Model-Based Dummy Feature Placement for Oxide Chemical–Mechanical Polishing Manufacturability

Ruiqi Tian, D. F. Wong, *Member, IEEE*, and Robert Boone, *Member, IEEE*

***Abstract*—Chemical–mechanical polishing (CMP) is an enabling technique used in deep-submicrometer VLSI manufacturing to achieve long range oxide planarization. Post-CMP oxide topography is highly related to local pattern density in the layout. To change local pattern density and, thus, ensure post-CMP planarization, dummy features are placed in the layout. Based on models that accurately describe the relation between local pattern density and post-CMP planarization by Stine *et al.* (1997), Ouma *et al.* (1998), and Yu *et al.* (1999), a two-step procedure of global density assignment followed by local insertion is proposed to solve the dummy feature placement problem in the fixed-dissection regime with both single-layer and multiple-layer considerations. Two experiments conducted with real design layouts gave excellent results by reducing simulated post-CMP topography variation from 767 Å to 152 Å in the single-layer formulation and by avoiding cumulative effect in the multiple-layer formulation. The simulation result from single-layer formulation compares very favorably both to the rule-based approach widely used in industry and to the algorithm by Kahng *et al.* (1999). The multiple-layer formulation has no previously published work.**

***Index Terms*—Chemical–mechanical polishing, design for manufacturability, dummy features, linear programming.**

I. Introduction

CHEMICAL–MECHANICAL polishing (CMP) is a standard technique in deep-submicrometer very large scale integration (VLSI) manufacturing for achieving long range planarization of oxide topography on wafer [1]. Long-range planarization enlarges process window, which results in a more robust manufacturing process. Therefore, a complementary metal–oxide–semiconductor production process normally have four to seven or more steps of oxide or doped oxide CMP, starting from shallow trench isolation (STI) for the active layer to interlevel dielectric (ILD) layers for metal interconnects at backend of line (BEOL).

Continued aggressive scaling down of VLSI feature size has constrained much of the manufacturing process window so that CMP for oxide planarization has become increasingly important for manufacturability. Many models were proposed to understand the CMP process. However, physical models of CMP have largely been compromised by the complexity of issues involved [9]. Despite the difficulty, all physical experiment has confirmed that post-CMP oxide thickness is highly correlated to pattern density distribution of features on a layer, where pattern density is defined as the ratio of raised area to total area for an area of a given size. Hence, one consideration to ensure CMP manufacturability arises from the fact that to achieve post-CMP oxide planarization, pattern density distribution in layout has to satisfy certain relations *prior* to the CMP process. Dummy features—features that are electrically inactive and are not for the purpose of optical assistance—are inserted into layout to change pattern density distribution. This insertion procedure is sometimes called "tiling" because the dummy features inserted typically are small polygons of similar shape. The small polygons are thus called "tiles" and they are usually squares or rectangles for simplicity. In the rest of this paper, the term "tiling" ("tiles") and dummy feature placement (dummy feature) are used interchangeably. Fig. 1 illustrates an exaggerated wafer cross section of two layers in which ILD thicknesses vary and tiling helps.

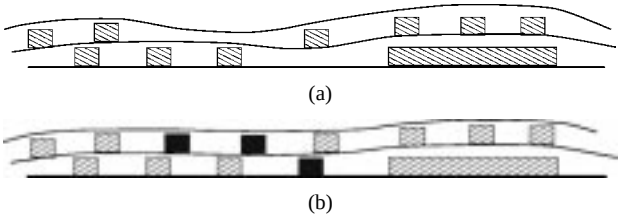


Fig. 1. Cross section of the wafer showing two layers. (a) ILD planarization problem without tiling is exaggerated for clarity. (b) Dummy features (dark) inserted.

Methods for tiling can be classified into two categories: rule-based and model-based. Rule-based tiling are from the experience that ILD thickness is directly proportional to local pattern density—pattern density for a small region of usually 50 by 50 to 500 by 500 μm^2; hence, physical design rules require local pattern density on a layer to be between a lower and an upper bound. Consequently, wherever there is open space large enough, tiles should be inserted to bring local pattern density within the bounds. This is usually done with Boolean operations to find the open space and fill it with tiles of a prescribed density. The problem with rule-based tiling is that the range for allowed density is usually fairly large, such that the density to prescribe must go through trial-and-error for every design and yet sometimes no single value works.

Compared to rule-based approach, model-based methods based on analytical expressions, which are not necessarily just simple proportionalities for the relation between local pattern density and post-CMP oxide thickness, allow both local tile

Manuscript received July 20, 2000; revised November 15, 2000. The work of D. F. Wong was supported in part by the National Science Foundation under Grant CCR-9912390. This paper was presented in part at the Design Automation Conference, Los Angeles, CA, June 2000. This paper was recommended by Associate Editor W. Schoenmaker.

R. Tian is with the Department of Computer Sciences, University of Texas at Austin, Austin, TX 78712 USA and also with Motorola Inc., Austin, TX 78721 USA (e-mail: ruiqi@cs.utexas.edu).

D. F. Wong is with the Department of Computer Sciences, University of Texas at Austin, Austin, TX 78712 USA (e-mail: wong@cs.utexas.edu).

R. Boone is with Motorola Inc., Austin, TX 78721 USA (e-mail: Robert_Boone@email.sps.mot.com).

Publisher Item Identifier S 0278-0070(01)04830-8.

0278–0070/01$10.00 © 2001 IEEE

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on March 20,2025 at 08:25:55 UTC from IEEE Xplore. Restrictions apply.

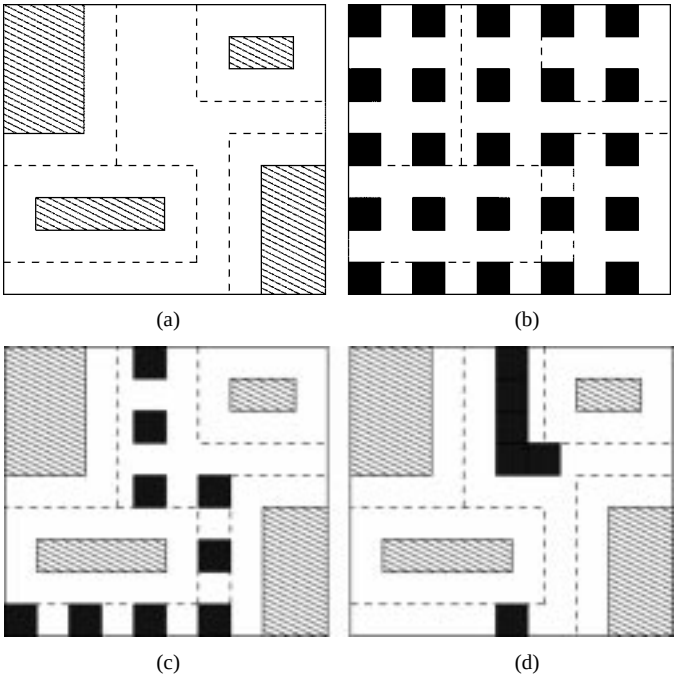


Fig. 2. Rule-based and model-based tiling. (a) Original layout with features lightly shaded and exclusion zones in dashed lines. (b) 25% dense tiling template in rule-based approach. (c) Result of the rule-based tiling after Boolean operations. (d) Possible model-based tiling result for the same layout. Notice tiles in (d) have different location and density from those in (c).

density and insertion location to vary. Fig. 2 illustrates the difference between rule-based and model-based tiling. Obvious to the observer, model-based methods provide more accuracy and efficiency.

