

A Dual-tree Complex Wavelet Transform Simulation Model for the Real-time SPP

Rahul Gupta, Nieqing Cao, Sang Won Yoon, Yu Jin, and Daehan Won

Abstract—This paper presents a dynamic simulation model for the stencil printing process (SPP) in surface mount technology (SMT) assembly lines, focusing on accurately replicating the real-time stencil printing while allowing adjustments to printer settings. The model offers a time and cost-effective alternative to the experiments and a reliable testing environment for researchers and technologists investigating advanced algorithms and strategic methodologies in SMT printing. Furthermore, it provides valuable insights by capturing the intricate relationships between input parameters and printing outcomes that can be used for machine parameter optimization. SPP is a critical procedure of SMT and plays a crucial role in determining the quality and yield of PCB production. The SPP is influenced by various factors, including printing speed, printing pressure, separation speed, cleaning cycle, and printing direction. However, an additional challenge arises from uncontrollable environmental noise that affects the printing quality, leading to uneven solder paste application and machine precision that brings randomness to the results. Recognizing the need to mitigate the effects of this environmental noise and enhance the accuracy of the simulator, the proposed simulation model incorporates a dual-tree complex wavelet transform (DTCWT) algorithm. This algorithm effectively captures the dynamic characteristics of the process. By modeling the uncontrollable noise, the simulation provides a realistic representation of the printing environment and its effect on quality.

Index Terms—Stencil Printing Process, Surface Mount Technology, Signal Processing, Wavelet, Simulation, Smart Manufacturing

I. INTRODUCTION

SURFACE mount technology (SMT) is a method for placing components on printed circuit boards (PCBs) and is favored especially because of the miniaturization trend in electronics [1]. A typical SMT-based PCB assembly process consists of three sections: solder paste printing (SPP), mounting, and solder reflow. SPP is a crucial step in SMT production lines that significantly affect PCB assembly success rates and accounts for more than 60% of observed defects. Optimizing the stencil printing parameters can improve the solder paste volume transfer efficiency (TE) [2]. The SPP's quality can be represented by the printed solder paste information detected by a solder paste inspection (SPI) machine after printing. Using the insights gained from the comprehensive design of experiments, several key factors significantly influenced the quality of the printing process. SPP is one of the critical procedures in the SMT, having around 52-71% of PCB fine pitch defects [3]. These pivotal factors encompass printing

Rahul Gupta, Nieqing Cao, Sang Won Yoon, Yu Jin, and Daehan Won are at the Department of Systems Science and Industrial Engineering, State University of New York at Binghamton, Binghamton, NY 13902 USA (e-mail: dhwon@binghamton.edu).

speed (PS), printing pressure (PP), separation speed (SS), printing direction, and cleaning cycle. Their profound effect extends particularly to critical aspects such as solder paste volumes and offsets, marking their significance in shaping the overall printing quality and accuracy. Offsets refer to the positional deviations between the intended placement of solder paste on a printed circuit board and its actual placement along the horizontal (X) and vertical (Y) axes, respectively.

Multiple offline models aim to optimize initial printing parameters to enhance the solder paste volume TE for PCB production [4, 5]. However, the models are less than perfectly accurate because of the inherent variability within production lines, changes in material properties, machine wear, and environmental factors. That variability undermines the uniformity of printing outcomes, diminishing the efficacy of static parameters, prompting the exploration of online learning models[6, 7] to adapt to real-time changes. Moreover, other challenges are in evaluating these models with limited manufacturing data is challenging, necessitating extensive numerical testing. Similarly, research models for stencil cleaning prediction and abnormal detection [1] lack adequate evaluation. To date, limited exploration has occurred regarding simulated dynamic SPP environments for testing solution modules. Addressing this gap, this study introduces the framework shown in Figure 1 that illustrates our dynamic system. Propose converting SPI data into 3D time sequence images as PCBs are printed sequentially, enabling real-time visualization of spatial volume TE. Additionally, the sequential images facilitate the comprehension of real-time production noise in SPP and allow the analysis of PCB inspection results using a wavelet-based model that approximates signals within the frequency domain [8]. In the proposed framework, a wavelet-based noise segmentation model is first used on lab experimental SPI data to obtain the noise characteristics of SPP, which is defined as the random variations in the data that are not caused by the parameter settings. Then, after adding the random generation of segmented noise to the filtered data, a dynamic SPP system is formulated from the prediction model on the printer settings for each pad size. The framework leverages data-driven models (i.e., machine learning) to predict the printing results for various parameters accurately, and it can generalize the simulation for different printer machines and aperture shapes, which are trained and estimated by the environment. The approach holds the potential for surpassing traditional lab tests in terms of efficiency and accuracy. The following are the main contributions of this research:

- 1) Through a DTCWT model, the SPP noise can be extracted, and sufficient simulation data can be generated

to build a dynamic SPP environment.

- 2) Based on the dynamic system, data-driven models (i.e., machine learning) are used to predict the printing results for various parameters accurately.
- 3) Furthermore, the proposed dynamic simulation model promises to provide sufficient data for evaluating and enhancing further SPP-related algorithmic studies.

The rest of this paper is organized as follows: Section 2 presents a review of related works about simulation methodologies; Section 3 introduces the methodology in this research; experimental settings and results are illustrated in Section 4; a conclusion and potential future work of this simulation framework are discussed in Section 5.

II. LITERATURE REVIEW

Extensive research has been dedicated to modern manufacturing assembly lines for PCBs, emphasizing the need for cutting-edge methodologies [9, 10, 11]. This includes investigations into cleaning profiles, solder volume behavior, and SPP optimization [1, 4, 6, 12]. These efforts focus on, improving the dynamic SPP, the crucial initial step in SMT procedures that requires a bridge between printing performance and parameters through dynamic simulation. To date, various simulation models are widely applied in other SMT assembly line processes to generate data and test real-life scenarios from simulation outcomes. For instance, a numerical model based on the shape of the rolling solder paste and the printing squeegee has been employed to study viscosity effects during SPP. The model's results, encompassing shear rate distribution, velocity field, and pressure profile, suggest the superiority of non-Newtonian fluid properties over Newtonian ones. However, that approach might overlook some influential factors in SPP, limiting a comprehensive understanding [13]. Computational Fluid Dynamics (CFD) is another effective avenue for simulations considering physical environments and fluid flow. Techniques such as the Finite Element Method (FEM) are used to solve CFD's mathematical equations. One prominent method, Volume of Fluid (VOF), captures multi-phase flow for different aperture sizes, presenting potential areas for enhancing printing quality [14].

