Process Variation Mitigation via Body Bias Technique

Neha Mishra¹, Shweta Agrawal², Prashant Bada¹
Research Scholar, Dept. of ECE., SRCEM Banmore, Morena, India
Assistant Professor, Dept. of ECE., SRCEM Banmore, Morena, India

ABSTRACT: The high performance VLSI designs are achieved via different technique where scaling is the primary approach used in last few decades. The process variation which is the outcome of limitation of fabrication technology, causing a significant yield failure. Process variations is deviation in devices parameter from their original value and becoming so severe that further scaling is not possible. The designs which are implemented without considering process variations fails to provide correct output. The modern devices not only demand high performance, energy efficient operation but also demand operation should be done accurately. Therefore, critical analysis is required, and corresponding mitigation technique has to be introduced to perform signal processing accurately. This paper first critically analyses different types of process variations and then evaluates the effectiveness of the body bias approach to mitigate the effect of process variation. The simulation results show the body bias approach significantly reduces the process variation severity.

KEYWORDS: Very Large Scale Integration, Fabrication, Process Variations, Technology scaling.

I. INTRODUCTION

The high performance devices are the outcome advancement in the VLSI technology. The prime approach of achieving the high performance, energy efficient designs are the device scaling. In past few decades, CMOS technology scaling [1] has been the primary engine that powered the ascent of semiconductor industry and provided a path toward both denser and faster integration an engineering achievement unmatched in human history. There are several research works that deal with the impact of technology scaling in various aspects of VLSI design where process variation (PV) has been found more of a concern and has received an increased amount of attention in recent years. Research studies on PV exposed that PV are expected to get worsen in future technologies, and this may result in unreliable circuits.

The assumption of deterministic device parameters fail as the technology moves towards sub-65nm design. The device behaviour below 65nm changes from deterministic to random. This fluctuation in the device characteristics is due to variation in process parameters such as change in length (L), width (W), oxide thickness (T_ox), number of dopant (N) in source and drain diffusion etc. The deviation in device parameter due to process is called process variation [2].

Different PVS are modelled in terms of variation in threshold voltage and as technology scales, the shift in threshold and spread of threshold variation increases [3]. Therefore, the PV is shown by a distribution graph showing variations in threshold and which in turn deviation in the delay and leakage parameters of the whole circuit. For the present ICs, the major factor for yield loss is only due to process variations [4]. The effect of process variation is so severe in the modern designs that design behaviour has changed from deterministic to probabilistic [5]. Therefore, a critical analysis of the PV of different technology is required.

Significant research has been devoted to achieve process aware designs and different techniques are developed. These techniques can be design time techniques or run-time techniques implemented at circuit/architectural level. Different techniques exhibit different merits and associated demerits in terms of area/power/delay overhead.

In order to design process aware logic, in this work an exhaustive literature review is done on the different sources of process variations and its classification based on its effect with time. Further body bias technique is exploited as mitigation technique.
II. PROCESS VARIATION AND MITIGATION TECHNIQUES

This section provides an exhaustive literature review on the various sources of process variations and the different mitigation techniques developed to make design process aware.

2.1 Process Variation
The process variation is deviation in device parameters due to the limitation of fabrication foundry. Its effect becoming severe below 90nm technology due the fact that at this technology small deviation results in large deviation in devices characteristics [6]. Process variation is a natural phenomenon of variation in the attributes of devices, such as L, W and Tox.

2.2 Process Variation classification
The PV can be classified into two major categories namely: Inter-die variation (die-to-die) and Intra-die variation (within die) variations.

2.2.1 die-to-die variations
The die-to-die variation shows the variation in the device parameters place on the different die or different wafer or lot [7], which ideally should be identical values. This deviation in device parameters is occurred due to sub-wavelength lithography process and exhibits symmetric variation. It is observed that beyond 45nm technology, the inter-die variation was more significant over intra-die i.e. yield loss is more due to inter-die than intra-die.

2.2.2 Within-die variation
The intra-die variation exhibits random nature of the variations. This random behavior occurs due to random nature of dopant atoms producing random dopant fluctuation, line edge/width roughness LER/LWR and Tox.