In general, given a model of the relationship between pattern density distribution and final oxide topography after CMP, the dummy feature placement problem is to determine the amount and location of dummy features to place into the layout so that certain constraints such as electrical and physical design rules are observed and certain objectives such as minimum or ranged variation are satisfied by post-CMP topography.

Recent models of CMP by Stine *et al.* [7], Ouma *et al.* [5], and Yu *et al.* [10] have enabled faster and more accurate prediction of ILD thickness from computing an effective initial pattern density by filtering local pattern density with a weighing function in a weighing region of millimeters squared. (See Section II for details on definition of effective density.) Based on these models, this paper proposes a two-step solution to the dummy feature placement problem with both single-layer and multiple-layer considerations in the fixed-dissection regime—one that divides the layout into a grid of small equal rectangles. The first step uses linear programming (LP) to compute the amount of dummy feature required in each small rectangle. The second step then places the calculated amount into each rectangle while optionally optimizing certain local properties.

In the sections that follow, the models used by this paper and the only previous work from [3] are reviewed in Section II for completeness. Section III describes notation and the two-step approach. Computational experience on tiling two real design layouts from Motorola, one using the single-layer and the other using the multiple-layer formulation from Section III, along with comparison to a rule-based approach and comparison to the algorithm in [3] are presented in Section IV. Finally, some concluding remarks are in Section V.

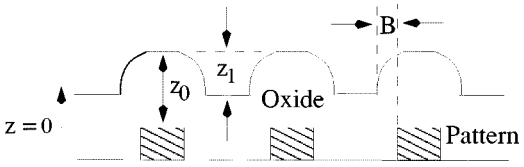


Fig. 3. Some variables in CMP (from [5]).

II. Models for the Oxide CMP Process

Several models were proposed for oxide planarization by CMP [4]. In contrast, the semiempirical model by Stine *et al.* based on Preston's equation is not computationally expensive nor difficult to calibrate [7]. In that model, ILD thickness z at location (x, y) is solved to be

$$z = \begin{cases} z_0 - [K_i t/\rho_0(x, y)] & t < (\rho_0 z_1/K_i) \\ z_0 - z_1 - K_i t + \rho_0(x, y) z_1 & t > (\rho_0 z_1/K_i) \end{cases} \quad (1)$$

where

K_i	blanket oxide polishing rate;
z_0	thickness of oxide deposition;
z_1	initial step height;
t	total polish time;
$\rho_0(x, y)$	initial oxide pattern density before CMP.

Fig. 3 shows a schematic for some of the variables. Specifically, to calculate oxide pattern density from the underlying feature pattern in layout, a bias B is first applied to model the *average* effect of oxide deposition. This model assumes straight sidewall profiles and perfect gap filling during oxide deposition so that resulting oxide pattern density is independent of z. All polygons in the layout are thus enlarged or shrunk from all sides by amount B. Values for B are process dependent and need calibration. Normally, total polish time t is larger than $(\rho_0 z_1/K_i)$, so final oxide thickness, by the second case of (1), is between 0 and $(z_0 - z_1)$. In addition, all K_i, z_0, z_1, and t are constants for a specific CMP process. As a result, the final topography is determined only by the initial oxide pattern density $\rho_0(x, y)$.

The simplest model uses the local oxide pattern density in a layout for $\rho_0(x, y)$. An algorithm by Kahng *et al.* [3] solves the dummy feature placement problem based on this model. In their algorithm, the objective of minimizing final ILD thickness variation is translated to minimizing pattern density variation within all possible floating rectangular regions of a given size (called *windows*). Obviously, this min-variation formulation is correct for the simple model and is very useful in the design phase to guarantee certain density range in the layout.

However, more accurate modeling by Ouma *et al.* considers the deformation of polishing pad during polish [5]. The initial oxide density $\rho_0(x, y)$ is no longer directly proportional to local oxide pattern density, but calculated as the summation of weighted local pattern density within a weighing region. $\rho_0(x, y)$ is then called the *effective* density from averaging. The weighing function $f(x, y)$, derived from elastic material under a circular uniform load applied normal to the material surface,

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on March 20,2025 at 08:25:55 UTC from IEEE Xplore. Restrictions apply.

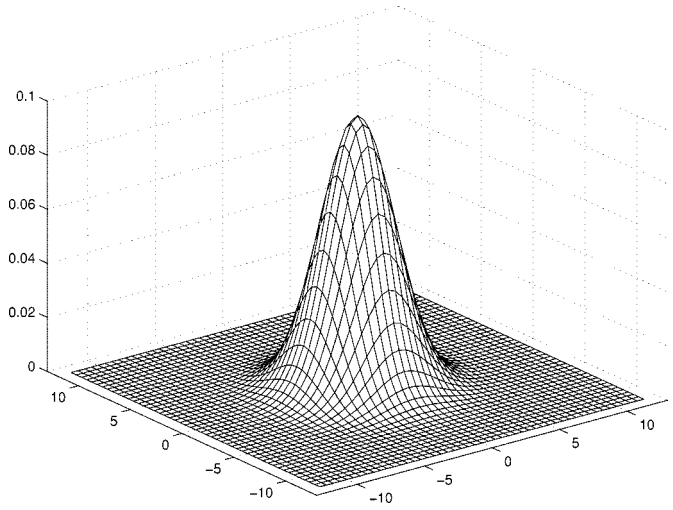


Fig. 4. Weighing function with $c_0 = 0.1$, $c_1 = -0.1$, and $c_2 = 1$. Units for both x and y are in millimeters.

is an elliptical function. Size of the weighing region depends on the interaction distance, which is the length at which the relative weight in $f(x, y)$ drops to certain negligible level. The interaction distance, whose value is typically several millimeters or more, depends on the specific condition of a CMP process, so calibration with direct measurements from test patterns is needed. The elliptical function does not go to zero as quickly as experimental data shows, so the function is cut off at a distance large enough from the center of the load. Further study by Travis *et al.* [8] assumed an approximation to the elliptical f to be

$$f(x, y) \approx c_0 \exp\left[c_1 \left(x^2 + y^2\right)^{c_2}\right]$$

where constants c_0, c_1, and c_2 are calibrated for each specific process. The approximation avoids the somewhat arbitrary cut off of f because its Gaussian-like behavior lets the function goes quickly to zero as distance increases. The interaction distance is then the length at which the relative weight drops to $1/e$ or $1/e^2$. Fig. 4 shows an example of the approximation.

