Physics Informed Neural Networks for Predicting Mold Filling in Roll-to-Roll Nanoimprinting Lithography

M. Stahr[1] and P.R. Pagilla[1]

[1]Texas AM University, Department of Mechanical Engineering, College Station, TX 77843

ABSTRACT

We describe a novel method for the selection of key process and transport parameters for Roll-to-Roll Nano-Imprinting Lithography (R2RNIL) process using Physics Informed Neural Networks (PINNs). We focus on the mold filling process of R2RNIL and discuss the use of PINNs to predict key parameters of R2RNIL in the mold filling governing equations. In particular, we discuss how to integrate real-world experimental measurements and results in addition to the model equations obtained from basic governing laws. We provide an outline of the specific methods for sample production, measurements, and how the height measurements of the mold pattern can be averaged across each line grating and interpreted for the purpose of integrating into the PINN. Keywords: R2R Processing, Nanoimprint Lithography, Physics Informed Neural Networks

1 INTRODUCTION

Roll-to-Roll Nano-Imprinting Lithography (R2RNIL) has been shown to be a viable process for manufacturing high-demand consumer goods such as super capacitors, solar cells, and flexible electronics including flexible displays, health monitors or other devices [1]–[4]. R2RNIL is a combination of Nanoimprinting Lithography, a proven method for fabricating micro and nanoscale patterns [5]–[7], and Roll to roll (R2R) manufacturing, which is a method for processing flexible materials (called webs), such as plastics, thin metal sheets, metal foils, fabrics, and coatings, while transporting the material continuously on rollers [1], [8].

However, many of the R2RNIL scientific and technical developments for scaling production to a wide variety of products are still at the lab scale because the transport speeds are still too low for mass production and profitable commercial production. One of the main hurdles for the process is the mold filling stage where the substrate coated with the fluid resin film enters the mold pattern in the mold roller. Within this stage there are three main problems that inhibit accurate reproduction of the patterns: The first is partial filling, which occurs when the resin partially fills the mold, but some residual air is trapped, resulting in craters being left on the final imprint pattern due to trapped residual air within the mold pattern. The second is template defects, which occur due to the small scale of the patterns; it is difficult to produce the highly accurate molds, and a slight defect has a significant impact due to the nanoscale of the features. The third problem is plugging which occurs when prints from a previous rotation of the mold roller do not properly debond, causing the mold to be filled by this now broken print, which prevents further filling of the mold in subsequent runs [9]. Due to the inability to measure many of the key parameters of the mold filling process, there are currently no known methods for direct feedback control.

In order to address the specific problem of mold filling, governing equations have been recently developed which can aid in identifying how machine parameters and process variables can be changed in order to allow the molds to fill at higher speeds [10]. These resulting governing equations are highly nonlinear and stiff PDEs, making standard numerical simulation of the results very difficult. Despite the nature of the governing equations, a linearization method and numerical procedure for solving the nondimensionalized equations are provided in [10]; while the numerical solver works well for a fully filled or almost filled pattern, the solver fails once the actual fill drops approximately below 70%. Another challenge in numerically solving these governing equations is the associated computational complexity. To address the aforementioned problems, a new method is needed for predicting the mold filling behavior for different process and transport conditions while also considering how the prints will be affected by real-world conditions. For addressing these types of issues, a method to use PINNs has been proposed. PINNs are a specialized form of neural networks; instead of being trained on a data set such as a library of images or predefined data points in a standard neural network, the neural networks in PINNs are setup so that the outputs and their derivatives are required to satisfy a set of governing equations obtained from physical governing laws, as well as their boundary and initial conditions. The expectation is that the fully trained PINN will be able to make mold pattern profile predictions and be able to take into account the results from previous successful experiments; thus, provide a mechanism to change material and machine parameters in real-time to facilitate proper mold filling.

