Modeling Within-Field Gate Length Spatial Variation for Process-Design Co-Optimization

Paul Friedberg, Yu Cao, Jason Cain, Ruth Wang, Jan Rabaey, and Costas Spanos
Department of Electrical Engineering and Computer Sciences,
University of California, Berkeley, CA 94720

ABSTRACT

Pelgrom's model [1] suggests that a spatial correlation structure is inherent in the physical properties of semiconductor devices; specifically, devices situated closely together will be subject to a higher degree of correlation than devices separated by larger distances. Since correlation of device gate length values caused by systematic variations in microlithographic processing is known to carry a significant impact on the variability of circuit performance, we attempt to extract and understand the nature of spatial correlation across an entire die. Based on exhaustive, full-wafer critical dimension measurements collected using electrical linewidth metrology for wafers processed in a standard 130nm lithography cell [8], we calculate a spatial correlation metric of gate length over a full-field range in both horizontal and vertical orientations. Using a rudimentary model fit to these results, we investigate the impact of correlation in the spatial distribution on the variability of circuit performance using a series of Monte Carlo analyses in HSPICE; it is confirmed that this correlation does indeed present a significant influence on performance variability.

From the same dataset, we also extract both the across-wafer (*AW*) and within-field (*WIF*) contributions to systematic variation. We find that the spatial correlation model's shape is strongly related to these two components of variation (both in magnitude as well as by spatial fingerprint). By artificially reducing each of these components of systematic variation—thereby simulating the effects of active, across-field process compensation—we show that spatial correlation is significantly reduced, nearly to zero. This implies that Pelgrom's model may not apply to die-scale separation distances, and that a more accurate correlation theory would combine Pelgrom's model over very short separation distances with a systematic variation model that captures variability over longer distances by means of non-stationary distributions.

Keywords: spatial correlation, process control, gate length variation, circuit performance, variability

1. INTRODUCTION

The aggressive scaling of silicon technology has enabled dramatic improvements in integrated circuit performance. However, as nominal device parameter values are rapidly reduced, control of semiconductor manufacturing processes has become increasingly difficult and expensive. In fact, the 2003 edition of the International Technology Roadmap for Semiconductors [3] lists the control of printed transistor gate length in the lithography process as falling short of expectations for the current technology generation; no known manufacturable solution is capable of achieving the target variation value. As variability in manufacturing processes has grown more severe, variations in device parameter values have grown in proportion to nominal values. In turn, wider distributions for device parameters have lead to increased variability in circuit performance, causing worsened yield degradation in successive technology generations. Therefore, the robustness of circuits has emerged as a roadblock (in addition to the problem of increasing power consumption) in modern IC design [2, 4].

In order to combat the growing negative impact of manufacturing variations on circuit performance, two approaches are being taken. The first approach is to apply a renewed focus on process control from a manufacturing perspective, in an effort to directly reduce the variations in device parameters; in the lithography process, for example, the development of integrated, scatterometry-based metrology [5] promises to improve gate CD control. The second approach comes from the design perspective, where practices can be developed to decrease circuit sensitivity to process variation. The most deleterious sources of device parameter variation may be identified through simulation; then, design styles that exhibit special robustness to those sources of variation can be selected.

In both approaches, there is great value in having a complete understanding of the underlying mechanisms that dictate how process variation affects circuit performance. In particular, an extremely useful tool in the arena of robust design is Monte Carlo simulation. Traditionally, Monte Carlo simulations have been performed by applying a statistical

variance to each device parameter and drawing random values for each simulation run in either (a) an entirely random fashion, with different values being chosen for each device in the given circuit, or (b) with total correlation—that is, each device is assigned the same randomly drawn device parameter value. These two limiting simplifications of the situation—namely setting the correlation factor ρ to either 0 (as in case "a") or 1 (as in case "b")—may lead to an undesirable exaggeration or underestimation of the resulting circuit performance variability, because Pelgrom's model suggests that spatial correlation has a structure that depends upon the separation of devices [1]. This paper addresses the problem of how spatial correlation should be modeled in Monte Carlo simulation by measuring spatial correlation in gate length over the span of the entire lithographic field and searching for the sources of that correlation. As we show in this work, while across-die correlation in the spatial distribution is a significant factor in circuit performance variability, it is a manifestation of systematic variation that can be modeled by means of appropriately scaled non-stationary distributions, and ultimately removed through process compensation measures across the field.

2. EMPIRICAL MODELING OF SPATIAL CORRELATION

In order to investigate the impact of spatial correlation on circuit performance, one must first investigate the nature of spatial correlation itself. To this end, this work focuses on the gate length for close spatial correlation analysis, since this parameter is one of the major culprits of process variation [6] and has shown some systematic patterns of variations due to lithography [7]. In three steps, gate length measurements were collected, spatial correlation characteristics were extracted, and a simple model was constructed to capture the general behavior of correlation for the purposes of circuit simulation.

2.1. Gate length measurement

Exhaustive critical dimension (CD) measurements were performed using electrical linewidth metrology (ELM) on a full 200mm wafer processed through an industrial 130nm manufacturing process [8]. The linewidth of a poly-silicon "gate" is measured via ELM by passing a precisely calibrated current through the gate and measuring the voltage across a subsection of the gate, as shown in **Fig. 1**.