Furthermore, a study explores varied scenarios in SPP using simulations with changing factors, such as printing speeds. Nonetheless, this study neglects broader SPP factors and dynamics, like noise and parameter settings' collective impact [15]. In a different study, a 3D simulation model developed via ANSYS software focused on squeegee angle optimization and print quality for SAC305 solder paste. Although it recommends an optimal angle, the investigation ignored other process parameters and their effect on stencil printing defects for process enhancement [16]. Moreover, a Monte-Carlo-based simulation model was developed to focus on positioning accuracy and indentation depth of components in solder paste, with more than one million boards simulated. While the model addresses material, geometrical, and process influences, it simplifies interactions and assumes linearity, potentially missing nuanced non-linear behaviors. The Monte Carlo method abstracts randomness but may not fully capture

real-world variability, affecting precision [17]. Another simulation model covers SMT processes from printing to solder reflow, emphasizing visual appearance and solder joint shape. However, it overlooks solder paste volume and pad offset information [18]. Three simulation models are employed, starting with printing and progressing to placement and reflow. A Deep Neural Network (DNN) generates solder joint images for automatic optical inspection. The limitation across studies is the inadequate treatment of environmental noise and the interplay of printing parameters in SPP simulation. In contrast, this study introduces a dynamic simulation model with the DTCWT algorithm to address dynamic noise and explores printer parameter configurations comprehensively. This approach provides a more precise emulation of real-world SPP conditions, enhancing efficiency in PCB production printing processes.

In manufacturing, time series data is processed using techniques such as wavelet transform (WT) and empirical mode decomposition (EMD) for simulation models [19]. EMD, known for decomposing signals into intrinsic mode functions (IMFs), is combined with wavelet packet decomposition (WPD) for simulating induction motors, and post-processing improves signal quality, validated by FEM analysis under various loads [20]. Prism Signal Processing (PSR) is another innovative method that generates output time-series signals based on input time series, proving versatile for applications such as wireless acceleration sensors and Industry 4.0 manufacturing systems [21, 22]. For assessing solder paste component accuracy and indentation depth, a Monte-Carlo-based simulation model is developed, considering material, geometry, and process factors. However, that model simplifies interactions and assumes linearity, potentially missing nuanced non-linear behaviors. Power quality disturbances are simulated using Fourier transform (FT), Short Time Fourier transform (STFT), continuous wavelet transform (CWT), and discrete wavelet transform (DWT), showing promise in monitoring power signals [23]. Dual-Tree Complex Wavelet Transform (DTCWT) is proposed as a noise reduction technique, enhancing DWT. DTCWT employs two real DWTs to represent the real and imaginary parts, meeting perfect reconstruction (PR) conditions. Unlike traditional DWT, DTCWT addresses issues of shift-invariance and directional selectivity, providing more precise results for 2D and multiple-dimension signals. The shift-variant problem of DWT is resolved by averaging the output reconstruction signals from the two trees [24, 25]. In this study, DTCWT is applied as a noise reduction and filtering method, with the effectiveness of noise reduction depending on thresholding techniques like VisuShrink, SureShrink, and BayesShrink, known for their efficacy in reducing noise in images [26].

This research uses machine learning models to comprehensively analyze printer behavior across various printer settings and aperture sizes. Decision Tree algorithms are frequently employed because of their robustness, efficiency, and ability to generate interpretable models with satisfactory accuracy in various domains. Like other machine learning methods, decision tree algorithms rely on hyperparameters that directly affect model performance. Given the multitude of possible

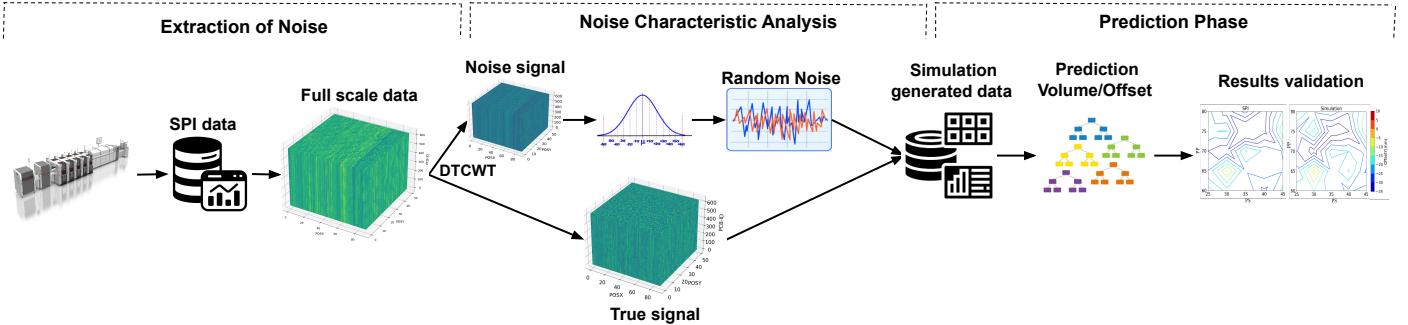


Fig. 1: Three-Phase Dynamic SPP Simulation Framework: Illustrates (i) Real production noise extraction using wavelet transform, (ii) Noise characteristics analysis using the detailed examination of statistical distribution, and (iii) Prediction phase for generalizing the simulator for different printer settings.

hyperparameter values, optimization techniques are commonly employed to identify optimal solutions that yield classifiers with superior predictive performance. [27]. Determining the appropriate stopping criterion during the tree-building process poses a significant challenge for decision tree algorithms. Many of the existing methods for constructing decision trees tend to produce complex models that overfit the training data. Overfitted trees not only exhibit limited predictive performance on new and unseen data, but they also present difficulties in terms of interpretability [28]. This lack of interpretability poses a significant obstacle to the practical application and widespread adoption of decision tree models [29]. Notably, the algorithm stands out by its inherent design that eliminates the need for hyperparameter optimization or regularization techniques. As a result, the training time for decision tree construction is substantially reduced [30].