The first significant implementation of PINN was provided in [11], where the behavior of the Schrodinger and Burgess equations were predicted; this was a significant step forward as the employed model “didn’t require prior statistical

assumptions, data, linearization or time stepping." Additional research on PINNs relevant to this work is the developments around solving Naiver-Stokes and flow related problems; for example, Naiver-Stokes Flow nets by [12] for solving the velocity-pressure and vorticity-velocity formulations of Navier-Stokes, which were used to predict steady, unsteady and turbulent flows. In [13] a self-adaptive loss balance was developed to solve complex models such as two dimensional steady Kovasznay flow, two-dimensional unsteady cylinder wake, and even three-dimensional unsteady Beltrami flow.

A significant amount of progress has been made to improve PINNs in the last couple of years. While PINNs as first established were shown to be able to solve some difficult problems, they did suffer from many of the same problems that standard numerical methods did, such as not being able to solve "stiff" PDEs [14], [15]. As demonstrated by Colby and Zhao, a standard PINN has very high error when attempting to solve a "stiff" PDE; the specific one they tested was the Allen-Cahn diffusion equation, and the standard PINN resulted in an l_2 error of 0.99. These results clearly show that using a standard PINN to solve stiff PDEs such as the mold filling equations are not an option, however they also presented a few different methods to address and overcome these problems. The methods they employed for solving these problems are mini-batching and time-adaptive learning which reduced the l_2 error to 0.0325 and 0.04, respectively. Another work that focused on this problem was by McClenny who implemented a soft attention mask to the points of interest with respect to the PDE. This method resulted in an average prediction error of 2.1% [16].

While all of these methods above were able to solve stiff PDEs, they were unfortunately much simpler compared to the mold filling problem, as they were only solving for one variable. In order to solve this problem a multibranched PINN can be used. This idea was shown by Niaki et al. [17] to be able to converge on the solution of both the curing percentage and temperature of tools inside an autoclave, while also accounting for the discrete medium change of the resin and autoclave wall.

This paper will outline how methods for improving convergence of PINNs can be applied to a set of governing equations established from basic governing laws, and with the aid of experimental results, to predict mold filling, which can enable real-time parameter optimization. The rest of this paper is organized as follows. A summary listing of the governing equations for the mold filling process in R2RNIL is provided in Section II. In Section III, the methods which aid in PINN convergence and the experimental setup for real-world data collection are described. The preliminary results from real-world experiments, and how the method for their integration into other segments of PINNs for the R2RNIL process are discussed in Section IV. Finally, a summary of the work and future work is provided in Section V.

2 MOLD FILLING GOVERNING EQUATIONS

The constitutive equations for the mold filling process of a substrate coated with a viscoelastic fluid film were derived in [10], and the generalized form of the equations are as follows:

$$\rho \frac{\partial^2 \vec{u}}{\partial t^2} = div(T_d) - \nabla \mathrm{p} \tag{1}$$

$$div(\vec{u}) = 0 \tag{2}$$

$$\lambda \frac{\partial T_d}{\partial t} + T_d = \mu \frac{\partial}{\partial t}(\nabla \vec{u} + \nabla \vec{u}^T) \tag{3}$$

where $\vec{u}$ is the displacement vector, T_d is the deviatoric stress, and p is pressure. The parameter in these equations are ρ, the density of the fluid, λ the stress relaxation time of the fluid, and μ, the fluid viscosity. The element-wise equations corresponding to the above constitutive equations that need to be solved are:

$$\rho \frac{\partial^2 u_x}{\partial t^2} - \frac{\partial T_{xx}}{\partial x} - \frac{\partial T_{xy}}{\partial y} + \frac{\partial p}{\partial x} = 0 \tag{4}$$

$$\rho \frac{\partial^2 u_y}{\partial t^2} - \frac{\partial T_{xy}}{\partial x} - \frac{\partial T_{yy}}{\partial y} + \frac{\partial p}{\partial y} = 0 \tag{5}$$

$$\lambda \frac{\partial T_{xx}}{\partial t} + T_{xx} - 2\mu \frac{\partial^2 u_x}{\partial t \partial x} = 0 \tag{6}$$

$$\lambda \frac{\partial T_{xy}}{\partial t} + T_{xy} - \mu \frac{\partial}{\partial t}\left(\frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x}\right) = 0 \tag{7}$$

