

## Neural network-based model of photoresist reflow

Charmaine Chia, Joel Martis, Stefanie S. Jeffrey, and Roger T. Howe

Citation: *Journal of Vacuum Science & Technology B* **37**, 061604 (2019); doi: 10.1116/1.5116857

View online: <https://doi.org/10.1116/1.5116857>

View Table of Contents: <https://avs.scitation.org/toc/jvb/37/6>

Published by the [American Vacuum Society](#)

---

### ARTICLES YOU MAY BE INTERESTED IN

[Atomic layer deposition of h-BN\(0001\) multilayers on Ni\(111\) and chemical vapor deposition of graphene on h-BN\(0001\)/Ni\(111\)](#)

*Journal of Vacuum Science & Technology A* **37**, 060903 (2019); <https://doi.org/10.1116/1.5120628>

[Corrosion protection of Cu by atomic layer deposition](#)

*Journal of Vacuum Science & Technology A* **37**, 060902 (2019); <https://doi.org/10.1116/1.5116136>

[Dark current–voltage characteristics of vacuum deposited multilayer amorphous selenium-alloy detectors and the effect of x-ray irradiation](#)

*Journal of Vacuum Science & Technology A* **37**, 061501 (2019); <https://doi.org/10.1116/1.5121197>

[Reflective metamaterial polarizer enabled by solid-immersion Lloyd's mirror interference lithography](#)

*Journal of Vacuum Science & Technology B* **37**, 061801 (2019); <https://doi.org/10.1116/1.5119138>

[Uniform and smooth molybdenum film produced through picosecond pulsed laser deposition](#)

*Journal of Vacuum Science & Technology A* **37**, 061506 (2019); <https://doi.org/10.1116/1.5113970>

[Correlation between sputter deposition parameters and I-V characteristics in double-barrier memristive devices](#)

*Journal of Vacuum Science & Technology B* **37**, 061203 (2019); <https://doi.org/10.1116/1.5119984>

---



## Instruments for Advanced Science

Contact Hiden Analytical for further details:  
**W** [www.HidenAnalytical.com](http://www.HidenAnalytical.com)  
**E** [info@hiden.co.uk](mailto:info@hiden.co.uk)

**CLICK TO VIEW** our product catalogue



### Gas Analysis

- dynamic measurement of reaction gas streams
- catalysis and thermal analysis
- molecular beam studies
- dissolved species probes
- fermentation, environmental and ecological studies



### Surface Science

- UHV-TPD
- SIMS
- end point detection in ion beam etch
- elemental imaging - surface mapping



### Plasma Diagnostics

- plasma source characterization
- etch and deposition process reaction kinetic studies
- analysis of neutral and radical species



### Vacuum Analysis

- partial pressure measurement and control of process gases
- reactive sputter process control
- vacuum diagnostics
- vacuum coating process monitoring



# Neural network-based model of photoresist reflow

Charmaine Chia,<sup>1,a)</sup> Joel Martis,<sup>1</sup> Stefanie S. Jeffrey,<sup>2</sup> and Roger T. Howe<sup>1</sup>

<sup>1</sup>*School of Engineering, Stanford University, Stanford, California 94305*

<sup>2</sup>*School of Medicine, Stanford University, Stanford, California 94305*

(Received 28 June 2019; accepted 17 September 2019; published 16 October 2019)

When polymers are heated above their glass transition temperature, they enter a viscous rubbery state that allows the polymer to be reshaped in a process called “reflow.” The final shape depends on the material, substrate, the initial dimensions of the structure, the reflow temperature, and time and is mostly governed by energy minimization. Most empirical models so far have used linear regression to predict scalar parameters like the thickness of the reflowed structure but do not account for intermediate shapes. In this work, the authors measure the profiles of photoresist patterns subjected to various reflow conditions, complementing results in the literature. Using shallow neural networks, they develop models to predict the type of shape produced after reflow and its full cross-sectional height profile. These models can serve as an aid for polymer engineering and fabrication and also demonstrate the usefulness of a neural network-based approach to physical optimization problems without analytical solutions. *Published by the AVS.*

<https://doi.org/10.1116/1.5116857>

## I. INTRODUCTION

Thermal reflow is the softening and material displacement that occurs in thermoplastic polymers when heated above their glass transition temperature,  $T_g$ . During reflow, polymer chains move to minimize the total energy of the system. This is essentially a creep process near  $T_g$  and a flow process at higher temperatures in the viscous state. A change in the shape of the original structure results—for example, a structure with a simple rectangular profile tends to evolve toward a curved geometry. Reflow behavior has been exploited in applications requiring complex three-dimensional structures that are difficult to obtain with standard lithography processes. An example is the reflow of photoresist, a class of photosensitive polymer used as lithographic patterning masks in the semiconductor and MEMS industries. Standard photoresist patterning gives rise to “binary” structures with rectangular profiles—that is, a layer of photoresist on a substrate is either completely preserved or removed in a selected region. By applying thermal reflow to circular photoresist structures, curved lens-shaped patterns can be formed, which may be transferred to the underlying substrate by a dry etching process with appropriate selectivities. These microlenses can be used for collimation, illumination, and imaging in areas like optical fiber communications, computing, image processing, laser and detector arrays, etc.<sup>1–4</sup> Resist reflow can also be used to make microfluidic channels with curved cross-sectional profiles to facilitate flow control using “push-up” or “push-down” valves.<sup>5</sup> “Wavy” grating structures patterned in a herringbone formation have also been used to capture circulating tumor cells.<sup>6</sup> The use of reflow techniques in conjunction with *grayscale* lithography, which allows for the photoresist to be patterned to different heights, extends the patterning toolkit further.<sup>7</sup> Localized reflow of thermoplastics like poly(methyl methacrylate) (PMMA) can

also be achieved by modifying  $T_g$  selectively,<sup>7–10</sup> allowing for structures comprising both smoothed and nanopatterned areas.

