Parameter-Specific Electronic Measurement and Analysis of Sources of Variation using Ring Oscillators

Lynn T.-N. Wang*[a], Liang-Teck Pang[a,b], Andrew R. Neureuther[a], Borivoje Nikolić[a]
[a]Dept of EECS, Univ. of California/Berkeley, Berkeley, CA, USA 94704
[b]IBM T.J. Watson Research Center

ABSTRACT

Parameter-specific and simulation-calibrated ring oscillator (RO) inverter layouts are described for identifying and quantitatively modeling sources of circuit performance variation from source/drain stress, shallow trench isolation (STI) stress, lithography, etch, and misalignment. This paper extends the RO approach by adding physical modeling/simulation of the sources of variability to tune the layouts of monitors for enhanced sensitivity and selectivity. Poly and diffusion layout choices have been guided by fast-CAD pattern matching. The accuracy of the fast-CAD estimate from the Pattern Matcher for these lithography issues is corroborated by simulations in Mentor Graphics Calibre. Generic conceptual results are given based on the experience from preparing of proprietary layouts that pass DRC check for a 45 nm test chip with ST Micro. Typical improvements in sensitivity of 2 fold are possible with layouts for lithography focus. A layout monitor for poly to diffusion misalignment based on programmable off-sets shows a 0.8% change in RO frequency per 1nm poly to diffusion off-set. Layouts are also described for characterizing stress effects associated with diffusion area size, asymmetry, vertical spacing, and multiple gate lengths.

Keywords: Variability, ring oscillators, lithography, DFM, Pattern Matching, focus, misalignment, stress

1. INTRODUCTION

Measurement of ring oscillator (RO) frequencies using scan-chains to select individual RO has been shown to be an efficient method to make thousands of electronic measurements using only a few input/output pads [1]. In recent studies, screening sets of RO with different layouts helped identify causes of layout-dependent circuit performance variability in 90 nm and 45 nm generation CMOS circuits [2, 3]. These designs have made possible thousands of measurements through which wafer-to-wafer, die-to-die, and layout dependent effects have been identified in the presence of random noise [4]. With each technology node, the dominant sources of circuit performance variation appear to change. At 45 nm, stress effects due to capping layers (5%) and STI (2%) have become significant and replaced lithography (now 2%) and etch (still 2%) as the dominant sources of circuit performance variation. However, even with large numbers of measurements, the confounding of parameter effects in RO frequencies makes it challenging to clearly identify the individual sources of performance variation.

A natural generalization of the scan-chain RO automated measurement approach is to introduce layout modifications that enhance the dependency of the electrical performance to particular parameters. For example, modifying the layout to double the electrical sensitivity to a parameter would allow four times fewer measurements to be made to quantify that parameter in the presence of random variations. A suitable starting point for enhancing the electrical sensitivity of layouts is to use the physical understanding of fabrication processes to design layouts to induce systematic levels of a given process effect, while also maintaining a nearly fixed response, or known response, to other parameters. The choice of layout changes can be guided by simulation tools such as Mentor Graphics Calibre for lithography and Synopsys Sentaurus for stress. The understanding of the physical effects must be moved from the wafer up to layout level where various choices in the layout can be adjusted to increase sensitivity to a given parameter while attempting to reduce the response to other parameters. First-cut approximate physical models at the mask plane are adequate for this purpose, such as the maximal lateral influence functions introduced with the Fast-CAD pattern matching technique [5], which can be viewed as a convolution of a lateral influence kernel and a mask layout. Of particular interest are the maximal lateral influence functions or kernels developed for lithography [6, 7]. Similar lateral influence kernels have yet to be developed for etch and stress.

Design for Manufacturability through Design-Process Integration III, edited by Vivek K. Singh, Michael L. Rieger,
Proc. of SPIE Vol. 7275, 72750L · © 2009 SPIE · CCC code: 0277-786X/09/$18 · doi: 10.1117/12.814227

This paper develops parameter-specific layout structures for ring oscillator monitoring of circuit performance at the 45 nm technology node. Thirty-two layout test structures that pass design rule checks are introduced. These aim to quantify the effects of lithography, etch, misalignment, STI stress, and source/drain capping layer stress. Maximum lateral influence kernels are used for lithography to guide the layout choices that are later corroborated by simulations in Mentor Graphics Calibre with generic process parameters. For stress where lateral influence kernels are not known, a set of structures is designed for characterizing the functional form of the lateral influence. The designs have been included in a testchip that is being fabricated by ST Micro. While only generic layouts and process parameters are used in this paper, the actual layouts checked with process models including OPC indicate that the expected performance is nearly the same as that shown here for generic assumptions.