III. METHODOLOGY

The dynamic environment simulation framework is structured into three primary phases: SPI data decomposition through DTCWT, volume generation involving noise and true data, and volume prediction considering new printing parameters. An overview of the framework's architecture is presented in Figure 1.

A. Dual-Tree Complex Wavelet Transform for SPI Data Decomposition

The DTCWT is employed to analyze and extract noise from time series data obtained using sequentially printed PCBs. This approach involves decomposition and reconstruction stages, as illustrated in Figure 2. During decomposition, the input SPI data $X(t)$ is processed using two parallel dual trees (DT) for signal processing. The upper tree handles the real part, while the lower tree manages the imaginary part, including the processing of Complex Wavelet Transform (CWT), i.e., DT-CWT or DTCWT [31]. Each tree constitutes a filter-bank tree with low-pass (h_0) and high-pass (h_1) filters for the real part in the upper tree and low-pass (g_0) and high-pass (g_1) filters for the imaginary part in the lower tree [32]. The DTCWT excels at processing complex signals, making it suitable for dynamic data like SPP-obtained time series data. Here, 3-D implementation involves separable filtering across rows,

columns, and slices, resulting in an octal-tree system with these bands LLL, HLL, LHL, HHL, LLH, HLH, LHH, and HHH. Those bands, which are similar to the conventional 3-D DWT as shown in Figure 3, are derived where "L" represents a low-pass filter capturing the coarse or average information of the signal, and "H" denotes a high-pass filter isolating the detailed or high-frequency components. Within the octal-tree system, each subband generates eight real coefficients (one from each tree) at every spatial location. These coefficients can be combined through straightforward arithmetic operations to create four directional subbands of complex coefficients [32]. These four subbands correspond to distinct quadrants within the 3-D spectral half-space, as depicted in Figure 3. This asymmetry aligns all filters at the same orthonormal set, enabling independent interpolation (shifting) of each subband of coefficients without affecting others [31]. Consequently, the shift variance is avoided, enhancing the reliability and accuracy of the DTCWT. For each level of the DT-CWT, excluding the LLL band, a total of $(8 - 1) * 4 = 28$ directional subbands are generated. These directional subbands arise from the eight-band structure. Specifically, let's consider the HHH band at a particular DT-CWT level (other bands follow a similar process). The mathematical representation of the 3-D wavelet for the first-quadrant HHH subband can be summarized as:

$$\begin{aligned} \psi_1(x, y, z) &= [\psi_a(x) + j\psi_b(x)] [\psi_a(y) + j\psi_b(y)] [\psi_a(z) + j\psi_b(z)] \\ &= [\psi_a(x)\psi_a(y)\psi_a(z) - \psi_b(x)\psi_b(y)\psi_b(z) \\ &\quad - \psi_a(x)\psi_b(y)\psi_b(z) - \psi_b(x)\psi_a(y)\psi_b(z)] \\ &\quad + j[\psi_a(x)\psi_a(y)\psi_b(z) - \psi_b(x)\psi_b(y)\psi_b(z)] \\ &\quad + \psi_a(x)\psi_b(y)\psi_a(z) + \psi_b(x)\psi_a(y)\psi_a(z)] \end{aligned} \quad (1)$$

In the equations above, subscripts a and b denote distinct trees representing the real and imaginary parts, as shown in Figure 2. The high-pass filters $\psi_a(i)$ and $\psi_b(i)$ are applied to the dimensions i (e.g., x, y, z), where x, y , and z represent rows, columns, and slices, respectively. It is noteworthy that the 1-D DTCWT filters effectively suppress the negative half of the frequency spectrum. However, real 3-D datasets encompass independent frequency components in all quadrants of the upper-half frequency space, as illustrated in Figure 3. The wavelets for the second, third, and fourth quadrants can be

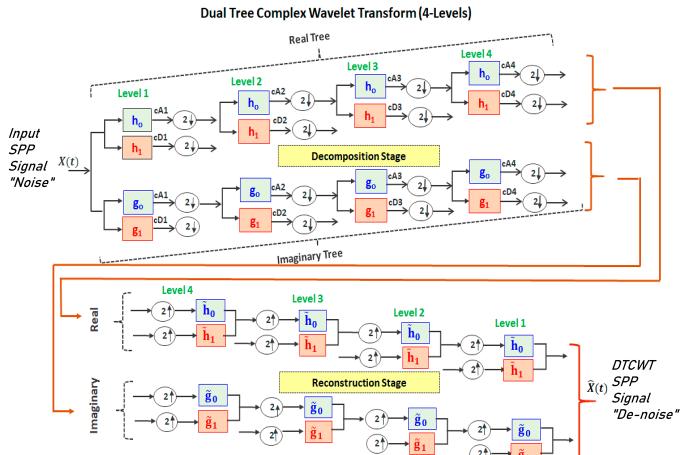


Fig. 2: Illustration of one-dimensional DTCWT showcasing the filter banks used in the decomposition and reconstruction phases [31].

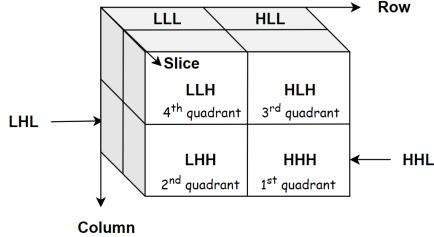


Fig. 3: Structure of 3-D DTCWT with eight bands, each cube representing subbands of DWTS. Filtering sequence follows rows, columns, and slices order, e.g., 'HLL' indicates high-pass filtering on rows, and low-pass on columns and slices.

derived using analogous equations.