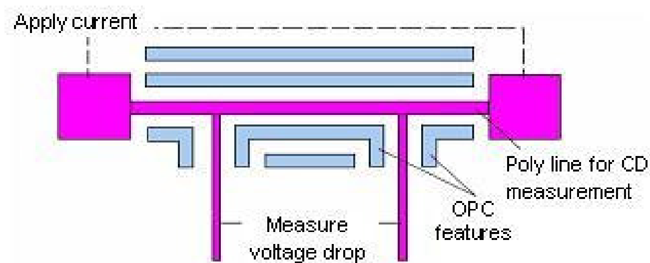

Figure 1. Kelvin test structure for ELM measurement. The measured voltage drop, combined with sheet resistance measurements from neighboring Van der Pauw structures can be used to extract linewidth.

The electrical measurement yields a numerical value with a sizable (~40nm) negative offset from values measured using more standard metrologies such as CDSEM or scatterometry, but the ELM measurements have been shown to have comparable precision to other metrologies and exceptional speed. Due to this speed of measurement, huge quantities of data may be taken—for example, this analysis makes use of 280 measurements per die over 35 dice, a total of roughly 10,000 measurements (**Fig. 2**). From this full-wafer dataset, the average die CD fingerprint may be extracted and modeled fairly well using a second-degree polynomial with coefficients calculated by least squares regression (**Fig. 3**). It is clear from this analysis that there exists a strong systematic within-field variation component, due to a combination of mask errors and systematic variation in the exposure tool. When this within-field variation component is removed from the full-wafer CD measurements, there is an obvious across-wafer systematic variation component, which can also be modeled well with a second-degree polynomial with coefficients determined by least squares regression (**Fig. 4**).

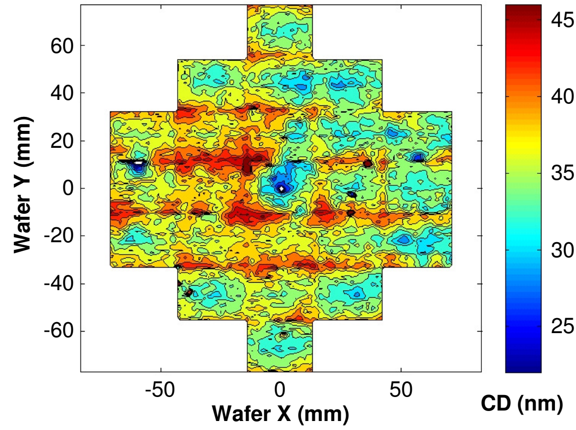

Figure 2. Full-wafer CD measurements.

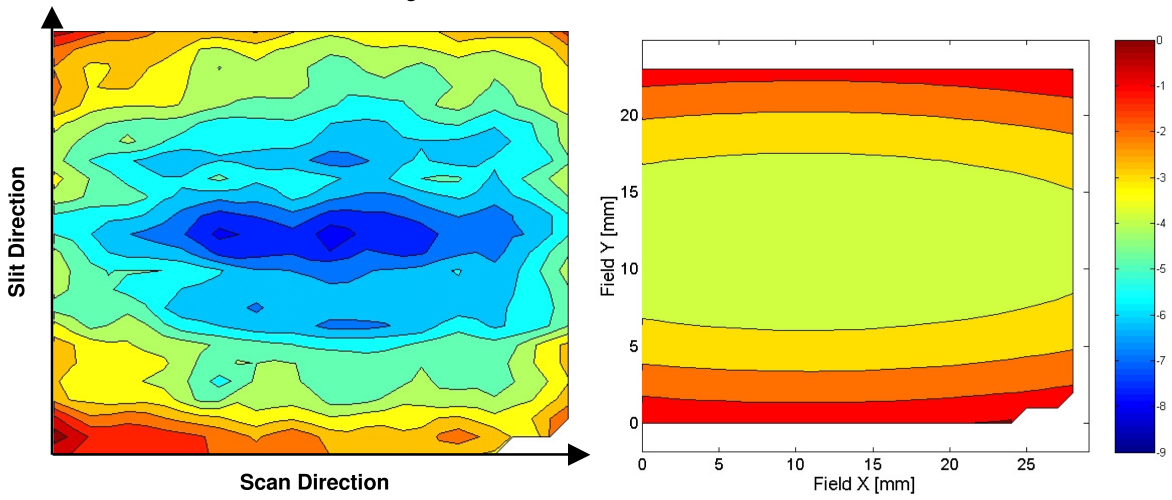

Figure 3. Average within-die CD fingerprint (left) and polynomial model of within-field variation (right). Cross-sections of this model dictate the nature of horizontal and vertical spatial correlation.

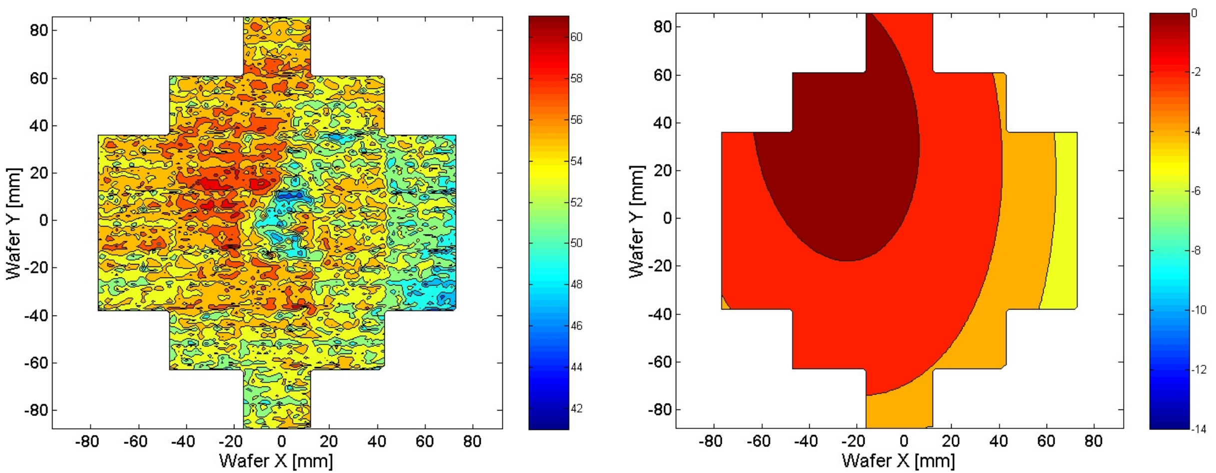

Figure 4. Full wafer CD map with average within-die CD fingerprint removed (left) and polynomial model of across-wafer variation (right).