$$\lambda \frac{\partial T_{yy}}{\partial t} + T_{yy} - 2\mu \frac{\partial^2 u_y}{\partial t \partial y} = 0 \tag{8}$$

$$\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} + \epsilon p = 0 \tag{9}$$

where u_x and u_y are the displacements in the x and y directions, respectively, T_{xx}, T_{yy}, and T_{xy} are the principal stresses in the x and y direction and the shear in the xy plane, respectively. The two boundary conditions that would apply to the free surface of the evolving pattern, and the maximum height are as follows, respectively:

$$\int_0^t e^{\frac{-t-s}{\lambda}} \left(\frac{\partial^2 u_x}{\partial y \partial s} + \frac{\partial^2 u_y}{\partial x \partial s}\right) ds \tag{10}$$

$$\gamma \frac{\partial^2 u_y}{\partial t^2} - 2\mu \int_0^t e^{\frac{-t-s}{\lambda}} \frac{\partial^2 u_y}{\partial x \partial s} ds \tag{11}$$

$$u_y = \left[\frac{h}{H} - \frac{2hL}{bH}|x|\right] H\left(\frac{h}{H} - \frac{2hL}{bH}|x|\right) \tag{12}$$

The boundary conditions for the displacement in the x and y directions on the surface of the web are as follows, respectively

$$u_y = H - h + R_0 \left[\left(1 - \frac{H-h}{R_0}\right)\cos(\omega t) + \sqrt{1 - \left(1 - \frac{H-h}{R_0}\right)^2}\sin(\omega t) - 1\right] \tag{13}$$

$$u_x = 0 \tag{14}$$

The initial conditions for all of the variables, and their derivatives with respect to time, are zero.

3 METHODS

The goal is to utilize a PINN to solve for the variables u_x, u_y, p, T_{xx}, T_{xy}, and T_{yy}, which are presented in equations (4)-(9). Since there is more than one output, a standard PINN with only a single branch is unlikely to provide the accurate predictions for each variable, therefore it would be difficult for governing equations to be satisfied. To address the multiple outputs case, a multi-branched PINN is considered which is shown in Fig. 1. This multi-branched PINN network is implemented for the mold filling problem and tested in TensorFlow. The standard PINN did not converge to any useful results beyond the initial conditions and several methods for PINN improvements available in the literature were considered.

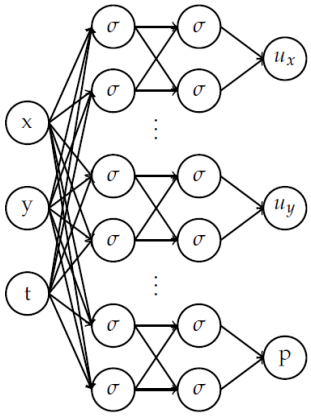

Fig. 1: Example of a multi-branched PINN, vertical dots represent extra branches for the variables not shown

The first specific method used was soft attention masking, which was applied to all of the collocation points, initial conditions, and periodic boundary condition points. The masking works by creating new weights that are assigned to every point of interest. These weights are then trained to increase the loss of the network, so that the places in both space and time that result in higher losses will have a higher impact on the adjustment of the network weights, thus helping to reduce the overall error [16]. The second method was mini-batching the training points. This method works by breaking down the space in which one wishes to train the PINN into smaller subsets. Previous work has shown that this method reduces the likelihood that the PINN will get trapped in a local minimum when compared to training the PINN with the full set [14]. In addition to just batching the spatial and temporal points, the same original set of points is kept constant, but each of the batches are randomly generated from these points until there are none left from the original set. This is done to prevent the chances of small, but minute cyclical changes that cause the network to undo the corrections from earlier in the batching process, thus preventing convergence on a solution. The last method is the time adaptive method of progressive time subset training, in which the time domain is divided into smaller subsets of equal size. For example, if the training time duration is divided into n sets, the PINN is first trained on only the first set (time interval, [0, t_1]) Once the solutions for all the equations converge on t_1, it is then trained on an extended time span of [0, t_2]. This process is repeated until the PINN has been trained on the full-time domain of n sets [14].

