electronics

MDPI

Article

A Study on the Pattern Effects of Chemical Mechanical Planarization with CNN-Based Models

Han Bao [1,2], Lan Chen [1,*] and Bowen Ren [1,2]

[1] The EDA Center, Institute of Microelectronics, Chinese Academy of Sciences, Beijing 100029, China; baohan@ime.ac.cn (H.B.); renbowen@ime.ac.cn (B.R.)

[2] School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: chenlan@ime.ac.cn; Tel.: +86-010-82995630

Received: 20 June 2020; Accepted: 14 July 2020; Published: 17 July 2020

Abstract: Chemical mechanical polishing (CMP) has become one of the most important process stages in the fabrication of advanced integrated circuits (IC). The CMP pattern effect strongly influences the planarization of the chip surface morphology after CMP, degrading the performance and the yield of the circuits. In this paper, we introduce a method to predict the post-CMP surface morphology with a convolutional neural network (CNN)-based CMP model. Then, CNN-based, density step height (DSH)-based, and common neural-network-based CMP models are built to compare the accuracy of the predictions. The test chips are designed and taped out and the predictions of the three models are compared with experimental results measured by an atomic force profiler (AFP) and scanning electron microscope (SEM). The results show that CNN-based CMP models have better accuracy by taking advantage of the CNN networks to extract features from images instead of the traditional equivalent pattern parameters. The effective planarization length (EPL) is introduced and defined to make better predictions with real-time CMP models and in dummy filling tasks. Experiments are designed to show a method to solve the EPL.

Keywords: chemical mechanical polishing (CMP); convolutional neural network (CNN); modeling; pattern effect; effective planarization length (EPL)

1. Introduction

Chemical mechanical polishing (CMP) is a physical-chemical process usually used to make locally and globally flat wafer surfaces. During the CMP process, a wafer is pressed faced down against a rotating polishing pad with a slurry and abrasives are added between the two surfaces. The redundant materials are removed via the combination of the chemical action caused by the slurry and the mechanical action caused by the rotation and pressure of the polisher [1,2]. CMP has been playing an increasingly important role in the fabrication of ultra-large-scale integrated (ULSI) circuits. The planarization of the post-CMP surface determines the depth of field (DOF), which will strongly affect the quality of the lithography during the fabrication of higher layers, while large CMP defects may cause chip failure [3,4]. The planarization of CMP faces tighter margins with the shrinking of the feature size of the transistors, because the defects in post-CMP surfaces are worse compared to the size of the transistors. CMP faces great challenges.

The planarization of post-CMP surfaces has a strong correlation with layout patterns, which are composed of transistors and interconnected lines [4]. Layout patterns cause the CMP pattern effect, which reflects how different layout patterns act on the CMP process and the final height of the post-CMP surface. By studying CMP pattern effects, we can predict defects [5] on the surfaces and

hotspots that may degrade the performance or even cause failure of the circuit. Then, the layout can be optimized for better performance and higher yields.

Currently, most CMP models are built based on pressure distribution and Preston's equation. The Massachusetts Institute of Technology (MIT) [6,7] and Microelectronics of the Chinese Academy of Sciences (IMECAS) [8,9] have built several pressure-distribution-based CMP models. These models predict the defects of post-CMP surfaces by taking advantage of the physical mechanisms of the fabrication process. The most commonly used models are based on density step height (DSH) theory. DSH models assume that the contact between the pad, the abrasives, and the chip surface fits Hooke's law [10]. The model divides the whole layout pattern into small process windows and extracts equivalent parameters from each process window. Then, Hooke's law and the linear approximation are used to calculate the removal height in each process window in a unit time period. The removal operation is iterated until the surface height reaches the desired value.

However, DSH-based CMP models have several limitations. First, the CMP process involves too many interfering factors, including the interactions among wafers, polishing pads, abrasives, and the slurry, for which some of the mechanisms are not yet well studied. Second, as they are bound by the computation speed, the process windows cannot be too small (usually 10 um $\times$ 10 um). The size of the process windows limits the resolution and accuracy of the CMP model. Third, Hooke's law and the linear approximation used in DSH models cannot accurately describe the real CMP process or cause many errors. Fourth, the deviation of the layout and the extraction of equivalent parameters lead to loss of detail regarding the layout pattern. These details can strongly influence the polishing results.

In our work, a convolutional neural network (CNN)-based post-CMP planarization model is designed. A CNN is a method based on deep learning. It learns to analyze how patterns act on outputs from labeled samples and then uses the learned knowledge to predict new unknown samples. A CNN-based CMP model extracts the features that have an impact on the post-CMP surface height and calculates how these features act on the final global surface height distribution. Our former work [11] showed a CNN-based CMP planarization model with a complete architecture. The practicality was discussed and the results showed that our CNN-based CMP model had good accuracy. In this article, test chips at the 28 nm process node are designed and fabricated by Semiconductor Manufacturing International Corporation (SMIC). An atomic force profiler (AFP) and a scanning electron microscope (SEM) are used to measure the experimental results. CNN-based, neural-network-based, and DSH-based CMP models are built to compare the predictions with the experimental results. The equivalent planarization length (EPL) is introduced to help make real-time CMP hotspot predictions. A method to solve the EPL using the CNN-based CMP model is then explained.

This paper is organized as follows. Section 2 introduces the pattern effects of the CMP process. Section 3 talks about the principles and methods used to build different CMP models, including DSH-based, neural-network-based, and CNN-based CMP models. Section 4 shows predictions of the three CMP models and compares them with experimental results. Section 5 talks about the EPL and using the CNN-based CMP model to solve the EPL. Section 6 summarizes the ideas.

2. CMP Pattern Effects

The materials on the surfaces of chips cannot be removed evenly. The removal rate is determined by the slurry, the abrasives, the pressure, the relative rotational speed between the chip and pad, and many other factors. The variance caused by CMP can be categorized into two types: random variation and system variation [3]. Random variation is made up of the fluctuations during fabrication processes [12], including random dopant fluctuations and line edge roughness. Random variation is generated from the fabrication process and is hard to mitigate via design methods. System variation is mainly caused by CMP pattern effects [13–15]. The layout pattern strongly influences the removal rate during the CMP process. CMP planarization models try to study these pattern effects to make precise predictions of the post-CMP surface height to help optimize the layout design.