The SPI obtained data is transformed into time sequence format, forming a 3-dimensional (3D) representation, with the x and y dimensions representing the pad positions ($PosX$ and $PosY$), and the z dimension representing the time dimension (PCB-id). By viewing the SPI data as a 3D dataset in the time domain, the wavelet transformation represents the signal as a series of approximations from the frequency domain [25]. Consequently, the structured time and spatial PCB data adopting DTCWT for noise segmentation promise high extraction and reconstruction accuracy, preserving the original pad distributions and pertaining scale invariance. For 3D data, DTCWT requires three dual trees in the "height," "width," & "time" directions. The application of the DTCWT in the simulation model enhances noise segmentation and extraction accuracy, providing a robust foundation for modeling the dynamic SPP environment and optimizing the printing process parameters under varying noise conditions.

B. Dynamic Environment Simulation

The numerical simulation process involves three pivotal stages, encompassing data integration, generation, and aggregation. It starts by processing the comprehensive experimental SPP data using DTCWT, an advanced signal analysis

technique that helps us separate high-frequency noise signals from the original data, revealing the true signal in the time domain. The true signal is converted into a continuous format within the temporal-spatial domain using uniformly segmented grids. This discretized true signal dataset aligns with the PCB board's pad layout, representing paste volume for individual pads across temporal sequences.

Preliminary noise characteristics are analyzed to ensure filtering and reconstruction accuracy within acceptable tolerances. The noise is studied based on filtering validation and calculating a residual by subtracting reconstructed noise and true signal data from the original dataset. The reconstructed signal, recognized as the true SPI result, is associated with printer settings and aggregated in terms of averages and standard deviations. Segmented noise accounts for random system variations because of unforeseen perturbations, enabling real-time simulation. To gain quantitative insights into noise characteristics, its distribution is studied, and random data is generated for simulation. Noise data analysis is conducted separately for various aperture shapes and in an aggregated form. Noise distribution is analyzed in dimensions of $PosX$, $PosY$, and *time*, segmented into different groups by area ratio (AR) and aspect ratio (ASR).

To create a dynamic simulation model, noise is simulated using random data generation based on its statistical distribution, which is then combined with the true volume of TE data. This dynamic model provides printing information for different settings within the solution space. Simulating dynamic printing outcomes under varied noise conditions involves generating random noise based on observed distribution characteristics for similar aperture shape groups. This meticulous process deepens the understanding of noise's intricate effects across diverse printing scenarios. The model encompasses true signal data, Pad ID, and AR & ASR values as features for predicting printing results. Generating random noise data based on normal distribution parameters contributes to the noise points per pad, enhancing the model's accuracy. By integrating this noise with true signal data, the simulated *volume/OffsetX/OffsetY* is obtained, providing comprehensive insights into various printing parameters. Additionally, *volume TE*, *OffsetX*, and *OffsetY* data are integrated for *Volume Avg* and *Volume Std* models, as well as *OffsetX/Y. Avg* and *OffsetX/Y. Std* models. These data points serve as training and validation inputs for real-time predictor models.

C. Prediction under Different Printing Parameters

The Dynamic Environment Simulation Framework applies data-driven prediction models to estimate the printing results for various printer parameter settings irrespective of the parameters included for creating the simulator, which helps to make it more generalized. The data is carefully filtered based on cleaning type and printing direction, encompassing critical factors such as printing pressure, printing speed, separation speed, cleaning age, area ratio, and aspect ratio. These filtered datasets serve as the training data for the prediction models, enabling accurate forecasts of printing performance under diverse scenarios. The framework's inherent flexibility allows

it to adapt to different noise conditions, providing valuable insights for improving the yield of printed circuit board manufacturing and the printing process. Here, the Decision Tree Regression (DTR) model predicts printing outcomes with good accuracy. To ensure optimal performance of the model, the hyperparameter tuning and cross-validation fine-tune its parameters and validate its predictive capabilities against unseen data. This approach empowers us to achieve precise predictions regarding the printing performance under a range of parameter settings, enhancing the robustness and reliability of the simulation framework. The simulation model is implemented using Python, providing a robust and flexible framework for dynamic environment simulation.

IV. EXPERIMENTAL RESULTS

A. Experiment Setup

To acquire noise analysis training data, an extensive experiment was conducted, encompassing a range of levels for print speed (PS), print pressure (PP), and squeeze speed (SS). PS levels were set at 25, 30, 35, 40, and 45 mm/s, PP levels at 60, 65, 70, 75, and 80 N, and SS levels at 3.0, 3.5, 4.0, 5.5, and 7.0 mm/s. That resulted in 125 total settings for the forward and backward printing directions. Before each printing setting combination, stencil cleaning was carried out before the initial board, followed by a 5-board cleaning cycle. Each printed circuit board (PCB) comprised 3 distinct AR groups, amounting to a total of 4,500 pads. The printing data were recorded by an SPI machine, capturing positional information (*PosX* and *PosY*) for each pad. Two key performance indices were used to evaluate the prediction model's performance: the root-mean-squared error (RMSE) and the mean absolute error (MAE). These metrics allowed for a quantitative assessment of the model's accuracy in predicting the printing results. The DTR model is considered a baseline model to accurately forecast the printing results, considering a wide range of parameter settings and diverse noise conditions.

B. 3D DTCWT-based Filtering Results and Analysis

The 3D DTCWT algorithm has been employed to address random noise issues in SPI results, as illustrated in Figure 4. Here, *PosX* and *PosY* represent the shapes of PCBs from a single printing iteration, and the third dimension corresponds to the time sequence of printed PCBs. The raw data, because of noise and outliers, may introduce bias into the prediction model during training. To preserve the authenticity of the printing information, applied a 2-level decomposition to the original dataset for noise segmentation, as visualized in Figure 5. In Figure 5(a), the LLL subband represents the signal, while Fig. 5(b)-(h) contain stable signal portions. To filter out unstable signals within the high-frequency subbands, apply the threshold of $\pi = 11.0$ for volume and $\pi = 7.52$ for the offset. The resultant signal, along with the LLL band, constitutes the true SPI data. Employing a reverse 3D DTCWT transformation converted the frequency domain signal into the time and spatial domain, leading to smoother SPI volume TE data as demonstrated in Figure 6. This transformation reduced the range of volume TE by eliminating outliers. Figure

7 highlights the noise obtained through noise reconstruction based on high-frequency subband data. Additionally, analyze the noise distribution to better comprehend its characteristics. By segmenting the noise data according to aperture shapes in dimensions of *PosX*, *PosY*, and *time* examined the noise level distribution in a single dimension, disregarding temporal and spatial information. The noise represents redundant volume TE not associated with the settings, where positive values signify redundancy, while negative values indicate insufficiency.