Despite all of the previous work on PINN improvements to solve "stiff" PDEs and other challenging fluid mechanics problems, the mold filling equations present a novel challenge, because PINNs for solving governing equations that involve the convolution integral have not been developed yet. The compounding error that occurs when evaluating the integral for each time period makes it challenging for traditional numerical methods to solve these governing equations. This compounding error will hinder the prediction over time, as the integral is being built upon samples with error which increases error in a cascading manner in subsequent sets. The reason PINNs may provide a significant improvement is that the network is iteratively tested across time non-sequentially, thus, potentially reducing this compound error, as the points are also tested for convergence thousands of time. Furthermore, by using the first-time adaptive approach alongside mini-batching and attention masking, the most troublesome points during each phase of the time step is to reduce the errors before moving onto further training. In addition, irrespective of the accuracy of the model and underlying governing equations, there will always be discrepancies in the real-world process because of noise and other parameter changes. Including these real-

world changes into the model, and subsequently into numerical methods, is generally difficult, whereas PINNs can be modified to include the error between solutions from the governing equations and real-world examples as one of the loss terms.

To generate the patterns for training the PINN, a custom R2RNIL machine was constructed. This machine consists of two segments: an experimental R2R platform containing unwind, rewind, and a NIL station containing gravure coating, mold and nip rollers, and UV lights for curing. The entire R2R system is controlled by an Allen-Bradley CompactLogix 5380 PLC, to maintain tension control (using load cells as feedback) and speed control by varying the speeds of the rewind and unwind roller. A picture of the entire setup and R2RNIL portion of the setup is shown in Figs. 2 and 3, respectively.

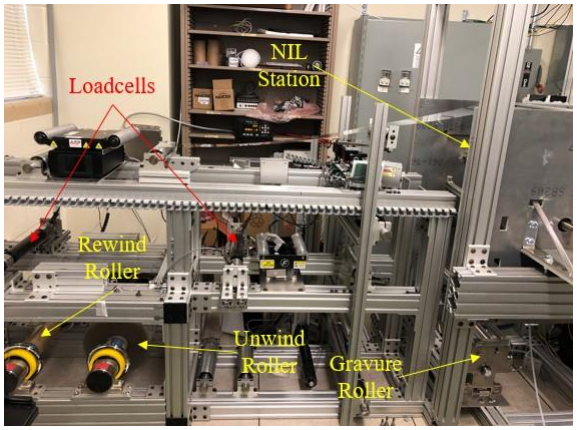

Fig. 2: Experimental R2R setup: material is released from the unwind roller, transported over a load cell, coated by the gravure roller before entering the NIL Station for patterning and curing, and then wound into a roll on to the rewind roller.

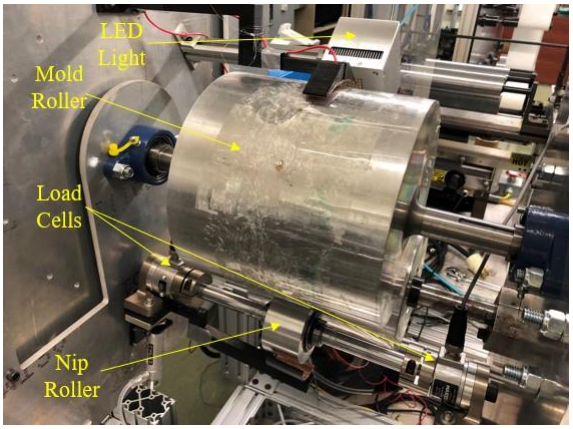

Fig. 3: High throughput NIL station: the nip roller is mounted on load cells on each end for force feedback, 6 ultraviolet (UV) lights cover the circumference of the mold roller for curing.