Section 2 introduces the RO array circuit design schematics and layout floor planning. Section 3 discusses the methodology of leveraging Pattern Matching to assess and guide the design of lithography-based test-structure layout, i.e. focus, etch, and misalignment. Section 4 describes the development of source/drain and STI stress test structure layouts.

2. RING OSCILLATOR TESTING: SCHEMATICS AND LAYOUT

The ring oscillator array developed by Boning [1] and adapted by Pang [2, 3] is grouped into basic units termed *tiles*. One ring oscillator *tile* consists of not only a ring oscillator but also additional control logic to enable the ring oscillator as well as to multiplex the ring oscillator signal from the ring oscillator output to the final output pad (shown in Figure 1(a)). These ring oscillator *tiles* are arrayed into a matrix shown in Figure 1(b). Vertical and horizontal scan-chains located along the perimeter of the final array are used to select the *tiles* (shown in Figure 1(b)). The vertical scan-chains select the row of the ring oscillator *tile*, and the horizontal scan-chains select the column of the ring oscillator *tile*. The on-chip frequency divider slows down the frequency of the ring oscillator before the ring oscillator signal reaches the final output pad. This slower signal improves the ease of off-chip automatic measurements. Note that just a few pads are required to measure thousands of ring oscillator circuits.

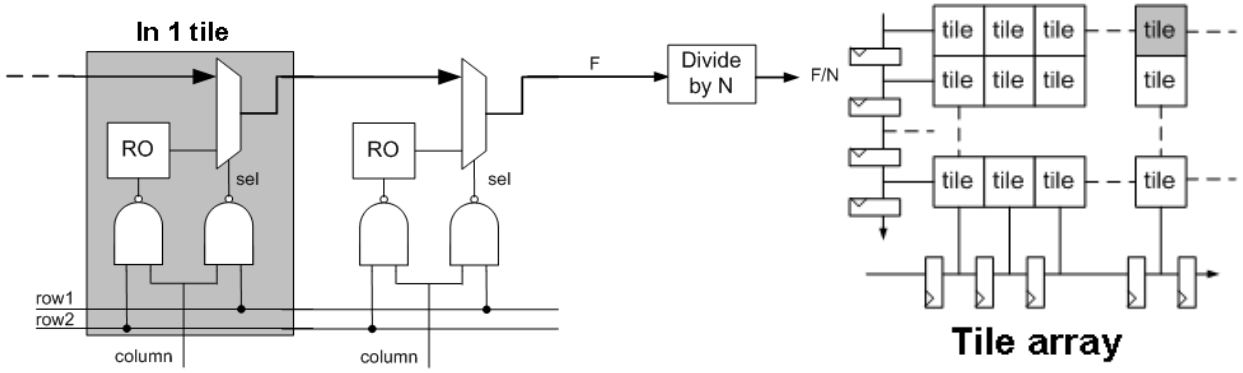


Figure 1(a) Figure 1(b)

Fig. 1. Figure 1 is the schematic view of the ring oscillator array floor plan and automatic electronic measurement setup. Figure 1(a) shows the schematic of a ring oscillator *tile* arrayed in a row. Figure 1(b) shows an example of a 2-D array with a set of vertical and horizontal scan-chains to enable row and column select, respectively.

The layout corresponding to the schematics in Figure 1 is shown in Figure 2(a). On the left of the layout is the set of vertical scan-chains to select the row *tiles*, and on the bottom of the layout is the set of horizontal scan-chains to select the column *tiles*. The smaller rectangular box represents a *tile*, and the larger rectangular box represents a set of 32 *tiles*, termed *big tile*. The *big tile* is replicated 3 times horizontally and four times vertically. On the right of the layout consists of a set of multiplexers, which route the chosen ring oscillator signal down to a single signal and then feeds into the on-chip frequency divider. The layout measures 230um by 270um and is replicated across a 2mm by 1mm die area in the manner shown in Figure 2(b) to capture the lithography scan and across slit direction variations.

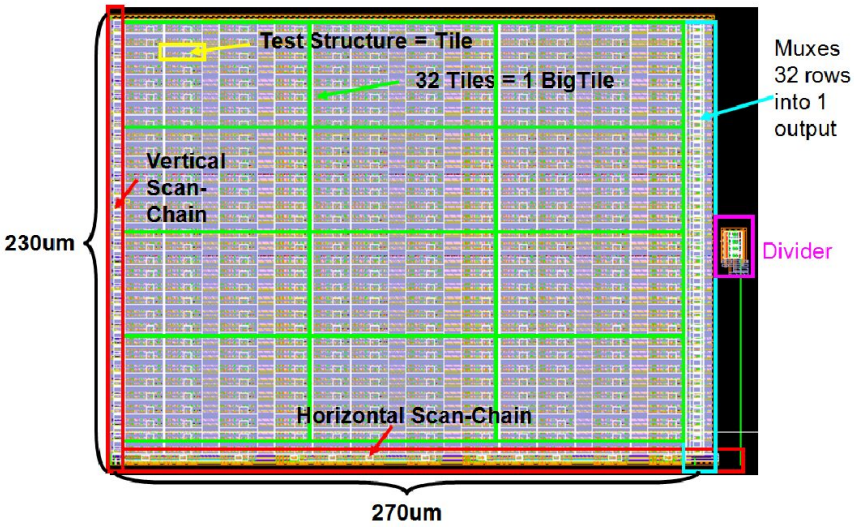


Figure 2(a) Figure 2(b)