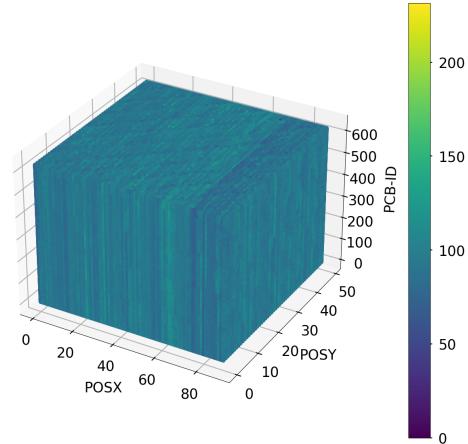


Fig. 4: The original SPI data collected from the SMT lab, which is visualized from the PCB *PosX*, *PosY*, and the time dimensions, to represent the volume TE in color codes.

C. Simulator Validation

Comprehensive validation tests were conducted to assess the reliability and accuracy of the proposed dynamic simulator for the SPP. Firstly, statistical tests were performed to validate the equality of two-sample variances and their means between the real and simulated data. The *F*-test was used to validate the equality of two-sample variances, ensuring that the variance of the simulated data is consistent with the variance of the real data. Additionally, a *t*-test was employed to validate the equality of two-sample means, ensuring that the mean values of the simulated data align closely with the mean values of the real data. The validation process for the *volume*, *OffsetX*, and *OffsetY* is illustrated in Figure 8. Here, we can see that the simulator has a very close distribution of features as compared with real data. Furthermore, to visualize and interpret the simulation results, 2D contour plots were generated, illustrating the relationship between the average volumes obtained from SPI and the simulator, in relation to varying printing speeds and pressures. The accuracy of these plots was quantified using the Wasserstein distance, yielding a value of 0.78. This indicates a reasonable level of agreement between the simulated and observed data. Similar plotting was performed for *volume STD*, *OffsetX*, and *OffsetY*, with Wasserstein distance values of 0.55, 0.10, and 0.13, respectively. These contour plots provided a comprehensive understanding for assessing process similarity. The contour plots for *Volume AVG*, *Volume STD*, *OffsetX*, and *OffsetY* are shown in Fig. 9, 10, 11, and 12, respectively. The simulator validation process underscored the robustness and fidelity of the proposed dynamic simulation framework. The statistical

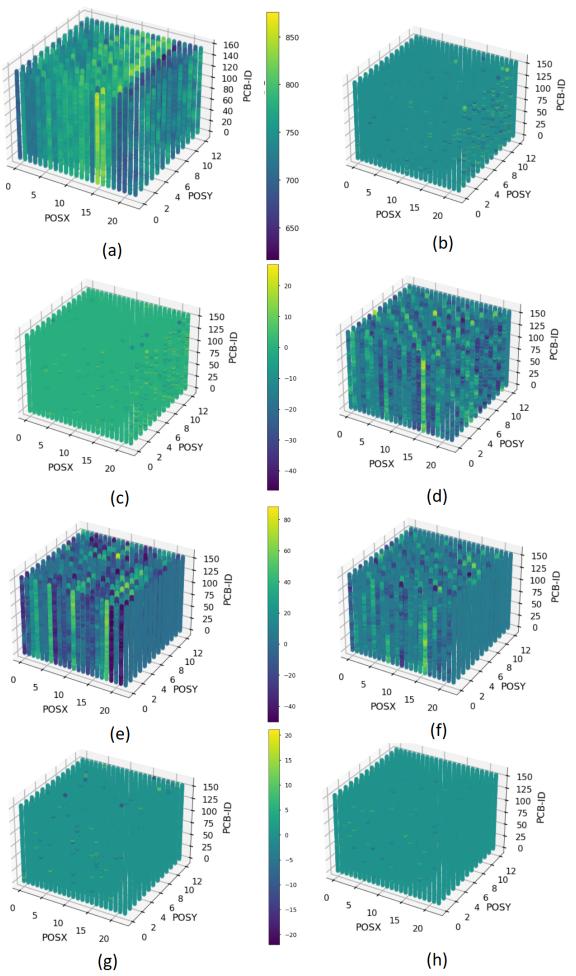


Fig. 5: The 8 subbands of data generated from the 3D DTCWT-based frequency domain decomposition process. (a) LLL, (b) LLH, (c) LHH, (d) LHL, (e) HLL, (f) HLH, (g) HHL, (h) HHH.

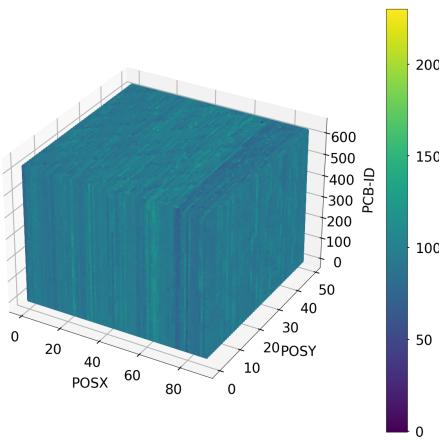


Fig. 6: The true SPI result reconstructed from frequency domain filtering, visualized from the PCB PosX, PosY, and the time.

tests and visualization techniques employed in the validation further validated the reliability of the simulator's predictions. A significant part of this validation involved a comparison

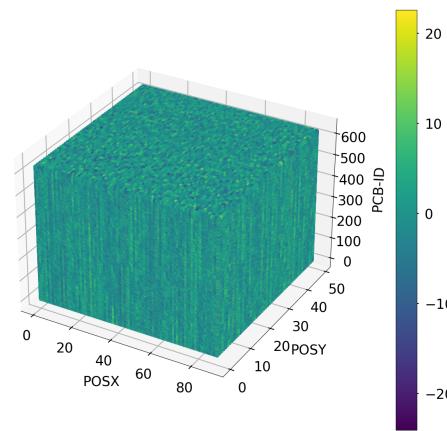


Fig. 7: The reconstructed noise from frequency domain segmentation, visualized from the PCB PosX, PosY, and the time.

