

# CMP Aware Shuttle Mask Floorplanning

Gang Xu

CS Department  
University of Texas at  
Austin  
Austin, TX 78712, USA  
xugang@cs.utexas.edu  
Tel: 1-512-471-9588

Ruiqi Tian

Freescale Semiconductor  
Austin, TX 78721, USA  
ruiqi.tian@freescale.com  
Tel: 1-512-933-7511

David Z. Pan

ECE Department  
University of Texas at  
Austin  
Austin, TX 78712  
dpan@ece.utexas.edu  
Tel: 1-512-471-1436  
Fax: 1-512-471-8967

Martin D.F. Wong

ECE Department  
University of Illinois at  
Urbana Champaign  
Urbana, IL 61801, USA  
mdfwong@uiuc.edu  
Tel: 1-217-244-1729

**Abstract - By putting different chips on the same mask, shuttle mask (or multiple project wafer) provides an economical solution for low volume designs and design prototypes to share the rising mask cost. A challenging floorplanning problem is to optimally pack these chips according to objectives and constraints related to cost and manufacturability. In this paper, we study the problem of CMP aware shuttle mask floorplanning, which is formulated as a rectangle packing problem with objectives of area and post-CMP topography variation minimization. We propose a 3-step procedure to solve the problem. First, we use the low-pass filter oxide CMP model to guide the simulated annealing search to minimize the topography variation. The result is then further improved by sliding each chip in its enclosing rectangle. Finally, we calculate the optimal amount of dummy feature needed with a linear programming method. Our experiment shows excellent results on real industry data.**

## I. Introduction

Aggressive scaling-down of VLSI feature size has led to new challenges to VLSI manufacturing among which sub-wavelength lithography is the most difficult. Advanced resolution enhancement technologies (RET) such as optical proximity correction (OPC) and phase shift mask (PSM) are widely used to solve the sub-wavelength lithography problem [1]. Unfortunately, RET dramatically increases the mask cost: nowadays the mask cost has soared and reached 1 million US dollars at 130-nm node, and 2 million per set at 90-nm node because of fine mask features required by these RET technologies. Particularly, for a low product volume design, for example, an ASIC prototype, such high cost is unfavorable and sometimes even unaffordable because the cost is impossible to amortize over the volume.

Shuttle mask, also known as multi-project wafer (MPW), provides an economical solution for low volume designs by putting different chips on the same mask. For example, in a simple mask cost model where each design is charged based on the area it occupies on the mask, the mask cost will be halved for each design if the mask is shared by two designs equally.

Of course in reality we have to consider the overhead such as the growth of data files, the increased complexity of the mask, and extra time or expense introduced by cutting

different chips from wafers. Nevertheless, the overhead is still much lower than the total costs of multiple mask sets, and thus can be easily compensated. Because of its cost advantage, shuttle mask service begins to proliferate.

It naturally follows a floorplanning problem of how to optimally pack different chips on the shuttle mask, which has drawn EDA community's attention recently. New objectives and constraints related to cost and manufacturability distinguish this shuttle mask floorplanning problem from the classical floorplanning problem in VLSI design, and make it more interesting and challenging. In the literature there have been a few papers studying the problem of shuttle mask floorplanning with different objectives and constraints such as die-to-die inspection constraint and wafer utilization objectives [2, 3, 4, 5].

However, none of the existing works on shuttle mask floorplanning considers manufacturability which forms our motivation. Among those factors impacting the VLSI circuit manufacturability, chemical-mechanical polishing (CMP) for oxide planarization is one of the most important, because the shallow-trench isolation (STI) process is now dominant in the deep sub-micron (DSM) regime. The STI process is important because the most challenging step of gate patterning in photolithography is immediately after STI. Typically, the most aggressive RETs are done on the gate layer to improve patterning. Hence, a minimum topography variation in the STI step will provide larger process margin by not consuming too much of the already minuscule depth of focus in gate patterning. Therefore, one of our primary objectives in shuttle mask floorplanning is to minimize post-CMP topography variation at STI.

Specifically, in this paper we study the problem of CMP aware shuttle mask floorplanning by predicting the post-CMP effect during floorplanning evaluation. To our best knowledge, we present the first study on this topic. We also consider area minimization, but not wafer utilization, because mask cost is dominant compared to wafer cost. In addition, our algorithm can handle the die-to-die inspection constraint, because the merge method in [3] is easily incorporated by treating two instances of the same block as one super block in our algorithm.

