

Modeling polymer reflow and shape evolution with neural networks

Charmaine Chia

Abstract: When polymers are heated, or baked, above their glass transition temperature (T_g), they undergo a phase change from a relatively hard amorphous state to a viscous rubbery state. This phenomenon can be applied to reshaping polymers in a process called ‘reflow’, whereby the shape of the polymer changes to minimize its total energy. The final shape depends on the material, substrate, the initial dimensions of the structure, the reflow temperature, and time. Most empirical models developed so far have used linear regression to predict scalar parameters like the final height of the reflowed structure—however these do not account for intermediate shapes and do not yield the full profile. In this paper, we present measured profiles of photoresist structures subjected to various reflow and initial shape conditions, complementing previously reported data. Several regimes of intermediate reflow shapes are identified. Using shallow neural networks, we develop a model to predict the type of shape produced after reflow, as well the full height vector describing the cross-sectional profile. These models further the description of the resist reflow process, providing a useful fabrication aid, while demonstrating the usefulness of a neural network-based approach to physics optimization problems with no simple analytical solutions.

1. Introduction

Thermal reflow is the softening and material displacement that occurs in some materials when heated above a threshold temperature. In polymers, this rubbery state occurs above the glass transition temperature, T_g . During reflow polymer chains move to minimize the total energy of the system, resulting in a change in the shape of the original structure. This behavior has been cheaply exploited in applications requiring complex three-dimensional structures that are difficult to obtain with standard processes. An example is the reflow of photoresist, a class of photosensitive polymer used as lithographic patterning masks in the microelectronics industry. 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. These microlenses could be used for collimation, illumination, and imaging in areas like optical fiber communications, image processing, laser and detector arrays, etc [1–4]. Resist reflow can also aid in the design of microfluidic channels with curved cross-sectional profiles to facilitate flow control using “push-up” or “push-down” valves. [5] ‘Wavy’ grating structures patterned in a herringbone formation have also been used to capture circulating tumor cells [6]. Polymer reflow can be a complex process involving several effects. The main driving force is the minimization of total energy, which has contributions from surface tension as well as interfacial tension. Generally, these act to produce a final equilibrium shape with 1) constant curvature, and 2) a macroscopic contact angle at the triple phase line of ambient, polymer and substrate that satisfies the surface free energies of the polymer and the substrate [11]. 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 [2]. Reflow shape is also affected by vaporization of lower M_w components (causing volume reduction), partial cross-linking of the polymer, substrate treatment, and edge stresses in the resist structure [7,12]. As such, the reflowed profile can be hard to predict.

2. Related work

Starting with a polymer structure with a rectangular profile, the simplest possible model for predicting the outcome of polymer reflow assumes that the equilibrium hemispherical reflowed profile is always achieved, and the final volume is equal to the initial volume multiplied by a volume reduction coefficient, and that the width of the structure stays constant [2], allowing the reflowed height to be directly calculated. Corrections to the model accounted for a critical contact angle between the resist and substrate, as well as deviations from the ‘ideal’ shape, via a 4th order polynomial [13]. Measurements on actual reflowed photoresist structures indicated that deviations from the simple model generally increased with the width of the structure [14]. A common cause of this non-ideality 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 the ideal dome-shaped profile [15]. For the photoresist SPR 220, it was found that the position of the bulge maxima from the edge, d_{max} , was proportional to initial resist height and increased with the cube root of reflow temperature. 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 [5]. A reasonable fit to the data was obtained using a multiple regression. In general, tall and narrow structures 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 polymer support with no triple phase line present [16,17]. A different approach is adopted by the soap film technique-based software, Surface Evolver (SE), which 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 reflow time, temperature and molecular weight, identifying