2.2. Spatial correlation characteristics

With these CD data, we wish to extract a metric that will capture the nature of spatial correlation. Unfortunately, our data set does not satisfy one critical statistical assumption in correlation theory. Specifically, the distribution of CD is a non-stationary distribution, meaning that subsets of the data taken from different spatial locations within a field or from field to field do not necessarily have the same shape. The impact of violating this assumption is that a standard calculation of correlation will contain information about the systematic spatial distribution of measurements—or, in layman's terms, the *shape* of the distribution of measurements—in addition to the true correlation of the parameter values. However, as a first attempt to understand correlation in this spatial distribution of measurements, we will perform this standard statistical correlation extraction, and address the effects of the violation of that assumption later. Also, we will refer to our approximation of spatial correlation with the term spatial correlation for ease of use and to avoid further confusion—however, it is important to keep in mind that *our* spatial correlation does not precisely match the traditional statistical definition of correlation.

To extract the spatial correlation characteristics for this lithography process, the CD measurements are standardized against the entire wafer-wide CD distribution using the relation

$$z_i = \left(x_i - \hat{\mu}\right) / \hat{\sigma}\,, \tag{1}$$

where x_i is the ith CD measurement and $\hat{\mu}$ and $\hat{\sigma}$ are the average and standard deviation of all CD measurements. Next, the spatial correlation approximation can be calculated and plotted as a function of separation distance using

$$\rho_{jk} = \left(\sum z_j * z_k\right) / n\,. \tag{2}$$

That is, for each separation distance l afforded by the density of measurements, all pairs of points (j ,k) separated by l that fall within the same die are included in the summation as dictated by Eq. 2. We plot the resulting spatial correlation for this data set in **Fig. 5** (displaying correlation versus separation measured in both the horizontal and vertical directions). In both cases, devices separated by relatively short distances have the highest correlation ρ; as the separation increases, ρ falls roughly linearly, then finds a minimum value, and finally begins to increase as the distance approaches the entire width (or length) of the die. It is clear from inspecting the within-field systematic variation fingerprint (shown in **Fig. 3**) that the increase in correlation at near-field-width separation distances is an artifact of this component of variation. That is, it is easy to see that gate length values at opposite ends of the die are closely related due to the bowl-shaped pattern of the average within die variation. While it is expected that this within-field variation fingerprint is not atypical, we will ignore, for the time being, this upswing-portion of the correlation plot, because we can assume that critical path lengths will rarely exceed roughly half the chip width.

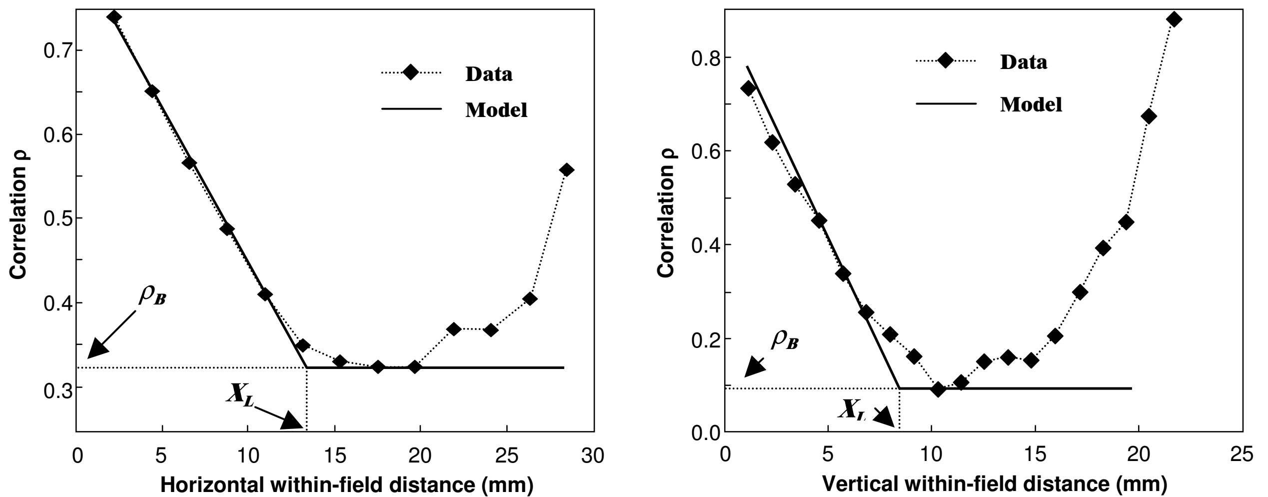

Figure 5. Spatial correlation dependences and models (horizontal on left, vertical on right).

