2019 29th International Conference on Field Programmable Logic and Applications (FPL)

Analysis of Performance Variation in 16nm FinFET FPGA Devices

Konstantinos Maragos, Endri Taka, George Lentaris, Ioannis Stratakos, Dimitrios Soudris
National Technical University of Athens

***Abstract*—Process variability is a challenging fabrication issue impacting, mainly, the reliability and performance of chips. Variability is already present in current technology nodes and is expected to become even more significant in the future. In this work, we focus on the study of performance variation in 16nm FinFET FPGAs. We devise a comprehensive assessment methodology based on multiple programmable sensors with diverse resource and delay characteristics. Additionally, we consider various voltage and temperature conditions and decouple variability to systematic and stochastic. The experimental results on Zynq XCZU7EV show up to 7.3% intra-die variation increasing to 9.9% for certain operating conditions. Our approach demonstrates that logic and interconnect resources present different variability, slightly uncorrelated, which highlights the necessity and way towards more sophisticated mitigation methods/tools.**

***Index Terms*—FPGA, Process Variability, Ring Oscillator**

I. Introduction

The well-known trend of further transistor scaling becomes more challenging and economically inefficient over the years [1]–[3]. As transistors reach the fundamental atomic dimensions and chip density increases, controlling process variability becomes more arduous and leads to great deviations in the electrical properties of transistors, such as threshold voltage and channel length [4]. These deviations translate to variations in maximum frequency and leakage power between supposedly identical dies/chips (inter-die variability), or between areas inside a single die/chip (intra-die variability). Process variation is classified into two main types: *systematic* and *stochastic* [4]–[6]. Systematic refers to deterministic, spatially correlated variations attributed to layout effects and equipment shifts during the manufacturing process (e.g., photoresist development, etching) [7]. Systematic sources of variation result in high likelihood for neighboring devices to present similar electrical properties. In contrast, stochastic refers to uncorrelated, unpredictable variations occurring in atomic scale due to stochastic fluctuations in the process (e.g., random dopant fluctuation, line-edge roughness) [8]. With stochastic variation, the properties of each device vary independently.

The FPGA is an appealing platform for assessing process variability, due to its regular architecture and the ability to program every single resource of its fabric at very low level. This enables performing built-in-self-tests (BIST) via the deployment of custom sensors and assessing the performance variation across the fabric [9]–[12]. The analysis of process variability and performance variation under different operating conditions, i.e., voltage and temperature, plays a key role for the implementation of methods targeting performance enhancement. There exists a number of works in the literature demonstrating the performance improvement by exploiting the process variability in FPGAs, either in-the-field via frequency/voltage scaling methods [13], [14] or by adapting the computer-aided design (CAD) tools to the specifics of the underlying chips [11], [15]–[17].

In this work, we focus on the analysis of performance variation in commercial state-of-the-art 16nm FinFET FPGAs. We assess the performance variation in logic and interconnect resources and provide variability maps, which we create by measuring custom sensors mapped across the FPGA fabric. We present a comprehensive assessment methodology based on: i) the employment of multiple type of sensors with diverse logic and interconnect characteristics, ii) the decoupling of variability into systematic and stochastic, and iii) the assessment of variability under different voltage and temperature conditions. In summary, the main contributions of this work are:

- The first variability analysis in the literature regarding 16nm FinFET commercial FPGAs.
- A comprehensive evaluation of performance variation under diverse voltage and temperature conditions.
- Presentation of results demonstrating unequal variability in logic and interconnect resources. These results highlight the importance of multifaceted evaluation of process variability and give insights for future implementations of more accurate methods/tools for variability exploitation.

According to our experimental results on four identical Zynq XCZU7EV FPGAs, we measured up to 7.3% and 8.3% intra- and inter-die variability, which increases to 9.9% and 12%, respectively, under particular operating conditions. Comparing the variability maps derived by sensors of varying mixture in logic and interconnect parts, we calculated relatively weak correlation with 3.6% maximum error in performance estimation.

The structure of the paper is as follows: Section 2 presents the related work, Section 3 describes the proposed methodology for analyzing variability, Section 4 provides our experimental results and, finally, Section 5 draws the conclusions.

II. Related Work

Performance variation in commercial FPGAs has been studied by several works in the past. The most established method relies on ring oscillator (RO) sensors. In [9] and [11], the authors employed ROs to analyze the stochastic and systematic intra-die process variability in 90nm Cyclone II and 65nm

Virtex-5 FPGAs, respectively. Furthermore, in [10], [13], [18], [19] the authors used ROs to measure the intra-die variation in 90-28nm Spartan-3E, Virtex-4 ,Virtex-5 and Zynq FPGAs.

Another method for variability evaluation is based on shadow registers. In [20], the authors evaluated the delay variation of 336 logic paths on a 65nm Virtex-5 FPGA by placing additional shadow registers alongside the main paths' registers. To estimate the minimum delay of the respective paths, they increase finely the clock frequency until an error is detected in the comparison between the data of main and shadow registers. In [21], using the method of negative-skewed shadow registers, they evaluated the delay of three different logic paths of a floating point adder circuit which was placed in five different locations on two 130nm Virtex-II FPGAs.