Our problem is formulated as a rectangle packing problem with area and topography variation minimization while meeting the constraint of die-to-die inspection. The objective function is a weighted combination of the area and the post-CMP topography variation of the floorplan. We also propose and implement a 3-step procedure to solve the problem. Our experiment shows excellent results on real

industry data.

The paper is organized as follows. Section 2 reviews the previous work on oxide topography minimization. Section 3 presents the technical details of the 3-step procedure. Section 4 demonstrates the experimental results. The conclusion and future work are in the last section.

## II. Topography Variation Minimization

Post-CMP oxide topography variation is closely related to feature density of the circuit layout [6, 7]. Several models were proposed to capture the correlation between topography variation and feature density, among which Ouma et al's 2-D low-pass filter model is inexpensive to compute, easy to calibrate, and reasonably accurate [8]. Therefore, this model is well accepted and widely used to estimate the oxide topography variation after CMP. In the 2-D low-pass filter model, oxide thickness  $z$  at location  $(x,y)$  satisfies the following equation:

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

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.

By discretizing the layout into grids of small squares called cells, the effective density can be calculated from the feature density of the layout using the following equation:

$$\rho_0(i, j) = IDFT[DFT[d(i, j) \cdot DFT[f(i, j)]]] \quad (2)$$

Tian et al [9] gave the following approximation of  $f(x,y)$ :

$$f(x, y) \approx c_0 \exp[c_1(x^2 + y^2)^{c_2}] \quad (3)$$

where constants  $c_0, c_1$  and  $c_2$  are calibrated for each specific process.

Topography variation can be reduced by inserting dummy features into the layout to change the feature density. Tian et al [9] rewrote Eq. (2) as a convolution:

$$\rho_0(i, j) = \sum_{i'=-L}^{i+L} \sum_{j'=-L}^{j+L} [(x_{i',j'} + x_{i',j'}^0) \cdot f(i' - i, j' - j)] \quad (4)$$

where  $x_{i',j'}$  is the variable representing the amount of dummy feature to be inserted, and  $x_{i',j'}^0$  is the feature density of cell  $(i, j)$ . They also presented a simple LP formulation to describe the problem of topography variation minimization, as:

$$\text{Minimize} \quad \rho^H - \rho^L \quad (5)$$

$$\text{subject to} \quad 0 \leq \rho^L \leq \rho_0(i, j) \leq \rho^H \leq 1$$

$$0 \leq x_{ij} \leq x_{ij}^a$$

where  $\rho^H$  and  $\rho^L$  are auxiliary variables and  $x_{ij}^a$  is the maximum capacity for dummy features at cell  $(i,j)$ .

In practice, the total amount of dummy feature inserted is also an important concern, because the smaller amount usually leads to higher polish rate and less impact on users' design. [9] also gave the following ranged-variation formulation which can be applied to the case in which less amount of dummy feature is preferred and near optimal variation is acceptable. The formulation is:

$$\text{Minimize} \quad \sum_{i,j} x_{ij} \quad (6)$$

$$0 \leq \rho^L \leq \rho_0(i, j) \leq \rho^H \leq 1$$

$$\text{subject to} \quad \rho^H - \rho^L \leq \varepsilon$$

$$0 \leq x_{ij} \leq x_{ij}^a$$

where  $\varepsilon$  is the variation budget parameter which describes how much variation can be afforded in order to get the minimum dummy fill. Obviously the budget must be larger than the solution to (5).

The oxide CMP model was later extended to model the shallow trench isolation (STI) process [10]. The STI model is more complex than the oxide CMP model, as it requires modeling of dual-material polish and local pad compression to be accurate. Thus, nonlinear programming formulations and iterative methods were proposed to minimize topography variation with dummy features [10]. Several improved versions of the LP method, as well as greedy and Monte-Carlo methods were introduced in [11] to improve solutions for the oxide and STI models.

Recently, Beckage et al. [12] provided an excellent engineering solution to the dummy fill problem for STI. Their solution treats the two stages in STI CMP separately with background and regional dummy fills by taking advantage of the oxide fill characteristics before CMP. With background dummy fill providing mostly nitride density only, the dummy fill problem for STI becomes an oxide CMP problem again, which can be solved optimally with LP as described above.

