Hierarchical Modeling of Spatial Variability with a 45nm Example

Kun Qian, Borivoje Nikolić and Costas J. Spanos
Dept. of EECS, University of California, 550 Cory Hall, Berkeley, CA USA 94720

ABSTRACT

In previous publications we have proposed a hierarchical variability model and verified it with 90nm test data. This model is now validated with a new set of 45nm test chips. A mixed sampling scheme with both sparse and exhaustive measurements is designed to capture both wafer level and chip level variations. Statistical analysis shows that the across-wafer systematic function can be sufficiently described as parabolic, while the within-die systematic variation is now very small, with no discernible systematic component. Analysis of pattern dependent effects on leakage current shows that systematic pattern-to-pattern L_{EFF} variation is almost eliminated by optical proximity correction (OPC), but stress-related variation is not. Intentionally introduced gate length offset between two wafers in our dataset provides insight to device parameter variability and sheds additional light on the underlying sources of process variation.

Keywords: circuit variability, across-wafer variation, 45nm, ring oscillator, pattern dependent effect.

1. INTRODUCTION

While pure random fluctuations impact all devices in modern ICs, significant spatial components across-die and across-wafers play a key role in the overall performance, leading to a hierarchical process variability model [1]. Used in conjunction with 90nm test data, this model has been shown to successfully address both local variability and variations across significant distances. Systematic chip-to-chip and within-chip variations are sufficiently described by deterministic spatial functions across-wafer and across-die. The residuals of these functions were shown to be identically, independently, normally distributed (IIND), rendering the concept of "spatial correlation" unnecessary.

To verify our model for state-of-the-art technology nodes, we applied it to a new, comprehensive 45nm data set. The test chip was designed by L.T. Pang [6] and manufactured on a state–of-the-art production line. On-chip test structures include arrays of ring-oscillators (RO) replicated with various layout styles, off-state transistors, and an SRAM array, each individually addressable. This allows for accurate and flexible estimation of the hierarchical components of the variability model.

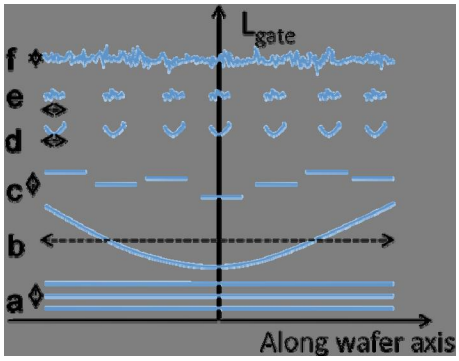

Fig. 1: Hierarchical process variability: (a) wafer-to-wafer (b) across-wafer (c) die-to-die (d) across-die (e) pattern dependent (f) local random noise [1]

This model is originally intended for modeling device parameters such as gate length (L_{EFF}), oxide thickness (T_{OX}), and threshold voltage (V_{TH}). However, many of these device parameters cannot be directly measured, because of the cost of the direct measurement when measuring large, statistically significant samples. Nonetheless, when the variation of device parameters is reasonably small, circuit performance metrics such as frequency may still be treated in the same fashion. However, due to the non-linear dependence on device parameters like L_{EFF} and V_{TH}, new assumptions are needed. For instance, wafer-to-wafer and pattern-to-pattern variation may no longer be modeled as simple additive

Design for Manufacturability through Design-Process Integration III, edited by Vivek K. Singh, Michael L. Rieger,
Proc. of SPIE Vol. 7275, 727505 · © 2009 SPIE · CCC code: 0277-786X/09/$18 · doi: 10.1117/12.814226

terms, and interactions can occur between wafer-to-wafer variability and all the other components, especially when wafers are processed under different process conditions.

2. TEST CHIP AND PROCESS TECHNOLOGY OVERVIEW

The test chips are fabricated using a 45nm low power CMOS process [7], [8], [9]. The die photo is shown in Fig. 2b. Each die contains a ring oscillator (RO) array with 18 x 16 identical tiles. Each tile consists of 17 ROs and 17 pairs of off-state NMOS and PMOS transistors for leakage measurements, each with the same device size embedded in a different pattern as shown in Fig. 2a. It should be noted that the pre-OPC patterns depicted in Fig. 2b are subject to OPC treatment prior to fabrication, and the specifics of this OPC treatment are not known to us. Measurement circuitry is adopted from the design of 90nm test chip [11]. RO frequency and corresponding off-state NMOS/PMOS transistor leakage currents are measured in our laboratory after the wafers have been diced and the chips packaged.

All transistor channels are oriented in the <100> direction, which enhances PMOS mobility and makes it insensitive to stress[13]. There are two major sources of stress in this process: strain caused by contact-etch stop layer (CESL) and the shallow trench isolation (STI) stress. Sub-atmospheric chemical vapor deposition oxide (SACVD) largely reduces usually strong compressive STI stress and turns it into a weak tensile one. CESL is formed by intentionally depositing a nitride layer on top of NMOS transistors, which introduces strong horizontal tensile strain that greatly enhances the electron mobility.

Another important feature of the new 45nm test chip fabrication is the different gate trimming treatment for the two wafers we have, aiming at a nominal 4nm reduction in gate CD from the slower wafer (#1) to the faster wafer (#2). Other minor changes in process may also exist. This allows us to more precisely identify the link between circuit performance and device parameters, and enables further investigation into the process variability.

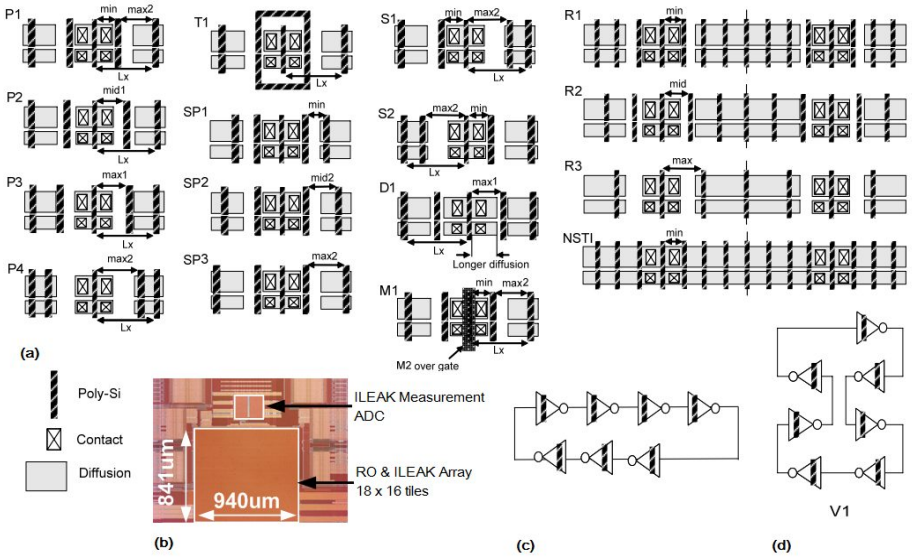

Fig. 2: (a) 16 pre-OPC layout configurations, all arranged horizontally. An additional configuration P1 is arranged vertically and not shown here; (b) 45nm test-chip die photo; (c) Horizontal arrangement; (d) Vertical arrangement [6]

3. MEASUREMENT SAMPLING SCHEME

To capture the wafer-level systematic variation we select chips aiming at wide spatial coverage across the wafer, but we only measure a small subset of the available samples on each of those chips. For instance, on each of the 45nm test wafers, half of the 90 chips on the wafer are sampled, and for each chip we only measured 8 of the available 288 tiles for RO frequency, reducing the overall measurement cost by a factor of 72. Statistical analysis shows that this sparse sampling scheme is reasonably good at capturing the average characteristics of a chip.

