Interval-Valued Statistical Modeling of Oxide Chemical-Mechanical Polishing

James D. Ma, Claire F. Fang, Rob A. Rutenbar
ECE Department, Carnegie Mellon University
{jdma, ffang, rutenbar}@ece.cmu.edu

Xiaolin Xie, Duane S. Boning
EECS Department, Massachusetts Institute of Technology
{xiaolin, boning}@mtl.mit.edu

***Abstract*— Technology-oriented tools provide the raw data needed to optimize the fabrication process itself, and to predict problematic variational impacts on silicon design. Unfortunately, even in these physics-oriented tools, statistically uncertain quantities appear as crucial inputs. To date, Monte Carlo techniques have been the dominant solution method. We suggest an alternative in which uncertainties are represented as correlated intervals, and interval-valued computations replace the standard scalar operations in the numerical algorithm for the tool. We use an oxide chemical-mechanical polishing tool as an example, and show how to "retrofit" workable statistical models on top of the original algorithm. Accuracies to within ∼1–10% of Monte Carlo simulation, and speedups of ∼10–100X can be achieved, depending on whether we choose a formulation which emphasizes accuracy, or efficiency.**

I. Introduction

Nanoscale technologies demand accurate predictive tools for how the physics of fabrication affects devices, circuits, interconnect, and indeed all levels of the design hierarchy. These technology-oriented tools often provide the raw data needed to optimize the fabrication process itself, and to predict problematic variational impacts on silicon design. One example of the former is statistical lithography simulation [1], used to predict the dependence of critical dimension (CD) on lithography variables such as focus, dose, mask CD, and resist thickness. Examples of the latter include chemical-mechanical polishing (CMP) models, which predict how non-uniformities in metal pattern density affect local oxide thickness, and thus critical interconnect parasitics, after polishing [2], [3], [4]; and 3-D device simulation strategies to predict how random dopant fluctuations give rise to wide V_{th} variations in nanoscale FETs [5], [6]. It is important to note that, even in these physics-oriented tools, statistically uncertain quantities appear as crucial *inputs*. That is, the tools we would like to rely on for predicting variational and correlation effects on familiar circuit/interconnect parameters are themselves reliant on models of variations in physical behaviors, equipment imperfections, layout geometry, etc. To date, Monte Carlo techniques (combined with some application-specific cleverness) seem to be the dominant strategy for handling these issues. For example, Postnikov *et al* [1] use smart sampling techniques to extract CD impacts from a tractable set of lithographic "corners." Frank *et al* stochastically place dopants in full 3-D MOSFET simulations to predict V_{th} and drain-induced barrier lowering (DIBL) impacts [5].

In this paper, we suggest an alternative to standard Monte Carlo sampling which is applicable to numerically oriented technology tools. Our idea is to replace individual numerical scalar values (e.g., $x = 5$) with *interval* values (e.g., $x = [3, 7]$) and then to propagate these through each atomic numerical operation. Interval calculation *per se* is not new, what is new are recent advances in interval representation that can explicitly handle correlations between different numerical quantities [7], and statistical extensions that allow such intervals to approximate not just range uncertainty, but also the actual probability distribution for each computed result [8], [9]. Early versions of our ideas have been successfully demonstrated in the application of "interval-valued" statistical model order reduction for linear RLC interconnect [8], [10]. In this paper, we study how these ideas might be exploited in physics-oriented tools, choosing CMP modeling as our experimental vehicle because of its importance in silicon design. We explore interval-valued statistical CMP models, with correlated interval values for either CMP modeling parameters such as initial step height, planarization length, and blanket removal rate (to be defined later on), or metal layout density. The new problem we face is that many physical models rely on profoundly nonlinear computations, which are even today challenging to manage in an interval-valued framework. We show a range of interval-centric engineering tradeoffs to combat this problem.

The paper is organized as follows. Section II reviews the mechanics of interval models for approximating statistics. Section III reviews deterministic CMP models. Section IV describes a set of different interval-valued statistical CMP modeling algorithms, and Section V shows experimental results. Finally, Section VI offers concluding remarks.

II. Modeling with Affine Intervals

Replacing scalars with intervals in numerical codes is an old idea [11] with, unfortunately, an equally old problem. If we only represent each uncertainty as a single range, e.g., $x = [3, 7]$, we do not represent critical correlations and dependencies among multiple random variables. As a result, deep chains of computations in which many cancellations should properly occur yield instead a set of intervals whose bounds are often uselessly large in comparison to the true

0-7803-9254-X/05/$20.00 ©2005 IEEE.

range of the uncertain quantity. The problem is called "range explosion," and seriously limits use of these classical methods.

Stolfi and de Figueiredo suggested a first-order solution with the *affine interval* representation [7]. Each uncertain quantity is now represented as a *central value*, added to which are a sum of scaled *uncertainty* terms:

$$\widehat{x} = x_0 + \sum_{i=1}^{n} x_i \varepsilon_i \quad (-1 \leqslant \varepsilon_i \leqslant 1) \tag{1}$$

The affine interval is symmetric about its central value (x_0) and explicitly depends on a number of independent uncertainty symbols (ε_i's) which can be shared among multiple variables. Cancellations can occur in numerical calculations, and variable dependencies can be explicitly accommodated. The result is a much tighter set of interval bounds.

Fang *et al* [12] took the first step toward connecting affine intervals with statistics by observing that, in the presence of more than just a few uncertainty terms in Equation 1, the Central Limit Theorem [13] suggests that the resulting distribution is asymptotically normal, independent of the distributions of individual uncertainty symbols. Classical confidence interval ideas were applied and used to statistically "tighten" the interval bounds, reflecting the fact that the probability of occurrence near the extreme edges of the standard affine interval form is exceptionally small. Fang *et al* [12] also showed how to trade off perfect conservatism for vastly improved practical bounding.