Polymer reflow and shape evolution can be a complex process involving several phenomena. The main driving force is the minimization of total energy, which has contributions from surface tension as well as interfacial tension. These act together to produce a final equilibrium shape with (1) constant curvature (i.e., a hemispherical shape) and (2) a macroscopic contact angle at the triple phase line of ambient, polymer, and substrate that balances the surface-free energies of the polymer and the substrate.<sup>11</sup> However, the progress of shape evolution also depends on temperature and time, and it is possible to obtain intermediate aspherical shapes if the polymer is cooled before it reaches its equilibrium shape.<sup>2</sup> The reflowed shape is also affected by vaporization of lower molecular weight ( $M_w$ ) components (causing volume reduction), partial cross-linking of the polymer, substrate treatment, and edge stresses in the resist structure.<sup>7,12</sup> As such, the reflowed profile can be challenging to predict.

Several studies have attempted to develop models for the outcome of polymer reflow. Starting with a polymer structure with a rectangular profile, the simplest possible model assumes that the equilibrium hemispherical reflowed profile is always achieved, the final volume is equal to the initial volume multiplied by a volume reduction coefficient, and the width of the structure stays constant,<sup>2</sup> allowing the reflowed height to be directly calculated. Corrections to the model accounted for a critical contact angle between the resist and the substrate, as well as deviations from the “ideal” shape, via a fourth order polynomial.<sup>13</sup> Measurements on actual reflowed photoresist structures indicated that deviations from the simple model generally increased with the width of the structure.<sup>14</sup> A common cause of this nonideality was found to be the “edge bulge effect,” the tendency for polymer rounding to begin at the edges of the structure, creating two bulges which eventually could converge at the center to give

<sup>a)</sup>Electronic mail: cchia@stanford.edu

the ideal dome-shaped profile. For the photoresist SPR 220, it was found that the position of the bulge maximum from the edge,  $d_{\max}$ , was proportional to the initial resist height and increased with the cube root of reflow temperature.<sup>15</sup> Fordyce *et al.* also examined the empirical dependence of final reflow dimensions (height, width) on initial shape parameters (vertical and lateral aspect ratio) of the photoresist AZ 501T, given fixed reflow conditions.<sup>5</sup> A reasonable fit to the data was obtained using a multiple regression. In general, tall and narrow structures (large vertical aspect ratio) were found to decrease in height after reflow, while low and wide structures did the opposite. Leveder *et al.* used an analytical spectral approach to describe the reflow behavior of polymer trenches and lines on a polymer substrate of the same material with no triple phase line present.<sup>16,17</sup> It was found that higher frequency components of the initial shape decayed faster, matching the observation that isolated polymer lines with rectangular profiles (i.e., higher order spatial frequencies) eventually attained a low-frequency sinusoidal reflowed shape. A different approach is adopted by SURFACE EVOLVER (SE), a software package that treats three-dimensional structures as surfaces governed purely by energy optimization criteria. Kirchner *et al.* adapted this technique to predict slow resist reflow as a function of time, temperature, and molecular weight, identifying a time constant for contact angle evolution and providing an additional shape constraint.<sup>11</sup> However, because the SE-based model is independent of explicit knowledge of material-specific parameters, it is not suited to predicting intermediate shapes for fast reflow processes and faces limitations when bulk effects such as elastic behavior are important. Finite-element volume methods could aid in accounting for free surface, interfacial, and bulk effects. This approach has been used to simulate two-dimensional viscous droplets and could be extended to the reflow problem by benchmarking model performance to real data, but it is also computationally more involved.<sup>18</sup>

In this study, we develop neural network (NN)-based models to predict the shape of simple resist structures after reflow, bringing together for the first time the initial dimensions of the polymer structure—i.e., (1) width and (2) lateral shape—as well as reflow (3) temperature and (4) time as input variables. The inclusion of reflow time as a variable enables us to simulate the diverse intermediate shapes that the structure evolves through as it reflows to its eventual stable shape at longer time scales. Specifically, we tackle two separate tasks: (i) predicting the *type* of shape obtained after reflow (a classification problem) and (ii) predicting the full cross-sectional height profile after reflow (a regression problem). A NN essentially maps the independent input variables to the prediction output via successive computational layers comprising one or more nodes. In each layer, the values at each node are multiplied by weights, summed, and fed into a nonlinear function to obtain the layer outputs, which in turn serve as the inputs to the nodes in the next layer. The predictive strength of NNs arises from their ability to model complex nonlinear relationships between multidimensional inputs and outputs. This approach differs from SE and finite-element methods as it does not require specific

physical knowledge (e.g., interface energy or contact angle) and can be flexibly trained to model different sets of input variables and nonidealities in the actual data.

The paper is structured as follows: in Sec. II, we describe the procedure used to fabricate photoresist structures for our reflow experiments, as well as how these were characterized to obtain the empirical height profile data. In Sec. III, we describe how these data were organized into datasets to be input into the neural networks, followed by how the NN models are set up, trained, and tested for both the (i) classification and (ii) regression tasks. In Sec. IV, we discuss model performances on these two NN tasks, using the metrics of (i) classification accuracy, for the first problem of post-reflow shape type classification and (ii) mean squared error (MSE), for the second problem of resist height profile regression. Figure 1 illustrates two sample post-reflow rectangular structures that we have fabricated. The top row shows SEM images taken at a cleaved cross section, while the bottom row shows height profile measurements taken with a profilometer. While both structures have undergone reflow for 120 s at 140 °C, their shapes appear quite different—structure (a) exhibits two lobes, while structure (b) is dome-like. Using the NN models we develop, we are able to predict, given a set of initial dimensions and a choice of reflow temperature and duration, for example, whether the reflowed structure will have one or two lobes (due to the edge bulge effect), the position and height of the lobes, when they will converge to form a single dome, and the final height profile of the dome at long time scales. This knowledge can guide the choice of a specific reflow program over another to obtain a desired outcome. While the specific models reported only apply to the type of photoresist we use, the methodology we describe can easily be extended to training NN-based models to make predictions about other resist types and structure geometries. Overall, we demonstrate the applicability of this new modeling approach to providing a fuller description of the reflow process to aid device design and fabrication.