Fig. 2. Figure 2(a) is the layout view of the 45nm ring oscillator array. Figure 2(b) shows the layout in Figure 2(a) replicated at 3 different locations across die in order to capture the variations due to the lithography scanner's slit and scan directions.

3. PATTERN MATCHING GUIDANCE IN DESIGNING OF TEST STRUCTURES

3.1 Physical Basis of Pattern Matching and Lithography Parameters

The concept of modeling process effects at the mask plane level as convolutions with maximal lateral influence kernel functions was introduced in a Pattern Matcher system by Gennari [5]. The influence functions for aberrations including focus and polarization were introduced by Robins [6] and McIntyre [7]. Extensions for illumination effects and for focus as a large aberration are included by Wang [8] and Rubinstein [9]. As shown in Figure 3, the Pattern Matcher contains a pattern generator that uses lens aberrations (including focus) and mask illumination to create the lateral influence pattern of kernel. The Pattern Matcher then makes an analog weighted convolution of the chip layout and reports the degree of similarity, or Pattern Match Factors (PMF). Circuit designers can use these outputs from the Pattern Matcher to determine the sensitivities their layouts to residual process effects.

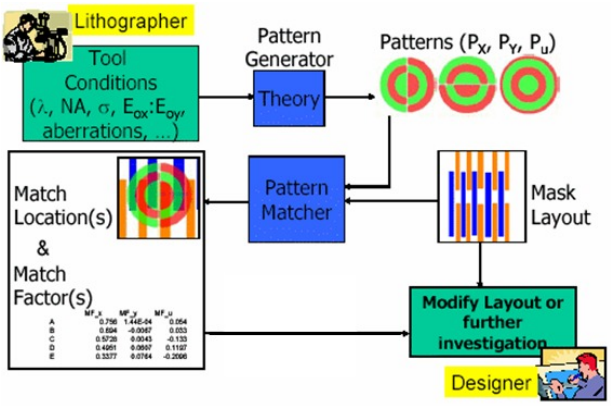


Fig. 3. The Pattern Matcher Concept and Flow

3.2 Using the Pattern Matcher to Optimize Test Structure Design

To create parameter-specific layouts in the current study, the starting point was to overlay the lateral influence function centered at the middle of the gate in the ring oscillator layout. Then by modifying the poly layout or diffusion layout, the similarity with the lateral influence function and the sensitivity could be enhanced. The increase in PMF, i.e. the convolution of the kernel with the layout, was used as a guide for the sensitivity improvement. Only changes that were consistent with DRC were allowed. A generic annular illumination optical model at the 45nm process was used in both generating the pattern and in assessing the sensitivity change using Mentor Graphics Calibre. The designs shown here were created without regard to OPC but with enhanced gate lengths. Proprietary versions of these designs for inclusion on a tapeout were tested with proprietary models including OPC and retained their sensitivity improvements.

Figure 4 illustrates the method for enhancing focus sensitivity of layouts. The goal in Figure 4(a) is to enhance the sensitivity of the poly gate length to focus. In this case, the maximum lateral influence kernel function for focus is overlaid and centered on the gate. The focus pattern suggests adding additional poly features to block the light in the green rings that reduce focus sensitivity, allowing the light through the red rings to increase focus sensitivity. The initial choices for additional rectangles were then adjusted in simulation with Calibre to obtain the layout design. The goal in Figure 4(b) is to increase the sensitivity of the transistor width with focus during printing of the diffusion layer. The test pattern is located where the poly crosses the active mask in this H-shaped diffusion area, which is prone to corner rounding with defocus. In this case, annular illumination is used, and the defocus maximum lateral kernel function has been multiplied by the mutual coherence for annular illumination [10]. This function is also oscillatory and produces more nulls and more green area in the kernel function.

