Process Variation Aware D-Flip-Flop Design using Regression Analysis

Shinichi Nishizawa† and Hidetoshi Onodera‡
†Graduate School of Science and Engineering, Saitama University, 255, Shimo-Ohkubo, Sakura-ku, Saitama, 338-8570 JAPAN
‡Graduate School of Informatics, Kyoto University, Yoshida-hommachi, Sakyo-ku, Kyoto 606-8501, JAPAN
E-Mail: nishizawa@mail.saitama-u.ac.jp, onodera@i.kyoto-u.ac.jp

Abstract—This paper describes a design methodology for process variation aware D-Flip-Flop (DFF) using regression analysis. We propose to use a regression analysis to model the worst-case delay characteristics of a DFF under process variation. We utilize the regression equations for transistor widths tuning of the DFF to improve its worst-case delay performance. Regression analysis can not only identify the performance-critical transistors inside the DFF, but also shows these impacts on DFF delay performance in quantitative form. Proposed design methodology is verified as a tool for a DFF design. We develop a delay model using regression analysis and highlight the transistors which have a similar or better delay characteristics in comparison with the DFF designed by an experienced cell designer.

I. INTRODUCTION

Standard cells are key components for designing high quality VLSIs. Performance of standard cells directly affect the final quality of VLSIs. Large number of D Flip-Flops (DFFs) are used in a VLSI circuit as storage elements for digital circuits. Delay, energy performances and cell area of DFF circuit have a strong impact on a VLSI design. DFF circuit is carefully designed by an experienced cell designer and provided by a semiconductor foundry as a library cell targeting its nominal supply voltage.

Recently, with increasing demands to reduce the energy consumption of VLSI circuits, low voltage operation of a VLSI circuit attracts more attention than before. On-currents of PMOS transistors and NMOS transistors are strongly affected by its supply voltage. Optimal design of a DFF circuit may change depending on its supply voltage. However, there are not enough discussions on the design methodology for DFF circuit with a few exceptions [1]–[3]. Identification method of the performance-critical transistors inside a DFF is not clear to improve a DFF delay performance.

As technologies are scaled down to deep sub-micrometer region, within-die random variation becomes critical and significantly impacts on circuit performance [4], [5]. Lowering a supply voltage increases the effect of process variation on circuit performance. Thus, variation tolerant DFF design methodology is required. DFF is composed of a few dozen of transistors, and the performance of the transistors and sub-circuit blocks affect the performance of other transistors and other sub-circuit blocks. Several articles reported random variation strongly affects the performance of the latch circuits to capture and keep the input data synchronized with the clock signal. References [6], [7] discuss the failure condition of DFFs caused by process variation. In these papers, they propose a method for identifying performance-critical transistors by Monte-Carlo analysis and improve functional yield by enlarging the widths of the identified transistors. However, there are no discussion on how to obtain an optimal width for each transistor under process variation. References [8], [9] evaluate the impact of process variation on its delay and energy characteristics in several DFF structures. However, these papers use only one constant width for all of transistors inside the DFF in the analysis. In actual DFF design, there are several variations in transistor widths and it is common to optimize those widths to achieve both fast operation and low energy consumption at the same time. It is not clear how to optimize the transistors width to achieve faster performance under process variation. Reference [10] proposes statistical framework for variation tolerant DFF design. This paper models the inverters and the transmission gates in a DFF circuit as a simple RC circuit, and the DFF delay is assumed as the sum of these logic gate delays for the worst case delay calculation without Monte-Carlo simulation. This model is simple, however this technique has several problems. It is difficult to assume all the transistors inside DFF affect to the whole DFF delay characteristics. In an actual DFF circuit, some performance-critical transistors are exists and these transistors determine the whole DFF delay characteristics. Also, this delay model exhibits large delay calculation error compared with transistor level simulation.

In this paper, we discuss a variation tolerant DFF design methodology for low voltage operation. We use regression analysis as a tool to obtain a delay model equation, and highlight the transistors which have a large impact on the DFF delay performance. The delay model is utilized to design a variation tolerant DFF, tuning the transistor widths inside the DFF, and verify its performance using Monte-Carlo simulation. We use a Transmission Gate DFF (TGFF) composed of minimum width transistors as an initial DFF circuit, and try to explore the best set of transistor widths for the TGFF to achieve the best delay performance at the 3σ worst case condition.