The mold roller used in these experiments is a 10-inch diameter steel roller with a wrap angle of 315°, giving a total curing distance of roughly 2.25 feet. The mold roller itself does not have any surface patterns on it, but it has patterned substrate film attached on the circumference of the mold roller. The patterned substrates which were used as mold sleeves were provided by Iscent Oy. The patterns are sinusoidal with a period of 1.2 μm, and a peak-to-peak variation of 100 nm as shown in Fig. 4. The substrate is first unwound and then passed over the first load cell, in order to gain feedback for tension control. From there it is sent to the gravure roller. The gravure roller is rotated in the opposite direction of web transport in order to deposit a more even coating of resin onto the web. Once the web is coated, it is transported to the nip region of the mold roller and nip roller. Once exiting the nip region, the resin coated web is exposed to UV lights that are spaced equally around the mold roller circumference in the region of wrap. During the curing process the resin is cured from the web to the top of the mold cavity. Once the web has traveled around the roller it is debonded from the roller, before winding on to the rewind roller.

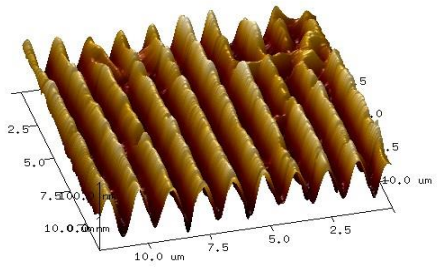

Fig. 4: 3D topological scan of the mold patterns provided by Iscent Oy (https://www.iscent.fi/)

Experiments were conducted at speeds of 3 and 7.6 MPM (10 and 25 FPM) and imprinted surface samples are then . These samples are then measured using an AFM scan, the results of which are used for feedback into the PINN.

4 RESULTS AND DISCUSSION

The setup was tested at both 3 and 7.6 MPM (10 and 25 FPM). The heights of the patterned AFM images in each line grating were averaged, and Fig. 5 shows the results for the experiment conducted at 3 MPM. It is evident that a majority of the sections are matching the amplitude and period of the original pattern, indicating a successful reproduction.

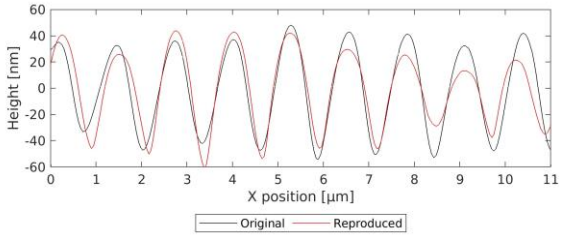

Fig. 5: Section averages of AFM scans (3 MPM samples)

These initial results are promising, as this indicates that the setup is capable of accurately reproducing patterns and the results align with both the numerical simulation and previous work.

Another important experiment is the production sample taken from speeds of 7.6 MPM with measurements shown in Fig. 6. Even at this higher speed of 7.6 MPM, the mold patterns are being replicated, with most of the defects likely due to damages from handling of the specimens post production. In addition to proving the speed capabilities of the imprinting setup, it also shows that the reproduced patterns are resilient enough to survive the entire transport process, as the prints shown were pulled from the bottom of the rewind roller.

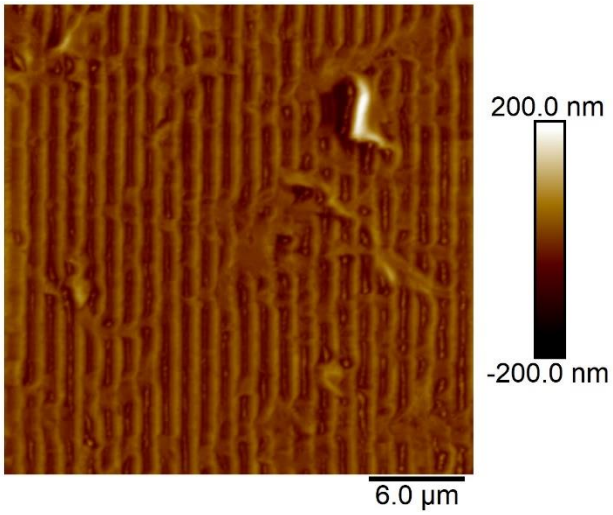

Fig. 6: Reproduction Prints, 7.6 MPM

These physical results serve a larger importance then to just show the capabilities of R2RNIL, they will also be used as feedback into the PINN so that the network can learn from the experimental observations as well as the governing equations, which would enable the ability to accurately predict how the molds would fill, thus allowing for real-time machine parameter optimization based on this model.

