In this paper, we present a Full-chip CMP simulation system. We discuss three problems in practical use of CMP simulation system: how to handle huge chip data, ECP model accuracy, and how to predict the errors effectively. We propose solutions to the problems as follows: First, we develop a data extraction tool from GDSII. Dummy fill insertion function of this tool can reduce the data size of GDSII considerably by removing dummy fill information. Secondly, we propose a refined ECP model for improving accuracy of simulation. Finally, we propose a new method of simulation to predict the errors in the presence of process parameter variations. The variations should be considered, because some of them vary widely and have bad impact on the final chip surface topography.

Experimental results show that our tool can extract information for ECP and CMP simulation from large size GDSII data, which a commercial tool may not be able to read. Some data sets that are too large even for our tool to extract from are reduced in size by reducing dummy fill information. Our tool can extract information from these data sets with dummy filling function. The results also show that the new ECP model we proposed has an error of less than 20nm. Our new simulation method can even find CMP process errors that the old method cannot.

Keywords: Planarization, CMP, ECP, simulation, DFM, DFY.

1. Introduction

Copper (Cu) interconnect was widely applied at 130nm and below, because of its smaller resistance as compared to aluminium. Copper interconnect, however, introduced dishing, dielectric erosion and thickness variation problems (Fig. 1) caused by Chemical Mechanical Polishing (CMP) during planarization. In the aluminium process, interconnect is patterned by dry etch process. On the other hand, in the copper process (also called the damascene process), interconnect trenches and via holes are etched after oxide material deposition. Then, the thin barrier metal layer, which facilitates copper film generation, is deposited as a seed layer. Next, copper is deposited to fill up trenches and holes on the whole wafer surface by Electro-Chemical Plating (ECP) process. Finally, copper outside the trenches is removed to generate interconnect patterns. CMP process is the technique to remove redundant copper and to planarize the surface of wafer. ECP and CMP processes are very sensitive to layout patterns. This is because the amount of copper deposited and removed depends on the patterns, and the pattern variations cause dishing, erosion and thickness variations that lead to interconnect resistance and capacitance variations, and to photolithographic problems [1, 2]. In deep submicron processes, depth of focus (DOF) requirements become stricter and chip surface variations make shapes of copper interconnects ambiguous.

To account for the dishing, erosion and thickness variations, dummy fill insertion is one of the solutions for designers and manufacturers [3]. Dummy fill insertion places dummy features that reduce the amount of dishing and erosion by uniformizing the pattern density, without affecting the
logic function of the circuit. However, it cannot always achieve the target density. Moreover, pattern density is not a sufficient metric for accurate prediction of thickness variations [4].

CMP simulation is an effective method to predict chip non-uniformity. It can predict dishing, erosion and thickness variations based on ECP and CMP process models before manufacturing. The advantage of model-based simulation is that it can be used for layout optimization for wafer planarity without measuring additional test chips to increase the chip uniformity and to improve timing yield.

Experimental data GDSII data extraction Calibrpekt input data ECP simulation CMP simulation

![Figure 2 CMP simulation system flow](image)

To simulate the final chip surface variations, accuracy of ECP and CMP process models is very important. Some CMP and ECP models considering pattern dependencies were proposed [5-7]. However, process model is not the only issue to be considered to make CMP simulator practical. The main contributions of this paper are as follows:

1. It provides a fast extraction technique to obtain dummy fill information. The size of GDSII containing dummy fill is so large that the extraction process from GDSII usually becomes a bottleneck in terms of the CPU time.
2. It points out deficiencies in the current ECP model and proposes a refined model.
3. We underscore information that is important for process designers and discuss features that should be incorporated in a CMP simulation system.

The rest of the paper is organized as follows: In Section 2 we provide the overview of a typical CMP simulation system. In Section 3 we highlight problems in the practical use of such a system. Our proposed solutions for these problems are described in Section 4. Section 5 presents the results on prediction of the final surface of some chips by our proposed CMP simulation system. Finally, Section 6 concludes this paper.

2. Overview of CMP simulation system

Fig. 2 shows a general flow of CMP simulation system. Each step of the flow is discussed in this section.