## II. EXPERIMENT

Megaposit SPR 220 photoresist was chosen as the polymer for these reflow studies. Photoresist was first spin-coated onto a 4 in. silicon wafer, then patterned by selective UV exposure using a digital mask (Heidelberg MLA150 Maskless Aligner) and developed to produce the resist test structures. These were patterned with widths ranging from 50 to 1000  $\mu\text{m}$ . The length of the structures was either equal to the width (i.e., a square) or 4.75 mm (i.e., a long rectangular strip). The former would show reflow behavior more similar to the circular resist patterns used to create microlenses, while the latter would resemble the rounded channels used in microfluidics. The resist structure height was 7  $\mu\text{m}$ , which is the thickness of the spin-coated film. Reflow was performed by heating the samples with the test structures on a hotplate prepared at three different temperatures: 120, 140, and 180 °C. Reflow time was controlled between 10 s and 3 h, after which the samples were cooled to room

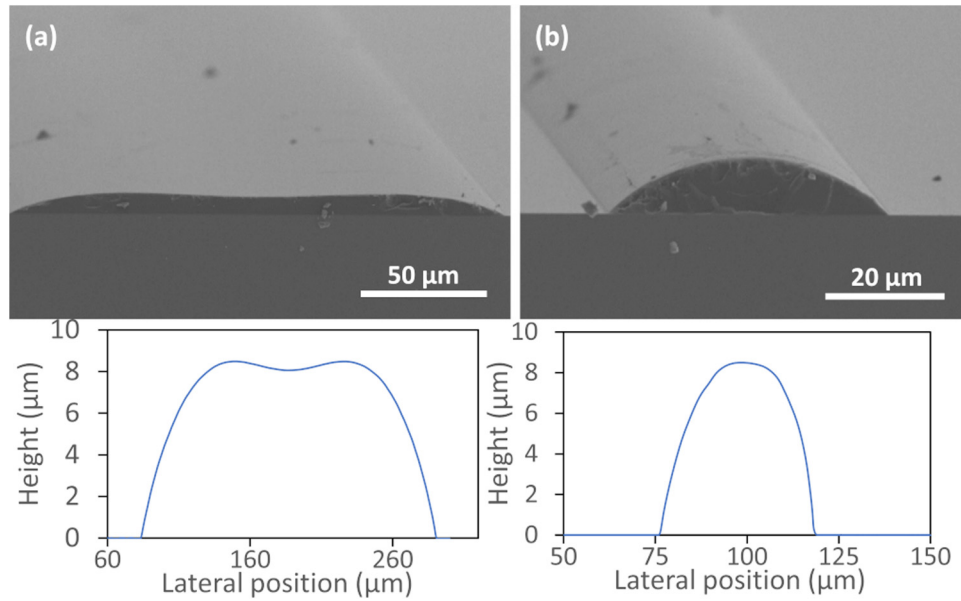


FIG. 1. SEM micrographs (top) and raw profilometer measurements (bottom) showing the curved cross-sectional profiles of two fabricated resist structures with widths of (a) 200 and (b) 50  $\mu\text{m}$ , after undergoing reflow for 120 s at 140  $^{\circ}\text{C}$ . The pre-reflow profiles of the structures were rectangular.

temperature on an aluminum slab. A total of  $m = 482$  distinct samples were fabricated.

The post-reflow profiles of the resist structures were characterized by profilometry (Dektak profilometer), yielding height data across the full width of the structures. The vertical resolution of the profilometer is on the order of 10 nm, while the lateral resolution is limited to about 0.5  $\mu\text{m}$ . The measurements were taken approximately across the middle of each square/long rectangular resist structure. For the long structures, the measured profiles were assumed to be representative of the cross section across most of the length. Profilometry was chosen as it is an accurate and fast way of characterizing the structure height. Other techniques like confocal and electron microscopy also allow for cross-sectional profile determination, but we found that absorption by the photoresist in the near-UV region compromises the accuracy of the optical measurements, while in the case of SEM imaging, additional image processing would be required to extract the height information.

### III. MODELING

#### A. Datasets

The independent input variables in the reflow tests were (1) the patterned width, (2) lateral shape (i.e., “long” or “square”), (3) reflow temperature, and (4) reflow time. The input values for all  $m$  samples acquired were normalized to lie in the range of  $[-1, 1]$  to speed up the convergence of the model training process and consolidated as  $X$ , a  $(4 \times m)$  matrix to be input into the neural network.

The raw profilometer data for each photoresist sample consist of a series of thousands of height measurements taken at various lateral positions as the profilometer tip is scanned across the resist structure. To obtain the shape *type* data, all  $m$  measured profiles were hand labeled according to

five shapes, as shown in Fig. 2. These labels were combined in  $Y$ , a  $(1 \times m)$  ground truth matrix for the classification task. In general, resist shape evolves with increasing reflow time through the different shapes in the following order: 0 (initial rectangular profile)  $\rightarrow 4 \rightarrow 3 \rightarrow 2 \rightarrow 1$  (dome shape). The intermediate shapes, 4, 3, and 2, are characterized by edge bulges with different amounts of separation. In shape 4, the

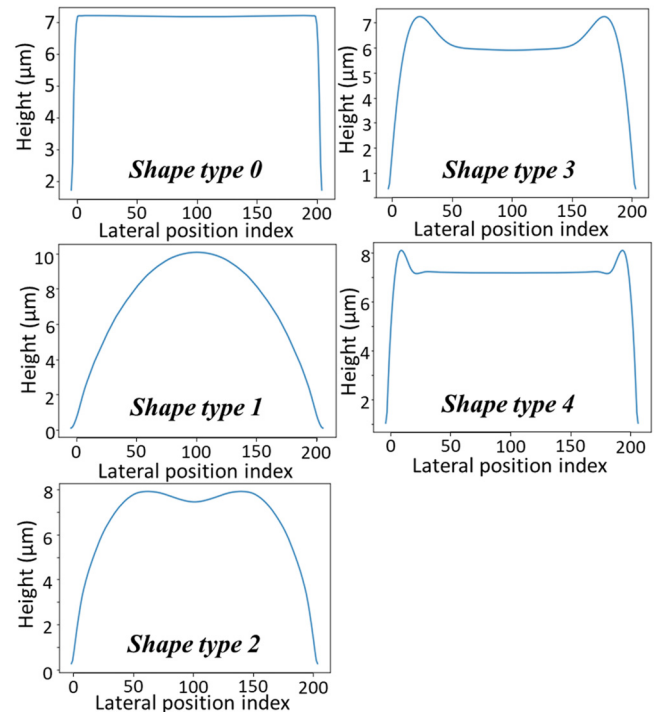


FIG. 2. Five types of shapes that occur during polymer reflow. The objective for the classification task is to accurately predict which shape type will be obtained after reflow with a given set of input variables.



edge bulges are at the edge of the resist structure and accompanied by subtler adjacent bulges reminiscent of ringing seen in electronic signals. In shape 3, the edge-most bulges are distinct while the ringing components have decayed, and in shape 2, the bulges have merged but not yet formed a single peak.

