Measurements and Analysis of Process Variability in 90 nm CMOS

Liang-Teck Pang, *Member, IEEE*, and Borivoje Nikolić, *Senior Member, IEEE*

***Abstract*—A test chip has been built to study the effects of circuit layout on variability, and to characterize within-die (WID) and die-to-die (D2D) variability of delay and leakage current in 90 nm CMOS technology. Delay is obtained through the measurement of ring oscillator frequencies, and the transistor leakage current is measured by an on-chip analog-to-digital converter (ADC). It has been found that the transistor performance depends strongly on the polysilicon (poly-Si) gate density, and the spatial correlation depends on the gate orientation and the direction of poly-Si spacing. WID variation is small with three standard deviations over a mean ($3\sigma/\mu$) of around 3.5%, whereas D2D and systematic layout-induced variations are significant, with a $3\sigma/\mu$ D2D variation of $\sim$15% and a maximum layout-induced frequency shift of 10%. Finally, a set of guidelines is proposed to help circuit designers mitigate the effects of process variations on CMOS performance.**

***Index Terms*—CMOS, leakage, polysilicon, ring oscillators, variability.**

I. Introduction

CMOS device scaling has increased the impact of process variability to the point where it is now regarded as a major roadblock to further scaling [1]. The control of process fluctuations has not kept pace with rapidly shrinking device dimensions. Furthermore, the drive to improve performance has enticed device and circuit designers to operate at conditions that are more sensitive to variability.

Process variations can be systematic or random, and are generally characterized as within-die (WID), die-to-die (D2D) and wafer-to-wafer (W2W) [2]. WID and D2D classifications reflect some of the spatial characteristics of the variations. Those that vary rapidly over distances smaller than the dimension of a die result in WID variations, whereas variations that change gradually over the wafer will cause D2D variations. W2W variations reflect both the spatial as well as temporal characteristics of the process and cause different wafers to have different properties. Systematic variations are deterministic shifts in process parameters, whereas random variations change the performance of any individual instance in the design in an unpredictable manner. In practice, although many of the systematic variations have a deterministic source, they are not known at the design time, or are too complex to model and are treated as random. The resulting random variation component will have a varying degree of spatial correlation. The primary sources of variability are the transistors, the interconnects, and the operating environment (supply and temperature) [3].

Many sources of systematic variability can be attributed to the particular characteristics of the manufacturing process. Deviations in nominal widths and lengths are caused by photolithography and etching. The variation in film thicknesses (e.g., oxide thickness, gate stacks, and wire and dielectric layer height) is caused by the deposition and growth process, as well as the chemical-mechanical planarization (CMP) step. Additional electrical properties of CMOS devices are affected by variations in the dosage of implants, as well as the temperature of annealing steps. Finally, random device parameter fluctuations stem from line-edge roughness [4], Si/SiO_2 and polysilicon (poly-Si) interface roughness [5], and doping fluctuations [6].

Presently, variability is captured in the design process through the use of simulation corners, where certain transistor parameters are varied by three standard deviations from their nominal values. This approach typically regards all variations as D2D, with all devices on a chip having correlated process parameters. As the corner spread increases with scaling, it becomes challenging to simultaneously satisfy performance, power and corner requirements. In order to better account for the variability, it is necessary to develop a deeper understanding of the contributions of random and systematic variations and to distinguish between WID and D2D components. It is also important to determine the spatial correlation distance of process parameters; yield can be improved by averaging out uncorrelated variations [7]. Detailed characterization of variations will allow circuit designers to incorporate design methodologies that mitigate the effects of systematic as well as random variations.

In this paper, various test structures are used to measure variability and analyze its systematic, random, WID and D2D components. Ring oscillators have been used in the past to study process variations, and are also adopted in our study. To decouple the impact of various process parameters on electrical variability, the ring oscillator measurements are augmented with the leakage current measurements of transistors in the off state. Our work features a set of simple layout variations, introduced to emphasize particular processing effects. The impact of different layout topologies on variation is investigated, and the spatial correlation of ring oscillator frequencies and leakage currents are characterized. Section II gives an overview of the

Manuscript received June 11, 2008; revised December 08, 2008. Current version published May 01, 2009. This work was supported by the National Science Foundation under NSF Infrastructure Grant No. 0403427, and the Center for Circuit and System Solutions (C2S2) Focus Center, one of five research centers funded under the Focus Center Research Program, a Semiconductor Research Corporation program. Wafer fabrication was provided by STMicroelectronics.

L.-T. Pang was with the Berkeley Wireless Research Center, University of California, Berkeley, CA 94704-1302 USA. He is now with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (e-mail: ltpang@eecs.berkeley.edu).

B. Nikolić is with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA 94720-1770 USA (e-mail: bora@eecs.berkeley.edu).

Digital Object Identifier 10.1109/JSSC.2009.2015789

0018-9200/$25.00 © 2009 IEEE

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on May 4, 2009 at 13:53 from IEEE Xplore. Restrictions apply.

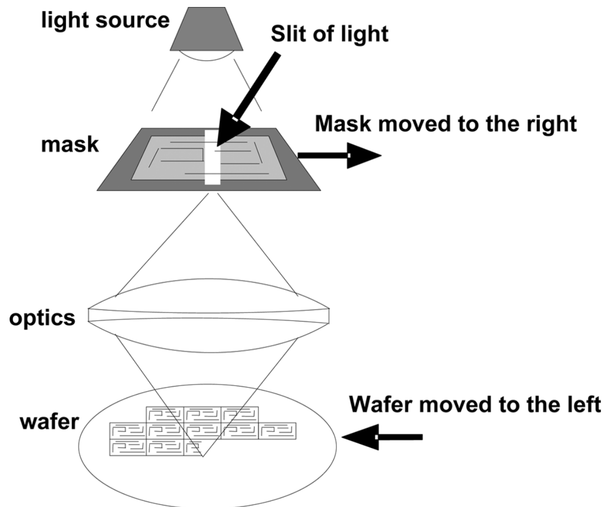


Fig. 1. Step and scan photolithography using a slit of light.

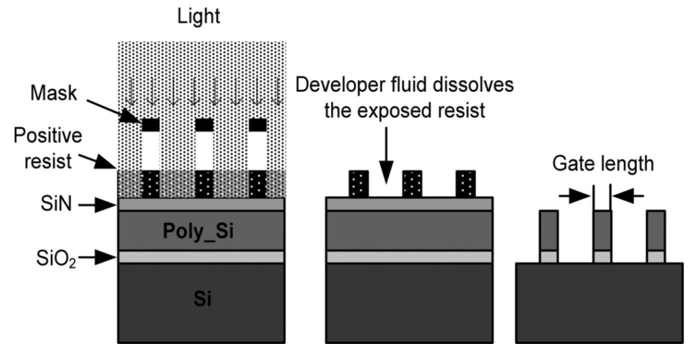


Fig. 2. Formation of poly-Si gates.