In the fixed-dissection regime in which layout area is divided into a grid of small rectangles, local oxide pattern density after biasing $d(i, j)$ is determined for each rectangle. (See Section III-A for notational details.) f is discretized accordingly with respect to the grid. The discretized effective local oxide pattern density $\rho_0(i, j)$ is then

$$\rho_0(i, j) = \mathrm{IFFT}[\mathrm{FFT}[d(i, j)] \cdot \mathrm{FFT}[f(i, j)]]. \qquad (2)$$

Therefore, in effect, the CMP process is modeled by (2) as a low-pass filter through which the local pattern density d not only contributes to immediate, but also short range ILD thickness within the weighing region defined by the interaction distance of f [5].

Furthermore, the variation in ILD thickness is cumulative from layer to layer, as shown in Fig. 1. Each layer except the first one cannot assume a perfectly flat starting surface. Therefore, tiling each layer individually may not achieve desired planarization when layers are stacked together later in manufacturing. In the model for cumulative effect by Yu *et al.* [10], topography

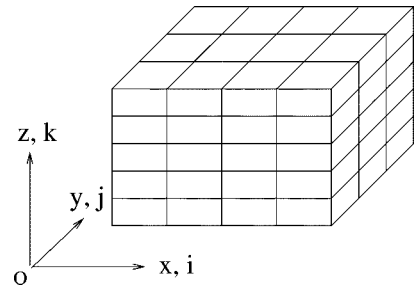


Fig. 5. Fixed-dissection of a layout.

variation of one layer attenuates through the subsequent CMP processes, each of which is modeled as a low-pass filter based on (1) and (2). Mathematically

$$\widehat{\rho_{0(k)}} = \begin{cases} \left[\widehat{d_k} + \left(\dfrac{z_{k-1}}{z_k}\right) \widehat{\rho_{0(k-1)}}\right] \times \hat{f} & \text{if } k > 1 \\ \widehat{d_1} \times \hat{f} & \text{if } k = 1 \end{cases} \qquad (3)$$

where

"^"	FFT operator;
$\rho_{0(k)}$	effective local density;
z_k	step height;
d_k	local density, all for layer k;
f	weighing function.

For example, the contribution of layer one to layer three is the ILD thickness of post-CMP oxide of layer one passed through two additional low-pass filters representing the CMP processes at layer two and layer three above, i.e.,

$$\widehat{\rho_{0(3)}} = \hat{f} \times \left(\widehat{d_3} + \left(\frac{z_2}{z_3}\right) \hat{f} \times \left(\widehat{d_2} + \left(\frac{z_1}{z_2}\right) \hat{f} \times \widehat{d_1}\right)\right).$$

III. Dummy Feature Placement

Equations (1)–(3) state post-CMP topography as a function of process condition (f, t, K_i, z_0, z_1, B) and local oxide pattern density (d). The process condition is fixed for a particular CMP process, so final topographies are differentiated only by different initial distributions of local oxide pattern density. In the rest of this section, the notation for the problem of dummy feature placement in the fixed-dissection regime is first introduced; then, based on (1)–(3), a two-step approach is used to solve the problem—global density assignment by LP followed by local dummy feature insertion. Both single-layer and multilayer formulation are presented.

A. General Assumption and Notation

Because oxide pattern is obtained by enlarging or shrinking all polygons in the layout with a bias amount B in all directions (see Fig. 3), any polygon from here on, unless otherwise noted, is assumed to be after the bias B is applied.

In the fixed-dissection regime, suppose each layer of K layers of layout area is divided into $(M \times N)$ small rectangles of equal dimensions and, thus, equal area A^0, then each small rectangle is called a *cell* and the ordered tuple (i, j, k) with $i \in [1, M]$, $j \in [1, N]$, and $k \in [1, K]$ is the coordinate of each cell. Fig. 5 shows the coordinate system for a layout in the fixed-dissection

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on March 20,2025 at 08:25:55 UTC from IEEE Xplore. Restrictions apply.

regime. For the cell at (i, j, k), let $P^0(i, j, k)$ denote all polygons inside the cell that are originally in the layout. If a polygon in the layout crosses the cell boundary, the cell boundary divides the polygon into smaller abutting polygons so that the resulting polygons are either inside or outside a cell. Similarly, let $P(i, j, k)$ denote all dummy features (polygons) inserted inside the same cell. In addition, let the function $\mathrm{Area}(p)$ return the area of an arbitrary polygon p. Therefore

$$x_{ijk} = \sum_{p \in P(i,j,k)} \left[\mathrm{Area}(p)/A^0\right]$$

denotes the density for the amount of dummy feature inserted into the cell at (i, j, k). Similarly, the density of features originally in the layout x^0_{ijk} is

$$x^0_{ijk} = \sum_{p \in P^0(i,j,k)} \left[\mathrm{Area}(p)/A^0\right].$$

Clearly, $d(i, j, k) = x_{ijk} + x^0_{ijk}$ in (2) for a given k.

To formulate constraints from electrical and physical design rules, cost zones (polygons) q are used to define regions of different costs $\mathrm{Cost}(q) \in [0, \infty)$ and $Q(i, j, k)$ denotes all cost zones in the cell at (i, j, k). Cost is uniform within a cost zone and a cost zone is called an exclusion zone if it has infinite cost so that no dummy feature can be placed inside. Exclusion zones are sufficient to implement physical design rules such as nonoverlapping rule and minimum spacing (buffer) rule. Hence, the total *available* density in each cell for inserting dummy feature is

$$x^a_{ijk} \le 1 - \sum_{\substack{q \in Q(i,j,k),\\ \mathrm{Cost}(q)=\infty}} \left[\mathrm{Area}(q)/A^0\right].$$

Furthermore, the average cost of inserting dummy feature into the cell at (i, j, k) is

$$\sigma_{ijk} = \sum_{\substack{q \in Q(i,j,k),\\ \mathrm{Cost}(q)\neq\infty}} \left(\mathrm{Cost}(q) \cdot \left[\mathrm{Area}(q)/A^0\right]\right).$$

Finally, if $K = 1$, then all subscript and variable k are dropped in the notation to reflect a two-dimensional layout.