The rest of this paper is organized as follows. Section II describes the regression analysis for DFF delay analysis. Section III describes the proposed DFF design methodology using regression formulation. Section IV describes experimental results of variation tolerant DFF design. Section V concludes this paper.

II. REGRESSION ANALYSIS FOR DFF VARIATION ANALYSIS

In this section, we describe the use of regression analysis as a tool for a DFF design. We develop a delay model using regression analysis and highlight the transistors which have a strong impact on DFF delay characteristics.

Figure 1 briefly describes the proposed variation-aware DFF design flow using regression analysis. In this design flow, we try to find performance-critical transistors inside a DFF and enlarge these widths to improve the DFF delay performance.
Proposed flow is composed of three main parts, and iteration method is used to generate a final DFF design. Firstly, Monte-Carlo simulation is performed and obtain transistor performance variations and DFF delay performance variations under process variation. Secondly, regression analysis is performed to model a relationship between the transistor performances and the DFF delay performance considering these mean and standard deviation. Thirdly, the transistor widths are tuned to achieve faster delay performance utilizing the result of regression analysis. If the performance is not enough or there are more budget to enlarge transistors width, we can iterate this flow to obtain faster DFF performance. Finally, optimal DFF design after transistor width tuning is obtained.

### A. Brief explanation of regression analysis

In this section, we briefly explain the regression analysis for DFF variation evaluation. Regression analysis is a statistical modeling method for estimating the relationships among the variables. Regression analysis tries to formulate regression function which expresses a relationship between a dependent variable and one or more independent variables. Regression function also indicates the impact of independent variables on the dependent variable. Thus, regression analysis helps designer to select and evaluate the best set of variables to be used for developing a predictive delay model.

### B. Setup for regression analysis of DFF delay analysis

We build a regression function which expresses the relationship between DFF delay characteristics and transistor performances. DFF delay characteristics is a dependent variable and transistor performances are independent variables in this regression analysis. In the regression analysis, some transistors have a strong relationship to the DFF delay characteristics, but other transistors may have a weak relationship to the DFF delay characteristics. We check probability of the coefficient of each independent variable in the regression function, and eliminate some of the variables which p-values (probability value) are less than significant level. The final result of the regression function is achieved when all of the p-values of reach to a significant level.

There are several important delay characteristics for DFF circuit, such as setup time (also called Clock-to-Data contamination delay), Clock-to-Q (C2Q) propagation delay and hold time characteristics. Data-to-Q propagation (D2Q) delay is defined as a sum of a setup time and a C2Q delay. In this paper, we build an objective function to minimize a D2Q delay, since (1) D2Q delay is composed of a setup time and a C2Q delay thus we can consider both the delay parameters, and (2) D2Q delay correspond an overhead in sequential circuit design since the maximum allowable logic delay is restricted by the clock cycle time minus the D2Q delay [11]. In this paper, we do not consider hold time characteristics since hold time violation can be eliminated in a circuit design [12].

Figure 2 shows a schematic of the target TGFF circuit. To obtain a relationship between the D2Q delay characteristics and transistor on-currents under the process variation, we perform Monte-Carlo simulation. In the evaluation, we use a transistor model which considers the threshold voltage dependence on the transistor channel area to reflect the Pelgrom’s model [13].

Our goal is to design a DFF circuit which has faster worst-case delay characteristics under process variation. Since transistor on-currents and circuit delay have an inverse relationship, it is difficult to build a regression function to estimate worst case delay under the transistor width tuning. We build a regression function which express relationship between operation speed (inverse of the D2Q delay) and transistor performances.

Figure 3 shows a simulation circuit for the DFF delay evaluation. The waveform generators composed with FO4
Divide each pull-up/down network and evaluate its on-current as "transistors performance".

Fig. 3. A DFF simulation circuit.

Fig. 4. Transistors current evaluation method.

loaded inverter cells generate data and clock signals for a target DFF. Each circuit contain parasitic RC from a layout to emulate realistic input slew and loading condition. Since D2Q delay strongly depends on a D2C delay, minimum values of D2Q delay for rise input case and fall input case are individually obtained changing the D2C delay for each Monte-Carlo trials.