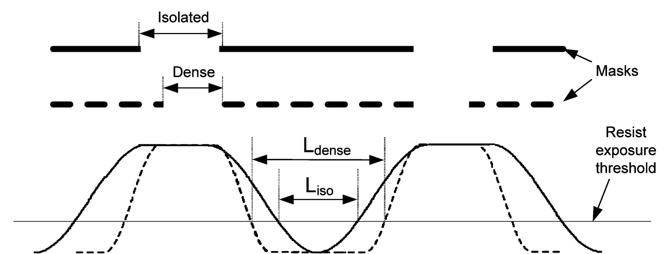


Fig. 3. Isolated and dense lines.

sources of variations in the manufacturing process. It begins with a discussion of patterning-induced variations, which have been observed to be the dominant source of systematic variations, and continues with a description of other manufacturing effects that contribute to systematic variations. The test chip is described in Section III, and the measurement and analysis results are presented in Section IV. Finally, Section V concludes with a list of guidelines for robust digital CMOS design.

II. Manufacturing-Induced Variations

Present lithography systems employ a step-and-scan method, where the stepping is used to move a wafer between the exposure fields. Within the exposure field, a narrow slit of light illuminates the mask, and the mask pattern is optically projected onto a wafer [8]. The mask and the wafer are moved simultaneously in opposite directions so that the slit of light scans the entire mask and projects the image onto the wafer. This is illustrated in Fig. 1. Movement of the wafer stage is controlled by look-ahead sensors.

In sub-wavelength lithography, the effective line width depends on the surrounding features [9]–[11]. The process step of patterning poly-Si gates is shown in Fig. 2. When exposed beyond a certain light intensity, positive resist will dissolve in the developer fluid. The exposed poly-Si gate stack layer will then be etched away, leaving behind the transistor gates. Narrow poly-Si lines with varying pitch will have different channel lengths when exposed with the 193 nm wavelength light, as illustrated in Fig. 3. Dense lines also have a higher depth of focus and are more immune to defocusing of the optical system [12]. Optical proximity correction techniques in the mask processing add sub-lithographic assist features to control the printed critical dimensions (CD). However, their effect may be limited due to the shallow depth of field.

After the resist is spun onto the wafer and exposed, it undergoes post-exposure bake (PEB). This step is essential to activate the photoactive compound and to set the resist exposure threshold [13]. If the temperature over the entire wafer is not even, this will result in different resist exposure thresholds over the wafer and cause D2D gate-length variation. This effect is mostly systematic, and wafers have exhibited a radial temperature profile during baking.

Dry etching is employed in the patterning of the poly-Si gates. This process suffers from microscopic loading effects [8] in which densely spaced poly-Si gates experience a lower etch rate than isolated poly-Si, resulting in dense poly-Si gates having longer gate-lengths than that of isolated poly-Si gates.

Modified illumination [9], which involves changing the partial coherence factor of the light source combined with off-axis illumination schemes such as dipole, quadrupole and annular illumination, is used to improve the resolution. This technique, however, can cause line-widths to vary according to their orientation [14] and can also exacerbate the effects of lens aberrations [15].

Lens imperfections are often characterized through aberrations (Fig. 4). Aberrations create optical path differences for each pair of rays through the imaging system. Effects such as astigmatism and spherical aberrations cause differences in exposed patterns at the level of the reticle. A coma effect [15] is an aberration caused by lens imperfection, which causes a gate surrounded by a non-symmetrical structure to print differently from its mirror image [16]. The scattering and reflection of light through the projection system results in flare, which causes variations in the effective CDs (Fig. 5). In general, the amount of flare depends on the local pattern density in the mask [15].

Another source of variability could come from e-beam mask stitching [17] discontinuity. E-beam lithography is employed in writing the optical mask. In order to cover the large reticle field with a small e-beam field, a mask needs to be constructed by drawing smaller e-beam fields and stitching the whole image together. Even though optics will reduce imperfections on the mask by a few times, aggressive scaling has made stitching discontinuities more significant. Similarly, variation in speeds

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on May 4, 2009 at 13:53 from IEEE Xplore. Restrictions apply.

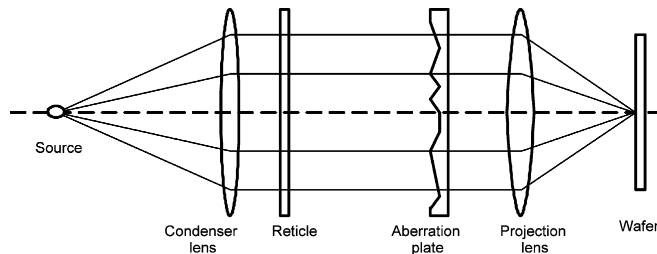


Fig. 4. Lens imperfections.

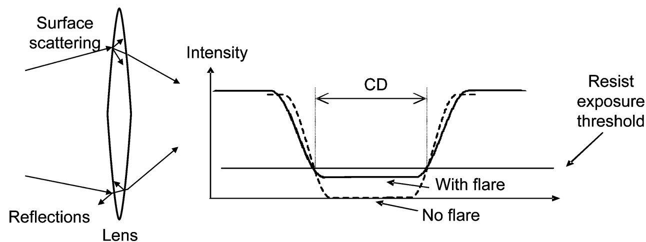


Fig. 5. Effects of flare.

during the scanning process may cause systematic variations over the reticle.

A mask overlay error due to a misalignment of the different masks can cause shifts in the position of the poly-Si gate with respect to diffusion and to source/drain contacts. Shadowing during pocket implantation can result in different properties in the source and drain of the transistors. These effects, like the coma effect, could cause symmetrical structures to have non-symmetrical properties.

Rapid thermal annealing (RTA) is used in activating implanted dopants and annealing defects without causing significant dopant diffusion [8]. The temperature distribution over the wafer surface could be non-uniform, resulting in wafer-level systematic variations. For example, the edge of the wafer could experience more cooling and hence a lower temperature, resulting in a radial temperature gradient across the wafer. This effect could also cause wafer warping that results in variation in substrate stress. Finally, local variation in reflectivity of the wafer surface could cause local non-uniform heating, resulting in variation in threshold voltage (V_T) and external resistance ($R_{\rm EXT}$), which includes mainly source-drain and contact resistance [18]. It has also been shown in [19] that differences in anneal temperature can cause shifts in V_T due to non-passivation of interface states.

Proximity effects, aberrations, flare, overlay errors, shadowing and RTA are usually not captured in the design process, and they induce layout-dependent systematic variations in the design.

III. Test Chip

A test chip has been implemented in a general-purpose 90 nm CMOS technology to evaluate the effects of lithography-induced variations, and to measure WID and D2D variations, as well as WID spatial correlation. This is done by measuring ring oscillator (RO) frequencies and transistor source-drain leakage currents ($I_{\rm LEAK}$) of an array of test-structures [20], [21].

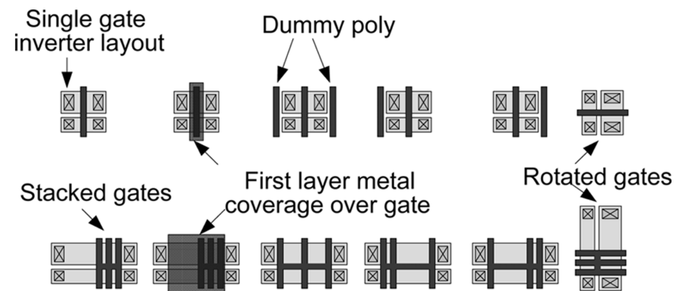


Fig. 6. Layout variations in the test chip, pictured here in the RO inverter stages.