Given the total measurement cost, there are various choices of chip selection in order to capture the across-wafer systematic nature of the variability. The most straightforward method would be a checkerboard sampling, as shown in Fig. 3a. This method, however, does not take advantage of the spatial property of the wafer-level process variation. As process conditions near the center of the wafer are usually better controlled than closer to the wafer edge, chips near the periphery of the wafer contribute more to the process variation, requiring a denser spatial pattern. On the other hand, chips near the center of the wafer are likely to be more uniform. An optimized sampling scheme taking into account of these effects is shown in Fig. 3b.

For within-die variability characterization, however, complete coverage over the die area is desired. Therefore, a small number (5~6) of chips from each wafer are exhaustively sampled by measuring all 288 tiles .

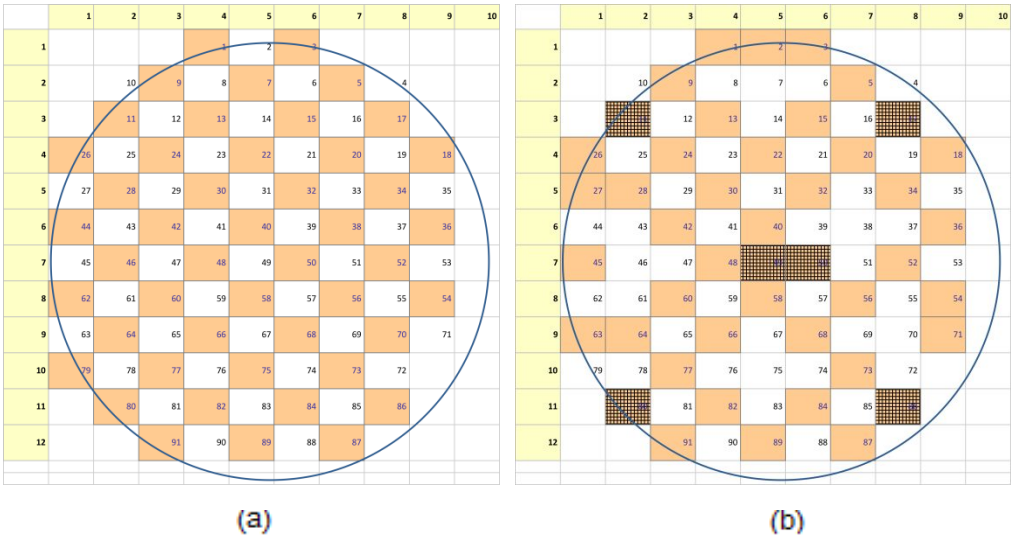

Fig. 3: (a) Checkerboard sampling scheme (b) weighted sampling scheme. Heavily shaded chips are measured exhaustively.

4. DEFINITION OF SYMBOLS

Measured RO frequency and leakage currents can be uniquely labeled as $f_0\langle T,D,W,P\rangle$, $I_{LEAKN,0}\langle T,D,W,P\rangle$, and $I_{LEAKP,0}\langle T,D,W,P\rangle$. To effectively compare the measurement data, RO frequency is normalized to SPICE simulations at a nominal L_{EFF} as per pattern using the corner model corresponding to the slower wafer in order to eliminate the impact of the difference of parasitic capacitance. Leakage current measurements are subjected to a log transformation prior to this analysis, followed by normalization to the average performance measured on pattern P1 of wafer #1. Similar practice is applied to SPICE simulations as well. The detailed definitions are listed in Table 1.

Table 1: Example definitions of symbols of measurement and simulation

$f_0\langle T,D,W,P\rangle$	Raw measurement of RO frequency of tile T, die D, wafer W, and pattern P
$fs_0\langle W,P,L_{EFF}\rangle$	Simulated RO frequency using corner corresponding to wafer W, pattern P, and effective gate length L_{EFF}
$fs_0\langle W1,P,48\rangle$	Simulated RO frequency for wafer #1, pattern P, at nominal gate length 48nm
$f\langle T,D,W,P\rangle$	Measured RO frequency normalized to SPICE simulation $\frac{f_0\langle T,D,W,P\rangle}{fs_0\langle W1,P,48\rangle}$

$fs\langle W,P,L_{EFF}\rangle$	Simulated RO frequency normalized to SPICE simulation $\frac{fs_0\langle W,P,L_{EFF}\rangle}{fs_0\langle W1,P,48\rangle}$
$\overline{f}\langle\bullet,D,W,P\rangle$	$f\langle T,D,W,P\rangle$ averaged over all tiles in die D on wafer W, with pattern P.
$\overline{f}\langle\bullet,D,W,\bullet\rangle$	$f\langle T,D,W,P\rangle$ averaged over all tiles and all patterns in die D on wafer W.
$\overline{f}\langle\bullet,D,W,P\rangle\big\vert_{AW}$	Fitted systematic across-wafer component of RO frequency of die D on wafer W with pattern P.
$I_{LEAKN,0}\langle T,D,W,P\rangle$	Raw measurement of NMOS leakage of tile T, die D, wafer W, and pattern P
$Is_{LEAKN,0}\langle W,L_{EFF}\rangle$	Simulated NMOS leakage using compact SPICE model[1] corresponding to wafer W and effective gate length L_{EFF}
$\overline{\log(I_{LEAKP,0})}\langle\bullet,\bullet,W1,P1\rangle$	$\log(I_{LEAKP,0})\langle T,D,W,P\rangle$ averaged over all devices with pattern P1 on wafer #1.
$\log(I_{LEAKN})\langle T,D,W,P\rangle$	Normalized NMOS leakage measurement: $\frac{\log(I_{LEAKN,0})\langle T,D,W,P\rangle}{\left\vert\overline{\log(I_{LEAKP,0})}\langle\bullet,\bullet,W1,P1\rangle\right\vert}$
$\log(I_{LEAKP})\langle T,D,W,P\rangle$	Normalized PMOS leakage measurement: $\frac{\log(I_{LEAKP,0})\langle T,D,W,P\rangle}{\left\vert\overline{\log(I_{LEAKP,0})}\langle\bullet,\bullet,W1,P1\rangle\right\vert}$

5. WAFER LEVEL VARIATION

Spatial process variation causes device and circuit performance to vary as a function of position within the wafer. According to our hierarchical variation model, the wafer level spatial variation can be decomposed into two parts: the systematic or deterministic across-wafer function and the random die-to-die variation. Depending on the performance metric we look at, the systematic variation has shown a dome or bowl shaped signature in various processes [1][12]. For the 90nm test data we used a fitted 2nd order polynomial function to capture the across-wafer shape. This methodology proves valid for the 45nm test chips. As we explain below, however, in the 45nm case the wafer-to-wafer and pattern-to-pattern variability components now *interact* with the shape of the across-wafer systematic function.