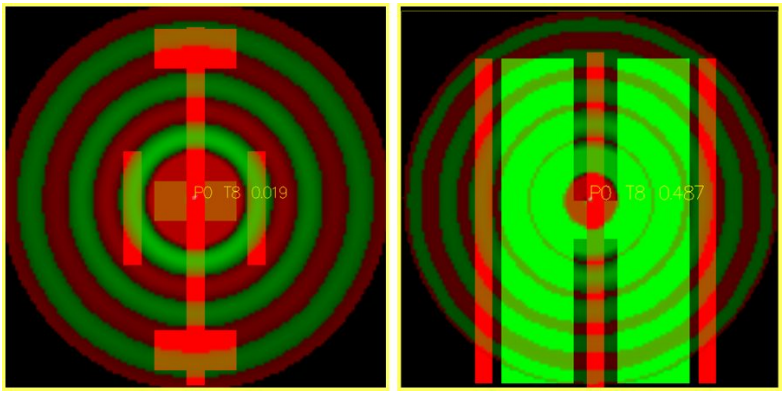


Figure 4(a) Figure 4(b)

Fig. 4. Figure 4(a) shows a defocus partial-coherent kernel matched to the poly layer of the layout. Figure 4(b) shows a defocus kernel with annular illumination matching to the diffusion layer of the layout.

3.3 Strehl Ratio Test for Mentor Graphics Calibre and Process Variation Band Calibration

To choose focus limits for simulation with Calibre with generic parameters, a calibration test was introduced. A Strehl ratio test was used which consisted of plotting the peak intensity from a transparent square measured (0.4)λ\NA on a side. An NA=1.15 and annular illumination with σ=0.9, 0.75 were used to set the focus limits (although top-hat illumination is also shown). The intensity peaks in focus and decreases on either side of best focus. For the generic simulation model used, the peak was shifted to -30nm. Since 2/3 of a Rayleigh Unit corresponds to a 0.9 intensity drop, the focus limits for the annular curve were found to be -67 nm to +07 nm. These focus settings were used throughout this work to assess layout sensitivity to focus for both poly and active layer imaging.

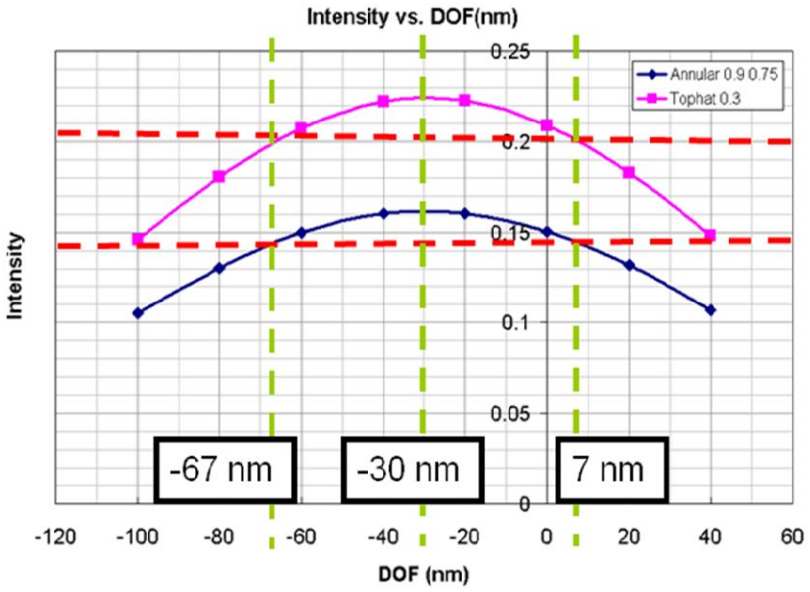


Fig. 5. Figure 5 shows the Strehl ratio test for a light square measuring (0.4)λ\NA per side at tophat and annular illumination via Calibre.

4. ETCH AND LITHOGRAPHY MONITORS

4.1 Layout Monitors for Etch Dependence of Poly

At 90nm, lithography/etch dependence on proximity to neighboring dummy gates was determined to be an issue [2]. For example, the ring oscillator frequency of an isolated line exceeded the range of the measurements; however, the addition of 2 neighbor dummy polys (one on each side of the gate-under-test) shifted the ring oscillator frequency by 10% compared to that of test structures with only one neighboring dummy poly. A similar test structure has been designed for the 45nm technology. To decouple the etch effects with the lithography effects prior tapeout, the optical effects on these layout test structures are calibrated with Calibre. The contours are obtained from aerial image simulations for 60 nm gates in the focus PV bands conditions determined above. Simulations in Figure 6 show nearly identical changes in gate length due to focus. Thus, it can be concluded that if there were any shifts in the measured ring oscillator frequency, it is more likely due to a contribution from etch rather than lithography focus.