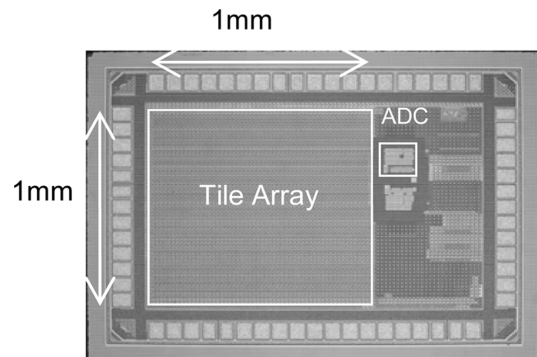


Fig. 7. Test chip die photo.

The chip contains 10×16 tiles, with a total area of 1 mm $\times$ 1 mm. Each tile has twelve 13-stage ROs and twelve transistors in the off-state, each with a different layout (Fig. 6). Some transistors in both the ROs and the off-state consist of a single poly-Si finger, while the rest are constructed with a stack of three fingers. The poly-Si pitch of neighboring dummy poly-Si lines is varied in the test structures to capture proximity effects. Poly-Si orientations, together with the properties of the two-dimensional tile array, are used to characterize spatial correlation. Non-symmetrical structures and their mirror image target measurements of the coma effect. The first layer metal coverage over certain gates are also varied in the layout to investigate the effects of anneal [19]. Dummy fills are not used on these layouts for all layers in order to preserve the topology.

The test chip shown in Fig. 7 includes an on-chip single-slope analog-to-digital converter (ADC) to measure transistor off-currents from 1 nA to 1 μA with 1 nA resolution. The ADC (Fig. 8) consists of a high-gain folded-cascode amplifier implemented with 2.5 V devices, a large on-chip metal fringe capacitor, and comparators. The switches, controlled by signals P1, P2, and P1b, provide the correct bias to the input and output of the amplifier, and they determine whether to measure the reference current ($I_{\rm ref}$) only or the sum of $I_{\rm ref}$ and I_s. The latter is the sum of all parasitic gate and substrate leakage currents and the $I_{\rm LEAK}$ of the selected device. By not selecting any of the devices, it is possible to measure the parasitic leakage currents and subtract them from the measurement. $I_{\rm ref}$ is obtained from an external source through current mirrors that divide the current down by a factor of ten.

During current integration, the output of the op-amp will ramp up, and as it crosses the voltages $\rm V_1$ and $\rm V_2$, the start and

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on May 4, 2009 at 13:53 from IEEE Xplore. Restrictions apply.

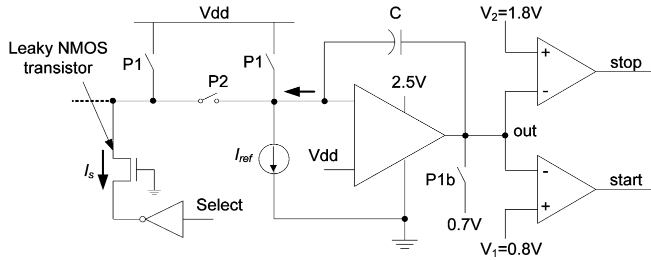
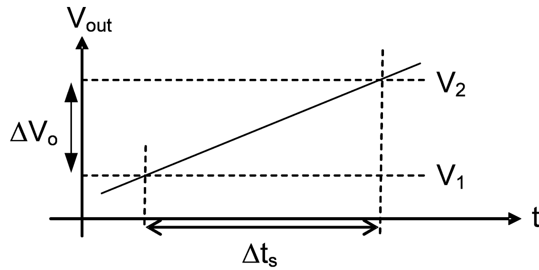


Fig. 8. Single slope ADC for current measurement.



$$\frac{\Delta t_{ref} - \Delta t_s}{\Delta t_s} \sim \frac{I_s}{I_{ref}}$$

Δt_{ref} = integration time when measuring I_{ref}

Δt_s = integration time when measuring $I_{ref} + I_s$

Fig. 9. Current measurement procedure.

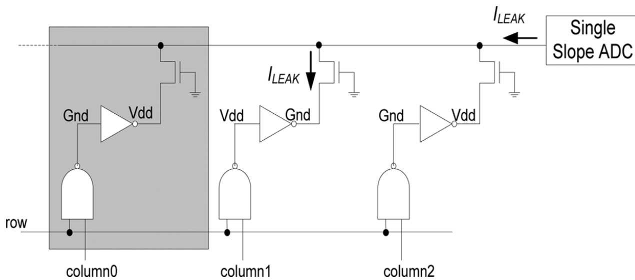


Fig. 10. Standby leakage current measurement.

stop signals will be generated. By measuring the time interval between start and stop signals the current is obtained (Fig. 9).

The selection of the device to be measured is done with row and column bits obtained from a shift register in order to reduce pad count. The selection of the NMOS transistor for leakage current measurement is done by applying either Vdd or Gnd to the source of the transistor using a large inverter (Fig. 10). For frequency measurements, two address bits per row and a column bit are used. These will enable the selected RO to go through a NAND gate placed in the ring, and select the multiplexer to output its own RO frequency instead of the previous one. Its output goes through a series of multiplexers to a frequency divider that outputs to a pad (Fig. 11) [22]–[24].

The chips are packaged in ceramic PGA packages and measured through a socket that is mounted on a custom PCB. Frequency measurements were done using a 20 GS/s oscilloscope, and a 150 ps resolution logic analyzer was used for measuring the delay between the start and stop signals of the ADC. In order

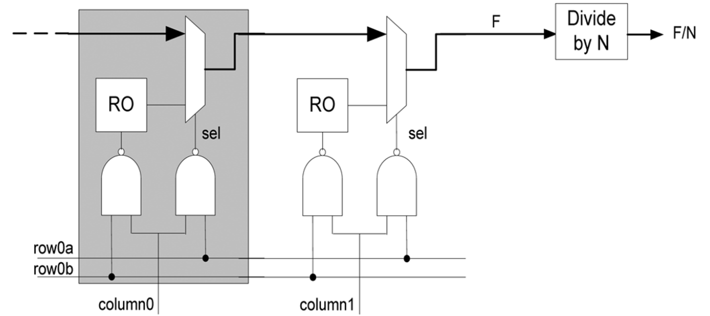


Fig. 11. RO frequency measurement.

to improve accuracy and reduce the impact of noise, several hundred measurements have been averaged for each data point.

IV. Measurement Results

In the analysis of the measurement results, tiles at the perimeter of the array are ignored to eliminate edge effects. The ROs with single isolated poly-Si fingers often exceed the fast corner and are excluded from the analysis.

To distinguish systematic effects from random effects, the results for different layouts are compared for each die, across all dies. The measurements for identical structures within a die are compared to obtain WID correlations. The averages for various dies are used to indicate the D2D spread. The averages across several dies are used to analyze the mask writing effects.

