Dielectric Spin Coating Characterization, Modeling, and Planarization
Using Fill Patterns for Advanced Packaging Technologies

by

Christopher Ilic Lang

S.B., Electrical Engineering and Computer Science, MIT, 2015

Submitted to the
Department of Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degree of
Master of Engineering in Electrical Engineering and Computer Science

at the

Massachusetts Institute of Technology

January 2017

© 2017 Massachusetts Institute of Technology. All rights reserved.

Author:

Department of Electrical Engineering and Computer Science
January 15, 2017

Certified by:

Duane S. Boning
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by:

Christopher Terman
Chairman, Masters of Engineering Thesis Committee
Department of Electrical Engineering and Computer Science
Dielectric Spin Coating Characterization, Modeling, and Planarization
Using Fill Patterns for Advanced Packaging Technologies

by

Christopher Ilic Lang

Submitted to the Department of Electrical Engineering and Computer Science
January 15, 2017
In Partial Fulfillment of the Requirements for the Degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

Redistribution layers (RDLs) are separate packaging layers dedicated to connecting dies with each other, and to external I/O ports in advanced 2.5D packaging technologies. These layers can be made smaller than the bulky metal traces in conventional substrate packaging, reducing electrical delay and power consumption. Currently, the damascene process is the most common method to create the copper traces in RDLs. However, due to the required inclusion of chemical mechanical polishing (CMP), this process is significantly more expensive than semi-additive electrochemical plating (ECP) and dielectric spin-coating (DSC) processes. The semi-additive techniques are typically avoided as, without CMP, they suffer from thickness variations following the fabrication of each layer. As multiple layers are fabricated, these variations compound, and can result in a structure with significant topographical and electrical performance concerns.

In this thesis, we model and predict the surface non-uniformities resulting from the DSC process applied to underlying topographies, and propose dummy fill and cheesing patterns which control the variations of the DSC process. We first design test vehicles (TVs) which represent topographies common in RDLs, most notably the copper lines and vias, and use these to experimentally determine the thickness variations caused by each process. We then develop empirical models based on these results. The DSC process is modeled as a convolution between the underlying topography (typically the copper lines) and an appropriately chosen impulse response. Finally, we present dummy fill and cheesing patterns that have the potential to control the variations of both processes for any arbitrary layout.

Thesis Supervisor: Duane S. Boning
Title: Professor of Electrical Engineering and Computer Science
Acknowledgements

Firstly, I would like to thank my research advisor, Duane Boning, for guiding me through my Master’s program. His advice has been instrumental in producing the following work, and his patience and enthusiasm have made the process a pleasure. I couldn’t ask for a better advisor, and look forward to working together in the future. Secondly, I would like to thank my parents, Jeffrey Lang and Marija Ilić. Both have given me invaluable advice and inspiration throughout my life, and I owe them everything that I have. I wouldn’t be where I am today without having them as parents. I would also like to thank all of the MTL staff, and in particular Vicky Diadiuk, Kurt Broderick, Dave Terry, Paul Tierney, Scott Poesse, Robert Bicchieri, Bernard Alamariu, and Paudely Zamora, who have each helped me during my research. Finally, I would like to thank Taiwan Semiconductor Manufacturing Company, who has in part sponsored this work, and in particular C.-C. Hsieh, J. Hu, J. Ho, C.-H. Tung, C.-H. Steven Lin, H.-P. Pu, and C.H. Yu all of whom have provided feedback throughout the process.
Contents

Chapter 1: Introduction .......................................................................................................................... 10
  1.1) Redistribution Layers in Advanced Packaging ............................................................................. 10
  1.2) Existing DSC Models ..................................................................................................................... 12
  1.3) Thesis Organization ...................................................................................................................... 13

Chapter 2: Experimental Setup and Results .......................................................................................... 15
  2.1) Test Vehicle Design and Fabrication ............................................................................................ 15
    2.1.1) Initial TV Design ....................................................................................................................... 16
    2.1.2) Transitions ............................................................................................................................... 16
    2.1.3) Vias .......................................................................................................................................... 18
    2.1.4) Dimensions .............................................................................................................................. 19
    2.1.5) Material Choice ....................................................................................................................... 19
    2.1.6) TV Fabrication and Characterization ....................................................................................... 21
  2.2) Spin Coating Results .................................................................................................................... 21
    2.2.1) Feature Scale Variation ............................................................................................................. 22
    2.2.2) Long Range Variation .............................................................................................................. 23
    2.2.3) Tip Pinch-Off and Measurement Fidelity .................................................................................. 25

Chapter 3: Modeling .............................................................................................................................. 29
  3.1) Model Form ................................................................................................................................... 29
  3.2) System Impulse, Step, and Frequency Response ........................................................................... 29
    3.2.1) Initial Fit ................................................................................................................................... 30
    3.2.2) Updated Fit ............................................................................................................................... 33
  3.3) Model Predictions .......................................................................................................................... 36
  3.4) Model Accuracy .............................................................................................................................. 38
  3.5) Model Extremal Cases ................................................................................................................... 40
    3.5.1) Penalty Function for Extremal Cases ....................................................................................... 41
  3.6) Model Trends .................................................................................................................................. 45
  3.7) Impact of Updated Fit and Coefficient Significance ...................................................................... 46
  3.8) Closed Form Analysis ................................................................................................................... 49
  3.9) Existing DSC Modeling Revisited ............................................................................................... 54
  3.10) Effects of Feature Location and Orientation .............................................................................. 56
  3.11) Effects of Coating vs. Curing ...................................................................................................... 58
Chapter 1: Introduction

This thesis explores some of the problems and potential solutions associated with the dielectric spin coating process when used in redistribution layer fabrication. In particular, we focus on modelling and controlling the non-planar surfaces resulting from coating existing topographies. In this chapter, we introduce the necessary background information and give an overview of the thesis as a whole. In Section 1.1 we review the necessary background information regarding integrated circuit (IC) packaging, redistribution layers fabrication, and the spin coating process. We then briefly cover the existing spin coating process models and their limitations in Section 1.2. While these existing models are only briefly discussed, we later return to them in Sections 3.9 through Section 3.11. Finally, we describe the overall structure of the thesis in Section 1.3.

1.1) Redistribution Layers in Advanced Packaging

Modern IC design is commonly driven by improving the electrical performance, cost, power efficiency, and system integration of the devices fabricated. Traditionally, designers and manufacturers have focused on improving these characteristics within an individual die. However, as the capability of the individual die continues to improve, a need for improved also packaging arises [1, 2]. Bulky metal traces in conventional substrate packaging, such as ball grid arrays (BGAs) and quad flat packages (QFP), limit interconnect density, are electrically slow, and consume significant power [3]. If not considered, these limitations can bottleneck the IC and system performance as a whole. Redistribution layers (RDLs) are separate packaging layers that use copper interconnects to connect multiple die to one another, as well as to external I/O ports, as seen in Fig. 1. Additionally, this form of packaging, known as 2.5D integration, can be more thermally and mechanically reliable compared to fully stacked, 3D integration [4], and can be more easily integrated with existing designs [5], making it a practical solution to current packaging needs.
Currently, the most common fabrication method for multi-level RDLs is the damascene process [5]. In this process, the dielectric is first deposited, then etched where copper lines will be formed. Copper is then deposited across the entire structure. Finally, chemical-mechanical polishing (CMP) removes the unwanted copper from above the dielectric, leaving a planar surface with embedded copper interconnect lines and vias.

While the damascene process produces well-formed structures, it is also costly [6]. A critical factor in adopting the 2.5D integration technology will be reducing this cost [5]. An alternative, semi-additive metallization process is a potential improvement, as it does not require CMP [2]. In this process, copper lines are first grown using semi-additive electro-chemical plating.
Then the dielectric is deposited, typically using dielectric spin coating (DSC). Finally, parts of the dielectric are etched to reveal the copper lines below. Both the damascene and semi-additive processes can be seen in Fig. 2.

While potentially less expensive than the damascene process, RDLs made using the semi-additive metallization process suffer from thickness variations, as the process does not use CMP [7]. Instead, as multiple layers are fabricated, thickness variations accumulate and can reduce the performance of the RDL structure. An illustration of this effect can be seen in Fig. 3. Additionally, large thickness variations across a layer require a large depth of focus in subsequent lithography steps, thus reducing the possible resolution and impacting the feature size for those layers [2, 7].

![Figure 3: Incorrectly fabricated RDL. Thickness variations in the ECP and DSC cause uneven final topography.](image)

In order to better understand the limitations of the conventional metallization process, this thesis explores a new model that predicts the resulting thickness variations after DSC based on the underlying topography. Using this model, metal layers can be designed such that the thickness variations after spin coating is limited, thus eliminating the need for CMP and reducing the fabrication cost of RDLs in advanced packaging.

1.2) Existing DSC Models

An existing DSC process model comes from Stillwagon and Larson [8], who predict the topography of a liquid dielectric during spin coating. However, this model suffers from two major problems. First, evaporation has a significant impact on the final topography of DSC, and this is not included in the initial model. In order to incorporate evaporation, a “two-stage” model is further developed by Peurrung and Graves [9]. During the first stage of this model, drying is assumed to have no effect on the shape of the film profile. Then after a reaching equilibrium during spinning,
the evaporation is modeled as a fractional loss. While this is the predominant DSC model, this thesis later shows that for polyimides, and other dielectrics which require separate curing stages, this model cannot be applied. In these cases, the film is often level after spin coating, and the curing stage dominates the final topography.

A second major problem with the two-stage model is that it is computationally infeasible to apply to the entire RDL. The model takes the form of a partial differential equation that must be solved using iterative methods. Applying this model to areas of 20mm x 20mm or larger, with feature sizes of 1 µm or less becomes computationally difficult, as there are approximately 400 million nodes. A DSC model is needed that can scale to larger areas.

A second model for the DSC process comes from Hirasawa et al. [10]. In this analysis, the authors focus on modeling the drying stage of the spin coating process. While more relevant to polyimides and other polymers that require separate curing stages, there are still significant differences between the Hirasawa model predictions and the experimental results presented in this thesis. It is believed that this discrepancy is due to simplifying assumptions used in the model, which are later explored in Section 3.11. Finally, the model again takes the form of a complex partial differential equation, limiting its feasibility for large scale applications.

1.3) Thesis Organization

The lack of an accurate and efficient model for the DSC process motivates further development. In this thesis, a newly proposed model is developed and compared to experimental results. Before presenting the model itself, the design and results of experiments used to develop and assess the model are presented in Chapter 2. This chapter begins in Section 2.1 by detailing the design and fabrication of Test Vehicles (TVs), structures which represent a range of topographies common in RDLs. After fabrication, the TV’s are coated with polyimide, and the resulting topographies are profiled and discussed in Section 2.2.

After discussing the experimental data, the predictive model for the process is presented in Chapter 3. The model is used to predict the profiles measured and presented in Chapter 2. The results of the model predictions are compared to the experimental data, and the model accuracy is quantified. Later in Chapter 3, the model is applied to two new types of underlying topographies, and its accuracy is examined. First, the models is used to predict coatings over structures with
multiple starting heights. Then, its form is expanded to two dimensions. Again, the results of the model in both cases are compared to experimental data. Chapter 4 then uses the results of the model to develop fill patterns which can limit the resulting dielectric surface variation of any existing RDL layout. Finally, Chapter 5 presents final conclusions and suggests future work.
Chapter 2: Experimental Setup and Results

In this chapter, we present the methodologies and results of our experimental findings used to develop the spin coating model. Section 2.1 details the development and final design of our Test Vehicle (TV). Here, we present both the layout of the structure, the fabrication techniques used, and the range of TV geometries used. Section 2.2 then presents experimental profiles of the spin coated TVs, and the feature scale and chip scale trends seen in these experimental profiles.

2.1) Test Vehicle Design and Fabrication

Here, the design and fabrication of the test vehicle is presented. A good TV includes structures common in RDLs, and the TV coated profiles are the primary data set used to develop and characterize the model. Therefore, the range of features present in the TV should not only represent those common in RDLs, but the TV features should be laid out in a way that facilitates the model development.

The most abundant and important features present in RDLs are the copper interconnects. Therefore, the TV is designed to include a large range of interconnect feature sizes. The geometry of each individual interconnect is defined by its line width and height, as illustrated in Fig. 5. While the line length also affects the coated profile, the profilometer used only measures height in one dimension, and thus the initial model only predicts these one dimensional coatings. For this reason, we use sufficiently long lines that reduce the effect of this second dimension (length), and do not initially consider its effect. We later expand our model to two dimensions in Section 3.14, which can then predict the coatings over lines of arbitrary lengths.

In addition to the geometry of each interconnect, the spacings of the interconnects are needed to define the entire RDL layout. Finally, the polyimide thickness is needed to define the entire coated RDL. These four parameters, height, width, spacing, and thickness (commonly abbreviated $H$, $W$, $S$, and $T$) are the main geometric parameters considered in this study. A common range of values for each of these variables is later defined and discretized. Finally, the TVs are designed, fabricated, and coated such that all combinations of these four discretizations are present.
2.1.1) Initial TV Design

The first iteration of the TV design aims to represent the major variables as simply as possible. This design consists of a matrix of uniform regions, where each region consists of parallel lines with a defined line width and line spacing. An overview of the initial TV design can be seen in Fig. 4 and an illustration of the uniform regions can be seen in Fig. 5. Along one axis of the TV, the line widths of the uniform regions increase, and along the other axis, the line spaces increase. This design ensures that there exists a uniform region for any combination of the defined line widths and spacings within each TV. Line height and polyimide thickness are not considered in this design, but are determined during the TV fabrication and subsequent coating. For this reason, multiple TVs are fabricated and coated such that there exists a range of combinations of line height and polyimide thickness. The TV layout, in combination with multiple TV fabrication runs, ensures that within the set of TVs, there exists a uniform region with each desired combination of the predefined line widths, spacings, heights, and polyimide thicknesses and that the ranges of interconnect dimensions common in RDLs are represented.

![Initial TV layout showing matrix of uniform line width and line spacing regions.](image1)

Figure 4: Initial TV layout showing matrix of uniform line width and line spacing regions.

![Diagram of lines in within a uniform region.](image2)

Figure 5: Diagram of lines in within a uniform region.

2.1.2) Transitions

While the initial design contains all desired combinations of line spaces, widths, and heights, it neglects the transitions between regions. In the initial design considered above, the line widths and spaces monotonically increase along each axis, and there is little variety in the transitions between the regions. Increasing the variety of feature changes between regions allows
us to better observe the area of effect or spatial range of each process, and helps develop a more general model. Here a re-ordering of the initial design is developed that is intended to maximize the range of feature transitions in the RDL.

In order to maximize the range of line width transitions, a staggered ordering is used, as is illustrated in Fig. 6. Instead of monotonically increasing line width, the largest width is followed by the smallest width, followed by the second largest width, followed by the smallest largest, etc. This layout results in transitions of decreasing intensity, and maximizes the range of transitions present in the TV.

Figure 6:  Staggered line widths. Note: $W_1$ through $W_N$ refers to relative line width.