There are two main reasons for CMP pattern effects. One is the original height distribution of the chip surface before polishing. Former steps such as electrochemical deposition, chemical vapor deposition, and etching form an initial height distribution, which influences the CMP removal rate. The other factor is the selectivity of the slurry used in CMP. The selectivity describes the ratio of the removal rates of two different materials. The surface forms a new height distribution during CMP, which is affected by the original height distribution and the selectivity.

CMP pattern effects have great impacts on chip fabrication. First, the unevenness of the chip surface influences the fabrication of the higher layers. Rougher surfaces require a larger depth of field (DOF), which reduces the resolution of the optical system. Second, the pattern effects also influence the performance of the chip [9]. The unevenness of the surface changes the resistance and the capacity of the circuit. For example, recessed copper lines increase the resistances of interconnects, while protruding copper lines decrease the resistances, meaning the delay of the circuit can be changed.

Works have been put forward to study CMP pattern effects [13–15]. There are theories that relate the local removal rate to the equivalent line width, line space, and density [10]. Progress has been made, however the accuracy is not high enough. One of the main reasons is that there are thousands of different patterns, which lead to different polishing results and may have the same equivalent parameters. The equivalent parameters do not have good pattern abstraction results, meaning that pattern effects cannot be precisely predicted by these theories.

To further study the CMP pattern effects and make more accurate predictions for post-CMP surface heights, CNN-based CMP planarization models were introduced. Test chips were designed and fabricated at the 28 nm process node. AFP and SEM were used to obtain experimental results. The experimental results were set as samples and used to train the network. Then, the trained network made predictions to confirm the precision. Our former work [11] showed that the CNN-based CMP model reduced the mean squared error (MSE) from 65.54 A when using the DSH model to 22.41 A. This article studies the most important reasons for the improvement of the accuracy and the additional usage of the CNN-based CMP model related to the equivalent planarization length.

3. CMP Planarization Models

CMP planarization models make predictions of the post-CMP surface height according to the input layout patterns and the process parameters. In practice, dishing and erosion defects [8] are more commonly used because the absolute height is hard to measure directly. Figure 1 shows that dishing is the height of the pattern oxide minus the height of the copper; erosion means the height of the field oxide minus the height of the pattern oxide.

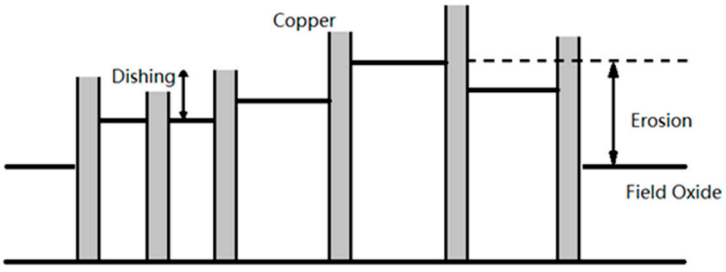

Figure 1. Illustration of dishing and erosion defects.

This section first talks about DSH-based CMP models, including dividing the layout into process windows; extracting the equivalent space, width, and density from each process window; and calculating the removal height in each process window. Then, the NN-based CMP model and the CNN-based CMP model used in our experiment are introduced.

3.1. DSH-Based CMP Models

DSH-based CMP models first divide the layout into square process windows and assume that the up and down areas of the surface bring unequal pressure distribution. The pressure is calculated for each process window based on the original height and the total force of the polishing pad. The equivalent parameters are then extracted for each process window. The pressure is calculated for each process window based on the equivalent parameters and the total force of the polishing pad according to Hooke's law. The calculated pressure is applied to obtain the material removal rate using Preston's equation. The removal height of a unit period time is calculated and the surface height is updated after each removal. The processes are repeated until the height reaches the desired value. The details for parameter extraction and removal are shown in the following paragraphs.

3.1.1. The Deviation of the Layout and the Extraction of Equivalent Parameters

The layout is usually divided into square process windows. The size of the process windows is decided by the computation speed. In our experiment, the size of the process windows was set to 10 um × 10 um and the computation time for each prediction was more than 1 h.

The equivalent parameters are then extracted [13,14]. The most commonly used equivalent parameters are the equivalent line space, line width, and density [16,17]. The steps are as follows [8]:

(1) In each window, record the perimeter C_i and the area S_i of each polygon in the window, where i is the identifier of each polygon. Substitute all patterns in the window by a single rectangle and solve Equation (1) to calculate the length Q_{il} and width Q_{iw} of the rectangle, which are the two roots of Q:

$$Q^2 - \left(\frac{C_i}{2}\right)Q + S_i = 0 \tag{1}$$

(2) The equivalent line width W of the rectangle is given by Equation (2) and named as the equivalent line width:

$$W = \frac{\sum_{i=1}^{n}(Q_{iw} \cdot S_i)}{\sum_{i=1}^{n} S_i},\ i = 1, 2, \ldots, n \tag{2}$$

(3) The equivalent density D:

$$D = \frac{\sum_{i=1}^{n} S_i}{square\ of\ window},\ i = 1, 2, \ldots, n \tag{3}$$

The parameters are calculated as the input for the removal process.

3.1.2. The Removal Process of DSH-Based CMP Models

DSH assumes that the polishing pad is an ideal plane composed of a collection of independent springs. The contact between the pad, abrasive, and surface fits Hooke's Law [10].

An improved CMP planarization model based on DSH and Greenwood-Williamson (GW) theory was studied [18,19]. The model assumes that the asperities of the polishing pad contacts the surface of the wafer following the Hertz contact theory [20]. First, the deformation of the surface is calculated according to the Hertz contact theory. Then, the feature-level pressure distribution is calculated considering the additional characteristics of asperities. The local removal rate is obtained based on Preston's equation. The removal rates for the up and down areas are given as follows [21]:

$$RR^U = K_p V P_{nom} \sqrt{1 + \frac{4R'_a(h^U - h^D)}{L^2}} \cdot exp\left(\frac{h^U - D^*}{\sigma}\right) \tag{4}$$

$$RR^D = K_p V P_{nom} \sqrt{1 + \frac{4R'_a(h^U - h^D)}{S^2}} \cdot exp\left(\frac{h^D - D^*}{\sigma}\right) \tag{5}$$

where RR^U and RR^D represent the removal rate of the up and down areas, respectively. Here, K_p denotes the Preston coefficient and V denotes the relative rotational speed between the wafer and the pad; P_{nom} and D^* relate to the material of the polishing pad; h^U and h^U are the heights of the up and down areas; R'_a is the equivalent curvature radius of the asperity; L is the line width and S is the line space. The model has good efficiency and the simulation results fit the experiment well.