The measured data shows several important trends that are analyzed in detail in this section. The largest impact of the layout on performance comes from the gate poly-Si density, which causes a systematic shift in frequency of up to 10%. The D2D variation is significant, resulting in a $3\times$ standard-deviation/median $(3\sigma/\mu)$ of 15% over half a wafer. Finally, the WID variation for identical structures is relatively small $(3\sigma/\mu \sim 3.5\%)$. The WID spatial correlation of RO frequency shows a dependency on the direction of spacing and the orientation of the gates. Each of these observations can be attributed to a particular step in the manufacturing process.

A. Effects of Layout on Frequency and Leakage

In order to investigate the systematic layout effects, the RO frequency and $I_{\rm LEAK}$ distribution of different layout structures is compared. These variations are divided into WID and D2D categories. Proximity effects due to different poly-Si gate pitches cause a shift in frequency of over 10%, and a 20x shift in $I_{\rm LEAK}$ (Figs. 12 and 13). More isolated gates have shorter gate lengths and thus have higher RO frequencies and higher leakage currents. This is much larger than the 1.1% frequency shift of the extracted layout predicted by Spice simulations, which only capture changes in parasitic capacitances. D2D and WID leakage current variations are reduced with increased poly-Si density. There is also a similar but smaller effect of poly-Si density on stacked gates. Fig. 14 plots the RO frequency distribution for stacked gates. In order to only show the proximity effects that are not captured by layout extraction, the measured frequencies are normalized to the TT corner. These plots show that the frequency of the densest stacked gate configuration is 5% slower than the fastest layout.

The difference between structures, which are mirror images of each other, is small and is likely caused by the coma effect, or

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on May 4, 2009 at 13:53 from IEEE Xplore. Restrictions apply.

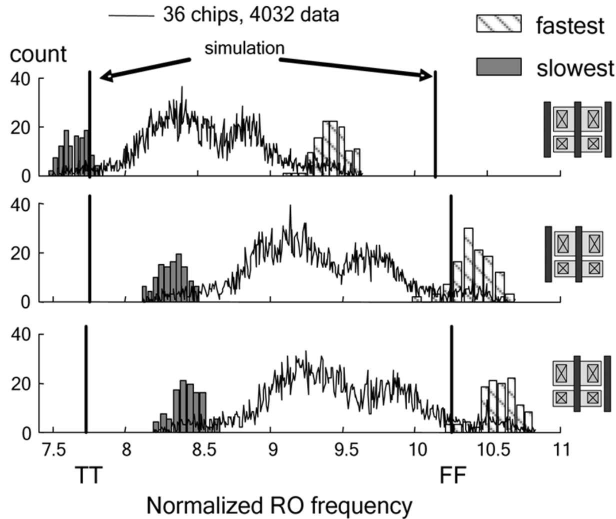


Fig. 12. Normalized frequency distribution for single-finger configurations. Vertical lines correspond to typical and fast corner simulation results. Bar plots correspond to the WID distribution of the fastest and slowest chip.

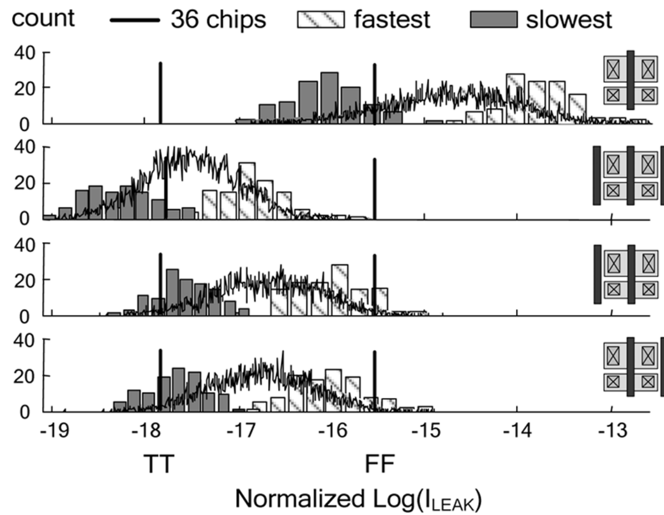


Fig. 13. $\mathrm{Log}(I_{\mathrm{LEAK}})$ distribution for single-finger configuration.

by the asymmetry in the dopant implantation. A roughly 1%–2% shift in the mean RO frequency with single gate stages was observed, as shown in the second and third plots of Fig. 12. The variation $(3\sigma/\mu)$ is smaller for rotated gates across all chips. The M1 coverage over gates has negligible effects on RO frequency and causes a small reduction in leakage current, indicating a small increase in V_T. It has been reported [19] that metal-covered gates suffer from a larger number of interface states because the metal coverage lowers the temperature on the gate during annealing, which results in V_T shifts of ~10%. In our layout, the metal 1 is in close proximity to a non metal-covered gate since the gate-length is small and the source and drain are fully covered with metal-1 contacts. It is likely that the temperature difference during annealing between metal-covered gates and non metal-covered gates is too small to create a significant shift in V_T. It is also possible that the process has sufficiently high annealing temperature to passivate the interface states of metal-covered gates, resulting in a small V_T shift due to different annealing temperatures.

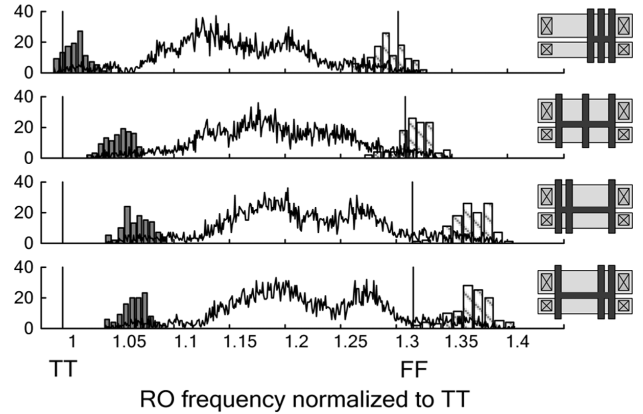


Fig. 14. Distribution plots of RO frequency normalized to its corresponding TT simulation corner for stacked gates. Vertical lines represent simulation corners from extracted layout. Less dense gates show a 5% increase in frequency compared to dense gates.

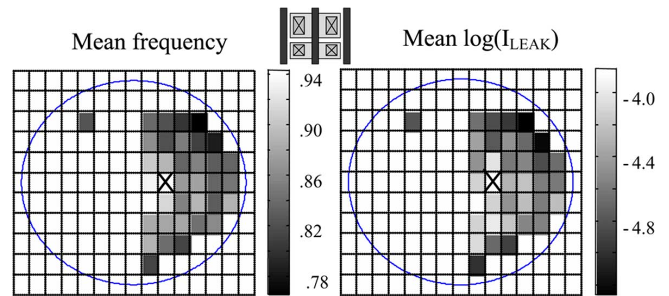


Fig. 15. Wafer maps of mean RO frequency and mean $\mathrm{Log}(I_{\mathrm{LEAK}})$. X marks a defective chip. Location of dies on only half the wafer are known.