B. Global Density Assignment

1) Single-Layer Consideration: From digital signal processing theory, the multiplication of d and f in frequency domain in (2) is the same as circular convolution of those two functions in physical domain, provided that d is periodically repeated extending to infinity, and f is a linear shift-invariant filter, i.e., contributions from different locations are linearly additive [2]. For the CMP process, the area defined by the layout, called the *reticle field*, is repeated in both the x and y directions many times on the wafer. If the reticle field is small compared to the total area on wafer, as it is almost always the case, the requirement for d repeating infinitely is satisfied. Meanwhile, the CMP process does not have a preferred location nor a preferred direction, so filter f satisfies the linear shift-invariant

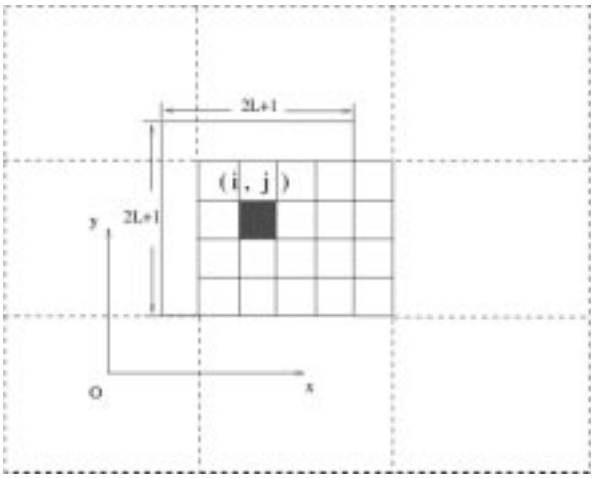


Fig. 6. Convolution implementation of filter f.

criterion. Therefore, assuming long range contribution outside the weighing region $[(-L, L) \times (-L, L)]$ is negligible, (2) can be rewritten as a circular convolution

$$\rho_0(i, j) = \sum_{i'=i-L}^{i+L} \sum_{j'=j-L}^{j+L} \left[\left(x_{i'j'} + x^0_{i'j'}\right) \cdot f(i' - i, j' - j)\right]. \tag{4}$$

An implementation of (4) is shown in Fig. 6 in which the reticle field is the center region with 5×4 cells and is repeated as dashed rectangles in both x and y directions. For the shaded cell at (i, j), (4) sums over the large square region marked with width $(2L + 1)$ and length $(2L + 1)$. Because the upper-left corner in the layout is actually adjacent to the lower-right corner in circular convolution, the summation in (4) is usually achieved by revolving indices of cells in the center region. The revolving index is not needed if the $(2L+1)\times(2L+1)$ convolution region is totally within a single die.

By subjecting $\rho_0(i, j)$ to linear constraints, (4) becomes a system of linear equations with $(M \times N)$ variables x_{ij}. The constraints are bounding the effective oxide pattern density, which scales to the ILD thickness with a constant factor z_1 by (1). Therefore, the dummy feature placement problem at the global stage, where all cells are considered, becomes an LP problem as x_{ij} in the LP solution means how much dummy feature is needed to place within the cell at location (i, j). Because $\rho_0(i, j)$, an effective density itself, is bounded from above by unity, the LP formulation for minimum topography variation (Min-Variation) is

minimize

$$\rho^H - \rho^L$$

subject to

$$0 \le \rho^L \le \rho_0(i, j) \le \rho^H \le 1$$
$$0 \le x_{ij} \le x^a_{ij} \tag{5}$$

where ρ^L and ρ^H are auxiliary variables and constraint (5) ensures that exclusion zones are observed as well as that no existing feature is deleted.

The Min-Variation formulation usually serves as a feasibility test for manufacturability since the objective $\rho^H - \rho^L$ in solution

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on March 20,2025 at 08:25:55 UTC from IEEE Xplore. Restrictions apply.

is the absolute minimum variation that can be achieved. In contrast, a *ranged* variation, where a process budget ϵ bounds the final topography, is required to control manufacturability. For the objective function, a linear function utilizing the cost zones or even the null function can be used. Hence, a more useful LP formulation for ranged variation (Ranged-Variation) is

minimize

$$\sum_{i,\,j}(\sigma_{ij}\cdot x_{ij}) \tag{6}$$

subject to

$$\begin{aligned} &0 \le \rho^L \le \rho_0(i,\,j) \le \rho^H \le 1\\ &\rho^H - \rho^L \le \epsilon\\ &0 \le x_{ij} \le x^a_{ij}. \end{aligned} \tag{7}$$

In this ranged-variation formulation, manufacturability is ensured by constraints in (7) and the cost zones for objective (6) can be used to model different considerations. For example, if all nonexclusion zones have constant cost, i.e., σ_{ij} is a constant, then the total amount of dummy feature inserted is minimized by (6) to maximize polish rate and thus throughput.

2) Multiple-Layer Consideration: By simple mathematical induction on the layer number k and due to the linearity of Fourier transforms, (3) can be written as

$$\widehat{\rho_{0(k)}} = \sum_{l=1}^{k}\left[\left(\frac{z_l}{z_k}\right)\hat{f}^{(k-l+1)}\times\hat{d}_l\right]. \tag{8}$$

For the height at a location $(i,\,j)$ on layer k, each term in the summation of (8), when converted to the physical domain, results in a multiple circular convolution

$$\begin{aligned} &\left[\mathrm{IFFT}\left(\hat{f}^{(\alpha)}\times\hat{d}_l\right)\right](i,\,j)\\ &=\left[\overbrace{(f\otimes f\cdots f)}^{\alpha}\otimes d_l\right](i,\,j)\\ &=\sum_{i_1}\sum_{j_1}\Bigg[f(i_1-i,\,j_1-j)\\ &\qquad\times\left(\sum_{i_2}\sum_{j_2}f(i_2-i_1,\,j_2-j_1)\right.\\ &\qquad\qquad\times\cdots\left(\sum_{i_\alpha}\sum_{j_\alpha}\left(f\left(i_\alpha-i_{\alpha-1},\,j_\alpha-j_{\alpha-1}\right)\right.\right.\\ &\qquad\qquad\qquad\left.\left.\left.\times\left(x_{i_\alpha j_\alpha l}+x^0_{i_\alpha j_\alpha l}\right)\right)\right)\right)\Bigg] \end{aligned} \tag{9}$$

where each double summation adds up the weighted contribution from a region of size $[(-L,\,L)\times(-L,\,L)]$ for each location in the next outer double summation.