2.3. Spatial correlation model

For future ease of Monte Carlo simulation, a rudimentary piecewise linear model (again, ignoring values of separation greater than half the width of the field) was imposed to more simply capture the spatial correlation relationship. Illustrated in **Fig. 5**, the model contains two parameters: X_L, a characteristic correlation length, and ρ_B, the characteristic correlation baseline. By definition, this model approximates the value of spatial correlation (ρ) given the separation distance (x or y) by:

$$\rho = \begin{cases} 1 - \dfrac{x}{X_L}(1-\rho_B), x \le X_L \\ \rho_B, x \ge X_L \end{cases} \tag{3}$$

These data are taken from a relatively large test chip covering the entire lithographic field (28 by 22 mm). However, the general shape of the spatial correlation curve will scale along with field size (since the polynomial model of systematic variation is a property of the exposure tool) and is therefore independent of field size. Thus, barring any specific tailoring of the within-field and across-wafer components of variation, and assuming one chip is printed per field, X_L will naturally fall at roughly half the chip length no matter what size field is printed.

3. IMPACT OF SPATIAL CORRELATION ON CIRCUIT PERFORMANCE

It is previously understood that delay variability is aggravated by high spatial correlation. A basic intuition of this effect can be attained through mathematical arguments. Consider the variance of the sum of n identically, normally distributed, correlated random samples:

$$\sigma^2_{tot} = n\sigma_{ind}^2 + 2\rho({}_nC_2)\sigma_{ind}^2 = [n + \rho(n)(n-1)]\sigma_{ind}^2 . \tag{4}$$

Using this relationship, we see that when $\rho = 1$, the condition of perfect correlation, $\sigma_{tot} = n\sigma_{ind}$, whereas when $\rho = 0$, the condition of perfect absence of correlation, $\sigma_{tot} = \sqrt{n}\sigma_{ind}$. If we approximate each stage delay in the critical path as a random sample (implicitly assuming that the delay of a given stage depends only on its associated gate length), we see that delay variability will decrease when either X_L or ρ_B is reduced (since this implies a reduction in ρ). This is a fairly accurate approximation—however, the actual impact of spatial correlation on circuit delay variability is slightly weaker (**Fig. 6**) because the assumption of perfect delay independence from one stage to the next is violated.

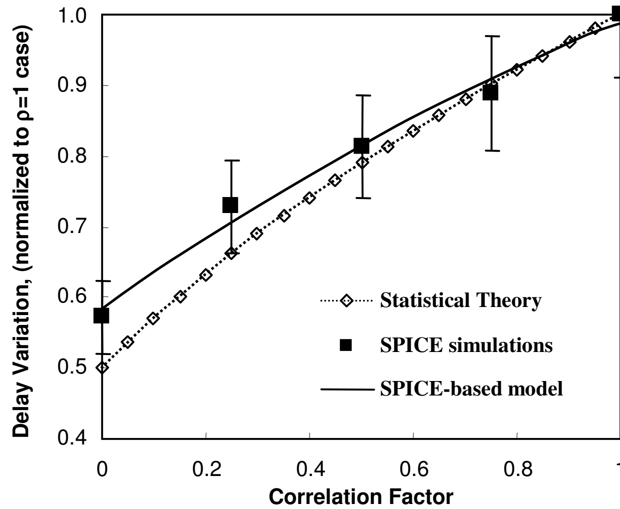

Figure 6. Comparison of statistical theory and SPICE simulation on 4-stage inverter chain.

That is, the discrepancy between mathematical theory and circuit simulation arises because the true delay of a given stage depends not only on its own gate length but also on the gate lengths of the preceding stage (due to input signal transition rate) and following stage (due to capacitive loading).

In the example of spatially distributed samples, the spatial correlation may not be fixed at one value, as assumed in the previous analysis for simplicity. For a more rigorous treatment, we need to perform calculations involving the covariance matrix, allowing for variable values of ρ for different separation distances. Using this more general approach, the variance matrix (a column matrix, where the ith entry corresponds to the variance associated with the ith sample) of the n samples is given by the relation:

$$variance\ matrix = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \sigma_1^{\,2} \\ \sigma_2^{\,2} \\ \sigma_3^{\,2} \\ \vdots \\ \sigma_n^{\,2} \end{bmatrix} \tag{5}$$

Here, ρ_1 is the correlation between samples separated by one unit, ρ_2 is the correlation between samples separated by two units, and so on. To calculate the total variance of the sum of the n samples, assuming all the individual variances are the same, we simply add the components of the resulting column vector to get:

$$\sigma_{tot}^{\,2} = \left(n + (n-1)(2)\rho_1 + (n-2)(2)\rho_2 + \ldots + 2\rho_{n-1}\right)\sigma_{ind}^{\,2} \tag{6}$$

From this analysis, we see that the correlation between samples situated close together is weighted more strongly in the total variance expression. This is intuitive since there are more pairs of samples separated by those closer distances.

To compare the improvements in delay variability achieved through the approach of reducing spatial correlation to those possible through the conventional approach of reducing gate length variation, simulations were performed for a canonical critical path subject to variable gate length under different combinations of X_L, ρ_B, and σ_L/μ_L. The canonical circuit (**Fig. 7**) consists of 10 equally-sized NAND stages, each loaded by a fan-out of two and separated from the following stage by 100μm of local interconnect. Nominal device parameter values and variances are based on an industrial 90nm technology node. The Berkeley Predictive Technology Model (BPTM) is employed to project the interconnect parameters and extract RC parasitics [9].

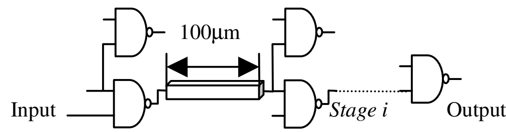

Figure 7. Canonical critical path circuit (10-stage NAND chain, Fan-Out=2).