In [12], an alternative technique is proposed for the evaluation of delay variability in FPGAs. The key idea is based on the placement of a combinatorial circuit under test (CUT) between a launch and a sampling register. A clock generator drives the clock of the registers and a stimuli generator provides inputs to the CUT. While stepping up the frequency, a custom circuit compares the output of the CUT with the data of the sampling register to detect the occurrence of timing errors. Consequently, the actual delay of the CUT is derived. Utilizing this technique, they measured the delay variation of LUTs, carry-chain units and embedded multipliers in Cyclone II and Cyclone III FPGAs. Similarly in [22], by using the same method they measured the intra-die delay variation of 1024 logic CUTs in 12 65nm Virtex-5 FPGAs.

Contrary to the aforementioned, the differentiating parts of our work are: i) we study the performance variation in 16nm finFET FPGAs under various voltage and temperature operating conditions, ii) we evaluate process variability in a multifaceted fashion considering diverse types of RO and interconnect sensors, iii) we analyze systematic and stochastic variability for both logic and interconnect resources, and iv) we perform correlation analysis on the variability results derived by the different sensors demonstrating the inconsistency in variation between the logic and interconnect resources.

III. Methodology

In this section, we describe our methodology for the analysis of variability. We assess the performance variation of configurable logic blocks (CLBs) and interconnects, which are the most prevalent resources in the FPGA fabric. The proposed methodology is based on the generation of multiple variability maps characterizing the silicon quality of the underlying FPGA and the performance variation under various operating conditions (voltage, temperature). The variability maps are extracted by monitoring multiple small sensors deployed across the FPGA fabric. For our analysis, we employ 16nm Zynq XCZU7EV FPGAs consisting of the Processing System (PS) and Programmable Logic (PL) parts (Fig. 1). The PS comprises a system of multi-core ARM CPUs (Cortex-A53, -R5) along with an embedded GPU and variety of peripherals, while the PL includes traditional resources of the FPGA fabric, i.e., CLBs, RAMBs, DSPs, etc. In our case, we use the ARM A53 CPU to control the operation of the sensors, collect their data and forward them to an external Host PC for further analysis.

A. Sensor Design & Network

Similar to other works [10], [13], the design of our sensors follows the ring oscillator (RO) approach. An RO is an asynchronous loop of N-stage inverter gates, where N is an odd number, such that a square wave signal is generated at the output. Upon measuring the rising edges of that generated signal for a predefined period of time T, we can determine the actual delay of the loop. The actual delay depends on the electrical properties of the region where the RO is mapped on. By placing multiple identical ROs across the FPGA fabric, we can evaluate the speed of the corresponding regions and hence, calculate the intra-die variability. It is essential that all the employed ROs must be constructed by the exact same resources (LUTs, pass-through DFFs) and routing connections to obtain precise results. To ensure this identity, we build an RO soft-macro block which is replicated onto multiple regions while applying the equivalent relative physical constraints for placement and routing. This leads to the creation of a sensor network covering the entire FPGA fabric. Both our sensor macro block and the sensor network are fully parametric in terms of RO stages, number of sensors and mapping locations.

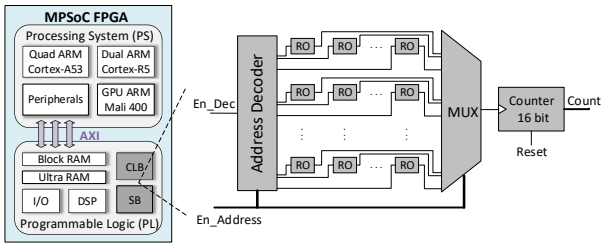


Fig. 1: Block diagram of proposed architecture.

To measure the delay of each individual RO sensor, we use a shared up-counter multiplexed with all the sensors of the network. The operation of ROs is performed sequentially to avoid potential voltage drops affecting the evaluation results. The sensor network architecture is illustrated in Fig. 1. The selection of a specific RO sensor for operation is specified by the address decoder unit which controlled by the ARM CPU via the *En_Address* signal. Notice that the same signal is also applied as select signal in the multiplexer driving the shared counter. The communication between the ARM CPU with the sensor network is realized via an AXI-Lite interface. The operation period T where each RO remains active is calculated by the private timer of ARM CPU and is selected to be 50 μs as proposed by [9] to avoid self-heating phenomena [23] and mitigate the error in the measurement process. When including the non-ideal timer operation, the quantization issue due to the non-aligned operation of timer and sensors [24], and the micro-fluctuations in voltage and temperature that affect RO operation, the overall measurement error is calculated to less

TABLE I: STA delay of various sensor configurations.

sensor conf.	delay of logic resources			delay of interconnects			total (ps)
	LUTs	DFFs	Total	intra-CLB	inter-CLB	Total	
7st_1sb	707 ps	463 ps	65.4%	295 ps	325 ps	34,6%	1790
7st_2sb	707 ps	463 ps	36.3%	295 ps	1762 ps	63,7%	3227
7st_inter	-	-	-	-	1437 ps	100%	1437
5st_1sb	582 ps	309 ps	67%	196 ps	244 ps	33%	1331
5st_2sb	582 ps	309 ps	37,2%	196 ps	1305 ps	62,8%	2392
5st_inter	-	-	-	-	1061 ps	100%	1061

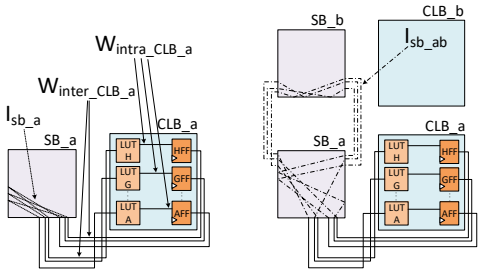
than 0.2%. To alleviate this error, we determine the RO delay as the average value over 10 consecutive 50 μs runs.