We also took the next step toward a statistical view by explicitly interpreting the affine interval form as a *symbolic* form [8], [10]. That is, a numerical algorithm has input variables of the affine form $\widehat{x} = x_0 + \sum_{i=1}^{n} x_i \varepsilon_i$, and we now interpret the uncertainty symbols as independent zero-mean unit-variance normal random variables. If the numerical procedure produces outputs of the affine form $\widehat{z} = z_0 + \sum_{i=1}^{n} z_i \varepsilon_i$, we still assume – an approximation of course – that this is a symbolic form combining a set of independent normal random variables. Roughly speaking, the heuristic is to assume that the careful mechanics of bound estimation, computed for each affine interval operation, can serve as a workable standin for how the statistics would actually move through the same computations, i.e., how the "essential bulk" of the distribution moves through these computations. For perfectly affine operations – addition/subtraction and scaling by a scalar constant – this is actually correct, and is essentially the same first-order form as that used to model delay in recent statistical static timing engines (e.g., [14], [15]). For other operations, this is an approximation, sometimes good, sometimes less so. For applications dominated by addition/subtraction (e.g., >90% of all operations), such as linear interconnect reduction, this simple heuristic is surprisingly workable [8]. One can also make intelligent engineering tradeoffs of where to use the intervals, and where to revert to standard Monte Carlo methods [10].

Recently, the connections between intervals and statistics have been further strengthened with a more sophisticated affine arithmetic that can handle ranges that are asymmetric about the central value, approximate distributions with some degree of skewness, and better predict the ranges and correlations for difficult nonlinear operations such as division and transcendental functions. Not surprisingly, this accuracy comes at the expense of much more computation time. See [9] for details.

Space does not permit a rigorous treatment of any of these interval models. Hence, let us demonstrate the essential ideas and tradeoffs with a concrete example – a small system of linear equations, which involves such basic arithmetic operations as addition/subtraction, multiplication, reciprocal, and division. We first start with solving the following three scalar-valued equations, based on backward substitutions.

$$\begin{bmatrix} 3 & 0 & 4 \\ 0 & 2 & 3 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = - \begin{bmatrix} 1 \\ 10 \\ 16 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ -4 \end{bmatrix} \tag{2}$$

Second, we assume classical interval-valued coefficients for the equations, and solve them by replacing every scalar operation in the backward substitution procedure with its classical interval counterpart [11]:

$$\begin{bmatrix} [1,5] & 0 & [1,7] \\ 0 & [1,3] & [1,5] \\ 0 & 0 & [2,6] \end{bmatrix} \begin{bmatrix} \overline{x}_1 \\ \overline{x}_2 \\ \overline{x}_3 \end{bmatrix} = - \begin{bmatrix} 1 \\ 10 \\ 16 \end{bmatrix}$$

$$\begin{bmatrix} \overline{x}_1 \\ \overline{x}_2 \\ \overline{x}_3 \end{bmatrix} = \begin{bmatrix} [0.3, 55] \\ [-7.3, 30] \\ [-8, -2.7] \end{bmatrix} \tag{3}$$

where the classical interval-valued solutions are denoted as $\overline{x}_1$, $\overline{x}_2$, and $\overline{x}_3$. Finally, we assume *correlated* affine interval-valued coefficients (sharing uncertainty symbols ε_1, ε_2, and ε_3) for these equations:

$$\begin{bmatrix} 3+\varepsilon_1+\varepsilon_2 & 0 & 4-\varepsilon_1+2\varepsilon_3 \\ 0 & 2+\varepsilon_1 & 3+\varepsilon_1-\varepsilon_3 \\ 0 & 0 & 4+\varepsilon_1-\varepsilon_3 \end{bmatrix} \begin{bmatrix} \widehat{x}_1 \\ \widehat{x}_2 \\ \widehat{x}_3 \end{bmatrix} = - \begin{bmatrix} 1 \\ 10 \\ 16 \end{bmatrix} \tag{4}$$

and solve them by replacing every scalar operation in the backward substitution procedure with its simple affine interval counterpart [7], [10]:

$$\begin{bmatrix} \widehat{x}_1 \\ \widehat{x}_2 \\ \widehat{x}_3 \end{bmatrix} = \begin{bmatrix} 5 - 1.7\varepsilon_1 - 1.7\varepsilon_2 + 1.3\varepsilon_3 - 2\varepsilon_5 \\ 1 + 3\varepsilon_1 + 0.5\varepsilon_3 - 2\varepsilon_4 \\ -4 - \varepsilon_1 + \varepsilon_3 \end{bmatrix} \tag{5}$$

and sophisticated affine interval counterpart [9]:

$$\begin{bmatrix} \widehat{x}_1 \\ \widehat{x}_2 \\ \widehat{x}_3 \end{bmatrix} = \begin{bmatrix} 5 - 0.8\varepsilon_1 - 0.8\varepsilon_2 + 0.6\varepsilon_3 + 0.1\varepsilon_6 + 0.2\varepsilon_7 \\ 1 + 1.6\varepsilon_1 - 1.8\varepsilon_3 + 0.1\varepsilon_4 + 0.4\varepsilon_5 \\ -4 - 0.5\varepsilon_1 + 0.5\varepsilon_3 \end{bmatrix} \tag{6}$$

respectively. Note that ε_4 to ε_7 are new uncertainty symbols that are introduced to approximate higher-order uncertainties and assumed to be independent of existing uncertainty symbols ε_1 to ε_3 [7], [10], [9].