## III. The Algorithm of CMP Aware Shuttle Mask Floorplanning

Based on the work in Section II, the accurate correlation model between feature density of VLSI layout and the topography variation after CMP for STI has been established as a low-pass filter model. In addition, the problem of optimal dummy feature insertion has been formulated and solved by the linear programming method. However, the linear programming method only works in the case of fixed layout. Chips may be packed on the shuttle mask in different ways to form different floorplans. Therefore the feature density and capacity distributions of the shuttle mask may vary that will affect post-CMP topography variation. The impact is not straightforward because of the complexity of the low-pass model. It is interesting and challenging to determine where each chip should be placed. The floorplanning algorithm must be CMP aware so as to guarantee the best manufacturability of the final floorplan.

### A. Algorithm Overview

We take the following strategies in our CMP aware floorplanning algorithm.

- We discretize the input chip design and extract its density and capacity distribution information to follow the low-pass model for CMP. A chip design is represented as an  $m \times n$  density matrix and an  $m \times n$  capacity matrix. A shuttle mask floorplan is now represented as a  $p \times q$  density matrix and a  $p \times q$  capacity matrix which contains sub-matrices corresponding to chips on the mask. White space is zero in the density matrix and one in the capacity

matrix. Fig 1 shows such an example.

- We choose the slicing floorplan to represent a shuttle mask. Slicing floorplans have a simple and nice binary rooted tree representation and a smaller solution space. Although a slicing floorplan is usually not as compact as a non-slicing one for the same input set, the result is still good enough.
- We use simulated annealing search to iteratively improve the result, as it worked well for previous floorplanning problems in most cases, if not always.

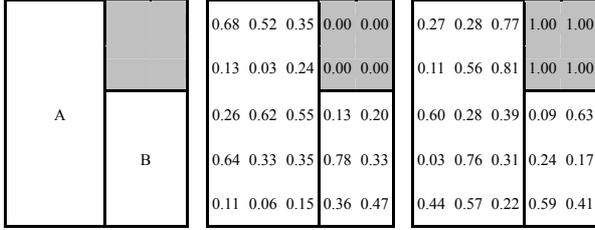


Figure 1 A floorplan and its density and capacity matrices

The CMP aware floorplanning algorithm is a 3-step procedure. First, we use the low pass filter model to guide the floorplanner to minimize the topography variation. Specifically, at each SA search move, the slicing tree is realized to its minimum area floorplan. For this floorplan, a cost function predicting the optimality of topography variation is evaluated. Notice that we cannot call LP in the SA search because of the high computation expense of the LP method. A fast predictive function is necessary instead. Second, when the SA search stops, the best result found by SA search will be further improved by sliding and rotating the chips, shown in Fig 2. Finally, we call the LP method to get the optimal amount of dummy features to be inserted. Since LP method is called only once, its computation expense is acceptable in this step. A pseudo code describing the algorithm is in Fig 3.

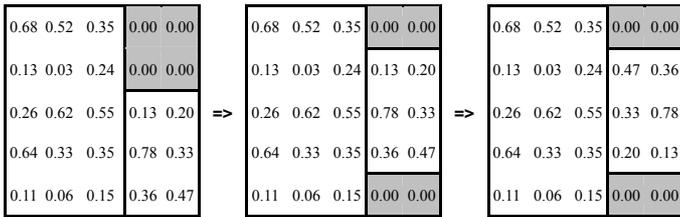


Figure 2: Slide B up, then rotate it by 180 degree

```

Begin
  x = InitialSolution;
  SA search with cost function f(x);
  SlidingAndRotation(x);
  DummyInsertion(x);
  output the best solution x;
end.

```

Figure 3 The 3-step procedure to find the optimal solution

## B. Predictive Functions

The cost function in our simulated annealing is a weighted sum of area and a predictive function. We develop

three functions to predict the topography variation in the SA search: *MaxDiff*, *SDH*, and *NSDH*. For these three functions, the less the value, the better the variation. In the following we use these notations:

$D^0 = (d_{i,j}^0)$ : the feature density matrix without dummy insertion.