While this technique increases the range of transitions in line widths between regions, it does not increase the transitions in line spacings. Staggering both axes would not achieve this, as the directionality of the TV interconnects allows profiles to be measured in only one direction. In order to maximize the range of transitions for both line spacing and width, the design is split into two halves. Each half includes either staggered line spacing, or staggered line width, while the other parameter is kept monotonically increasing, as is illustrated in Fig. 7. Both halves of the design still include the same combinations of line widths and spaces; however, the width of each uniform region is reduced by a factor of two. This final modification allows for a more complete study of line width and space transitions, without increasing the total area of the TV, or reducing the total area devoted to each line and space combination.
2.1.3) Vias

While copper lines account for the majority of the features in RDLs, copper vias are also present. For this reason, vias are included along the outer side of the TV. Both square and circular vias, as well as positive (raised), and negative (recessed) vias are included in the design. The inclusion of both positive and negative vias allows the same mask to be used for both positive and negative resists when fabricating the TV, and also gives additional information that can be considered during the model development. The final layout of the TV can be seen in Fig. 8.

2.1.4) Dimensions

The overall TV dimensions are 22mm x 22mm, as this is the maximum image field size of the Nikon stepper lithography system used at the MIT Microsystems Technology Laboratories (MTL). Structures larger than this require additional steps to fabricate, and this is not necessary for the TV design.

The line widths and spaces chosen range from 0.5 µm to 256 µm. This range is chosen as feature sizes smaller than 0.5 µm are not common in RDLs [2], and features spaced greater than 250 µm apart can be modeled as isolated features. Ten values, geometrically spaced from 0.5 µm to 250 µm, are used for the TV line widths and spaces. This gives each uniform region an area of 2 mm x 1 mm, and line width and spaces of 0.5 µm, 1 µm, 2 µm, 4 µm, 8 µm, 16 µm, 32 µm, 64 µm, 128 µm, and 256 µm. These values are also used for the via sizes, while the distances between the center of the vias is kept constant at 200 µm (except for the 256 µm vias, where distances of 500 µm are used instead). The nominal feature heights are 2 µm, 4 µm, 5 µm, 7 µm, and 9 µm.

2.1.5) Material Choice

The final decision in the TV design is the choice of material used to create the mock interconnects. Direct deep silicon etching into the wafer will not result in an exact or uniform depth, as the DRIE process to do so has its own line width and spacing and other pattern dependencies [11]. It is necessary to select a process which will accurately replicate the TV design with a relatively uniform step height across the TV pattern features. Four methods are considered, and their pros and cons evaluated.

2.1.5.1) Polyimide

Constructing the TV out of polyimide, a typical dielectric used in RDL fabrication, is the first option considered. This technique has several advantages. First, because it is used in the RDL itself, polyimide accurately represents one of the materials found in the underlying topographies. Secondly, many polyimides are photosensitive, so there is no need for a separate lithography and etching step. Simply exposing and developing the polyimide itself produces the desired TV surface structure. However, the use of polyimide (or any other photoresist) is ultimately rejected, as its optical properties are too similar to the dielectric to be used in the spin-on experiments. This would make it difficult to use optical measurements to determine the absolute thickness of the dielectric
after spin coating.

2.1.5.2) Copper Electroplating

The use of copper electroplating is also considered to fabricate the TV. Again, this material is found in RDLs, but has unique optical constants, making spectroscopy an option. However, this too is initially dismissed, as the only electroplating facilities in MTL are in the Exploratory Materials Laboratory (EML), and therefore the TV would not be allowed back in the Integrated Circuits Laboratory (ICL), where the remainder of the processing takes place. Additionally, prior studies have shown that the copper growth rate during electroplating has its own pattern dependencies, which would result in a TV with non-uniform line heights [12]. While not used for the initial TV fabrication, we do note that it is later used when fabricating multi-layer TVs in Section 3.12. This takes place after the DSC model is developed, and so the pattern dependencies of each process are not confused in this later case.

2.1.5.3) Buried Oxide

The use of buried oxide wafers is also considered for TV fabrication. These silicon wafers have a layer of oxide inside of them, which can be used as a stopper during the etching process. The silicon and oxide thicknesses are highly accurate, and etching the TV design into them would create a consistent thickness. However, these wafers are costly, and a cheaper solution is desired.

2.1.5.4) Oxide Deposition

The final option, and the process ultimately chosen, consists of an oxide deposition on a silicon wafer, followed by a plasma etch using a 3:1 mixture of CF$_4$ to CHF$_3$. Depositing the oxide on a silicon wafer is less expensive than purchasing buried oxide wafers, and still has the same advantages. In this technique, the silicon is used as an etch stopper, and the high selectivity allows for well-defined line heights [13]. Additionally, the optical constants of silicon and polyimide are significantly different, and spectroscopy can be used to determine the absolute dielectric heights during spin coating experiments [14, 15]. For these reasons, an oxide deposition is chosen for the TV fabrication.
2.1.6) TV Fabrication and Characterization

TVs are fabricated using nominal line heights of 2 µm, 4 µm, 5 µm, 7 µm and 9 µm. For the larger line heights, overetching and underetching become more common. To ensure that the final TVs are properly etched, the thickness of the oxide is optically measured after its deposition. Then, following the etch the thicknesses of the exposed areas and areas covered by photoresist are again measured to verify that the oxide is etched where desired and that the heights of the covered areas remain unchanged. Finally, SEM images of the TVs are taken for the 2 µm step height TV. These images can be seen in Fig. 9. Cross-sectional SEM images after coating are later provided in Section 2.2.3 to confirm the profilometry scans.

![SEM image of TV. Both have 2 µm line height and 1 µm line widths. The left image has 1 µm and 2 µm spacings while the right has 0.5 µm line spacing.](image)

2.2) Spin Coating Results

After fabrication, the TVs are spin coated with dielectric. Polyimide is a commonly used dielectric in RDLs [16-18], and is the dielectric of choice in this study. The polyimide chosen is HD-4110 by HD Microsystems.

For each line height, two TV wafers are fabricated. This allows us to use two different polyimide thicknesses for each line width, spacing, and height combination. Nominal polyimide thicknesses of 8 µm and 13 µm are used. After spin coating, the TVs undergo a soft bake at 100°C for 10 minutes, followed by a cure at 300°C for 1 hour.

After spin coating, the surface of the coated TVs is profiled. For each TV, only the upper half (staggered spacing) is profiled. For feature sizes less than 4 µm the surface is found to be almost completely planar, and features with either the widths or spacings above 64 µm appear to
be independent. For this reason only a subsection of each TV is profiled. This subsection includes all widths between 4 µm and 128 µm, and all spacings between 2 µm and 64 µm.

As previously described, each TV consists of a matrix of uniform regions of parallel lines. While each uniform region could be individually profiled, instead multiple uniform regions are profiled in a single scan. This allows both the feature scale variations, as well as the transitions between uniform regions, to be measured. As only the upper half of the TV is profiled, the line width within each scan is constant, while the line spacings change within each scan. Between scans, the line width changes, while the ordering of the spacings is constant. The ordering of the spacings in each scan is 16 µm, 8 µm, 32 µm, 4 µm, 64 µm, 2 µm. Sample scans can be seen in Fig. 10, while a complete collection of the scans for all combinations of line widths, spacings, heights, and polyimide thicknesses are presented in the appendix alongside their model predictions.

![Figure 10: Sample experimental profiles for W = 4 µm (left) and W = 64 µm (right). Both are from the H = 5 µm, T = 8 µm TV.](image)

### 2.2.1) Feature Scale Variation

After profiling each TV, the scans are analyzed. A key statistic extracted is the feature scale variation. The feature scale variation is defined as the average difference between the peak and the troughs within a given uniform region and is found to be a function of line width, spacing, height, and polyimide thickness. To calculate this value, only the middle third of each uniform region is analyzed. This minimizes the impact of the adjacent uniform regions. This middle third is then divided into smaller segments of length \( W + S \) (the pitch of the uniform region in the TV), and the minimum and maximum value are extracted from each period. Finally the differences between
these minima and maxima is averaged over all periods for each uniform region to find the average feature scale variation.

While the feature scale variation is a function of both line width and spacing, these two variables can be simplified into a new parameter defined as the effective feature width, $W_{eff}$:

$$W_{eff} = \frac{w \cdot s}{w + s}$$

All combinations of widths and spacings with equivalent $W_{eff}$ will have equal feature scale variations. This value represents the variation due to an isolated feature with width $W_{eff}$. As $S$ approaches $\infty$, the feature becomes isolated, and $W_{eff}$ simplifies to $W$, the width of the feature. A similar analysis shows that this is also the variation due to an isolated trench of “spacing” $S$. It should be noted that $W_{eff} \leq W, S$ for all combinations of $W$ and $S$. After later developing the spin coating model, the development of this simplification is expanded upon in Section 3.8.

The variation in each uniform region as a function of $W_{eff}$ can be seen in Fig. 11 for $H = 2 \ \mu m$ with $T = 8 \ \mu m$ and for $H = 2 \ \mu m$ with $T = 13 \ \mu m$. The variations for the remainder of the coated TVs can be seen in the appendix next to the corresponding model predictions.

![Figure 11: Feature scale variations vs. $w_{eff}$ for $H = 2 \ \mu m$ with $T = 8 \ \mu m$ (left), and $13 \ \mu m$ (right).](image)

2.2.2) Long Range Variation

In addition to calculating the feature scale variations for each region, the long range, or chip scale, effects of the underlying topography are also considered. Here, a complete 2D profile for the coated TV is constructed, then the average heights within the uniform regions are calculated. To build this 2D profile, the previous profile scans, similar to those of Fig. 12, are
combined with a single column scan of the TV. This single column scan determines the relative heights between the row measurements, thus allowing the relative heights between all profiled regions to be known. It should be noted that the column measurement is taken from the $S = 2 \, \mu\text{m}$ column. This is the column with the lowest feature scale variation, thus minimizing the uncertainty of the 2D profile. Fig. 13 shows the location of each measurement, while Fig. 14 shows the complete TV profile.

![Figure 12: Example scan of a TV Row.](image1)

![Figure 13: Illustration of scan locations: row scans (left), column scan (center), and complete scans (right).](image2)

![Figure 14: Sample resulting 2D TV profile.](image3)
The average height of each uniform region is calculated to quantify these chip scale variations. It is found that the height variation between uniform regions is approximately linear with pattern density \( \frac{W}{W+S} \). Intuitively this makes sense, and suggests a conservation of mass during spin coating. Each uniform region contains the same amount of polyimide; however, its distribution within the region may differ based on the underlying topography. Therefore, the relative average height after spin coating is the same as the relative average height before spin coating, \( H \cdot \frac{W}{W+S} \). Plots for the average height of each uniform region for \( H = 1.8 \) and \( 4.2 \) um and \( T = 8 \), and \( 11 \) um, can be seen in Fig. 15.

![Pattern Density vs Height](image1)

![Pattern Density vs Height](image2)

![Pattern Density vs Height](image3)

![Pattern Density vs Height](image4)

*Figure 15: Average heights as a function of pattern density. \( H = 1.8 \) um (top), \( H = 4.2 \) um (bottom), \( T = 8 \) um (left), \( T = 11 \) um (right).*

### 2.2.3) Tip Pinch-Off and Measurement Fidelity

For high aspect ratio trenches, tip pinch off during profilometry scans was encountered. As the line height increases, the stylus tip no longer reaches the bottom of the high aspect ratio
trenches as illustrated in Fig. 16.

Figure 16: Tip pinch off for high aspect ratio trench.

This was first noticed while developing an initial version of the model that systematically overpredicted the feature scale variation when compared to the experimental results. It was then noticed that the effective feature width simplification would also break down for high aspect ratio trenches. These two phenomenon suggested that tip pinch off was occurring. The original tip radius used was 12.5 µm, but after purchasing a new tip of radius 2 µm, both of these problems were resolved. The tip manufacturer, Bruker Corporation, provides the maximum trench depth possible to resolve given the trench width for common tip radii, which can be seen in Fig. 17. For 8 µm spacings, all of the trench feature scale variations are measured to be smaller than 4 µm, and for 4 µm spacings, all feature scale variations are under 1.5 µm, except one (line height 9 µm, polyimide thickness 8 µm). As the tip radii is 2 µm, these should all be under the maximum resolvable trench depth.

Figure 17: Maximum resolvable trench depth as a function of trench width and stylus radius [Bruker Nano Inc., personal communication, April 21, 2016].
To confirm this, cross sectional images are taken and compared to the profilometry results. SEM images for a coated TV of line height of 5 µm, line width of 32 µm, polyimide thickness of 8 µm and spacings of 2 µm and 4 µm are shown in Figs. 18 and 19.

Figure 18: Cross sectional SEM image of $H = 5 \, \mu m$, $T = 8 \, \mu m$, $W = 32 \, \mu m$, $S = 2 \, \mu m$.

Figure 19: Cross sectional SEM image of $H = 5 \, \mu m$, $T = 8 \, \mu m$, $W = 32 \, \mu m$, $S = 2 \, \mu m$.

The feature scale variations measured using the SEM images are compared to those of the profilometer, and are presented in Fig. 20. The variation measured by the SEM is consistently lower than that of the profilometer, especially for small spacings. This suggests that the profilometry scans are not suffering from pinch off, otherwise the SEM scans would give consistently larger variations. This discrepancy is most likely due to the truncation of the SEM measurements. Distances measured using the SEM are truncated at two significant figures, and therefore the peak measurements are all 10 µm (as opposed to 10.5 µm, for example). This lowers the variation measured by the SEM.
Figure 20: Surface variation measured using SEM vs. Profilometer for $H = 5 \mu m$, $T = 8 \mu m$.

Using the 2 $\mu$m profilometer tip, the effective feature width simplification works well. If the profilometry scans were to still suffer from pinch off, they would produce smaller variations for small spacings and large line widths compared to the variation from large spacings and small line widths. The fact that this is not the case provides the final piece of evidence to confirm that the profilometry scans accurately represent the true coated profiles.
Chapter 3: Modeling

In this chapter, the predictive model for the spin coated surfaces is developed. At a high level, the coated profile is modeled as a convolution between the underlying topography and an appropriately chosen impulse response. This modeling technique has been used for similar pattern dependent processes [19, 20].

After presenting the model form in Section 3.1, the impulse response for each coated TV is experimentally determined in Section 3.2. Then, the experimental data is fit to an analytic form, and the surface of the coated TV is predicted and presented in Section 3.3. These predictions are then compared to the experimental data in Section 3.4, and an optional addition to the model is presented in Section 3.5 that improves the accuracy for extremal cases. The trends seen in the model coefficients and a closed form analysis are then discussed in Sections 3.6 through Section 3.8. We revisit the existing models and compare them to the one presented in this thesis in Sections 3.9 through 3.11. Finally, the model is expanded to two dimensions, and the coating of non-binary and two-dimensional surfaces are predicted in Sections 3.12 through 3.14.

3.1) Model Form

As mentioned, the coated surface is modeled as a convolution between an appropriately chosen impulse response and the starting topography. From a signal processing perspective, the input signal is the underlying profile, \( p_u(x) \), the output is the coated profile, \( p_c(x) \), and the system represents the act of coating the wafer, with an impulse response \( h(x) \).

\[
p_c(x) = \int p_u(x - x') \cdot h(x') \, dx' = p_u(x) * h(x)
\]

Here, the convention used is that the reference height for \( p_c(x) \) is the coated, but un-patterned, area. To find the height above the silicon wafer surface, the thicknesses of each polyimide layer on a flat surface must also be added.