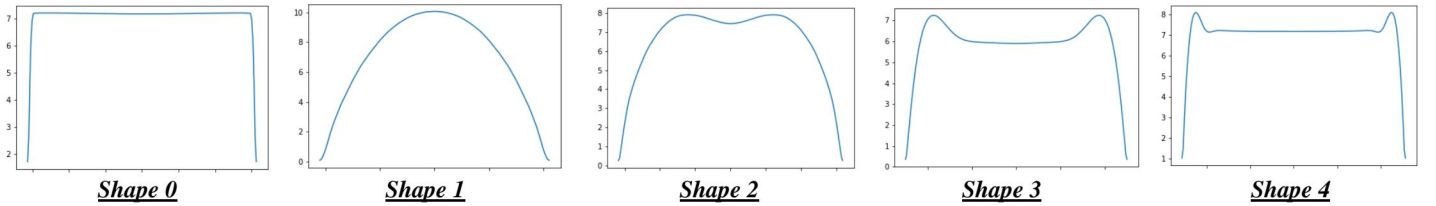
a time constant for contact angle evolution, providing an additional shape constraint [11]. While the SE-based model is thus independent of explicit knowledge of material-specific parameters, it is not suited to predicting intermediate shapes for fast reflow processes, and faces limitations when bulk effects such as elastic behavior are important. Finite element volume methods could aid in the accounting for both free surface, interfacial *and* bulk effects. This has been used to simulate two-dimensional viscous droplets and could be extended to the reflow problem by comparing model performance to real data, but it is also computationally more involved [18].

This study complements the literature by drawing together the contribution of reflow temperature, time and the initial dimensions of the polymer structure. While these parameters have been individually investigated, there is yet a model that combines their effects while also predicting the diverse types of intermediate reflow shapes arising from the edge bulge effect. To do this, we introduce the use of shallow neural networks to tackle the tasks of predicting: 1) the type of shape that results from given reflow conditions (classification), and 2) the height vector tracing the cross-sectional profile of the shape (regression). A neural net (NN) essentially maps the independent to dependent variables via successive computational layers in which the layer inputs are weighted, summed, and fed into a non-linearity to obtain the layer outputs, which in turn serve as the inputs to the next layer. The predictive strength of NN's arises from their ability to model complex non-linear relationships between multidimensional inputs and outputs. The model parameters that define the mapping, i.e. the weights in each layer, are learnt through backpropagation, an iterative process by which the weights are incrementally adjusted to approach the minima of a pre-defined loss function. This approach is easier to apply than the SE and finite element methods as it does not require specific physical knowledge to set up, and can be flexibly trained to model different sets of parameters and non-idealities in the actual data. Overall, we demonstrate the applicability of this new approach to provide a fuller description of the reflow process that can be used to aid in device design and fabrication.

3. Dataset and Features

To prepare the test structures, Megaposit SPR 220 photoresist was spun onto a 4" silicon wafer, then patterned and developed to produce the rectangular resist test structures. These had widths ranging from 50 μm to 1000 μm . The length of the structures was either equal to the width (i.e. a square), or 4.75 mm (i.e. a long strip). Resist structure height was 7 μm . Reflow was performed by heating the samples with the test structures on a hotplate prepared at three different temperatures: 120 $^{\circ}\text{C}$, 140 $^{\circ}\text{C}$ and 180 $^{\circ}\text{C}$. Reflow time was controlled between 10 seconds and 3 hours, after which the samples were cooled to room temperature on an aluminum slab. The post-reflow profiles of the resist structures were characterized by profilometry, yielding height data across the full width of the structures.

The independent variables in the reflow tests comprised the patterned width, lateral shape (i.e. 'long' or 'square'), reflow temperature and reflow time. The parameters for all m samples acquired (where $m = 482$) were normalized to lie between -1 and 1 to improve learning, and consolidated as X , a $(4 \times m)$ matrix of inputs. The raw measured cross-sectional profiles of the reflowed photoresist structures were hand labeled according to five shapes. These labels constitute the output Y , a $(1 \times m)$ ground truth matrix for the classification task. Typical shapes are shown in Fig. 3. 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 (hemispherical shape). The intermediate shapes, 4, 3, and 2, are characterized by edge bulges with different amounts of separation. The raw data from each sample consists of a vector with thousands of points. To reduce dimensionality and noise, data was truncated at the edges of the structure and averaged about the midpoint to obtain a symmetric profile. 201 points were sampled at regular intervals across the width to compress the vector while retaining the shape of the structure. The processed vectors were normalized to lie between -1 and 1 and aggregated in Z , a $(201 \times m)$ matrix which serves as ground truth for the regression task. Training and test matrices were obtained by randomizing the order of the samples, then splitting the data in a 90-10 ratio.