To standardize the profilometer data from various samples, the data were cropped at the edges of each resist structure, averaged about the midpoint, normalized to lie in the range of  $[-1, 1]$ , and finally subsampled at regular intervals to obtain a compressed series of 201 points. The height data for all  $m$  samples were then combined in a  $(201 \times m)$   $Z$  matrix, which serves as ground truth for the regression task. The order of the samples was then randomized, and train and test matrices constructed out of 90% and 10% of the data, respectively.

## B. Setting up the neural networks

### 1. Classification of post-reflow shape type

Given the input variables  $X_i$ , we would like to predict the type of reflow shape produced in terms of shape label  $Y_i$ ,  $i = 1, \dots, m_{\text{train}}$ . To do this, an NN with three fully-connected (FC) hidden layers was used, with layer sizes (i.e., number of nodes) of  $N_1 = 11$ ,  $N_2 = 9$ , and  $N_3 = 7$ . The values entering each hidden node were multiplied by weights, which were initialized using random noise prior to training the model. These were then summed at each node of the successive layer and fed through a Rectified Linear Unit (ReLU) activation. The output layer ( $N_o = 5$ ) consisted of a softmax activation over five shape types ( $j = 0, 1, \dots, 4$ ), with each neuron  $j$  outputting  $P_j$ , the modeled probability that  $Y_i = j$ . Figure 3 illustrates the NN architecture used. The learning rate was 0.0001. The cost function was a sum of the cross entropy loss and a regularization term. Regularization essentially penalizes the learning of an overly complex model by constraining the weights, thus helping to minimize overfitting. L1 and L2 norms on the model weights were explored for regularization, and regularization parameter  $\lambda$  was varied

between 0.0005 and 0.01 to control the amount of penalty applied. Training the neural network involves feeding the inputs forward through the network (forward propagation), calculating the error between the predicted and actual output values in terms of the cost function, then adjusting the weights in the NN based on the gradient of the cost function with respect to the weights (backward propagation). Each step of forward and backward propagation constitutes one epoch. Ideally, over multiple epochs of training, the NN weights are reconfigured by gradient descent and the cost function eventually decreases to a steady minimum value. Model performance was evaluated from the classification accuracy (i.e., the percentage of shape type labels accurately predicted for the training and test data) and was also benchmarked against the performance of multiclass logistic regression and support vector machine (SVM) models.

### 2. Regression of post-reflow height profile

Given the input variables  $X_i$ , we would like to predict the normalized 201-point height profile  $Z_i$  produced after reflow. The actual height values can be obtained by multiplying the output by the normalization constant. Two distinct architectures with six hidden layers were explored for this task—the first was a fully-connected NN with layer sizes  $N_1 = 9$ ,  $N_2 = 15$ ,  $N_3 = 25$ ,  $N_4 = 41$ ,  $N_5 = 71$ ,  $N_6 = 111$ ,  $N_7 = 141$ , and ReLU activations. The second architecture employed two FC layers with  $N_1 = 15$  and  $N_2 = 25$ , followed by either four 1D deconvolution layers or two 1D deconvolution layers followed by two more FC layers. The deconvolution layers were explored as a method to increase output dimensionality. 1D batch normalization was applied to all the hidden layers. Tanh and ReLU activations were tested for the output layer, which yields the predicted height profile. The initial learning rate was 0.0001. The cost function was calculated using a MSE loss between actual height profiles and the ones predicted at the output layer ( $N_o = 201$ ) of the NN, along with some amount of regularization. The mean squared error was also used to gauge model performance on both the train and test data.

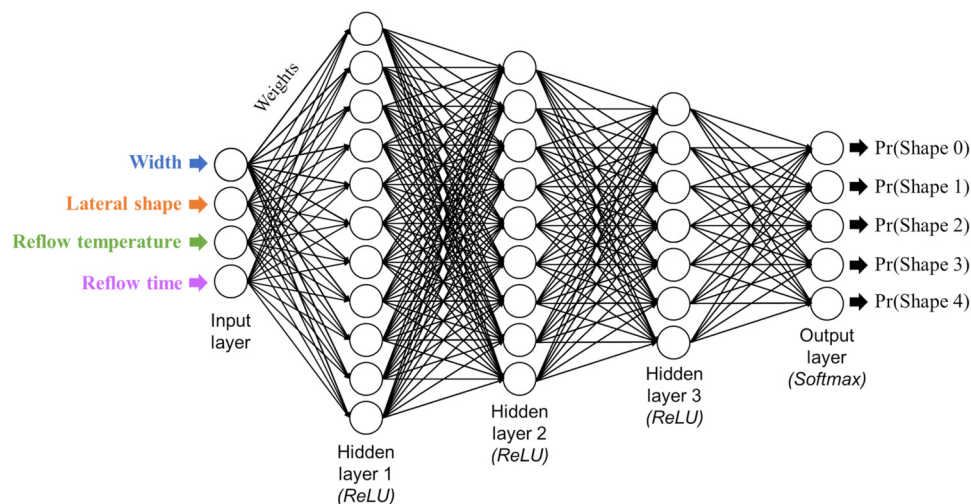


FIG. 3. NN architecture used to train the classification model for predicting the post-reflow resist shape type.