To compare the results, let us look at the joint distribution of $(\widehat{x}_1, \widehat{x}_2)$. To do this, we first randomly, normally sample 10,000 times over the uncertainty symbols ε_1 to ε_5 in Equation

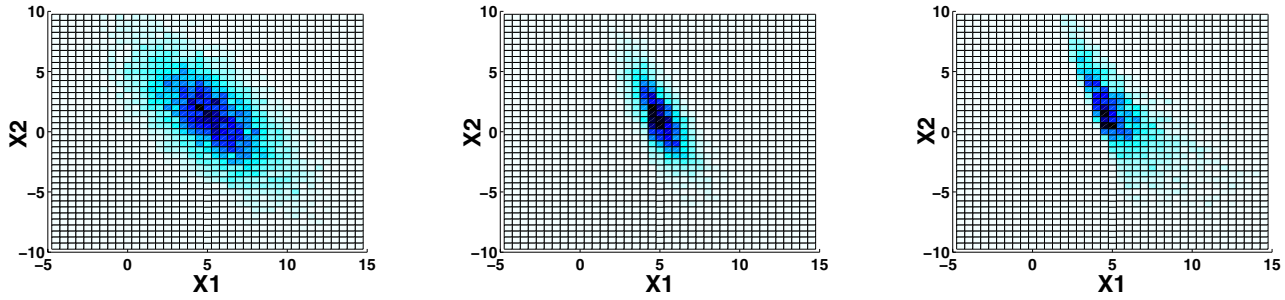


Fig. 1. Comparison of the joint distributions of (x_1, x_2) solution points among (a) simple affine interval solve, (b) sophisticated affine interval solve, and (c) Monte Carlo simulation.

5 for the simple affine interval solve. Then we follow the sampling techniques suggested in [9] over the uncertainty symbols ε_1 to ε_7 for the sophisticated affine interval solve. Figures 1(a) and 1(b) show the distributions of joint solution points on the x_1–x_2 plane using false color shading, for the simple and sophisticated affine interval solves, respectively. More (x_1, x_2) solution points are distributed in darker regions. We also normally sample over the original uncertainty symbols ε_1 to ε_3 in Equation 4, solve for each set of scalar-valued sample equations, and plot the *real* joint distribution of (x_1, x_2) solution points as shown in Figure 1(c). Both the simple and sophisticated interval models do a workable job of modeling the most likely joint values. The more complex intervals yield an *asymmetric* distribution that more accurately predicts the true skewed distribution, albeit at much more computational cost. In contrast, the classical interval result, which is simply the rectangle $[0.3, 55] \times [-7.3, 30]$, is too pessimistically large even to fit on the above plots. The optimal choice of which interval-based representation to use, and even where to deploy intervals to advantage, is an engineering decision. One can trade off efficiency for accuracy. In the sequel, we discuss how to make these tradeoffs in the CMP application.

III. Review of Deterministic CMP Modeling

Chemical-mechanical polishing (CMP) has become the dominant technique for inter-layer dielectric (ILD) planarization, which in turn has great impact on the overall circuit performance and yield [16]. Even though the fundamental physical principles behind CMP are complex and under active investigation, the basic mechanisms for removing oxide are well understood as resulting from the interplay of a chemical reaction which softens the oxide surface and mechanical abrasion by slurry particles deposited on a porous pad against which the wafers rotate [17].

Even after CMP, however, the oxide thickness across the dies (and wafers) may still be non-uniform. CMP is highly sensitive to the underlying metal layout. Intuitively speaking, the higher the layout pattern density, the larger the contact area with the pad, and the lower the localized pressure on the raised oxide. Therefore, the oxide removal rate varies *inversely* with the pattern density [17], which leads to locally planar but globally non-planar oxide thickness. The length scale over which this global non-planarity exists is called the *planarization length*, which is dependent on the specific CMP process and can be determined empirically. This phenomenon is conceptually illustrated in Figure 2(a).

Assuming fixed layout pattern density, CMP modeling and simulation have been extensively studied (e.g., in [2], [3], [4]) to characterize the profile of oxide being removed. Figure 2(b) defines the terms in a typical basic CMP model. The thickness of oxide removed by CMP is generally expressed as a nonlinear function of initial oxide step height z_1, blanket removal rate K, polishing time t, and pattern density $\rho(x, y)$ for a lateral location (x, y).

Of particular interest to us is the model described in [17], [4] that introduced the concept of *effective* pattern density, instead of using the original pattern density directly, for oxide thickness calculation. The reason is to accommodate the effect of neighboring patterns on a given polishing site due to pad deformation. Neighboring patterns are often referred to as those patterns that are within *planarization length distance* of the point of interest. Given a planarization length, an appropriate "weighting function" can be constructed to take the neighboring patterns and pad deformation into account, and "transform" the original pattern density into effective pattern density. More specifically, the effective pattern density $\rho(x, y)$ can be obtained by convolving the original real pattern density $d(x, y)$ with the weighting function $f(x, y)$:

$$\rho(x, y) = d(x, y) \otimes f(x, y) \tag{7}$$

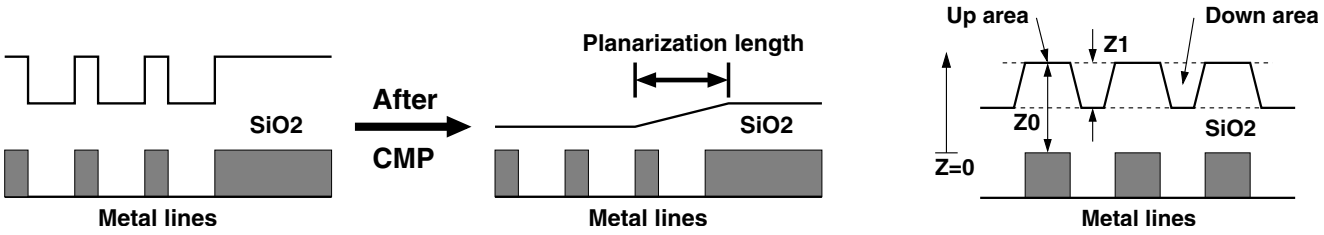
The convolution can be done efficiently via two-dimensional (discretized) Fast Fourier Transform (FFT) and inverse FFT (IFFT):

$$\rho(x, y) = \text{IFFT}(\text{FFT}(d(x, y)) \cdot \text{FFT}(f(x, y))) \tag{8}$$

An example of a weighting function we study in this paper is of Gaussian form:

$$f(x, y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{9}$$

where σ is one third of the planarization length, and is also the standard deviation of the Gaussian function. Then the scalar-valued thickness of trench and active oxide removed (i.e., the



(a) (b)

Fig. 2. (a) Conceptual cross-section view of oxide non-planarity after CMP. (b) Definitions in a basic CMP model.