2.1. Extraction

Copper density and line width of interconnect are extracted as parameters for ECP and CMP process models, because accuracy of the models heavily depends on these parameters. Since it is too time-consuming to calculate the height of each interconnect accurately in the entire chip, usually the chip area is divided into small meshes (around 20um to 40um square), and the average copper density and average line width are calculated in each mesh. The extraction step itself is time-consuming if we use the GDSII data file format as it is, because the data size is often too large. Moreover, its data structure is not suitable for extracting layout information in each mesh. We will discuss this issue in Section 3.
2.2. ECP simulation

ECP process should be simulated to accurately calculate copper thickness variation on the whole chip before CMP process, because the ECP thickness variations strongly affect the final surface topography. In [5], Park proposed an ECP model that uses two parameters: line width $W_L$ and line space $W_S$. In [7], Luo et al. proposed a refined ECP model considering physical mechanism. Both models need to be calibrated with the data of ECP experiments to meet accuracy requirements. This will be discussed in Section 2.5.

2.3. CMP simulation

CMP process is simulated with the input of the layout pattern parameters (as explained in Section 2.1) and the results of the ECP simulation. In [6], Tugbawa proposed a CMP simulation model that consists of two phases in calculation: the global and local processes. During CMP simulation, the global and local processes are applied iteratively at each time step. First, the global process calculates the distribution of pressure at each mesh. Each mesh has two height values: the height of the top surface and the height of the bottom surface (Fig. 3). Next, the local process calculates the removal rates of both surfaces under the pressure distribution computed earlier. Then, the local process updates the heights of both surfaces based on the obtained removal rates of each mesh.

To build an appropriate CMP model, several control parameters, such as layout pattern, polishing time, slurry, shape of pad, pressure, temperature, etc., are required. It is, however, impractical to consider all these parameters, since the run-time can be exorbitant. Therefore, the CMP model should consist of only those parameters to which the results of the CMP process are most sensitive. The other parameters are fixed in calibration, and the CMP simulation is executed assuming they are fixed. In future, if the fixed parameters need to be changed to cope with the new process conditions, another calibration would be required for model parameter fitting.

2.4. Layout modification

After ECP and CMP simulations, the chip surface data is analyzed to check for hot-spots. A hot-spot is defined as the point which may degrade timing / manufacturing yield. Thickness range of the whole chip, copper density variation, and line width variation are usually used to identify hot-spots. When hot-spot errors are found after ECP and CMP simulations, the layout patterns and dummy fills should be modified to obtain a target surface planarity. To identify hot-spots efficiently, viewer tools are required to view the chip surface and analyze the results. 2D /3D profiles help us to identify the problems in the chip layout. Also, a what-if analysis, which checks the effect of the layout modifications without re-generating GDSII data files, is very helpful for designers and manufactures.

2.5. Calibration

ECP and CMP models require calibration, because the mechanism of these processes is too complex to make accurate simulation models. Calibration adjusts model parameters to minimize the errors between simulation results and experimental data. To obtain accurate models for practical use, test chips should contain various layout patterns. Our test chips include line arrays with features in the range from 0.14um to 25um, density in the range from 5% to 90%, module size from 30um to 500um, and isolated lines and pad modules with various sizes. Usually CMP process consists of multiple steps, with each step having a different process condition. Therefore, the experimental data for calibration is measured before each step, after each step, and halfway during each step.

3. Problems of current CMP simulation system

To make a CMP simulation system practical, the following issues must be addressed.

A. Handling the huge chip data

Some chips are larger than 20mm square in area and have more than 10 interconnect layers. Since a huge number of copper rectangles are additionally inserted as dummy fill patterns, the size of the GDSII data that contains this dummy fill information can be very large (e.g., over 20GB). Extraction process is one of the most time-consuming processes with such large data. To do a what-if analysis (as discussed in Section 2.4), only dummy fill information should be updated without
modifying the GDSII data file. Fast extraction from GDSII data files is an essential requirement to improve the productivities of designers and manufacturers.

