. Measurement method of process variations

On measuring the process variations, a signal is rushing through LUTs from the first LB in a square region, which is captured by the SDFF in each LB. LUTs are configured as follows during the measurement.

- The LUT in the first LB is configured to become true at any input value.
- The LUT in the second LB is configured to become true if the input **B** from the previous **Sout** becomes true.
- The LUTs in the other LBs are configured to become true only if the input **A** from **Mout** becomes true.

Applying a clock pulse to SDFFs under the above LUT configuration, **Sout** of the first LB becomes true, which is transmitted through LBs. During the transmission, let us apply another clock pulse to SDFFs. Then the SDFFs in the LBs where the true signal have been transmitted become true. If WID variations are observed, number of transmitted LBs will be different in each square region as shown in Fig. 4.

Fig. 5 shows a micrograph of a fabricated LSI and Table I shows its specifications.

#### C. Measurement results from fabricated chips.

We have measured 25 fabricated dice on a single 300mm wafer. Each die is 2.5mm×2.5mm and taken from a 20mm×25mm reticle. The location of each chip on the wafer and on the reticle is unknown. Fig. 6 shows the result of a measurement. Measurements are done at the resolution of 64 (8×8) LBs. It measures the performance by counting the number of transmitted LBs as in Fig. 4. Process variations

TABLE I  
SPECIFICATIONS OF THE FABRICATED LUT ARRAY CHIP.

Process	6M-IP 90nm CMOS
Wafer	300mm
Chip Size	2.5mm×2.5mm
# of Trs./LB	1,950
# of Trs. (Total)	1.56M
Area of the LUT array	.997mm <sup>2</sup>
Power Dissipation	2.45mW(@50MHz)

appear very small since the transistor speed is quantized as the number of LBs.

To avoid the quantization and measure the difference more clearly, varying clock cycle (time to transmission) from 4.0ns to 8.0ns at 0.1ns interval, measurements have been repeated 100 times per cycle. Total 4200 measurements are done per chip at the resolution of 16 (4×4) LBs. Then 100 results are averaged at every cycle. The obtained average value is regarded as the number of transmissions at the cycle. By setting the clock cycle on the horizontal axis and the average number of transmissions on the vertical axis, the gradient is calculated using the least square method. The gradient depends on the performance of each block of LBs. If its speed is doubled, the gradient is doubled. The ratio of the gradients is equivalent to the ratio of the speeds. We can regard these gradients as the performance indicator.

Fig. 7 shows the statistic of WID variations from a fabricated chip. The peripheral LBs tend to be fast and the central LBs are slow. The other 24 chips have the same tendency. The possible reasons of the concave delay curve are as follows.

- Systematic within-die variations because of the aberrations in the stepper lens etc.[3].
- Central portions degradation caused by IR drop.
- Distributions of the wire length.

It is unlikely that systematic within-die variations cause such a sharp tendency to all the chip in the same way since the chip size is very small. The area of the LUT array is about 1mm<sup>2</sup> as in Fig. 5. To distinguish the WID and D2D variations, statistics from the 25 dice are averaged for every 16 LBs. These averaged gradients are called the reference delays. The average value of the residual errors between the measured and reference delays on a die is regarded as the D2D process variation, which is shown in Fig. 8. Fig. 9 shows WID variations of the slowest, typical and fastest chips. Each distribution is obtained to subtract measured delays from the reference delays. The three distributions are very similar to the Gaussian distribution. Therefore the above residual-based method is practical to extract WID variations. The D2D variations can be compensated using well-known body-bias techniques, which are effective for D2D variations. However, it is difficult to compensate the entire WID variations, since the number of controllable body bias voltages is limited. Fig. 8 and 9 reveals that the WID and D2D variations have the same order in the 90nm process. Even if D2D process variations are compensated, WID variations may degrade the speed and yield in these conditions. In the

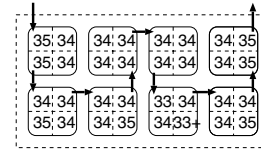


Fig. 6. The number of transmitted LBs from a measurement of a fabricated chip at the resolution of 64 (8×8) logic blocks (LBs). Each decimal number means the number of transmitted LBs at 20ns (clock frequency 50MHz). The + indicator means the number of transmitted LBs changes on each measurement.