2.2.2.1 Random Dopant Fluctuations
The RDF occurs due to random motion of the dopant atom while forming the source and drain diffusion as shown in Fig. 1.

![Fig. 1: Illustration of random dopant fluctuation.](image)

During the process of ion implantation for forming two diffusions, dopant atoms are bombarded on the p-type substrate [8]. In this process the location of the dopant atom depends of the kinetic energy of the atom, time of diffusion and environment temperature. Small deviation in the temperature and other parameter will change the concentration of dopant atoms [8].

2.2.2.2 LER and LWR
The line-edge roughness [9] is process variation occurred due to imperfect etching process as shown in Fig. 2. It refers to the deviation in the poly width that forms the gate of the device. This deviation in the device width results in the shorter or larger transistors channel length. This may cause the mismatch in the device channel length of two transistors on the same chip. This intra-die variation changes the device behavior from deterministic to stochastic.
2.2.3 Temporal Variations
There is another kind of variation is the temporal variation the come within the device with its age i.e. its severity increases with device usage. In this PV device degrade with time and do not retain its original specifications. Following subsection details different types of temporal variations and their severity will be analyzed.

2.2.3.1 NBTI and PBTI
The Negative bias temperature instability is the major source of temporal variation occurred in the PMOS device when biased negatively for a long time. The stress due to negative biasing causes defect generation at the Si-SiO₂ interface and increase with time. These defect increases the threshold of PMOS which increases the delay of the device as shown in Fig. 3.

This process variation exhibits partial recovery property i.e. the created dangling bonds due stress recovers when the stress is removed partially. Since the recovery mechanism is partial, there is some defects still remain in the oxide layer which degrade the characteristic of the MOS device.

Similar to the NBTI, the PMOS device exhibits positive bias temperature instability (PBTI) [6] in NMOS devices. Although, in older technologies the effect of NBTI was more severe as compared to PBTI and was ignored, in the modern designs due to the use of high-k material, PBTI is also severe as NBTI.

2.2.3.2 Hot Carrier Injection (HCI)
The high electric field in the device due to reduced device dimension make the charge carrier’s kinetic energy very high. This carrier gets heated and result in impact ionization as shown in Fig. 4. The HCI induces defect at Si-SiO₂ interface or in the bulk and degrades the device performance similar to the NBTI.
These heated carriers generated due to the impact ionization move in to the oxide layer and degrades the device characteristics. Further the effect of HCI is more significant in NMOS as compared to PMOS as the electrons have higher mobility over the holes. Moreover, the effect of HCI is more when the NMOS device is switched from low-to-high.

2.3 Process Variation Mitigation Techniques
There are different circuit/architectural level techniques developed to mitigate the effect of process variations. These techniques can be broadly classified as design-time techniques and run-time techniques [10]. The design time-techniques are different circuit/architecture and device level techniques that are employed while designing the process aware designs whereas the run-time techniques are additional circuit/logic which will be effective during run-time.

2.3.1 Design-time techniques
There are number of circuit/architectural level techniques that are used to reduce the effect of process variations. The following subsection detail the different circuit/architectural level techniques of process variation aware design.

2.3.1.1. Circuit level techniques
The circuit level techniques include conservative approach of considering large design margin to mitigate process variation. this method of design considers original delay and the worst case delay due to variations while designing the circuit. For example, consider the frequency of the operation for the design after adding delay all margins as shown in Fig. 5

From the Fig. 6 it can be seen that delay due to the PVT variation causes failure which can be eliminated by designing logic with reduced delay by supply upscaling or by transistor resizing. Further, another approach can be reduction in the frequency of design to consider all delay due to different variation which in turn provides reduced performance design. Therefore, in both the cases this method does not provide an efficient design [11]. Further, logic
level resizing can be used to modify the delay of logic such that it can reduce the amount of delay increased due to process variation. The dual threshold logic is also employed as the technique to mitigate the effect of process variations. In this technique, leaky designs are made with higher threshold logic whereas the low performance/high-delay designs are lower threshold logic to mitigate the effect of process variation.