From the reproduced prints, pattern height measurements are recorded. A significant number of height measurements of the reproduced patterns can then be combined to create a depiction of the average height profile that is reproduced with respect to the *x* position. The same can be done with the original patterns in order to determine what the ideal height profile with respect to the x position is. The height profile from the original molds can then be used to replace the upper surface height limit, and an extra loss term can be used in the PINN to compare the predicted final height of the free surface to the real one. An illustration of this process is provided in Fig. 7.

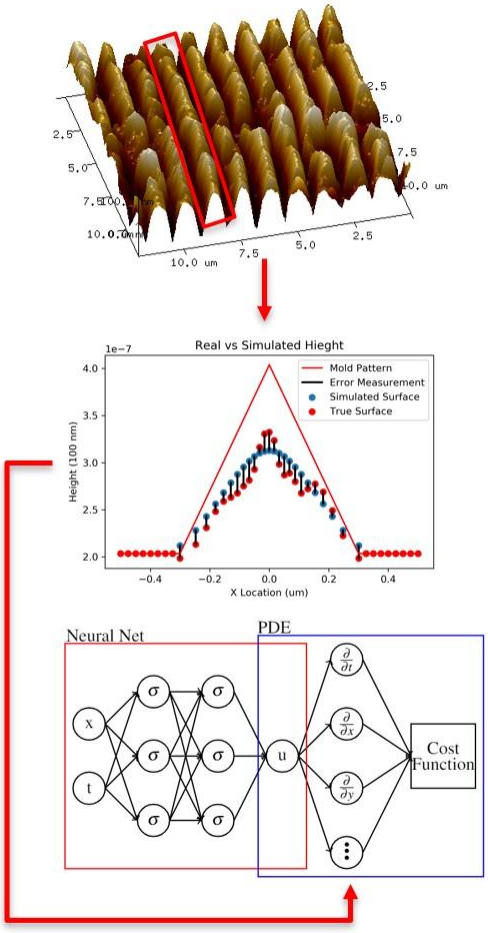

Fig. 7: The rows from the AFM scans will be isolated, and averaged, then compared to the PINNs prediction, which will be used as another loss term for training the network

The methods discussed for utilizing experimental feedback and measurements with PINN to predict mold filling can be applied to another, equally as important and hard to measure phase of R2RNIL, the phase transition portion of R2RNIL process. Like the mold filling portion of the process, there are currently no methods for measuring the parameters in real-

time, but there are however governing equations that have been derived to describe the process.

5 CONCLUSIONS

In this paper we have discussed the use of physics-informed neural networks for predicting mold filling in the R2RNIL process, in order to improve the quality of imprint patters and aid in real-time control of machine parameters.

Future work on this project will involve gathering more measurements from a wider range of machine running conditions, and material parameters, to aid in the PINNs ability to predict a large variation of potential setups and machine adjustments. For this, prints will be produced at speeds ranging from 5 to 60 MPM in 5 MPM intervals. This range was chosen due to previous research [18] having identified 30 MPM being the speed at which their specially designed resin failed to produce 100% fill. The experiments will be completed up to 60 MPM as to allow for a lower percentage of the mold to fill (>50%), which will allow for testing the PINNs ability to converge on a wide range of machine variations. Once the PINN for prediction of the mold filling is completed and feedback has been implemented, it will be compared to the current linearized numerical model in terms of accuracy and speed. The long-term goals are to implement the fully trained PINN as a method for accurately predicting real-time production accuracy, and then identify the optimal changes to the machine or material parameters to allow for proper mold filling.

ACKNOWLEDGMENTS

This work was supported in part by the NSF grant number 1635636. We want to thank Dr. Raimo Korhonen at Iscent Oy (https://www.iscent.fi/) for graciously providing the nanoimprint mold films utilized in the R2R NIL experimentation.

REFERENCES

[1] J. Park, K. Shin, and C. Lee, "Roll-to-roll coating technology and its applications: a review," *International Journal of Precision Engineering and Manufacturing*, vol. 17, no. 4, pp. 537–550, 2016.