3.2. CNN-Based CMP Planarization Models

3.2.1. The Practicality of CNN-Based CMP Models

The convolutional neural network is a popular deep learning algorithm. It is widely used in object detection and image classification tasks, such as facial and text recognition [22–24]. CNN is good at image analysis. A CNN-based model can extract features from image samples and verify how these features act on the output. Then, the CNN network uses its knowledge to make predictions of new samples with high accuracy.

The CNN consists of three kinds of typical layers to construct a complete network: convolutional layers, pooling layers, and fully connected layers. A convolutional layer maps a multilayer input image to a multilayer output image with a kernel. The kernel is usually a small multilayer matrix (usually 5×5). The parameters of the kernel are trained with samples. A pooling layer goes through the image with a stride and applies a pooling operation on each unit window (usually 2×2 or 3×3). A pooling operation picks the max or the average value of the unit. A fully connected layer flattens a 2D or 3D input into a 1D array and then makes linear combinations of each element.

The workflow of a CNN-based model is as follows. The model first needs an architecture of several neural layers. For each sample, the input goes through the network from the first layer to the last layer. Most of the layers are convolutional layers, which means that each element of each layer needs a convolution operation with the kernel [22,24] to pass to the next layer. The labels of the sample are then compared with the predictions of the network to update the parameters of the kernels. The process lasts until the predictions of samples of the CNN network reach the desired accuracy. More details can be found in works by [22,25] Lecun et al. [23,24] and Hinton et al.

Three important characters make CNN more accommodating to CMP models: sparse interactions, parameter sharing, and equivariant representations. Sparse interactions mean that each element of each convolutional layer only interacts with the parameters in the kernel (usually 3×3 or 5×5) of the layer. This matches the CMP mechanism, whereby the removal of a given region is only strongly influenced by neighbors. Parameter sharing means that each layer only has one kernel. Equivariant representations mean that when the input is shifted, the output will be shifted by the same operation and the same value will be kept. These two characters guarantee that the same input pattern will have the same output in a CMP planarization model.

The key ideas behind DSH- and CNN-based CMP models are different. Pressure-distribution- and DSH-based CMP models assume that the mechanism and details of CMP process are well studied. The mechanisms are used to calculate the removal rate so that the post-CMP surface height can be predicted. However, the real CMP process can be very complicated, as it involves a polishing pad, slurry, abrasives, pad conditioner, and other environmental factors. The shape of asperities on the polishing pad and the size of the abrasives are random, meaning the post-CMP height cannot be predicted precisely. On the one hand, the mechanisms of the polishing process are not very clear yet. Models based on fluid shear stress, solid-solid contact, and chemical reaction removal have been built [10,17,26], but none of these models can match the CMP process perfectly. On the other hand, the pattern effect of the CMP process is not well studied yet. Existing CMP planarization models either ignore the impact of neighbor regions during material removal or simply average the values, limiting the accuracy of the predictions.

A CNN-based CMP planarization model relies on a Bayesian equation:

$$\mathrm{P}(H_i|P) = \frac{\mathrm{P}(H_i)\mathrm{P}(\mathrm{P}|H_i)}{\sum_{j=1}^{n}\mathrm{P}\left(H_j\right)\mathrm{P}\left(P_j\middle|H_j\right)} \tag{6}$$

In Equation (6), P refers to the layout pattern and H refers to the height of the post-CMP surface. The experimental measured data are used a posteriori to estimate the prior value, which is implicit in the equation, using a maximum likelihood estimate (MLE). Then, the implicit a priori value is used to predict the new samples. Correspondingly, the CNN networks use the a posteriori samples and learns the a priori value from the a posteriori value to update the parameters. Then, the networks can predict the former knowledge.

3.2.2. The Architecture of the CNN-Based CMP Model

The final architecture of our CNN-based CMP model was designed based on AlexNet [27]. AlexNet was chosen because of its fast training speed compared with Inception- and ResNet-based networks, while the predictions of the three structures have close mean squared errors (MSE) [11]. AlexNet was originally designed to deal with image classification tasks. We redesigned the architecture of AlexNet to fit the CMP problem.

First, consider the inputs to the model. For each input graph, the CMP model needs to generate an output value of the post-CMP surface height of the middle pixel. During the CMP process, the removal rate of each position mainly relates to the surrounding areas, so the inputs of the networks are designed as square matrices centered on the target pixels. The size of the input matrices is decided by the effective planarization length (EPL) and is set as a hyperparameter of the model. The EPL describes the maximum length between two positions that interact with each other during the CMP process and is further studied in Section 5.

The test chips used in our experiments were designed using a 28 nm standard cell library. The typical line width and line space are 0.03 um–0.15 um and 0.09 um–0.48 um, respectively. The values are mostly integer multiples of 0.03 um. Therefore, we set 0.03 um × 0.03 um as a unit square. Each square is either blank or occupied, depending on whether 50% of the unit is occupied. The average height of the unit square is extracted as an element of the input matrix.

Conv0 and Pool0 were designed for data preprocessing. In our experiment, the input size of a single prediction operation was set to $1364 \times 1364 \times 1$, representing an area of 40.92 um × 40.92 um with a single layer. The Conv0 receives $1364 \times 1364 \times 1$ and implements a convolution operation with a kernel size of $5 \times 5 \times 3$ and stride 3 to output $454 \times 454 \times 3$. Then, the Pool0 changes the data into $227 \times 227 \times 3$, so that the format of the data suits the middle layer of AlexNet.

The middle part of the model adopts an AlexNet-like architecture for the network construction [11,27]. Figure 2 shows the complete network structure diagram. Here, a convolutional layer is denoted as conv, a pooling layer is denoted as pool, and a fully connected layer is denoted as FC. The network includes 6 convolutional layers and 4 pooling layers in the middle part of the network for feature extraction from the layout.

Finally, consider the output part of the model. In the CNN network with conventional image processing, the output is usually processed by a soft-max layer to obtain the probability output of the graphic classification. However, the output of the CMP model should be a height distribution, which is a scalar value. The architecture should be designed to fit a regression task [28–30].