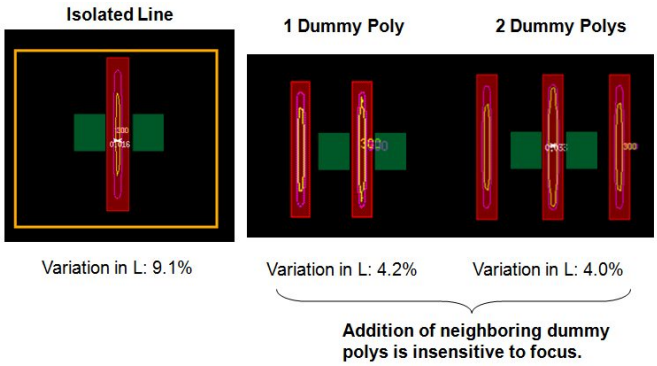


Fig. 6. Isolated line (shown on the left) is 2 times more sensitive to focus variations than that of the test structures with dummy structures.

4.2 Layout Monitors for Focus Dependence of Poly

The focus monitors are designed at two different gate lengths. The lateral influence kernel for focus was overlaid as shown in Figure 4, and the layout was adjusted to increase the similarity of the poly overlap to the color of the central red area. To make these structures specifically more sensitive to focus, hammerheads are added to the gate and the dummy polys are truncated where they are on green. The new structures for both linewidths show a sensitivity increase to focus of 2 fold and still meet design rule requirements.

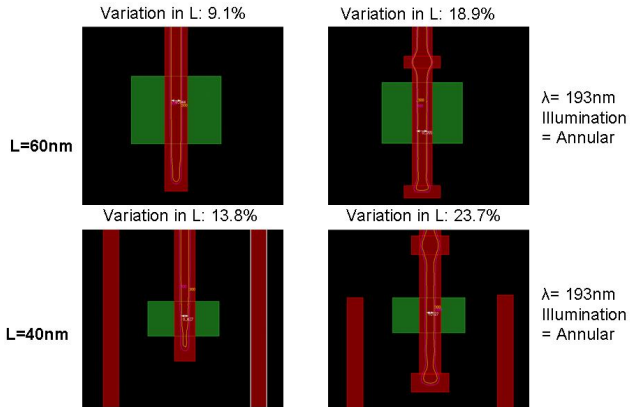


Figure 7(a) Figure 7(b)

Fig. 7. Focus layout structures. Figure 7(a) contains the control test structures, and Figure 7(b) contains the 2X focus sensitive test structures.

4.3 Layout Monitors for Diffusion Focus and Poly/Diffusion Misalignment Using Programmed Off-Sets

H-shape and wedge-shape programmable misalignment monitors are presented. Each structure is designed with five pre-programmed misalignment off-sets. Typical images of a rectangular diffusion transistor that result from the H-shape diffusion layout in Figure 4 are shown in Figure 8. The yellow rectangles indicate alternative pre-programmed layout positions at offsets of ±15nm. The curvature of the diffusion area in Figure 8(b) indicates that the transistor gate width will increase out of focus.

Contours from image simulation for the rectangular diffusion area show only a 1% change in gate width with a deliberate shift of 15nm from the center of the gate. As shown in Figure 8(b) in the H-shape monitors, the addition of the large diffusion area increases the curvature, and the same amount of shift of the gate location shows a change 20% in width, which increases the sensitivity by 20 fold compared to that of the minimum width gate.

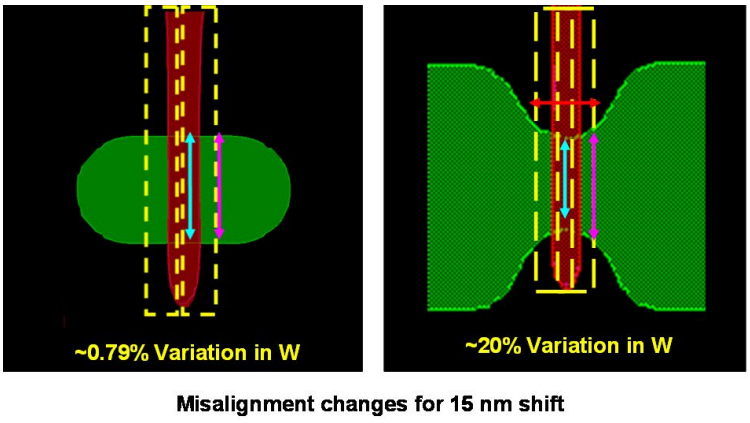


Figure 8(a) Figure 8(b)

Fig. 8. Figure 8(a) shows the simulation contours for a minimum-sized gate. Figure 8(b) presents the H-shape misalignment test structure.

The relative sizing of transistors in SRAMs often introduces wedge-structures and thus creates sensitivity to alignment. This inspired the creation of the wedge-shape structure in Figure 9. Here the active area decreases with gate position shift from left to right. The wedge-shape structures are designed with offsets of 0nm, ±5nm, and ±10nm. A shift of the gate by 10nm shows a gate width change of 13%, which increases the sensitivity by 13 times compared to that of the rectangular diffusion layout at minimum width gate in Figure 9(a).