Figure 4 shows a simulation circuit for the transistor performance evaluation. There are many candidates as a metrics of transistor performance, such as transistor width, threshold voltage, and on-current. We use static DC on-current as a metric of transistor performance, since the use of DC on-current is suitable to avoid multicollinearity problem in regression analysis. Pull-up and pull-down paths inside the DFF are divided, and these DC on-currents are individually evaluated. Stacked transistors are treated as single equivalent transistor and its on-current is used as variable to avoid the multicollinearity problem. Also, on-current of a PMOS transistor is used to represent the transmission gate and treated as one independent variable, since parallel connected transistors show similar characteristics to transistor width modulation and this leads the multicollinearity problem. This transformation reduces the independent variable from 24 transistors to 17 variables as initial set of independent variables in the regression analysis.

III. Regression function for DFF delay evaluation

In this section, we obtain a regression function which estimates the DFF operation speed at the worst case condition under process variation. The regression function is used to estimate the DFF operation speed considering the transistor width tuning on the target DFF. We propose an equation to estimate mean delay and its standard deviation to estimate the worst case 3σ delay under the transistor width tuning.

A. Building the regression function for DFF circuit design

We create data set of DFF performances and transistor on-currents using Monte-Carlo simulation, and build regression functions for both the D2Q rise and fall operations.

First, we build a normalized regression function to extract the performance-critical transistors. DFFs operation speed and transistor on-currents values have different units and mean values. Each value is normalized, and the following expression is obtained as normalized regression expression.

\[ O_{\text{norm}} = K_{1,\text{norm}} I_{1,\text{norm}} + K_{2,\text{norm}} I_{2,\text{norm}} + \cdots + K_{i,\text{norm}} I_{i,\text{norm}} \]

where \( O_{\text{norm}} \) is a normalized value of an operation speed of the DFF, \( K_{i,\text{norm}} \) is a normalized regression coefficient and \( I_{i,\text{norm}} \) is the normalized value of on-current of the \( i \)-th transistor, respectively.

If we assume the current of the \( j \)-th transistor \( I_{j,\text{norm}} \) and DFF operation speed \( O_{\text{norm}} \) follows a Gaussian distribution, we can express the relationships of mean and standard deviation of operation speed and transistor on-currents as following.

\[ \mu_{O,\text{norm}} = \sum_{j=1}^{i} (K_{i,\text{norm}} \cdot \mu_{I,j,\text{norm}}) \]

\[ \sigma_{O,\text{norm}}^2 = \sum_{j=1}^{i} (K_{i,\text{norm}} \cdot \sigma_{I,j,\text{norm}})^2, \]

where \( \mu_{O,\text{norm}} \) and \( \sigma_{O,\text{norm}} \) are the mean and the standard deviation of the normalized DFF operation speed, and \( \mu_{I,j,\text{norm}} \) and \( \sigma_{I,j,\text{norm}} \) are the mean and the standard deviation of the normalized \( j \)-th transistors on-current, respectively.

In the regression analysis, there are some statistical metrics to evaluate the obtained regression function. When some independent variables shows larger p-value then we can eliminate these independent variables from regression function and build more simple and compact function. Adjusted R-squared is an index value how the obtained regression function well describes the input data set. Normalized regression analysis is also utilized to evaluate the impact of each transistor on DFF operation speed comparing the regression coefficients. Transistors with larger regression coefficients have larger impact on DFF operation. On the other hand, transistors with smaller regression coefficients have less impact on DFF operation thus we can eliminate these transistors form delay estimation in regression analysis.

Next, we build a regression function to estimate worst case operation speed of DFF. Operation speed and each on-current value can be expressed as following regression expression.

\[ \mu_{O} = \sum_{j=1}^{i} (K_{j} \cdot \mu_{I,j}) \]

\[ \sigma_{O}^2 = \sum_{j=1}^{i} (K_{j} \cdot \sigma_{I,j})^2, \]
where $\mu_0$ and $\sigma_0$ are the mean and the standard deviation of DFF operation speed, and $\mu_{ij}$, $\sigma_{ij}$ and $K_i$ are the mean and the standard deviation of $j$th the transistors on-current, and its regression coefficient, respectively.