B. Assessment Approach

Our variability assessment methodology includes i) various sensor configurations, ii) decoupling of variability to systematic and stochastic, iii) diverse voltage and temperature conditions. More details are given in the following paragraphs.

1) Variety of sensor configurations: Owing to the parametric implementation of our RO sensor, we utilize various configurations with different resource and delay characteristics. This serves a twofold purpose. First, we need to investigate how the derived variability results are affected by the footprint of the sensor. Second, we need to analyze the impact of variability on logic and interconnect resources. To do so, we utilize RO configurations with different fraction of logic and interconnect resources. The delay attributed to logic and interconnect resources is specified via the custom mapping of the sensor on the FPGA fabric by using the floorplan utility of Xilinx Vivado tool. We clarify that the term "interconnects" in our work refers to intra-, inter-CLB wires and switch boxes (SB). Since the Vivado tool does not distinguish between inter-CLB wires and SBs, we put their delays together under the same category of inter-CLB interconnects. In Table I, we provide details regarding sensor configurations. We employ ROs of N=5 and N=7 stages using different routing topologies with M=1 and M=2 SBs, named "*Nst_Msb*". We note that, all SB-SB routing is based on short wire segments (direct connection between the SB tiles) [25]. For each sensor configuration, we distinguish the delay attributed to logic and interconnect resources as reported by static timing analysis (STA) tool. Notice that the configurations consisting of a single SB imply less routing, hence the logic delay dominates the total delay, i.e., 65,4% and 67% for 7 and 5 stages, respectively. The opposite applies in the case of two SBs, i.e., interconnects dominate the total delay with 63.7% and 62,8%, respectively.

An important feature of our sensors is that the configurations having the same RO stages (e.g., 7st_1sb, 7st_2sb) share the exact same CLB resources for a given location. Fig. 2 depicts such a case, with $W_{intra\_CLB\_a}$ and $W_{inter\_CLB\_a}$ being identical in both sensors. Taking advantage of our RO feature, by subtracting the measured delays of the two different sensors, we isolate and calculate the delay of the remaining inter-CLB interconnects, i.e., $I_{sb\_ab} - I_{sb\_a}$. The I_{sb_ab} and I_{sb_a} have been carefully selected to avoid any overlap between their routing. Thus, we create extra sensors, named *7st_inter* and *5st_inter* in Table I, which enable us to measure the interconnects alone.



(a) Nst_1sb (b) Nst_2sb

Fig. 2: RO architectures with identical CLB resources and different routing resources.

2) Decoupling of variability: We decouple the total measured variability into systematic and stochastic to study the impact of each individual type separately. In presence of variability, the delay of a path can be expressed as a random variable following the first order canonical form [26]:

$$T_d = T_d^S + T_d^R \tag{1}$$

where T_d^S represents the systematic part and T_d^R the random/stochastic part. For the purpose of our analysis, we model the FPGA fabric as a X-Y grid, with each point representing a sensor. The T_d^S is spatially correlated and can be expressed as a function of (x, y), while T_d^R has no spatial correlation and can be expressed as a random variable following a normal distribution $(0, \sigma^2)$. According to the grid model [27], we assume that all resources of a sensor have perfectly correlated spatial variation (they are closely located, our assumption was also verified in practice). When specifically considering the interconnects, subtracting RO delays to derive the Nst_inter sensor results also in a random variable and is expressed as:

$$\begin{aligned} T_{inter} &= T_{inter}^S + T_{inter}^R \\ &= \left(T_{sb\_ab}^S - T_{sb\_a}^S\right) + \left(T_{sb\_ab}^R - T_{sb\_a}^R\right) \end{aligned} \tag{2}$$

because T_{inter} is generated by the subtraction of two random variables (from Nst_2sb and Nst_1sb) describing delay paths with systematic and stochastic parts (eq. 1). Notice that these variables/paths have overlapped parts (identical CLB resources, Fig. 2), but their subtraction leads to a random variable, T_{inter}, derived by two independent parts (T_{sb_ab} and T_{sb_a} with no physical overlap). The subtraction inside T_{inter}^S expresses accurately the systematic part of T_{inter} as the spatial correlation of $T_{sb\_ab}^S$ and $T_{sb\_a}^S$ is assumed 1 (as mentioned above). That is to say, directly subtracting RO delays is sufficient for calculating T_{inter}^S. However, when considering the stochastic parts, since T_{inter}^R is the difference of two *statistically independent* variables, the subtraction of individual RO delays would typically follow normal distribution $(0, \sigma_{sb\_ab}^2 + \sigma_{sb\_a}^2)$ and would not be correct for our analysis. Instead, we need to derive $\sigma_{sb\_ab}^2 - \sigma_{sb\_a}^2$ [27]. Thus, we first calculate the variances of Nst_2sb and Nst_1sb, independently for each RO set, and we subtract them afterwards.