**B. The accuracy of the ECP model**

We evaluated the ECP model proposed in [7] and the CMP model proposed in [6] in our CMP simulation system. The CMP model showed good results of chip surface topography, but we found that the ECP model has large errors, which have a bad impact on the final chip surface in some layout patterns. This is because the ECP model is too compact to express all possible corner conditions. The ECP model must be improved to handle the practical chip data.

**C. Effective methods for analysis of the results**

It is important to find hot-spots and to avoid errors in the chip surface analysis phase. But there are few cases to find errors in the data without breaking design rules under the typical condition because proper rules and ECP/CMP process recipe are usually established. Most errors occur only in specific regions of wafer because die-to-die variations have great impact on final chip surface variation. We should develop effective methods to predict these errors.

**4. Proposed approach**

**4.1. Chip data extraction**

The importance of input file extraction from GDSII data was discussed in Sections 2 and 3. Copper density and perimeter of copper line of each mesh are required to be extracted in our system to calculate copper line width and line space. They can be extracted with commercial design rule checker (DRC) tools, but we decided to develop our own extraction tool for the following two reasons. The first reason is the performance in speed and the capacity in memory usage. Since commercial DRC tools have multiple functions and their main purpose is to check design rule violations in chip layout, they consume excessive CPU time and memory. Therefore, we needed a fast and compact extraction tool.

The other is the need for an incremental dummy fill insertion function. Since dummy fill occupies large amount of GDSII data, the data size can be reduced if the extraction tool itself can insert dummy fill into the extracted data for simulation without inserting the dummy fill into the GDSII file.

After analyzing the chip surface variations, generally dummy fill patterns are to be refined instead of interconnect layout, because dummy fill modification has less impact on chip performance. The incremental dummy fill insertion function enables designers and manufacturers to carry out what-if dummy fill analysis easily and quickly.

We now briefly describe the algorithm of our extraction tool to realize the incremental function. First, each copper object is registered in all the meshes where the object lies. Then, in each mesh, copper area and perimeter length are calculated with a line sweep method. The boundaries of each copper object are extended by the dummy-to-object spacing size defined in the design rules for counting dummy inhibit area, and the total blank area in the mesh is calculated as a dummy filling area. The dummy filling area, illustrated in Fig.4, is considered to be the total area where dummy fill can be placed. Copper area and perimeter length per unit area of the dummy fill pattern are used...
to calculate copper area and perimeter length of the virtually inserted dummy fill. They will be saved in addition to the original copper area and perimeter length calculated without the dummy fills in each mesh. This method causes some errors because the shapes of the dummy fills and the dummy filling area are not considered. However, the errors are too small to impact the final chip surface. When tuning the values of the copper area and perimeter length for layout optimization, we believe that it would be faster and easier not to handle the dummy fill shape and the dummy filling area as they are, and to control the values directly.

4.2. Refined ECP model

The wafer is coated with copper in chemical solution in ECP process. To achieve void-free copper film, three additives, accelerators, suppressors, and levellers, were introduced [8]. Reid et al. [9] proposed an ECP model that explains additive’s behaviour. In the ECP model in [7], which is based on Reid’s theory, total volume of copper deposition is regarded proportional to the amount of accelerators. Since accelerators adhere to the whole chip surface, total volume is proportional to the surface area. In each mesh of the chip, surface area consists of top / bottom surfaces and sidewall of trenches. Therefore copper volume $V$ can be formulated as

$$V = H_0(\ell L + D^2),$$

(1)

where $H_0$ is the copper thickness of the mesh that has no interconnect, $L$ is the perimeter length of interconnects in the mesh, and $D$ is the mesh size. The effective trench height $T_e$ is smaller than trench height $T$, because the amount of accelerators on the sidewall may be smaller than that on top/bottom surfaces.

According to the layout pattern, there are three cases of copper deposition topographies (Fig.5). In each case, copper thickness $H$ and step height $S$ are to be computed. $H$ is defined as the copper film thickness above the oxide area, and $S$ is defined as the gap between the height of copper above the oxide area and the height of copper above the trench. $S$ is positive when the height of copper above the oxide is larger than that above the trench (Fig.5 (A)).