4. Methods

Classification: Given input X_i , we would like to predict the shape label Y_i , $i = 1, \dots, m_{\text{train}}$. To do this, a fully-connected NN with three fully-connected (FC) hidden layers was used, with layer sizes $N_1 = 11$, $N_2 = 9$ and $N_3 = 7$, and ReLU activations. The output layer consisted of a softmax activation over five shape classes ($j = 0, 1, \dots, 4$), with each neuron j outputting P_j , the

modelled probability that $Y_i = j$. The learning rate was 0.0001. The cost function was a sum of the cross entropy loss and the a regularization term. L1 and L2 norms on the weight parameters were explored for regularization, along with the regularization coefficients λ ranging from 0.0005 and 0.01. The cost function was plotted each iteration, and training proceeded until the cost ceased to decrease. Model performance was evaluated from the accuracy of labels predicted for the training and test data, and also benchmarked against the performance of multi-class logistic regression and support vector machine (SVM) models.

Regression: Given four-dimensional input X_i , we would like to predict the 201-dimensional shape vector Z_i . 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 1D shape vector. The initial learning rate was 0.0001. The cost function used a mean squared error (MSE) loss between the predicted and actual vectors. This was also used gauge model performance on both the train and test data.

5. Experiments / Results / Discussion

5.1. Classification of reflow shape

The goal of the classification task was to obtain reasonable decision boundaries separating the different phases of photoresist shape evolution. Model results were visualized by plotting a ‘phase map’ with two of the independent variables along the X and Y axes (we chose temperature and the initial width of the structure, respectively) while the other two (lateral aspect ratio, temperature) being fixed. Ground truth shape classes of measured data were plotted as color-coded points. The XY space was also divided into a mesh and fed through the model to predict the shape class for each point. This in turn allowed for the background of the phase map to be color-coded into five classes according to model predictions. Fig. 4a shows an example of decision boundaries output by training of the model with no regularization added to the cost function, at reflow temperature of 120 °C. This model achieved an accuracy of 99% on training data and 97% on test data, 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 and the resulting predictions at 120 °C can be seen in Fig. 4b. While the decision boundaries are very smooth, the model appears too simple to capture some of 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 Fig. 4c that the decision boundaries are arbitrarily complex and abrupt especially in the region of $200 \text{ s} < \text{reflow time} < 1000 \text{ s}$ where the data is sparse, and thus likely not a good representation of actual reflow evolution.

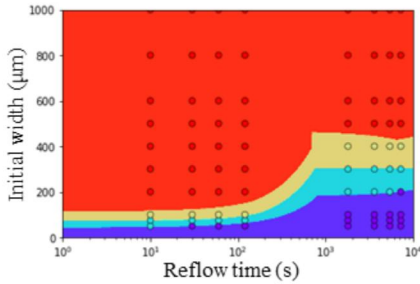


Fig. 4a: NN model predictions (no regularization)