B. D2D Variations

Wafer maps of the mean frequency and the leakage of each die for the layout configuration shown in Fig. 15 show a typical radial pattern that can be attributed to non-uniform resist development [25], [26] or systematic temperature gradient during RTA [8]. Faster and leakier chips are located at the center of the wafer. The D2D variation is significant, resulting in a $3\sigma/\mu$ of 15% over half a wafer for the densest single gate structure. For the other single gate structures, it slightly increases to around 17%.

C. WID Variations

The WID variation of identical layout structures is small $(3\sigma/\mu \sim 3.5\%)$ and weakly dependent on the layout. For each layout structure, the data for each chip is normalized to zero mean and unit variance before being used to compute its spatial correlation. Confidence intervals for the correlation are computed using Fisher's z-transformation to convert Pearson's correlation to a normally distributed random variable. In the frequency measurements, the spatial correlation depends only on the direction of spacing and the orientation of the gates.

The spatial correlation is higher in the horizontal direction of the chip than in the vertical direction. In particular, for horizontally spaced ROs (Fig. 16(a)), correlation is higher for vertically oriented gates, whereas for vertically spaced ROs (Fig. 16(b)), correlation is higher for horizontally oriented gates. This data is summarized in Table I, showing that horizontally spaced ROs

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on May 4, 2009 at 13:53 from IEEE Xplore. Restrictions apply.

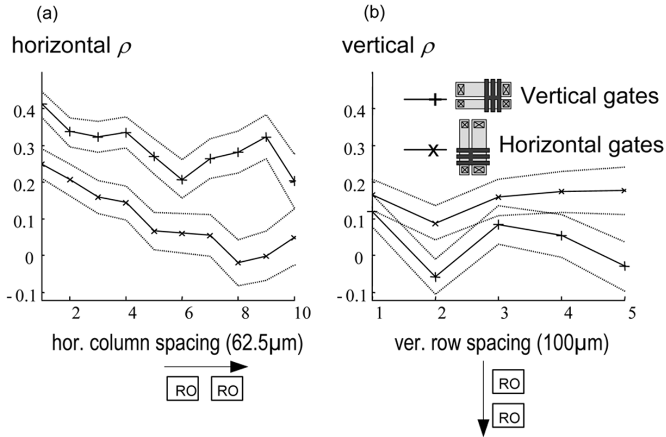


Fig. 16. Spatial correlation coefficient versus (a) horizontal column spacing and (b) vertical row spacing for vertical and horizontal gates. Dotted lines represent 99% confidence bounds for ρ.

TABLE I
SPATIAL CORRELATION COEFFICIENT OF NEIGHBORING ROs FOR DIFFERENT GATE ORIENTATION AND SPACING DIRECTION

Slit direction = horizontal; Scan direction = vertical		Horizontal spacing	Vertical spacing
RO Freq.	Vertical Gates	0.4	0.1
	Horizontal Gates	0.25	0.2
I_{LEAK}		0	0

with vertically oriented gates have a spatial correlation of 0.4 between immediate neighbors, whereas horizontally spaced ROs with horizontally oriented gates have a spatial correlation of 0.25.

This dependence on direction could be explained by the step-and-scan photolithography. The horizontal direction in this chip is along the slit of light in the stepper and is subject to lens aberrations and curvature, resulting in more correlated features. The vertical direction is along the scan direction, which is subject to variation in scan speeds and stage vibration that affects focus and light dosage, resulting in less correlated features.

Other manufacturing effects such as e-beam mask stitching [17] discontinuity or a systematic shift in the wafer stage scan speed could also account for the observed dependence on direction. In order to investigate the systematic reticle field effects, the normalized data from 36 chips is averaged, and Fig. 17 shows the resulting surface plot of RO frequency and $I_{\rm LEAK}$. There is a sharp systematic change of 1% in frequency between the 4th and 5th rows for RO structures with vertically oriented gates and a small shift in $I_{\rm LEAK}$. This could be attributed to a discontinuity at the level of the mask reticle, which suggests e-beam mask stitching defects. Another reason could be a change in the scan speed of the wafer/mask stage during scanning of the reticle image onto the wafer. Since the scan speed is adjusted simultaneously during a scan, based on look-ahead sensors, it is possible that there is discontinuity in the sensor readings, causing the scan speed to shift from the 4th to 5th row.

TABLE II
MEASUREMENT CONDITIONS

No.	Vdd (V)	Vbs (V)	T (°C)
1	1	0	25
2	**1.4**	0	25
3	1	**- 0.2**	25
4	1	0	**60**

The leakage current has a much smaller spatial correlation because it is less sensitive to L but more sensitive to V_T than RO frequency.

Varying degree of correlation in the two directions can be utilized by statistical timing analysis tools and can have an impact on memory and analog layouts.

D. Inferring Process Parameters

The scatter plot of $\mathrm{Log}(I_{\rm LEAK})$ versus frequency shows a strong positive correlation between $I_{\rm LEAK}$ and the frequency for D2D and layout-to-layout (L2L) variations (Fig. 18). This means that these variations are dominated by process parameters that cause both $\mathrm{Log}(I_{\rm LEAK})$ and frequency to vary in the same manner. The WID scatter plot shows no significant correlation, suggesting that it is caused by a combination of process parameters.

In order to relate these variations to process parameters, we use the least squares method and the BSIM3 model to infer the variation in gate length (L), gate oxide thickness $(T_{\rm ox})$ and channel dopant concentration $(N_{\rm ch})$ from the frequency (F) and leakage current $(I = \mathrm{Log} I_{\rm LEAK})$ measurements. Since the sensitivities of the measured data to process parameters vary with different conditions of supply voltage $(V_{\rm dd})$, substrate voltage $(V_{\rm bs})$ and temperature (T), the accuracy of the inference has been improved by taking the measurements under the different conditions listed in Table II.

The equations below are used in the estimation process. If index i represents the four different measurement conditions, then:

- F_m, I_m: Estimated mean F and I
- L_m, Tox_m, Nch_m: Estimated mean process parameters
- F_0, I_0: Mean measurement data for 1 die and 1 layout
- $L_{\rm typ}, Tox_{\rm typ}, Nch_{\rm typ}$: Typical values of L, *Tox* and *Nch*

$$Y_i = \begin{bmatrix} \frac{(F-F_m)}{F_0} \\ \frac{(I-I_m)}{I_0} \end{bmatrix}_i, \quad X = \begin{bmatrix} \frac{(L-L_m)}{L_{\rm typ}} \\ \frac{(Tox-Tox_m)}{Tox_{\rm typ}} \\ \frac{(Nch-Nch_m)}{Nch_{\rm typ}} \end{bmatrix},$$

$$M_i = \begin{bmatrix} \frac{\partial\left(\frac{F}{F_0}\right)}{\partial\left(\frac{L}{L_{\rm typ}}\right)} & \frac{\partial\left(\frac{F}{F_0}\right)}{\partial\left(\frac{Tox}{Tox_{\rm typ}}\right)} & \frac{\partial\left(\frac{F}{F_0}\right)}{\partial\left(\frac{Nch}{Nch_{\rm typ}}\right)} \\ \frac{\partial\left(\frac{I}{I_0}\right)}{\partial\left(\frac{L}{L_{\rm typ}}\right)} & \frac{\partial\left(\frac{I}{I_0}\right)}{\partial\left(\frac{Tox}{Tox_{\rm typ}}\right)} & \frac{\partial\left(\frac{I}{I_0}\right)}{\partial\left(\frac{Nch}{Nch_{\rm typ}}\right)} \end{bmatrix}_i,$$

$$\Rightarrow Y_i = M_i \cdot X \cdot \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} M_1 \\ M_2 \\ M_3 \\ M_4 \end{bmatrix} .X \Rightarrow X = M^{+}Y.$$

Fig. 19 illustrates how the estimated normalization constants, F_m, I_m, and L_m are obtained in the case where only a single parameter L is to be estimated. By using Spice simulations, F and I are plotted versus L. The region of L that results in a

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on May 4, 2009 at 13:53 from IEEE Xplore. Restrictions apply.