[2] K. Fukuda and T. Someya, "Recent progress in the development of printed thin-film transistors and circuits with highresolution printing technology," *Advanced Materials*, vol. 29, no. 25, p. 1602736, 2017.

[3] A. Gusain, A. Thankappan, and S. Thomas, "Roll-to-roll printing of polymer and perovskite solar cells: compatible materials and processes," *Journal of Materials Science*, pp. 1– 53, 2020.

[4] P. Maury, D. Turkenburg, N. Stroeks, P. Giesen, I. Barbu, E. Meinders, A. Van Bremen, N. Iosad, R. Van der Werf, and H. Onvlee, "Roll-to-roll uv imprint lithography for flexible electronics," *Microelectronic Engineering*, vol. 88, no. 8, pp. 2052–2055, 2011.

[5] S. Y. Chou, P. R. Krauss, and P. J. Renstrom, "Nanoimprint lithography," *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena*, vol. 14, no. 6, pp. 4129–4133, 1996.

[6] J. Shao, X. Chen, X. Li, H. Tian, C. Wang, and B. Lu, "Nanoimprint lithography for the manufacturing of flexible electronics," *Science China Technological Sciences*, vol. 62, no. 2, pp. 175–198, 2019.

[7] T. Lee, C. Lee, D. K. Oh, T. Badloe, J. G. Ok, and J. Rho, "Scalable and high-throughput top-down manufacturing of optical metasurfaces," *Sensors*, vol. 20, no. 15, 2020.

[8] A. Seshadri, P. R. Pagilla, and J. E. Lynch, "Modeling print registration in roll-to-roll printing presses," *Journal of dynamic systems, measurement, and control*, vol. 135, no. 3, p. 031016, 2013.

[9] J. J. Dumond and H. Yee Low, "Recent developments and design challenges in continuous roller micro-and nanoimprinting," *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena*, vol. 30, no. 1, p. 010801, 2012.

[10] J. Gomez-Constante, P. Pagilla, and K. Rajagopal, "A thermomechanical description of the mold filling process in roll-toroll nanoimprinting lithography," *Applications in Engineering Science*, vol. 1, p. 100001, 2020.

[11] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations," *arXiv preprint arXiv:1711.10561*, 2017.

[12] X. Jin, S. Cai, H. Li, and G. E. Karniadakis, "Nsfnets (navierstokes flow nets): Physics-informed neural networks for the incompressible navier-stokes equations," *Journal of Computational Physics*, vol. 426, p. 109951, 2021.

[13] Z. Xiang, W. Peng, X. Zheng, X. Zhao, and W. Yao, "Selfadaptive loss balanced physics-informed neural networks for the incompressible navier-stokes equations," *arXiv preprint arXiv:2104.06217*, 2021.

[14] C. L. Wight and J. Zhao, "Solving allen-cahn and cahnhilliard equations using the adaptive physics informed neural networks," *arXiv preprint arXiv:2007.04542*, 2020.

[15] S. Wang, X. Yu, and P. Perdikaris, "When and why pinns fail to train: A neural tangent kernel perspective," *arXiv preprint arXiv:2007.14527*, 2020.

[16] L. McClenny and U. Braga-Neto, "Self-adaptive physicsinformed neural networks using a soft attention mechanism," *arXiv preprint arXiv:2009.04544*, 2020.

[17] S. A. Niaki, E. Haghighat, T. Campbell, A. Poursartip, and R. Vaziri, "Physics-informed neural network for modelling the thermochemical curing process of composite-tool systems during manufacture," *Computer Methods in Applied Mechanics and Engineering*, vol. 384, p. 113959, 2021.

[18] M. W. Thesen, M. Rumler, F. Schlachter, S. Grützner, C. Moormann, M. Rommel, D. Nees, S. Ruttloff, S. Pfirrmann, M. Vogler, *et al.*, "Enabling large area and high throughput roll-to-roll nil by novel inkjetable and photo-curable nil resists," in *Alternative Lithographic Technologies VI*, vol. 9049, p. 90490H, International Society for Optics and Photonics, 2014.