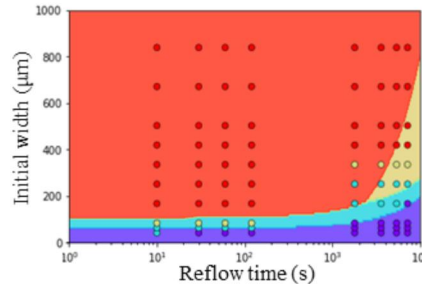


Fig. 4b: Logistic regression model predictions

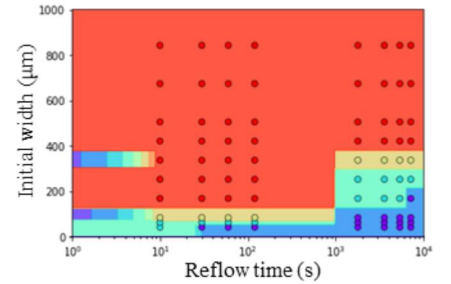


Fig. 4c: SVM with decision trees kernel predictions

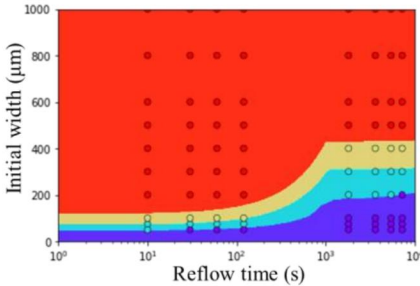


Fig. 4d: Best NN model predictions.
Reflow temperature: 120 °C

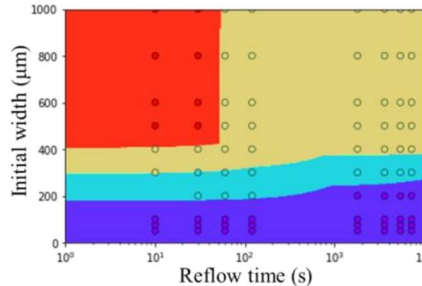


Fig. 4e: Best NN model predictions.
Reflow temperature: 140 °C

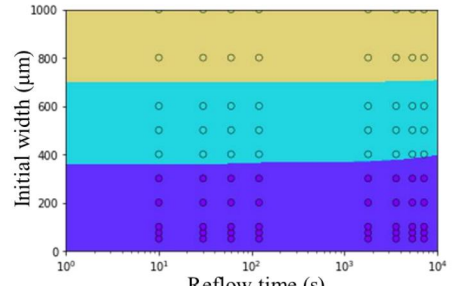


Fig. 4f: Best NN model predictions.
Reflow temperature: 180 °C

To improve on the results from the NN model, the effect of different regularization coefficients and types of regularization were investigated. Fig. 4d, e and f (for reflow temperatures of 120 °C, 140 °C and 180 °C respectively), shows the best results obtained so far. These were obtained by conducting further training on the no-regularization model used in Fig. 4a, using a combination of L1 and L2 weight regularization, where $\lambda = 0.00075$. The improvement in the smoothness of boundaries is clear seen from a comparison of Fig. 4a and d. L1 regularization was found to have the effect of creating simpler decision boundaries, while L2 regularization had the effect of smoothing them. In general, we observe that the larger the initial width of the resist structure, the longer the reflow time and the higher the reflow temperature needed to evolve to the final dome shape, if it even gets there. The vertical aspect ratio determines whether the Laplace effects that attempt to relax the stress in the edges (i.e. edge rounding) or the interfacial effects (wetting) dominate the overall shape transformation.

5.2. Regression of reflow shape vector

Comparing the two architectures based on A) only fully connected layers 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, given that it was both easier to train (each iteration took significantly less time), and output much smoother reflowed cross-sectional profiles with lower mean squared error (MSE). Fig 5a and b show a comparison of the cross-sectional profiles output by architectures A and, with predicted shapes in orange overlaid on top of the actual shapes in blue. Going forward, we will be tuning the performance of only architecture A. Table 1 shows a summary of the hyperparameter experiments performed.

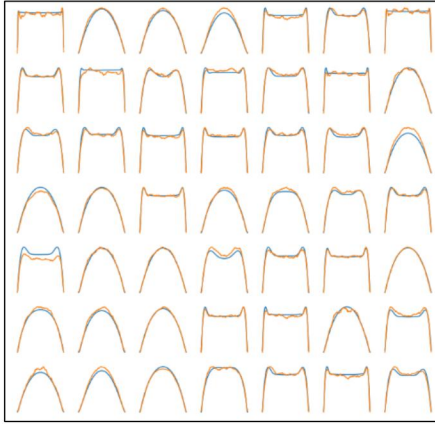


Fig. 5a: Output profiles from model B (with 4 1D deconvolution layers)

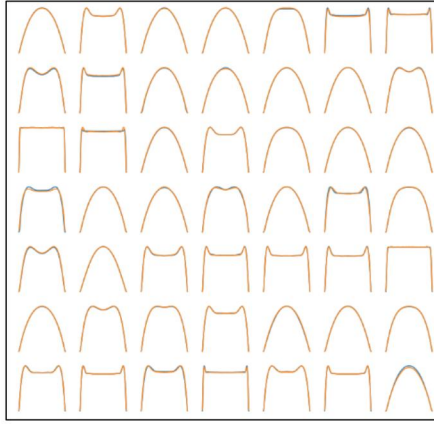


Fig. 5b: Output train set profiles from model A (FC layers only). No regularization