Because the multiple convolution written as a series of summations in (9) is linear in term of x, the min-variation formulation for single layer is easily extended for multiple layers as Multiple-Min-Variation

minimize

$$\sum_{k}\left(\rho^H_k-\rho^L_k\right)$$

subject to

$$\begin{aligned} &0 \le \rho^L_k \le \rho_0(i,\,j,\,k) \le \rho^H_k \le \sum_{l=1}^{k}\left(\frac{z_l}{z_k}\right)\\ &0 \le x_{ijk} \le x^a_{ijk} \end{aligned} \tag{10}$$

where $\sum_{l=1}^{k}(z_l/z_k)$ is the upper bound on the cumulative height for layer k. Similarly, the ranged-variation formulation translate to Multiple-Ranged-Variation

minimize

$$\sum_{i,\,j,\,k}(\sigma_{ijk}\cdot x_{ijk})$$

subject to

$$\begin{aligned} &0 \le \rho^L_k \le \rho_0(i,\,j,\,k) \le \rho^L_k \le \sum_{l=1}^{k}\left(\frac{z_l}{z_k}\right)\\ &\rho^H_k-\rho^L_k \le \epsilon_k\\ &0 \le x_{ijk} \le x^a_{ijk} \end{aligned}$$

where ϵ_k and σ_{ijk} can differ among layers to achieve a solution that emphasize particular (usually lower) layers. Also, if process uniformity in terms of deviation of maximum within-layer variation from layer to layer is desired to be within a positive amount β, circular constraints

$$-\beta \le \left(\rho^H_k-\rho^L_k\right)-\left(\rho^H_{k-1}-\rho^L_{k-1}\right) \le \beta\ (k\in[2,\,K])$$

and

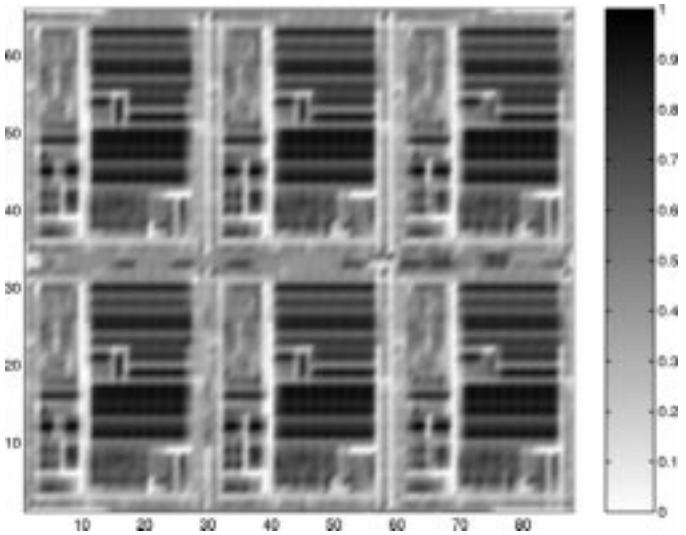
$$-\beta \le \left(\rho^H_1-\rho^L_1\right)-\left(\rho^H_K-\rho^L_K\right) \le \beta$$

can be added to the formulation above to control the variation between layers.

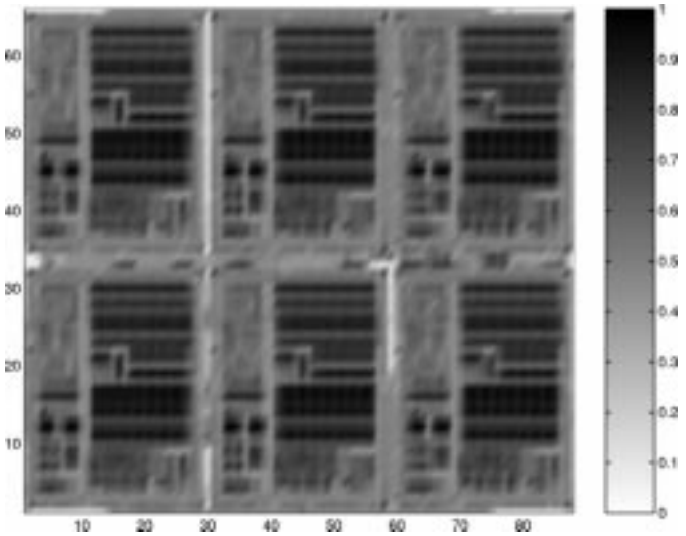
C. Local Dummy Feature Insertion

After the density x_{ijk} is determined by LP globally, the total area of dummy feature needed for a single cell is $A_{ijk}=x_{ijk}\cdot A^0$, and A_{ijk} should be feasible by LP constraints in (10). Then at the local level in layout, dummy features are distributed within available areas of a cell $(\mathrm{Cost}(q)\neq\infty)$ such that total area inserted is A_{ijk}. Available areas, constrained by electrical and physical design rules, may be highly irregular in shape. The dummy features of choice therefore are very small polygons (tiles), so that arbitrary shapes can be covered easily. To avoid design rule violations and pathological geometric difficulties during insertion, available area x^a_{ijk} should be determined *before* LP by sizing and Boolean operations. Moreover, actual location of tiles inside a cell does not change the original LP computation if the cell is small and insertion is random, because CMP interaction length (millimeters) is of orders larger than anything on the local level, thus making CMP intrinsically a global effect.

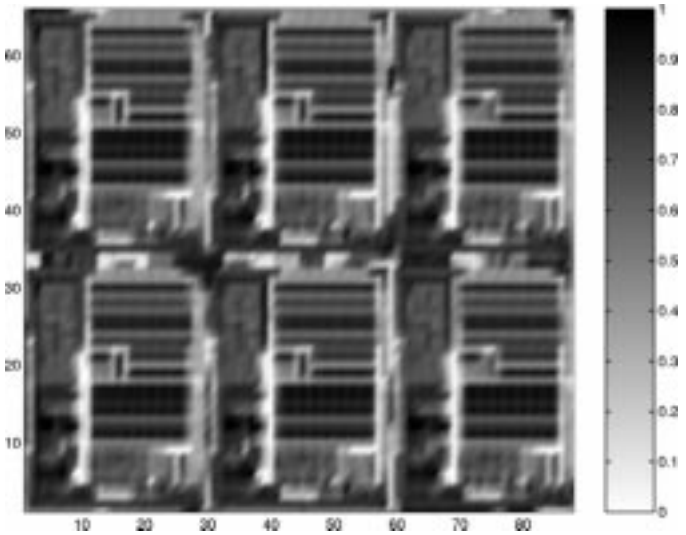
Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on March 20,2025 at 08:25:55 UTC from IEEE Xplore. Restrictions apply.