Additionally to the above analysis of the interconnects' delay, we consider two distinct methods for the decoupling of

TABLE II: Measured total performance variation results for nominal conditions.

Sensor	device 1 (psec)				device 2 (psec)				device 3 (psec)				device 4 (psec)				inter-die
	σ_{total}	μ	vs STA	$range/min$	σ_{total}	μ	vs STA	$range/min$	σ_{total}	μ	vs STA	$range/min$	σ_{total}	μ	vs STA	$range/min$	$range/min$
7st_1sb	11.1	1255.3	29.9%	5.51%	15.8	1227.2	31.4%	6.55%	14	1243.3	30.5%	6.15%	5.7	1216.3	32.1%	3.95%	8%
7st_2sb	19	2172.5	32.7%	5.05%	21.7	2133	33.9%	4.95%	19.7	2144.7	33.5%	4.77%	7.4	2112.7	34.5%	2.92%	7.01%
7st_inter	8.9	917.2	36.2%	4.82%	6.4	905.8	37%	3.85%	6.2	901.3	37.3%	4.04%	3	896.3	37.6%	2.62%	6.44%
5st_1sb	8.2	922.4	30.7%	5.8%	11.8	901.4	32.3%	7.3%	10.4	913.3	31.4%	6.29%	4.4	893.1	32.9%	4.09%	8.31%
5st_2sb	14.1	1611.7	32.6%	5.02%	16.1	1581.4	33.9%	4.96%	14.7	1589.9	33.5%	4.82%	5.6	1566.2	34.5%	2.94%	6.93%
5st_inter	6.8	689.4	35%	5.19%	4.8	680	35.9%	4.15%	4.7	676.6	36.2%	3.98%	2.4	673.1	36.6%	2.76%	6.83%

variability, as proposed in the literature: regression method [9] and down-sampled moving average estimator (DSMA) [11]. In regression method, systematic variation is modelled by a quadratic polynomial function of x and y, where coefficients are computed by a least-square curve fitting algorithm. In contrast, DSMA applies a moving average window across the die to calculate the values for each (x, y); a 5x5 window size was found to be the optimal according to [11] and our own experimentation. In either methods, the residuals are utilized to estimate the stochastic variability. We test both methods and choose the most accurate one for our analysis.

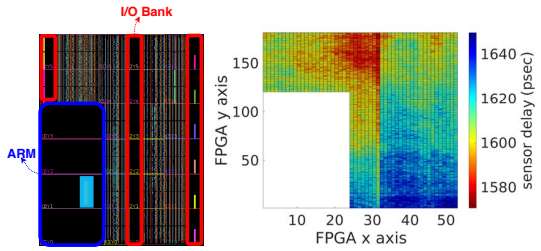
3) Diverse operating conditions: We assess the performance variation under different voltage and temperature operating conditions. The assessment regards all sensor configurations and decoupling of variability. To perform voltage scaling, we utilize the on-board power controller (IRPS5401M) of the ZCU104 board to alter the supply voltage, V_{ccint}, via the I2C interface of Zynq PS. To modify temperature, we employ a thermal chamber of uniform thermal distribution.

IV. Variability Analysis

Following the methodology described in Section III, we perform the assessment of variability in four supposedly identical Zynq XCZU7EV FPGAs. For each FPGA, we generate multiple variability maps by utilizing the different sensors of Table I. For Nst_1sb RO configurations we deploy 13200 sensors, while for the counterparts Nst_2sb we deploy 6600 sensors due to their larger footprint on the fabric (Fig. 2). All sensors are uniformly placed over the fabric to sufficiently cover the die. An example of variability map with the corresponding floorplan view for the arbitrary selected device 1 is illustrated in Fig. 3. The red color denotes the faster regions of the device (smaller RO delay) while the blue color denotes the slower regions. We observe a smooth distribution and change of performance across the die due to systematic variation, with a noticeable discontinuity in the middle column of our variability map (Fig. 3b), which is explained by observing the floorplan of the FPGA (Fig. 3a): the corresponding physical area is utilized by I/O Banks, not considered in this work.

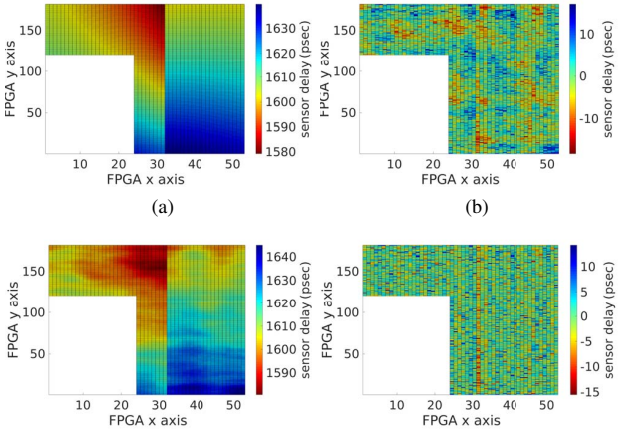
We begin our analysis by examining fixed nominal conditions: supply voltage, $V_{ccint} = 0.85V$, and junction temperature, $T_j = 30°C$. The values of V_{ccint} and T_j are monitored by Xilinx integrated system monitor (SYSMON). The measured error in V_{ccint} among the four devices is negligible, i.e., 0.3%.