In our CMP model, 4 fully connected layers were set following convolutional layers. The output of the last layer is a one-dimensional scalar value. The purpose of these 4 fully connected layers is to summarize the features extracted by convolutional layers and to measure the importance of each feature to the final output. The loss function for a classification task is usually cross-entropy loss to predict the probability for each class that the sample may belong to. By changing the loss function to mean squared error (MSE), the network can compare the difference between the continuous output

and the label. Using stochastic gradient descent as the optimization method, the MSE loss function is used to carry out the backpropagation to complete the network training and make predictions with continuous variables.

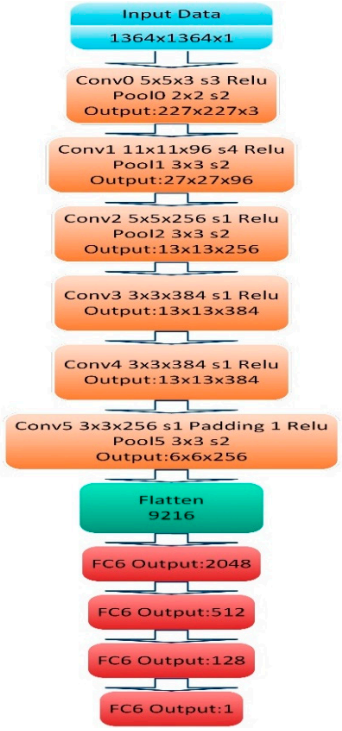

Figure 2. Complete network structure of the chemical mechanical polishing (CMP) surface height model with convolutional neural network (CNN).

3.2.3. A CMP Model Based on Equivalent Parameters and a Neural Network

As a comparison, we set up a CMP model based on the equivalent parameters and using the common neural network (NN) architecture. The grid division method of the NN-based model is the same as the DSH model. The input parameters of the model are the equivalent parameters extracted using the aforementioned method, including the equivalent line width, the equivalent line spacing, and the equivalent density. The intermediate parameters, including the total polygon number in the grid and the polygon circumference, are also used in calculating these parameters. The middle layer of the model adopts a fully connected, 4-layer neural network structure. The input has 8 parameters; the middle has three fully connected layers, all of which have 256 neurons; and the final output is a scalar output.

The common NN-based model is used to find the most important reason to improve the accuracy. If the accuracy of the NN-based model is closer to the CNN-based model, the application of the neural network is the most important factor. In contrast, if the accuracy of the NN-based model is closer to the DSH-based model, the reason for the precision difference lies in the ability to extract graphic features provided by CNN.

4. Accuracy Analysis of CMP Models

4.1. The Test Layout Patterns

The test layout patterns applied in our experiment were designed based on the 28 nm standard cell library from Semiconductor Manufacturing International Corporation (SMIC). Most of the structures of the logical units in the cells are lines or spaces, and less than 5% of the cells have gates with other structures, including crossing gates, L-shaped gates, and U-shaped gates.

The test patterns can be categorized into 3 parts. The patterns of part 1 are 200 um × 200 um squares with straight vertical gates. The gate length is set to 200 um, gate widths range from 0.03 um to 0.15 um, and the gate spaces are 0.10 um–0.50 um. The patterns in part 2 are similar to part 1, but horizontal gates are added. The width of the horizontal gates is either 0.06 um, 0.08 um, or 0.12 um, and each horizontal gate passes through 3 to 20 vertical gates. Part 3 involves 200 um × 200 um squares with straight vertical gates, whose gate length and gate width are variant within each square. The 200-um horizontal edge is divided into eight 25-um pieces, with each piece retaining a constant gate length and gate width. This part is designed to confirm the impacts of the neighbors of the pattern effect. Figure 3a shows one of the test chips and Figure 3b shows a single input image in the CNN network. The label and the CNN output of Figure 3b is the height of the middle pixel.

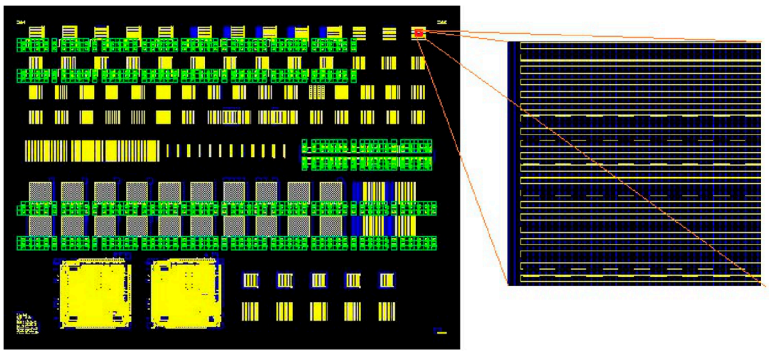

Figure 3. A test chip (**a**) and an example of a single input image (**b**).

For each test pattern, AFP and SEM were used to measure the dishing defect of the center position. For the DSH-based CMP model, equivalent parameters of the test patterns were extracted and used to make predictions. For the CNN-based CMP model, the test pattern and measured defects of the training set were used to train the network and predictions were made for other test patterns. The predictions of the two models were compared. Figure 4 shows an example of a SEM picture of a single sample. The labeled dishing defect of the sample was calculated using the measured results.

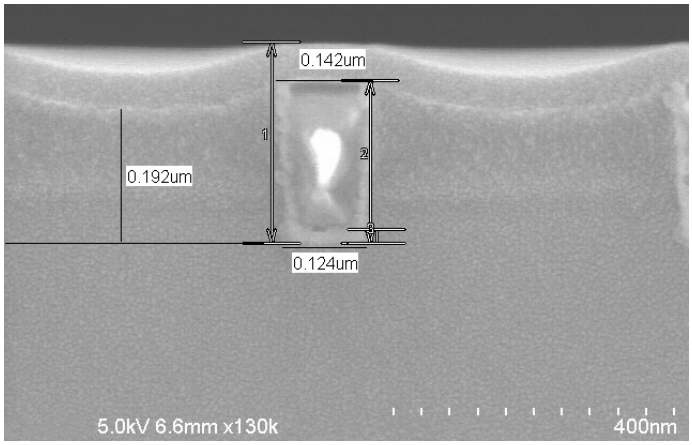

Figure 4. Scanning electron microscope (SEM) picture of a single sample.

4.2. Process Description.

The test chips were designed and fabricated, then AFP and SEM were used to measure the dishing and erosion defects. Our experiments mainly studied the dishing defects. The test patterns and measured results were used to test the accuracy of the DSH-based, CNN-based, and common NN-based CMP models. The processes were as outlined below.