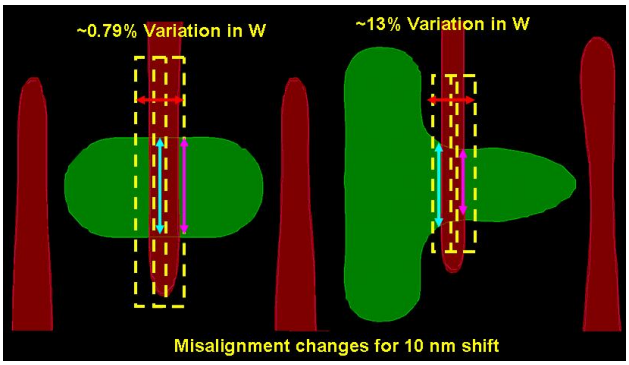
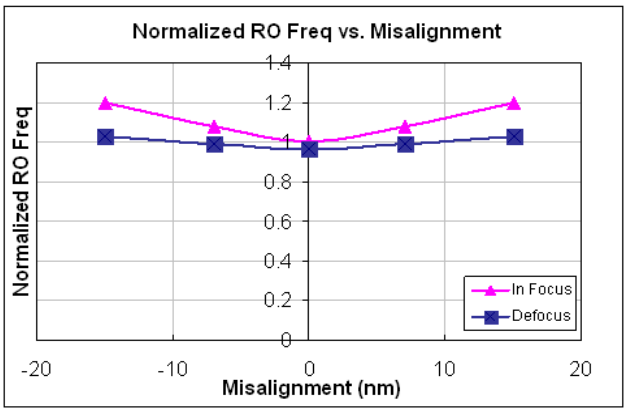


Figure 9(a) Figure 9(b)

Fig. 9. Figure 9(a) shows the simulation contours for a minimum-sized gate. Figure 9(b) presents the wedge-shape misalignment test structure.

Programmed misalignment mapped to ring oscillator frequency is shown in Figure 10. For each programmed offset, the gate width at the offset is extracted from simulation contours in Calibre. This gate width is then used in ELDO simulations to determine relative ring oscillator frequency. The H-shape layout monitor is characterized by a parabolic curve, for which the minimum location identifies the best alignment. The curvature indicates the level of defocus as shown in Figure 10(a) for a defocus level of 0nm and -67nm.

The wedge-shape structure returns a characteristic curve with a slope of 0.8% change in RO freq per 1nm defocus at 0nm off-set as shown in Figure 10(b). Despite its sensitivity to misalignment, it lacks a focus reference as well as an optimum alignment reference due to the test structure's asymmetric properties.



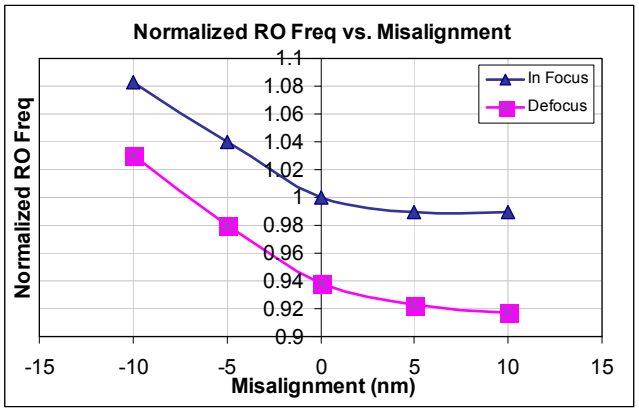


Figure 10(a) Figure 10(b)

Fig. 10. Ring oscillator frequency versus misalignment (a) for the H- shape structure at offsets of 0nm, ±7nm, and ±15nm and (b) the wedge-shape structure, at offsets of 0nm, ±5nm, and ±10nm.

For many test structures, the change in loading due to gate capacitance tends to reduce the change in ring oscillator frequency since the change in gate capacitance induces a change in gate drive current. This effect can be mitigated by adding sizeable capacitance loading so that the fractional change in capacitance is small. This mitigation can be achieved by inserting large, fixed capacitors in between the inverter stages. One option is to implement these capacitors using transistors with source and drain tied to the power supply or ground as shown in the schematic in Figure 11. Figure 11 also shows a plot of RO frequency versus normalized gate width via simulations in ELDO. Here the drawn

inverter layout width is varied from 80% to 120% to observe the frequency sensitivity to gate width change. Without the addition of capacitors, the sensitivity at (1,1) is 0.2. With the addition of capacitors, the sensitivity increases to 0.8. While the ring oscillator frequency is about an order of magnitude slower with the additional of these capacitors, there is more than sufficient accuracy to reliably measure to fractions of a percent.