"down area" *ARd* and "up area" *ARu*, respectively, as shown in Figure 2(b)) at location (x, y) can be calculated as [18]:

$$ARd(x,y) = \log\left[1 + e^{-\frac{\rho(x,y)z_1}{z_c}}\left(e^{\frac{Kt}{z_c}} - 1\right)\right] z_c \qquad (10)$$

$$ARu(x,y) = \frac{Kt}{\rho(x,y)} - \frac{1-\rho(x,y)}{\rho(x,y)} ARd(x,y) \qquad (11)$$

where z_c is the *contact height* – as the CMP progresses, the oxide step height is eventually reduced to contact height, at which point the pad touches the down area, and down area removal begins. The contact height can also be estimated empirically.

IV. Interval-Valued Statistical CMP Modeling

We start with representing uncertain CMP modeling parameters in affine intervals in Section IV-A, and build statistical CMP models with these interval-valued CMP modeling parameters in Section IV-B. Then as a "twin" problem to this, we build statistical CMP models with interval-valued pattern density in Section IV-C.

A. Representing Variational CMP Model Parameters with Affine Intervals

The polishing process is subject to manufacturing variations, and many aforementioned modeling parameters should thus be represented as random variables. In this paper, we consider initial step height, planarization length, and blanket removal rate, which generally have greater variations than other modeling parameters. The true physical and chemical sources of their variations are multi-fold and complex. Here we assume a simple hierarchical structure of spatially correlated variations similar to that in [19] (see Figure 3). For example, the initial step height values in the entire CMP area are affected similarly by a "global," possibly wafer-level variation source. Then the area is divided into four regions, with the initial step height values in each region being affected similarly by a "local," possibly reticle-level variation source. Obviously, this hierarchical structure can be extended to more levels of variations. As a result, as shown in Figure 3, the initial step height values in regions "5" and "6" are affected by different die-level variation sources, but share a common reticle-level variation in region "1" and a wafer-level variation in region "0." They can both be represented in the following affine forms:

$$\widehat{z}_{1,5} = z_{1,5} + \Delta z_{1,5,0}\varepsilon_0 + \Delta z_{1,5,1}\varepsilon_1 + \Delta z_{1,5,5}\varepsilon_5 \qquad (12)$$

$$\widehat{z}_{1,6} = z_{1,6} + \Delta z_{1,6,0}\varepsilon_0 + \Delta z_{1,6,1}\varepsilon_1 + \Delta z_{1,6,6}\varepsilon_6 \qquad (13)$$

$z_{1,5}$ and $z_{1,6}$ are the nominal values. Each unique source of global or local variation is modeled by an uncertainty symbol ε and its coefficient (with a 'Δ') is the *linearized* magnitude of the parameter uncertainty due to that particular source of variation [20]. Subscripts denote the contributing regions up this spatial hierarchy. Again, any individual source of variation may contribute to more than one parameter and hence direct correlations among these parameters. This is a simple model, but it can capture global and local variations and first-order correlations, and it maps perfectly onto our preferred affine interval model of computation.

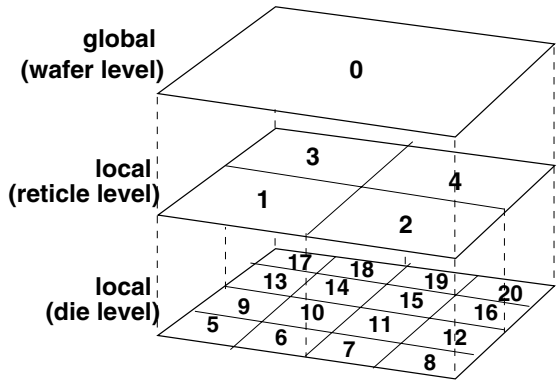


Fig. 3. A simple hierarchical structure of spatially correlated variations.

We model variational planarization length and blanket removal rate in affine forms analogously, and *retrofit* the deterministic CMP numerical model [17], [18], [4] to handle such variations in the following sub-section, based on the sophisticated affine arithmetic developed in [9].

B. Coping with CMP Model Parameter Variations

1) Pervasive Interval Computation Scheme: First we develop a statistical CMP modeling algorithm using affine interval computations *pervasively*. Put another way, we will strive to use interval-valued calculations as much as possible in the model. Thus, we expect to attain the fastest possible statistical model, but with some concomitant loss of fidelity. We start with the interval-valued weighting function:

$$\widehat{f}(x,y) = \frac{1}{\sqrt{2\pi}\widehat{\sigma}} e^{-\frac{x^2+y^2}{2\widehat{\sigma}^2}} \qquad (14)$$

where $\widehat{\sigma}$ is one third of our interval-valued planarization length. Then the interval-valued effective pattern density is obtained by convolving the fixed, scalar-valued original pattern density with the interval-valued weighting function:

$$\begin{aligned}\widehat{\rho}(x,y) &= d(x,y) \otimes \widehat{f}(x,y) \\ &= \text{IFFT}(\text{FFT}(d(x,y)) \cdot \text{FFT}(\widehat{f}(x,y)))\end{aligned} \tag{15}$$

Finally, the interval-valued thickness of trench and active oxide removed is calculated by

$$\widehat{ARd}(x,y) = \log\left[1 + e^{-\frac{\widehat{\rho}(x,y)\widehat{z}_1}{z_c}}\left(e^{\frac{\widehat{K}t}{z_c}} - 1\right)\right] z_c \tag{16}$$

$$\widehat{ARu}(x,y) = \frac{\widehat{K}t}{\widehat{\rho}(x,y)} - \frac{1-\widehat{\rho}(x,y)}{\widehat{\rho}(x,y)}\widehat{ARd}(x,y) \tag{17}$$

where $\widehat{z_1}$ is the interval-valued initial step height and $\widehat{K}$ is the interval-valued blanket removal rate.