Using this canonical circuit, 3000 Monte Carlo simulations were performed under nominal conditions of X_L = 2mm, ρ_B = 0.2, and σ_L/μ_L = 10%. (The value of X_L was intentionally reduced when compared to the values extracted from the 130nm-generation test structures to reflect the smaller chip size.) The results, shown in **Fig. 8**, indicate that delay variability is most sensitive to the tuning of σ_L/μ_L; however, reduction of both X_L and ρ_B does display some benefit in terms of reducing delay variability. While the improvement due to direct reduction of parameter variation is linear, the improvement due to reduction of spatial correlation parameter X_L increases gradually as the scaling factor is reduced, achieving over half of the total potential improvement with the final 25% reduction of X_L. Also notable is the fact that the greatest gains in reducing ρ_B are seen when X_L is also strongly reduced.

These data show that indeed spatial correlation provides us with a significant lever for reducing circuit delay variation. If process control schemes could allow us to tune X_L and ρ_B arbitrarily, we would expect to see up to a *4x* improvement in delay variability. Although this method may be less efficient than reducing process variation directly,

that strategy has long been a well-known manufacturing goal. Thus, the potential for gains in this area have already planned for on industry roadmaps, whereas spatial correlation concerns are as of yet an untapped resource for robust design. The wisest approach for leveraging our understanding of the effects of spatial correlation in critical path optimization will be to use process control schemes in manufacturing to tune X_L and ρ_B lower as well as to incorporate circuit design rules to increasing the separation between successive stages to minimize the degree of correlation "seen" by the circuit.

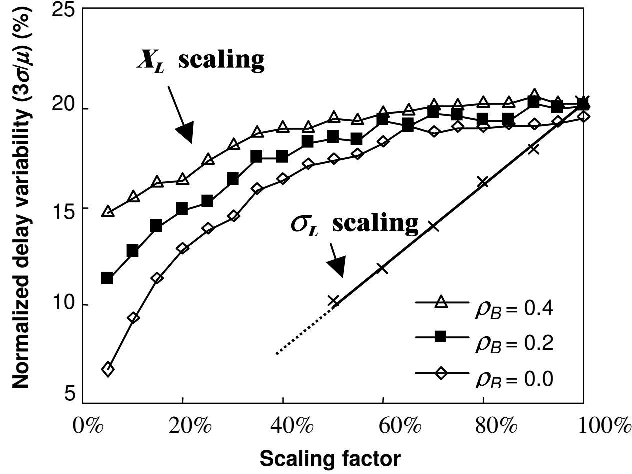

Figure 8. Comparison of impacts of spatial correlation and normalized variation in gate length on critical path delay variability.

4. POTENTIAL IMPACT OF PROCESS CONTROL SCHEMES ON SPATIAL CORRELATION

Motivated by an understanding of the significant effects of spatial correlation on circuit performance variation, we would now like to determine how various forms of process control in the lithography cell might impact the spatial correlation dependence. This will serve to show us what we can expect to gain in terms of correlation by removing the various sources of controllable, systematic variation, and will also serve to eventually render the data distribution stationary, which will allow us to finally gauge the natural level of correlation in the data without violating any statistical assumptions. To accomplish this goal, we decompose the CD measurements into five components: the average CD (μ), mask errors (*mask*), within-field systematic variation (*WIF*), across-wafer systematic variation (*AW*), and random variation (ε):

$$CD_i = \mu + mask_i + WIF_i + AW_i + \varepsilon_i \tag{7}$$

where the mask errors are measured directly from the reticle used to generate the patterns and the *WIF* and *AW* components are extracted from the data by least-square fitting to a 2nd-order polynomial

$$fit = \beta_1 x^2 + \beta_2 y^2 + \beta_3 x + \beta_4 y \tag{8}$$

Then, using this decomposition, we can artificially remove each component of variation from the group of measurements to simulate a specific example of process control and then recalculate the spatial correlation dependence curves and compare to the original dependence to evaluate the impact of that process control effort. For example, by removing either the *WIF* component or *AW* component of variation, we see a significant change in the spatial correlation curves (**Figs 9 and 10**, respectively).

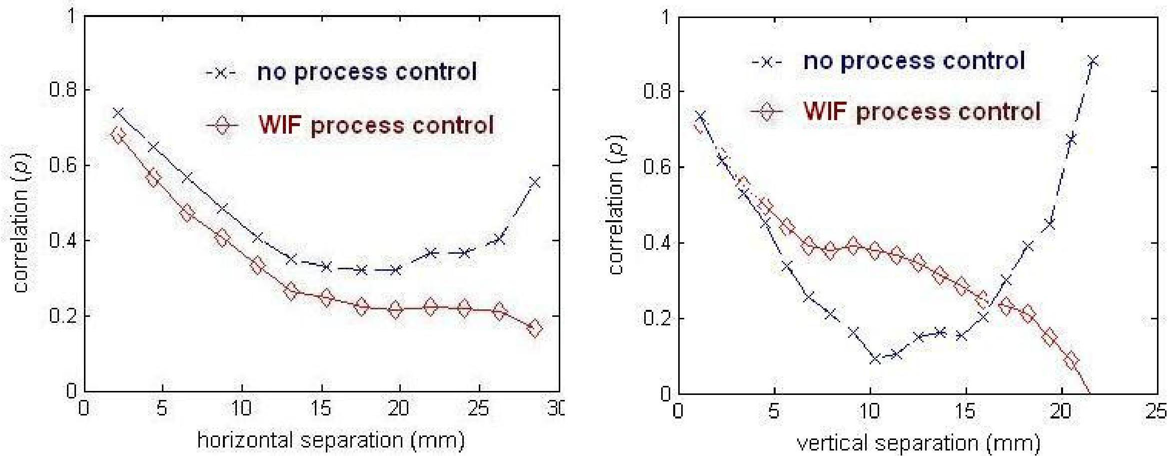

Figure 9. Comparison of spatial correlation characteristics under within-field process control to original characteristics.