$P^0 = (\rho_{i,j}^0)$ : the effective density matrix without dummy insertion, which is derived from the above feature density matrix according to Eq (2).

$C = (c_{i,j})$ : the capacity matrix.

The function *MaxDiff* is defined as:

$$MaxDiff = \max\{\rho_{i,j}^0\} - \min\{\rho_{i,j}^0\} \quad (7)$$

This function represents the maximal difference between the effective densities of cells in the floorplan. By using *MaxDiff* function, we actually use the topography variation before the dummy feature insertion to predict the topography variation after the dummy feature insertion. This function is necessary when C is a sparse matrix, which corresponds to the case that chips on the mask have strong restriction on dummy insertion. For example, sensitive circuits hand crafted by designers, like analog circuits, forbid automatic dummy insertion in the mask floorplanning stage after circuit tape-out.

The prediction of *MaxDiff* is not very reliable because it ignores the dummy feature insertion. The second function *SDH*, representing "sigma delta height", is proposed to improve the prediction. It is defined as:

$$SDH = \sum (1 - c_{i,j})(\rho_{i,j}^0 - \min\{\rho_{i,j}^0\}) \quad (8)$$

The definition of *SDH* is based on the following considerations:

- We expect a cell with large variation to have large capacity, which means more flexibility to adjust its feature density.
- We expect the total weighted variation to be small, which suggests the current floorplan is more flat.

We also consider the case in which we have variation budget and want minimum dummy to be inserted. According to (4), the effective density at cell (i,j) is most impacted by the feature density at cell (i,j). Therefore, to achieve the minimum dummy fill objective, a natural idea is to add dummy features directly to the cells which have low effective density as much as possible. High capacity is thus preferred at the cells with low effective density. In addition, large white space is not preferred, because the white space cell also needs to be filled. More white space cells may indicate more dummy features to be inserted.

Therefore, we modify *SDH* to get the third function *NSDH*, which stands for "new sigma delta height". It is defined as:

$$NSDH = \sum (2 - c_{i,j}) [1 + (\rho_{i,j}^0 - \min\{\rho_{i,j}^0\}) / (\max\{\rho_{i,j}^0\} - \min\{\rho_{i,j}^0\})] \quad (9)$$

The motivation is that *SDH* is not ideal to predict the direction of less white space. This is because the capacity of white space cell is 1, and thus does not contribute to the function value. Also, *SDH* is not ideal for cells with minimum effective density for the same reason. In addition, we normalize the variation of each cell related to the minimum effective density in order to make a fair comparison between different floorplans. Without

normalization the function may lead the search to minimum variation objective, instead of the minimum dummy fill objective that we desire.

#### IV. Experimental results

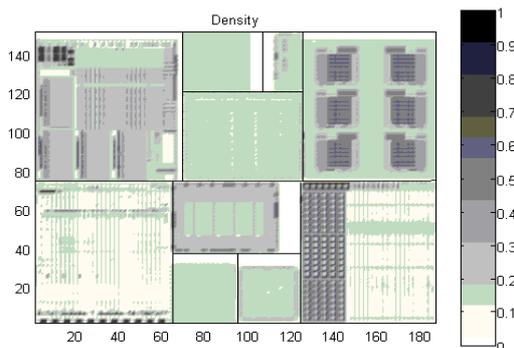
We implement a CMP aware floorplanner based on Wong-Liu floorplanner [13]. The code is written in C for efficiency and flexibility consideration. We use FFTW3.0.1 to compute Fourier transformation. In the final step, we use CPLEX as the LP solver. The code runs on a Pentium-4 Linux workstation with a P4 2.4G Hz CPU and 1G DRAM. We test a data set from a real industry mask for the 90nm technology node which consists of 10 chips. We use the typical industrial process parameter as reported in [10].

Table 1 shows the comparison among different cost functions. WS represents the white space rate, VwoD represents minimum variation without dummy insertion. The unit of the variation is angstrom. VwithD represents the minimum variation with dummy insertion, which is the topography variation from solving LP with minimum variation objective (Eq. (5)). DAmount represents the minimum dummy fill amount obtained by solving the LP with minimum fill objective, i.e., Eq. (6). The value unit does not matter. The variation budget in the LP is obtained by rounding the minimum topography variation in the previous column to the next 10's, e.g., in the case of area+SDH, 64 is rounded to 70 to form the minimum dummy fill problem.

