Parameter-Specific Ring Oscillator for Process Monitoring at the 45 nm Node

L. T.-N. Wang, N. Xu, S.-O. Toh, A. R. Neureuther, T.-J. King Liu, B. Nikolic
Department of Electrical Engineering and Computer Sciences, University of California, Berkeley
Berkeley, CA 94720 USA

***Abstract*-Parameter-specific ring oscillator (RO) experimental results are reported, demonstrating the ability to electronically distinguish and quantify sources of variations from gate lithography focus, gate-to-active overlay, nitride contact etch stop layer (CESL) strain, and Shallow Trench Isolation (STI) stress. A 2% RO frequency change due to gate focus variations, a three-four nm overlay error, a 20% increase in RO frequency per 1 um increase in length of diffusion (LOD), and a 3% speed-up per 0.3 um change in STI width are measured. Typical standard-deviation/mean (σ/μ) among 36 ROs within-chip is 0.2-0.3%.**

1. INTRODUCTION

Ring oscillators (RO) are good Design-for-Manufacturability (DfM) monitors because they reflect circuit performance variability caused by variations from physical effects. They are small in area, simple to design, and easy to implement. Prior works have shown RO monitoring to be an efficient method to screen for possible causes and levels of layout-dependent variability in 90-nm, 65-nm, and 45-nm generation circuits [1][2][3][4][5]. In this paper, experimental RO frequency sensitivity is reported for gate lithography focus, gate-to-active overlay, nitride CESL-strain induced, and STI-stress induced monitors.

2. RO DESIGN AND TESTING

Each RO consists of 13 inverter stages and is designed following standard logic design rules. The inverters are designed with systematic pre-programmed variation in one or more layout parameters, as shown in Fig. 1, to capture changes in RO frequency due to gate lithography focus (2 layouts), gate-to-active overlay (5 layouts), CESL strain (8 layouts), and STI stress (3 layouts). Overlay and STI monitors have large capacitors inserted between each inverter stage to increase the sensitivity of RO frequency to gate width [6]. These capacitors are made large such that they are less susceptible to process variations.

In this work, the nomenclature used for the RO monitors is described in Table 1. The ROs are grouped into categories according to the process parameter that is being monitored: C= Control (with one dummy poly on each side of the gate), I = Isolated gate, F = Focus, O = Overlay, N=Nitride CESL strain, and S= STI stress. The naming convention is as follows: (1) category, (2) gate width (W for wide gate width, where W_{NMOS} = 300nm and S for short gate width, where W_{NMOS} = 130nm; W_{PMOS}/W_{NMOS} = 1.4), (3) drawn gate length (40-nm, 50-nm, and 60-nm), and (4) subscripts enumerating different variations of the same RO monitor. Monitor names without subscripts indicate that there is only one variation of that monitor. For example, $OS40_1$ represents an overlay monitor with a small gate width and a drawn gate length of 40-nm.

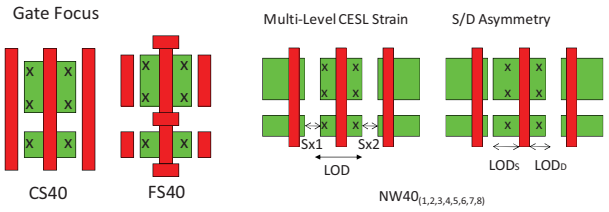

Fig. 1(a) Focus monitors Fig.1(b) CESL-induced strain monitors

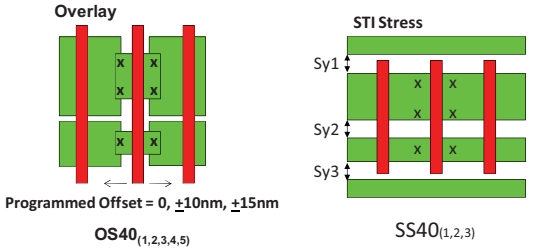

Fig.1(c) Overlay monitors Fig.1(d) STI-induced stress monitors

Fig. 1. Overview of parameter-specific monitors

Category	Gate Width(nm)	Drawn Gate Length(nm)
C = Control	W: W_{NMOS} = 300	40
I= Isolated gate	S: W_{NMOS} = 130	50
F = Focus		60
O=Overlay		
N= Nitride CESL		
S= STI		

Table 1. Nomenclature used for RO monitors

Each RO is replicated 12 times with other circuits in a local block, and the block is repeated 3 times within the 2x2 mm^2 chip (Fig. 2) so that there are 36 instances of each RO.