Table II provides detailed total variability results. We report the mean sensor delay (μ), the standard deviation (σ_{total}), the difference between the STA estimation (Table I) versus the actual mean sensor delay (vs STA), as well as the estimation



(a) Floorplan (b) Variability map

Fig. 3: FPGA floorplan and variability map of sensor $5st\_2sb$.

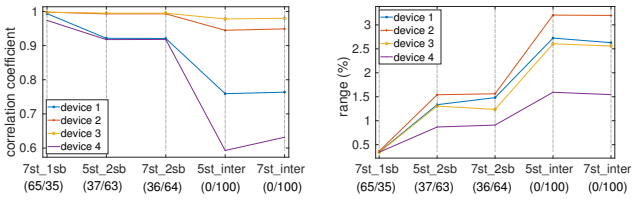


(a) (b) (c) (d)

Fig. 4: Systematic and stochastic variability maps of sensor $5st\_2sb$ using regression (a),(b) and DSMA (c),(d) methods.

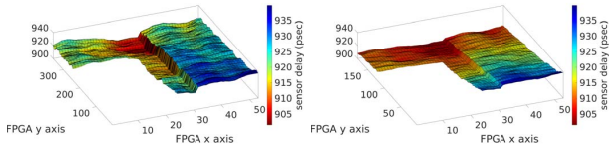
of variability expressed by the $range/min$ metric, where $range = max - min$ refers to the maximum and minimum sensor delays. The first important observation is the great difference between the STA and the actual measured delay of the sensors, which tends to rise as the portion of delay attributed to interconnects increases, reaching up to 37.6% for $7st\_inter$. Essentially, this indicates that the STA tool introduces more pessimism to interconnects rather than logic.

Regarding variability, the highest intra-die variation is measured with the smaller RO configuration ($5st\_1sb$), reaching up to 7.3%. By comparing the $7st\_inter$ and $5st\_1sb$ sensors which have similar mean delay (implying fair comparison), we observe that interconnects present smaller intra-die variation in all devices, with the difference being 0.98-3.45%. The same applies for inter-die variability as well; inter-die variation decreases when considering sensors with more interconnects, however, to strongly support this statement a greater number



(a) Correlation (b) Performance estimation error

Fig. 5: Correlation results and performance estimation error between $5st\_1sb$ (67/33) and rest sensors (logic/interconnect).



(a) (b)

Fig. 6: Systematic variability maps of $5st\_1sb$ (a) and 7_inter (b) sensors (device 1, common scale)

of devices is required. Note that Table II includes a negligible error in the σ_{total}/μ of interconnects, i.e., 0.03%, due to our approximation in the stochastic parts of variation (eq. 2), and hence, more accurate results are derived by the separate analysis of systematic and stochastic variability in Sec. IV-A.

A. Analysis of Systematic and Stochastic Variation

A step further in our analysis, as described in Section III-B, we decouple the variability into systematic and stochastic using both regression and DSMA methods. To address the discontinuity in variability maps (Fig. 3), we empirically insert null columns until the systematic impact on the stochastic maps get minimized. In Fig. 4, we show the extracted systematic (Fig. 4a, 4c) and stochastic (Fig. 4b, 4d) variability maps of device 1. Comparing the two methods, we deduce that DSMA formulates more precisely the systematic shape of the initial map (Fig. 3b) and highlights the random nature of stochastic variation. Thus, in the remainder of the paper, we continue our analysis based on the DSMA method.

Our study of systematic variation is based on correlation analysis. We use the Pearson coefficient to correlate the systematic variability maps derived by the different sensors. We note that, for this calculation, the shown discontinuity in our maps was removed manually to avoid artificial biasing and obtain more accurate/fair correlation coefficients. In Fig. 5a, using as reference the smaller RO sensor, $5st\_1sb$, we depict how the correlation varies with respect to the other sensors. The correlation weakens as the ratio of logic/interconnects sensor delay decreases reaching down to 0.59 for interconnect-only sensors. Moreover, in Fig. 5b we illustrate the maximum difference (error) in relative performance estimation between the various systematic variability maps. Keeping the $5st\_1sb$ sensor as reference, we calculate that the average error of the entire map ranges in 0.01-1.22% and extends to 0.34-3.2% considering particular (x, y) points on the maps. Fig. 6 depicts the systematic variability maps of $5st\_1sb$ and 7_inter sensors

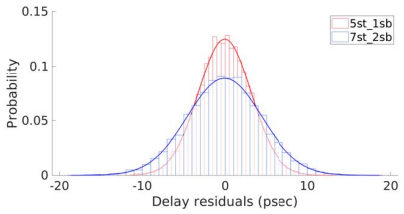


Fig. 7: Probability distribution of stochastic variation.

TABLE III: Systematic and stochastic variability results.