Case A: Step height $S$ is greater than 0 and accelerators outside the trench shrink trench width by $\delta_s$. Then, the volume of copper is

$$V = HD^2 - SD^2 \delta_s + TD^2 \rho, \quad (2)$$

where $\rho$ is the area of interconnects, $\delta_s$ is the shrunk density after ECP process, and $H$ is the copper thickness of the mesh. Since the shrinks of the trench only depend on the accelerators outside the trench,

$$H_0 D^2 (1-\rho) = HD^2 (1-\delta_s), \quad (3)$$

From equations (1), (2), and (3), $H$ and $S$ are obtained.

Case B: Step height is smaller than 0 and accelerators in the trench expand trench width by $\delta_e$. Then, the volume of copper is

$$V = HD^2 - SD^2 \rho_e + TD^2 \rho, \quad (4)$$

where $\rho_e$ is the expand density after ECP process. Since the expansion of the trench only depends on the accelerators in the trench,

$$H = H_0, \quad (5)$$

From equations (1), (4), and (5), $H$ and $S$ are computed.

Case C: Step height equals to 0.

$$S = 0 \quad (6)$$

Then, the volume of copper is

$$V = HD^2 + TD^2 \rho, \quad (7)$$

From equations (1), (6), and (7), $H$ and $S$ are obtained.

With the experimental ECP data of test chip, parameter $T_e$, $\delta_s$ and $\rho_e$ are calibrated. Still, we found large errors in some layout patterns. So we increased the number of calibration parameters to calculate the ECP topography more accurately.

Fig.6 shows the surface profiles of some test modules in the test chip. In Fig.6 (1) line width $W_L$ / line space $W_S$ is 0.14um/0.56um. In this case, $S = 0$ and $T_e = 0.144T$ are calculated from experimental data. In Fig.6 (3), $W_L/W_S = 0.14um/1.26um$, $S = 0$ and $T_e = 0.143T$ are obtained. In
Fig. 6 (2)), \( \frac{W_L}{W_S} = 1.2 \mu m / 0.13 \mu m \), S takes non-zero value. Whichever type of topography in Fig. 5 this module takes, the range that \( T_e \) can take is from 0.455T to 0.619T. Therefore \( T_e \) can be defined as a function of \( W_L \). \( T_e \) takes a fixed value when \( W_L \) is larger than a specific size. Therefore, \( T_e \) is defined as

\[
T_e = \min(\frac{\gamma_1}{\gamma_1}, \frac{\gamma_2}{\gamma_2}, \frac{\gamma_3}{\gamma_3}),
\]

where \( \gamma_1, \gamma_2, \) and \( \gamma_3 \) are calibration parameters. It is natural that \( T_e \) is proportional to the line width because additives may easily adhere to the sidewall in a wide line.

\( \delta_s \) is defined as the calibration parameter proportional to line space \( L_S \) because shrinks of the trench only depend on the accelerators outside the trench. Therefore \( \delta_s \) is defined as

\[
\delta_s = \alpha_2 W_S \beta_2
\]

In the same way, \( \delta_e \) is defined as

\[
\delta_e = \alpha_3 W_L \beta_3
\]

\( \alpha_2, \alpha_3, \beta_2, \) and \( \beta_3 \) are calibration parameters. We calibrated these 7 parameters with experimental data.

4.3. Simulation method

As was discussed in Section 3, proper design rules and CMP process recipe seldom cause errors under ideal conditions of ECP and CMP processes. One of the most important purposes of CMP simulation is to check non-uniformity of the chip surface and modify it. So as not to miss hot-spots and errors under various conditions, die-to-die variations should be considered during ECP and CMP simulations. Among die-to-die variations, the variations of copper thickness after ECP process and of over-polishing time have the largest impact on the final chip surface topography.

Initial copper thickness variations after ECP process have great influence on non-uniformity of the chip surface in terms of on-chip and die-to-die variations. Large on-chip variations after ECP may cause final chip non-uniformity because initial variations are propagated through the entire CMP process (Fig. 8 (a)). However, on-chip variations can be predicted with the accurate ECP model proposed in Section 4.2. On the other hand, the difference of initial height caused by die-to-die variations changes polishing time, that has an impact on the final surface (Fig. 8 (b)). Fig. 7 shows experimental results on the intra-wafer variations and wafer-to-wafer variations of copper thickness after ECP process. We can see that the variations are large enough to be considered in the
ECP simulation. We used three values of copper thickness (thin, typical, thick) as corner conditions in simulation.