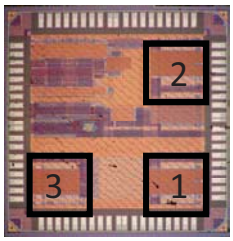

Fig. 2. 2x2mm^2 chip photo shows RO replicated 3 times across chip.

Measured RO frequency is collected for all 36 RO on each of 17 chips spaced at least one exposure field apart on one wafer

978-1-4244-5760-1/10/$26.00 ©2010 IEEE

(Fig. 3). Within-chip variation of average RO frequency between blocks has a range of ~1.5%. This range is on the order of 6(σ/μ) within a block (12 instances), which is ~1.2%. Hence, throughout this paper, chip-level averages (based on 36 instances) are used to assess the process-specific sensitivity for each RO design.

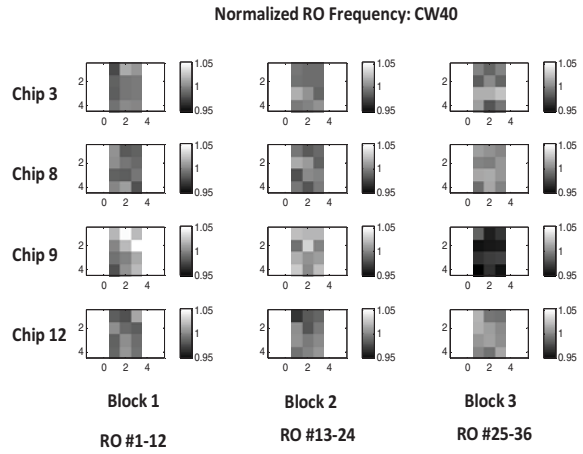

Fig. 3. Measured RO frequency normalized to the mean for that chip shows random within-block variation.

The average RO frequency, normalized to the average across 17 chip, is shown in Fig. 4 for drawn gate lengths of 40, 50 and 60 nm for two different gate widths as control-case layouts. The largest range is about 11.11%, occurring for the smallest gate length and smallest gate width, i.e. the smallest channel area. A strong systematic oscillatory behavior is observed, which reflects that the chips are most likely packaged in sequence along the rows of a wafer.

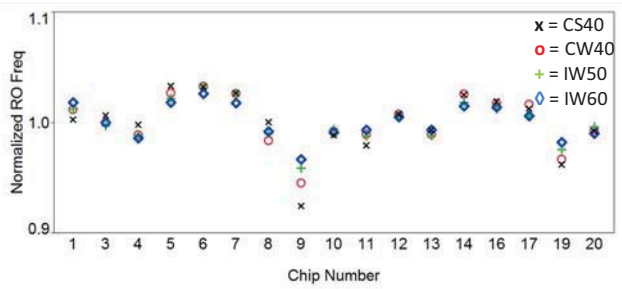

Fig. 4. Mean measured RO frequency normalized to the mean for that RO monitor designs CS40, CW40, IW50, and IW60 show across-wafer variation range of 11.11%, 8.84%, 7.35%, and 5.99%, respectively.

The $(\sigma/\mu)_{36}$ for each chip is shown in Fig. 5 for the three gate lengths and two gate widths control-case layouts in Fig. 4. Here $(\sigma/\mu)_{36}$ is about 0.2% for most chips, but is unusually high for chips 4 and 9. Excluding chips 4 and 9, the average $(\sigma/\mu)_{36}$ across 17 chips for CS40, CW40, IW50, and IW60 are: 0.30%, 0.20%, 0.19%, and 0.18%, respectively. For the same gate length, $(\sigma/\mu)_{36}$ for ROs with a smaller gate width is 0.1% higher than the $(\sigma/\mu)_{36}$ for ROs with the large gate width.

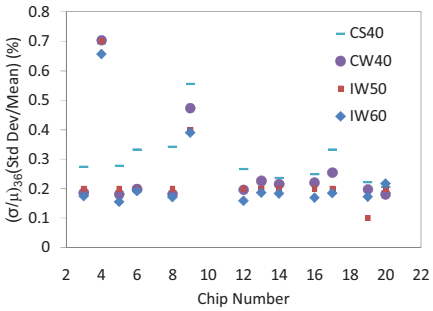

Fig.5. Within-chip $(\sigma/\mu)_{36}$ distribution shows that Chip 4 and 9 are noisy compared to typical chips.

3. PROCESS-SPECIFIC MONITORING WITH RO

A. RO Sensitivity to Gate Lithography Focus

Gate lithography focus RO monitors are designed such that the gate patterns have hammerheads and dummy gates placed at specific locations to exacerbate focus effects as guided by Pattern Matching and confirmed by rigorous aerial image simulations using Mentor Graphics Calibre [6]. RO frequency sensitivity from gate lithography focus monitors is then compared to that of control-case RO to extract the specific parameters.