It is easy to see that we follow the same computations of deterministic CMP modeling as outlined in Section III and replace every scalar operation with its affine interval counterpart. It is important to point out that the interval-valued FFT and IFFT are the most computation-intensive steps in the statistical CMP model, but only involve affine addition and scaling by scalar constants, which are perfect affine operations and do not create any estimation errors. The more accurate affine arithmetic of [9] is used for such nonlinear operations as multiplication, reciprocal, divisions, $exp(\cdot)$, and $log(\cdot)$. Further, note that very shallow nonlinear computation depth is required, i.e., Equations 14, 16, and 17 only, thus minimizing mis-estimations.

2) Hybrid Interval/Scalar Computation Scheme: As discussed in our previous work [10], it is not essential to use intervals in every step of the statistical CMP modeling algorithm. Approximation of affine interval endpoints and correlations creates errors rather like floating point roundoff errors, but more macroscopic and not so easy to "ignore." More nonlinear operations inevitably accrue more estimation errors. Therefore, recognizing the nonlinear operations in Equations 16 and 17 (even though at a rather shallow computation depth), we study another statistical CMP model that employs a *hybrid* interval/scalar computation strategy to trade off efficiency for accuracy.

Once we obtain the interval-valued effective pattern densities (by Equations 14 and 15), we stop interval-valued computation and switch to sampling over the uncertainty symbols in the pattern density intervals, as well as in the interval-valued initial step height and blanket removal rate, again using the sampling techniques suggested in [9] to produce asymmetric distributions. For each set of scalar-valued pattern density, initial step height, and blanket removal rate samples, we compute the scalar-valued thickness of trench and active oxide removed according to Equations 10 and 11, just like the deterministic CMP model. Enough random samples generate distributions, but not intervals, for the variational thickness of trench and active oxide removed. Therefore, this hybrid strategy, which exploits accurate linear interval-valued computations for the large-scale FFT and IFFT steps and scalar-valued sampling and computations that avoid interval estimation errors at the very last two nonlinear steps, turns out to be quite worthwhile for improving the accuracy, without much speed degradation.

C. Coping with Pattern Density Variations

The above two interval-valued statistical CMP models both assume a fixed layout pattern density as input and consider manufacturing uncertainties associated with some CMP modeling parameters. On the other hand, concerns exist that a layout that (marginally) meets design specifications (according to simulation) may still yield failing chips during the real manufacturing process, because the actual pattern density deviates from its nominal value due to manufacturing fluctuations, and may further increase ILD non-uniformity after CMP to an extent the manufacturing process cannot tolerate. Conventionally, this issue is addressed at a post-layout stage by either inserting dummy fill (to increase the density) or reverse etch-back (to decrease the density) [17]. In this paper, we propose to first analyze the pattern density variations at an early design stage from a circuit layout designer's point of view. In other words, we leverage the CMP model to consider *random variational* layout pattern density and predict its impact on the oxide removed by CMP, such that the prediction outcome can enable the designers to improve their layout and leave proper design margins to accommodate random manufacturing fluctuations. Therefore, we develop statistical CMP models for interval-valued pattern density in this sub-section. For simplicity, we assume the other CMP modeling parameters are deterministic at this early design stage.

First we represent the variational layout pattern density in affine forms similar to Section IV-A. That is, the interval pattern density values are weighted linear combinations of hierarchical global and local variations with first-order spatial correlations. Then we adopt a similar pervasive interval computation scheme, starting with the interval-valued effective pattern density:

$$\begin{aligned}\widehat{\rho}(x,y) &= \widehat{d}(x,y) \otimes f(x,y) \\ &= \text{IFFT}(\text{FFT}(\widehat{d}(x,y)) \cdot \text{FFT}(f(x,y)))\end{aligned} \tag{18}$$

followed by still using Equations 16 and 17 to compute the interval-valued thickness of trench and active oxide removed, based on the sophisticated affine arithmetic.

We also develop a similar hybrid interval/scalar computation scheme to improve the accuracy. We stop interval-valued computation once we obtain the interval-valued effective pattern density according to Equation 18, switch to Monte Carlo sampling over the effective pattern density intervals, and for each sample continue scalar-valued computations based on Equations 10 and 11. Again, enough random samples generate distributions for the variational oxide of trench and active oxide removed.

It is easy to see that the above two statistical CMP models with interval-valued pattern density are just twins of the two algorithms in Section IV-B, with even fewer affine approximations as only a scalar-valued weighting function is needed in