## IV. RESULTS

### A. Classification of post-reflow shape type

The goal of the classification task is to accurately predict the type of shape resulting from reflow, given a set of input parameters. Model classification results were visualized by plotting a “phase map” with two of the independent variables along the horizontal and vertical axes (here, we chose to plot temperature and the initial width of the structure) while the other two (lateral aspect ratio, reflow temperature) were fixed. Ground truth shape type labels describing the actual data were plotted as color-coded circles, while model predictions for each point in the variable space were represented via the color of the background. Figure 4(a) shows a phase map for an NN model trained without regularization, at a reflow temperature of 120 °C. This model achieved a shape type classification accuracy of 99% on training dataset and 97% on test dataset, a very good fit. For comparison, the same classification task was also performed with logistic regression and SVM, using various kernels. Logistic regression yielded 86% accuracy on the training data and 82% accuracy on test data. The prediction results at 120 °C are shown in Fig. 4(b). While the decision boundaries separating regions of different shape types are very smooth, the model appears too simple to capture some of the patterns in the data. Most SVM kernels gave a similar accuracy performance, but one notable exception was the decision trees kernel, which gave 100% accuracy on the training data and 97% accuracy on the test data. However, we can see from its plotted results in Fig. 4(c) that the decision boundaries are arbitrarily complex and abrupt—this is likely not a good representation of actual reflow evolution. This is also somewhat the case for the NN boundaries seen in Fig. 4(a) in the

region of  $200\text{ s} < \text{reflow time} < 1000\text{ s}$  where the data are sparse.

To improve on the results from the NN model, the effect of different regularization coefficients and types of regularization was investigated. Figures 4(d)–4(f) (for reflow temperatures of 120, 140, and 180 °C, respectively) show the best results obtained. These were obtained by conducting further training on the no-regularization model used in Fig. 4(a), using a combination of L1 and L2 weight regularization, where  $\lambda = 0.00075$ . The improvement in the smoothness of the decision boundaries is clearly seen from a comparison of Figs. 4(a) and 4(d). L1 regularization was found to have the effect of creating simpler decision boundaries, while L2 regularization had the effect of smoothing them.

From Figs. 4(d)–4(f), we observe that the larger the initial width of the resist structure, the longer the reflow time and the higher the reflow temperature needed for it to evolve from a shape with widely separated lobes (shape type 4) to the one in which the lobes have started to converge (shape types 3 and 2), to a single dome (shape type 1). However, for many of the wider structures, e.g., those with widths  $> 300\text{ }\mu\text{m}$ , shape type 1 is not attained even after a long reflow time of 3 h at 180 °C. That is, the ratio of width to height of the resist structure seems to determine whether the Laplace effects that attempt to relax the stress at the edges (edge rounding) or the interfacial effects (wetting) dominate the overall shape transformation. Using these predictions, we can infer, for example, that if we would like to achieve a dome-shaped post-reflow structure with a width of  $200\text{ }\mu\text{m}$  within 0.5 h of reflow, we would require a reflow temperature of  $> 140\text{ }^\circ\text{C}$ . Shape predictions for input parameters other than those appearing in the training data (e.g.,  $T_{\text{reflow}} = 156\text{ }^\circ\text{C}$ ) can

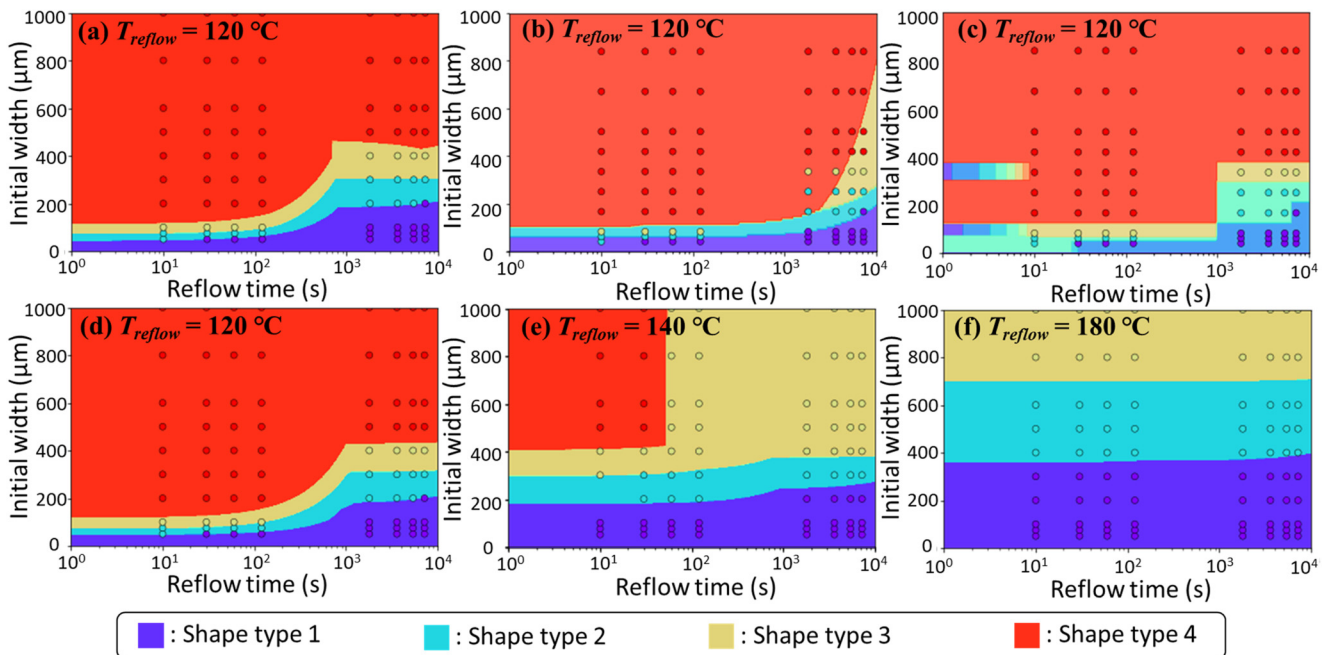


Fig. 4. Shape type classification results for rectangular resist structures for a fixed reflow temperature, obtained using various models: (a) NN with no regularization; (b) logistic regression; (c) SVM using decision trees kernel; and (d)–(f) the best NN model with regularization, for  $T_{\text{reflow}} = 120, 140,$  and  $180\text{ }^\circ\text{C}$ , respectively.

also be made by plugging these values into the trained model.