5.1 Across-wafer variation

To extract the across-wafer variation component, we first calculate the chip averages for each pattern style on both wafers. By taking the mean frequency of all devices with the same pattern on one chip, we obtain one data point for each measured chip on the wafer. Since leakage current changes exponentially with threshold voltage, $\log(I_{LEAK})$ really responds mostly to threshold voltage variation, and not so much to linear factors such as mobility enhancement. Measured frequency and leakage current maps across the wafer for pattern P2 on wafer #2 are shown in Fig. 4.

There is a strong correlation among the across-wafer function of frequency, NMOS leakage and PMOS leakage, as shown in Fig. 5. All three types of across-wafer variations can be approximated by a dome-shaped deterministic function. This can be explained by a systematic bowl-shape gate length variation across the wafer, and the corresponding threshold voltage roll-off: RO frequency increases when L_{EFF} gets shorter, and both NMOS and PMOS threshold voltages are likely to drop in magnitude due to short-channel effects. Thus $\log(I_{LEAKN})$ and $\log(I_{LEAKP})$ are correlated by the fact that NMOS and PMOS share the same gate length, while the RO frequency is determined by both NMOS and PMOS threshold voltage and the inverse proportional dependence on gate length.

[1] Since the two wafers in our experiment were produced as different process "splits", each is modeled with a different compact SPICE model, as suggested by the manufacturer.

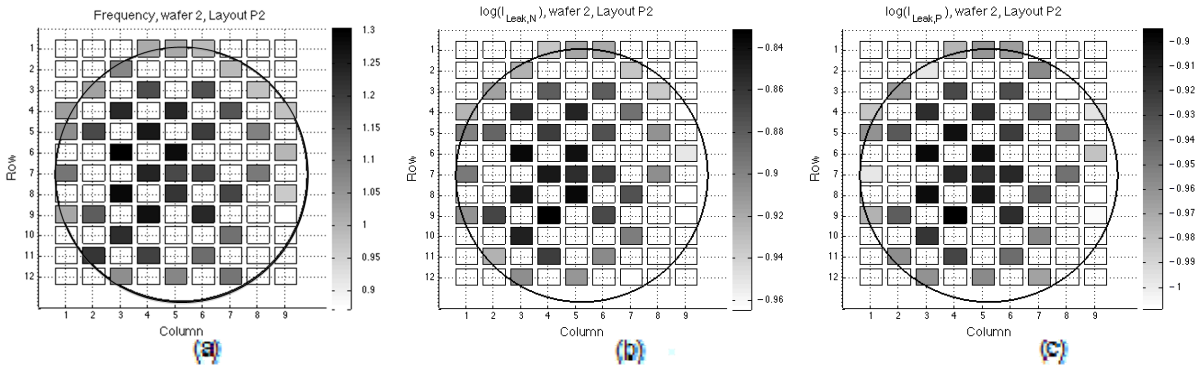

Fig. 4: Wafer maps of (a) $\overline{f}\langle \bullet,D,W2,P2\rangle$ (b) $\overline{\log(I_{LEAKN})}\langle \bullet,D,W2,P2\rangle$ (c) $\overline{\log(I_{LEAKP})}\langle \bullet,D,W2,P2\rangle$. Symbols indicate data averaging across each measured chip. (Note that both leakage plot numbers are in the negative regime)

Past analysis of 90nm data shows that the across-wafer systematic variation can be approximated by a 2nd order polynomial function of the chip coordinates *f(x,y)*. This is still true for the new 45nm test chips. The fitted across-wafer functions of RO frequency and leakage currents are shown in Fig. 6. Comparisons of fitted functions to measurement data are shown in Fig. 7.

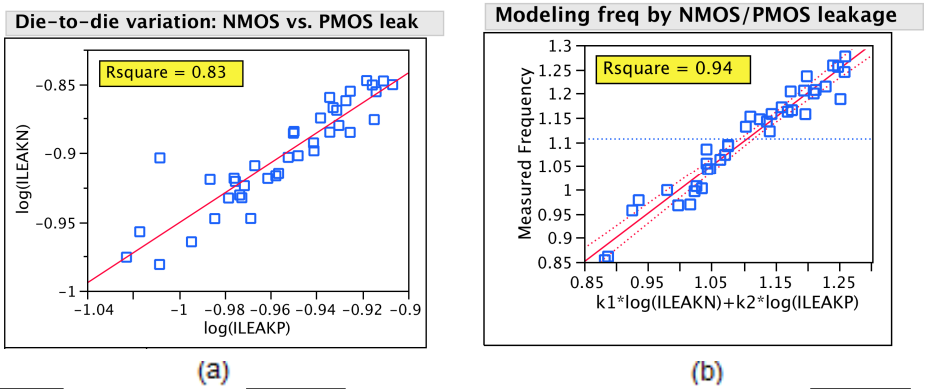

Fig. 5: (a) $\overline{\log(I_{LEAKN})}\langle \bullet,D,W2,P2\rangle$ vs. $\overline{\log(I_{LEAKP})}\langle \bullet,D,W2,P2\rangle$ (b) Modeling $\overline{f}\langle \bullet,D,W2,P2\rangle$ by $\overline{\log(I_{LEAKN})}\langle \bullet,D,W2,P2\rangle$ and $\overline{\log(I_{LEAKP})}\langle \bullet,D,W2,P2\rangle$. Symbols indicate data averaging across each measured chip.

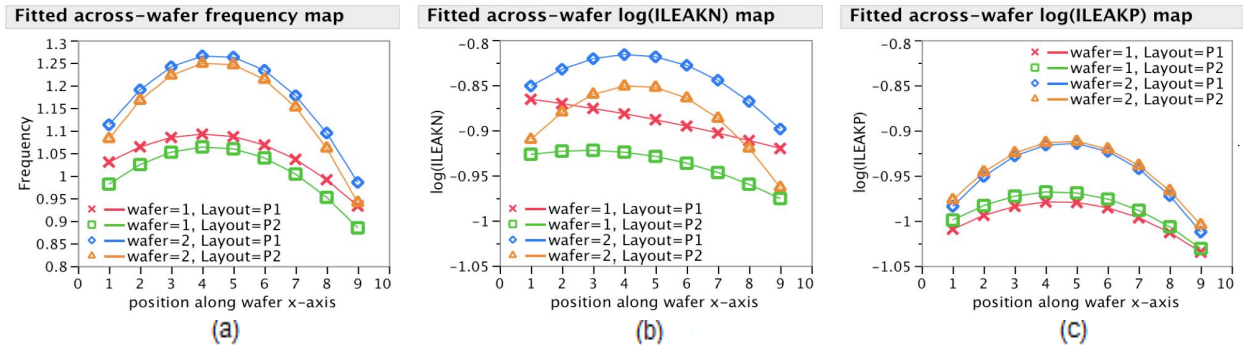

Fig. 6: Fitted across-wafer functions along the central x-axis: (a) $\overline{f}\langle \bullet,D,W,P\rangle\big|_{AW}$, (b) $\overline{\log(I_{LEAKN})}\langle \bullet,D,W,P\rangle\big|_{AW}$, (c) $\overline{\log(I_{LEAKP})}\langle \bullet,D,W,P\rangle\big|_{AW}$. Symbols indicate data averaging across each measured chip.