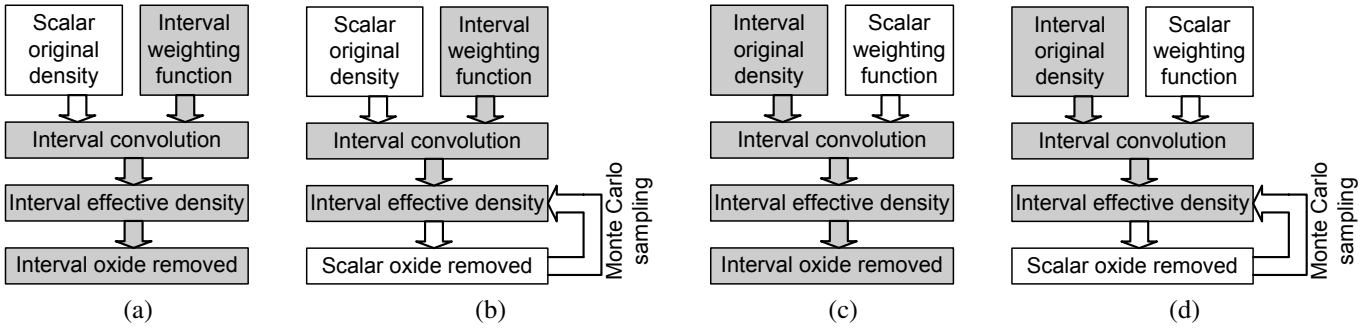


Fig. 4. Statistical CMP modeling algorithms using (a) pervasive interval computation scheme with interval-valued modeling parameters, (b) hybrid interval/scalar computation scheme with interval-valued modeling parameters, (c) pervasive interval computation scheme with interval-valued pattern density, and (d) hybrid interval/scalar computation scheme with interval-valued pattern density.

this case. We summarize and compare the four interval-valued statistical CMP modeling algorithms developed in Sections IV-B and IV-C in Figure 4.

V. Experimental Results

The affine arithmetic library and the four interval-valued statistical CMP modeling algorithms have been implemented in C/C++ and tested on a 1.0GHz UNIX machine using two layout pattern density maps, i.e., layout area divided into a regular grid of small square cells (see Table I). When representing interval-valued CMP modeling parameters and pattern density map in affine forms (e.g., Equations 12 and 13), we choose 3 to 4 levels of hierarchical uncertainties that correspond to 21 (=1+4+16) and 85 (=1+4+16+64) uncertainty symbols, respectively, according to Figure 3. Obviously, one symbol is for global variation and others for local variations distributed in hierarchical levels. For the interval-valued CMP model parameters, we assume 15%/5% and 5%/15% for the relative σ of global and local variations, and for the interval-valued pattern density, we assume 10%/10% and 5%/15%.

	number of cells	cell size
map1	128 × 128	$40\mu m \times 40\mu m$
map2	512 × 512	$40\mu m \times 40\mu m$

TABLE I
Numbers of cells and cell size for the two pattern density maps.

We compute the variational thickness of trench and active oxide removed using our proposed algorithms and compare the results with those of C/C++ based complete Monte Carlo simulations, i.e., randomly varying the CMP modeling parameters or pattern density according to the specified relative σ and repeating many times the deterministic CMP modeling procedure outlined in Section III[1]. For the sake of fairness and efficiency, we determine how many samples for each approach, parameter setting, and experiment using standard confidence interval methods [21]. We use samples sufficient to guarantee a 99% confidence level with 1% error. This is conducted for both the complete Monte Carlo simulations and the sampling steps in our proposed hybrid interval/scalar computation schemes. For almost all our experiments, this turns out to be between 3,000 and 4,000 samples. The results are shown in Tables II and III. The mean and standard deviation (std) given in columns 3, 4, 6, 7, 12, and 13 in both tables are the average mean and average standard deviation of the thickness of oxide removed, respectively, over all the cells in the density maps. And the data in columns 9–11 and 15–17 in both tables (i.e., the mean and standard deviation errors, and run time speedups) are the relative difference of our proposed approaches with respect to their corresponding complete Monte Carlo simulations.

Table II shows the results for variational thickness of active oxide removed with variational CMP modeling parameters. Compared to complete Monte Carlo simulation, the average mean error is 0.7%, the average standard deviation error is 9.6%, and the speedup is 90 for the pervasive interval computation scheme; the average mean error is 0.4%, the average standard deviation error is 4.5%, and the speedup is 9 for the hybrid interval/scalar computation scheme. Comparing between these two statistical CMP models, the pervasive interval computation scheme achieves about 10X more speedup, while the hybrid interval/scalar computation scheme achieves roughly half of the error, hence a clear tradeoff between accuracy and efficiency. Similar results are obtained for the variational thickness of trench oxide removed.

Table III shows the results for variational thickness of active oxide removed with variational pattern density. Compared to complete Monte Carlo simulation, the average mean error is 1.1%, the average standard deviation error is 5.0%, and the speedup is 91 for the pervasive interval computation scheme; the average mean error is 0.1%, the average standard deviation error is 0.2%, and the speedup is 9 for the hybrid interval/scalar computation scheme. Comparing between these two statistical CMP models, the pervasive interval computation scheme again achieves about 10X more speedup, while

[1]It is worth noting that we have, in fact, only *one* C/C++ CMP code, and that by overloading each numerical operator with the appropriate interval representation from [9], we "retrofit" an interval-valued statistical view on the original deterministic code. Thus, we actually compare against the deterministic version of the code with original scalar operators in a Monte Carlo loop.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
global / local variation	# of ε	Monte Carlo (MC) mean (Å)	Monte Carlo (MC) std (Å)	Monte Carlo (MC) time (s)	Pervasive scheme mean (Å)	Pervasive scheme std (Å)	Pervasive scheme time (s)	Pervasive vs. MC mean error	Pervasive vs. MC std error	Pervasive vs. MC speed -up	Hybrid scheme mean (Å)	Hybrid scheme std (Å)	Hybrid scheme time (s)	Hybrid vs. MC mean error	Hybrid vs. MC std error	Hybrid vs. MC speed -up
map1																
15% / 5%	21	3504.6	480.6	1487	3531.4	525.9	14	0.8%	9.4%	105	3516.9	500.5	112	0.4%	4.1%	12
	85	3505.2	481.1	1520	3538.9	527.3	23	1.0%	9.6%	65	3522.2	503.2	251	0.5%	4.6%	5
5% / 15%	21	3503.4	479.3	1493	3528.7	527.4	15	0.7%	10.0%	99	3515.1	499.5	109	0.3%	4.2%	13
	85	3504.7	480.5	1509	3536.6	528.2	26	0.9%	9.9%	57	3521.3	500.7	245	0.5%	4.2%	5
map2																
15% / 5%	21	3652.3	484.4	18390	3674.2	537.2	144	0.6%	10.9%	134	3659.6	507.7	1394	0.2%	4.8%	12
	85	3655.1	487.5	18482	3684.3	532.4	291	0.8%	9.2%	66	3669.7	508.5	2831	0.4%	4.3%	6
5% / 15%	21	3650.5	484.6	18378	3672.4	525.3	157	0.6%	8.4%	122	3665.1	508.8	1380	0.4%	5.0%	12
	85	3655.1	486.8	18463	3673.4	534.0	278	0.5%	9.7%	69	3673.4	509.7	2795	0.5%	4.7%	6