Dev.	Metric	7st_1sb	7st_2sb	7st_inter	5st_1sb	5st_2sb	5st_inter
Systematic							
1	$range/min$	4.28%	4.05%	4.01%	4.25%	4.07%	4.12%
2		5.47%	4.11%	2.78%	5.54%	4.14%	2.83%
1-4	$range/min$ (inter-die)	6.69%	5.73%	5.56%	6.73%	5.86%	5.69%
Stochastic							
1	$3\sigma_{rand}/\mu$	0.90%	0.62%	0.79%	1.04%	0.70%	0.84%
2		0.86%	0.59%	0.77%	0.98%	0.66%	0.81%

for device 1. There is a noticeable difference in their maps denoting that interconnects (mainly wires) follow a different process than logic components (mainly transistors). Overall, these results demonstrate the importance of having multiple sensors, with diverse characteristics, to accurately analyze variability and predict the performance variation in FPGAs.

Having calculated the systematic variation, we continue with the analysis of stochastic variation. First, we verify that the residuals derived by the subtraction of total measured and extracted systematic variability maps, indicating the stochastic variation, follow a Gaussian distribution in all cases (Fig. 7). Table III shows representative systematic and stochastic variability results. The maximum systematic variability is calculated to 5.5% ($range/min$) while the maximum stochastic to 1.1% ($3\sigma_{rand}/\mu$). Comparing the $5st\_1sb$ and $7st\_inter$ sensors which have similar actual delays, we observe that systematic variation in $7st\_inter$ is lower than $5st\_1sb$ in all tested devices with the difference being in the range of 0.24-2.76%. The same applies for stochastic variation, i.e., 0.18-0.25%. Considering all sensors, our results verify that as the footprint of the sensor increases and total delay becomes higher, the σ^2_{rand} increases (Fig. 7) but stochastic variation as ratio of mean value ($3\sigma_{rand}/\mu$) attenuates due to averaging over multiple gate and interconnect delays [11].

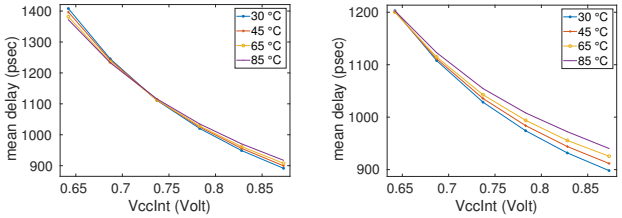
B. Variability Under Voltage and Temperature Change

We continue the analysis of performance variation considering diverse voltage and temperature operating conditions: V_{ccint} in range of 0.640-0.875 V and T_j in 30-85 C°. For each specific V_{ccint} value, we retrieve variability maps for different T_j with a step of ∼20 C°. Before we proceed to the analysis of results, for better understanding, we mention the alpha-power law model of the CMOS logic gate delay [28]:

$$t_d = \frac{V_{ccint}}{K \cdot (V_{ccint} - V_{th})^a} \quad (3)$$

where, V_{ccint} is the supply and V_{th} is the threshold voltage, a is a fitting parameter and K is a process dependent parameter.

Fig. 8 illustrates the mean sensor delay for all different voltage-temperature combinations for representative RO



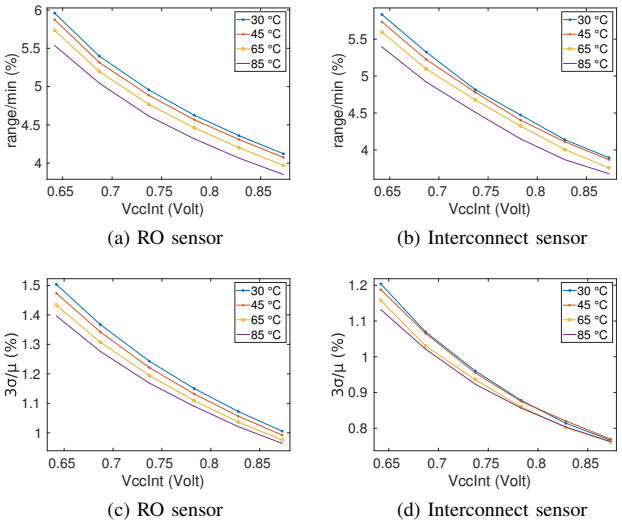
(a) RO sensor (b) Interconnect sensor

Fig. 8: Maximum performance w.r.t. voltage and temperature.

($5st\_1sb$) and interconnect ($7st\_inter$) sensors in device 1. In RO sensor (Fig. 8a), the delay decreases almost linearly ($a \approx 1$) for higher V_{ccint} values (0.81-0.875V) and then with a factor of $a \approx 1.2$ for lower V_{ccint} [29]. The estimation of a is calculated by fitting models in MATLAB using measurements with fine change of V_{ccint}. Below a certain V_{ccint} value (0.72V), the *temperature inversion* phenomenon occurs [30]; delay decreases with elevated T_j. The temperature inversion point as well as the value of fitting parameter a vary depending on the sensor configuration. In sensors with higher portion of logic delay (greater amount of transistors), we measure higher performance degradation and temperature inversion manifests in higher V_{ccint} values. Fig. 8b depicts the case of interconnect sensor, which is composed mainly by wires and the transistors residing in the SBs of the FPGA. In contrast to the RO sensor, the interconnect sensor shows lower performance degradation with voltage decrease and a higher degradation with temperature increase; resistance of wires becomes higher [31]. Considering all devices and sensors, the performance degradation due to voltage ranges up to 33.9% ($7st\_inter$) - 57.9% ($7st\_1sb$), while the degradation due to temperature up to 2.9% ($5st\_1sb$) - 4.8% ($5st\_inter$).