3.2) System Impulse, Step, and Frequency Response

The impulse response of the system cannot be directly measured, but can instead be derived from the step response, \( E(x) \), of the system. As the model takes the form of an LTI system, the
impulse response is derived by taking the derivative of the step response, as shown in Fig. 21. These step responses are obtained by measuring the profile of a coated “step” in height on the TV. Here, \( p_u(x) = 1 \) for \( x \geq 0 \) and \( p_u(x) = 0 \) for \( x < 0 \).

When measuring this step response, it is best to measure a step which is as isolated as possible. This reduces the effects that surrounding regions have on the step response. In practice, the steps measured are at least 2000 \( \mu \)m away from any other feature, and the responses were truncated 1000 \( \mu \)m away from the feature itself.

![Figure 21: Example unprocessed step response (left), and derived impulse response (right).](image)

To reduce noise in the measurements, six step responses are averaged for each line height and polyimide thickness combination. Here, the peaks of each impulse response are used to align the step responses. Two of these step responses are measured 1 cm away from the center of the wafer, two 3 cm away, and two 5 cm away. It should be noted that later in Section 3.10, the impulse response is shown to be independent of the radial distance from the center of the wafer, allowing the responses to be averaged at multiple locations. Finally, the step responses are leveled in post processing by setting the slopes of the responses at each edge to be zero. This is done by fitting the first and last 10% of each step response to a first order polynomial, and subtracting the average of these two fits from the experimental data.

3.2.1) Initial Fit

After experimentally collecting the step and impulse responses, they are fit to an analytic function. While the form of the impulse response is not immediately obvious, its Fourier Transform, the spatial system frequency response, is easier to observe. As shown below, the spatial
frequency response of the system fits nicely to the form:

\[ \mathcal{H}(k) = e^{-\beta |k|} \]

where \( k \) is the spatial frequency, and \( \beta \) is a constant that encapsulates the system parameters. Sample system response fits using this form are shown in Fig. 22.

![Figure 22: System spatial frequency response fit to \( e^{-\beta |k|} \). \( H = 1.8 \text{ um} \) (top), \( H = 4.2 \text{ um} \) (bottom), \( T = 8 \text{ um} \) (left), \( T = 13 \text{ um} \) (right).](image)

Using the inverse Fourier Transform, the impulse response takes the form:

\[ h(x) = \frac{\beta}{\pi \cdot (\beta^2 + x^2)} \]

Fig. 23 shows the fits for the experimental impulse responses. Finally, integrating \( h(x) \) with respect to \( x \) gives the form of the normalized step response:

\[ E(x) = \frac{\tan^{-1}(\frac{x}{\beta})}{\pi} + \frac{1}{2} \]
This agrees with intuitive expectation for the systems step response. Near the feature there is a steep slope, and further away, the height asymptotically approaches either 0 or 1.

![Figure 23: Impulse responses fit to $\frac{\beta}{\pi(\beta^2+x^2)}$.](image)

While the frequency response fits nicely to its proposed form, a discrepancy exists near $k = 0$. By definition, $\mathcal{H}(0) = 1$, as the impulse response must integrate to 1. However, the fit does not intersect with this point. Instead there appears to be an impulse at $k = 0$. If this is real, the new frequency response takes the form:

$$\mathcal{H}(k) = e^{-\beta |k|} + c \delta(k)$$

From this it follows that:
\[ h(x) = \frac{\beta}{\pi \left( \beta^2 + x^2 \right)} + c \]

and:

\[ E(x) = \frac{\tan^{-1}\left(\frac{x}{\beta}\right)}{\pi} + \frac{1}{2} + c \cdot x \]

This corresponds to a constant offset in the impulse response, and therefore an additional constant slope in the step response. This implies that infinitely far away from a feature, the coating thickness is infinite, which does not make intuitive sense. Instead, the zero frequency offset is attributed to a leveling and scaling issue during the step response measurements.

3.2.2) Updated Fit

While this form fits the impulse response well, the corresponding step response fit is less accurate than desired. A sample fit for the step response, as well as the resulting impulse response fit can be seen in Fig. 24. Here, fitting the model to the impulse response results in a different fit than fitting the model to the step response. In the first case, the mean squared error (MSE) of the impulse response fit is minimized, while in the second case the MSE of the integral of the impulse response fit is minimized. These are two separate optimization problems with different answers, explaining why the impulse response fit looks poor when the step response fit is optimized and vice-versa.

![Figure 24: Step response fit, normalized by line height, and impulse response fit.](image-url)

*Figure 24: Step response fit, normalized by line height, and impulse response fit.*
Clearly, fitting the step response to the hypothesized model does not produce the most accurate results for the corresponding impulse response. While this fit does capture the coating shape immediately near the feature, its accuracy quickly degrades as the distance from the feature increases. Therefore, when using this form of the model, the long range, or chip scale, effects are typically lost when predicting coated profiles. This phenomenon is further explored in Section 3.7 after presenting the results of the updated model.

The error between the step fit and experimental profile is presented in Fig. 25. Here, the error on each side of the feature can be approximated as a decaying exponential. To improve our model, this term is added to the system response fit to account for this error.

When this term is added, the new impulse response takes the form:

\[ h(x) = a \cdot \frac{b}{\pi \cdot (b^2 + x^2)} + (1 - a) \cdot e^{-\frac{|x|}{c}} \cdot \frac{1}{2 \cdot c} \]

And the step response takes the form:

\[ E(x) = a \cdot \left( \frac{\arctan \left( \frac{x}{b} \right)}{\pi} + \frac{1}{2} \right) + (1 - a) \cdot e^{\frac{x}{c}} \cdot \frac{1}{2} \quad \text{for } x \leq 0 \]
\[ E(x) = a \cdot \left( \frac{\tan \left( \frac{x}{b} \right)}{\pi} + \frac{1}{2} \right) + (1 - a) \cdot (1 - e^{-\frac{x}{c}} \cdot \frac{1}{2}) \] for \( x \geq 0 \)

Here, \( b \) is analogous to \( \beta \) in the original model, \( c \) represents the decay constant of the exponential term, and \( a \) is the weighting between the two terms. Using this updated model, the fit of both the step response and its corresponding impulse response is much closer to the experimental results. In Fig. 26, the mean squared error is plotted for the original and updated step response fit. Averaged across all combinations of \( H \) and \( T \), the MSE for the updated fit is \( 6.75 \cdot 10^{-5} \text{um} \), 11.82% of that of the original MSE.

![Figure 26: Mean squared error of step response fit using original (left) and updated (right) form.](image)

In this updated model, the first term captures the impact of coating very close the feature, giving a good fit to the impulse response and to the feature scale variations, while the second term captures the impact of coating far away from the feature, improving the step response fit and capturing the longer range process effects, such as the transitions between uniform regions. The significance of these two terms is later explored in Section 3.7 after first discussing the results of the model Section 3.4.
3.3) Model Predictions

After developing the process model and fitting it to the measured step responses, the coated TV profiles are predicted. For each coated TV, the same profiles as shown in Section 2.2 are predicted. Sample coating predictions compared to their corresponding experimental scans can be seen in Fig. 28, and sample feature scale variation predictions compared to their corresponding experimental results can be seen in Fig. 29. The remainder of the predicted coatings, predicted feature scale variations, and both sets of experimental results can be seen in the appendix.
Figure 28: Experimental (left) and predicted (right) scans for $H = 5 \, \text{um}$, $T = 8 \, \text{um}$, $W = 4 \, \text{um}$ (top), $W = 64 \, \text{um}$ (bottom).
3.4) Model Accuracy

Ideally, the accuracy of the model would be quantified on a point-by-point basis, using either the mean-squared error or cross correlation value, but in practice, this is difficult. The first difficulty arises from the fact that the experimental scans are not level. Adding slopes to the experimental scans creates large differences between the two sets of scans, giving the model deceptively poor accuracy. A second problem with this approach comes from the angular offset between the two scans. The predicted scans are theoretical, and are perfectly perpendicular to the underlying lines; however, this is not true for the experimental scans. Due to this angular offset, the experimental scans are slightly stretched, as these scans must be longer to cover the same number of underlying lines. While both of these problems could be corrected for using optimization techniques, a simpler method is used to quantify the accuracy.
Instead, the models accuracy is assessed by comparing the predicted and experimental feature scale variations as defined in Section 2.2.1. For each combination of line width, spacing, height and polyimide thickness, the error is calculated by taking the difference between the predicted and experimental variations and normalizing by the line height.

The resulting assessment of the models accuracy for all combinations of line width, spacing, height, and polyimide thickness are shown in Fig. 30. The mean error for these predictions is 4.4%, and the standard deviation is 0.043. Additionally the maximum prediction error is 23.5%. This is a surprisingly high maximum prediction error, and exposes a limitation of the model. The model consistently under-predicts the variation for lines with small width and large spacings. However, this is only the case for large line heights and small polyimide thicknesses, and these extremal cases are relatively uncommon in RDLs.
3.5) Model Extremal Cases

To analyze the source of this error, consider a thin, isolated, interconnect, whose height \((H)\) is larger than the polyimide thickness on a flat surface \((T)\). Here, the copper line will be poking through the polyimide surface. Clearly, the surface variation cannot be less than the difference between the height of the interconnect and the polyimide thickness. However, by making the interconnect arbitrarily narrow, the model will predict an arbitrarily small surface variation. The
predicted coated profile will be:

\[ p_c(x) = \int p_u(x - x') \cdot h(x') \, dx' \]

\[ = \int_{-\Delta}^{\Delta} 1 \cdot h(x') \, dx \]

\[ = 0 \text{ for all } x \text{ as } \Delta \to 0 \]

Where \( h(x) \) is the impulse response, \( p_c \) is the coated profile, and \( p_u \) is the underlying profile whose height is 1 between \(-\Delta\) and \(\Delta\) and 0 elsewhere.

While this seems like a major limitation of the model, in practice, it is an unlikely extremal case. In order for this error to be significant, three conditions must be simultaneously met:

1) The line height must be larger than the line width.
2) The line height must be larger than the polyimide thickness.
3) The spacing of the lines must be significantly larger than the width of the lines (~2x).

The first condition is required for the model to predict approximately zero surface variation, as the width of the feature must be small. The second condition must be met in order for the lines to rise above the flat polyimide surface. While the error occurs when \( T = 8 \) µm, it is unseen when \( T = 13 \) µm. Additionally, while the nominal thickness is 8 µm for the thinner polyimide, the cross sectional SEM images in Figs. 18 and 19 show that the real value is closer to 6.5 µm, exacerbating this effect. Finally, the third requirement is necessary as lines with small spacing and line width do not show this error under any conditions.

3.5.1) Penalty Function for Extremal Cases

While these extremal cases are relatively uncommon in typical RDL applications, we propose a simple addition to the model that can improve predictions in these extremal cases and leave correct predictions unaffected. This improvement is not typically necessary for predicting surface variations in RDLs, however, it may be useful for alternative spin coating applications.

This correction to the model is applied after the standard predictions are made, and attempts to compensate for the extremal cases similar to those of Section 3.5. After predicting the surface using the standard model, we define the uncompensated dielectric thickness, \( r(x) \), at each point:
\[ r(x) = p_c(x) - p_u(x) \]

This is the distance from the underlying topography to the predicted surface. If this value is less than 0, then our model has predicted a negative dielectric thickness, and we know that our model has under-predicted the coated surface.

We then compute a penalty function \( f_p(r) \), and add this back to the prediction in order to compensate:

\[ p_c = p_c + f_p(r) \]

There are many options for choosing the correct form of \( f_p(r) \), but all viable forms must meet two requirements. First, \( p_c(x) > p_u(x) \) for all \( x \). This corresponds to the physical constraint of having the predicted coating be higher than the underlying topography. Secondly, if \( p_c(x) \gg p_u(x) \) then \( f_p(r) \approx 0 \). This corresponds to cases where the predicted coating is significantly higher than the underlying topography, as in regimes where the model already accurately predicts the surface. Therefore, these cases should remain unaffected by \( f_p(r) \).

While the optimal form of \( f_p(r) \) is not known, we propose a form which meets all of these requirements, and is successful in compensating for the extremal cases:

\[ f_p(r) = y_{anch} \cdot e^{-\frac{(r-x_{anch})^2}{d}} \]

The terms \( x_{anch} \) and \( y_{anch} \) are “anchor points” where the value of \( f_p(r) \) is approximately known. By choosing these points, we set \( f_p(x_{anch}) = y_{anch} \). To determine both values, consider two cases.

First, consider the case when \( H \gg T \). Here, the prediction is flat, but the underlying topography “pokes out” through our prediction, as illustrated in Fig. 31. In this case, we can assign \( x_{anch} = T - H \), and \( y_{anch} = H - T \). By setting this, we ensure that when \( r \approx T - H \) and the line pokes through the surface, our prediction is compensated by this distance, ensuring that \( p_c(x) > p_u(x) \).
Second, consider the case when $H \ll T$. Here, our predictions are always significantly higher than the underlying topography. In these cases, our unmodified model is correct, and thus we do not want to make significant changes. We can again set $x_{\text{anch}} = T - H$, but this time set $y_{\text{anch}} = 0$. By doing this, we ensure that $f_p(r) = 0$ for all $r$, and our correct predictions are not altered.

We can then combine these two cases by always setting $x_{\text{anch}} = T - H$, and could use $y_{\text{anch}} = \max(0, H - T)$. In order to use a smooth function for computing $y_{\text{anch}}$, we can instead use a smooth maximum:

$$y_{\text{anch}} = \log(1 + e^{H-T})$$

By choosing these points, we set $x_{\text{anch}}$ as the minimum uncompensated dielectric thickness possible. Because the predicted surface cannot be lower than $T$, and the underlying topography cannot be higher than $H$, the uncompensated thickness cannot be smaller than $T - H$. Therefore, this penalty function will have a maximum value of $y_{\text{anch}} \approx \max(0, H - T)$ in the worst case prediction (the flat prediction with an infinitely narrow interconnect analyzed in the previous section). Additionally, the impact of $f_p(r)$ will diminish as the uncompensated dielectric thickness increases, and as the original model becomes more accurate.

The final choice in determining $f_p(r)$ is then to determine the decay constant $d$. Here, we select a value which minimizes the MSE between the prediction and measurements over a complex topography. We optimize over all TVs used, and select the value $d = 25 \, \mu m^2$. While it is likely the case that each set of process parameters should have its own corresponding value of $d$, we did not find it necessary to optimize $d$ separately for each coated TV to show the effectiveness of the
penalty function.

By using this version of $f_p(r)$, we reduce the max feature scale error over all TV’s from 23.5% to 12% of the line height, and the average error from 4.4% to 3.5%. This is a significant improvement, especially in the extremal cases where the unmodified version breaks down. In Fig. 32, we present three sets of scans that each show the experimental, unmodified, and modified predictions. Example 1 shows the profiles for $H = 9 \, \mu m$, $T = 8 \, \mu m$ and $W = 4 \, \mu m$. Example 2 shows the profiles for $H = 7 \, \mu m$, $T = 8 \, \mu m$ and $W = 4 \, \mu m$. Example 3 shows the profiles for $H = 5 \, \mu m$, $T = 13 \, \mu m$ and $W = 16 \, \mu m$. Two of the three scans show an improvement, while the third shows a relatively correct and unmodified prediction.