TABLE II

RESULTS FOR THE VARIATIONAL THICKNESS OF ACTIVE OXIDE REMOVED, WITH VARIATIONAL CMP MODELING PARAMETERS.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
global / local variation	# of ε	Monte Carlo (MC) mean (Å)	Monte Carlo (MC) std (Å)	Monte Carlo (MC) time (s)	Pervasive scheme mean (Å)	Pervasive scheme std (Å)	Pervasive scheme time (s)	Pervasive vs. MC mean error	Pervasive vs. MC std error	Pervasive vs. MC speed -up	Hybrid scheme mean (Å)	Hybrid scheme std (Å)	Hybrid scheme time (s)	Hybrid vs. MC mean error	Hybrid vs. MC std error	Hybrid vs. MC speed -up
map1																
10% / 10%	21	3511.4	255.7	1455	3546.5	267.0	14	1.0%	4.4%	103	3513.8	256.0	104	0.1%	0.1%	13
	85	3514.0	257.1	1487	3556.2	270.5	24	1.2%	5.2%	61	3517.9	258.5	242	0.1%	0.5%	5
5% / 15%	21	3510.9	255.6	1466	3553.0	268.1	12	1.2%	4.9%	121	3513.2	255.9	105	0.1%	0.1%	13
	85	3512.6	258.2	1492	3547.7	272.4	23	1.0%	5.5%	64	3513.7	258.7	234	0.0%	0.2%	5
map2																
10% / 10%	21	3660.3	267.0	18005	3693.6	279.5	146	0.9%	4.7%	122	3663.0	267.5	1349	0.1%	0.2%	12
	85	3665.9	267.5	18094	3709.3	280.6	265	1.2%	4.9%	67	3668.2	268.3	2712	0.1%	0.3%	6
5% / 15%	21	3661.8	269.2	18017	3699.4	284.5	139	1.0%	5.7%	129	3662.4	269.4	1364	0.0%	0.1%	12
	85	3666.4	270.1	18092	3708.1	282.5	279	1.1%	4.6%	64	3669.6	270.6	2788	0.1%	0.2%	5

TABLE III

RESULTS FOR THE VARIATIONAL THICKNESS OF ACTIVE OXIDE REMOVED, WITH VARIATIONAL PATTERN DENSITY.

the hybrid interval/scalar computation scheme now achieves roughly 10–25 times less error, hence again a clear tradeoff between accuracy and efficiency. Similar results are obtained for the variational thickness of trench oxide removed. Note that the modeling errors in the case of variational pattern density are much smaller than those in the case of variational physical parameters, as more nonlinear operations are needed to construct the interval-valued weighting function in the latter case.

To offer some visual evidence on the accuracy of our proposed approaches, we plot the means and standard deviations of the variational thickness of active oxide removed for *all* the cells of map1, with 21 uncertainty symbols in 3 levels, 5% global variation, and 15% local variation for the pattern density. In Figure 5 we compare our proposed approaches (both the pervasive interval scheme and the hybrid interval/scalar scheme) with complete Monte Carlo simulation. The correspondence is good, especially between the complete Monte Carlo simulation and the accurate, hybrid interval/scalar computation scheme. Note that the four periodic bumps in all the plots for standard deviations are artifacts of the simple hierarchical structure of uncertainties that we impose on the pattern density map.

VI. CONCLUSIONS

Even in physics-oriented modeling tools, statistically uncertain quantities appear as crucial inputs. We showed how we can approximate the necessary expensive statistical calculations by representing them as correlated intervals, and employing our recent ideas in interval-valued numerical computation. Using CMP as an example, we demonstrated how to "retrofit" workable statistical models on top of the original algorithm. Accuracies to within $\sim$1–10% of Monte Carlo simulation, and speedups of $\sim$10–100X were achieved, depending on whether we chose to trade off accuracy, or efficiency. Our current work focuses on further accuracy and efficiency improvements in interval-valued computation, and demonstration in other important design-for-manufacturing (DFM) oriented physical models.

VII. ACKNOWLEDGMENTS

We thank Julia Fei at CMU for helpful discussions on CMP. The work was in part supported by C2S2, the MARCO Focus Center for Circuit & System Solutions, under MARCO contract 2003-CT-888.

REFERENCES

[1] S. V. Postnikov, K. Lucas, K. Wimmer, V. Ivin, and A. Rogov, "Monte Carlo method for highly efficient and accurate statistical lithography simulations," in *Proc. SPIE, Optical Microlithography XV*, 2002.

[2] B. E. Stine, D. O. Ouma, R. R. Divecha, D. S. Boning, J. E. Chung, D. L. Hetherington, C. R. Harwood, O. S. Nakagawa, and S.-Y. Oh, "Rapid characterization and modeling of pattern-dependent variation in chemical-mechanical polishing," *IEEE Trans. on Semiconductor Manufacturing*, 1998.