Fig. 9 presents the intra-die systematic and stochastic performance variation for the reported voltage-temperature conditions. In both cases, variability increases with the decrease of V_{ccint}. This was expected because, according to eq. 3, as V_{ccint} scales down, the delay of slower transistors (higher V_{th}) increases relatively higher than faster transistors (lower V_{th}) thus, leading to higher variability. On the other hand, variability decreases with the elevation of T_j: the V_{th} decreases almost linearly to T_j increase [31], i.e., the V_{ccint}–V_{th} increases more for slow transistors hence, decreasing their t_d more (see eq. 3) than fast transistors. Considering all devices and operating conditions, the systematic variability is increased up to 5.9% ($7st\_inter$) - 7.3% ($5st\_1sb$) and the stochastic variability is increased up to 1.41% ($7st\_inter$) - 1.53% ($5st\_1sb$). The corresponding intra- and inter-die total variability is increased up to 7.4-9.9% and 9.5-12%, respectively.

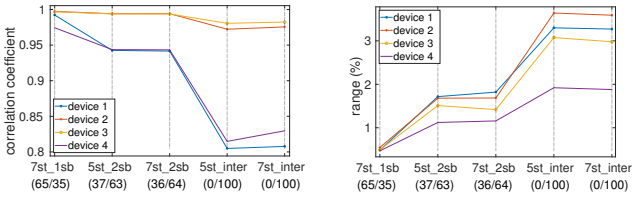
Fig. 10a shows the Pearson coefficients and the maximum performance estimation error between the $5st\_1sb$ reference and the rest sensors for $V_{ccint} = 0.64V$. In contrast to the corresponding plots in Fig. 5 for nominal conditions ($V_{ccint} = 0.85V$), the correlation between RO sensors remains almost the same, however, the correlation with the interconnect sensors has been greatly increased to 0.82 (from 0.59). The latter is explained by the fact that, with voltage under-scaling



(a) RO sensor (b) Interconnect sensor

(c) RO sensor (d) Interconnect sensor

Fig. 9: Systematic (a,b) and stochastic (c,d) variability as a function of voltage and temperature.



(a) Correlation (b) Performance estimation error

Fig. 10: Correlation results and performance estimation error between $5st\_1sb$ (67/33) and other sensors for 0.640V, 30°C.

the change in interconnect delay is attributed mainly to the transistors residing in SBs, hence, the variability maps tend to follow the behavior of transistors. Nevertheless, note that even though the correlation is improved, the error in performance estimation (Fig. 10b) is increased to 3.6% (from 3.2%).

V. Conclusion

In this work, we studied the performance variation in 16nm FinFET commercial FPGAs. We employed multiple type of sensors to measure variability in logic and interconnect resources, we analyzed the systematic and stochastic process variability, we evaluated the impact of diverse voltage and temperature conditions. Our experimental results showed up to 9.9% intra-die and 12% inter-die performance variation under certain operating conditions. Moreover, we deduced that logic and interconnect resources present different variation, with low correlation, and a maximum error of 3.6% in performance estimation. Our results highlight the importance of a multifaceted assessment of variability in FPGAs and provide insights for the implementation of more sophisticated mitigation methods.

VI. Acknowledgment

This work was partially supported by Xilinx and the "SDK4ED, Project No 780572" of the European Union's 2020 research and innovation programme.

REFERENCES

[1] ITRS, "International technology roadmap for semiconductors," http://www.itrs2.net/itrs-reports.html, 2015.

[2] S. Alam and G. Douglass, "Moore…or less?" https://www.accenture.com/t20180511T035547Z__w__/us-en/_acnmedia/PDF-61/Accenture-Moore-Less-POV.pdf, Accenture, 2017.

[3] S. Mittal, "A survey of architectural techniques for managing process variation," *ACM Computing Surveys*, vol. 48, no. 4, pp. 54:1–54:29, 2016.

[4] K. Bernstein, D. J. Frank, A. E. Gattiker, W. Haensch, B. L. Ji, S. R. Nassif, E. J. Nowak, D. J. Pearson, and N. J. Rohrer, "High-performance CMOS variability in the 65-nm regime and beyond," *IBM Journal of Research and Development*, vol. 50, no. 4.5, pp. 433–449, 2006.

[5] K. J. Kuhn, M. D. Giles, D. Becher, P. Kolar, A. Kornfeld, R. Kotlyar, S. T. Ma, A. Maheshwari, and S. Mudanai, "Process technology variation," *IEEE Transactions on Electron Devices*, vol. 58, no. 8, pp. 2197–2208, 2011.

[6] S. K. Saha, "Modeling process variability in scaled CMOS technology," *IEEE Design & Test of Computers*, vol. 27, no. 2, pp. 8–16, 2010.

[7] L.-T. Pang, K. Qian, C. J. Spanos, and B. Nikolic, "Measurement and analysis of variability in 45 nm strained-Si CMOS technology," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 8, pp. 2233–2243, 2009.

[8] E. Baravelli, A. Dixit, R. Rooyackers, M. Jurczak, N. Speciale, and K. De Meyer, "Impact of line-edge roughness on FinFET matching performance," *IEEE Transactions on Electron Devices*, vol. 54, no. 9, pp. 2466–2474, 2007.