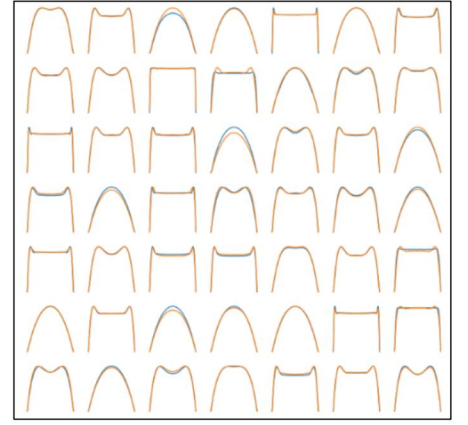


Fig. 5c: Output test set profiles from model A (FC layers only). No regularization

	Output activation	Regularization	λ	Training epochs	α	Train set MSE	Test set MSE
1	Tanh	-	-	100k	1e-4	0.103	1.14
2	Tanh	L2	0.001	100k	1e-4	0.100	0.75
3	Tanh	L2	0.005	100k	1e-4	0.093	0.99
4	Tanh	L2	0.01	100k	1e-4	0.12	0.66
5	Tanh	L2	0.01	100k	1e-4 \rightarrow 2e-5	0.12	0.55
6	Tanh	L2	0.015	100k	1e-4	0.16	0.71
7	Tanh	L2 + Δarea^2	0.001, 0.00025	100k	1e-4	0.17	0.92
8	Tanh	L2 + Δarea^2	0.001, 0.00005	+150k	1e-4	0.11	0.90
9	Tanh	L2 + Δarea^2	0.0015, 0.00001	250k	1e-4	0.032	0.84
10	ReLU	L2	0.0002	100k	1e-4	0.26	0.82

Table 1: Summary of hyperparameter tuning experiments

After training for ~100 epochs without regularization, a MSE error of ~0.1 was obtained for the training data, and ~1.14 for the test data. To give a sense of what these errors mean, note that the reflow profile has a normalized vertical range of $[-1, 1]$ and spans 201 points; a MSE of 0.1 per sample gives an average error of $\sqrt{(0.1/201)} \approx 0.02$, or 1% of the total vertical range. To reduce this overfitting on the training data, L2 weight regularization was added. Optimal λ was found to be ~ 0.01 (see Table 1) 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. Fig. 5b and c show sample output profiles

for the train and test sets side by side. The main contributions to MSE include: Type 1 (hemispherical) shapes with inaccurate heights, Type 0 (rectangular) shapes not being predicted as reflowed profiles, Type 3 & 4 shapes being output with edge bulges of inaccurate height and position. To encourage better reflow profile match, we also explored adding a second regularization term based the squared difference in cross-sectional area, Δarea^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. Neither of these changes seemed to have much positive effect on the test set error; in fact, it seemed like the ReLU activation learnt the profile shapes more slowly. Finally, we implemented learning rate decay based on the observation that as the MSE $\rightarrow 0$, the loss decrease became noisier (see Fig. 5a). By reducing α from 0.0001 to 0.00002 after 60k epochs, we were able to decrease the final test MSE from 0.66 to 0.55, while train error remained at 0.12. This difference in 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) on the test error.

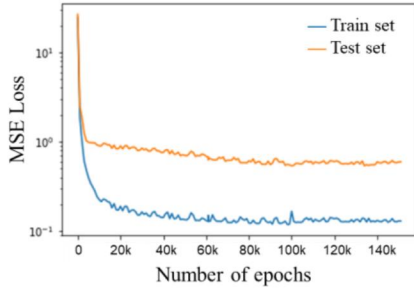


Fig. 5a: MSE loss over training process for experiment 5

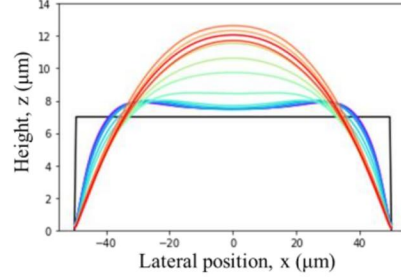


Fig. 5b: Reflow evolution prediction by NN for 100 μm wide structure, at 120 $^{\circ}\text{C}$