Then, we build an operation speed estimation function utilizing the equations (4) and (5). To consider the impact of transistor width tuning on both operation speed and its standard deviation, we use following equations to estimate both the mean operation speed and its standard deviation after the transistor width tuning.

$$\mu_{ij,\text{pred.}} = \frac{L_{ij,\text{original}} W_{ij,\text{pred.}}}{L_{ij,\text{pred.}} W_{ij,\text{original}}}$$

$$\sigma_{ij,\text{pred.}} = \sigma_{ij} \sqrt{\frac{L_{ij,\text{original}} W_{ij,\text{original}}}{L_{ij,\text{pred.}} W_{ij,\text{pred.}}}}.$$  

where $\mu_{ij,\text{pred.}}$ and $\sigma_{ij,\text{pred.}}$ are the predicted mean on-current and its standard deviation of the $j$th transistor after the width modulation. $L_{ij,\text{pred.}}$, $L_{ij,\text{original}}$, $W_{ij,\text{pred.}}$, and $W_{ij,\text{original}}$ are the length and width of the $j$th transistor before and after the transistor width modulation, respectively. The $\sigma_{ij,\text{pred.}}$ value itself is an estimated value calculated from transistor width dependence of the random variation based on Pelgrom’s model.

B. Objective function for transistor width tuning

Objective of this paper is to design a DFF with faster delay performance under process variation. The worst case operation speed ($\mu-3\sigma$ point operation speed $O_{\mu-3\sigma}$ in this paper) can be expressed utilizing the equations (4),(5),(6) and (7) as follows,

$$O_{\mu-3\sigma} = \mu_{O,\text{pred.}} - 3\sigma_{O,\text{pred.}}$$

$$= \sum_{j=1}^{i} K_j \cdot \mu_{j,\text{pred.}} - 3 \sqrt{\sum_{j=1}^{i} K_j^2 \cdot \sigma_{j,\text{pred.}}^2}.\quad (9)$$

Transistor widths are tuned to achieve DFF with better worst-case delay performance under process variation. If the performance-critical transistors are appropriately selected and their width are enlarged, transistor width tuning has an possibility to improve the DFFs worst-case operation speed. To maximize the operation speed for the both rise and fall input data cases, we minimize the objective function $D_{\mu+3\sigma,\text{obj}}$ as follows,

$$D_{\mu+3\sigma,\text{obj}} = \sqrt{D_{\mu+3\sigma,\text{rise}}^2 + D_{\mu+3\sigma,\text{fall}}^2},\quad (10)$$

where assuming

$$D_{\mu+3\sigma,\text{rise}} = \frac{1}{O_{\mu-3\sigma,\text{rise}}},\quad (11)$$

$$D_{\mu+3\sigma,\text{fall}} = \frac{1}{O_{\mu-3\sigma,\text{fall}}}.\quad (12)$$

$O_{\mu-3\sigma,\text{rise}}$ and $O_{\mu-3\sigma,\text{fall}}$ are the worst case rise and fall operation speed calculated using equation (9).

C. Iteration of the regression analysis and the transistor width tuning

After transistor widths are updated, regression analysis is recursively applied for new DFF circuit. Some of the performance-critical transistors may not be a performance-critical after the width tuning, and the other transistors may become a performance-critical in next design. We need to iterate the regression analysis and transistor width tuning to improve the final DFF performance. Parasitic RC strongly affects to the DFF operation thus it is required to extract parasitics from layout information. If there are small modifications from previous trial, it is possible to reuse the simulation netlist with parasitics for the next trial updating the transistor widths and diffusion parasitic capacitance. If the large modifications are required, it is better to re-extract a netlist from a new layout. We partially utilize cell layout generator [14] to update the simulation netlist when the layout structure has been modified. This iteration requires Monte-Carlo simulation for each iteration and it requires large simulation cost. We evaluate the required number of iterations and compare the final result of DFF in next section.

IV. Experimental Results

A. Implementation result

The target DFF in this paper is TGFF designed using 28 nm LP CMOS process. Nominal supply voltage for this process is 1.0 V. We target 0.5 V supply voltage for the evaluation of the DFF at the low voltage operation.