To make predictions with the pressure-distribution- and DSH-based CMP model, the process window was first divided and the equivalent parameters were extracted. The inputs of the model were the extracted equivalent parameters and the origin height of the process windows. The model calculated the removal height according to Preston's equation during a unit time period, updated the height of the surface, and then repeated these steps until the height reached the desired value.

As for the CNN-based and common NN-based CMP models, the first step was to divide the dataset to train the model, as below:

(1) Split the dataset into the training set, validation set, and test set as 60%, 20%, and 20%, respectively;
(2) Use the training set to train the network;
(3) Use the validation set to optimize the hyperparameters of the network;
(4) Use the training set and validation set together to train the optimized network;
(5) Make predictions for the test set and record the results. Guarantee that the test set does not have any contact with the network before making predictions;
(6) Initialize the network and repeat steps 1 to 5 up to 500 times to make other predictions;
(7) The final prediction for each pattern is the average of the recorded predictions.

This process regularizes the algorithm in a similar way to the random forest approach. The averages of 500 models are utilized to reduce the impact of the exception value caused by randomness. The process also keeps the network from knowing the results of the test set, so the performance of the model can be correctly examined.

Common NN-based and CNN-based models have different input data formats. The inputs of the common NN-based model are the equivalent parameters used in the DSH-based model. Besides the equivalent width, the space and density, total perimeter, and amounts of polygenes and crossings are also used as input features. During the training, the network uses MSE as a loss function and compares the prediction and experimental data to update the parameters by back propagation. CNN-based CMP models use complete patterns rather than artificial parameters as inputs. The network learns the pattern effect from the layout pattern and experimental values by itself. A 0.03 um $\times$ 0.03 um square is set as a pixel. The typical length, width, and space are all integral multiples of 0.03 um, so that every pixel is complete filled or blank. A sample uses a 1364×1364 pattern centered on the target position as the input. The CNN layers extract features that have impacts on the output removal results and analyze the weight of each feature. The parallel computing of the CNN network means the model is efficient.

4.3. The Predictions of the Models

CMP models based on DSH, common NN, and CNN were then built. The SEM data were used to calibrate the DSH-based model and to train NN- and CNN-based models. Predictions of a group of locations for all three models were collected and compared with the measured results. The overall root mean squared error (RMSE) values of the DSH-based module, NN-based module, and CNN-based module were 4.17 A, 3.56 A, and 2.02 A, respectively.

The results were then visualized, which are shown in Figures 5–8. The post-CMP heights of patterns whose equivalent line spaces ranged from 0.12 um to 0.48 um and line widths ranged from 0.03 um to 0.15 um were especially relevant. The measured data and predictions for these patterns are shown in the figures.

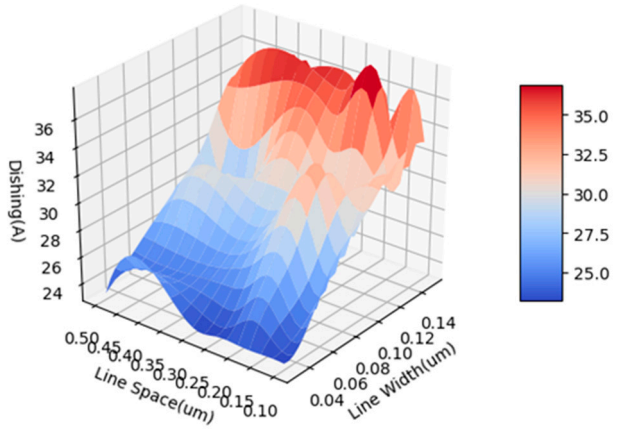

Figure 5. The dishing defects of the SEM data.

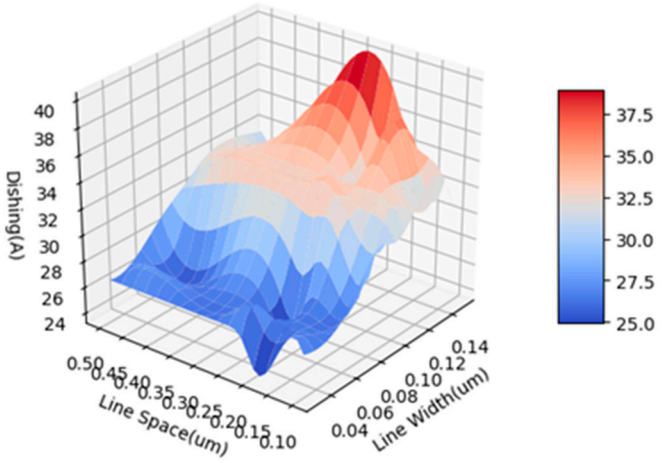

Figure 6. The dishing defects of predictions of the density-step-height (DSH) based chemical mechanical polishing (CMP) model. The root mean square error (RMSE) is 4.17 A.

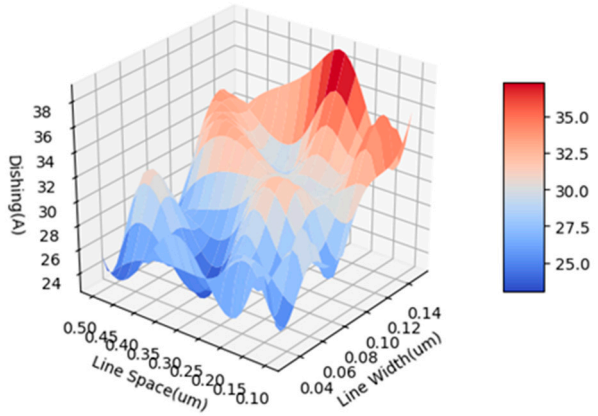

Figure 7. The dishing defects of the predictions of the common NN-based CMP model. The RMSE is 3.56 A.

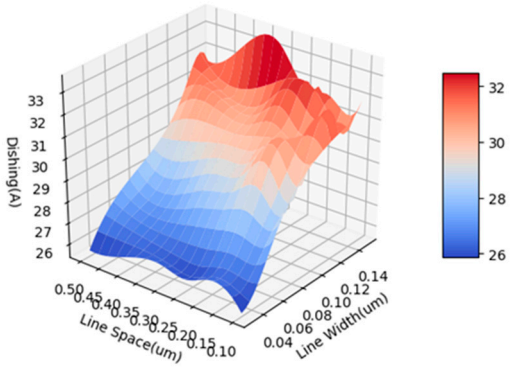

Figure 8. The dishing defects of the predictions of CNN-based CMP model. The RMSE is 2.02 A.