### B. Regression of post-reflow height profile

To obtain actual height values of the post-reflow structure rather than just the shape *type*, we would need to employ a model trained for the regression task. Comparing the architectures described in Sec. III B 2 based on (A) FC layers only and (B) a combination of FC layers and deconvolution layers, we found that the first architecture was in general better suited to the regression problem, converging more quickly and also predicting smoother cross-sectional profiles with lower MSEs relative to the actual profile data. A  $7 \times 7$  grid of sample profiles from the training dataset can be seen in Fig. 5(a), with the NN predictions (print version: light line; online version: orange line) using architecture A overlaid on top of the actual profile data (print version: dark line;

online version: blue line). There is very good alignment between prediction and measurement, giving a low MSE. Figure 5(d) shows a closeup example of measured and predicted profiles for a single resist sample, prepared by reflow of a  $1000\text{ }\mu\text{m}$  wide rectangular structure at  $120\text{ }^\circ\text{C}$  for 120 s. In contrast, the profiles seen in Fig. 5(c) were output by NN architecture B after training for the same number of epochs. These clearly appear more jagged and asymmetrical, giving a higher MSE. Subsequently, we focused only on optimizing the performance of architecture A by tuning hyperparameters like the output activation function, the learning rate  $\alpha$ , the type of regularization, and the regularization parameter  $\lambda$ .

A summary of the hyperparameter experiments performed is presented in Table I. For the baseline FC model (no regularization) whose training output was described in Fig. 5(a), the MSE per sample after 100k training epochs was  $\sim 0.1$  for the train data and  $\sim 1.14$  for the test data (see row 1). The significantly higher test error is undesirable and indicates that

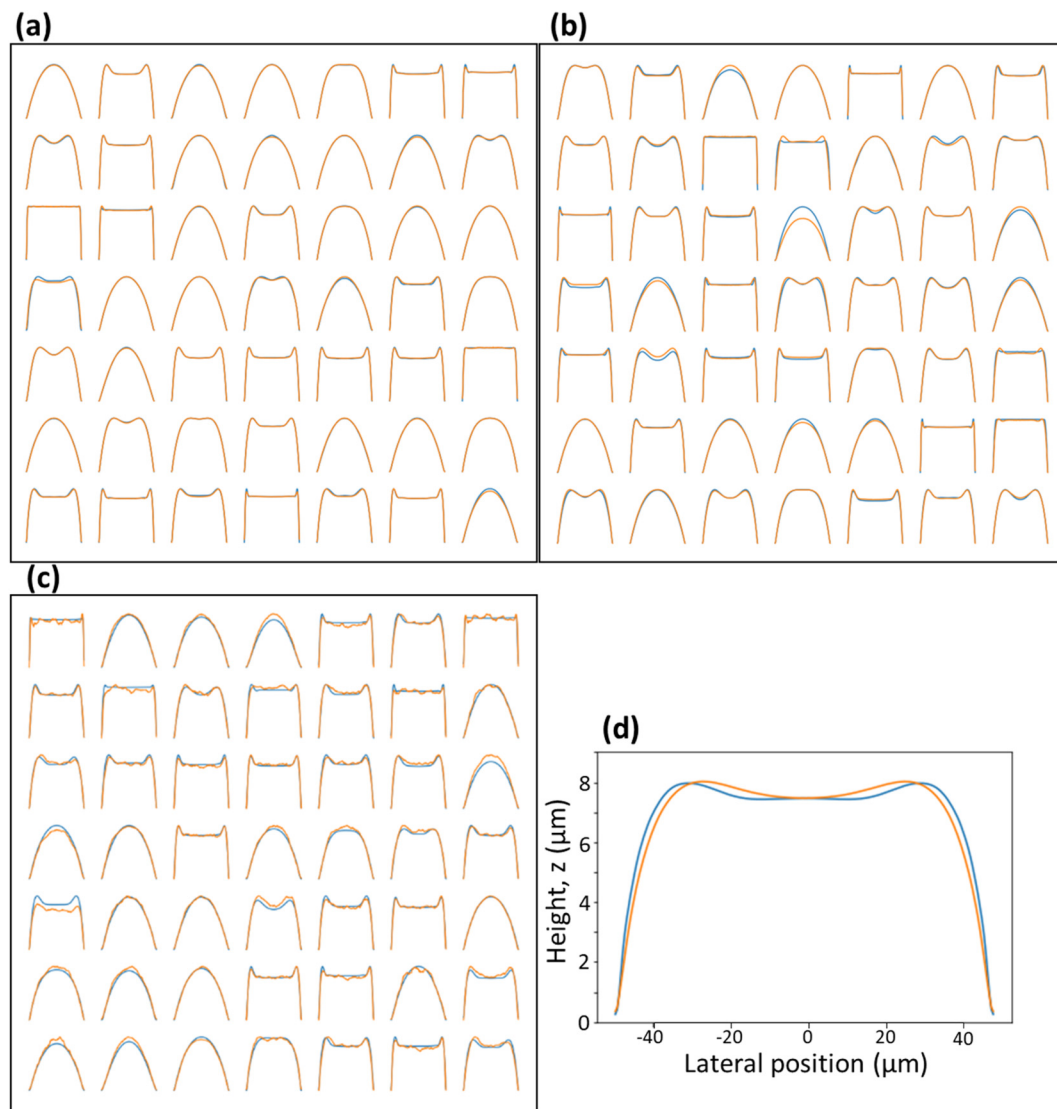


FIG. 5. Measured profile (print version: dark line; online version: blue line) and predicted profile (print version: light line; online version: orange line).  $7 \times 7$  grid of profiles from (a) train set, predicted using architecture A (FC layers only, no regularization); (b) test set, predicted using architecture A (FC layers only, no regularization); (c) train set, predicted using architecture B (with four 1D deconvolution layers). For the profiles plotted in (a)–(c), the y-axis represents the normalized height and the x-axis represents the lateral position. (d) Sample profile predicted using architecture A, for reflow at  $120\text{ }^\circ\text{C}$  for 120 s. Training of both models was done over 100k epochs.