As we can see, the predictive functions serve well in variation optimization and minimum dummy fill. The variation is improved by around 30% in all the three functions. With the same amount of dummy feature insertion, area+NSDH obtains a little larger variation than the results of the area+SDH. However, this function obtains the minimum white space as we expect. If we consider all three metrics of area, topography variation, and amount of dummy feature insertion, area+NSDH performs the best. Fig 4 shows the floorplan obtained by area + NSDH.

**Table 1 Comparison among different cost functions.**

Function	WS	VwoD	VwithD	DAmount
Area only	2.82%	818	92	340
Area+MaxDiff	6.87%	612	67	338
Area+SDH	8.27%	588	64	298
Area+NSDH	6.04%	751	67	298



**Fig 4 A shuttle mask floorplan by area+NSDH**

#### V. Conclusions and Future Work

In this paper, we propose a novel problem formulation of CMP aware shuttle mask floorplanning and present an effective 3-step procedure to solve this problem. The experimental results on real industry shuttle mask data set show a 30% reduction in the optimal topography variation.

Currently our approach focuses on the topography variation minimization of the active layer, the most critical layer, because its planarity requirement is much more stringent than the metal layers. Without a flat active layer the yield will be very low. Therefore it is necessary to consider the active layer first before considering metal layers.

Extensions of this work may include: addition of metal layers, faster or more accurate predictive functions, etc.

#### References

- [1] Scheffer, L. Physical CAD changes to incorporate design for lithography and manufacturability. *Proc of ASPDAC* (Jan 2004).
- [2] Chen, S. and Lynn, E. C. Effective placement of chips on a shuttle mask. *Proc of SPIE, 5130* (2003), 681-688.
- [3] Xu, G., Tian, R., Wong, D. F., and Reich, A. Shuttle mask floorplanning. *Proc of SPIE, 5256* (2003), 185-194.
- [4] Andersson, M., Gudmundsson, J., and Levcopoulos, C. Chips on wafer. *Proc. of Workshop on Algorithms and Data Structures* (2003).
- [5] Kahng, A. B., Mandoiu, I. I., Wang, Q., Xu, X., and Zelikovsky, A. Multi-project reticle floorplanning and wafer dicing. *Proc. of ISPD* (2004).
- [6] Prasad, S., Loh, W., Kapoor, A., Chang, E., Stine, B., Boning, D., and Chung, J. Statistical metrology for characterizing CMP processes. *Microelectron. Eng., 33* (1997), 231-240.
- [7] Stine, B. E., Ouma, D. O., Divecha, R. R., Boning, D. S., Chung, J. E., Hetherington, D. L., Harwood, C. R., Nakagawa, O. S., and Oh, S.-Y. Rapid characterization and modeling of pattern-dependent variation in chemical-mechanical polishing. *IEEE Trans. Semiconduct. Manufact., 11* (1998), 129-140.
- [8] Ouma, D., Boning, D., Chung, J., Shinn, G., Olsen, L., and Clark, J. An integrated characterization and modeling methodology for CMP dielectric planarization. *Proc. Int. Interconnect Technology Conf.* (June 1998), 67-69.
- [9] Tian, R., Wong, D. F., and Boone, R. Model-based dummy feature placement for oxide chemical-mechanical polishing manufacturability. *TCAD, 20* (July 2001), 902-910.
- [10] Tian, R., Tang, X., and Wong, D. F. Dummy feature placement for chemical-mechanical polishing uniformity in a shallow trench isolation process. *TCAD 21*, (Jan 2002), 63-71.
- [11] Chen, Y., A.B. Kahng, G. Robins, and A. Zelikovsky. Area fill synthesis for uniform layout density. *TCAD 21* (Oct 2002), 1132-1147.
- [12] Beckage, P., T. Brown, R. Tian, A. Phillips, C. Thomas, and E. Travis. Implementation of model-based tiling at STI CMP for 90nm technology. *Proc. 9th CMP-MIC* (Feb. 2004), 157-162.
- [13] Wong, D.F., and Liu, C.L. A new algorithm for floorplan design. *Proc of the 23rd ACM/IEEE Design Automation Conference* (1986), 101-107.