The model based on CNN predicted the CMP results more accurately, for which there are some more details to note.

First, in the predictions of the DSH-based model, we can notice that for the moderate equivalent spacing, the trend of dishing defects changing with the equivalent line width is in accordance with that of the SEM measured data. However, in the case of bigger or smaller equivalent spacing, the predictions errors are large. The reason may be that when dealing with patterns with large spacing and small spacing, the relative relation between the size of the equivalent spacing and the size of the process windows changes, meaning the extracted equivalent parameters no longer meet the approximate conditions of the pressure calculation in the DSH model. At the same time, the change of spacing changes the size relationship with the abrasive particles, making the removal mechanism different from the patterns with 0.30 um spacing. In the model based on NN, although the equivalent parameters are also applied as inputs, the model is corrected with samples that have bigger or smaller spaces during the training process. Therefore, this phenomenon does not occur in the common model.

Second, in the NN-based model, the trend of the dishing defects changing with the changes of the equivalent line width and the equivalent spacing seems to be in poor agreement with the measured data, but still has a lower root mean square error than the DSH-based model. This is because in the training of the NN-based CMP model, the influence of other parameters besides the equivalent line width and spacing, such as the total circumference and the number of polygons, means that it makes sense to remove the results. The NN-based CMP model not only learned the relationships between the output dishing defects and the established physical mechanism, but also learned the relationships with other unknown mechanisms. However, the overall accuracy was greatly improved. The current method for the process window division and parameter extraction loses a lot of information that reflects the features of layout patterns. It is difficult to establish a precise CMP model with only these equivalent parameters.

The CNN-based CMP model has improved accuracy by minimizing the RMSE to 2.02 A. This model makes use of the ability of CNN network to extract features from the patterns and to analyze the weight of each feature on the final output. The experiments showed that the predictions of CNN-based model are in better accordance with the SEM data and can indeed improve the accuracy. The excellent CNN architecture, parallel computing, and parameter reuse also make it possible for the model to have a much smaller unit element in computing, with acceptable time consumption. Therefore, it is an effective method to carry out layout pattern effect analysis based on the CNN network and establish CMP or even other process models that are related to the layout pattern.

5. The Study of the Effective Planarization Length

5.1. The Definition of the Effective Planarization Length

The effective planarization length (EPL) is a quantitative index that describes the range of lengths between positions that interact with each other during the CMP process. To be more precise, it is the

maximum range of the surrounding patterns that can significantly influence the removal rate of the target position. The EPL has many applications. For example, the EPL can be used to make real-time predictions of post-CMP surfaces by only updating the planarization information within the EPL range when the layout is locally modified. Additionally, the EPL can also be used to examine whether the inserted materials will have a significant influence on the CMP defects during dummy filling.

The removal rate of a position in the layout during CMP is determined by two factors: the local height and surrounding patterns. The effect of the local height has been accurately described by elastic deformation theory and Preston's equation. However, there is no effective theory that can precisely predict the impact of surrounding patterns made on the target position. The EPL tries to study the impact range.

In our CNN-based CMP model, the EPL is defined in the following way: for a specific position, when the patterns outside the range L are arbitrary changed and the patterns within L are kept, if the defect of the specific position always varies within a certain range (5% in this article), the minimum distance L that guarantees this condition for all positions on the layout is called the effective planarization length. It must be noted that during a complete polishing process, the variation of the removal height will continue to spread to other areas over time. Therefore, an EPL works under specific process conditions and with a specific removal duration. For a new CMP environment, a new EPL should be solved.

5.2. The Test Patterns and Experiment Process

By using the CNN-based CMP models, the EPL can be efficiently calculated using test layouts. The EPL can then be applied to real-time post-CMP surface prediction and dummy filling to further optimize the design.

The line-space structure accounts for the majority of the SMIC 28 nm standard units, so the test layout is mainly composed of line–space patterns. Figure 9 is a diagram of a test pattern. The main structure of the test pattern is repeated vertical gates with equal spacing and a 40-um gate length. The center of the pattern is selected as the target position. Vertical gates of the same width are arranged at equal intervals on both sides. In our CNN-based CMP model, the size of a single input pattern is $1364 \times 1364 \times 1$ unit squares, while the size of each unit square is 0.03 um × 0.03 um. Therefore, the size of an input image is 40.92 um × 40.92 um.

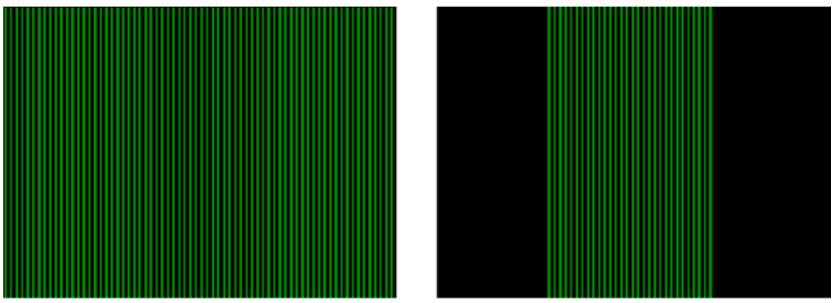

Figure 9. The effective planarization length test pattern. (**a**) original layout pattern, (**b**) the layout pattern that outer vertical gates are removed.

Three groups of test patterns were designed with gate spacings of 0.15, 0.3, and 0.45 um. For each group, the complete original pattern (as Figure 9a) was first input to the trained CNN network and the prediction of the dishing defect D_0 was recorded as the basis. Then, the vertical gates were gradually removed from far to near (Figure 9b) in 2 um steps. For each step i, the maximum horizontal length from the vertical gates to the center was set as d_i. Each new pattern corresponding to d_i was input to the network to predict a new dishing defect D_i; D_0 and D_is were then summarized and are shown in Figure 9.