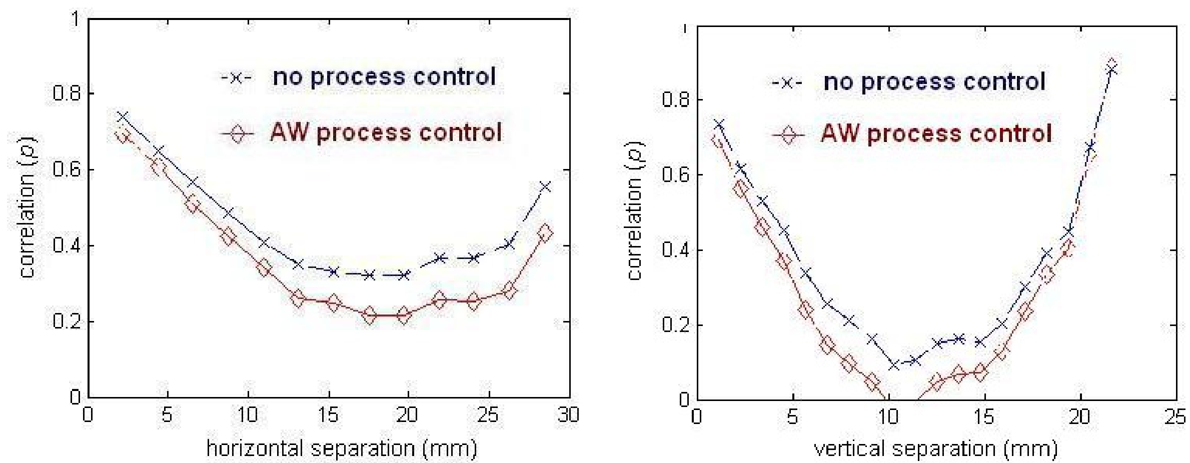

Figure 10. Comparison of spatial correlation characteristics under across-wafer process control to original characteristics.

From these two sets of plots, we can see that the two types of process control (*WIF* and *AW*) have distinctly different impacts on the spatial correlation characteristics. *WIF* control changes the shape of the curves, particularly for large separation distances. This effect is due to the reduction in the curved shape of the *WIF* systematic variation effected by the process control. By removing much of this curvature, the correlation between points on opposite sides of the field is strongly reduced. In the case of the vertical spatial correlation curve, it is interesting to note that the spatial correlation actually increases in the middle range of the field. This may be due to the fact that a much stronger systematic variation exists across the field in the vertical direction; when this variation is reconciled, the remaining *AW* variation serves to "group" the measurements within the die in the vertical direction such that the correlation among points within a single field is actually increased. While the relative contribution of spatial correlation to circuit performance variability will increase in this instance, the overall level of performance variability will likely decrease since such a large component of variation has been reconciled. The fact that we see such a strong change in shape of the spatial correlation curves as we impose *WIF* control serves as a clear indicator that indeed our violation of the assumption of stationary distribution has a profound effect on our correlation calculations. Thus, we must be careful not to conclude that *WIF* process control is deleterious in terms of spatial correlation based on this example alone. With simulation of *AW* process control, the spatial correlation is reduced across the entire range of the lithographic field.

Finally, by applying both *WIF* and *AW* process control schemes artificially, we create a third set of hypothetical spatial correlation curves:

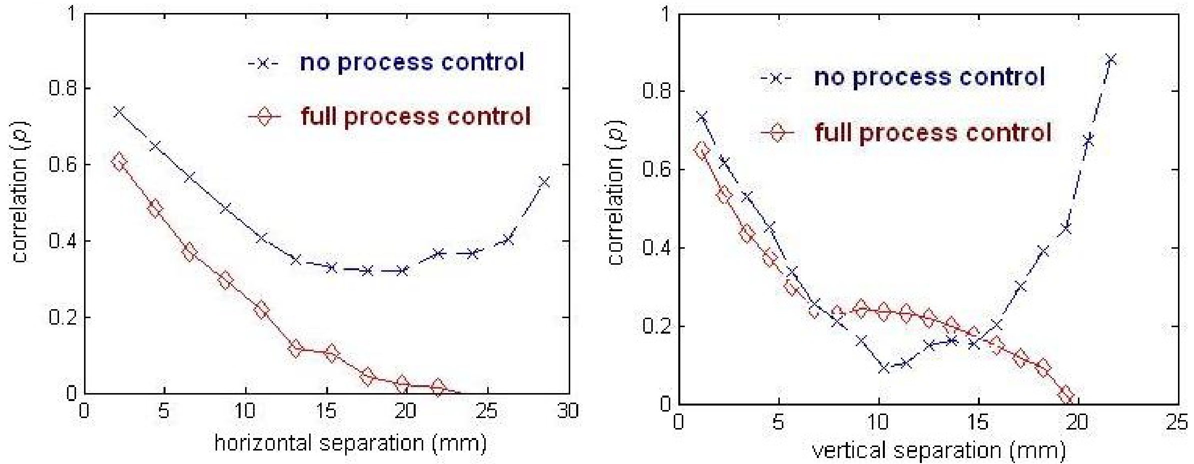

Figure 11. Comparison of spatial correlation characteristics under both *AW* and *WIF* process control to original characteristics.