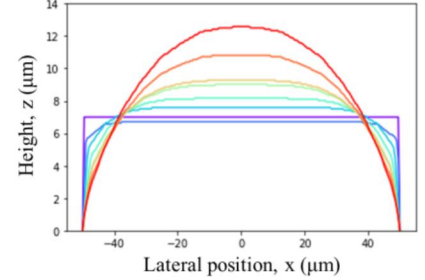


Fig. 5c: Reflow evolution prediction by Surface Evolver for 100 μm wide structure

Apart from using Surface Evolver or finite element methods, there is not really a good way of generating the full post-reflow polymer profile. Even so, SE itself does not take into account effects other than energy minimization, and is unable to simulate the dependence of shape evolution on temperature and time. Fig. 5c shows its prediction for a 100 μm wide structure, which differs considerably from our measured and predicted structures (Fig. 5b). A contact angle evolution constraint could be added, but this approach requires careful data acquisition, and still does not guarantee that the rest of the profile matches the prediction. 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, contact angle, etc., thereby allowing the full profile to be reconstructed using spline interpolation. However, the fit to thus obtained is generally worse than that achieved by the NN model. Using neural nets essentially allows us to perform ‘non-linear’ regression on all the individual points while considering relationships between adjacent points and optimizing over losses over the entire structure at once, so that we also end up obtaining a smooth profile. Having said that, the predictive power the NN-trained model is limited by having a representative body of training data. For example, the model presented above would not be suitable for predicting the reflow behavior for a different type of polymer; nonetheless, it would be entirely feasible to extend the model to do that by training on additional data from other polymers (transfer learning), or perhaps adding an input that accounts for different molecular weights or glass transition temperatures.

6. Conclusion / Future Work

In summary, we have acquired extensive data on the reflow of a certain photoresist and introduced a new method for flexibly modelling reflow behavior. Based on shallow fully-connected neural networks, this method allows us to predict not only the general shape of the reflowed structure given input parameters of initial shape, width, reflow temperature and time, but also the full cross-sectional profile as it evolves in time. This is an outcome that simple linear regression models are unable to achieve, and which more complex surface-based or finite-element based models face limitations in modeling accurately as they often rely on precise knowledge of physical parameters, which is not always possible. The NN models developed here further the description of the resist reflow process, providing a useful fabrication aid, while paving the way to more general descriptions that take into account more input parameters. Future work could further enhance the accuracy of the profile predictions; for example, learning could be improved close to the end, or the dataset could be augmented with more examples of structures the model found harder to learn, such as the box-shaped structures at ‘reflow time = 0’. More generally, this demonstrates the usefulness of neural networks to solve physics optimization problems with no simple analytical solutions.

7. Acknowledgements

I'd like to thank Joel Martis for help in performing the reflow experiments, and Prof. Roger Howe and Dr. Swaroop Kommara for advice. Work was performed in part in the nano@Stanford labs, which are supported by the National Science Foundation as part of the National Nanotechnology Coordinated Infrastructure under award ECCS-1542152.