We use a data set of 300 trials of Monte-Carlo simulation and build a regression function of operation speed at both rise data input and fall data input cases. Transistor width tuning is performed to maximize a 3$\sigma$ point operation speed estimated based on the regression function. We recursively perform this process updating the DFF design. When this process finishes, we obtain the 3$\sigma$ point operation speed of the DFF designed based on the proposed regression analysis, and compare the result with the DFF designed by an experienced designer.

Utilizing the results of regression analysis, we tune the transistor widths inside the DFF to maximize its worst case operation speed. There are 17 transistors as independent variables for the regression function thus we select some transistors and enlarge the width to maximize the operation speed of the DFF utilizing the objective function in equation (10).

We start from a DFF (named DFF$_{\text{min}}$) which is composed of transistors with minimum width as an initial DFF. Applying proposed DFF design methodology, we obtain a DFF (named DFFRA) which targets to maximize worst-case operation speed. Another DFF (named DFF$_{\text{manual}}$) which is designed by experienced cell designer is used as competitor. We assume same total channel area for both DFFRA and DFF$_{\text{manual}}$ as a transistor width tuning constraint.

In the experiment, we use Synopsys HSPICE for the transistor level simulation and R for the regression analysis. In the regression analysis, we check p-values of each independent variable and eliminate some independent variables from regression function which show larger than 0.1% estimation error.
B. Design examples of DFFs with proposed design flow

Figure 5 shows evaluation result of the worst-case operation speeds of DFFs for rise and fall data inputs, which are obtained by Monte-Carlo simulations and estimated by regression analysis. In the proposed design flow, transistor widths are updated recursively. We perform 20 trials to improve the worst-case operation speed while the total channel area inside the DFF is less than a constraint value. Monte-Carlo simulation is performed for initial DFF and worst case delay is obtained (MC0). Then, regression analysis estimates worst case delay after the transistor width tuning (RA1). In the next trial, Monte-Carlo result (MC1) does not reach to the estimated point (RA1), since the performance-critical transistors are changed due to the transistor width tuning at first trial. However, Monte-Carlo simulation result shows both the rise and fall operation speeds are improved constantly, since regression analysis succeed to extract the performance-critical transistors at each trial.

Figure 6 shows evaluation result of the worst-case operation speed of DFFs with different number of iterations. Total channel area constraint is constant, thus each DFF has different amount of additional channel area for transistor width tuning in each trials. If we use only one trial (assign all channel area constraint to initial result), operation speed degraded from original DFF. As the number of trials increased, the final result of DFFs show better operation speed. In this case, more than 10 iterations show almost the same performance.

C. Comparison result

Table I shows the transistor widths of DFFs. DFF\textsubscript{manual} is designed by experienced designer which targets 0.5 V operation. DFF\textsubscript{RA,0.5 V} is designed with proposed design method targets 0.5 V operation. DFF\textsubscript{RA,1.0 V} is designed with proposed design method but its target supply voltage is 1.0 V. DFF\textsubscript{RA,1.0 V} is a design example how supply voltage difference affects to transistor width tuning result. Two DFFs designed by the proposed method uses 20 times iteration in the design flow, and these have almost same total channel area as DFF\textsubscript{manual}. In the table, averaged value is used for series connected transistor widths. Result shows DFF\textsubscript{RA,0.5 V} has almost same set of transistor widths as DFF\textsubscript{manual}. On the other hand, DFF\textsubscript{RA,1.0 V} shows different set of transistor widths. This result shows the optimal set of transistor widths varies depending on its supply voltages.

Table II shows the cell layout height and width of each DFF. Note that each cell size is normalized by the library basic cell (unit cell). Since there is no cell area constraint in the proposed design flow and objective function, some transistors requires two-finger structure thus it requires larger cell width compare to DFF\textsubscript{manual}.