(a)



(b)



(c)

Fig. 7. Density d of final masks for a DSP chip from Motorola. (a) Original layout. Dark area represents the memory, gray for random logic, and white for open spaces. (b) Rule-based tiling. (c) Model-based tiling. Units for x and y are 0.25 mm.

The placement of tiles inside a cell has many solutions. The simplest method is to grid the cell and place tiles of same shape and size (usually squares or donuts) at each intersection. After the tiles outside available area are removed by Boolean operations, tiles not removed are selected at random until A_{ijk} is reached. Attention is needed in implementation for the boundary condition of very small available areas (slivers). Deletion of slivers before the available area calculation in LP simplifies local insertion.

Although local tile insertion has little global impact, more useful approaches consider its local impact of increased interconnect capacitance due to coupling. Assuming tiles are small and tile-to-tile coupling negligible, the coupling in tiling is inversely proportional to the (buffer) distance between tiles and original feature [6]. Many algorithms and heuristics, such as greedy, iterative improvement, or simulated annealing, can be employed to minimize coupling by maximizing the buffer distance inside a cell under the constraint of available area.

IV. Computational Experience

A. Single Layer

The mask in the first experiment, which includes six same dies in the reticle field with approximately two million polygons in each die, is for STI of a real design of a digital signal processor (DSP) from Motorola. The layout has embedded memory, random logic, and some empty spaces in each die and various process control and alignment structures between the dies. The layout is tiled with both rule-based and model-based tiling algorithms for comparison. Fig. 7 shows the oxide pattern density maps for both the rule-based and model-based tiling as measured from final masks. Fig. 7 also show the map of the original untiled layout. Clearly, the rule-based approach inserted tiles in all open spaces so that the density map in Fig. 7(b) has an overall shade of gray comparing to Fig. 7(a). In contrast, the model-based method puts tile at different location with very different densities, as shown in Fig. 7(c).

The two-step process in model-based tiling was performed with Ranged-Variation LP formulation from Section III using commercial design rule checker (DRC) tools for local density acquisition, commercial LP package for LP computation, and internally developed geometry engine and software for polygon manipulation in local insertion. The reticle field was divided into $M = 68 \times N = 88$ squares with approximately 0.25-mm sides each, so the size of the LP problem was $M \times N = 5808$ variables with 5808 bounds and $2 \times M \times N + 3 = 11\,619$ constraints. The parameter ϵ in the ranged-variation LP formulation was 200 Å/7000 Å. All softwares were executed on a Sun Enterprise 3000 server with 200-MHz CPUs and 6 GB of memory. CPU time is approximately 5 min for density acquisition, approximately 10 min for LP, and approximately 10 min for local tile insertion. For comparison, the algorithm with min-variation LP formulation in [3] was also performed on the same data with comparable CPU time observed.

Table I lists the resulting ρ_0 ranges from the different approaches. The value for range $(z) = z_1 \cdot [\max(\rho_0) - \min(\rho_0)]$ is calculated with step height $z_1 = 7000$ Å. Fig. 8 also shows the

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on March 20,2025 at 08:25:55 UTC from IEEE Xplore. Restrictions apply.

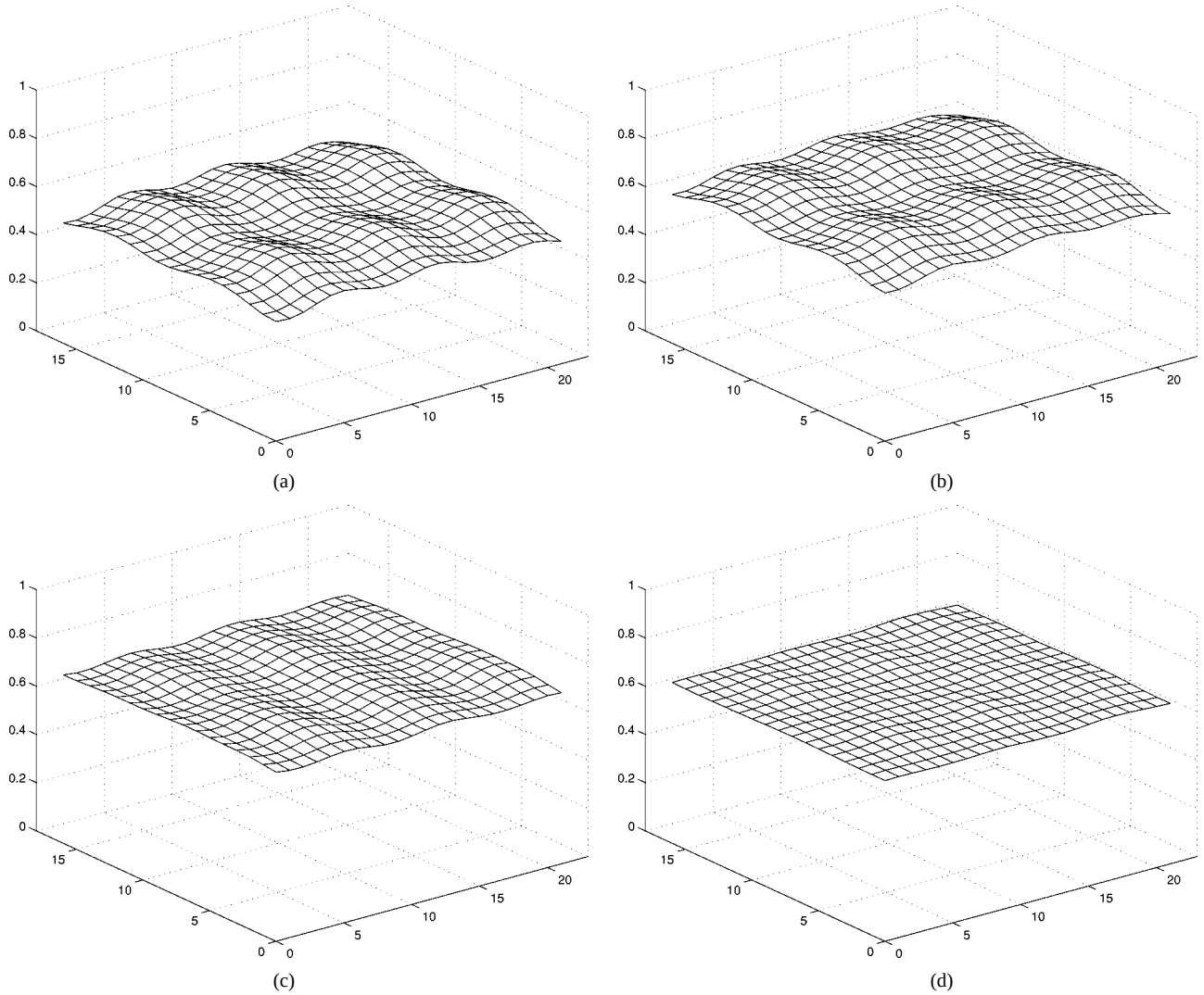


Fig. 8. Post-CMP topographies represented by ρ_0 (in z direction). Units for x and y are millimeters. (a) Original layout. (b) Rule-based approach. (c) Previous work in [3]. (d) Current work.