<table>
<thead>
<tr>
<th>Example</th>
<th>Experimental</th>
<th>Prior</th>
<th>Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1" alt="Experimental Scan" /></td>
<td><img src="image2" alt="Prior Scan" /></td>
<td><img src="image3" alt="Modified Scan" /></td>
</tr>
<tr>
<td>2</td>
<td><img src="image1" alt="Experimental Scan" /></td>
<td><img src="image2" alt="Prior Scan" /></td>
<td><img src="image3" alt="Modified Scan" /></td>
</tr>
<tr>
<td>3</td>
<td><img src="image1" alt="Experimental Scan" /></td>
<td><img src="image2" alt="Prior Scan" /></td>
<td><img src="image3" alt="Modified Scan" /></td>
</tr>
</tbody>
</table>

*Figure 32:* Experimental, prior, and modified scans using penalty function. First two (1,2) show desired change, and the Third (already correct) remains unaffected.

Additionally, we can see the improvement in the accuracy from error plots first shown in Section 3.4. In particular, the cases of $T = 8 \, \mu m$ and $H = 7$ or $9 \, \mu m$ show significant...
improvement. In Fig. 33, we include both the previous errors, and the errors after implementing the penalty function. We note that for all other cases, the errors were not significantly affected as $H < T$ for these cases.

In Fig. 33, we include both the previous errors, and the errors after implementing the penalty function. We note that for all other cases, the errors were not significantly affected as $H < T$ for these cases.

**Figure 33:** Uncompensated (left) and corrected (right) feature scale errors for $T = 8 \mu m$ with $H = 7 \mu m$ (top), and $T = 8 \mu m$ with $H = 9 \mu m$ (bottom).

While the inclusion of the penalty function does improve the extremal cases, we note that its inclusion is not required, and for the remainder of this thesis, we will ignore it for simplicity. Finally, we note that while the form of $f_p(r)$ does improve results, our proposed form is likely sub-optimal, and further improvements to its form can be made.

### 3.6) Model Trends

The $a, b, \text{and } c$ coefficients present in our modeled system responses empirically encapsulates the effects of all spin coating parameters, including line height, coating thickness, spin speed, and the material properties of the polyimide. The measured coefficients as a function of coating thickness and step height can be seen below in Fig. 34.

Ideally, these parameters could be predicted from material and procedural constants. If so,
the optimum polyimide, line height and spin speed could be selected to minimize surface variation while meeting other design criteria. Unfortunately, the extracted parameters do not show any clear trends from which an empirical model can be developed that would predict the model coefficients from material and procedural constants.

While we cannot quantitatively predict the change, we believe that increasing the coating thickness will increase the $b$ coefficient, as this effect is seen in all cases where the line height is kept constant. This increase in $b$ leads to a more planar coated surface, as the step response term $\tan\left(\frac{x}{b}\right)$ will be stretched horizontally. Intuitively this makes sense, as increasing the polyimide thickness should smooth out the underlying topography more heavily.

Besides this, there are no other clear trends in the model coefficients. While $b$ appears to decrease with line height for $T = 8 \, \mu\text{m}$, it is constant with line height for $T = 13 \, \mu\text{m}$. Additionally, both $a$ and $c$ seem to be constant with both line height and coating thickness; however, there is not enough data to fully confirm this. We believe that the lack of clear coefficient trends is likely due to differences in the curing procedure between TVs. This effect is later revisited and explained in Section 3.11.

![Figure 34: $a$, $b$, and $c$ coefficients as a function of line height for $T = 1.8 \, \mu\text{m}$ (blue), and $13 \, \mu\text{m}$ (red).](image)

**Figure 34:** $a$, $b$, and $c$ coefficients as a function of line height for $T = 1.8 \, \mu\text{m}$ (blue), and $13 \, \mu\text{m}$ (red).

### 3.7) Impact of Updated Fit and Coefficient Significance

As previously described in Section 3.2.2, the original step and impulse response fits were
updated to include an additional term to address long-range variation effects. Now, the significance of this update and of the model terms is investigated. First, sample predictions using the original model are shown and compared to their updated versions. Then the impact of both system response terms is analytically examined both near the feature, and far away from the feature. Together, the predictions and the closed form analysis will give intuition for each model term.

In the original model, the impulse response takes the form:

\[
h(x) = \frac{b}{\pi (b^2 + x^2)}
\]

While this fits well near the feature itself, its accuracy commonly degrades outside of this small range as can be seen in Fig. 35.

![Figure 35: Impulse response showing good fit near center, while the accuracy degrades further away.](image)

Using the simpler version of the model, the predictions lack many of the long range effects present in both the experimental scans, and the predictions using the updated fit. Fig. 36 shows four profiles that highlight the importance of the model update. Fig. 36a shows the experimental scan; Fig. 36b shows the prediction using the original impulse response fit; Fig. 36c shows the prediction using the updated impulse response fit; finally, Fig. 36d shows the prediction using the experimental impulse response. In this case the raw, unfit impulse response is used during the
convolution. For all figures, \( H = 5 \ \mu m \), \( T = 13 \ \mu m \), and \( W = 32 \ \mu m \). The spacings change within each graph, and are 16 \( \mu m \), 8 \( \mu m \), 32 \( \mu m \), 4 \( \mu m \), 64 \( \mu m \), and 2 \( \mu m \) in that order.

![Profile graphs](image)

**Figure 36:** Sample experimental scan (upper left), prediction using original form (upper right), prediction using updated form (lower left), prediction using experimental impulse response (lower right).

In all profiles, the feature scale variations within each region are approximately the same. The most significant difference between these profiles is that the profile predicted using the original fit lacks the gradual transitions between uniform regions. This signifies that the \( \frac{b}{\pi(b^2 + x^2)} \) term captures the feature scale variations, while the \( e^{-\frac{|x|}{c}} \cdot \frac{1}{2c} \) term captures the long range, or chip scale, variations.
Now, the effect of each term is analytically examined near the feature and far away from the feature in order to show the impact of each. Here, it is assumed that $b \ll c$, as this is the case in all observed fits. At the feature itself ($x = 0$), the feature scale term has a value $\frac{b}{\pi (b^2 + 0^2)} = \frac{1}{\pi b}$, and the long range term has a value $e^{-\frac{|0|}{c}} \cdot \frac{1}{2 \times c} = \frac{1}{2c}$. As $b \ll c$, it is clear that the feature scale term dominates near $x = 0$. Additionally, when $x \approx c$, relatively far away from the feature, the feature scale term has a value $\frac{b}{\pi (b^2 + c^2)} \approx \frac{b}{c^2}$, and the long range term has a value $e^{-\frac{|c|}{c}} \cdot \frac{1}{2c} \approx \frac{1}{c}$. In this case, $b \ll c$, thus $\frac{b}{c^2} \ll \frac{1}{c}$, and in this case the long range term dominates. This analysis demonstrates why the original model can predict the feature scale variations, while it was unable to predict the gradual transitions between regions.

Finally, it is also worth noting that the profiles using the updated fit and experimental impulse responses are almost identical. This is expected, as the MSE between the two step responses is $6.75 \cdot 10^{-5}$ um. Therefore, the impulse responses and predicted profiles are similarly close, and further updates to the fit are not necessary.

However, the similarity between the predictions using the updated fit and those using experimental data raises an interesting question: Why is fitting necessary if using the experimental impulse response results in a similarly accurate prediction? The answer to this is twofold. First, by fitting the data to a compact model, the amount of information needed to quantify the process is greatly reduced. Using the fit, only three values are needed to quantify the system, one for each of the model coefficients. However, using the experimental data, a value for every sampled point would be needed. Therefore, the model fit provides a convenient way to compare which process will result in more planar structures. Secondly, by fitting the impulse responses, the noise from the experimental data is reduced. This will be important when the model is expanded to two dimensions and 2D impulse responses are generated from the 1D responses. The algorithm used for this has a low noise tolerance and will be expanded upon in Section 3.14.

### 3.8) Closed Form Analysis

Here, closed form results are developed for regions with uniform line height, spacing, and width. This will explain the effective feature size approximation first introduced in Section 2.2.1,
and will help establish design rule guidelines which will later be expanded upon in Chapter 4. In this analysis, only relatively small feature sizes are considered, particularly those where $w, s \ll c$, the coefficient related to the long range variations.

To approach this problem, first consider the coating over an individual feature as shown in Fig. 37.

![Figure 37: Ideal coating over an isolated interconnect.](image)

To find the variation in the coated profile, $p_c$, the minimum and maximum heights of the coated profile, $P_{\text{min}}$ and $P_{\text{max}}$, must be determined. The minimum height of the coated profile occurs infinitely far away from the feature, and $T$ $\mu$m above the wafer surface. Using the original convention, this point has an absolute height of zero. Additionally, $x = 0$ is the point where $p_c(x) = P_{\text{max}}$.

Because the model takes the form of a LTI system, the underlying profile, $p_u(x)$, can be separated into two parts, a step up, followed by a step down. Then, the resulting profiles can be summed to find $p_c(x)$. Here, $u(x)$ refers to the Heaviside (step) function.

$$p_u(x) = H \cdot \left( u \left( x + \frac{w}{2} \right) - u \left( x - \frac{w}{2} \right) \right)$$

Then, using the closed form step responses:
\[ P_{\text{max}} = p_c(0) = H \left( a \cdot \frac{\tan \left( \frac{w}{2b} \right)}{\pi} - a \cdot \frac{\tan \left( -\frac{w}{2b} \right)}{\pi} \right) + (1 - a) \left( 1 - \frac{e^{-w/c} \cdot 1}{2} \right) - (1 - a) \cdot \frac{w}{2c} \cdot \frac{1}{2} \]

Because \( x \ll c \), this can be approximated by:

\[ P_{\text{max}} \approx H \left( a \cdot \frac{\tan \left( \frac{w}{2b} \right)}{\pi} - a \cdot \frac{\tan \left( -\frac{w}{2b} \right)}{\pi} \right) \]

Finally, as \( \tan(x) \) is an odd function, \( \tan(-x) = -\tan(x) \) and:

\[ P_{\text{max}} = \frac{2 \cdot H \cdot a}{\pi} \cdot \frac{w}{2b} \]

This gives a closed form solution to the variation from a single feature with \( w \ll c \). In Fig. 38, this approximation is plotted against the experimental variations as a function of \( W_{\text{eff}} \) for \( H = 4 \ \mu m \), and \( T = 13 \ \mu m \). As previously described, \( W_{\text{eff}} = \frac{w \cdot s}{w + s} \) and is the effective width for a given line width and spacing. Here, the variation from a uniform region of lines with width \( w \) and spacing \( s \) should match the variation caused by an isolated feature of width \( W_{\text{eff}} \).

![Figure 38: Closed form approximation for feature scale variation vs. \( W_{\text{eff}} \).](image_url)

This closed form analysis is now expanded to the coatings over an infinite array of lines,
similar to those in the TV, as illustrated schematically in Fig. 39.

![Diagram of ideal coating over an infinite array of lines]

**Figure 39**: Ideal coating over an infinite array of lines.

Again, the feature scale variation is found by taking the difference between $P_{\text{min}}$ and $P_{\text{max}}$.

First, the starting profile is broken apart into individual steps:

$$p_u(x) = H \sum_{i = -\infty}^{\infty} u \left( x + \frac{w}{2} + i \cdot (w + s) \right) - u \left( x - \frac{w}{2} + i \cdot (w + s) \right)$$

The simplification $w \ll c$ is again used to approximate the coated profile:

$$p_c(x) \approx \frac{H \cdot a}{\pi} \sum_{i = -\infty}^{\infty} \tan \left( \frac{x + \frac{w}{2} + i \cdot (w + s)}{b} \right) - \tan \left( \frac{x - \frac{w}{2} + i \cdot (w + s)}{b} \right)$$

Again, $P_{\text{max}}$ occurs directly above the feature at $x = 0$, but now $P_{\text{min}}$ occurs in the spaces between features, at $x = \frac{w + s}{2}$:

$$P_{\text{max}} = \frac{H \cdot a}{\pi} \sum_{i = -\infty}^{\infty} \tan \left( \frac{\frac{w}{2} + i \cdot (w + s)}{b} \right) - \tan \left( \frac{-\frac{w}{2} + i \cdot (w + s)}{b} \right)$$

$$P_{\text{min}} = \frac{H \cdot a}{\pi} \sum_{i = -\infty}^{\infty} \tan \left( \frac{-\frac{s}{2} + i \cdot (w + s)}{b} \right) - \tan \left( \frac{\frac{s}{2} + i \cdot (w + s)}{b} \right)$$

52
The feature scale variation is the difference between these minima and maxima:

\[
P_{\text{max}} - P_{\text{min}} = \frac{H \cdot a}{\pi} \sum_{i = -\infty}^{\infty} \text{atan} \left( \frac{w + 2i \cdot (w + s)}{2b} \right)
- \text{atan} \left( \frac{-w + 2i \cdot (w + s)}{2b} \right)
- \text{atan} \left( \frac{-s + 2i \cdot (w + s)}{2b} \right)
+ \text{atan} \left( \frac{s + 2i \cdot (w + s)}{2b} \right)
\]

Next, this can be simplified by recognizing that for every \(i\) in the summation, there also exists \(-i\):

\[
P_{\text{max}} - P_{\text{min}} = \frac{2H}{\pi} \cdot \sum_{i = -\infty}^{\infty} \text{atan} \left( \frac{w + 2i \cdot (w + s)}{2b} \right) + \text{atan} \left( \frac{s + 2i \cdot (w + s)}{2b} \right)
\]

Finally, using the identity \(\text{atan}(a) + \text{atan}(b) = \text{atan}\left(\frac{ab-1}{a+b}\right)(\text{mod}\ \pi)\):

\[
P_{\text{max}} - P_{\text{min}} = \frac{2 \cdot H \cdot a}{\pi} \sum_{i = -\infty}^{\infty} \text{atan} \left( \frac{2 \beta \cdot \left( \frac{(w + 2i \cdot (w + s)) \cdot (s + 2i \cdot (w + s)) - 1}{4b^2} \right)}{(w + 2i \cdot (w + s) + s + 2i \cdot (w + s))} \right)
\]

\[
= \frac{2 \cdot H \cdot a}{\pi} \sum_{i = -\infty}^{\infty} \text{atan} \left( \frac{(w + 2i \cdot (w + s)) \cdot (s + 2i \cdot (w + s)) - 4b^2}{2b \cdot (4i + 1) \cdot (w + s)} \right)
\]

There are likely other approximations which can be made which will further simply this expression to the form \(P_{\text{max}} - P_{\text{min}} \approx f(W_{\text{eff}})\); however, these are currently unknown. Most likely this simplification will use the same identity, \(\text{atan}(a) + \text{atan}(b) = \text{atan}\left(\frac{ab-1}{a+b}\right)\), to simplify the infinite sum. However, from this infinite summation, the emergence of the \(\frac{w+s}{w+s}\) term can be seen inside of the atan function.