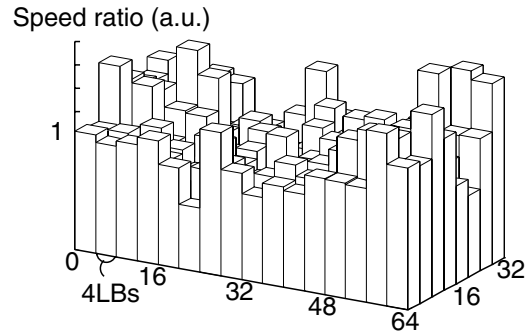


Fig. 7. Statistics of a fabricated dice by regarding the gradient from the least square method as the performance indicator. Peripheral LBs are fast and central ones are slow.

following section we evaluate speed and yield enhancement obtained by the proposed placement optimization according to these process variations.

#### IV. EXPERIMENTAL RESULTS OF YIELD AND SPEED ENHANCEMENT BY RECONFIGURATION

The main idea of the proposed scheme is to optimize placement of circuit blocks on reconfigurable devices according to the transistor performance. We evaluate speed and yield enhancement obtained from placement optimization compared with a conventional fixed placement.

Suppose that a reconfigurable device contains 2048 logic blocks, which is same number as the fabricated chips shown in section III. Each functional circuit block occupies 16 (4×4) logic blocks but has a different critical-path length one by one. The critical path lengths are quantized according

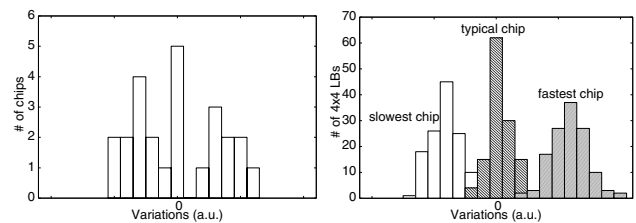


Fig. 8. Observed D2D variations, Fig. 9. Extracted WID variations which is obtained from the average values from three distinguished chips. Left: slowest, Center: typical, Right: values and averaged representative fastest. They are similar to the Gaussian distribution. The scale of the x axis is same as Fig. 8

TABLE II

YIELD AND SPEED ENHANCEMENT OBTAINED FROM ONE MILLION DICE.

allocation	WID: results of Sect. III		WID: tripled	
	Yield	Ave. Speed	Yield	Ave. Speed
Fixed	73.0%	1.010	54.3%	1.004
Opt (ave)	99.3%	1.040	66.3%	1.020
Opt (each)	100.0%	1.051	100.0%	1.095

to the normal distribution from 0 to 8. Critical path length 0 means that the 16 logic blocks is not used. The array at the left side in Fig. 10 shows these functional circuit blocks. The array at the right side in Fig. 10 shows a transistor performance distribution of 16 ( $4 \times 4$ ) logic blocks converted to acceptable critical path lengths. Performances are different on every chip because of WID variations and IR drop etc. Here we ignore D2D variations that can be compensated. We model the performance variations using the measurement result of section III. The reference delay is referred as the average performance. WID variations are modeled by Gaussian distributions estimated from Fig. 9.

Here, each functional circuit block is assigned to a 16 ( $4 \times 4$ ) logic block. If transistor performance of the logic blocks is below the critical path length of the assigned functional block, the entire chip does not work with the target operating speed. Its initial allocation can not be exchanged on the conventional FPGAs. On the other hand, we assume that they can be exchanged within a small region in the proposed scheme like Fig. 10. We suppose that  $4 \times 4$  functional blocks can be exchanged without overhead of wire length. These three schemes are compared

- 1) Fixed to the initial allocation.
- 2) Optimized within  $4 \times 4$  to suit to the reference delays. This is called "Opt (ave)".
- 3) Optimized within  $4 \times 4$  to each chip considering both the reference delays and WID variations. This is called "Opt (each)".