TABLE I
COMPARISON OF RESULTS FOR DIFFERENT TILING APPROACHES

Data	$\min(\rho_0)$	$\max(\rho_0)$	range(z)
original	0.4521	0.5618	767Å
rule-based approach	0.5724	0.6728	702Å
LP formulation in [3]	0.6499	0.7011	358Å
current work	0.6221	0.6438	152Å

resulting distributions of ρ_0 from the different approaches. Obviously, the rule-based method is not adequate. Also, in comparison to the approach presented in this paper, the LP formulation from [3] is inserting more dummy feature, thus having higher overall ρ_0 values and yet the final variation in topography is still very large.

B. Multiple Layers

The experiment for tiling multiple layers used three layers of metals of another real design from Motorola with six dies in the reticle field and various process control structures between the dies. In this layout, there are approximately 32 million polygons total on layer one, 572 thousand total on layer two, and 444 thousand total on layer three. The density maps for the layers are omitted here. The step heights for the three layers are $z_1 = 7000$ Å and $z_2 = z_3 = 8000$ Å.

In first part of the experiment, each of the three layers is tiled *individually* using the ranged-variation formulation from Section III with a target maximum variation of $\epsilon = 300$ Å/z_k for the three layers of $z_k = z_1, z_2, z_3$ and then the cumulative effect of the resulting layers is calculated. In the second part, the three layers are tiled using the Multiple-Ranged-Variation formulation from Section III with target maximum variations for each layer being $\epsilon_k = 300$ Å/$z_k (k = 1, 2, 3)$. Fig. 9 shows the results of both parts of the experiment. All $\rho_{0(k)}$ is normalized with formula $\rho_{0(k)} = \rho_{0(k)} - \min(\rho_{0(k)}) + 0.3(k-1) + 0.2$. This normalization is to show only the variation for each layer and to shift the variations to fit a total range of [0, 1]. Clearly, the cumulative effect of multiple layers degrades the results of tiling each layer individually. At the top layer, the cumulative variation is about 540 Å, almost doubling the target intended. In contrast, tiling considering the cumulative effect has every layer within the target bound of 300 Å.

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on March 20,2025 at 08:25:55 UTC from IEEE Xplore. Restrictions apply.

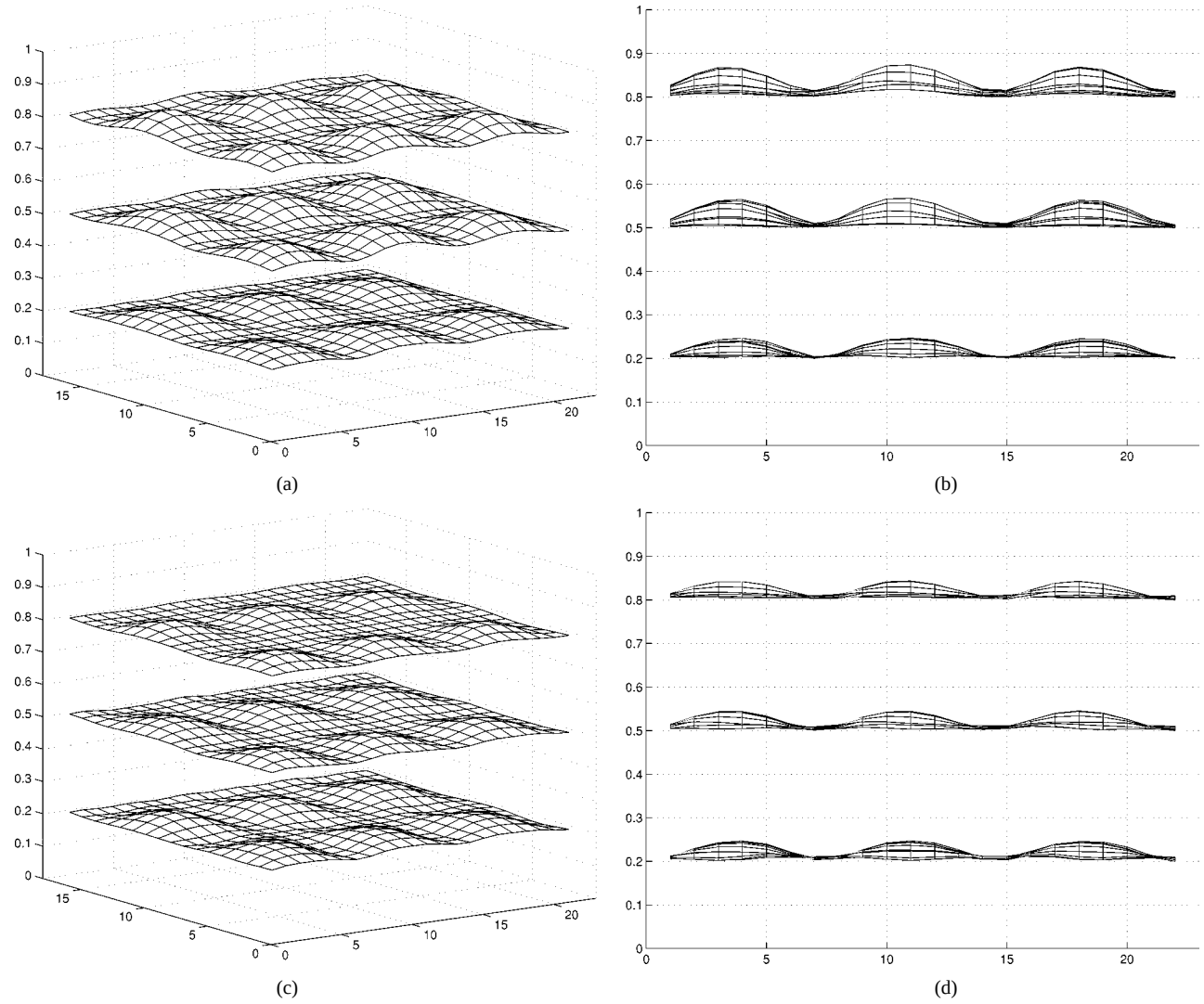


Fig. 9. Experiment for tiling multiple layers. (a) Cumulative topography $\rho_{0_{(k)}}$ for considering each layer individually. (b) Projection of (a) on to the x–z plane. (c) Result from Multiple-Ranged-Variation formulation. (d) x–z projection of (c). Unit for x and y is millimeters.