[9] P. Sedcole and P. Y. Cheung, "Within-die delay variability in 90nm FPGAs and beyond," in *International Conference on Field Programmable Technology (FPT)*. IEEE, 2006, pp. 97–104.

[10] K. M. Zick and J. P. Hayes, "On-line sensing for healthier FPGA systems," in *International Symposium on Field Programmable Gate Arrays (FPGA)*. ACM, 2010, pp. 239–248.

[11] T. Tuan, A. Lesea, C. Kingsley, and S. Trimberger, "Analysis of within-die process variation in 65nm FPGAs," in *International Symposium on Quality Electronic Design (ISQED)*. IEEE, 2011, pp. 1–5.

[12] J. S. Wong, P. Sedcole, and P. Y. Cheung, "Self-measurement of combinatorial circuit delays in FPGAs," *ACM Transactions on Reconfigurable Technology and Systems*, vol. 2, no. 2, pp. 10:1–10:22, 2009.

[13] K. Maragos, G. Lentaris, and Soudris, "In-the-field mitigation of process variability for improved FPGA performance," *IEEE transactions on Computers*, vol. 68, no. 7, pp. 1049–1063, 2019.

[14] K. Maragos, G. Lentaris, I. Stratakos, and D. Soudris, "A framework exploiting process variability to improve energy efficiency in FPGA applications," in *Great Lakes Symposium on VLSI (GLSVLSI)*. ACM, 2018, pp. 87–92.

[15] L. Cheng, J. Xiong, L. He, and M. Hutton, "FPGA performance optimization via chipwise placement considering process variations," in *International Conference on Field Programmable Logic and Applications (FPL)*. IEEE, 2006, pp. 1–6.

[16] Z. Guan, J. S. Wong, S. Chaudhuri, G. Constantinides, and P. Y. Cheung, "A two-stage variation-aware placement method for FPGAs exploiting variation maps classification," in *International Conference on Field Programmable Logic and Applications (FPL)*. IEEE, 2012, pp. 519–522.

[17] Z. Guan, J. S. Wong, S. Chaudhuri, G. Constantinides, and P. Cheung, "Exploiting stochastic delay variability on FPGAs with adaptive partial rerouting," in *International Conference on Field-Programmable Technology (FPT)*. IEEE, 2013, pp. 254–261.

[18] Y. Pino, V. Jyothi, and M. French, "Intra-die process variation aware anomaly detection in FPGAs," in *International Test Conference (ITC)*. IEEE, 2014, pp. 1–6.

[19] H. Yu, Q. Xu, and P. H. Leong, "Fine-grained characterization of process variation in FPGAs," in *International Conference on Field-Programmable Technology (FPT)*. IEEE, 2010, pp. 138–145.

[20] P. Sedcole, J. S. Wong, and P. Y. Cheung, "Characterisation of FPGA clock variability," in *Computer Society Annual Symposium on VLSI (ISVLSI)*. IEEE, 2008, pp. 322–328.

[21] J. Li and J. Lach, "Negative-skewed shadow registers for at-speed delay variation characterization," in *International Conference on Computer Design (ICCD)*. IEEE, 2007, pp. 354–359.

[22] M. Majzoobi, E. Dyer, A. Elnably, and F. Koushanfar, "Rapid FPGA delay characterization using clock synthesis and sparse sampling," in *International Test Conference (ITC)*. IEEE, 2010, pp. 1–10.

[23] T.-Y. Chiang, B. Shieh, and K. Saraswat, "Impact of Joule heating on scaling of deep sub-micron Cu/low-k interconnects," in *Symposium on VLSI Technology. Digest of Technical Papers (Cat. No. 01CH37303)*. IEEE, 2002, pp. 38–39.

[24] K. M. Zick and J. P. Hayes, "Low-cost sensing with ring oscillator arrays for healthier reconfigurable systems," *ACM Transactions on Reconfigurable Technology and Systems*, vol. 5, no. 1, pp. 1:1–1:26, 2012.

[25] V. Betz, J. Rose, and A. Marquardt, *Architecture and CAD for deep-submicron FPGAs*. Springer Science & Business Media, 2012, vol. 497.

[26] C. Visweswariah, K. Ravindran, K. Kalafala, S. Walker, S. Narayan, D. Beece, P. Piaget, I. Zamek, J. Fan, J. Beetner, N. Venkateswaran, and J. Hemmett, "First-order incremental block-based statistical timing analysis," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 25, no. 10, pp. 2170–2180, 2006.

[27] D. Blaauw, K. Chopra, A. Srivastava, and L. Scheffer, "Statistical timing analysis: From basic principles to state of the art," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 27, no. 4, pp. 589–607, 2008.

[28] T. Sakurai and A. Newton, "Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas," *IEEE Journal of Solid-State Circuits*, vol. 25, no. 2, pp. 584–594, 1990.

[29] M. Wirnshofe, *Variation-aware adaptive voltage scaling for digital CMOS circuits*. Springer, 2013.

[30] R. Kumar and V. Kursun, "Reversed temperature-dependent propagation delay characteristics in nanometer CMOS circuits," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 53, no. 10, pp. 1078–1082, 2006.

[31] N. Weste and D. Harris, *CMOS VLSI design: a circuits and systems perspective*. Pearson Education, Inc., 2011.