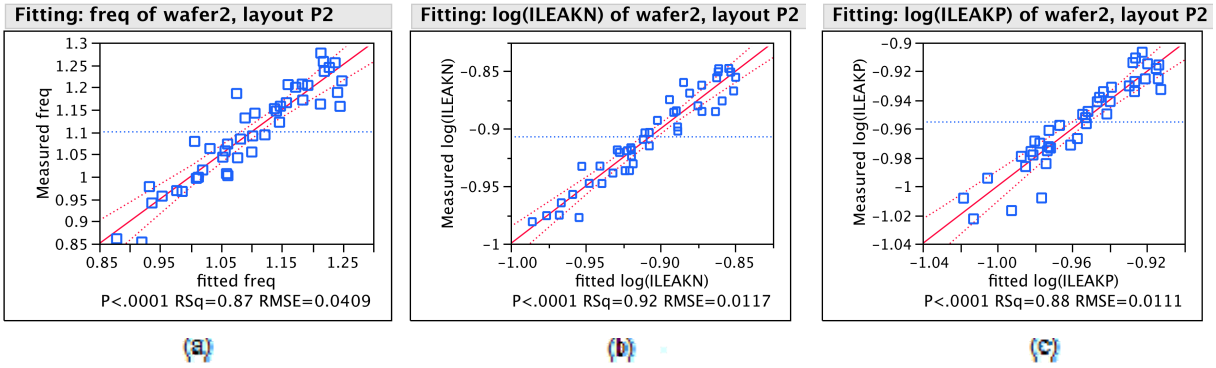

Fig. 7: Fitting across-wafer function for (a) $\overline{f}\langle\bullet,D,W2,P2\rangle$, (b) $\overline{\log(I_{LEAKN})}\langle\bullet,D,W2,P2\rangle$ and (c) $\overline{\log(I_{LEAKP})}\langle\bullet,D,W2,P2\rangle$. Symbols indicate data averaging across each measured chip.

5.2 Pattern dependent effects

In our previous work with 90nm data, the pattern dependent effect was modeled as a simple additive component. This means that the shape of the across-wafer and the across die functions were identical for all measured layout patterns. This assumption still holds true within each chip for the 45nm data, and if we normalize these constants to their chip average, the numbers are fairly consistent from chip to chip as shown in Fig. 8. This, however, is no longer true across the wafer. As shown in Fig. 9, the within-die pattern-to-pattern variation range is approximately proportional to the die average for frequency and NMOS leakage measurements. Meanwhile the two are almost uncorrelated for the PMOS leakage current measurement. This implies that for RO frequency and NMOS leakage current, layout effects now *interact* with the shape of the across-wafer systematic variability function. It is noteworthy that due to the apparent effectiveness of the OPC treatment the PMOS leakage current has little pattern to pattern variation and can still be modeled as an additive component.

Fig. 8 helps explain this behavior. Within-die pattern-to-pattern variation is first normalized to the die average so that the impact of die-to-die or wafer-to-wafer variation is excluded, then averaged over all the dies on the same wafer. Systematic pattern dependency is observed. Indeed, since PMOS leakage current does not vary much from one pattern to the next, we must conclude that the OPC algorithm was successful in removing most of the patterning related effects on L_{EFF}. This means that the pattern dependent variation observed in the NMOS leakage current and in RO frequency is likely to be the result of stress related effects on the NMOS devices. (Recall that PMOS current is insensitive to those effects in our configuration). This is further reinforced by the fact that $\log(I_{LEAKN})$ correlates strongly with RO frequency, but the trend of $\log(I_{LEAKP})$ does not. If L_{EFF} was the underlying reason, one would expect similar behavior from the frequency and both leakage currents..

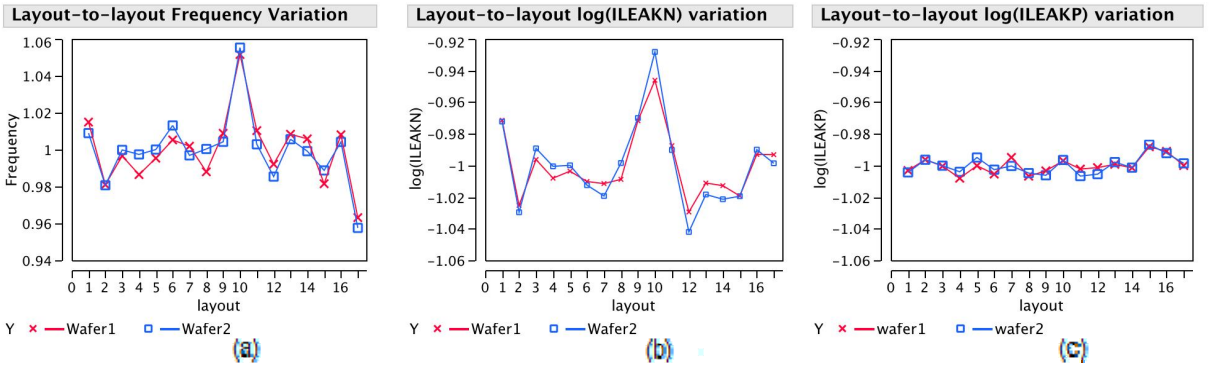

Fig. 8: Mean pattern-to-pattern variations: (a) $\overline{fL}\langle\bullet,W,P\rangle$, (b) $\overline{INL}\langle\bullet,W,P\rangle$, (c) $\overline{IPL}\langle\bullet,W,P\rangle$, here $fL\langle D,W,P\rangle \equiv \frac{\overline{f}\langle\bullet,D,W,P\rangle}{\overline{f}\langle\bullet,D,W,\bullet\rangle}$, $INL\langle D,W,P\rangle \equiv \frac{\overline{\log(I_{LEAKN})}\langle\bullet,D,W,P\rangle}{\overline{\log(I_{LEAKN})}\langle\bullet,D,W,\bullet\rangle}$ and $IPL\langle D,W,P\rangle \equiv \frac{\overline{\log(I_{LEAKP})}\langle\bullet,D,W,P\rangle}{\overline{\log(I_{LEAKP})}\langle\bullet,D,W,\bullet\rangle}$.

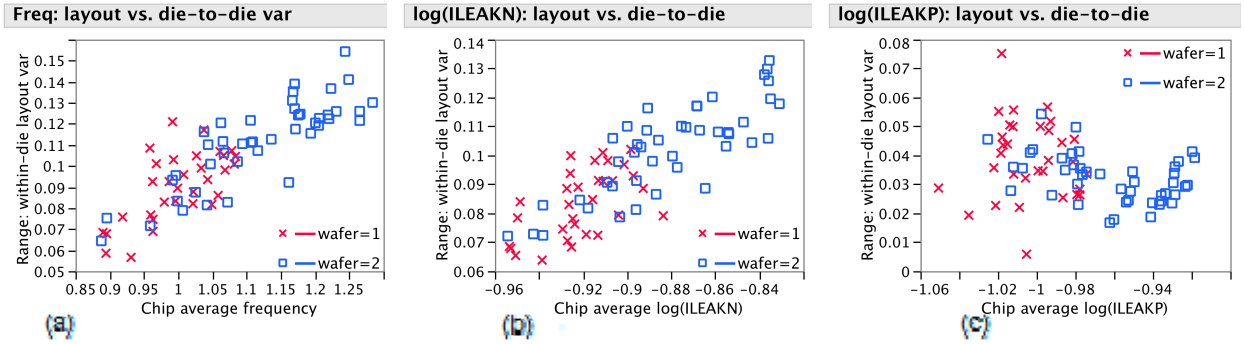

Fig. 9: (a) $\underset{P}{\text{range}}\,\overline{f}\langle\bullet,D,W,P\rangle$ vs. $\overline{f}\langle\bullet,D,W,\bullet\rangle$, (b) $\underset{P}{\text{range}}\,\overline{\log(I_{LEAKN})}\langle\bullet,D,W,P\rangle$ vs. $\overline{\log(I_{LEAKN})}\langle\bullet,D,W,\bullet\rangle$, (c) $\underset{P}{\text{range}}\,\overline{\log(I_{LEAKP})}\langle\bullet,D,W,P\rangle$ vs. $\overline{\log(I_{LEAKP})}\langle\bullet,D,W,\bullet\rangle$