The combination of the two process control schemes actually reduces the spatial correlation values by a greater amount than the sum of the two schemes taken individually. A significant reduction in spatial correlation is seen for the entire width of the chip in the horizontal direction; in the vertical direction, a slight reduction is seen initially, but the middle-range is still subject to higher spatial correlation as a result of the *WIF* control. Casting these results in terms of the spatial correlation model, we see that neither form of process control has a strong impact on the characteristic correlation length X_L, but that the characteristic correlation baseline parameter ρ_B can be significantly reduced. The fact that X_L is not significantly changes by either form of process control may suggest that this parameter is relatively robust to process control changes as long as even a slight amount of systematic variation persists.

To explore this possibility, once the *WIF* and *AW* components of variation modeled with second-degree polynomials have been removed, the full-wafer data set is examined again to search for remaining systematic variation that might be removed by other forms of process control. As shown in **Fig. 12**, the new full-wafer data set still contains some clear systematic variation. In particular, it appears that the average CD for each field may still vary from field to field. Under suspicion that this variation is caused by errors in dose control from shot to shot, we can apply a field-by-field correction term [10] that would simulate the reconciliation of this source of variation (**Fig. 13**). Finally, if we plot the spatial correlation characteristics one last time with this final form of process control, we see that the spatial correlation is very nearly zero across the entire field (**Fig. 14**).

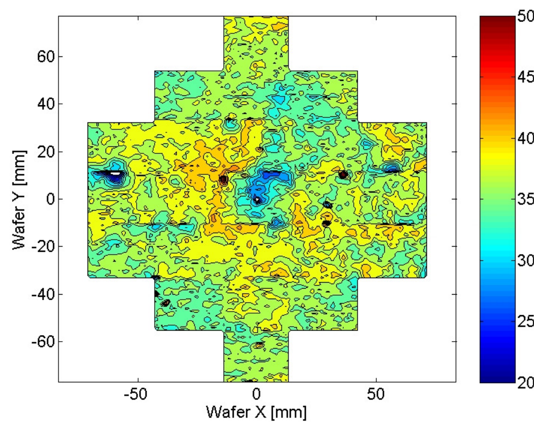

Figure 12. Full-wafer CD map with *WIF* and *AW* polynomial model systematic variation components removed.

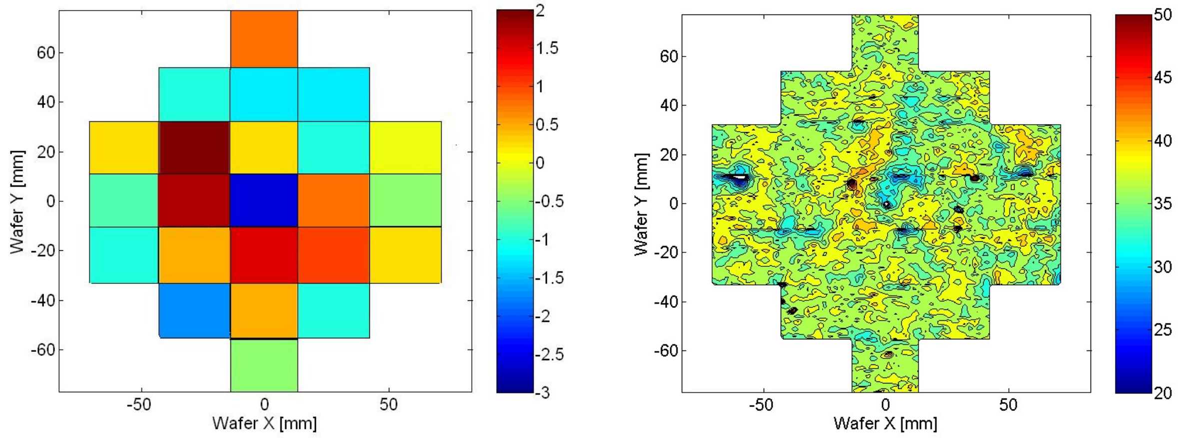

Figure 13. Die-to-die correction dose-error correction offsets (left) and resulting full-wafer CD map when these corrections are applied to the data set (right).