2.3.1.2. Architectural level techniques:
There are number of architectural level techniques that can be used to eliminate/reduce the effect of process variations. Following subsection details some of the important process mitigation techniques.

Pipelined unbalancing: The variation in the data-path delay in the pipeline architectures significantly affect the performance of the architecture. The over design to mitigate the process variations results in inefficient designs. There are different frameworks have been developed to determine the optimal delay between pipeline stage under process variation. In order to balance the different pipeline stages, in one approach for example, reduce the delay by supply upscaling or increase the size of the transistor. On the other hand, transistors on the non-critical path the operated with reduced supply to maintain the power budget.

CRISTA: It is design paradigm for to develop low power higher performance circuit or system for the error resilient applications. In this approach aggressive supply voltage scaling [12] is done to reduce power while maintaining the operative frequency. This supply voltage reduction results in the timing error in the design which may cause failure. Therefore, in this approach supply voltage scaling is done for the path which are not critical. The concept of CRISTA timing is shown in Fig. 6.

![Fig. 6: Timing diagram of CRIST approach.](image)

It can be seen from the Fig. 6 that if timing error occurred in the critical path, an additional clock cycle is given to complete the operation without any error. Further, if the probability of critical time activation is low, it will significantly improve the performance without loss accuracy.

2.3.2 Run-time techniques
Following subsection details the run-time techniques to compensate the effect of process variations [13]. The advantage of the run-time technique is that the technique can be applied in the existing designs also with some additional circuit/logic.

2.3.1.2. Circuit-level techniques
Number of techniques at the circuit level have been introduced [14] in the literature over which body bias is the most frequent and effectively used technique. The principle of body bias technique to mitigate the process variation is the dependency of threshold voltage on the body voltage. As we know that, threshold voltage of the transistor can be increased or decreased by reverse or forward biasing the body over the source terminal. Therefore, for the die which exhibits lower threshold voltage due to process variation can be applied with forward by bias to mitigate the increased threshold due to process variations. Similarly, applying reverse body bias increases the threshold voltage which gets increased due to process variations. Further due to process variations, the die with reduced threshold voltage are considered as too leaky whereas the die with higher threshold are very slow over the typical ICs as shown in Fig. 7.
Fig. 7: Body bias effect shown on the die exhibiting high process variation.

The principle of the body bias to apply appropriate supply voltage such that leaky die and slower die will fall under the typical design range. Slower die from bin 3 and too leaky die from bin 1 can be moved to bin 2 by applying adaptive body bias.

2.3.1.2. Architectural-level techniques:
Along with the circuit level technique, architectural level technique that can be applied during run-time are also introduced. RAZOR is the run time technique at architectural level as discussed in details below.

**Razor [9]:** It is the dynamic voltage scaling technique in which supply voltage is dynamically scaled and corresponding timing error is detected and corrected during run time. The working principle of the RAZOR technique lies in the detection of timing error due to process variations and due to scaling and the apply a corrective action such that time errors are avoided. In this technique a special RAZOR flip-flop is used which double sample the original signal with small delay as shown in Fig. 8. The RAZOR flip-flop consists of main flip-flop along with the shadow latch and a comparator to compare the two sampled signal for the discrepancy. If there is the different an error signal will be generated and corresponding recovery mechanism will be activated.

Fig. 8: Body bias effect shown on the die exhibiting high process variation.

If there is error between the two sampled signal, it reflects the error propagation and then corresponding action will be taken. In this technique if the sampled signal shows an error then modified misspeculation recovery mechanism is active and will recover from the error.

III. PROPOSED WORK

The section details on the proposed work on process variations severity analysis and then presents body bias technique to mitigate the effect of process variation.

3.1 Analysis of process effect on performance
The methodology followed to evaluate the severity of the process variations at different technology is given by the flow chart in Fig. 9.
In order to evaluate the process variation severity on the path delay of the logic, a single threshold inverter is considered as shown in Fig. 9. The threshold of this MSOFETs within this inverter is varied due to process variation. Further using this inverter, inverter chain consists of 100 inverters and 500 inverters are also considered to compute the net effect on the logic delay. Although, the procedure to compute the variation statistic is similar to the single inverter, consideration of large chain of inverter cancel out the random intra-die variations while provide more accurate inter-die variations. This inverter chain accurately models the general path of the logic. The simulation results in the next section illustrates the severity of the process variation on different inverter chain at different technology.