5.3 Wafer-to-wafer variation

We measured 45nm test chips from two wafers. According to direct gate CD measurement data provided by the line that produced our test chips, there is a nominal 4nm split in the average gate length between the two wafers, and the spread of gate CD of the two wafers is about the same.

This split in gate length enables us to explore and validate our assumption that L_{EFF} variation is causing the across-wafer systematic variation. As shown in Fig. 6, the slower wafer (#1) and the faster wafer (#2) share a similar across-wafer function for RO frequency, $\log(I_{LEAKN})$, and $\log(I_{LEAKP})$. Because the sensitivity of the RO frequency to L_{EFF} increases as L_{EFF} gets shorter, the wafer with smaller L_{EFF} should have greater curvature in the across-wafer function. This phenomenon is indeed observed on Fig. 10a, where the across-wafer function of wafer #2 shows almost twice the range of that of wafer #1. While higher RO frequencies correspond to wider frequency range when comparing wafers due to non-linear dependency on L_{EFF}, this does not always apply to two patterns with different speeds, which again suggests the pattern-to-pattern difference is not likely to be caused by L_{EFF} change, as discussed in section 5.2. The L_{EFF} split is also helpful in identifying the accuracy and effectiveness of the physical models used for SPICE simulation, which will be discussed in the next section.

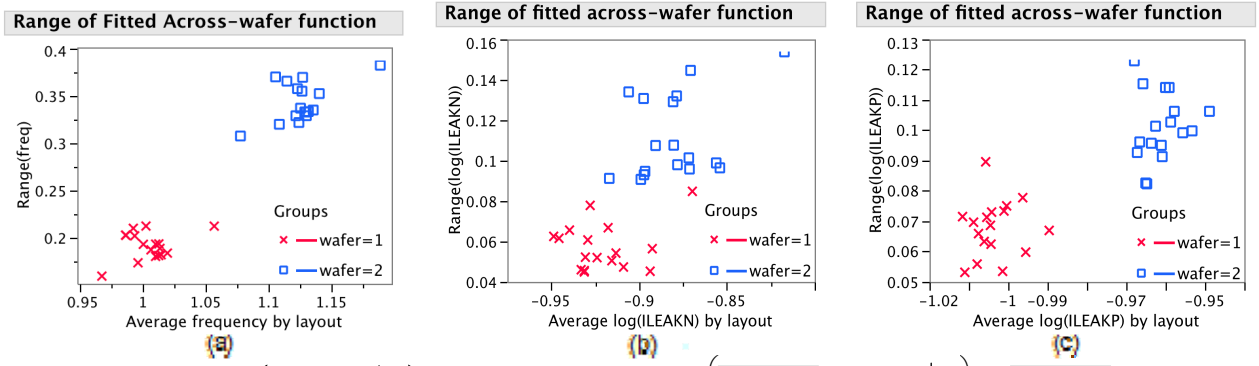

Fig. 10: (a) $\underset{D}{\text{range}}\left(\overline{f}\langle\bullet,D,W,P\rangle\big|_{AW}\right)$ vs. $\overline{f}\langle\bullet,\bullet,W,P\rangle$, (b) $\underset{D}{\text{range}}\left(\overline{\log(I_{LEAKN})}\langle\bullet,D,W,P\rangle\big|_{AW}\right)$ vs. $\overline{\log(I_{LEAKN})}\langle\bullet,\bullet,W,P\rangle$, (c) $\underset{D}{\text{range}}\left(\overline{\log(I_{LEAKP})}\langle\bullet,D,W,P\rangle\big|_{AW}\right)$ vs. $\overline{\log(I_{LEAKP})}\langle\bullet,\bullet,W,P\rangle$

As shown in Fig. 10b &c, the across-wafer systematic leakage variation is still larger for the faster wafer, though the split between two wafers is smaller compared to RO frequency. This can be explained by the V_{TH} roll-off characteristics. Since leakage current is linearly proportional to mobility but changes exponentially with threshold voltage, $\log(I_{LEAK})$ really reflects the threshold voltage variation while suppresses minor factors like mobility enhancement. PMOS transistors, in particular, are insensitive to stress, so that one may expect the gate length dependent threshold voltage roll-off effect will be the dominant mechanism. The slope of the V_{TH} roll-off curve is likely to be steeper for a smaller L_{EFF}, hence the same across-wafer L_{EFF} variation will result in larger change in V_{TH}.

Die-to-die random variations are obtained by removing the systematic across-wafer function from the chip average measurements. Die-to-die variation includes two components: the systematic residual from the polynomial fitting procedure, and the "true random" die-to-die variation. Since there is no practical way of modeling completely the systematic wafer function with limited measurements, we can only lump the two parts together as a single Gaussian variable. The standard deviation of this die-to-die random variable is plotted in Fig. 11. As for frequency (Fig. 11a), a very similar pattern is observed compared to the across-wafer range plot Fig. 10a, indicating that the systematic residual may be dominant. NMOS leakage and PMOS leakage, on the other hand, have die-to-die variation that is largely independent of average wafer speed and comparable between the two wafers, showing more of the random components which could be the result of field to field litho exposure or defocus variation.