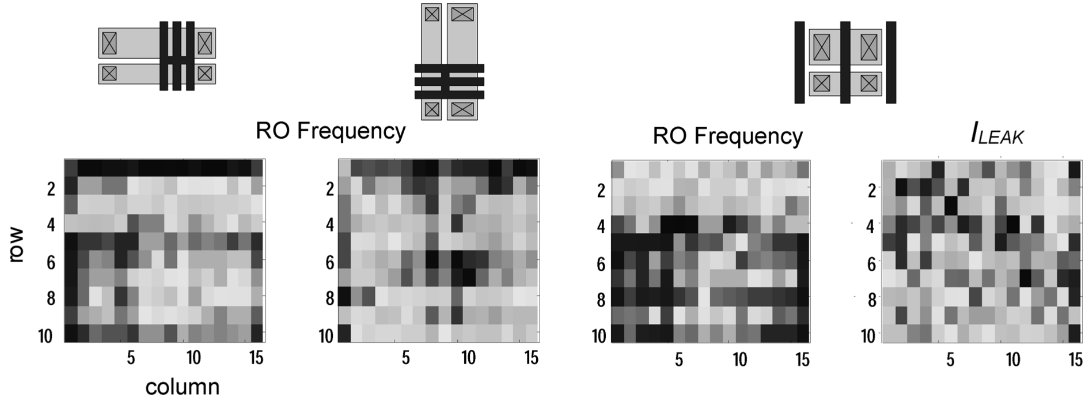


Fig. 17. Surface plot of normalized data of 36 chips, RO of layout with vertically oriented gates show a 1% frequency shift between rows 4 and 5.

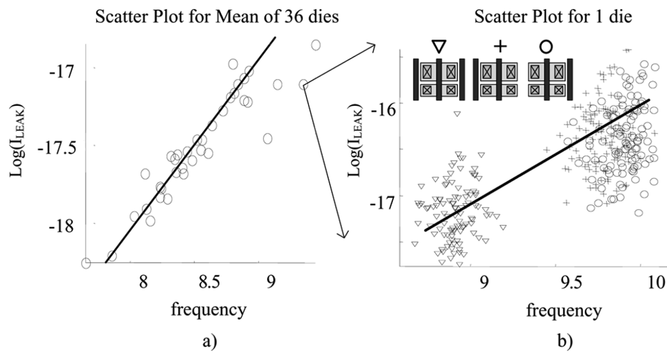


Fig. 18. Scatter plot of $\mathrm{Log}(I_{\mathrm{LEAK}})$ versus frequency showing (a) mean $\mathrm{Log}(I_{\mathrm{LEAK}})$ and mean RO frequency for 36 dies and (b) all $\mathrm{Log}(I_{\mathrm{LEAK}})$ and RO frequency for three layouts on one die.

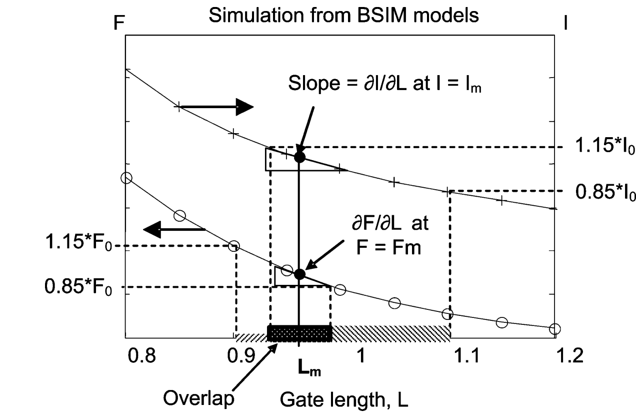


Fig. 19. Illustrates the estimation of the nominal value of $L(L_m)$ in the case where only one parameter is inferred. I_0 and F_0 are the measurement data. L_m is used to obtain the gradients and to find I_m and F_m, which are used to normalize the data in the least squares estimation.

TABLE III
SENSITIVITIES USED IN PROCESS PARAMETERS EXTRACTION FOR 1 CHIP AND 1 LAYOUT STRUCTURE

$$\text{If } FNch = \frac{\partial\left(F/F_0\right)}{\partial\left(Nch/Nch_{typ}\right)}, \quad FTox = \frac{\partial\left(F/F_0\right)}{\partial\left(Tox/Tox_{typ}\right)}, \quad FL = \frac{\partial\left(F/F_0\right)}{\partial\left(L/L_{typ}\right)},$$

$$INch = \frac{\partial\left(I/I_0\right)}{\partial\left(Nch/Nch_{typ}\right)}, \quad ITox = \frac{\partial\left(I/I_0\right)}{\partial\left(Tox/Tox_{typ}\right)}, \quad FL = \frac{\partial\left(I/I_0\right)}{\partial\left(L/L_{typ}\right)}$$

	Default	Vdd = 1.4V	Vbs = − 0.2V	T = 60 °C
FNch	-0.052	-0.053	-0.039	-0.049
FTox	-0.35	-0.35	-0.14	-0.39
FL	-1.79	-1.83	-1.65	-1.86
INch	0.061	0.066	0.090	0.050
ITox	-0.045	-0.056	-0.098	-0.028
IL	1.01	1.07	1.23	0.92

$\pm 15\%$ change in the mean measured data, F_0 and I_0, is found. The mean of this overlap region gives L_m, which is used to obtain the gradients and the normalization constants (F_m, I_m). The parameters are then normalized, and process parameters are inferred using least squares estimation.

This method is used for three process parameters, L, T_{ox} and N_{ch}. Spice simulations provide plots of F and I versus L, T_{ox} and N_{ch} over a small range around their typical values for the different conditions found in Table II. The measured F and I data is mapped to an overlapping range of values of the process parameters. The mean of this range $(L_m, T_{\mathrm{ox}m}, N_{\mathrm{ch}m})$ is used as the nominal value of X to obtain the gradients in M (Table III) and the normalization constants (F_m, I_m). We then normalize the components in X and Y. Finally, we apply the least squares estimation to estimate X. This is repeated for each die and each layout configuration since the small amount of variation assures a good linear approximation.

Fig. 20 shows the distribution of normalized L, which is obtained from the inference process. The sensitivities of F and I to L are larger than to other process parameters as shown in Table III. This results in significantly more accuracy in the inferred L. The plots of a normalized L show a strong correlation to the plots in Figs. 12 and 13. Faster and leakier ROs have a shorter L. The mean values of the L distributions depend significantly on the layout, showing that systematic layout-to-layout variation is attributed primarily to L shifts. The D2D spreads of L are significant since the bar-plots of the fastest and slowest chip are spaced far apart.