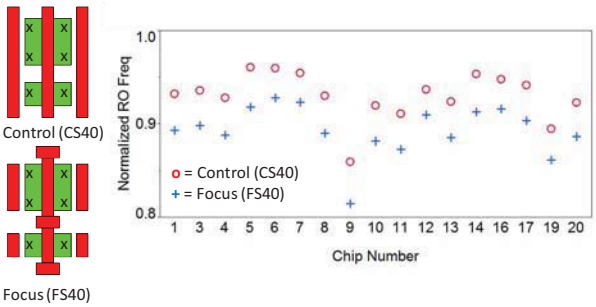

Fig. 6. Mean measured RO frequency shows that the hammerheads lower RO frequency by 4.05%. The across-wafer peak-to-peak variations for the control (CS40) and focus monitor (FS40) are 11 % and 13%, respectively.

The measurements show that the FS40 RO monitor has a 2% greater peak-to-peak range than that of the control-case RO monitor. The hammerheads slow down the RO frequency by 4%. $(\sigma/\mu)_{36}$ = 0.3% for both the CS40 and FS40 RO monitors.

For comparison purposes, across-wafer variation in RO frequency is simulated as follows: First, Process Variation (PV) Band simulation (provided by ST Micro) is used to extract changes in gate length (ΔL) using Mentor Graphics Calibre. Then, the RO frequency versus L relationship is simulated with SPICE. Using this relationship from SPICE, ΔL is then mapped to change in RO frequency (Δf_{RO}). Table 2 compares the simulation results with the experimental results.

Monitor Name	Simulated ΔL from PV	Simulated Δf_{RO} from PV and SPICE	Measured Across-Wafer Peak-to-Peak Variation (Δf_{RO})
CS40	11%	11%	11%
FS40	18%	16%	13%

Table 2. Comparison of simulated with measured across-wafer variation

As indicated by simulation results, the RO monitors are designed to show 1.5× increase in sensitivity compared to that of the control RO. However, measurements show only a 1.2× increase in sensitivity. This discrepancy may be attributed to the fact that there is little defocus in the gate lithography process.

B. RO Sensitivity to Gate-to-Active Overlay

Programmable overlay monitors ($OS40_{1,2,3,4,5}$) are designed based on lithographic rounding of an H-shaped active area (Fig. 7).

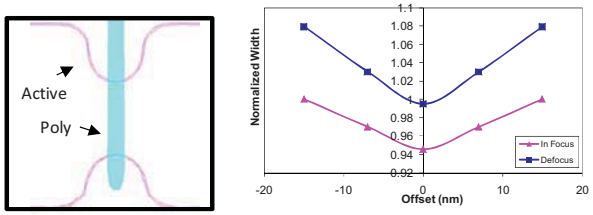

Fig. 7. Simulated aerial image using Mentor Graphics Calibre of overlay monitor shows that best alignment and focus produce minimum width.

Five ROs are designed with pre-programmed gate-to-active offsets at 0 nm, ±10nm, and ±15nm. A plot of the RO frequency versus programmed offset in Fig. 8 shows a parabolic shape; the location of the minimum quantifies overlay error.

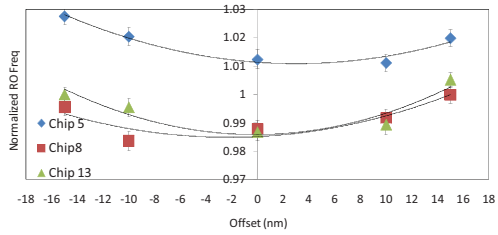

Fig.8. Overlay measurement results for 3 chips show a 1.5% difference in RO frequency and $(\sigma/\mu)_{36}$ =0.3%.

The frequency dip for best overlay is about 1.5%, compared to a $(\sigma/\mu)_{36}$ of 0.2% for each data point in the curve. The overlay versus chip number in Fig. 9 is double-humped following the average RO frequency versus chip number in Fig. 4.

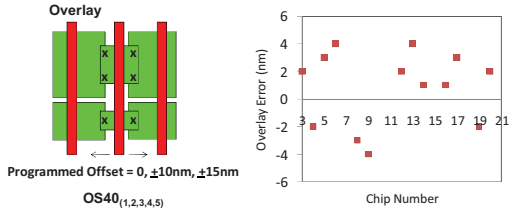

Fig.9. Summary of measured overlay error.