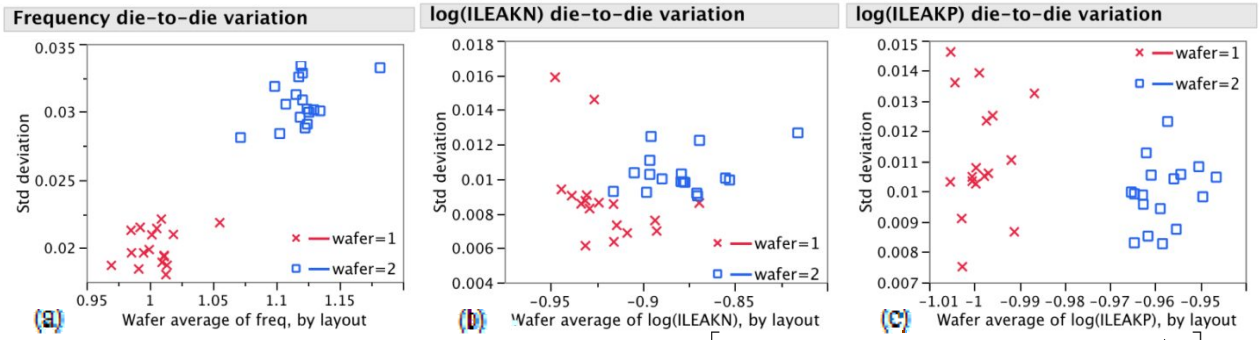

Fig. 11: (a) $\sigma_D\left(\overline{f}\langle\bullet,D,W,P\rangle-\overline{f}\langle\bullet,D,W,P\rangle\Big|_{AW}\right)$ vs. $\overline{f}\langle\bullet,\bullet,W,P\rangle$, (b) $\sigma_D\left[\overline{\log(I_{LEAKN})}\langle\bullet,D,W,P\rangle-\overline{\log(I_{LEAKN})}\langle\bullet,D,W,P\rangle\Big|_{AW}\right]$ vs. $\overline{\log(I_{LEAKN})}\langle\bullet,\bullet,W,P\rangle$, (c) $\sigma_D\left[\overline{\log(I_{LEAKP})}\langle\bullet,D,W,P\rangle-\overline{\log(I_{LEAKP})}\langle\bullet,D,W,P\rangle\Big|_{AW}\right]$ vs. $\overline{\log(I_{LEAKP})}\langle\bullet,\bullet,W,P\rangle$

5.4 Physical explanations and SPICE model examination

Two sources may account for the across-wafer systematic gate L_{EFF} variation. During post-exposure-bake (PEB), the wafer temperature is non-uniform during the rapid heating step [2]. Also during plasma etching, higher temperatures near the center of the wafer typically cause over etch, leading to faster devices [3].

While gate length variability accounts for the underlying mechanism of the across-wafer performance variation, it cannot explain the significant difference between the measured range of across-wafer variation and the SPICE simulated boundaries produced with L_{EFF} measurement data, as shown in Fig. 12. While the simulated frequency fits the actual frequency well near the nominal L_{EFF} of wafer #1 and #2 respectively, the faster or slower chips seem to be beyond the range where SPICE can provide accurate prediction.

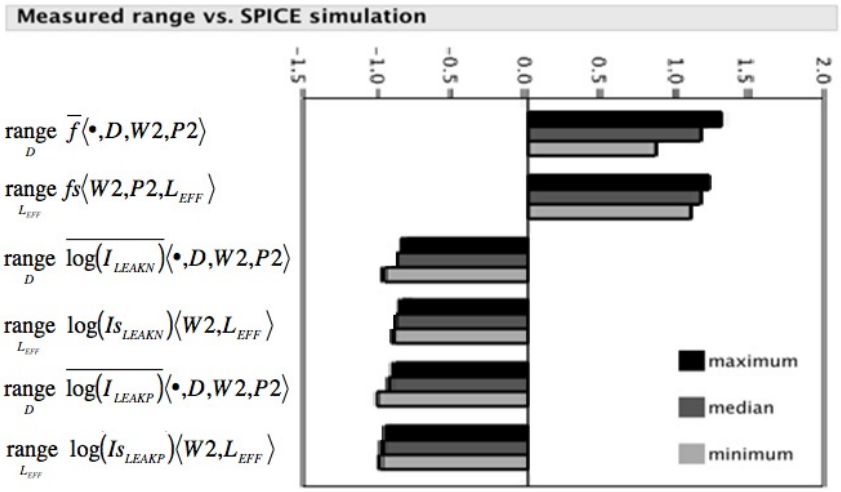

Fig. 12: Range of measurement vs. SPICE simulation (based on direct gate CD measurement) of RO frequency, $\log(I_{LEAKN})$ and $\log(I_{LEAKP})$.

This can be explained by the fact that the SPICE decks available to us did not yet have an accurate model for the strain effect introduced by the nitride capping-layer. Qualitatively, with the same amount of strain applied on the channel, transistors with shorter gate length will be subject to more uniaxial strain per unit channel length, therefore exhibiting a greater change in mobility. When L_{EFF} is very different from the target, the change in mobility due to strain becomes significant and results in more variability. Another potential reason is that the doping condition is not exactly the same as those defined in the SPICE corners, thus the real threshold voltage roll-off characteristics is different from the simulation. The greater range of the systematic across-wafer variation suggests that the actual slope of the V_{TH} roll-off curve is larger than what is defined in the model. These observations provide a simple approach to examine the effectiveness of SPICE model, and may help building a better physical model in the future.

6. WITHIN-DIE VARIATION

Die-level variation, or within-die variation, is the spatial variability that occurs in the range of the size of a die, and can also be decomposed into two components: the systematic (or deterministic) part, and the random local noise from device to device. Based on the hierarchical model, the total systematic variability of a die is the combination of a segment of slow varying across-wafer function and the within-die systematic variation. Due to the very small size of the die, the contribution of across-wafer function is expected to be negligible. Therefore, within-die variation can be obtained by simply removing the chip average from the raw data. Since the systematic within-die variation is assumed to be identical for all chips (an assumption that our data does not contradict), it can be extracted by taking the average of measurements of all the dies. Systematic within-die frequency and leakage measurement results are shown in Fig. 13 for pattern P2 from wafer #2 specifically. There is no significant systematic within-die variation in our dataset, however, possibly due to the relatively small area of our die (0.841mm x 0.94mm).

As shown in Fig. 14, standard deviation and mean of RO frequency and leakage currents for each chip is calculated for each pattern wafer combination, and plotted against each other. Within-die RO frequency variation is proportional to the average chip frequency, confirming that circuit performance of chips with shorter gate length is more susceptible to process variations. NMOS leakage however, does not show much correlation between variation and speed. Within-die random variation usually comes from such sources of randomness as threshold voltage variation introduced by random dopant fluctuations (RDF), SiO_2 interface noise[14], and line edge roughness (LER). According to Pelgrom's model [4], σ_{VTH} is inversely proportional to the square root of device size W*L. As the gate length L is varying from chip to chip across the wafer, missing this phenomenon indicates that the RDF effect cannot be the dominant mechanism for within-die random variation. For PMOS transistors, leakier chips actually have smaller within-die variation. While this is certainly counter-intuitive, it is most likely to be due to the limited accuracy of our measurement, especially for the very weak PMOS leakage currents provided by our test circuits.

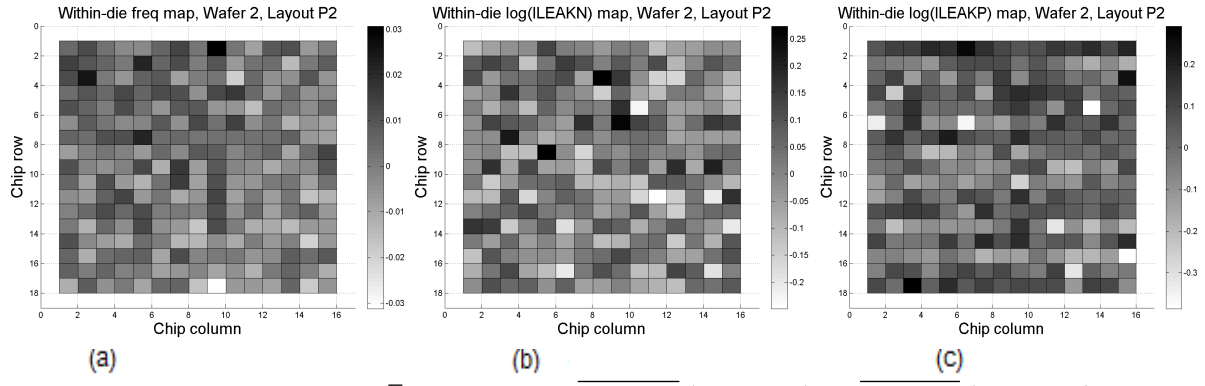

Fig. 13: Within-chip variation map: (a) $\overline{f}\langle T,\bullet,W2,P2\rangle$, (b) $\overline{\log(I_{LEAKN})}\langle T,\bullet,W2,P2\rangle$, (c) $\overline{\log(I_{LEAKP})}\langle T,\bullet,W2,P2\rangle$. Symbols indicate pattern P2 data averaged across all chips measured from wafer #2.