Fig. 21 shows a typical surface plot of a die, showing the WID variation of measured F and I and the corresponding inferred normalized L. We see that the systematic WID variation observed in F is strongly reflected in L.

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on May 4, 2009 at 13:53 from IEEE Xplore. Restrictions apply.

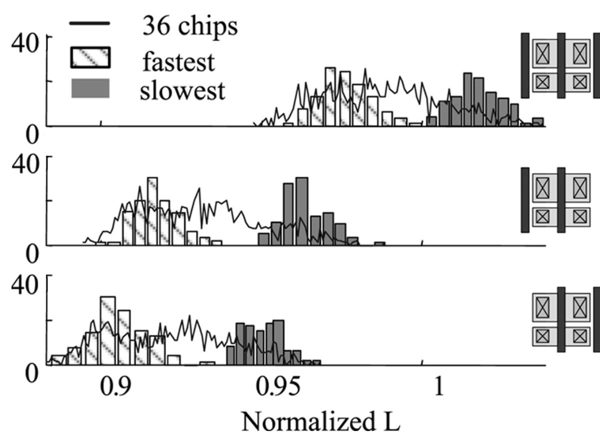


Fig. 20. De-embedded distribution of L.

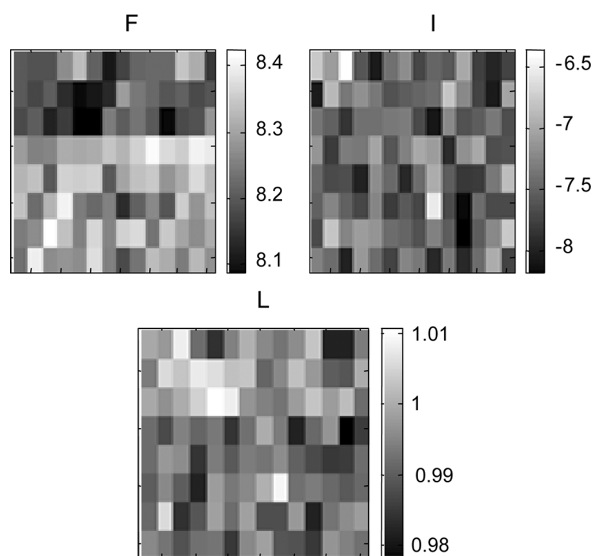


Fig. 21. Top plots: surface plot of within-die RO frequency and $\mathrm{Log}(I_{\mathrm{LEAK}})$ of a typical die. Bottom plot: corresponding surface plots of estimated normalized L.

V. Conclusion

Layout has a significant impact on lithography-induced variability in 90 nm technology. Analysis of the inferred process parameters indicates that poly-Si gate pitch has the strongest effect on the effective transistor gate length, resulting in up to a 10% variation in RO frequency for inverters laid out with isolated poly-Si lines. This systematic effect can be compensated by using layout extraction tools that account for proximity effects [16] or better OPC [27]. A simpler method would be to use regular layouts [28] or more restrictive layout rules that only allow transistors with a few possible gate pitches, together with an extractor that maps each gate pitch to its respective gate length.

The use of step-and-scan lithography induces a stronger correlation between gates placed parallel to the direction of the slit of light than those placed parallel to the direction of the scan. This effect can be exploited in the layout of regular datapaths and memory [1], [29]. By placing gates of the same path in the low correlation direction and by placing parallel paths in the high correlation direction, we can obtain a tighter performance spread.

Analysis of the spatial correlation of ROs and inverters can be used to model the spatial correlation coefficient of circuit paths and interconnected gates, allowing for more accurate statistical timing analysis.

For the same layout, systematic D2D variations dominate the total variations for small chips, indicating that the use of die-level adjustable supply voltages and substrate biasing can improve parametric yield significantly [2].

Acknowledgment

The authors wish to acknowledge the contributions of the students, faculty, and sponsors of the Berkeley Wireless Research Center, the National Science Foundation, the wafer fabrication donation of STMicroelectronics, and the support of the Center for Circuit and System Solutions (C2S2) Focus Center. The authors would also like to thank E. Perea and the engineers at STMicroelectronics, Crolles, France, for their support, Prof. A. Neureuther and Prof. C. Spanos and their students for helpful discussions, and C. Siow for his help with the measurements.

References

[1] K. A. Bowman, S. G. Duvall, and J. D. Meindl, "Impact of die-to-die and within-die parameter fluctuations on the maximum clock frequency distribution for gigascale integration," *IEEE J. Solid-State Circuits*, vol. 37, no. 2, pp. 183–190, Feb. 2002.

[2] J. W. Tschanz *et al.*, "Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage," *IEEE J. Solid-State Circuits*, vol. 37, no. 11, pp. 1396–1402, Nov. 2002.

[3] K. Bernstein *et al.*, "High-performance CMOS variability in the 65-nm regime and beyond," *IBM J. Res. Develop.*, vol. 50, pp. 433–449, Jul./Sep. 2006.

[4] P. Oldiges, Q. H. Lin, K. Petrillo, M. Sanchez, M. Ieong, and M. Hargrove, "Modeling line edge roughness effects in sub 100 nanometer gate length devices," in *Proc. Int. Conf. Simulation of Semiconductor Processes and Devices*, Seattle, WA, Sep. 2000, pp. 131–134.

[5] A. Asenov, S. Kaya, and J. H. Davies, "Intrinsic threshold voltage fluctuations in decanano MOSFETs due to local oxide thickness variations," *IEEE Trans. Electron Devices*, vol. 49, pp. 112–119, 2002.

[6] D. J. Frank, Y. Taur, M. Ieong, and H.-S. P. Wong, "Monte Carlo modeling of threshold variation due to dopant fluctuations," in *1999 Symp. VLSI Circuits Dig. Tech. Papers*, Kyoto, Japan, Jun. 1999, pp. 171–172.

[7] H. Masuda, S. Ohkawa, A. Kurokawa, and M. Aoki, "Challenge: Variability characterization and modeling for 65- to 90-nm processes," in *Proc. IEEE Custom Integrated Circuits Conf.*, Sep. 2005, pp. 593–599.

[8] J. D. Plummer, M. D. Deal, and P. B. Griffin, *Silicon VLSI Technology*. Upper Saddle River, NJ: Prentice-Hall, 2000.

[9] A. K.-K. Wong, *Resolution Enhancement Techniques in Optical Lithography*. Bellingham, WA: SPIE Press, 2001.

[10] M. Orshansky, L. Milor, P. Chen, K. Keutzer, and C. Hu, "Impact of spatial intrachip gate length variability on the performance of high-speed digital circuits," *IEEE Trans. Comput.-Aided Des.*, vol. 21, pp. 544–553, May 2002.

[11] M. Orshansky, L. Milor, and C. Hu, "Characterization of spatial intrafield gate CD variability, its impact on circuit performance, and spatial mask-level correction," *IEEE Trans. Semiconduct. Manufact.*, vol. 17, no. 1, pp. 2–11, Feb. 2004.