8. References

- [1] Z.D. Popovic, R.A. Sprague, G.A. Neville Connell, Technique for monolithic fabrication of microlens arrays, *Appl. Opt.* 27 (1988) 1281. doi:10.1364/AO.27.001281.
- [2] F.T. O'Neill, J.T. Sheridan, Photoresist reflow method of microlens production Part I: Background and experiments, *Optik (Stuttg.)* 113 (2002) 391–404. doi:10.1078/0030-4026-00186.
- [3] Z.J. Lian, S.Y. Hung, M.H. Shen, H. Yang, Rapid fabrication of semiellipsoid microlens using thermal reflow with two different photoresists, *Microelectron. Eng.* 115 (2014) 46–50. doi:10.1016/j.mee.2013.10.025.
- [4] H. Liu, S. Reilly, J. Herrnsdorf, E. Xie, V.G. Savitski, A.J. Kemp, E. Gu, M.D. Dawson, Large radius of curvature micro-lenses on single crystal diamond for application in monolithic diamond Raman lasers, *Diam. Relat. Mater.* 65 (2016) 37–41. doi:10.1016/j.diamond.2016.01.016.
- [5] P.M. Fordyce, C.A. Diaz-Botia, J.L. Derisi, R. Gomez-Sjoberg, Systematic characterization of feature dimensions and closing pressures for microfluidic valves produced via photoresist reflow, *Lab Chip*. 12 (2012) 4287–4295. doi:10.1039/c2lc40414a.
- [6] R. He, S. Wang, G. Andrews, W. Shi, Y. Liu, Generation of Customizable Micro-wavy Pattern through Grayscale Direct Image Lithography, *Sci. Rep.* 6 (2016) 1–9. doi:10.1038/srep21621.
- [7] R. Kirchner, H. Schiff, Thermal reflow of polymers for innovative and smart 3D structures: A review, *Mater. Sci. Semicond. Process.* 92 (2019) 58–72. doi:10.1016/j.mssp.2018.07.032.
- [8] H. Schiff, C. Spreu, A. Schleunitz, J. Lee, Shape control of polymer reflow structures fabricated by nanoimprint lithography, *Microelectron. Eng.* 88 (2011) 87–92. doi:10.1016/j.mee.2010.08.029.
- [9] A. Schleunitz, V.A. Guzenko, A. Schander, M. Vogler, H. Schiff, Selective profile transformation of electron-beam exposed multilevel resist structures based on a molecular weight dependent thermal reflow, *J. Vac. Sci. Technol. B, Nanotechnol. Microelectron. Mater. Process. Meas. Phenom.* 29 (2011) 06F302. doi:10.1116/1.3634013.
- [10] A. Schleunitz, V.A. Guzenko, M. Messerschmidt, H. Atasoy, R. Kirchner, H. Schiff, Novel 3D micro- and nanofabrication method using thermally activated selective topography equilibration (TASTE) of polymers, *Nano Converg.* 1 (2014) 7. doi:10.1186/s40580-014-0007-5.
- [11] R. Kirchner, A. Schleunitz, H. Schiff, Energy-based thermal reflow simulation for 3D polymer shape prediction using Surface Evolver, *J. Micromechanics Microengineering*. 24 (2014). doi:10.1088/0960-1317/24/5/055010.
- [12] J.-F. Gravel, E. Roy, T. Veres, R. Peytavi, B. Voisin, D. Boudreau, Microlens array fabrication by enhanced thermal reflow process: Towards efficient collection of fluorescence light from microarrays, *Microelectron. Eng.* 86 (2009) 2255–2261. doi:10.1016/j.mee.2009.04.001.
- [13] F.T. O'Neill, J.T. Sheridan, Photoresist reflow method of microlens production Part II: Analytic models, *Optik (Stuttg.)* 113 (2006) 405–420. doi:10.1078/0030-4026-00187.
- [14] M. Ashraf, C. Gupta, F. Chollet, S.V. Springham, R.S. Rawat, Geometrical characterization techniques for microlens made by thermal reflow of photoresist cylinder, *Opt. Lasers Eng.* 46 (2008) 711–720. doi:10.1016/j.optlaseng.2008.05.008.
- [15] H. Liu, J. Herrnsdorf, E. Gu, M.D. Dawson, Control of edge bulge evolution during photoresist reflow and its application to diamond microlens fabrication, *J. Vac. Sci. Technol. B, Nanotechnol. Microelectron. Mater. Process. Meas. Phenom.* 34 (2016) 021602. doi:10.1116/1.4943558.
- [16] T. Leveder, S. Landis, L. Davoust, Reflow dynamics of thin patterned viscous films, *Appl. Phys. Lett.* 92 (2008) 013107. doi:10.1063/1.2828986.
- [17] T. Leveder, E. Rognin, S. Landis, L. Davoust, Reflow of supported sub-100 nm polymer films as a characterization process for NanoImprint lithography, *Microelectron. Eng.* 88 (2011) 1867–1870. doi:10.1016/J.MEE.2011.01.078.
- [18] H. Teyssèdre, P. Gilormini, Extension of the natural element method to surface tension and wettability for the simulation of polymer flows at the micro and nano scales, *J. Nonnewton. Fluid Mech.* 200 (2013) 9–16. doi:10.1016/J.JNNFM.2012.10.003.