IV. EXPERIMENTAL RESULT & ANALYSIS

The performance deviations due to PV is computed by implementing inverter chains are designed on Tanner. Spice netlists are extracted and simulated with Monte-Carlo simulation [15]. The designs are implemented with different size based on the technology considered e.g. in 90nm technology $L_{MN}$ considered is 90nm while in 32nm design $L_{MN}$ can be taken is 32nm. The technology files considered are the 90nm, 65nm, 45nm, and 32nm. It is observed that at 90nm, 65nm, 45nm, and 32nm technologies, the change in threshold variations due to process variations are 10%, 20%, 40%, 70% respectively. Finally, these spice netlists are simulated and performance metrics are extracted.

4.1 Implementation of design on Tanner:
The schematic of the single inverter designed on the Tanner is shown in Fig. 10. It consists of two transistor sized properly based on the kind of technology considered while simulation.
Fig. 10: Schematic of inverter on Tanner.

The functionality of the implemented designs is also verified by simulation. Further, a symbol for the inverter is created to use this inverter for making inverter chain. The schematic of the 100 inverter chain is shown in Fig. 11. The inverter chain utilizes the same inverter designed on the Tanner. The primary input to this inverter is the A while the output is Y.

Fig. 11: Schematic of inverter chain (100 inverters).

Further, an inverter-chain consists of 500 inverters is also designed to evaluate the effect of process variation on the longer path. The spice netlists for these designs are extracted and simulated with the specified technology files. The simulation results for different technology are given in the next section.

4.2 Simulation Results and Analysis
In order to compute the performance deviation due to process variation, different $V_{TH}$ variations are considered at different technology. The statistic of the threshold variation as illustrated in Table 1 for pMOS transistor reflect that the variance increases with increasing scaling i.e. at latest technology (32nm) node deviations are much more over the 90nm.
Table 1: PMOS $V_{TH}$ statistic (mv) at different technology.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Technology</th>
<th>90nm</th>
<th>65nm</th>
<th>45nm</th>
<th>32nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td></td>
<td>-338.0816</td>
<td>-362.8707</td>
<td>-229.5111</td>
<td>278.4249</td>
</tr>
<tr>
<td>Median</td>
<td></td>
<td>-339.9022</td>
<td>-367.0919</td>
<td>-232.8989</td>
<td>287.4947</td>
</tr>
<tr>
<td>Variance</td>
<td></td>
<td>0.1113068</td>
<td>0.5983488</td>
<td>0.3854291</td>
<td>2.7624</td>
</tr>
<tr>
<td>Sigma</td>
<td></td>
<td>10.5502</td>
<td>24.4611</td>
<td>19.6323</td>
<td>52.5586</td>
</tr>
</tbody>
</table>

Similarly, to the PMOS device statistics for the nMOS considered. With this statistics of the devices, the designs (single stage inverter, inverter chain of 100 inverters and 500 inverters) are simulated.

4.2.1 Simulation results for single inverter

The delay distribution metrics for single stage inverter (SSI) simulated with different technology node are given in Table 2. The simulation results show that mean of delay is reducing with each technology which is due to outcome of the scaling. But, with this reduced delay, deviation in the delay with respect to the mean value is increasing with each technology which is very severe and degrades the yield of the designs. As the variation from the mean value reduces, some of the IC may fails due to exhibiting longer critical path. From the simulation results show that this deviation increase by more than 6X in at 32nm over 90nm technology.

Table 2: Delay statistic of SSI at various technology.

<table>
<thead>
<tr>
<th>Delay (ps)</th>
<th>Technology</th>
<th>90nm</th>
<th>65nm</th>
<th>45nm</th>
<th>32nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td></td>
<td>23.7</td>
<td>23.6</td>
<td>44.4</td>
<td>2.4.7</td>
</tr>
<tr>
<td>Sigma</td>
<td></td>
<td>3.39</td>
<td>6.13</td>
<td>10.8</td>
<td>20.7</td>
</tr>
</tbody>
</table>

Finally, the delay distribution as shown as in Fig. 12 reflect a comparative increase in the standard deviation in the scaled technologies.