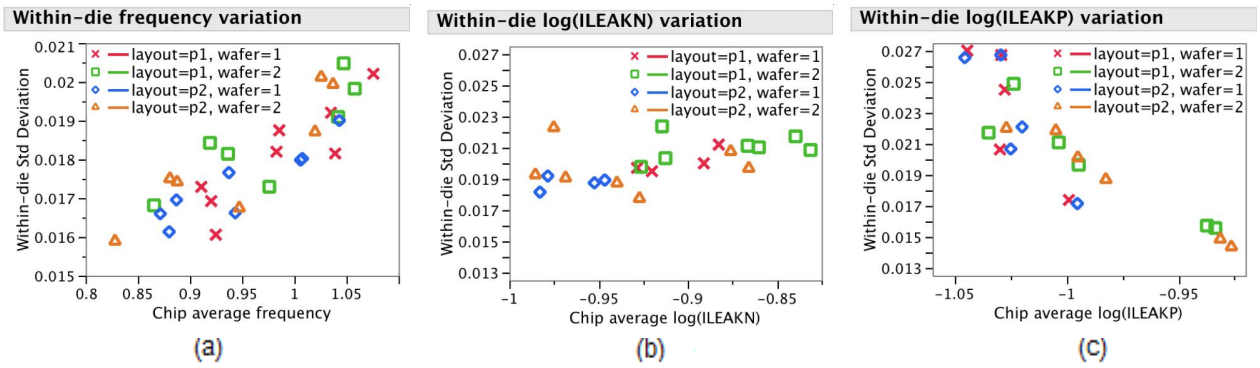

Fig. 14: (a) $\sigma_T f\langle T,D,W,P\rangle$ vs. $\overline{f}\langle\bullet,D,W,P\rangle$ (b) $\sigma_T \log(I_{LEAKN})\langle T,D,W,P\rangle$ vs. $\overline{\log(I_{LEAKN})}\langle\bullet,D,W,P\rangle$ (c) $\sigma_T \log(I_{LEAKP})\langle T,D,W,P\rangle$ vs. $\overline{\log(I_{LEAKP})}\langle\bullet,D,W,P\rangle$

7. MODEL SUMMARY AND APPLICATIONS

So far our hierarchical model captures the spatial variability of the 45nm test chips fairly well. Looking at wafer #2 only, the wafer-to-wafer component is then excluded. For the convenience of comparison, we treat the pattern-to-pattern variation as an additive component within the die instead of the multiplying factor from die to die, hence a single across-wafer function is used for all the patterns. Fig. 15 compares the relative standard deviation of the four components of the model. A significant amount of pattern dependent effects are observed in NMOS leakage but not PMOS, which supports our conclusion that NMOS transistors should account for the pattern dependency. Systematic across-wafer variation turns out to be the single most important term, especially for RO frequency measurements. This suggests that a reasonably good frequency prediction can be made if the across-wafer systematic function can be correctly calculated. Fig. 16 illustrates the ideal procedure of predicting the statistics of ROs with any given patterns from different wafers using a minimal amount of measurement. As discussed in section 5.4, however, there are some discrepancies between the SPICE model and the actual process. We compensate this by building empirical "layout effect" model using the exhaustive measurement for all the patterns on a small number of chips from wafer#1, and approximate the "process corner" using the across-wafer functions obtained from pattern P1 of both wafers. Thus we are capable of re-constructing the statistics of pattern P2 (or any other pattern!) on wafer #2, as shown in Fig. 17. The error in predicting the standard deviation is about 12%.

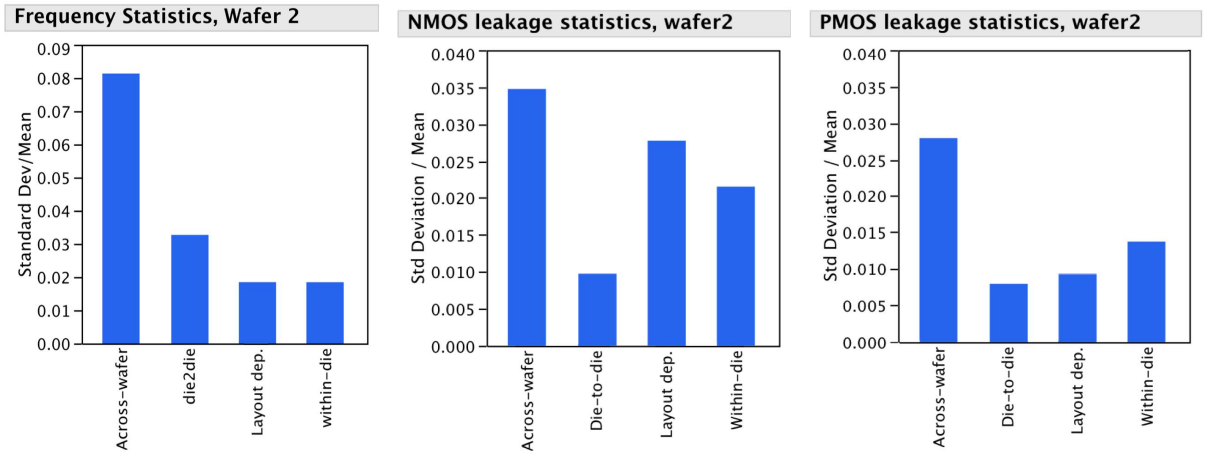

Fig. 15: Statistics of variability components for frequency and leakage measurement of wafer #2.

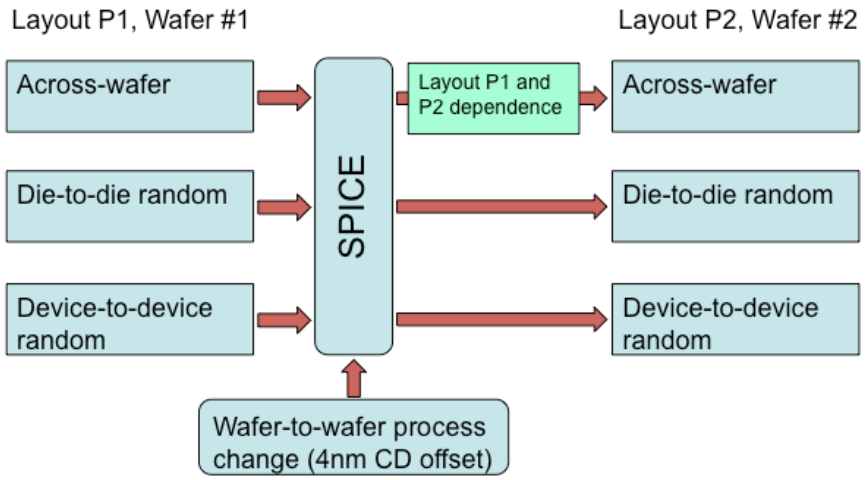

Fig. 16: Prediction procedure for RO frequency in ideal case, assuming SPICE model can correctly simulate the pattern dependent effects and the corners conditions are properly calibrated for each wafer.