Instead of further simplification, this form is normalized by \(H\) and \(a\), and plotted as a function of \(W_{\text{eff}}\) for the common value of \(b = 8 \mu m\) in Fig. 40. Here, both \(w\) and \(s\) are swept from 1 \(\mu m\) to 100 \(\mu m\), and each line in the graph represents a different line width. Finally \(i = -99: 100\)
is used to approximate the infinite summation. This shows that while the feature scale variation is not an exact function $W_{\text{eff}}$, it is a reasonable approximation, and the form developed for isolated features can be used as a rule of thumb for any width and space combination where $w, s < c$:

$$P_{\text{max}} - P_{\text{min}} \approx \frac{2 \cdot H \cdot a}{\pi} \cdot \text{atan} \left( \frac{W_{\text{eff}}}{2b} \right)$$

![Normalized Variation for b = 8 um](image)

*Figure 40: Ideal feature scale variation vs. $W_{\text{eff}}$ for all reasonable spacing and width combinations.*

3.9) Existing DSC Modeling Revisited

In this section, the DSC models from existing literature and discussed in Section 1.2 are revisited. The most significant prior contribution comes from Stillwagon and Larson [8]. In this model, Stillwagon and Larson develop a model for non-volatile fluids that cure during the coating process. Using this model, they predict that the system step response can take one of two forms depending on the feature orientation. For steps facing the center of the wafer (the direction of the step is from the low side to the high side), the step response should be:

$$E(x) = \frac{H}{2} \cdot e^{\lambda x} \quad \text{for } x < 0$$

$$E(x) = H \cdot \left(1 - \frac{e^{-\lambda x}}{2}\right) \quad \text{for } x > 0$$

where
\[
\lambda = \left( \frac{3 \rho \omega^2 r_0}{\gamma t_b} \right)^{1/3}
\]

\(\rho\) is liquid density, \(\omega\) is spin speed, \(r_0\) is distance from the center, \(\gamma\) is film surface tension, \(t_b\) is thickness before curing (on a flat surface), \(t_f \approx t_b \cdot s\), and \(s\) is the percentage of solid material in solution. Taking the derivative of their proposed step response and normalizing by the line height gives the impulse response:

\[
h(x) = \frac{\lambda e^{-\lambda|x|}}{2}
\]

And taking the Fourier Transform of this gives the spatial frequency response:

\[
\mathcal{H}(k) = \frac{\lambda}{\pi (\lambda^2 + x^2)}
\]

This impulse response and the one presented in this thesis are now compared. In particular, the values of each at \(x = 0\) are assumed to be equal. This will allow the coefficients in both models to be compared.

As \(b \ll c\), the following simplification is made to the model presented in this paper:

\[
h(0) = a \cdot \frac{b}{\pi \cdot b^2} + (1 - a) \cdot \frac{1}{2 \cdot c} \approx \frac{a}{b}
\]

Additionally, because \(a\) has been shown to be constant across all parameters investigated, this can be further approximated:

\[
h(0) \propto \frac{1}{b}
\]

And in the Stillwagon and Larson model,

\[
h(0) \propto \lambda
\]

Therefore,

\[
b \propto \frac{1}{\lambda}
\]

Or alternatively:
This analysis suggests two phenomena. First, according to the Stillwagon and Larson model, the $b$ coefficient should decrease with increasing radial distance from the center of the wafer. Secondly, the shape of the step (and therefore the value of the fit $b$ coefficient) should change based on the orientation of the step relative to the center. This second point comes from the fact that the step response would take a different form (it is similar, but not included in this analysis) based on the orientation.

3.10) Effects of Feature Location and Orientation

The effects of the location and orientation of the experimentally coated profile is now investigated. Here six test structures are fabricated, with step heights 2 $\mu$m and 4 $\mu$m, and coating thicknesses of 8 $\mu$m, 11 $\mu$m, and 13 $\mu$m. The step response is measured at three different distances from the center (1, 3, and 5 cm), and four different orientations (0°, 90°, 180°, and 270° from the radial direction). Here, the direction of the feature is defined as the vector from the low side to the high side of the line array, and 0° indicates a feature pointing towards the top of the wafer. Fig. 41 illustrates the location and orientation of the features measured. Fig. 42 shows the measured $b$ as a function of radial distance, averaged across all orientations, and as a function of orientation, averaged across all radial distances.

\[ b \propto \left( \frac{\gamma t_b}{\rho \omega^2 r_0} \right)^{\frac{1}{3}} \]

Figure 41: Location and orientation of measurements. The red line indicates the location and direction of each measured step.
The results of this analysis suggest that neither position nor orientation have any significant effect on the coefficients of the model, or on the coated features in our experimental data. This simplifies the model, as a single impulse response can be used to predict the entire coated profile, regardless of these factors. However, these results also suggest that the Stillwagon and Larson analysis does not accurately describe the physics of the DSC process under the experimental conditions used in this thesis.

While the forms of the Stillwagon and Larson system responses are similar to the ones proposed in this thesis, the two are not identical. Interestingly, the forms of the two proposed impulse and frequency responses are inverted. To determine which one is more accurate, the fit of each proposed frequency response is plotted in Fig. 43. While similar, the fits suggest that the form of the system response proposed is this thesis are more accurate.
3.11) Effects of Coating vs. Curing

The discrepancies between the Stillwagon and Larson model and the results of this thesis can be explained by the different curing procedures. In their experiments, Stillwagon and Larson spin coat an epoxy that cures while the wafer spins. In the case of this thesis, the polyimide is first spun on, then cured in a separate stage, up to several minutes later. This suggests that the DSC process is highly impacted by the curing process. In fact, Stillwagon and Larson recognize that their model only applies during the spin coating process itself, and not for processes that require a separate curing stage [8]. Additional work has been conducted which investigates evaporation during spin coating, and suggests that the evaporation profile will influence the final coating [21, 22]. However, there has been limited work that investigates a true two stage spin coating process. Therefore, the model proposed in this thesis is primarily intended for spin coating processes that require a cure well above room temperature (300º C for the HD-4110).

3.11.1) Experimental Results

Here, profiles of the coated wafer curing are shown both after spin coating but before curing, and again after curing. Traditional contact profilometry cannot be used in this case as the wafers are still wet. Instead, they are optically profiled using a WYKO NT3300 interferometer. Fig. 44 shows scans before curing of areas similar to those done in Section 2.2, while Fig. 45 shows scans of coated steps before curing. Finally, Fig. 46 shows a sample optical profile after curing.
Figure 44: Optical scans of coated TV before curing for $W = 64$ um, $S = 16, 8, 32, 4$ um (both) and $H = 4$ um, $T = 8$ um (left) and $H = 9$ um, $T = 13$ um (right).

Figure 45: Optical scans of coated steps before curing for $H = 4$ um, $T = 8$ um (left) and $H = 9$ um, $T = 13$ um (right).
Before discussing the significance of these scans, it is worth noting the differences between these profiles and those measured with the contact profilometer. Here, the long scans, Figs. 44 and 46, are comprised of approximately 80 smaller scans stitched together. During the stitching process, it is common for the measurements to drift over time. This explains the long term drifting which was not seen in the more reliable profilometry scans.

The results of this experiment suggest that the uncured coated profile is essentially flat. The surface variations for the long scans, as seen in Fig. 44 are consistently under 0.1 µm before curing. It is not until after curing that the variations reach values common in the post-curing profilometry scans as seen in Fig. 46. The same is true for the step responses in Fig. 45. If the coating stage dominated the final profile, a height difference of up to 4 µm and 9 µm should be seen in the two step responses, respectively. Instead, the maximum height differences are approximately 0.1 and 0.3 microns, a factor of 30 less than what is expected if the coating stage of the process determines the final surface variation.

These scans strongly suggest that the polyimide coating is flat before curing, and that the curing process dominates the final profile. This also explains the discrepancies between the
previous experimental results and the Stillwagon and Larson analysis. Finally, it should be noted that these coated TVs are not cured or optically profiled immediately after spin coating. There is a brief period after spin coating which is needed to calibrate the optical profiler and to bring the oven up to temperature. It is possible that during this period the (liquid) coating formed during the spin process itself has a chance to planarize.

3.11.2) Hirasawa Analysis

While models exist that incorporate evaporation during the spin coating process [9, 23], little research has been conducted describing the curing process after the coating has stopped, and the dielectric has planarized. One model for the curing process, proposed by Hirasawa et al. [10], does attempt to explain this phenomenon. In this analysis they assume that the coating thickness is uniform after spin coating. Then the profile is leveled due to surface tension while the dielectric cures. This takes the form of the P.D.E.:

$$\frac{\partial z_1}{\partial t} = -\frac{\partial}{\partial x} \left\{ \frac{(z_1 - z_0)^3}{3\mu} \left( \frac{\partial}{\partial x} \left[ \sigma \frac{\partial^2 z_1}{\partial x^2} \left( \frac{3}{2} \right) \right] \right) \right\} - v$$

where $z_0$ is the underlying topography, $z_1$ is the coating height (including $z_0$), $\mu$ is the viscosity, $\sigma$ is the surface tension, $v$ is the evaporation rate and $\mu$ and $v$ are proportional to the solvent content $y$.

This model is simulated using the Crank-Nicholson method and it is determined that the surface should planarize well before curing should occur. Using the physical constants of the HD-4110, and evaporation rates similar to those in the referenced paper, the leveling process is simulated for periodic lines having a width and spacing of 200 µm, height of 5 µm, and final polyimide thickness of 8 µm. The first 4.2 seconds of this process can be seen in Fig. 47, after which the surface profile remains constant. Only half of the profile is shown, as it is symmetric. The time to fully cure the dielectric is on the order of 1 to 5 minutes depending on coating thickness and other process parameters.
Figure 47: Simulated leveling over 5 seconds using Hirisawa et al. model.

This model supports the hypothesis that separating the coating and curing stages allows the profile to first planarize, and then generate the surface topography during curing. Thus the curing process dominates the final profile. However, there are also significant differences between the previously reported experimental results and the results of this model. According to this model, if the dielectric is allowed to planarize, and is later cured, the final profile remains flat regardless of the underlying topography. This can be seen from the P.D.E. Once the profile has flattened, $\frac{\partial^2 z_1}{\partial x^2} = 0$ across the entire profile, and $\frac{\partial z_1}{\partial t} = -v$. This indicates that the profile is shrinking constantly in the vertical direction with a rate equal to the evaporation rate. This effect was not previously seen, as the experimental profiles partially conform to the underlying profile.

The discrepancy between this model and the experimental results likely comes from one of the simplifying assumptions in the model. They assume that the solvent content is constant throughout the dielectric, however this is not the case. Consider a step coated in a planar, uncured dielectric, such as that in Fig. 47. As this profile cures, the solvent initially evaporates evenly across the surface. However, if an equal volume of solvent is taken from above the step and from above the trench, there is proportionally less solvent left above the step. This should decrease further evaporation rate above the step. However, this effect is neglected in the model, as the solid
content is assumed to be uniform throughout the dielectric, leading to the discrepancy between the experimental results and the predictions in this model. In reality, the solvent diffuses over time, but this is neglected in the Hirisawa et al. model.

Additionally, the shape of the predicted impulse response is not consistent with the previously reported experimental results, and is shown in Fig. 48. This simulated impulse response does not resemble the form presented in Section 3.2, but instead appears to take the form of a triangle wave. If the form was similar, this model could be used to predict the $a$, $b$, and $c$ constants based on the process parameters; however, this is not the case.

![Sample simulated impulse response](image)

**Figure 48:** Simulated impulse response using Hirisawa et al. model.

While the Hirisawa et al. model cannot reliably predict coated profiles, it does give an order of magnitude estimate for the time it takes for the polyimide to level without curing (less than a minute). This gives additional evidence that the curing, as opposed to the spin process itself, is the dominant force in determining the final profile in the DSC process.

Finally, the Hirisawa et al. model gives additional insight into the process parameters that influence the $a$, $b$, and $c$ parameters. Besides material properties which are already assumed to play a role, the evaporation rate has a large impact on the final profile in this model. The evaporation rate must be a function of temperature, as the polyimide does not solidify at room temperature. Therefore, this model suggests that the exact temperature profile that the dielectric is subjected to will also have a large impact on the final profile. Gradually raising the temperature over time will
slowly remove the solvent, allowing the dielectric more time to level in the process, and resulting in a more planar profile. In contrast, quickly raising the temperature to the same maximum value will drive out the solvent more quickly, giving a relatively conformal profile.

This suggests that the cause of the seemingly erratic model coefficient trends in Section 3.6 may be due to differences in the exact curing procedure. While line height and polyimide thickness do likely play a role in the coefficient values, the curing temperature profile was not well controlled during these experiments, skewing the model coefficient trends. However, the profiles of the coated TV and their corresponding step responses were taken from the same wafer, thus both were subjected to identical curing conditions. Therefore, the same model coefficients extracted from the step response for any given TV wafer can be used for predicting the profiles at all locations on that coated TV. Further work should revisit the relationship between the model coefficients, the curing temperature profile, and the other process parameters.

3.12) Coatings Over Non-Binary Topographies

The experiments in Chapter 2 only consider coatings over topographies with binary heights (either raised by a constant amount, or not raised at all). While this is often the case for the first layer of the RDL, subsequent layers do not follow this form, as variations from either the spin coated surface, or the copper lines themselves, are carried through to subsequent layers, as illustrated schematically in Fig. 49. In this section, the model is applied to non-binary starting topographies, and its accuracy is confirmed under these extended conditions.

Figure 49: Non uniform starting topography for P1 leads to non uniform dielectric after coating (left). Spin coating the second level dielectric (P2), sees a complex starting topography due to variation in P1 thickness, even if all metal lines uniform (right). These two effects may be present simultaneously and compound with increasing layers (not shown).
To explore the effect of complex starting topography, two experiments are presented. First a four-leveled structure is fabricated by first re-creating the original TV. Then two additional copies of the TV are fabricated above the previous layer. For each additional layer, the structure is shifted over by a fixed amount to increase the variation in the test structure. In this structure, the layers are labeled M1, M2, M3 and P1 as illustrated conceptually in Fig. 50.

![Figure 50: Top perspective of M1, M2, M3 (P1 not shown) (left) and cross section including P1 (right).](image)

The second experiment is similar to the first, except polyimide is spin coated between each copy of the TV. In this experiment, only one additional copy of the TV is used. Beyond this, lower layers of the test structure routinely peeled off due to the internal stresses in the structure. The layers in this experiment are labeled M1, P1, M2, and P2 and are illustrated in Fig. 51.

![Figure 51: Cross section of M1, P1, M2, P2 structure.](image)

In these experiments, the test structures are no longer fabricated using an oxide deposition and etch to mimic the copper lines. Instead, the interconnects are formed using semi-additive electroplating. While these structures more closely match real RDLs, they are also significantly less uniform than the previous TVs. One reason for this is that the electroplating process has its
own layout dependencies, as the copper growth rate is impacted by the structure pattern density [12]. Secondly, the electroplating process did not consistently resolve line spacings of 8 µm or less. We suspect this is due to the copper sulfate in the plating solution etching the photoresist during electroplating. This phenomenon also creates line spacings larger than their nominal values. Because many of the smaller feature do not resolve, the length of the experimental scans is increased to include additional features. Each of the measured profiles still have a unique, fixed width, but the range of spacings within the scans is increased.