Endpoint detection is the key process to achieve planarity of the chip surface. In the copper removal step of the CMP process, a sensor in the CMP machine searches the point where barrier metal was detected during polishing. This technique adjusts polishing time properly to achieve a target height even though initial height and removal rate vary widely. After an endpoint is detected, removal process continues as over-polishing to clearly remove overburden copper. If over-polishing time is too long, removal rate of barrier metal is much smaller than that of copper so that dishing grows more serious. If over-polishing time is too short, copper residue may cause short errors in interconnect. Thus over-polishing time has a great impact on CMP results. When an endpoint is detected on a certain chip, some chips on the same wafer may already have reached the endpoint, but not others. This is because the sensor does not always detect the first endpoint of the wafer. We used three values of over-polishing time (over, typical, under) as corner conditions, and ran simulation in each case.

5. Experimental Results

Experimental results of our CMP simulation system are discussed in this section. All the experiments were performed on a 2.8GHz Opteron Linux machine with 16GB memory.

We obtained model parameters of the proposed ECP model of Section 4.2 with calibration. Fig.9 shows simulation results vs. experimental results of copper height $H$ and step height $S$. These graphs include results of line arrays with features in range from 0.14um to 25um, and density in range from 5% to 90%. The results of the model of [7] are also plotted in Fig.9. Diagonal line indicates the points where there is no difference between simulation results and experimental results. Using our model, the average RMS errors of $H$ and $S$ are 12.2nm and 18.9nm respectively, whereas the model of [7] has RMS errors 38.5nm and 38.2nm respectively. CMP model is also calibrated with experimental data.

<table>
<thead>
<tr>
<th>Table 1 Comparison with commercial tool in extraction process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chip data</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>A(15.5x16.0mm, 6.2MB)</td>
</tr>
<tr>
<td>B(4.8x3.8mm, 54MB)</td>
</tr>
<tr>
<td>C(9.9x9.9mm, 26GB)</td>
</tr>
<tr>
<td>C w/o dummy fill (77MB)</td>
</tr>
<tr>
<td>D(8.0x6.4mm, 4.7GB)</td>
</tr>
<tr>
<td>E(21.3x20.9mm, 6.0GB)</td>
</tr>
</tbody>
</table>
Input data for ECP simulation is extracted from GDSII files. We first compare our tool with a commercial tool on some real chip data. Table 1 show that our tool is less time-consuming and uses less memory than commercial tool (Cadence Assura). Our tool cannot extract input data from GDSII data C, but GDSII data C without dummy fill information can be handled with dummy fill insertion function. The maximum error of copper density is 5.4%.

After extraction, the final chip surface data is obtained by running ECP and CMP simulation. To take into account die-to-die variations, simulation is carried out under different conditions of initial copper height and over-polishing time.

In the analysis phase, chip surface corresponding to these conditions is checked with our viewer tool. Fig. 10 shows snapshots of the chip surface. The chip surface corresponding to the typical condition has small variation (Fig. 10 (A)), but the gap between top and bottom of the chip surface is increased and uncleared copper causes short errors in another condition (Fig. 10 (B)).

![Figure 10 Chip surface variation](image)

In order to avoid short errors, we modify the dummy fill layout. We extract the input file of ECP simulation without modifying GDSII data and run the simulation again. Extraction time is only 7 minutes. This is much smaller than the time for initial extraction, 389 minutes. The reason is that dummy modification is done only in the layer where errors exist. Fig. 10 (C) shows the final chip surface of the modified layout under the same condition. Chip surface variation is decreased and short errors are avoided.

### 6. Conclusion

In this paper, we presented our CMP simulation system and proposed approaches to solve the problems that arise during its practical use. A data extraction tool that can handle large-size chip data within reasonable memory size and CPU time is developed. The ECP model is improved, which results in reduced average errors in both copper height and step height. A new simulation method considering die-to-die variations is proposed. Finally we showed experimentally how to simulate and modify practical chip layouts.

### References