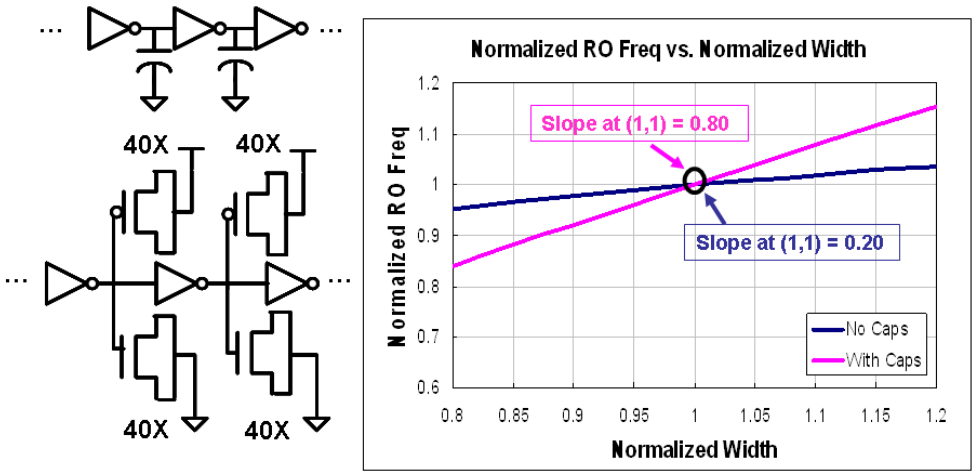


Fig. 11. Schematic for adding a large fixed capacitance between stages and the normalized ring oscillator frequency as a function of gate width, with and without capacitive load.

5. SOURCE/DRAIN STRESS AND STI STRESS MONITORS

The 45nm process utilizes strained-Si CMOS technology with several process techniques to enhance NMOS and PMOS mobility. The 45nm process techniques are summarized by Pang in [10]. The nitride CESL capping layer induces a strong uniaxial tensile strain in the NMOS devices to increase NMOS mobility, without affecting the PMOS performance. The transistor channels are oriented in the <100> direction, which decreases transistors' sensitivity to stress. STI strain is made to exert a weak tensile force on the transistor channels due to the use of sub-atmospheric chemical vapor deposition oxide (SACVD) for trench isolation.

To characterize the impact of strained-Si technology and to evaluate the impact thereof, test structures are designed at 45nm technology to examine RO freq variations due to source/drain stress, STI stress, source/drain asymmetry, and gate stress. Figure 12 is a summary of the stress monitors.

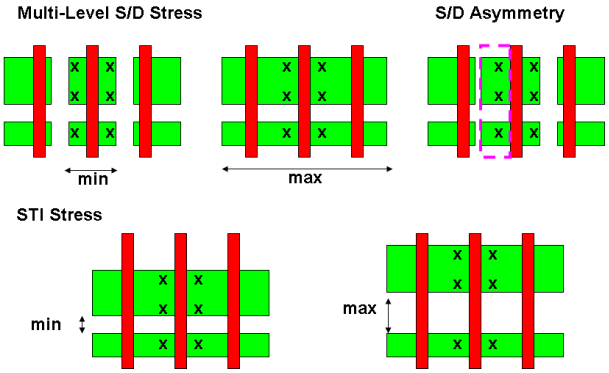


Fig. 12. Test structures that monitor the following sources of variability are shown: multi-level source/drain stress, source/drain asymmetry, and vertical/horizontal STI test structures.

Previous results from the 45nm test chip developed by Pang demonstrated that an increase of the source/drain area induced at 5% change in RO frequency [12]. However, this experimental screening was done at two levels as indicated in Figure 13. Six new test structures with multiple levels of source/drain area are designed to systematically characterize

the cumulative impact effect. By further interpreting the effect of the cumulative impact curve, the maximum lateral impact function can be obtained. This lateral impact function can be translated into a pattern or kernel and incorporated into the Pattern Matcher.