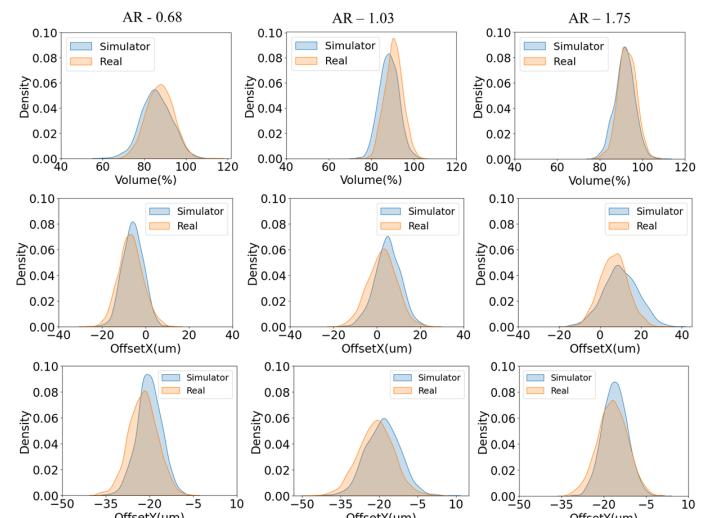


Fig. 8: Validation of simulator and actual SPI volume, offsetX for different aperture shapes.

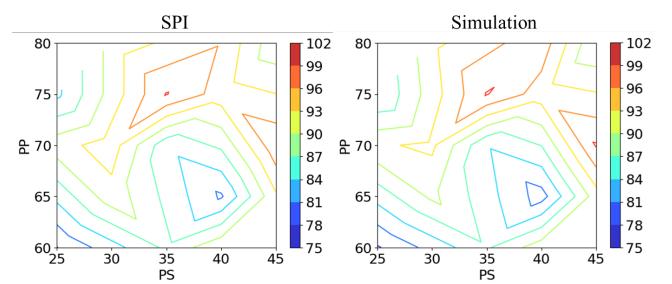


Fig. 9: Contour plot illustrating the comparison between real and simulated Vol. AVG for different printing settings.

with the Bayesian statistical simulator. When assessing the prediction accuracy for volume TE, our simulator consistently displayed superior performance. Detailed analyses indicated that our simulator's algorithms and methodologies provided more accurate and consistent predictions compared to the Bayesian model. This comparison emphasizes the reliability and robustness of our simulator. Performance evaluations were conducted using data from 30 boards, the results of which are delineated in Table I displaying the RMSE value.

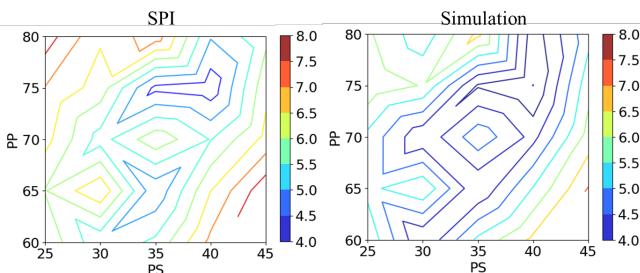


Fig. 10: Contour plot illustrating the comparison between simulated and real Vol. STD for different printing settings.

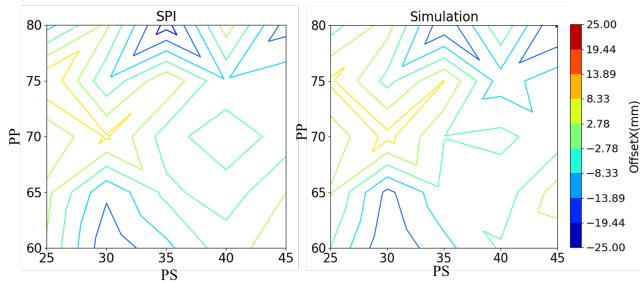


Fig. 11: Contour plot illustrating the comparison between simulated and real OffsetX for different printing settings.

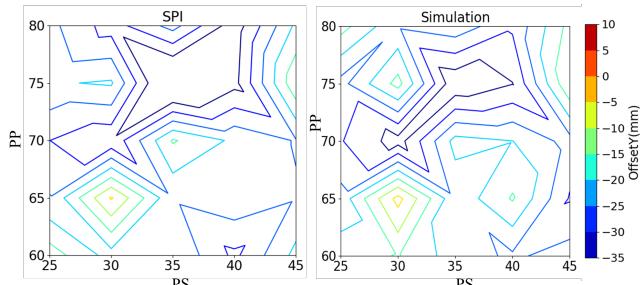


Fig. 12: Contour plot illustrating the comparison between simulated and real OffsetY for different printing settings

D. Evaluation of Prediction Results

Using supervised classification improves the generalizability of the simulator. Machine learning incorporates critical input features such as printing speed, pressure, separation speed, cleaning age, and geometric attributes (AR and ASR). Dedicated models for different printing outcomes (*Volume*, *OffsetX*, and *OffsetY*) efficiently capture the relationship between process variables and printing results. The Decision Tree Regressor stands out as the most accurate model for predicting printing outcomes. To enhance the model's performance, a comprehensive hyperparameter tuning process was conducted, involving critical parameters such as “*max_depth*,” “*min_samples_split*,” “*max_features*,” “*min_weight_fraction_leaf*,” “*max_leaf_nodes*,” and “*ccp_alpha*”. The tuning process used cross-validation