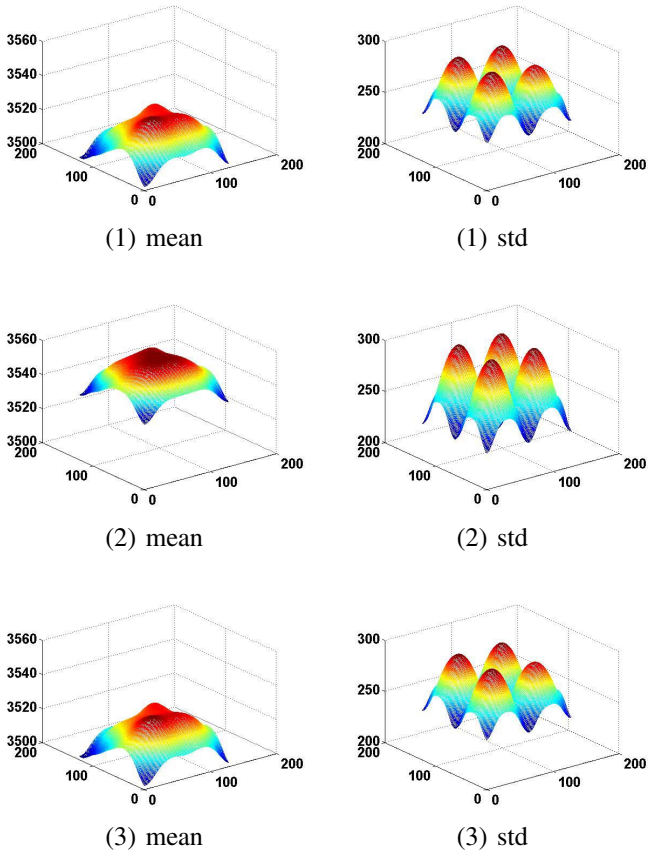


Fig. 5. Comparison among (1) complete Monte Carlo simulation, (2) pervasive interval scheme, and (3) hybrid interval/scalar scheme with variational pattern density for map1.

[3] C. Ouyang, K. Ryu, L. Milor, W. Maly, G. Hill, and Y.-K. Peng, "An analytical model of multiple ILD thickness variation induced by interaction of layout pattern and CMP process," *IEEE Trans. on Semiconductor Manufacturing*, 2000.

[4] D. O. Ouma, D. S. Boning, J. E. Chung, W. G. Easter, V. Saxena, S. Misra, and A. Crevasse, "Characterization and modeling of oxide chemical-mechanical polishing using planarization length and pattern density concepts," *IEEE Trans. on Semiconductor Manufacturing*, 2002.

[5] D. J. Frank, Y. Taur, M. Ieong, and H.-S. P. Wong, "Monte Carlo modeling of threshold variation due to dopant fluctuations," in *Proc. Symp. on VLSI Technology*, 1999.

[6] D. J. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Y. Taur, and H.-S. P. Wong, "Device scaling limits of Si MOSFETs and their application dependencies," in *Proc. IEEE*, 2001.

[7] J. Stolfi and L. H. de Figueiredo, *Self-Validated Numerical Methods and Applications*. Brazilian Mathematics Colloquium Monograph, IMPA, Rio De Janeiro, Brazil, 1997.

[8] J. D. Ma and R. A. Rutenbar, "Interval-valued reduced order statistical interconnect modeling," in *Proc. Int. Conf. on Computer Aided Design*, 2004.

[9] C. F. Fang, "Probabilistic interval-valued computation: Representing and reasoning about uncertainty in DSP and VLSI designs," Ph.D. dissertation, CMU, 2005.

[10] J. D. Ma and R. A. Rutenbar, "Fast interval-valued statistical interconnect modeling and reduction," in *Proc. Int. Symp. on Physical Design*, 2005.

[11] R. E. Moore, *Interval Analysis*. Prentice-Hall, 1966.

[12] C. F. Fang, R. A. Rutenbar, M. Püschel, and T. Chen, "Toward efficient static analysis of finite precision effects in DSP applications via affine arithmetic modeling," in *Proc. Design Automation Conf.*, 2003.

[13] M. R. Spiegel, *Theory and Problems of Probability and Statistics*. McGraw-Hill, 1992.

[14] H. Chang and S. Sapatnekar, "Statistical timing analysis considering spatial correlations using a single PERT-like traversal," in *Proc. Int. Conf. on Computer Aided Design*, 2003.

[15] C. Visweswariah, K. Ravindran, K. Kalafala, S. G. Walker, and S. Narayan, "First-order incremental block-based statistical timing analysis," in *Proc. Design Automation Conf.*, 2004.

[16] G. Nanz and L. Camilletti, "Modeling of chemical-mechanical polishing: A review," *IEEE Trans. on Semiconductor Manufacturing*, 1995.

[17] D. O. Ouma, "Modeling of chemical mechanical polishing for dielectric planarization," Ph.D. dissertation, MIT, 1998.

[18] X. Xie, T. Park, D. Boning, A. Smith, P. Allard, and N. Patel, "Characterizing STI CMP processes with an STI test mask having realistic geometric shapes," in *Chemical-Mechanical Polishing Symp., MRS Spring Meeting*, 2004.

[19] A. Agarwal, D. Blaauw, V. Zolotov, S. Sundareswaran, M. Zhao, K. Gala, and R. Panda, "Statistical delay computation considering spatial correlations," in *Proc. Asia South Pacific Design Automation Conf.*, 2003.

[20] V. Mehrotra, S. Nassif, D. Boning, and J. Chung, "Modeling the effects of manufacturing variation on high-speed microprocessor interconnect performance," in *Proc. IEEE Int. Electron Devices Meeting*, 1998.

[21] R. Burch, F. N. Najm, P. Yang, and T. N. Trick, "A Monte Carlo approach for power estimation," *IEEE Trans. on Very Large Scale Integration Systems*, 1993.