5.3. Experimental Results and Conclusions

Patterns with 0.06 um, 0.09 um, and 0.13 um gate widths; and 0.15 um, 0.30 um, and 0.45 um gate spaces were designed and input to the trained CNN-based CMP model, as Section 5.2 shows. The results are shown in Figures 10–12 for 0.15 um, 0.30 um, and 0.45 um spacing, respectively. The horizontal ordinate is the horizontal length d_i within which the vertical gates are kept. The vertical ordinate is the rate of change of the dishing defects compared with the basis D_0. Table 1 shows the average minimum d_i, which keeps the rate of change of the dishing defect D_i within 5% compared with D_0.

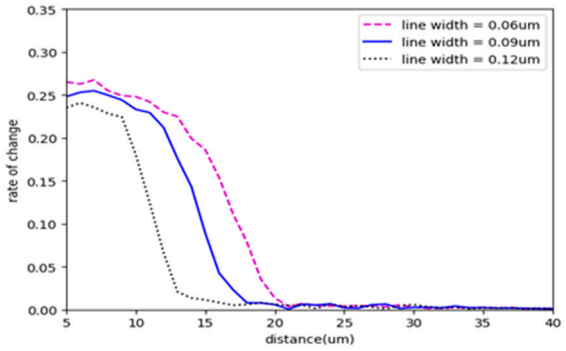

Figure 10. The effective planarization length test with 0.15 um gate spacing.

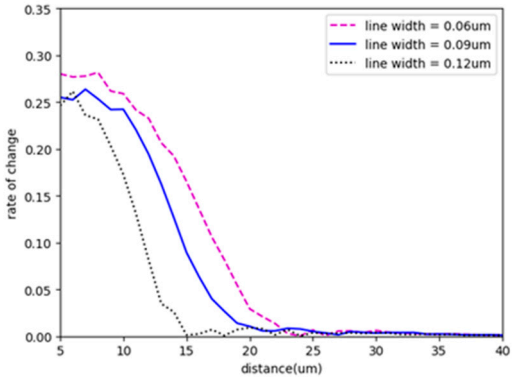

Figure 11. The effective planarization length test with 0.30 um gate spacing.

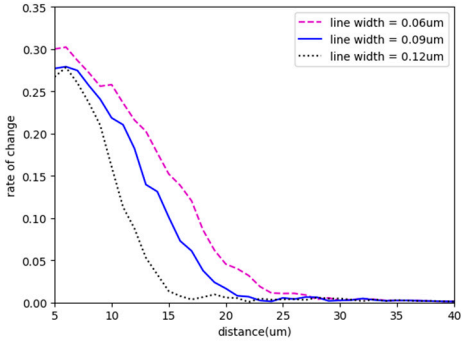

Figure 12. The effective planarization length test with 0.45 um gate spacing.

Table 1. The average minimum d_i, which keeps the rate of change of the dishing defect D_i within 5%.

Average Minimum d_i	Line Width 0.06 um	Line Width 0.09 um	Line Width 0.12 um
Line space 0.15 um	18.45 um	15.27 um	12.52 um
Line space 0.30 um	19.02 um	16.35 um	12.45 um
Line space 0.45 um	19.12 um	17.23 um	12.88 um

From the figures and Table 1, we can see that when the distance reaches 20 um, the impact of the outer regions on the center position is less than 5% in our test chips. Outer regions with other patterns were also tested and the distances were close to the results shown. In general, when considering the pattern effects of CMP for the standard cell library used in our experiment, we can ignore the influence of the patterns beyond the distance of 20 um to the center region within the error range of 5%. With 5% as the standard error range, the maximum effective flattening length of this group of test patterns is 20 um.

Furthermore, there are other conclusions we can draw from the experimental results. First, in several combinations of the gate width and gate spacing, it can be observed that when the distance d_i is between 5 um and 10 um, the outer regions have a 20% to 30% impact on the rate of change of the dishing defects. When d_i is in the range of 10 um to 20 um, the influence of the rate of change drops sharply and reaches about 5% at 20 um d_i. Second, the influence of the gate width on the dishing defects is larger than that of the gate spacing. Patterns with smaller gate widths are usually more easily impacted by surrounding patterns. Third, the gate spacing has smaller impact on the dishing defects than the gate width. Dishing defects of patterns with smaller line widths are less impacted by their surroundings.

In summary, we conclude that in the process environment used in our experiment, the effective flattening length range was about 20 um when a 5% change of the dishing defect rate was used as the threshold. When the horizontal distance d_i between the target position and the boundary of the kept patterns ranged from 5 um to 10 um, the rate of change of the dishing defects was usually 20% to 30%. When d_i changed from 10 um to 20 um, the rate of change decreased from 20% to 5%. Additionally, the rates of change of the dishing defects of patterns with smaller gate spaces, larger gate widths, or larger pattern densities were less affected by changes to the surrounding patterns.

6. The Discussion and Conclusion

In this paper, we can see the influence of the pattern effect on the CMP planarization process. A pressure-distribution- and DSH-based CMP model that is commonly used in Technology Computer Aided Design (TCAD) and Electronic design automation (EDA) tools is introduced. Then, a CMP planarization model based on a convolutional neural network is designed and proposed. The practicality and architecture of the CNN-based CMP model are then discussed.

Test chips are designed and fabricated based on the SMIC 28 nm standard cell library. AFP and SEM are used to measure the height and dishing defects of the post-CMP surface. The results show that the CNN-based CMP planarization model has better accuracy and can make good predictions of the CMP pattern effects. The overall root mean squared error (RMSE) of the CNN-based CMP model is 2.02 A, while the RMSE of the DSH-based model is 4.17 A. The comparison between the CNN-based method and the common neural network method shows that the key factor for improving the accuracy of the CMP models is to better extract new features from the layout patterns instead of using equivalent parameters, such as the equivalent line width, spacing, and density.

The effective planarization length is also introduced and studied. The EPL plays an important role in dummy filling and real-time prediction during layout modification. The test layout and experiment are designed and the CNN-based CMP model is used to study the EPL. The results show that the maximum planarization length is about 20 um in our process environment, with a 5% rate of change of the dishing defects as the threshold. The rate of change of the dishing defects changes with the

horizontal distance d_i from the center when the patterns of the outer regions of d_i are removed. The rate of change varies from 20% to 5% when d_i varies from 10 um to 20 um. The dishing defects of patterns with smaller line spaces, larger line width, and greater density are less affected by the surrounding patterns.

The CMP model based on CNN can effectively analyze layout patterns and can assess the impacts of the extracted features on the planarization output. The CNN-based model has great advantages in the analysis of local and regional pattern effects and is an excellent choice for post-CMP surface height prediction. It can be further used in dummy filling and real-time local CMP prediction.