In this experiment, each of the three layer was divided into 34×44 squares with approximately 0.5-mm sides so the size of the multiranged-variation LP formulation in the second part of the experiment was $M \times N \times K = 4356$ variables with 4356 bounds and $(2 \times M \times N + 3) \times K = 26\,145$ constraints. All software and hardware were the same as those used for the single layer experiment. The runtime was approximately 8 min total to collect density information for three layers, approximately 40 min for LP calculation, and approximately 10 min per layer for local tile insertion. This LP problem has 25% less variables and 125% more constraints than the LP problem presented in the earlier section on single layer, but the CPU time increased approximately 300%. Because algorithms for LP has polynomial time complexity in term of number of variables, this indicates that time complexity for LP problems also depends on problem structures.

V. Conclusion

A solution can only be as good as the modeling. The models in Section II are theoretically sound and experimentally verified [5], [7], [8], [10]. Therefore, the two-step solution in Section III, which builds on those models, can achieve excellent results for post-CMP ILD topography reduction. As shown in Section IV, experiments on real industry design layouts produce simulation results much better than those from the rule-base method and those in [3]. The experiment for multiple layers avoided the cumulative effect by using the multilayer formulation in Section III. Future work needed includes direct wafer measurements to compare the effectiveness of current work with other approaches.

Acknowledgment

The authors would like to thank A. Dengi, E. Travis, S. Chheda, and A. Reich of Motorola for helpful discussions and encouragements, L. Lasdon of the Department of Management Science and Information Systems, University of Texas at Austin for help on optimization, and the anonymous reviewers for their comments and suggestions.

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on March 20,2025 at 08:25:55 UTC from IEEE Xplore. Restrictions apply.

REFERENCES

[1] I. Ali, S. Roy, and G. Shinn, "Chemical–mechanical polishing of interlayer dielectri: A review," *Solid State Technol.*, vol. 37, pp. 63–70, Oct. 1994.

[2] C. S. Burrus and T. W. Parks, *DFT/FFT and Convolution Algorithms: Theory and Implementation*. New York: Wiley, 1985.

[3] A. B. Kahng, G. Robins, A. Singh, and A. Zelikovsky, "Filling algorithms and analyzes for layout density control," *IEEE Trans. Compuer-Aided Design*, vol. 18, pp. 445–462, Apr. 1999.

[4] G. Nanz and L. E. Camilletti, "Modeling of chemical-mechanical polishing: A review," *IEEE Trans. Semiconduct. Manufact.*, vol. 8, pp. 382–389, Nov. 1995.

[5] D. Ouma, D. Boning, J. Chung, G. Shinn, L. Olsen, and J. Clark, "An integrated characterization and modeling methodology for CMP dielectric planarization," in *Proc. IEEE Int. Interconnect Technology Conf.*, Feb. 1998, pp. 67–69.

[6] B. Stine, D. Boning, J. Chung, L. Camilletti, F. Kruppa, E. Equi, W. Loh, S. Prasad, M. Muthukrishnan, D. Towery, M. Berman, and A. Kapoor, "The physical and electrical effects of metal-fill patterning practices for oxide chemical–mechanical polishing processes," *IEEE Trans. Electron Devices*, vol. 45, pp. 665–679, Mar. 1998.

[7] B. Stine, D. Ouma, R. Divecha, D. Boning, J. Chung, D. Hetherington, I. Ali, G. Shinn, J. Clark, O. Nakagawa, and S. Oh, "A closed-form analytical model for ILD thickness variation in CMP processes," in *Proc. International Chemical Mechanical Polishing for ULSI Multilevel Interconnection Conf.*, Santa Clara, CA, Feb. 1997, pp. 266–273.

[8] E. Travis, private communication.

[9] M. Tomozawa, "Oxide CMP mechanisms," *Solid State Technol.*, vol. 40, pp. 169–175, July 1997.

[10] T. Yu, S. Chheda, J. Ko, M. Roberton, A. Dengi, and E. Travis, "A two-dimensional low pass filter model for die-level topography variation resulting from chemical mechanical polishing of ILD films," in *Proc. Int. Electron Devices Meeting*, Washington, DC, Dec. 1999, pp. 909–912.

D. F. Wong (M'88) received the B.Sc. degree in mathematics from the University of Toronto, Toronto, ON, Canada, and the M.S. degree in mathematics and the Ph.D. degree in computer science from the University of Illinois, Urbana-Champaign, in 1987.

He is currently a Professor of Computer Sciences at the University of Texas, Austin. His main research interest is computer-aided design of VLSI. He has authored or coauthored over 200 technical papers and the book *Simulated Annealing for VLSI Design* (Norwell, MA: Kluwer, 1988) in addition to two invited articles in the Wiley Encyclopedia of Electrical and Engineering (1999). He has supervised 22 Ph.D. dissertations. He is on the Editorial Boards of *ACM Transactions on Design Automation of Electronic Systems* and the *International Journal on Applied Mathematics*. He was the Technical Program Chair and the General Chair of the 1998 and 1999 ACM International Symposium on Physical Design. He also has served on the technical program committees of many other VLSI conferences.

Dr. Wong received the 2000 IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS Best Paper Award. He also received Best Papers Awards at the 1986 Design Automation Conference and the 1995 International Conference on Computer Design for his work on floorplan design and FPGA routing, respectively. Dr. Wong has served as a Guest Editor of three special issues of IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS.

Ruiqi Tian received the B.S. degree in computer sciences and the B.S. and M.A. degrees in physics from the University of Texas, Austin. He is working toward the Ph.D. degree in computer sciences at the same university.

He is currently with Motorola Inc., Austin, TX. His current research interest is computer-aided design for VLSI, especially in the area of design for manufacturability.

Robert Boone (A'96–S'97–M'99) received the B.S. degree in mathematics from the University of Texas, San Antonio, in 1988 and the M.A. degree in gemanic languages from the University of Texas, Austin, in 1991.

Before joining Motorola, he was with Lockheed Martin as a Software and Systems Engineer. He is currently the Manager of Advanced Computer-Aided Design Group at Motorola's Digital DNA Laboratories, Austin, TX.

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on March 20,2025 at 08:25:55 UTC from IEEE Xplore. Restrictions apply.