C. RO Sensitivity to Nitride CESL Strain

Nitride CESL strain monitors ($NW40_{1,2,3,4,5,6,7,8}$) are designed with 40 nm gates and varying lengths of source/drain diffusion (LOD). Normalization of raw data to simulation data to correct for parasitic effects shows that RO frequency can increase by 20% for long LOD compared to minimum LOD (Fig. 10) due to increased CESL-induced strain for larger LOD. Doubling the minimum LOD raises the frequency by 5% (>> $(\sigma/\mu)_{36}$ = 0.3%).

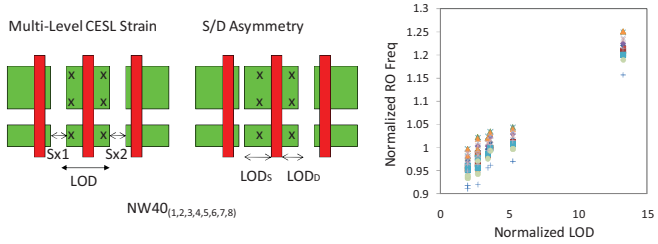

Fig. 10. RO Freq vs. LOD for 17 chips shows 20% increase in RO frequency for long LOD compared to minimum LOD.

Asymmetrical source/drain designs after correction for the change in source/drain parasitic capacitance show 3% greater speed-up for longer LOD on the source side versus the drain side (Fig. 11). These observed effects are significant, considering that the noise level of the measurement is $(\sigma/\mu)_{36}$ = 0.2% . This RO frequency speedup is due to the source of a device having a larger impact on transistor injection velocity than that of the drain [7].

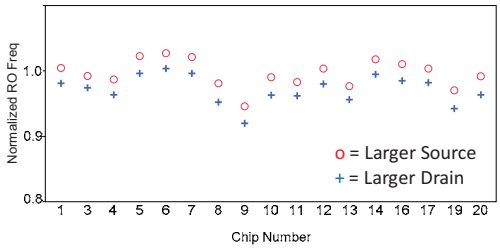

Fig.11. For the same LOD, asymmetrical monitors with bigger sources operate 3% faster than monitors with bigger drains.

D. RO Sensitivity to STI Stress

STI monitors ($SS40_{1,2,3}$) are designed with various lengths of STI between the fixed power rail and PMOS active, between PMOS active and NMOS active, and between NMOS active and the fixed ground rail. Increasing the STI by 0.3um on NMOS devices results in 3% (>> $(\sigma/\mu)_{36}$ = 0.3%) RO frequency increase, while the same change for PMOS devices has little impact (Fig. 12).

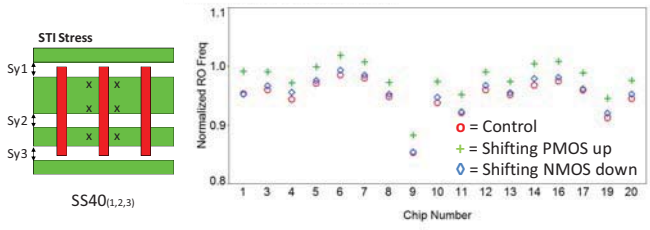

Fig. 12. Across-wafer RO frequency measurements for STI monitors show that shifting NMOS down by 0.3um has negligible

effects. Shifting PMOS up by 0.3um speeds up the RO frequency by 3%.

Since the NMOS devices show more sensitivity to STI stress, bulk silicon piezoresistance coefficients suggest that the devices are oriented with <100> channel direction [4] [8].

E. Gate length, Gate Oxide Thickness, and Doping Effects

Measurements under various operating voltage (V_{dd}) and temperature (T) conditions are used to distinguish the effects of gate length variation (ΔL), gate-oxide thickness variation (ΔT_{ox}), and random dopant fluctuations (ΔN_{ch}). As shown in Fig. 13, variation in RO frequency increases and diverges for different chips, as V_{dd} is increased from 0.8 to 1.5V. Chip 9 has higher $(\sigma/\mu)_{36}$ and more divergence in $(\sigma/\mu)_{36}$ than that of typical chips.

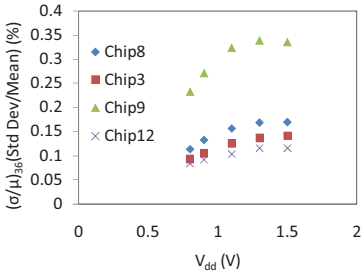

Fig.13. Measured $(\sigma/\mu)_{36}$ for RO monitor CW40 from V_{dd} =0.8 to 1.5 V for 4 chips show that Chip 9 has the biggest and the most vertical divergence in $(\sigma/\mu)_{36}$.