Fig. 12 Delay variation of SSI at various technology.

In order to more accurately check the effect on the path delay in the logic, inverter chain is considered.
4.2.2 Simulation results for inverter chain
The simulation result of the delay metric variation is computed for inverter chain at different technology and illustrated in Table 3. It can be seen from the Table that delay is getting reduced with each technology but at the same time spread of delay distribution increases with technology. Therefore, it can be seen that inter-die variation causes delay variation increases with each scaled technology and reaches to an intolerable limit.

Table 3: Delay statistic of 100 inverter chain.

<table>
<thead>
<tr>
<th>metrics</th>
<th>Technology node</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>90nm</td>
</tr>
<tr>
<td>Mean</td>
<td>4.34E-09</td>
</tr>
<tr>
<td>Sigma</td>
<td>5.75E-11</td>
</tr>
</tbody>
</table>

Fig. 13 shows comparative analysis of delay variation at different technology nodes.

4.3 Simulation results with process variation and body bias
When the body bias is applied to the inverter, it significantly reduces the deviation. The simulation results with different body bias voltage when applied to the inverter under process variation is given in Table 4. It can be seen from the simulation results (mean, sigma and variance) that value of delay distribution (sigma value) reduces significantly over to the sigma value of the delay when the body bias voltage was not applied. The comparative, analysis at different value of mean value for varying body bias voltage is given in Fig. 14.
Table 4: Illustration of metrics under PV at different body-bias voltage.

<table>
<thead>
<tr>
<th>Technology</th>
<th>Metrics</th>
<th>Vb, Vbn</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>µ</td>
<td>3.97E-11 3.79E-11 3.61E-11</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>4.05E-22 3.68E-22 3.41E-22</td>
</tr>
<tr>
<td></td>
<td>Sigma</td>
<td>4.05E-22 1.92E-11 1.85E-11</td>
</tr>
<tr>
<td>45</td>
<td>Mean</td>
<td>1.14E-12 1.83E-12 2.54E-12</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>1.16E-22 1.15E-22 1.13E-22</td>
</tr>
<tr>
<td></td>
<td>Sigma</td>
<td>1.08E-11 1.07E-11 1.06E-11</td>
</tr>
<tr>
<td>65</td>
<td>Mean</td>
<td>2.20E-11 2.00E-11 1.78E-11</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>4.14E-23 4.32E-23 4.10E-23</td>
</tr>
<tr>
<td></td>
<td>Sigma</td>
<td>6.44E-12 6.58E-12 6.41E-12</td>
</tr>
<tr>
<td>90</td>
<td>Mean</td>
<td>2.11E-11 1.85E-11 1.61E-11</td>
</tr>
<tr>
<td></td>
<td>Variance</td>
<td>1.19E-23 1.08E-23 9.03E-24</td>
</tr>
<tr>
<td></td>
<td>Sigma</td>
<td>3.45E-12 3.28E-12 3.00E-12</td>
</tr>
</tbody>
</table>

Fig. 14: Comparison of mean value of delay for various body-bias voltage.
V. CONCLUSION

In this paper an exhaustive analysis of process variation with technology scaling is done by considering inverter chain and simulated under process variations at different technology. To achieve this, different size inverter chain (100 inverters and 500 inverters) is considered and implemented on Tanner. These inverter chains are simulated under process variations with different technology files (90nm, 65nm, 45nm, 32nm). From the simulation results it can be seen that this deviation increase by more than $6\times$ in the designs at 32nm technology over the design at 90nm technology.

Further, body bias at different technology node is evaluated to compensate the effect of process variation. It is observed that this deviation increase by more than $6\times$ process variation with different technology nodes (90nm, 65nm, 45nm, 32nm). From the simulation results it can be seen that this deviation increase by more than $6\times$ in the designs at 32nm technology over the design at 90nm technology.

It is observed that body bias technique significantly reduces the effect of process variation.