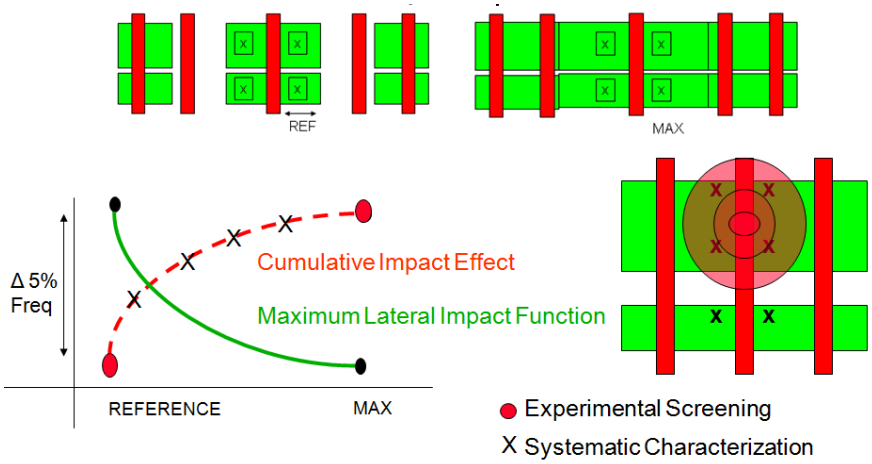


Fig. 13. The multi-level source/drain stress test structures serve to characterize the cumulative impact effect at six different intermediate levels of source/drain areas. By interpreting the cumulative impact curve, the maximum lateral impact function can be obtained and mapped into a kernel in the Pattern Matcher.

Khakifirooz demonstrated that with scaling of devices, the source of a device has a larger impact on transistor mobility than that of the drain [11]. Test structures with asymmetric source/drain area are designed to corroborate Khakifirooz observations. In Figure 12, the test structure for source/drain asymmetry shows a larger source than the drain. It is hypothesized that the larger source will have more severe impact on the ring oscillator frequency than that of a larger drain.

STI test structures are designed to study the effect of STI influence in the lateral and vertical direction of the channel. In the lateral direction, the monitors are designed with various levels of STI widths that lie in between the transistor and its adjacent dummy gates. In the vertical direction, test structures are designed by varying the STI widths that lie in between NMOS and PMOS devices.

6. CONCLUSIONS

Circuit performance using ring oscillators on scan-chains has been generalized by adjusting layout features to enhance their response to particular physical effects to create parameter-specific quantitative monitors. Pattern Matching with known lateral spillover functions gives helpful guidance. Combinations involving layout changes on multiple layers and multiple levels of programmed off-sets are very effective. For example by adding poly features outside the device area, the sensitivity of poly to focus was doubled. Wedge and H-shape misalignment monitors in diffusion were used to produce a sensitivity of 0.8% change in RO frequency per 1 nm alignment, and in the case of the H-shape, the defocus level could simultaneously be determined. Stress is a new frontier where intuitive layout changes are being utilized and will later be coupled with systematic process simulation.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the contributions of the students, faculty and sponsors of the Berkeley Wireless Research Center (BWRC), the National Science Foundation Infrastructure Grant No. 0403427, wafer fabrication donation of ST Microelectronics.

This research was partially supported by the UC Discovery Grant ele07-10283 under the IMPACT program, and by SRC 1443.

REFERENCES

[1] Boning, D., Panganiban J., Gonzalez-Valentin, K., Nassif S., McDowell, C., Gattiker, A., Liu, F., "Test Structures for Delay Variability," TAU, (2002).

[2] Nikolic, B. and Pang, L.-T., "Measurements and Analysis of Process Variability in 90nm CMOS", 8th International Conference on Solid-State and Integrated-Circuit Technology, (2006).

[3] Pang, L.-T., "Measurement and analysis of variability in CMOS circuits," Ph.D. Dissertation, University of California, Berkeley, (2008)

[4] Qian, K., Spanos, C., Nikolic, B., "Hierarchical Modeling of Spatial Variability with a 45nm Example," SPIE *Advanced Lithography,* (2009).

[5] Neureuther A., Gennari, F., "No-Fault Assurance: Linking Fast Process CAD and EDA," SPIE, Vol. 4889 (2002).

[6] Robins, G., Neureuther, A. R., Adam, K., "Measuring Optical Image Aberrations with Pattern and probe Based Targets," J. Vac. Sci. Technol. B., Vol. 20, No.1, pp. 338-343, (2002).

[7] McIntyre, G., Holwill, J., Neureuther, A., Capodieci L., Zou Y., Levinson H., and Kye J., "Screening layouts for high-NA polarization effects using pattern matching," J. Vac Sci. Technol., (2005).

[8] Wang, L.-T., Yeh, A., Kem, L., and Neureuther, A.R., "Illustration of Illumination Effects on Proximity, Focus Spillover, and Design Rules," SPIE *Advanced Lithography,* (2009).

[9] Rubinstein J., Neureuther A. R., "Through-focus pattern matching applied to double patterning," SPIE *Advanced Lithography,* (2009).

[10] Pang, L-T., Nikolic, B., "Measurements and Analysis of Variability in 45nm Strained-Si Technology," CICC, (2008).

[11] Khakifirooz A., Antoniadis D. A., "Transistor Performance Scaling: The Role of Virtual Source Velocity and Its Mobility Dependence," IEDM, (2006).