We perform 1000 trials of Monte-Carlo simulation and evaluate the mean D2Q delay and standard deviation. Table III summarizes the final values of objective function, mean and standard deviation of the D2Q delay, and the calculated worst

<table>
<thead>
<tr>
<th>Tr. id#</th>
<th>DFF\textsubscript{min}</th>
<th>DFF\textsubscript{manual}</th>
<th>DFF\textsubscript{RA,0.5 V} (iteration: 20)</th>
<th>DFF\textsubscript{RA,1.0 V} (iteration: 20)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>mp1/mn1</td>
<td>80 PMOS</td>
<td>80 NMOS</td>
<td>270 PMOS</td>
<td>180 NMOS</td>
<td>270 PMOS</td>
</tr>
<tr>
<td>mp2/mn2</td>
<td>80 PMOS</td>
<td>80 NMOS</td>
<td>300 PMOS</td>
<td>200 NMOS</td>
<td>320 PMOS</td>
</tr>
<tr>
<td>mp3/mn3</td>
<td>80 PMOS</td>
<td>80 NMOS</td>
<td>80 PMOS</td>
<td>80 NMOS</td>
<td>80 PMOS</td>
</tr>
<tr>
<td>mp4/mn4</td>
<td>80 PMOS</td>
<td>80 NMOS</td>
<td>180 PMOS</td>
<td>120 NMOS</td>
<td>290 PMOS</td>
</tr>
<tr>
<td>mp5/mn5</td>
<td>80 PMOS</td>
<td>80 NMOS</td>
<td>300 PMOS</td>
<td>200 NMOS</td>
<td>290 PMOS</td>
</tr>
<tr>
<td>mp6/mn6</td>
<td>80 PMOS</td>
<td>80 NMOS</td>
<td>80 PMOS</td>
<td>80 NMOS</td>
<td>80 PMOS</td>
</tr>
<tr>
<td>mp7/mn7</td>
<td>80 PMOS</td>
<td>80 NMOS</td>
<td>360 PMOS</td>
<td>240 NMOS</td>
<td>420 PMOS</td>
</tr>
<tr>
<td>mp8/mn8</td>
<td>80 PMOS</td>
<td>80 NMOS</td>
<td>240 PMOS</td>
<td>160 NMOS</td>
<td>80 PMOS</td>
</tr>
<tr>
<td>mp9/mn9</td>
<td>80 PMOS</td>
<td>80 NMOS</td>
<td>240 PMOS</td>
<td>160 NMOS</td>
<td>220 PMOS</td>
</tr>
</tbody>
</table>

Total channel area [\mu m^2]: 0.058, 0.126, 0.122, 0.124
case delay at the 3σ variation point. Objective function of DFF_{RA,0.5 V} achieves 7.2% smaller value than that of DFF_{manual} at 0.5 V operation. Figure 7 shows the 3σ point worst case rise and fall delay. Result shows DFF_{RA,0.5 V} achieves 14% faster 3σ rise delay than DFF_{manual}, at the cost of 3.0% slower 3σ point fall delay. DFF_{RA,0.5 V} shows 10% faster rise slew and 2.9% faster fall slew than that of DFF_{manual}, since DFF_{RA,0.5 V} has wider transistors for output inverter. Enlarging the transistor widths increases gate capacitances and diffusion capacitances thus it increase circuit operation energy. DFF_{RA,0.5 V} and DFF_{manual} consumes almost twice energy than DFF_{min}, but there are less difference in energy consumption between DFF_{RA,0.5 V} and DFF_{manual}, since total channel areas are almost same. Results show proposed design method can generate a DFF which has almost equivalent performance as a DFF which is designed by an experienced cell designer.

V. CONCLUSION

This paper discuss a design methodology for variation aware DFF, using regression analysis to express the DFF operation speed using the on-current of transistors inside the DFF. DFF circuit is carefully hand-crafted by an experienced cell designer, however there are not enough discussion that how to design or tune the transistor size inside a DFF to achieve faster delay performance. Regression analysis not only identifies the transistors which strongly affect to DFF delay characteristics, but also shows its amount of impact on delay in quantitative form. Proposed design methodology is verified via DFF design experiment, and result shows the DFF designed with proposed methodology has similar or faster delay characteristics as a DFF designed by an experienced cell designer.

Our future work is to consider both energy consumption and area constraint into objective function to achieve energy efficient, compact and faster delay characteristics for DFF circuit considering process variation effect.

This work has been partly supported by JSPS KAKENHI JP16H01713 and JP17K12657. This work is also partly supported by VLSI Design and Education Center (VDEC), the University of Tokyo in collaboration with Synopsys, Inc., Cadence Design Systems, Inc., and Menter Graphics, Inc.

ACKNOWLEDGMENTS

REFERENCES