Principal Component Analysis is applied using five combinations of V_{dd} and T measurement conditions (Table 3).

M Gradient Matrix	$\frac{\partial(F)}{\partial(L)}$	$\frac{\partial(F)}{\partial(Tox)}$	$\frac{\partial(F)}{\partial(Nch)}$
Vdd=0.8, T= 25 deg C	-0.47	-0.75	-0.53
Vdd=0.9, T= 25 deg C	-0.71	-1.03	-0.60
Vdd=1.1, T= 25 deg C	-1.17	-1.36	-0.73
Vdd=1.1, T = 40 deg C	-1.15	-1.22	-0.72
Vdd=1.1, T = 90 deg C	-1.14	-1.15	-0.67

Table 3. Normalized gradient matrix for Chip 9 at five operating conditions

Using SPICE and PSP/BSIM4, variations corresponding to $\pm$5% of mean measured RO frequency are used to compute the normalized gradient matrix M in Table 2 in a linear mean square problem, MX=Y [9]. The M matrix shows that as operating voltage increases, RO frequency is more sensitive to changes in L, T_{ox}, and N_{ch} because drain-induced barrier lowering (DIBL) effects have more prominent increase at differing rates at higher operating voltages. Here, Y is a vector of measurements, X, is the normalized solution for changes in L, N_{ch}, and T_{ox} computed via SVD for each RO in one chip.

The 3 solution parameters for each 36 RO in Chip 9 are shown in Fig. 14 for layout CW40. In this case, to reduce block-to-block effects, the average RO frequency for each RO block has been used to normalize measured RO frequency within each block. These variations show that the source of within-chip variations is mainly attributed to random dopant fluctuation effects with L, T_{ox}, and channel doping contributing $\pm$ 1 %, $\pm$ 1%, and $\pm$ 6%, respectively.

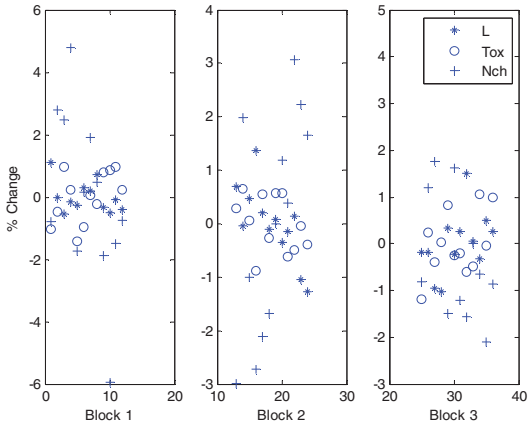

Fig. 14. % Change in L, T_{ox}, and N_{ch} vs. RO Number for Chip 9 RO monitor CW40. N_{ch} varies $\pm$ 6%. Block-to-block variation (every 12 RO) is present.

4. CONCLUSIONS

Lithography variations are shown, for the first time, to be electronically measurable, which demonstrate that specific process parameters can be identified and quantified through electronic measurements. Adjustments in inverter layout design parameters can enhance the sensitivity of RO frequency to variations in focus, overlay, CESL-induced strain effects, and STI-induced stress effects to be well above the residual ~0.2%-0.3% level of variation due to random sources. In a 45 nm technology, layout dependent effects are most significant for nitride CESL-induced stress. While parameter-specific RO monitors provide a permanent record of process effects, they are best used during process development and calibration, when less stringent design rules and no-OPC drop-ins can be accommodated, yielding inverter layouts with higher sensitivities to process variations.

5. ACKNOWLEDGMENTS

The authors wish to acknowledge the contributions of the students, faculty and sponsors of the Berkeley Wireless Research Center (BWRC), C2S2, and the wafer fabrication donation of ST Microelectronics. This research is partially supported by the UC Discovery Grant ele07-10283 under the IMPACT project and by SRC Research Task 1443 and 1600.

6. REFERENCES

[1] D. Boning, TAU 2002
[2] M. Kanno, VLSI 2007
[3] L.-T. Pang, ISSCC 2006
[4] L.-T. Pang, CICC 2008
[5] B. Wan, IEEE Workshop on Test Structure for Variability Characterization 2008
[6] L. T.-N. Wang, SPIE 2009
[7] K. Rim, IEDM Short Course 2007
[8] A. Khakifirooz, IEDM, 2006
[9] L.-T. Pang, UC Berkeley PhD Thesis 2008