While these non-uniformities can be seen as a positive, as they enable us to test the model against even more complex structures, the starting topographies cannot be predicted, as this needs a complete model for the ECP process. For this reason, the post-coatings thicknesses cannot be predicted from only the masks. Instead, the structures are profiled before spin coating, then this experimental profile is used as the starting topography when predicting the resulting coating. Therefore, the effect of minor errors in the model cannot be seen to accumulate over multiple coating layers.

The following sets of predictions use the same convolution model as presented in Section 3.1, however, in these predictions, the model $a$, $b$, and $c$ parameters are adjusted. Using the same parameters as measured in the previous sections results in sub-optimal predictions. Additionally, the large non-uniformities in the copper growth, especially near the edges of large features, prevent clean steps where the model coefficients can be extracted. Instead, the coefficients are chosen by hand to best match the complex profiles. Ideally, an optimization method such as gradient descent would select the optimal coefficients, however this is difficult as computing the error on a point-by-point basis is difficult for the reasons previously described in Section 3.4. The coefficients chosen are $a = 0.4$, $b = 12 \mu m$ and $c = 300 \mu m$ for the first data set (M1-M3, P1), and $a = 0.6$, $b = 25 \mu m$, and $c = 350 \mu m$ for the second data set (M1, P1, M2, P2).

The necessity to change the model coefficients between runs again suggests that these coefficients are heavily impacted by the curing temperature profile, which was not well controlled. Alternatively, the change in optimal coefficients may be due to the change in line height. This seems less likely, however, as it is now shown that a single set of coefficients is able to accurately predict the final profiles for a wide variety of line heights within a single coating.
Fig. 52 shows an example scan for the M1, M2, M3 underlying topography, while Fig. 53 shows the experimental coated profile, and its predicted coated profile. Below these, we include example scans from the M1, P1, M2, P2 experiments. Fig. 54 shows an example scan for the underlying topography of M1 and M2 while Figs. 55 and Figs 56 show the experimental coated profile, and the corresponding predicted coated profile.

Figure 52: Sample underlying profile for M1, M2, M3.

Figure 53: Sample experimental (left) and predicted (right) profiles for M1, M2, M3, P1.
Figure 54: Experimental underlying profiles, M1 (left) and M2 (right).

Figure 55: Experimental (left) and predicted (right) P1 layers.

Figure 56: Experimental (left) and predicted (right) P2 layers.
In these two sets of experiments, there is significant wafer warpage after spin coating, which was previously unseen. We believe this is due to the recrystallization of copper during the curing process, and is corrected for in post processing of the scanned profiles. To correct the scans, three points are selected whose distances above the wafer are known to be relatively equal. The points are then fit to a second order polynomial, and this polynomial is subtracted from the measured profiles. The profiled areas between 4000 µm and 5000 µm, between 8000 µm and 9000 µm, and finally between 12000 µm and 13000 µm are known to always be solid copper, as the ECP process was unable to resolve the fine spacings in these areas. An example of a raw scan can be seen in Fig. 57, and can be compared to the corrected coated profile in Fig. 58. While the resulting profiles are not perfectly flattened, they more accurately represent the distance from the silicon wafer to the surface of the polyimide.

![Coated Profile for w =16 µm](image1)

![Coated Profile for w =16 µm](image2)

*Figure 57: Raw, un-flattened sample scan (left) and corresponding flattened scan (right).*

Unfortunately, quantifying the model accuracy for these experiments is difficult. Using either the mean-squared error or cross correlation functions suffer from the same problems as described in Section 3.4, and these problems are exacerbated by the wafer warpage. Additionally, both error functions are more significantly impacted by the average heights of areas, compared to the feature scale variations. As both are of interest, neither point-based function can properly quantify the model accuracy. In the future, improvements to quantifying the model accuracy will be investigated. In the meantime, the model can only be qualitatively assessed.
While the model predictions generally mirror the experimental results, there are some discrepancies. Some of these can be explained as artifacts of the experiments, while others may be limitations of the model. One such discrepancy occurs commonly in the high, flat areas which are used for de-warping. Here, bumps are present in the experimental scans which do not appear in the predictions, for example in Fig. 53. These are small bubbles formed either in the polyimide itself, or possibly in between the polyimide and the copper. It is unclear as to why these do not appear in the original data sets, but they are not believed to be limitations of the model.

Secondly, the right hand edges of the M1, M2, M3, P1 predictions drop off faster than the experimental results, as seen in Fig. 56. This is due to long range interactions with the unknown topography to the right of these profiles which are not taken into account in the predictions. In reality, additional features exist beyond the range of our starting topography scan; however, they are intentionally excluded from the predictions. These areas contain vias, which are 2D features, and are difficult to profile using 1D profilometer. Therefore, the topography is assumed to be flat beyond the scans, thus skewing the predictions near the edges.

One discrepancy that may be a limitation of the model is the average slope of some of the regions in the profiles. The model occasionally under predicts this slope, losing some of the long range process effects, as is seen in 56. While this may be an issue of choosing the optimal coefficients, particularly the long range \( c \) coefficient, a better fit could not be found. Perhaps better coefficients could be chosen using true step responses, or through more robust optimization techniques.

Regardless of these discrepancies, the model captures both the long range as well as the feature scale variations. The relative heights between the major peaks and troughs of the predicted profiles are typically within 5-10% of those in the corresponding experimental scans. Additionally, the predicted feature scale variations match the experimental results for both large and small starting heights.

3.13) Model Sensitivity

In order to explain the generalization of the model to coatings over non-binary heights, we analyze the sensitivity of the model to its chosen coefficients. In Fig. 58, the normalized feature scale variation is plotted versus \( W_{\text{eff}} \) for common values of \( b \) using the technique described in
Section 3.8. Additionally, the difference between the variations for \( b = 5 \mu m \) and \( b = 10 \mu m \) is also included to show the error if the coefficient is incorrectly chosen. These two values of \( b \) represent upper and lower bounds for observed values of \( b \), and the difference between the resulting variations represents the maximum error possible as long as the coefficients are chosen within reason.

Figure 58: Feature scale variation vs. \( W_{eff} \) for reasonable values of \( b \) (left), and corresponding maximum error (right).

In addition to estimating the error for incorrectly chosen \( b \) values, Fig. 58 also explains the ability to predict coatings over non-binary starting heights. Given that the observed \( b \) values in Section 3.6 are not constant across different TVs with different line heights, it is conceivable that a single impulse response and convolution may not be able to accurately predict the entire coating for multi-height TVs. However, given the results of the multi-level predictions, it seems that this is not the case, and we get good predictions across the range of resolved feature heights. Upon closer inspection, the results of the sensitivity analysis suggest this as well.

Consider coating a layer of metal interconnects fabricated above previous layers in the RDL (i.e., not M1), and assume that the impulse response is fit to match the coating over new metal lines on a flat underlying structure. In this case the coatings over existing features with heights similar to those of the new copper lines should be correctly predicted. Conversely, the model may not be correctly fit for feature heights dissimilar to the new copper lines. Here, there are two cases, existing feature whose heights are significantly smaller than the new line heights, and existing features with heights significantly larger than the new line heights.
In the case of the smaller heights, prediction errors of 20%, the maximum error indicated by Fig. 58, are negligible compared to the variation caused by the new metal lines, as the starting heights are relatively small. Now, consider the case of the large features which have been previously coated. Because these variations are large after the previous coating, the effective feature width must also be large, as the variation increases with increasing $W_{ef}$ . In this case the prediction error, as indicated by Fig. 58, will be small, as the error decreases with increasing feature size. Because the prediction error is relatively small in both cases, only a single correctly chosen impulse response is able to predict non-binary coatings with relatively modest error.

While the sensitivity analysis helps explain the ability of the model to predict coatings over non-binary heights, it should be noted that the value of $b$ is impacted most significantly by coating thickness, not line height, as indicated in Fig. 34. When coating a single layer, each feature is subjected to the same process conditions including coating thickness. This means that the range of common model parameters analyzed in Fig. 58 is most likely an over-estimate, and that the real range of model parameters possible to observe on a single layer is most likely tighter. Finally, as previously mentioned in Section 3.11, it is believed that the specific curing procedure for each layer has the most significant impact on model parameters after coating thickness. Again, as each feature being coated is subjected to the same curing procedure, there is likely even less variation among the model parameters optimized for each feature on a single layer.

3.14) Two Dimensional Model Form

Until this point, only the one dimensional form of the model has been discussed. In this section its form is expanded into two dimensions and applied it to a newly chosen test structure. This new test structure is the calibration test mask for the Nikon stepper in MTL. It serves as a good example complex test pattern as it is specifically designed to contain a wide variety of feature shapes and sizes. Using this layout as a mask, the structure is etched into 4.2 µm thick oxide and coated with 8 µm thick polyimide. Again, the Wyko interferometer is used to optically profile the structure both before and after spin coating. These profiles can be seen in Figs. 59 and 60. Unfortunately, the Wyko does not accurately measure sidewalls. For this reason, many of the walls in the coated profile appear unnaturally sharp, but this is simply an artifact of the measurement system.
The model for predicting the 2D coated profile still consists of a convolution between the system impulse response and the underlying topography; however, there are two key distinctions. First, the convolution is now a two-dimensional convolution, instead of a one-dimensional
convolution; this is expected. The second point is more subtle: obtaining the two dimensional impulse response is not as trivial as the one dimensional case, as it cannot be directly measured from a topography step response. Instead, the two dimensional impulse response can be inferred from the one dimensional impulse response, or from more complex 2D data. Here, an approximation for the 2D impulse response is first presented, and a more complete solution is later formed. For the first attempt, the 2D impulse responses is approximated as:

\[ h_2(x, y) = h_1(x) \cdot h_1(y) \]

This form is desirable as it reproduces the original 1D impulse response when integrated along either the x or y direction. To show why this is required, consider the 1D step response, \( E(x) \). It can be expressed in either one or two dimensions. First, it is the convolution of the 1D impulse response \( h_1(x) \) with a 1D step \( u(x) \). Secondly, it is also the result of the 2D convolution between a 2D step \( u(x, y) \), and the 2D impulse response \( h_2(x, y) \). Here, \( u(x, y) = 1 \) for \( x > 0 \) and 0 otherwise.

\[ E_1(x) = \int h_1(x') \cdot u(x - x')dx' = h_1(x) \ast u(x) \]

\[ E_2(x, y) = \int \int h_2(x', y') \cdot u(x - x', y - y') dy' dx' = h_2(x) \ast u(x, y) \]

Here, as the 2D step response is only a function of \( x \), \( u(x, y_1) = u(x, y_2) \) for all \( y_1, y_2 \), we can rewrite \( E_2(x, y) \) as:

\[ \int \int u(x - x', y - y') \cdot h_2(x', y') dy' dx' = \int u(x - x') \cdot \int h_2(x', y') dy' dx' \]

And the 1D and 2D step responses can be related by:

\[ E_1(x) = \int u(x - x') \cdot h_1(x')dx' = \int u(x - x') \cdot \int h_2(x', y') dy' dx' = E_2(x) \]

Thus:

\[ h_1(x) = \int h_2(x, y)dy \]

While approximating \( h_2(x, y) = h_1(x) \cdot h_1(y) \) does result in the first desired property, this method results in an impulse response which is not radially symmetric. Because of this, the predicted profile changes based on the coordinate system. Additionally, when this form is used, the predicted profile has unnatural artifacts, especially near corners, as seen in Fig. 61 for the approximated 2D impulse response and in Figs. 62 for the resulting predicted profile.
Figure 61: Approximate 2D impulse response.

Figure 62: Predicted coating using the approximate 2D impulse response. Artifacts are observed in corners and as unexpected ridges.

To obtain a radially symmetric two dimensional impulse response, we first consider a rotation of the one dimensional case:

\[ h_2(x, y) = h_1 \left( \sqrt{x^2 + y^2} \right) = \frac{\beta}{\pi (\beta^2 + x^2 + y^2)} \]
While this form has the property of being rotationally symmetric, it does not reproduce the original 1D impulse response when integrated along the $x$ or $y$ direction:

$$\int \frac{\beta}{\pi (\beta^2 + x^2 + y^2)} dy = \frac{\beta}{\pi \sqrt{\beta^2 + x^2}} \neq \frac{\beta}{\pi (\beta^2 + x^2)}$$

Therefore, the use of this technique results in significantly worse predictions when compared to the previous approximation.

In order to determine a consistent two dimensional rotationally symmetric impulse response, the following iterative algorithm is developed. Here, we use gradient descent to find a 2D radially symmetric impulse response which integrates back to the 1D form:

1) Initialize a guess for the 2D impulse response as a function of $r$. Here, the rotated 1D response may be used:

$$h_2(r) = h_1(x)$$

2) Determine the resulting 1D impulse response as a function of the proposed 2D response by integrating along one dimension:

$$h'_1(x) = \int h_2(r) \, dx$$

3) Subtract a fraction, $\eta$, of the 1D error from the current estimate for the 2D impulse response to obtain an improved 2D response:

$$h_2(r) = h_2(r) - \eta \cdot (h'_1(x) - h_1(x))$$

4) Repeat steps 2-3 until convergence.

It should be noted that in this algorithm the learning rate, $\eta$, should be kept very small ($\eta < 0.05$ was used here). Additionally, the target and initial 1D impulse responses used in this algorithm should not be raw data, but should be fit to the proposed 1D form. This eliminates the effect of noise on the algorithm, which may significantly alter its final form, and possibly prevent convergence. Ideally, this algorithm could be used to determine an analytical form of the two dimensional impulse response; however, this has not yet been completed. Until then, this algorithm will approximate the two dimensional impulse response to any desired accuracy. In Fig. 63, the 2D impulse response can be seen, resulting from a 1D impulse response extracted from the same
wafer as the Nikon calibration test structure.

![2D impulse response generated using the algorithm.](image)

**Figure 63:** 2D impulse response generated using the algorithm.

Using this rotationally symmetric impulse response, the two dimensional coated structure can be predicted. The resulting predicted surface, shown in Fig. 65, corresponds well to the measured surface seen in Fig. 64. We note that because the predictions are generated from measured, non-ideal, underlying topographies, variations in the widths of the lines often carry through into the predictions. This can be seen in the areas such as the “ribs”, which have non-uniform widths.

In these figures the heights of key features are labelled. Ideally, the overall prediction error could be exactly quantified, but this is difficult for the same reasons as outlined in Section 3.4 and Section 3.12. Additionally, because the measured profile is generated from multiple images stitched together (approximately 80), phase shifts are again present within the measured profile, which further skews point-based error quantifications.
Figure 64: Measured coated profile.
Figure 65: Predicted coated profile.
Chapter 4: Model Applications

In this chapter, the results of the proposed model are applied in order to control the surface variations of coatings over existing layouts using dummy fill techniques. Section 4.1 explores design guidelines intended to control both feature scale and long range spin coating variations. These guidelines serve only as rules of thumb, and as is shown, these rules of thumb have significant limitations. In order to truly control the height variations after spin coating, fill patterns are developed which can be applied to existing layouts. These fill patterns are more flexible and powerful than the design guidelines, and are explored in Section 4.2. Finally, the fill patterns are expanded upon in Section 4.3 by also adding including a cheesing pattern. With this final addition, we limiting both feature scale and long range, average height surface variations to a small value constrained primarily by the allowed feature size and desired pattern density.