[12] T. Brunner, D. Corliss, S. Butt, T. Wiltshire, and C. P. Ausschnitt, "Laser bandwidth and other sources of focus blur in lithography," in *Proc. SPIE—Optical Microlithography XIX*, D. G. Flagello, Ed., Mar. 2006, vol. 6154.

[13] D. Steele, A. Coniglio, C. Tang, and B. Singh, "Characterizing post exposure bake processing for transient and steady state conditions, in the context of critical dimension control," in *Proc. SPIE-Metrology, Inspection and Process Control for Microlithography XVI*, D. J. Herr, Ed., Jul. 2002, vol. 4689, pp. 517–530.

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on May 4, 2009 at 13:53 from IEEE Xplore. Restrictions apply.

[14] I. Matthew, C. E. Tabery, T. Lukanc, M. Plat, M. Takahashi, and A. Wilkison, "Design restrictions for patterning with off-axis illumination," in *Proc. SPIE—Optical Microlithography XVIII*, B. W. Smith, Ed., May 2004, vol. 5754, pp. 1574–1585.

[15] T. A. Brunner, "Impact of lens aberrations on optical lithography," *IBM J. Res. Develop.*, vol. 41, pp. 57–67, Jan./Mar. 1997.

[16] M. Choi and L. Milor, "Impact on circuit performance of deterministic within-die variation in nanoscale semiconductor manufacturing," *IEEE Trans. Comput.-Aided Des.*, vol. 25, no. 7, pp. 1350–1367, Jul. 2006.

[17] T. Kjellberg and R. Schatz, "Effect of stitching errors on the performance of DFB lasers fabricated using e-beam lithography," *J. Lightw. Technol.*, vol. 10, no. 9, pp. 1256–1266, Sep. 1992.

[18] I. Ahsan *et al.*, "RTA-driven intra-die variations in stage delay, and parametric sensitivities for 65 nm technology," in *2006 Symp. VLSI Technology Dig. Tech. Papers*, Honolulu, HI, 2006, pp. 170–171.

[19] H. Tuinhout, M. Pelgrom, R. Penning de Vries, and M. Vertregt, "Effects of metal coverage on MOSFET matching," in *Int. Electron Devices Meeting (IEDM) Tech. Dig.*, San Francisco, CA, 1996, pp. 735–738.

[20] L.-T. Pang and B. Nikolic, "Impact of layout on 90 nm CMOS process parameter fluctuations," in *2006 Symp. VLSI Circuits Dig. Tech. Papers*, Honolulu, HI, 2006, pp. 69–70.

[21] B. Nikolic and L.-T. Pang, "Measurements and analysis of process variability in 90 nm CMOS," in *Proc. 8th Int. Conf. Solid-State and Integrated Circuit Technology*, Shanghai, China, 2006, pp. 505–508.

[22] K. Gonzalez-Valentin, "Extraction of variation sources due to layout practices," M.S. thesis, Massachusetts Inst. Technol., Cambridge, MA, 2002.

[23] J. S. Panganiban, "A Ring Oscillator Based Variation Test Chip," M.S. thesis, Massachusetts Inst. Technol., Cambridge, MA, 2002.

[24] D. Boning, J. Panganiban, K. Gonzalez-Valentin, C. McDowell, A. Gattiker, and F. Liu, "Test structures for delay variability," presented at the 2002 ACM/IEEE Tau Workshop, Monterey, CA, 2002.

[25] J. P. Cain and C. J. Spanos, "Electrical linewidth metrology for systematic CD variation characterization and causal analysis," in *Proc. SPIE-Metrology, Inspection, and Process Control for Microlithography XVII*, D. J. Herr, Ed., Jun. 2003, vol. 5038, pp. 350–361.

[26] P. Friedberg, Y. Cao, J. Cain, R. Wang, J. Rabaey, and C. Spanos, "Modeling within-field gate length spatial variation for process-design co-optimization," in *Proc. SPIE-Design and Process Integration for Microelectronic Manufacturing III*, L. W. Liebmann, Ed., May 2005, vol. 5756, pp. 178–188.

[27] D. Beale, J. Shiely, L. Melvin III, and M. Rieger, "Advanced model formulations for optical and process proximity correction," in *Proc. SPIE—Optical Microlithography XVII*, B. W. Smith, Ed., May 2004, vol. 5377, no. 12, pp. 721–729.

[28] V. Kheterpal *et al.*, "Design methodology for IC manufacturability based on regular logic-bricks," in *Proc. Design Automation Conf. (DAC)*, Jun. 2005, pp. 353–358.

[29] H. Onodera, "Variability: Modeling and its impact on design," *IEICE Trans. Electronics*, vol. E89-C, pp. 342–348, Mar. 2006.

Liang-Teck Pang (S'02–M'08) received the Dipl.-Ing. from Ecole Centrale de Paris, France, and the Master of Philosophy (M.Phil.) from Cambridge University, Cambridge, U.K., in 1997. Between 1998 and 2002, he worked in the DSO National Labs in Singapore on VLSI implementation of signal processing algorithms and high performance micro-architecture and circuit design. In 2002, he started the Ph.D. program in the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley. In the summers of 2005–2007, he interned in IBM Austin Research Labs where he was an inventor on several U.S. patent applications. He was presented with the IBM Invention Achievement Award in 2008 in recognition of those contributions.

Upon completion of the Ph.D. program in 2008, he joined the IBM T. J. Watson Research Center, Yorktown Heights, NY, as a research staff member. His Ph.D. research involved designing circuits to measure and characterize CMOS performance variability due to fluctuations in the manufacturing process. His emphasis is on the effects of layout on CMOS performance, measurement of the spatial correlation of logic gates, and analysis of variability data.

Borivoje Nikolić (S'93–M'99–SM'05) received the Dipl.Ing. and M.Sc. degrees in electrical engineering from the University of Belgrade, Serbia, in 1992 and 1994, respectively, and the Ph.D. degree from the University of California at Davis in 1999.

He lectured electronics courses at the University of Belgrade from 1992 to 1996. He spent two years with Silicon Systems, Inc., Texas Instruments Storage Products Group, San Jose, CA, working on disk-drive signal processing electronics. In 1999, he joined the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, where he is now a Professor. His research activities include digital and analog integrated circuit design and VLSI implementation of communications and signal processing algorithms. He is a coauthor of the book *Digital Integrated Circuits: A Design Perspective* (2nd ed., Prentice-Hall, 2003).

Dr. Nikolić received the NSF CAREER Award in 2003, the College of Engineering Best Doctoral Dissertation Prize and Anil K. Jain Prize for the Best Doctoral Dissertation in Electrical and Computer Engineering at the University of California at Davis in 1999, as well as the City of Belgrade Award for the Best Diploma Thesis in 1992. For work with his students and colleagues he received the Best Paper Award at the ACM/IEEE International Symposium of Low-Power Electronics in 2005, and the 2004 Jack Kilby Award for the Outstanding Student Paper at the IEEE International Solid-State Circuits Conference.

Authorized licensed use limited to: Univ of Calif Berkeley. Downloaded on May 4, 2009 at 13:53 from IEEE Xplore. Restrictions apply.