One million chips are generated according to the measured WID variations. Fig. 11 shows distributions of operating speed normalized by the target speed. Table II shows the yield and average chip speed obtained from the above three placement schemes. When WID variations are obtained from the measurement results in Sect. III, the yield becomes 99.3% by optimizing placement using the average variation (referred as Opt.(ave)). If WID variations are tripled, the yield is not increased by Opt.(ave). But, it is enhanced to 100% by optimizing placement chip by chip according to the chip-oriented WID variation (referred as Opt.(each)).

## V. CONCLUSION

We propose a yield and speed enhancement scheme using a reconfigurable device. An LUT array LSI is fabricated on a 90nm process to measure process variations of LUTs. D2D and WID variations are clearly observed on the fabricated chip.

Reconfiguration (placement optimization) according to the measured WID variations boosts yield and also increases the average operating speed by 4.1%. In addition, availability of

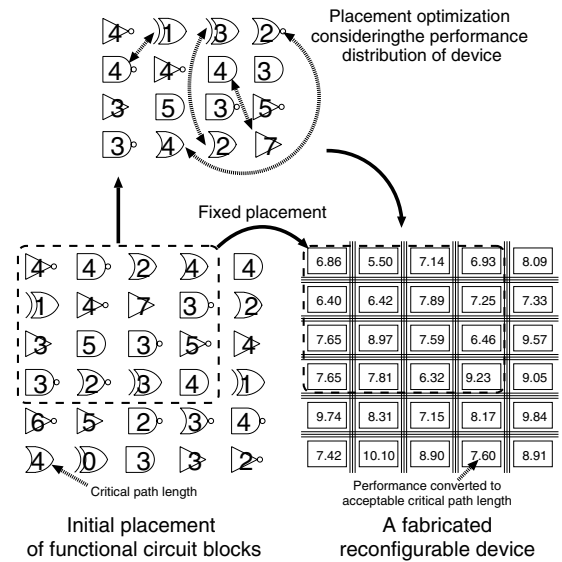


Fig. 10. Placement optimization scheme according to the performance of a fabricated chip.

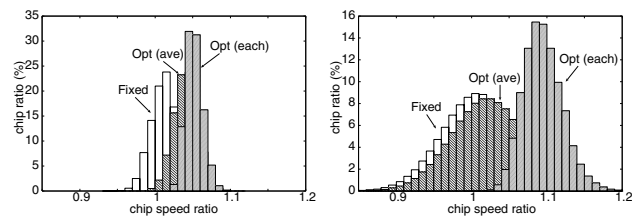


Fig. 11. Distributions of the operating speed normalized by the target operating speed. Left: WID variations are estimated by the measurement results of section III. Right: WID variations are tripled.

the proposed method increases according to the expansion of WID variations. Effect to speed improvement become more than double (4.1% to 9.1%) if the WID variations are tripled in a fine future process.

## ACKNOWLEDGMENT

The VLSI chip in this study has been fabricated in the chip fabrication program of VLSI Design and Education Center (VDEC), the University of Tokyo in collaboration with ASPLA Corp., Synopsys Inc., Cadence Design Systems Inc. and Mentor Graphics Inc.

## REFERENCES

- [1] Naoki Izami, Hiroji Ozaki, Yoshikazu Nakagawa, Naoki Kasai, and Tsunetoshi Arikado. Evaluation of Transistor Property Variations Within Chips on 300-mm Wafers Using a New MOSFET Array Test Structure. *IEEE Transactions on Semiconductor Manufacturing*, Vol.17, No.3, pages 248–254, 2004.
- [2] S. Ohkawa, M. Aoki, and H. Masuda. Analysis and Characterization of Device Variations in an LSI Chip Using an Integrated Device Matrix Array. *IEEE Transactions on Semiconductor Manufacturing*, Vol.17, No.2, pages 155–165, 2004.
- [3] Keith A. Bowman, Steven G. Duvall, and James D. Meindl. Impact of Die-to-Die and Within-Die Parameter Fluctuations on the Maximum Clock Frequency Distribution for Gigascale Integration. *Journal of Solid-State Circuits*, vol. 37, no. 2, pages 183–190, 2002.
- [4] Samie B. Samaan. The Impact of Device Parameter Variations on the Frequency and Performance of VLSI Chips. In *ICCAD2004*, pages 343–346, 2004.