Author Contributions: H.B. conceived the idea, helped with programming and writing the original draft; L.C. made contributions to conception, design, analysis and experimental verification; B.R. helped with the original draft and formal analysis. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Major Science and Technology Projects of China grant 2017ZX02301007-001 and the Academy of Integrated Circuit Innovation grant Y9YC067001.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study: in the collection, analysis, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

References

1. Tsujimura, M. The way to zeros: The future of semiconductor device and chemical mechanical polishing technologies. *Jpn. J. Appl. Phys.* **2016**, *55*, 6S3. [CrossRef]
2. Wu, L.; Yan, C. Effects of Polishing Parameters on Evolution of Different Wafer Patterns During Cu CMP. *IEEE Trans. on Semicond. Manuf.* **2015**, *28*, 106–116. [CrossRef]
3. Ma, T.; Chen, L.; Fang, J. Study of Optimal Dummy Fill Modes in Chemical–Mechanical Polishing Process. *IEEE Trans. Compon. Packag. Manuf. Technol.* **2012**, *2*, 1043–1047.
4. Katakamsetty, U.; Colin, H.; Yeo, S.; Valerio, P.; Qing, Y.; Fong, Q.S.; Aravind, N.S.; Matthias, R.; Roberto, S. Scanner correction capabilities aware CMP lithography hotspot analysis. Design-Process-Technology Co-optimization for Manufacturability VIII. International Society for Optics and Photonics. 2014. Available online: https://doi.org/10.1117/12.2053035/ (accessed on 17 July 2020).
5. Wu, L.; Hahn, S.; Yan, C. Investigation of Evolution Processes of Wafer Profiles with Edge over Erosion in Copper CMP. *IEEE Trans. Semicond. Manuf.* **2017**, *30*, 69–77. [CrossRef]
6. Bardon, M.G.; Moroz, V.; Eneman, G.; Schuddinck, P.; Dehan, M.; Yakimets, D.; Jang, D.; Van der Plas, G.; Mercha, A.; Thean, A.; et al. Layout-induced stress effects in 14nm&10nm FinFETs and their impact on performance. In Proceedings of the 2013 Symposium on VLSI Circuits, Kyoto, Japan, 12–14 June 2013; pp. T114–T115.
7. Park, T.H. Characterization and modeling of pattern dependencies in copper interconnects for integrated circuits. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2002.
8. Ruan, W.; Chen, L.; Ma, T.; Fang, J.; Zhang, H.; Ye, T. Optimization of a Cu CMP process modeling parameters of nanometer integrated circuits. *J. Semicond.* **2012**, *33*, 127–134. [CrossRef]
9. Ma, T.; Chen, L.; Cao, H.; Fang, J. Re-examining Chemical Mechanical Polishing Pattern Effects Considering Slurry Selectivity. *IEEE Trans. Semicond. Manuf.* **2013**, *26*, 549–555. [CrossRef]
10. Tugbawa, T. Chip-scale Modeling of Pattern Dependencies in Copper Chemical Mechanical Polishing Processes. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, February 1999.
11. Bao, H.; Chen, L. A CNN-Based CMP Planarization Model Considering LDE Effect. *IEEE Trans. Compon. Packag. Manuf. Technol.* **2020**, *10*, 723–729. [CrossRef]
12. Chen, Z.; Sun, F.; Vacassy, R. An Empirical Dielectric Erosion Formula in Metal Chemical Mechanical Planarization. *J. Electrochem. Soc.* **2006**, *153*, G582–G586. [CrossRef]
13. Cai, H. Modeling of Pattern Dependencies in the Fabrication of Multilevel Copper Metallization. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2007.
14. Xie, X. Physical Understanding and Modeling of Chemical Mechanical Planarization in Dielectric Materials. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2007.

15. Chen, D.C.; Lin, D.S.; Lee, T.H.; Lee, R.; Liu, Y.C.; Wang, M.F.; Cheng, Y.C.; Wu, D.Y. Compact Modeling Solution of Layout Dependent Effect for FinFET Technology. In Proceedings of the 2015 International Conference on Microelectronic Test Structures, Tempe, AZ, USA, 23–26 March 2015; pp. 110–115.
16. Lee, B. Modeling of chemical mechanical polishing for shallow trench isolation. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2002.
17. Ouma, D.O. Modeling of Chemical Mechanical Polishing for Dielectric Planarization. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, February 1999.
18. Greenwood, J.A.; Williamson, J.B.P. Contact of nominally flat surfaces. *Proc. R. Soc. Lond. Ser. A Math. Phys. Sci.* **1966**, *295*, 300–319.
19. Greenwood, J.A.; Tripp, J.H. The Contact of Two Nominally Flat Rough Surfaces. *Proc. Inst. Mech. Eng.* **1970**, *185*, 625–633. [CrossRef]
20. Preston, F. The Theory and Design of Plate Glass Polishing Machines. *J. Soc. Glass Technol.* **1927**, *11*, 214.
21. Yang, Z.; Xu, Q.; Chen, L. A Chemical Mechanical Planarization Model Including Global Pressure Distribution and Feature Size Effects. *Compon. Packag. Manuf. Technol. IEEE Trans.* **2016**, *6*, 177–184. [CrossRef]
22. Lecun, Y.; Boser, B.; Denker, J.S.; Henderson, D. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Comput.* **2014**, *1*, 541–551. [CrossRef]
23. Krizhevsky, A.; Hinton, G.E. Using very deep autoencoders for content-based image retrieval. In Proceedings of the 19th European Symposium on Artificial Neural Networks, Bruges, Belgium, 27–29 April 2011.
24. Hinton, G.E.; Srivastava, N.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R.R. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv* **2012**, arXiv:1207.0580.
25. Jarrett, K.; Kavukcuoglu, K.; Ranzato, M.A.; LeCun, Y. What is the best multi-stage architecture for object recognition? In Proceedings of the International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 2146–2153.
26. Luo, J.; Dornfeld, D.A. Material Removal Mechanism in Chemical Mechanical Polishing: Theory and Modeling. *IEEE Trans. Semicond. Manuf.* **2001**, *14*, 112–133.
27. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. *arXiv* **2016**, arXiv:1602.07360.
28. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
29. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2961–2969.
30. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*; NeurIPS Foundation: Montreal, QC, Canada, 2015; pp. 91–99.