TABLE I. Summary of hyperparameter tuning experiments for architecture A. Boldface values correspond to best model trained.

	Output activation	Regularization	$\lambda$	Training epochs	$\alpha$	Train set MSE	Test set MSE
1	Tanh	—	—	100k	$1 \times 10^{-4}$	0.103	1.14
2	Tanh	L2	0.001	100k	$1 \times 10^{-4}$	0.100	0.75
3	Tanh	L2	0.005	100k	$1 \times 10^{-4}$	0.093	0.99
4	Tanh	L2	0.01	100k	$1 \times 10^{-4}$	0.12	0.66
5	Tanh	L2	0.01	100k	$1 \times 10^{-4} \rightarrow 2 \times 10^{-5}$	0.12	0.55
6	Tanh	L2	0.015	100k	$1 \times 10^{-4}$	0.16	0.71
7	Tanh	$L2 + \Delta \text{area}^2$	0.001, 0.000 25	100k	$1 \times 10^{-4}$	0.17	0.92
8	Tanh	$L2 + \Delta \text{area}^2$	0.001, 0.000 05	+150k	$1 \times 10^{-4}$	0.11	0.90
9	Tanh	$L2 + \Delta \text{area}^2$	0.0015, 0.000 01	250k	$1 \times 10^{-4}$	0.032	0.84
10	ReLU	L2	0.000 2	100k	$1 \times 10^{-4}$	0.26	0.82
11	ReLU	L2	0.001	100k	$1 \times 10^{-3} \rightarrow 1 \times 10^{-4}$	0.17	0.66
12	<b>ReLU</b>	<b>L2</b>	<b>0.01</b>	<b>100k</b>	<b><math>1 \times 10^{-3} \rightarrow 1 \times 10^{-4}</math></b>	<b>0.12</b>	<b>0.44</b>

the model is overfitting to the training data. To get a sense of how large these errors are, we can estimate the average deviation *per point* between the predicted and actual profiles by dividing by the MSE value by the 201 data points in each profile and taking the square root. For the train error, this gives  $\sqrt{(0.1/201)} \approx 0.02$ . Given that the height data have been normalized to lie within a vertical range of  $[-1, 1]$ , a deviation of 0.02 is  $\sim 1\%$  of the total vertical range, which is pretty small.

To reduce overfitting and obtain both a low MSE on *both* the train and test datasets, L2 weight regularization was added. Optimal  $\lambda$  was found to be  $\sim 0.01$  (see Table I), giving a significantly lower test error of 0.66. However, as this was higher than the train set error, error analysis was done to evaluate what the sources of deviation between actual and predicted structures were. This was done by examining the sample output profiles on the test dataset [Fig. 5(b)]. The main contributions to MSE include shape type 0 (rectangular) profiles being predicted to be curved, type 1 (domelike) profiles with inaccurate heights, and type 3 and 4 profiles being output with edge bulges of inaccurate height and position. To encourage better reflow profile match, we explored adding a second regularization term based on the squared difference in cross-sectional area,  $\Delta \text{area}^2$  (calculated by integrating under the profile). We also looked at the effect of replacing the tanh activation function at the output with a ReLU one. While the  $\Delta \text{area}^2$  regularization did not have much positive effect on the test set, the ReLU activation brought it down. Note that a higher learning rate  $\alpha$  was needed when using the ReLU activation at the output as initial learning was slower. Finally, we also implemented learning rate decay based on the observation that as the MSE  $\rightarrow 0$ , the loss decrease became noisier. By reducing  $\alpha$  from 0.0001 to 0.000 02 after 60k epochs, we were able to decrease the final test MSE from 0.66 to 0.55 using a tanh activation at the output. Using the ReLU activation and decreasing  $\alpha$  from 0.001 to 0.0001 after 60k epochs, an even lower test MSE of 0.44 was achieved, while train MSE remained at 0.12. This difference in train and test error is likely due in part to the small dataset, which increases the impact of noise (e.g., imperfections in the fabricated

photoresist structures, measurement uncertainty, etc.) on the test error. All in all, the best NN model for predicting the post-reflow height profile from the input parameters employed: fully-connected layers only, ReLU activations at the output layer, L2 regularization with  $\lambda = 0.01$ ,  $\alpha = 1 \times 10^{-3}$  for the first 60k epochs, and  $\alpha = 1 \times 10^{-4}$  for subsequent epochs.

## V. DISCUSSION

Currently, the only other methods capable of generating the full height profile of a polymer structure after reflow are SURFACE EVOLVER and finite-element methods. However, SE does not take into account effects other than energy minimization and is unable to simulate the dependence of shape evolution on temperature and time. A contact angle evolution constraint could be added, but this approach requires careful data acquisition, and does not guarantee that the contact angle applied is consistent with the rest of the profile predicted. An example of reflow profile evolution predicted by SE can be seen in Fig. 6(b). The series of profiles shown was acquired by allowing the SE optimization algorithm to run for a range of iterations from 0 to 25 000. In contrast, the profile evolution predicted by our NN-based model at a reflow temperature of 120 °C for the same 100  $\mu\text{m}$  wide rectangular structure is shown in Fig. 6(a). The differences are significant—for example, the NN model predicts (accurately) that the structure initially evolves through shapes with edge bulges, while SE fails to do so.