TABLE II: Result Performance Table

Features	Statistics	RMSE	MAE
Volume	Avg.	5.05	4.05
	Std.	0.64	0.53
OffsetX	Avg.	6.93	5.5
	Std.	0.76	0.64
OffsetY	Avg.	7.42	6.13
	Std.	1.89	1.13

to ensure the model's robustness and prevent overfitting. The implementation of the DTR model and hyperparameter tuning was executed using the GridSearchCV function. After evaluating various hyperparameter combinations, the optimal set of hyperparameters was identified, and the final DTR model was trained using these optimized values. The effectiveness of the DTR model in predicting printing outcomes was demonstrated for each different AR and ASR combination. The simulation results, including *Volume AVG* (mean values), *Volume STD* (standard deviations), *OffsetX/Y AVG*, and *OffsetX/Y STD* exhibited high accuracy, illustrating the model's capability to capture the complex relationships between input parameters and printing performance. Those results can provide valuable insights into process optimization and defect reduction in SMT-based printed circuit board manufacturing. In the study, the simulator's performance was rigorously assessed using data from 30 boards, with the findings presented in Table II. The table delineates the predictions for solder paste *volume*, *OffsetX*, and *OffsetY*. Evaluation metrics, namely the RMSE and MAE, were employed to gauge the accuracy of these predictions. The results underscore that the simulator's predictions are increasingly consistent with the real printer, especially in terms of *volume*, *OffsetX*, and *OffsetY* accuracy. Furthermore, the enhanced model demonstrated a 36% improvement in *Volume AVG* accuracy on RMSE and a 62% rise in *Volume STD*, marking a significant advancement over the original simulator.

V. CONCLUSION AND FUTURE WORK

The objective of this research is to successfully develop a dynamic simulation model aimed at generating sufficient data for the study of SPP. To account for the inherent variability observed in real assembly printing outcomes within identical settings, introduced randomness into the system by simulating SPP noise through a DTCWT-based noise filtering approach. Through frequency domain filtering and reconstruction isolate the true volume directly associated with setting variations as the signal. Concurrently investigated the statistical distribution characteristics of the random noise, characterized by high-frequency variations. By simulating noise generation based on its distribution properties and integrating it with the true volume TE data, our dynamic simulation model was constructed, enabling the provision of printing information for various settings within the solution space. This system accurately mirrors the characteristics observed in real-world stencil printing processes. In future research, the current simulation model presented in this study will be primarily limited to specific equipment conditions. As the behavior of PCBs can vary significantly based on factors such as design, size, and

TABLE I: Simulation Comparison Result

Features	Statistical Simulator	Proposed Simulator
Volume Avg.	5.84	1.86
Volume Std.	7.76	4.86

material composition, there is a need for further validation encompassing a diverse range of PCB types. Moreover, it is essential to extend the evaluation of these algorithms to encompass various initial printer settings, including parameters like snap-off distance, stencil thickness, and material composition. This broader scope of validation will provide a more comprehensive understanding of algorithm performance across a spectrum of real-world scenarios.