4.1) Design Guidelines

To explore our design guidelines, a simple set of requirements is first proposed. Lines of height $H$ must be coated using a fixed polyimide recipe whose model coefficients are $a$, $b$, and $c$. It is required that after coating, the feature scale variations must be kept under $V_f$, and the long range, average height variations must not exceed $V_a$. Together, these two requirements ensure that the global variation, $V_g$, will not exceed $V_f + V_a$. An illustration of these variations can be seen in Fig. 66.

![Figure 66: Feature scale ($V_f$), average ($V_a$), and global ($V_g$) variations.](image-url)
First, consider a naïve approach to this problem. In order to limit the feature scale variations, the results of Section 3.8 suggest that this can be done by limiting the $W_{\text{eff}}$ of all features. For areas of uniform line spacing and width:

$$P_{\text{max}} - P_{\text{min}} \approx \frac{2 \cdot H \cdot a}{\pi} \cdot \tan \left( \frac{W_{\text{eff}}}{2 \cdot b} \right)$$

As $\tan(x)$ is a strictly increasing function, keeping the $W_{\text{eff}}$ of all features under a fixed value, $W_{\text{eff}, \text{max}}$, will limit these feature scale variations to $V_f$.

$$2 \cdot b \cdot \tan \left( \frac{\pi V_f}{2 \cdot H \cdot a} \right) \approx W_{\text{eff}, \text{max}}$$

Now, consider the expression for $W_{\text{eff}}$:

$$W_{\text{eff}} = \frac{w \cdot s}{w + s}$$

Here, it is shown that $W_{\text{eff}} \geq w, s$. Both are strictly positive and finite for simplicity.

Therefore:

$$w + s > \max(w, s)$$

Thus,

$$W_{\text{eff}} < \frac{w \cdot s}{\max(w, s)} = \min(w, s)$$

This leads to the first proposed design guideline. Limiting either all line widths or all spacings to less than $W_{\text{eff}, \text{max}}$ will limit feature scale variations to less than $V_f$ for uniform regions.

Next, the average height variations are considered. As first suggested in Section 2.2.2, the average height of a coated region, $H_c$, is unchanged after coating (as always, the reference height is always the height of a coated, flat surface). This is most easily explained in the frequency domain.

Consider an uncoated area with an average height $H_u$. For binary height topographies with height $H$, and pattern density $\rho$, $H_u = H \cdot \rho$, as exactly $\rho$ of the profile has a height $H$, and the remainder has zero height. The DC value of the impulse response, $\mathcal{H}(0)$, is 1 by definition, as it must integrate to a step response, and far away from the step, the thicknesses of the high and low sides must be equivalent.
\[ H(0) = \int h(x)dx = E(\infty) - E(-\infty) = 1 \]

As this is a LTI system, the average value of the coated profile is:

\[ H_c = H_u \cdot H(0) = H_u \]

This leads to the second proposed design guideline. Limiting the range of pattern densities to \( \rho_r = \rho_{\text{max}} - \rho_{\text{min}} \) will limit the long range, average height variations, \( V_a \), to \( H \cdot \rho_r \).

4.2) Fill Pattern

Now, these guidelines are applied to the original requirements. Let \( H = 5 \, \mu m, a, b, \) and \( c \) be the recorded model coefficients for \( H = 5 \, \mu m \) and \( T = 13 \, \mu m \), and let \( V_f = 1 \, \mu m \) and \( V_a = 2.5 \, \mu m \). Using the results of the previous section, \( W_{eff,\text{max}} = 8 \, \mu m \) and \( \rho_r = .5 \).

To meet all of these requirements, fill patterns for the spacings are designed. For simplicity, this pattern consists of parallel lines of uniform width and spacing. In reality, 2D fill patterns are more common [24, 25], but the simple 1D case will demonstrate the fill effectiveness. By setting the spacing of the fill pattern to \( 8 \, \mu m \), and its pattern density to 50%, it is ensured that throughout the wafer all spacings will be no larger than \( 8 \, \mu m \), and that the range of pattern densities will be under 50% (\( \rho = 100\% \) for all lines, \( \rho = 50\% \) for all spacings). Finally, this fill pattern is applied to the TV, and the simulated resulting coatings with and without the fill patterns are plotted for the \( W = 128 \) scan in Fig. 67.

![Figure 67: Normal predicted profile (left) and prediction using fill pattern (right).](image-url)
While this pattern meets the requirements for the majority of the features, the success is not universal, revealing the limitations of the guidelines and of the naïve fill pattern. The feature scale variations exceed the tolerance limits for areas with originally large feature sizes, such as the \( w = 128 \, \mu m, s = 64 \, \mu m \) region, between 8000 \( \mu m \) and 10000 \( \mu m \) in Fig. 67. The reason is that the feature scale variation reductions are only guaranteed for uniform regions. As soon as multiple line widths exist, such as the fill and original pattern lines, the analysis from Section 3.8 breaks down.

These large starting features can also be viewed as areas of 100% pattern density (the lines) and 0% pattern density (the spacings). Interestingly, when looking at the starting features from this perspective, the guidelines do work, as the variation between the lines and spacings is actually a long range, average height variation (with 2.5 \( \mu m \) tolerance), as opposed to a feature scale variation (with 1 \( \mu m \) tolerance). This example demonstrates the ambiguity of the two types of variations. Unfortunately, these long range, average height variations can appear at relatively small feature sizes (~ 100 \( \mu m \) in the previous example).

A naïve solution to this problem is to raise the pattern density of the fill pattern. Raising \( \rho_{\text{min}} \) to 90\% ensures 0.5 \( \mu m \) average height variations, and dropping the feature size to 4 \( \mu m \) keeps the feature scale variation to 0.5 \( \mu m \) as well. This would ensure a global variation no larger than 1 \( \mu m \). For reference, the coating using this fill pattern can be seen below in Fig. 68. While simple, this approach is highly undesirable for a number of reasons. First, increasing the pattern density of the fill greatly increases the capacitance between lines. Secondly, increasing the metal pattern density often increases the warpage caused after electroplating [26]. Because of this, it is impractical to use near 100\% pattern density throughout the wafer.
4.3) Fill and Cheesing Patterns

While the original fill pattern provides only limited success, a modified version is now presented which meets the original requirements. As was shown, limiting both feature scale, and average height variations requires both a constant pattern density, and minimum feature size across the wafer. The previous pattern could only do so by raising the pattern density to nearly 100% after fill. While this reduces variation, it is also impractical.

Instead, the pattern density of the lines can also be reduced by removing area from the interior of the lines in a process known as “cheesing” [27]. This allows the pattern density and the feature sizes of both the lines and spaces to be independently and arbitrarily set after fill and cheesing. To simplify this process, the layout is discretized into equal sized “pixels”. Then each pixel is assigned either the fill pattern for spaces, or the cheesing pattern for lines. This allows the fill and cheesing patterns to be easily applied to any existing RDL layout.

The fill and cheesing pixels are both the same size, with edges of length $P$ as illustrated in Fig. 69. The line pixel is similar to a picture frame. Its border is completely metal, ensuring conductivity to adjacent pixels, while the center is a hole of size $H \times H$, allowing the pattern density to be controlled. The spacing pixel is the inverse of the line pixel. Its border is empty, ensuring a lack of conductivity to adjacent pixels, while the center is filled with a square of size $F \times F$, and
sets the pattern density.

![Figure 69: Fill pixel to replace an empty space (left) and cheesing pixel to replace copper lines (right).](image)

Both of these pixels must be constructed so that the pattern density is constant, ensuring zero long range, average height variation, and the feature size must be kept small, ensuring minimal feature scale variation. Here, we assume that the minimum feature size, \( M \), is set by the lithography process, and that the desired pattern density \( \rho \) is set by the maximum allowable warpage or added capacitance. To ensure equal pattern density:

\[
\frac{p^2}{\rho^2} = \frac{p^2 - H^2}{p^2}
\]

Thus

\[
F = \sqrt{\rho} \cdot P
\]

\[
H = \sqrt{1 - \rho} \cdot P
\]

Then to meet lithography requirements:

\[
M \leq F, H, \frac{P - F}{2}, \frac{P - H}{2} = P \cdot \sqrt{\rho} \cdot P \cdot \sqrt{1 - \rho}, \frac{P \cdot (1 - \sqrt{\rho})}{2}, \frac{P \cdot (1 - \sqrt{1 - \rho})}{2}
\]

Finally, solving for \( P \), we use this value to find \( F \) and \( H \).
\[ P \geq \frac{M}{\min\left(\sqrt{\rho}, \sqrt{1-\rho}, \frac{1-\sqrt{\rho}}{2}, \frac{1-\sqrt{1-\rho}}{2}\right)} \]

It is worth noting that it may be possible to achieve dimensions twice as small as originally suggested. If the top edge of each pixel is moved to the bottom, and the left edge moved to the right, a new design is constructed where the minimum feature size becomes twice as large, as seen in Fig. 70. In a bulk material, these pixels create the exact same patterns as the previously mentioned counterparts. However, when isolated, the minimum feature size is twice as big, as the opposing edges are combined.

![Figure 70: Alternative fill (left), and cheesing (right) pixels.](image)

Fig. 71 shows four pixels using this alternative pattern. The upper left pixel is the spacing pixel, while the remaining three are the line pixels. Just as before, the line pixel will connect adjacent line pixels, assuming rectilinear interconnects, and the spacing pixel will not connect adjacent pixels.
Figure 71: Four adjacent pixels using the alternative patterns. The upper left is a spacing pixel and is disconnected from the other three.

However, these alternative pixels may potentially create unwanted connections. Consider four pixels of alternating types in a grid, as seen in Fig. 72. While the original design may or may not connect the lines, depending on the lithography and ECP results, the alternative design is guaranteed too. While this may produce unwanted results, this configuration is not recommended for either pixel type (or the traditional 0%, 100% fill patterns). For this reason, the alternative pixel design may be appropriate to use in all cases, but it is especially preferred when lithography limits become significant.

Figure 72: Four pixels of alternating types in a grid. The conventional (left) may or may not connect diagonal pixels, while the alternative (right) will typically connect.

Applying these patterns to an existing layout is straightforward. The RDL layout is first
designed without restriction, then these fill patterns are substituted for the typical 0% and 100% pattern density space and line patterns. However, there are two potential drawbacks with this approach. First, unknown parasitics, particularly capacitances, are added from the fill patterns, which must be considered [24, 25]. Secondly, as the cheesing removes metal, the resistance of the lines increase.

Assume the completely filled material has a sheet resistance of $R_s$. The new effective sheet resistance $R_{s,eff}$ is calculated as follows. The top and bottom edge of the frame (including the corners) together contribute a resistance:

$$R_{T+B} = R_s \cdot \frac{P - H}{P}$$

While the right and left edges contribute:

$$R_{L+R} = R_s \cdot \frac{H}{P - H}$$

Together, the new effective sheet resistance is:

$$R_{s,eff} = R_s \cdot \left( \frac{P - H}{P} + \frac{H}{P - H} \right)$$

To test the effectiveness of the fill pattern, a coating over a new test structure, made using the space fill and cheesing patterns, is simulated. This structure is divided into four quadrants: cheesing, space fill, lines, and random. The cheesing and fill quadrants are filled with their respective patterns. The lines quadrant is filled with alternating columns of the space fill and cheesing pattern. Finally, the random quadrant is filled with randomly selected pixels of either the cheesing or fill pattern. Each quadrant is filled with 100 x 100 pixels of each type. Finally, the outside is padded with 500 μm of the space fill pattern to reduce edge effects. A desired pattern density of $\rho = 36\%$, and a minimum feature size of $M = 2 \mu m$ is selected (Note: this can be increased to $M = 4 \mu m$ with the alternative pixel structure). This results in pixel dimensions of $P = 20 \mu m$, $F = 12 \mu m$, and $H = 16 \mu m$. The coatings are simulated using the typical values $a = 0.6, b = 8 \mu m$ and $c = 200 \mu m$. Finally, all resulting heights are normalized by the starting line height. An overview of the entire structure, as well as close ups of the individual regions, are shown in Figs. 73 through 76.
Figure 73: Overview of the entire simulated structure after coating.

Figure 74: Simulation of coated cheesing region.
The results of these simulations are quite successful, and demonstrate the effectiveness of the test patterns. Each region has the desired 36% pattern density, controlling any long range, average height variations. Without this, the fill-in quadrant would have a height of zero, the cheesing quadrant would have a height of 1, and the remaining two quadrants would have average heights of 50%. Finally, the feature scale variation is approximately 10%. The minimum value on
the entire wafer is 31.6% of the starting line height, and the maximum value is 41.0% of the starting line height.

To expand the analysis of the fill and cheesing patterns, the coatings over a more complete set of test structures are simulated using the same model parameters. These simulations are intended to determine the maximum possible variation for any test structure using the proposed fill and cheesing patterns, as a function of \( M \) and \( \rho \).

In these simulations, each test structure is 1mm x 1mm in size, and is filled with a single uniform region of lines and spacings, just as in the TV uniform regions. This time, the lines and spacings are now replaced with their respective fill and cheesing patterns. We design and simulate the coating over these test structures for all combinations of \( M \) ranging from 1 to 16, \( \rho = 0.1, 0.25, 0.36, 0.49 \); and line widths and spacings 1, 2 4, and 8 pixels wide. Then, the maximum variation possible is calculated for each combination of \( M \) and \( \rho \) over all width and spacing combinations. We note that calculating the coatings above \( \rho = 0.5 \) results in symmetric maximum variations around \( \rho = 0.5 \).

The resulting maximum global variation can be seen below in Fig. 77 It should be noted that near \( M = 0 \), rounding and discretization cause fill and cheesing pixels with non-equal pattern densities, leading to non-zero variations near this point. Finally, these simulations use the “traditional” fill and cheesing pixels, and the minimum feature size can effectively be doubled using the alternative fill pattern. The results of this simulation suggest that the long range, feature scale, and subsequently global variations of any RDL layout can be made as small as desired using these fill and cheesing patterns, limited only by the minimum allowable feature size and resolution.
Figure 77: Maximum global variation vs. $M$, $\rho$ using fill and cheesing patterns.
Chapter 5: Conclusion and Future Work

The purpose of this thesis is to propose a simulation method that can accurately predict spin coatings over arbitrary topographies. While this model is shown to apply to a variety of topographies, its primary focus is to predict polyimide or other dielectric coatings over copper interconnects. This study is motivated by the hope of bypassing CMP during RDL fabrication, thereby significantly reducing the fabrication cost. The contributions of this thesis are first summarized, then suggestions for future work are presented.

5.1) Contributions

In order to develop the new spin coating model, test vehicles are designed and fabricated. These TVs cover a wide range of line widths, spacings, heights, and polyimide thicknesses common on RDLs. After spin coating, the TV surfaces are profiled, and the feature scale and long range variations are analyzed.