One reason for this discrepancy is that unlike linear thermoplastics like PMMA, photoresist undergoing reflow does not behave like a Newtonian liquid, making it challenging to model accurately using just surface energy considerations. Resist also exhibits cross-linking above a certain temperature, and its contact angle is highly dependent on reflow conditions, making model predictions even more complex. Kirchner *et al.*'s approach to extracting scalar features like the height of the structure using linear regression could be extended to features like the height and position of edge bulges and contact angle, thereby allowing the full profile to be reconstructed using spline interpolation. However, the



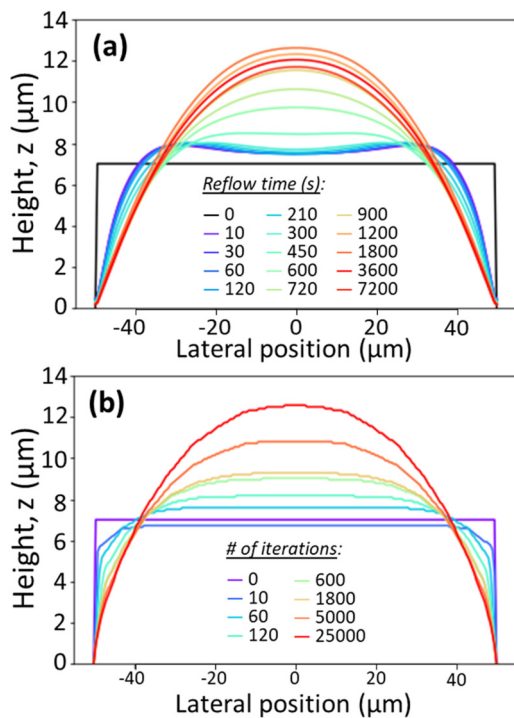


FIG. 6. Reflow evolution prediction for a  $100\text{ }\mu\text{m}$  wide rectangular structure using (a) NN for  $T_{\text{reflow}} = 120\text{ }^{\circ}\text{C}$  and (b) SURFACE EVOLVER.

profiles thus obtained are generally also less accurate compared to those predicted by the NN model.

Using neural networks essentially allows us to perform “nonlinear” regression on all the individual points in the height profile while taking into account correlations between adjacent points. By optimizing over errors for the entire structure at once, neural networks are able to produce smooth profile predictions from which reasonable secondary parameters (e.g., curvature, contact angle, etc.) can be extracted, in a way that cannot be achieved by regressing individual points in isolation. Nevertheless, the predictive power of a neural network is limited by the scope of the dataset used to train the model—for example, the model presented here would not be suitable for predicting the reflow behavior for a different type of polymer. Having said that, it would be entirely feasible to extend the model by training on additional data from other polymers or add an input that accounts for different molecular weights or glass transition temperatures.

## VI. SUMMARY AND CONCLUSIONS

We have acquired extensive empirical data on the reflow of SPR 220, a common photoresist, and introduced a new method for modeling reflow behavior based on fully-

connected neural networks. Given the input variables of initial shape, width, reflow temperature, and time, the models developed can be used to predict both the type of shape of the reflowed structure and the full cross-sectional height profile as it evolves in time. Simple linear regression models are unable to achieve this output, while more complex surface-based or finite-element based models require complete knowledge of physical parameters. Future work could focus on reducing the error in the profile predictions, for example, by augmenting the dataset with more examples of structures that the model found harder to learn. NN models could also be developed to simulate the reflow behavior of other types of polymers and more complex structures. More generally, this work demonstrates the usefulness of neural networks in solving complex nanofabrication problems with a variety of applications from optics to biomedical devices.

## ACKNOWLEDGMENTS

The authors thank Swaroop Kommera for his technical advice over the course of the project. Part of this work was performed at the Stanford Nano Shared Facilities (SNSF)/Stanford Nanofabrication Facility (SNF), supported by the National Science Foundation (NSF) under Award No. ECCS-1542152.

- <sup>1</sup>Z. D. Popovic, R. A. Sprague, and G. A. Neville Connell, *Appl. Opt.* **27**, 1281 (1988).
- <sup>2</sup>F. T. O'Neill and J. T. Sheridan, *Optik (Stuttg)* **113**, 391 (2002).
- <sup>3</sup>Z. J. Lian, S. Y. Hung, M. H. Shen, and H. Yang, *Microelectron. Eng.* **115**, 46 (2014).
- <sup>4</sup>H. Liu, S. Reilly, J. Hermsdorf, E. Xie, V. G. Savitski, A. J. Kemp, E. Gu, and M. D. Dawson, *Diam. Relat. Mater.* **65**, 37 (2016).
- <sup>5</sup>P. M. Fordyce, C. A. Diaz-Botia, J. L. Derisi, and R. Gomez-Sjoberg, *Lab Chip* **12**, 4287 (2012).
- <sup>6</sup>R. He, S. Wang, G. Andrews, W. Shi, and Y. Liu, *Sci. Rep.* **6**, 1 (2016).
- <sup>7</sup>R. Kirchner and H. Schift, *Mater. Sci. Semicond. Process.* **92**, 58 (2019).
- <sup>8</sup>H. Schift, C. Spreu, A. Schleunitz, and J. Lee, *Microelectron. Eng.* **88**, 87 (2011).
- <sup>9</sup>A. Schleunitz, V. A. Guzenko, A. Schander, M. Vogler, and H. Schift, *J. Vac. Sci. Technol. B* **29**, 06F302 (2011).
- <sup>10</sup>A. Schleunitz, V. A. Guzenko, M. Messerschmidt, H. Atasoy, R. Kirchner, and H. Schift, *Nano Converg.* **1**, 7 (2014).
- <sup>11</sup>R. Kirchner, A. Schleunitz, and H. Schift, *J. Micromech. Microeng.* **24**, 055010 (2014).
- <sup>12</sup>J.-F. Gravel, E. Roy, T. Veres, R. Peytavi, B. Voisin, and D. Boudreau, *Microelectron. Eng.* **86**, 2255 (2009).
- <sup>13</sup>F. T. O'Neill and J. T. Sheridan, *Optik (Stuttg)* **113**, 405 (2006).
- <sup>14</sup>M. Ashraf, C. Gupta, F. Chollet, S. V. Springham, and R. S. Rawat, *Opt. Lasers Eng.* **46**, 711 (2008).
- <sup>15</sup>H. Liu, J. Hermsdorf, E. Gu, and M. D. Dawson, *J. Vac. Sci. Technol. B* **34**, 021602 (2016).
- <sup>16</sup>T. Leveder, S. Landis, and L. Davoust, *Appl. Phys. Lett.* **92**, 013107 (2008).
- <sup>17</sup>T. Leveder, E. Rognin, S. Landis, and L. Davoust, *Microelectron. Eng.* **88**, 1867 (2011).
- <sup>18</sup>H. Teyssèdre and P. Gilormini, *J. Nonnewton. Fluid Mech.* **200**, 9 (2013).