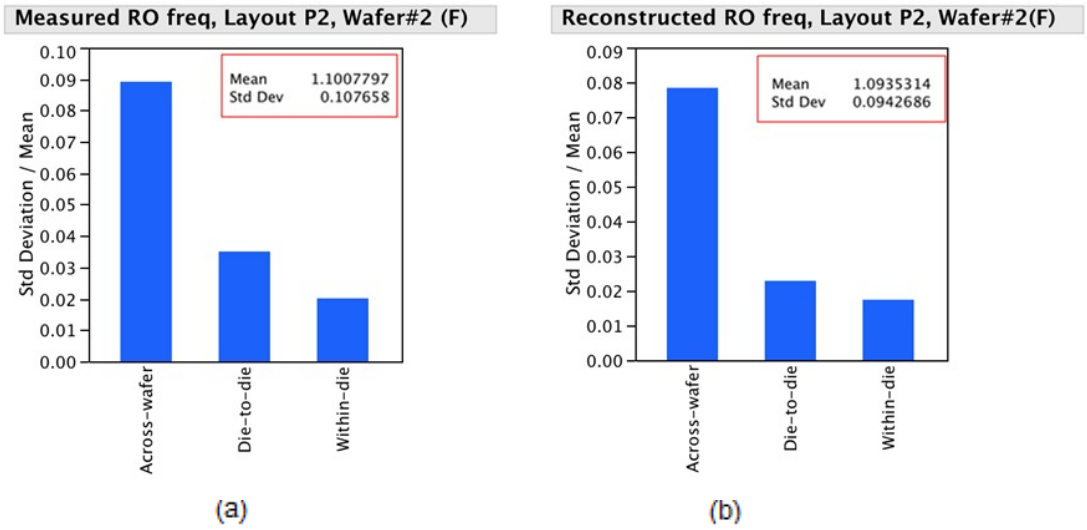

Fig. 17: (a) Measured freq of wafer #2 pattern P2 statistics vs. (b) reconstructed statistics

8. CONCLUSION

The hierarchical variability model is valid for the 45nm test chips with a few modifications on the additive assumptions of pattern-to-pattern and wafer-to-wafer variations. Systematic across-wafer variation contributes most to the total spatial variability, and can be sufficiently described by a 2nd order polynomial of the (x, y) position on the wafer. Further, examining two wafers that were processed under different process conditions, we see that the shape of the systematic variability does depend on the underlying process conditions. This is most likely due to the non-linear dependence between underlying variables, such as L_{EFF} who probably have identical deterministic shapes, and the higher level observed variables, such as RO frequency and leakage current. We also see two different types of pattern dependent effects: the OPC residual in gate length and the stress introduced threshold voltage change. The first type does not impact the other components much, while the latter seems to contribute to interact with across-wafer systematic and die-to-die random variation. Within-die systematic variation cannot be observed in this dataset, possibly due to the relatively

small size of our die, therefore all within-die variability is treated as random. In addition, the local variability in RO frequency is proportional to the average speed of the chip, but not so much for leakage current. Given these observations, our hierarchical model shows the capability of predicting the statistics of devices of any pattern from a given wafer based on a minimal set of measurements.

9. ACKNOWLEDGEMENT

The authors wish to acknowledge the contributions of the students, faculty and sponsors of the Berkeley Wireless Research Center, and wafer fabrication donation of STMicroelectronics. They would also like to thank Liang-Teck Pang and the National Science Foundation Infrastructure Grant No. 0403427 for the measurements. This work was funded by the U.C. Discovery Grant ele07-10283 and a number of participating companies based in California.

REFERENCES

[1] Kun Qian, Costas J. Spanos, "A Comprehensive Model of Process Variability for Statistical Timing Optimization", Proc. SPIE Int. Soc. Opt. Eng. 6925, 69251G (2008), DOI:10.1117/12.772980

[2] Qiaolin Zhang et al. "Across Wafer Critical Dimension Uniformity Enhancement Through Lithography and Etch Process Sequence: Concept, Approach, Modeling, and Experiment", Semiconductor Manufacturing, IEEE Transactions on (2007) vol. 20 (4) pp. 488-505

[3] Kanno et al. "Controlling gate-CD uniformity by means of a CD prediction model and wafer-temperature distribution control", Thin Solid Films (2007) vol. 515 (12) pp. 4941-4944

[4] M. Pelgrom, A. Duinmaijer, and A. Welbers, "Matching properties of MOS transistors," IEEE J. Solid-State Circuits, vol. 24, no. 5, pp. 1433– 1440, Oct. 1989.

[5] LT. Pang, K. Qian, CJ. Spanos, B. Nikolic, "Measurement and Analysis of Variability in 45nm Strained-Si CMOS Technology", JSSC 2008, ***to be published***

[6] L.-T. Pang, B. Nikolić, "Measurement and analysis of variability in 45nm strained-Si CMOS technology," in Custom Integrated Circuits Conference Dig. of Tech. Papers, Sept. 2008, pp. 129 – 132.

[7] E. Josse, et al, "A cost-effective low power platform for the 45-nm technology node," in *2006 IEEE Int. Electron Devices Meeting Tech. Dig.*, San Francisco, CA, USA, Dec 2006, pp. 1–4.

[8] C. Cam et al, "A low cost drive current enhancement technique using shallow trench isolation induced stress for 45-nm node," in *2006 Symp. VLSI Technology Dig. of Tech. Papers*, 2006, pp. 82–83.

[9] B. L. Gratiet, et al, "Process control for 45nm cmos logic gate patterning," in *Metrology, Inspection, and Process Control for Microlithography XXII*, ser. Proc. SPIE, J. A. Allgair and C. J. Raymond, Eds., vol. 6922, March 2008.

[10] Thompson et al. "A 90-nm logic technology featuring strained-silicon". *Electron Devices, IEEE Transactions on* (2004) vol. 51 (11) pp. 1790 – 1797

[11] L.-T. Pang, B. Nikolić, "Impact of Layout on 90nm CMOS Process Parameter Fluctuations," Dig. of Tech. Papers, 2006 Symp. VLSI Circuits, Hawaii, USA, pp. 69 – 70, 2006

[12] Verhaegen et al. "Compensating measured intra-wafer ring oscillator stage delay with intra-wafer exposure dose Corrections", Proceedings of SPIE (2006)

[13] A.V-Y Thean, et al., "Uniaxial-Biaxial Stress Hybridization For Super-Critical Strained-Si Directly On Insulator (SC-SSOI) PMOS With Different Channel Orientations," in *2005 IEEE Int. Electron Devices Meeting Tech. Dig.*, San Francisco, CA, USA, Dec 2005, pp. 509-512.

[14] Cellere et al. "Plasma-induced Si/SiO2 interface damage in CMOS", Microelectronic Engineering (2002)