It is found that the long range variations after spin coating are generated, in the form of differences in the average heights from region to region. This average height of a region is proportional to the pattern density of the lines before spin coating. Intuitively this makes sense, as it suggests a conservation of mass during the process. While the same amount of polyimide coats each area, the distribution of material may change based on the underlying topography. Additionally, the feature scale variations are found to be a function of the effective feature width, \( W_{eff} \), defined as \( \frac{w \cdot s}{w + s} \), where \( w \) is the line width, and \( s \) is the line spacing. These two results suggest that in order to ensure planarization, the pattern density of the design must be kept constant, and the feature sizes must be kept as small as possible.

After collecting the experimental data, the model itself is developed. The model takes the form of a convolution between the underlying topography and an appropriately chosen impulse response. Similar models have previously been shown to work for other pattern dependent processes such as nano-embossing and CMP. The results of the model are compared to the experimental results, and the mean square error (MSE) of the feature scale variations are approximately 5%. It is worth noting that the model breaks down for a single extremal case. When the height of the copper line is taller than the polyimide thickness, the model under-predicts the variation caused by high aspect ratio lines. An optional penalty function is proposed that can
compensate for these extremal cases.

The impact of the curing stage on the final topography is investigated. By optically profiling coated wafers, it is found that the uncured surfaces are almost perfectly planar. This suggests that the curing stage, and not the spin coating stage itself, is the step which dominates the final profile. This is in contrast to the prevailing spin coating model proposed by Stillwagon and Larson, which focuses on the spin step and almost entirely ignores the curing process. This is further confirmed by the observed independence between the model parameters and the orientation and location of features, a discrepancy between our experimental results and the Stillwagon and Larson analysis. An alternative model proposed by Hirasawa et al. suggests that the curing indeed dominates the process as a whole. However, this model has its own limitations, as overly simplified approximations combined with computational complexity prevent this model from being used practically.

The model developed in this thesis is also used to predict coatings over non-binary heights. Two non-binary structures are experimentally coated then simulated. Each of these structures are formed by taking the original TV and superimposing additional copies onto the underlying layers. The first structure analyzed has only a single polyimide layer, while the second has two. Here, the model is found to predict these non-binary cases well. While this may seem surprising, as the line height might be expected to influence the model coefficients, sensitivity analysis shows that the expected error from mismatched coefficients is smaller than one might originally expect. Additionally, it is believed that other variables, such as the curing procedure and polyimide thickness, play a larger role in determining the optimal model coefficients, and that these are held constant in the non-binary case.

The new spin coating model is then expanded to the two dimensional case, and is compared to a newly chosen complex structure. Here, an algorithm is proposed that converts the one dimensional impulse response to a two dimensional, radially symmetric impulse response that integrates back to the one dimensional form. Here, the predictions again match the coated experimental structure; however, quantification of the error is still needed.

Finally, the implications of the model are analyzed with the goal of limiting the surface variation after spin coating. Design guidelines are proposed, and while their effectiveness is limited, they do provide rule of thumb metrics for limiting variation. As these design metrics do
not guarantee universal planarization, fill and cheesing patterns are instead designed which can be applied to existing designs. First, a naïve implementation is proposed, which only includes fills. It is found that in order to effectively limit surface variations, nearly 100% pattern density must be achieved, an impractical requirement. Instead, this fill-in pattern is supplemented with a cheesing patterns. This allows the pattern density to be controlled across the design, and for the feature scale variations to be kept under a specified value. While further investigation must be done to ensure success for all geometries, the initial results are positive.

The models and applications described in this thesis are capable of potentially eliminating the need for CMP not only during RDL fabrication, but in many other potential spin coating applications. Additionally, many of the modeling techniques and fill and cheesing patterns may also be applicable to other pattern dependent processes.

5.2) Future Work

While the current results of the spin coating model are promising, additional work is still needed to further verify its accuracy in future cases. Most importantly, the model would benefit from verification of model predictions for coatings in real-world RDLs. Testing the model against unknown layouts would help to confirm that the model applies not only to uniform test structures, but also generalizes to future designs. Similarly, the model must also be tested using a wider range of dielectric materials. Currently, it has only been tested using a single polyimide, and it is still necessary to confirm its validity under a wider range of process conditions.

Additionally, we hope to apply our fill and cheesing technique to existing RDL layouts in order to further verify both the model and the planarization method. All current results of this planarization technique presented in this thesis are results of simulations, and not of experiments. If we apply our method to an existing RDL layout and are able to limit the global variation to the predicted value, we would further confirm the effectiveness and applicability of our planarization method and of the model as a whole. Additionally, by applying this method to a real-world RDL layout, we may better understand the consequences of the fill and cheesing patterns, such as changes to the electrical parasitics of the system and to the wafer warpage.

Besides verifying the model under these unknown conditions, there still exists improvements which can be made to the model itself. One such improvement would be to develop
a relationship between the empirical model coefficients and the known process parameters, such as the dielectric material properties and the dielectric spin speed. This would avoid the need to re-fit the model coefficients for every set of process parameters, and would also allow us to optimize the spin coating process for maximum surface planarization. Additionally, if the relationship is derived using fundamental physics, this would further validate the model as a whole. A final improvement to the model would be to improve its accuracy for the extremal cases discussed in Section 3.5. While we hypothesize a penalty function that may be used in these cases, we believe its form is sub-optimal, and can be improved upon. Additionally, future work may also focus on determining the exact conditions under which the unmodified model break down occurs.

In addition to future work relating to the spin coating process, we also hope to develop a similar model and planarization technique for the electro-chemical plating process. As is seen in Section 3.12, and in other studies [12], the ECP process has its own pattern dependencies that result in copper interconnects growing at different rates within a single layout. While these variations are not as significant as those resulting from the DSC process, they are still important to model and control in order to ensure planar RDLs. If accurate models and planarization techniques using dummy fills can be developed for both processes, then CMP can potentially be avoided during RDL fabrication. This would significantly reducing the cost of fabrication not only the RDLs, but subsequently of ICs as a whole.
References


Appendix A: Experimental Scans and Predictions

Here, we present the full datasets for our experimental scans alongside their corresponding predictions, similar to those of Section 2.2 and Section 3.3. We first present the results for $T = 8 \, \mu m$, then present those for $T = 13 \, \mu m$. For each polyimide thickness, we include tables for $H = 2 \, \mu m, 4 \, \mu m, 5 \, \mu m, 7 \, \mu m,$ and $9 \, \mu m,$ and within each table, both profiles are shown for $W = 4 \, \mu m, 8 \, \mu m, 16 \, \mu m, 32 \, \mu m, 64 \, \mu m,$ and $128 \, \mu m.$ Finally, within each profile, the line spacings transition between $S = 8 \, \mu m, 16 \, \mu m, 32 \, \mu m, 4 \, \mu m, 64 \, \mu m,$ and $2 \, \mu m$ at $2000 \, \mu m, 4000 \, \mu m, 8000 \, \mu m,$ and $10000 \, \mu m$ respectively.

$T = 8 \, \mu m$

<table>
<thead>
<tr>
<th>$W$</th>
<th>Experimental Result</th>
<th>Model Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W = 4 , \mu m$</td>
<td><img src="image1.png" alt="Graph" /></td>
<td><img src="image2.png" alt="Graph" /></td>
</tr>
<tr>
<td>$W = 8 , \mu m$</td>
<td><img src="image3.png" alt="Graph" /></td>
<td><img src="image4.png" alt="Graph" /></td>
</tr>
<tr>
<td>Width (µm)</td>
<td>Graph 1</td>
<td>Graph 2</td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>16</td>
<td><img src="image1.png" alt="Graph" /></td>
<td><img src="image2.png" alt="Graph" /></td>
</tr>
<tr>
<td>32</td>
<td><img src="image3.png" alt="Graph" /></td>
<td><img src="image4.png" alt="Graph" /></td>
</tr>
<tr>
<td>64</td>
<td><img src="image5.png" alt="Graph" /></td>
<td><img src="image6.png" alt="Graph" /></td>
</tr>
<tr>
<td>128</td>
<td><img src="image7.png" alt="Graph" /></td>
<td><img src="image8.png" alt="Graph" /></td>
</tr>
</tbody>
</table>
$T = 8 \, \mu m, \quad H = 4 \, \mu m$

<table>
<thead>
<tr>
<th>$W$</th>
<th>Experimental Result</th>
<th>Model Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W = 4 , \mu m$</td>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
</tr>
<tr>
<td>$W = 8 , \mu m$</td>
<td><img src="image3" alt="Graph" /></td>
<td><img src="image4" alt="Graph" /></td>
</tr>
<tr>
<td>$W = 16 , \mu m$</td>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
</tr>
</tbody>
</table>
$W = 32 \, \mu m$

$W = 64 \, \mu m$

$W = 128 \, \mu m$
$T = 8 \mu m, H = 5 \mu m$

<table>
<thead>
<tr>
<th>$W$</th>
<th>Experimental Result</th>
<th>Model Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$4 \mu m$</td>
<td><img src="image1.png" alt="Graph" /></td>
<td><img src="image2.png" alt="Graph" /></td>
</tr>
<tr>
<td>$8 \mu m$</td>
<td><img src="image3.png" alt="Graph" /></td>
<td><img src="image4.png" alt="Graph" /></td>
</tr>
<tr>
<td>$16 \mu m$</td>
<td><img src="image5.png" alt="Graph" /></td>
<td><img src="image6.png" alt="Graph" /></td>
</tr>
</tbody>
</table>
$W = 32 \, \mu m$

$W = 64 \, \mu m$

$W = 128 \, \mu m$
### Experimental Result vs. Model Prediction

**$T = 8 \, \mu m$, $H = 7 \, \mu m$**

<table>
<thead>
<tr>
<th>$W$</th>
<th>Experimental Result</th>
<th>Model Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W = 4 , \mu m$</td>
<td><img src="image1" alt="Graph 1" /></td>
<td><img src="image2" alt="Graph 2" /></td>
</tr>
<tr>
<td>$W = 8 , \mu m$</td>
<td><img src="image3" alt="Graph 3" /></td>
<td><img src="image4" alt="Graph 4" /></td>
</tr>
<tr>
<td>$W = 16 , \mu m$</td>
<td><img src="image5" alt="Graph 5" /></td>
<td><img src="image6" alt="Graph 6" /></td>
</tr>
</tbody>
</table>
$T = 8 \mu m, H = 9 \mu m$

<table>
<thead>
<tr>
<th>W</th>
<th>Experimental Result</th>
<th>Model Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W = 4 \mu m$</td>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
</tr>
<tr>
<td>$W = 8 \mu m$</td>
<td><img src="image3" alt="Graph" /></td>
<td><img src="image4" alt="Graph" /></td>
</tr>
<tr>
<td>$W = 16 \mu m$</td>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
</tr>
</tbody>
</table>
W = 32 µm

W = 64 µm

W = 128 µm
\( T = 13 \, \mu m \)

\( T = 13 \, \mu m, \, H = 2 \, \mu m \)

<table>
<thead>
<tr>
<th>W</th>
<th>Experimental Result</th>
<th>Model Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>( W = 4 , \mu m )</td>
<td><img src="image1.png" alt="Graph" /></td>
<td><img src="image2.png" alt="Graph" /></td>
</tr>
<tr>
<td>( W = 8 , \mu m )</td>
<td><img src="image3.png" alt="Graph" /></td>
<td><img src="image4.png" alt="Graph" /></td>
</tr>
<tr>
<td>( W = 16 , \mu m )</td>
<td><img src="image5.png" alt="Graph" /></td>
<td><img src="image6.png" alt="Graph" /></td>
</tr>
<tr>
<td>$W$</td>
<td>Experimental Result</td>
<td>Model Prediction</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>$W = 4 \mu m$</td>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
</tr>
<tr>
<td>$W = 8 \mu m$</td>
<td><img src="image3" alt="Graph" /></td>
<td><img src="image4" alt="Graph" /></td>
</tr>
<tr>
<td>$W = 16 \mu m$</td>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
</tr>
</tbody>
</table>

$T = 13 \mu m$, $H = 4 \mu m$
\[ T = 13 \, \mu m, \; H = 5 \, \mu m \]

<table>
<thead>
<tr>
<th>W \ (um)</th>
<th>Experimental Result</th>
<th>Model Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td><img src="image1.png" alt="Graph" /></td>
<td><img src="image2.png" alt="Graph" /></td>
</tr>
<tr>
<td>8</td>
<td><img src="image3.png" alt="Graph" /></td>
<td><img src="image4.png" alt="Graph" /></td>
</tr>
<tr>
<td>16</td>
<td><img src="image5.png" alt="Graph" /></td>
<td><img src="image6.png" alt="Graph" /></td>
</tr>
<tr>
<td>Width (µm)</td>
<td>Height in um</td>
<td>Position in um</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>----------------</td>
</tr>
<tr>
<td>32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>128</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
$T = 13 \, \mu m, \, H = 7 \, \mu m$

<table>
<thead>
<tr>
<th></th>
<th>Experimental Result</th>
<th>Model Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W = 4 , \mu m$</td>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
</tr>
<tr>
<td>$W = 8 , \mu m$</td>
<td><img src="image3" alt="Graph" /></td>
<td><img src="image4" alt="Graph" /></td>
</tr>
<tr>
<td>$W = 16 , \mu m$</td>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
</tr>
<tr>
<td>$W$</td>
<td>Experimental Result</td>
<td>Model Prediction</td>
</tr>
<tr>
<td>-------</td>
<td>---------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>$4 \mu m$</td>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
</tr>
<tr>
<td>$8 \mu m$</td>
<td><img src="image3" alt="Graph" /></td>
<td><img src="image4" alt="Graph" /></td>
</tr>
<tr>
<td>$16 \mu m$</td>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
</tr>
</tbody>
</table>

$T = 13 \mu m, H = 9 \mu m$
Appendix B: Feature Scale Results and Predictions

Here, we present the experimental feature scale variations alongside their corresponding model predictions for all polyimide thicknesses and line heights, similar to those of Section 2.2.1 and Section 3.3. We first present those for $T = 8 \, \mu m$, then present those for $T = 13 \, \mu m$ in two separate tables. Within each table, we include both sets of variations as a function of $W_{eff}$ for $H = 2 \, \mu m, 4 \, \mu m, 5 \, \mu m, 7 \, \mu m, \text{ and } 9 \, \mu m$.

$T = 8 \, \mu m$

<table>
<thead>
<tr>
<th>$H$</th>
<th>Experimental Result</th>
<th>Model Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H = 2 , \mu m$</td>
<td><img src="image1.png" alt="Graph" /></td>
<td><img src="image2.png" alt="Graph" /></td>
</tr>
<tr>
<td>$H = 4 , \mu m$</td>
<td><img src="image3.png" alt="Graph" /></td>
<td><img src="image4.png" alt="Graph" /></td>
</tr>
</tbody>
</table>
$T = 13 \, \mu m$

<table>
<thead>
<tr>
<th>$H$</th>
<th>Experimental Result</th>
<th>Model Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H = 2 , \mu m$</td>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
</tr>
<tr>
<td>$H = 4 , \mu m$</td>
<td><img src="image3" alt="Graph" /></td>
<td><img src="image4" alt="Graph" /></td>
</tr>
<tr>
<td>$H = 5 , \mu m$</td>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
</tr>
</tbody>
</table>
$H = 7 \, \mu m$

$H = 9 \, \mu m$