REFERENCES

- [1] Alelaumi, Shrouq, et al. "Residue buildup predictive modeling for stencil cleaning profile decision-making using recurrent neural network," *Robotics and Computer-Integrated Manufacturing* 68, (2021): 102041.
- [2] N. Khader and S. W. Yoon, "Adaptive optimal control of stencil printing process using reinforcement learning," *Robotics and Computer-Integrated Manufacturing* 71 (2021): 102132.
- [3] Pan, Jianbiao, et al. "Critical variables of solder paste stencil printing for micro-BGA and fine-pitch QFP," *IEEE Transactions on Electronics Packaging Manufacturing* 27.2 (2004): 125-132.
- [4] N. Khader and S. W. Yoon, "Stencil Printing Process Optimization to Control Solder Paste Volume Transfer Efficiency," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 8, no. 9, Sep. 2018.
- [5] Lu, Hongya, et al. "A guided evolutionary search approach for real-time stencil printing optimization," *IEEE Transactions on Components, Packaging and Manufacturing Technology* 11.2 (2020): 333-341.
- [6] H. Lu, H. Wang, S. W. Yoon, and D. Won, "Real-Time Stencil Printing Optimization Using a Hybrid Multi-Layer Online Sequential Extreme Learning and Evolutionary Search Approach," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 9, no. 12, Dec. 2019.
- [7] Lu, Hongya, et al. "Dynamic predictive modeling of solder paste volume with real time memory update in a stencil printing process," *Procedia Manufacturing* 38 (2019): 108-116.
- [8] H. Wang, H. Lu, S. M. Alelaumi, and S. W. Yoon, "A wavelet-based multi-dimensional temporal recurrent neural network for stencil printing performance prediction," *Robotics and Computer-Integrated Manufacturing*, vol. 71, Oct. 2021.
- [9] T. He, D. Li, and S. W. Yoon, "A multi-phase planning heuristic for a dual-delivery SMT placement machine optimization," *Robotics and Computer-Integrated Manufacturing*, vol. 47, Oct. 2017.
- [10] K. P. Ellis, J. E. Kobza, and F. J. Vittes, "Development of a placement time estimator function for a turret style surface mount placement machine," *Robotics and Computer-Integrated Manufacturing*, vol. 18, no. 3, Jun. 2002.
- [11] I. I. Kochegarov, A. Tynda, and N. Goryachev, "The method of obtaining the thermal model of the printing circuit assembly," *2016 XIX IEEE International Conference on Soft Computing and Measurements (SCM)*, St. Petersburg, Russia, 2016.
- [12] T.-N. Tsai, "Improving the fine-pitch stencil printing capability using the Taguchi method and Taguchi fuzzy-based model," *Robotics and Computer-Integrated Manufacturing*, vol. 27, no. 4, Aug. 2011.
- [13] O. Krammer, T. Al-Ma'iteh, and P. Martinek, "Investigating the Effect of Viscosity Models on the Stencil Printing by Numerical Modelling," *2019 42nd International Spring Seminar on Electronics Technology (ISSE)*, Wrocław, Poland, 2019.
- [14] M. S. Rusdi et al., "Multiphase Flow in Solder Paste Stencil Printing Process using CFD approach," *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences*, vol. 46, no. 1, 2018, Accessed: Jul. 08, 2023.
- [15] O. Krammer, T. I. Al-Ma'iteh, B. Illes, D. Bušek, and K. Dušek, "Numerical investigation on the effect of solder paste rheological behaviour and printing speed on stencil printing," *Soldering Surface Mount Technology*, vol. 32, no. 4, Jun. 2020.
- [16] M. S. Rusdi et al., "Numerical Investigation on the Effect of Squeegee Angle during Stencil Printing Process," vol. 1082, Aug. 2018.
- [17] H. Wohlrabe, H. Fritzsche and S. Lüngen, "Simulation based optimization of a SMD-assembling process," *2016 39th International Spring Seminar on Electronics Technology (ISSE)*, Pilsen, Czech Republic, 2016.
- [18] J. Nau, J. Richter, D. Streiterd and M. Kirchhoff, "Simulating the Printed Circuit Board Assembly Process for Image Generation," *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, Madrid, Spain, 2020.
- [19] Shi, Junbing, et al. "Extraction method of weak underwater acoustic signal based on the combination of wavelet transform and empirical mode decomposition," *International Journal of Metrology and Quality Engineering* 12 (2021): 7.
- [20] K. Tian, T. Zhang, Y. Ai, and W. Zhang, "Induction Motors Dynamic Eccentricity Fault Diagnosis Based on the Combined Use of WPD and EMD-Simulation Study," *Applied Sciences*, vol. 8, no. 10, Sep. 2018.
- [21] M. Henry and V. V. Sinitis, "Prism signal processing for machine condition monitoring I: Design and simulation," *2018 IEEE Industrial Cyber-Physical Systems (ICPS)*, St. Petersburg, Russia, 2018.
- [22] M. Henry and V. V. Sinitis, "Prism signal processing for machine condition monitoring II: Exp data and fault detection," *2018 IEEE Industrial Cyber-Physical Systems (ICPS)*, St. Petersburg, Russia, 2018.
- [23] F. M. Arrabal-Campos, F. G. Montoya, R. Baños, J. Martínez-Lao and A. Alcayde, "Simulation of power quality disturbances through the wavelet transform," *2018 18th International Conference on Harmonics and Quality of Power (ICHQP)*, Ljubljana, Slovenia, 2018.
- [24] I. W. Selesnick, R. G. Baraniuk and N. C. Kingsbury, "The dual-tree complex wavelet transform," in *IEEE Signal Processing Magazine*, vol. 22, no. 6, Nov. 2005.
- [25] A. Kushwaha, A. Khare, O. Prakash, Jong-In Song and Moongu Jeon, "3D medical image fusion using dual tree complex wavelet transform," *2015 International Conference on Control, Automation and Information Sciences (ICCAIS)*, Changshu, China, 2015.
- [26] Jana, Debalina, and Kaushik Sinha. "Wavelet thresholding for image noise removal," *International Journal on Recent and Innovation Trends in Computing and Communication* 2.6, (2014): 1400-1405.
- [27] R. G. Mantovani, T. Horváth, R. Cerri, J. Vanschoren and A. C. P. L. F. d. Carvalho, "Hyper-Parameter Tuning of a Decision Tree Induction Algorithm," *2016 5th Brazilian Conference on Intelligent Systems (BRACIS), Recife, Brazil*, 2016.
- [28] Cao, Nieqing, Daehan Won, and Sang Won Yoon. "EXPLAINS: Explainable Anomaly Prediction for SMT Solder Joints Using SPI Data," *International Conference on Flexible Automation and Intelligent Manufacturing*. Cham: Springer Nature Switzerland, 2023.
- [29] Schaaf, Nina, Marco Huber, and Johannes Maucher. "Enhancing decision tree based interpretation of deep neural networks through 11-orthogonal regularization," *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*. IEEE, 2019.
- [30] R. García Leiva, A. Fernández Anta, V. Mancuso and P. Casari, "A Novel Hyperparameter-Free Approach to Decision Tree Construction That Avoids Overfitting by Design," in *IEEE Access*, vol. 7, 2019.
- [31] Kingsbury, Nick G. "The dual-tree complex wavelet transform: a new technique for shift invariance and directional filters." *IEEE digital signal processing workshop*. Vol. 86. Citeseer, 1998.
- [32] Al-Naami, Bassam, et al. "Assessment of Dual-Tree Complex Wavelet Transform to Improve SNR in Collaboration with Neuro-Fuzzy System for Heart-Sound Identification," *Electronics* 11.6, (2022): 938.
- [33] I. W. Selesnick, R. G. Baraniuk, and N. G. Kingsbury, "The dual-tree complex wavelet transform," *IEEE Signal Process. Mag.*, vol. 22, no.6, Nov. 2005.

Rahul Gupta is currently pursuing a master's degree at the Department of Systems Science and Industrial Engineering, the State University of New York at Binghamton, Binghamton, NY, USA. Her research interests include data predictive modeling and systems optimization in manufacturing.

Nieqing Cao is currently pursuing a Ph.D. degree at the Department of Systems Science and Industrial Engineering, the State University of New York at Binghamton, Binghamton, NY, USA. Her research interests include large-scale data predictive modeling and systems optimization in manufacturing.

Daehan Won is an Associate Professor at the Department of Systems Science and Industrial Engineering, the State University of New York at Binghamton, Binghamton, NY, USA. His current research interests include large-scale mathematical programming, data analytics/mining in healthcare, and designing smart manufacturing systems to advance Industry 4.0.

Sang Won Yoon is currently a Professor at the Department of Systems Science and Industrial Engineering, State University of New York at Binghamton, Binghamton, NY, USA. His current research interests include distributed decision-making, coordination protocol design, collaborative control theory, and large-scale data analytics and predictive modeling.

Yu (Chelsea) Jin is currently an Assistant Professor at the Department of Systems Science and Industrial Engineering, State University of New York at Binghamton, Binghamton, NY, USA. Her current research interests focus on sensing and analytics, optimization, and simulation for advanced manufacturing and service applications.