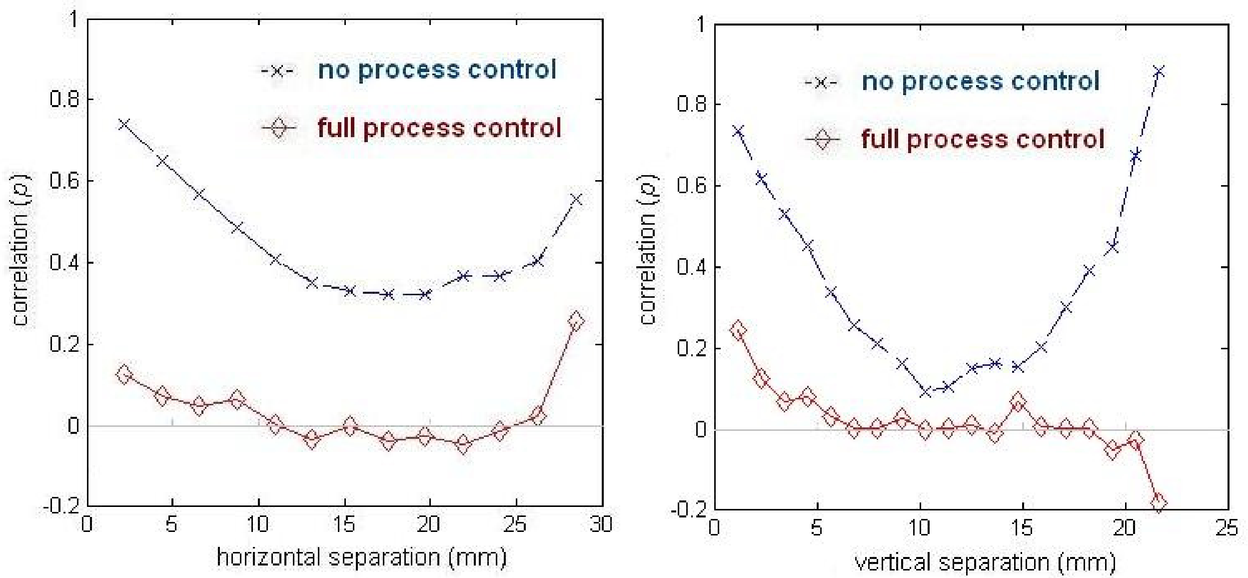

Figure 14. Spatial correlation dependences with full across-field, across-wafer, and field-by-field process control applied for horizontal (left) and vertical (right) separation.

By removing the last substantial portion of this systematic variation, the correlation has been strongly reduced. In addition, these results show that instead of affecting variation and correlation independently, process control will serve to reduce both effects simultaneously. In reference to **Fig. 8**, process control will move down both the variance and correlation curves simultaneously as the implementation of process control becomes more complete. Finally, it is quite significant to observe that the removal of the last component of systematic variation leads to the greatest reduction of remaining spatial correlation. This demonstrates the importance of process control, and furthermore demonstrates the advantage of *complete* process control over partial process control.

5. CONCLUSION

A simple analytical model for spatial correlation is derived from experimental CD measurement data. Based on this model, we find that in order to minimize critical path delay variability, the process should be tuned to minimize the correlation between stages. By imposing artificial process control on these data, we show that the spatial correlation

structure is due nearly entirely to the various sources of systematic variation introduced in processing. That is, by rendering the data set stationary through process control, so that our correlation measurement captures correlation in the honest, statistical sense—we find that the remaining correlation in the data is negligible. This suggests that most of the correlation between device parameter values over the separation distances examined here is due to systematic variation caused by processing that can potentially be removed by standard forms of process control, when implemented with extreme accuracy. In turn, this may suggest that Pelgrom's model applies only over very short separation distances that cannot be seen here given the granularity of our data set. That is, over short distances, true statistical correlation may exist and decrease with increasing separation as Pelgrom has previously found; over larger distances, however, correlation is only an artifact of systematic variation. This hypothesis will be the subject of continued research in this area. If we can find the limiting distance over which true correlation exists, we may propose circuit design rules to minimize that correlation by separating successive stages by more than that limiting distance. Finally, these results suggest that for the purposes of Monte Carlo simulation, correlation should be manufactured through the application of a long-range systematic variation shape, with the possible addition of a short-range Pelgrom's model-based correlation to be investigated later.

ACKNOWLEDGEMENTS

This work was funded by Advanced Micro Devices, Applied Materials, Atmel, Cadence, Canon, Cymer, DuPont, Ebara, Intel, KLA-TENCOR, Mentor Graphics, Nikon Research, Novellus Systems, Panoramic Technologies, Synopsis, Tokyo Electron, and the UC Discovery Grant.

REFERENCES

[1] M. Pelgrom, A. Duinmaijer, and A. Welbers, "Matching properties of MOS transistors," *IEEE Journal of Solid-State Circuits*, vol. 24, no. 1, pp. 1433-1439, Oct. 1989.
[2] B. E. Stine, D. S. Boning, and J. E. Chung, "Analysis and decomposition of spatial variation in integrated circuit processes and devices," *IEEE Transactions on Semiconductor Manufacturing*, vol. 10, no. 1, Feb. 19.
[3] International Technology Roadmap for Semiconductors, 2003.
[4] S. Nassif, "Design for manufacturability in DSM technologies," *IEEE International Symposium on Quality Electronics Design*, pp. 451-454, 2000.
[5] X. Niu, N. Jakatdar, J. Bao, and C. Spanos, "Specular Spectroscopic Scatterometry," *IEEE Transactions on Semiconductor Manufacturing*, vol. 14, no. 2, pp. 97-111, 2001.
[6] Y. Cao, et al., "Design sensitivities to variability: extrapolations and assessments in nanometer VLSI," *IEEE International ASIC/SoC Conference*, pp. 411-415, 2002.
[7] M. Orshansky, et al., "Impact of spatial intra-chip gate length variability on the performance of high speed digital circuits," *IEEE Transactions on Computer-Aided Design*, vol. 21, pp. 544-553, May 2002.
[8] J. Cain and C. Spanos, "Electrical linewidth metrology for systematic CD variation characterization and causal analysis," *Metrology, Inspection, and Process Control for Microlithography XVII*, Proceedings of SPIE vol. 5038, pp. 350-361, 2003.
[9] Y. Cao, et al., "New paradigm of predictive MOSFET and interconnect modeling for early circuit design," IEEE Custom Integrated Circuits Conference, pp. 201-204, 2000.
[10] J. van Schoot, et al, "CD uniformity improvement by active scanner corrections," *Optical Microlithography XV*, Proceedings of SPIE vol. 4691, pp. 304-312, 2002.